UNIVERSIDADE DA BEIRA INTERIOR
Engineering

Biometric Recognition in Unconstrained Environments

Gil Melfe Mateus Santos

Thesis for obtaining the degree of Doctor of Philosophy in
Computer Science and Engineering
(3^{rd} Cycle Studies)

Supervisor: Prof. Doutor Hugo Pedro Martins Carriço Proença

Covilhã, October 2014

Thesis prepared at *IT - Instituto de Telecomunicações*, within the Pattern and Image Analysis - Covilhã Group, and submitted to University of Beira Interior for defense in a public examination session.

Work financed by the Portuguese *Fundação para a Ciência e Tecnologia* through grant contract SFRH/BD/80182/2011 under the program QREN - POPH - Type 4.1 – Advanced Training, co-funded by the European Social Fund and by national funds from the Portuguese *Ministério da Educação e Ciência*.

Dedicatory

To my family.
For all the love and support.

Acknowledgments

This doctoral research program would not be possible without the help of many people, to whom I am sincerely grateful.

First and foremost, I would like to thank my family for being by my side, with encouraging words and supporting attitude during these last years. In particular my parents, my sister and my fiancée, for all the love and support. I should also not forgot to mention the encouragement from my grand-mothers, aunts, uncles, cousins and mother-in-law. And, of course, the cheering from my stepchildren. I am grateful for their understanding during all the time I was absent due to this research work.

Second I would like to thank all my friends, who stood by my side all these years, always cheering me up and giving me the support I needed to carry on. I am also grateful to my colleagues at SOCIA Lab, NMCG and Release, who shared some concerns and a lot of coffee on the endless days and nights of hard work.

I must also mention my gratitude though the people at Optics Center, in particular Professor Paulo Fiadeiro, that not only made the logistics for the experiments possible, but was always very supportive. I could not forget to mention all the people that contributed as volunteers in the construction of the working datasets used during my doctoral research.

Last, but not least, I would like to express my gratitude to my supervisor, Prof. Hugo Proença, for the guidance and trust, specially in the most difficult moments. I can now conclude that our discussions over the last years strongly contributed to the development of this thesis.

People think that computer science is the art of geniuses but the actual reality is the opposite, just many people doing things that build on each-other, like a wall of mini stones.

Donald Knuth

List of Publications

Articles included in the thesis resulting from this doctoral research program

Journal Articles

1. **Fusing Color and Shape Descriptors in the Recognition of Degraded Iris Images at Visible Wavelengths**
   Hugo Proença and Gil Santos
   Computer Vision and Image Understanding, 116, 168-178, 2012.
   DOI: 10.1016/j.cviu.2011.10.008

2. **On Periocular Biometrics: A Comprehensive Outline**
   Gil Santos and Hugo Proença
   Artificial Intelligence Review, subbmitted for consideration, 2013.

3. **A Fusion Approach to Unconstrained Iris Recognition**
   Gil Santos and Edmundo Hoyle
   Pattern Recognition Letters, 33(8), 984-990, June 2012.
   DOI: 10.1016/j.patrec.2011.08.17

4. **Fusing Iris and Periocular Information for Cross-sensor Recognition**
   Gil Santos, Emanuel Grancho, Marco V. Bernardo and Paulo T. Fiadeiro
   Pattern Recognition Letters, 2014.
   DOI: 10.1016/j.patrec.2014.09.012

5. **BioHDD: A Dataset for Studying Biometric Identification of Heavily Degraded Data**
   Gil Santos, Paulo T. Fiadeiro and Hugo Proença
   IET Biometrics, 2014.
   DOI: 10.1049/iet-bmt.2014.0045

International Conferences

1. **Iris Recognition: Preliminary Assessment about the Discriminating Capacity of Visible Wavelength Data**
   Gil Santos, Marco Bernardo, Paulo Fiadeiro and Hugo Proença
   6th IEEE Workshop on Multimedia Information Processing and Retrieval - MIPR '10, December 13-15, 2010, Taiwan.
   ISBN: 978-0-7695-4217-1

2. **Periocular Biometrics: An Emerging Technology for Unconstrained Scenarios**
   Gil Santos and Hugo Proença
   IEEE Symposium on Computational Intelligence in Biometrics and Identity Management - CIBIM '13, April 16-19, 2013, Singapore.
   ISBN: 978-1-4673-5879-8/13

3. **A Robust Eye-Corner Detection Method for *Real-World* Data**
   Gil Santos and Hugo Proença
   IEEE International Joint Conference on Biometrics – IJCB '11, October 11-13, 2011, Washington DC, USA.
   DOI: 10.1109/ijcb.2011.6117596

4. **Segmenting the Periocular Region using a Hierarchical Graphical Model Fed by Texture / Shape Information and Geometrical Constraints**
   Hugo Proença, João C. Neves and Gil Santos
   IEEE International Joint Conference on Biometrics – IJCB '14, September 19 - October 02, 2014, Clearwater, Florida, USA.

5. **A Dual-Step Approach to Head Landmark Detection In The Wild**
   Gil Santos and Hugo Proença
   submitted to the 8^{th} *IAPR International Conference on Biometrics – ICB-2015*, May 19-22, 2015, Phuket, Thailand.

6. ***Quis-Campi*: Extending In The Wild Biometric Recognition to Surveillance Environments**
   Gil Santos, João C. Neves, Sílvio Filipe, Emanuel Grancho, Silvio Barra and Fabio Narducci
   submitted to the 8^{th} *IAPR International Conference on Biometrics – ICB-2015*, May 19-22, 2015, Phuket, Thailand.

Other publications resulting from this doctoral program not included in the thesis

Book Chapters

1. **Using Ocular Data for Unconstrained Biometric Recognition**
   Hugo Proença, Gil Santos, João C. Neves
   Face Recognition in Adverse Conditions, Maria de Marsico, Michele Nappi, Massimo Tistarelli (Eds.), IGI Global, 2013.
   DOI: 10.4018/978-1-4666-5966-7

Journal Articles

1. **Robust Periocular Recognition by Fusing Sparse Representations of Color and Geometry Information**
   Juan C. Moreno, V. B. Surya Prasath, Gil Santos and Hugo Proença
   Journal of Signal Processing Systems, subbmitted for consideration, 2014.

International Conferences

1. **Iris Recognition: Analysing the Distribution of the Iriscodes Concordant Bits**
   Gil Santos and Hugo Proença
   IEEE Proceedings of the 3rd International Congress on Image and Signal Processing - CISP 2010, October 16-18, 2010, Yantai, China.

2. **Facial Expressions: Discriminability of Facial Regions and Relationship to Biometrics Recognition**
   Elisa Barroso, Gil Santos and Hugo Proença
   IEEE Symposium on Computational Intelligence in Biometrics and Identity Management - CIBIM '13, April 16-19, 2013, Singapore.
   ISBN: 978-1-4673-5879-8/13

Resumo

Ao ser humano está associado, pela sua natureza, um conjunto de características físicas e comportamentais que o caracterizam. O estudo dessas características permitiu o desenvolvimento de um considerável número de sistemas e aplicações – sistemas biométricos.

A utilização de sistemas biométricos tem vindo a aumentar ao longo dos últimos anos, principalmente na área da segurança: autenticação, controlo de acesso, identificação criminal, etc. Sendo um sector de elevada exigência, é natural que se dê maior destaque às características biométricas que permitam atingir uma maior discriminação entre os sujeitos, ao mesmo tempo que são menos propícias a falsificação. Contudo, estas restrições acarretam um impacto significativo tanto na usabilidade do sistema como na sua flexibilidade, necessitando de um elevado grau de cooperação por parte do utilizador. É neste contexto que a íris se apresenta como a característica biométrica por excelência. Os sistemas de reconhecimento biométrico que utilizam a íris como característica principal baseiam-se essencialmente na abordagem pioneira proposta por John Daugman. Esta demonstrou ser uma excelente opção para cenários de reconhecimento cooperativo em que as imagens possam ser adquiridas no infravermelho.

Contudo, nem sempre a cooperação por parte dos indivíduos é expectável e, nesses casos, sistemas com elevadas restrições na aquisição deixam de ser viáveis. Linhas de investigação mais recentes tentam então contornar esta problemática seguindo duas possíveis abordagens: adaptando os métodos existentes aos novos cenários e desafios, ou encontrando novas características biométricas que melhores se adaptem a esta realidade. Nesta última abordagem, a região periocular (i.e., o olho e a região circundante) é uma das características mais promissoras: aproxima-se do método de reconhecimento usado naturalmente e de forma espontânea pelo ser humano; tem uma localização privilegiada em relação à íris, facilitando a aquisição simultânea de ambas as características; e tem, tal como corroborado pela literatura, um conjunto de características promissoras que podem ser usadas para efeitos de reconhecimento.

O objetivo principal destes trabalhos de doutoramento é então adaptar ou desenvolver um sistema de reconhecimento biométrico, especialmente adequado para ambientes não-controlados (*in the wild*). Esse sistema, pelos seus requisitos e especificidades, deverá usar a região periocular como característica preferencial de reconhecimento, dado que esta permite uma maior flexibilidade e facilidade na aquisição em condições particularmente adversas, por forma a minimizar as restrições de funcionamento. Os indivíduos poderão ser reconhecidos a distâncias superiores, em movimento, com condições de iluminação irregulares, e usando informação adquirida no espectro de luz visível.

Por forma a atingir este objetivo, uma série de etapas intermédias foi estabelecida. Começou-se por estudar a íris enquanto sinal biométrico, prestando especial atenção à vertente não-cooperativa e ao funcionamento no comprimento de onda visível. Este estudo englobou também os efeitos da luz visível no reconhecimento, tendo sido levada a cabo uma análise da reflectância da íris em função do comprimento de onda de diferentes iluminantes. Tendo em conta que a íris não se apresenta como sinal ideal ao reconhecimento *in the wild*, foram estudadas características biométricas emergentes, prestando especial atenção à região periocular. Da literatura analisada, os métodos mais relevantes foram implementados e testados contra um mesmo con-

junto de dados. Finalmente, várias contribuições foram propostas e aceites pela comunidade científica, com aplicação em diferentes ambientes não-controlados, tendo sido a última a conceptualização de um sistema biométrico capaz de trabalhar nas condições desafiantes a que nos propúnhamos.

Palavras-chave

análise de erro; análise de textura; aprendizagem automática; aprendizagem supervisionada; avaliação da performance; bases de dados de imagens; biometria; biometria multi-modal; biometria não-cooperativa; biometria ocular; comprimento de luz visível; dados degradados; decisão baseada na fusão de resultados; descritores globais; descritores locais; extracção de características; identificação de pessoas; privacidade; reconhecimento periocular; reconhecimento biométrico da íris; região de interesse; robustez do reconhecimento; selecção de características; variação de pose; vídeo-vigilância.

Resumo alargado

Introdução

Esta tese tem por objectivo abordar a problemática associada ao reconhecimento biométrico em ambientes não-controlados. Compreendendo que este é um objetivo desafiante, optámos por seguir uma abordagem por etapas. Desta forma, começámos por analisámos a viabilidade do reconhecimento biométrico em cenários não-cooperativos tendo a íris como principal sinal biométrico. De seguida procurámos alternativas mais viáveis, tendo estabelecido a região periocular como a característica biométrica mais promissora para os cenários alvo e, como tal, conduzimos um estudo mais aprofundado sobre a mesma. Ao longo dos trabalhos desenvolvidos foram propostas várias contribuições científicas e, por fim, é apresentado um sistema de reconhecimento.

Enquadramento da Tese

Na sociedade atual tem vindo a registar-se uma crescente preocupação com a segurança, quer individual quer global, impulsionando a pesquisa científica na área da biometria. Esta tendência de crescimento é visível tanto nos métodos e técnicas desenvolvidos, como em aplicações, estando presente hoje em dia num vasto número de dispositivos. O utilizador comum pode facilmente encontrar modos de autenticação baseados em biometria em telemóveis e computadores pessoais, mas é nos sistemas de controlo de acesso a um mais alto nível (e.g., postos de controlo fronteiriços) que se foca a maioria dos desenvolvimentos. Neste último grupo de aplicações, e devido aos elevados padrões de segurança, a maior parte da pesquisa é centrada em características biométricas que permitam uma maior discriminação dos indivíduos, ao mesmo tempo que minimizam a possibilidade de contrafação, escolha esta que pode comprometer a usabilidade e flexibilidade do sistema.

O reconhecimento biométrico é na sua essência um problema de reconhecimento de padrões, na medida em que o objetivo é reconhecer um determinado indivíduo com base num conjunto de características físicas ou comportamentais. Entende-se por reconhecimento de padrões a disciplina científica responsável pelo processamento dos dados "crus" recolhidos de um determinado objeto, e obtenção da informação (i.e., padrões) necessária para que este possa ser atribuído a uma determinada classe. Tendo sido uma disciplina meramente teórica até aos anos 60, a evolução dos sistemas computacionais trouxe consigo um vasto leque de aplicações, tornando-se numa área extremamente ativa no domínio da aprendizagem automática e em visão computacional, mineração e dados e extração de conhecimento em particular [1]. A título de exemplo, a Figura 1.1 ilustra duas classes de um problema típico de reconhecimento de padrões: tendo uma representação de um dado objeto, determinar a que classe ele pertence. Neste caso particular o objetivo seria analisar uma imagem médica contendo uma região devidamente delimitada, e decidir se esta correspondia a uma lesão benigna (classe A) ou era na realidade um cancro (classe B). Para efeitos ilustrativos, assumamos que estamos na posse de uma base-de-dados contendo imagens deste tipo, e que dentro da mesma classe estas são similares. Para que possamos classificar uma nova instância de dados é necessário que as características que tornam as classes distintas sejam identificadas. Neste exemplo ilustrativo, e como é possível perceber por inspeção visual, a intensidade dos pixeis difere significativamente de uma classe

para a outra. Assim sendo, usando métricas tão simples como a média (P) e o desvio padrão (Q) dos valores de intensidades, conseguimos obter uma ótima separabilidade entre as classes – Figura 1.2. Havendo este tipo de distinção, um *threshold* de classificação pode ser inequivocamente definido (Figura 1.2, linha contínua). Ao *adquirir dados* de uma nova observação, procede-se à *extração de características* tal como realizada anteriormente (P e Q). Traçando a nova observação no gráfico (*) podemos então *classificar* a mesma como pertencendo à classe A, uma vez que se encontra acima do limiar de classificação. São estes os passos gerais na resolução de um problema de classificação de padrões.

Dos sinais biométricos existentes, os sistemas implementados comercialmente dão preferência aos que conduzam a taxas de erro mínimas, pelo que a íris se apresenta como uma característica biométrica primordial. Contudo, a captura de imagens da íris está condicionada por rígidos protocolos de aquisição, o que se deve essencialmente ao eu tamanho reduzido e propriedades móveis, não sendo portanto um sinal biométrico ideal para um sistema de funcionamento subreptício em ambientes não-controlados. Em oposição aos sistemas biométricos clássicos que requerem um elevado nível de cooperação por parte dos utilizadores, em aplicações *reais* sob típicos cenários de vídeo-vigilância nem sempre tal é expectável. Contrariamente a crenças populares, não existe ainda um sistema capaz de levar a cabo o reconhecimento biométrico em condições tão adversas. Neste contexto a região periocular apresenta-se como um sinal biométrico emergente, havendo um crescente número de publicações sobre o seu potencial de reconhecimento, tanto de forma autónoma, como em fusão com outras características biométricas. Neste último cenário, a íris assume-se como um candidato particularmente interessante, visto que devido à sua localização ambas podem ser adquiridas simultaneamente recorrendo apenas a um único sensor.

Os desafios à comunidade científica apresentam-se então sob duas formas: 1) por um lado estudam-se alternativas que permitam relaxar os protocolos de aquisição, adaptando os sistemas existentes para o funcionamento não-cooperativo; 2) e por outro, explora-se a utilização de novas características biométricas que melhor satisfaçam as atuais exigências. Estes trabalhos doutorais debruçam-se sobre os dois eixos, na medida em que avaliam a usabilidade de caraterísticas biométricas tradicionais (e.g., íris) adaptando-as aos requisitos do funcionamento não-cooperativo, e combinando-as com sinais biométricos emergentes (i.e., região periocular). O nosso objetivo principal é propor um sistema de reconhecimento biométrico adaptado ao funcionamento em cenários não-cooperativos, capaz de lidar com a aquisição de indivíduos a maiores distâncias, em movimento, sob condições de iluminação variáveis e adquirindo a informação no comprimento de onda visível. A motivação para estes trabalhos de doutoramento converge ainda com outra tendência de investigação, na medida em que combina o reconhecimento biométrico com cenários normalmente associados ao funcionamento de sistemas de vídeo-vigilância, tendo em vista o desenvolvimento de soluções *híbridas* capazes de identificar utilizadores *alvo* em cenários não-controlados.

Descrição do Problema e Objetivos de Investigação

De uma forma geral, esta tese foca-se no problema do reconhecimento biométrico em ambientes não-controlados. Assim sendo, o objetivo central destes trabalhos de doutoramento é adaptar ou desenvolver um sistema de reconhecimento biométrico capaz de lidar com este tipo de cenários. Tendo por base a pesquisa desenvolvida, a região periocular assume-se como carac-

terística central preferencial, devido à sua flexibilidade e facilidade de aquisição nas condições que o sistema exige.

Por forma a atingir o objetivo proposto foi estabelecido um conjunto de metas intermédias para melhor estruturação do trabalho que seria necessário levar a cabo:

1. Estudar a íris enquanto sinal biométrico, dando particular ênfase aos métodos desenvolvidos para ambientes não-cooperativos que funcionem no comprimento de onda visível. Este estudo deve focar-se em métodos de reconhecimento que visem a redução das restrições de aquisição, ou que de outro modo sejam capazes de lidar com informação especialmente degradada.

2. Uma vez que a informação adquirida no comprimento de onda visível introduz novos fatores de ruído com um impacto considerável no reconhecimento biométrico da íris, é necessário analisar a sua reflectância ao ser iluminada por diferentes iluminantes, em diferentes comprimentos de onda. As propriedades discriminativas da íris devem ser analisadas sobre três eixos: o tipo de iluminante presente na cena; o nível de luminância desse mesmo iluminante; e o nível de pigmentação da íris.

3. Estudar sinais biométricos emergentes, com especial destaque aos que minimizem os condicionantes associados a ambientes não-controlados (e.g., região periocular). A região periocular uma característica biométrica apropriada por três motivos: 1) é uma região utilizada diária e intrinsecamente pelo ser humano ao fazer o reconhecimento; 2) possui uma localização privilegiada em relação à íris, permitindo a aquisição simultânea destes dois sinais biométricos com o mesmo sensor; 3) possui um vasto leque de características passíveis de serem exploradas, tal como corroborado pela literatura.

4. Implementar e avaliar o desempenho dos métodos mais relevantes analisados no tópico anterior, avaliação esta que deve ser levada a cabo sobre o mesmo conjunto de dados. Pretende-se ainda estabelecer uma nova base-de-dados com informação capturada simultaneamente nos comprimentos de onda do infravermelho e luz visível, que embora adquirida em condições cooperativas introduza um conjunto de ruídos que simulem os que são encontrados em ambientes não-controlados.

5. Por fim, a maioria dos sistemas biométricos identifica (ou verifica a identidade) positivamente um determinado indivíduo de entre uma galeria de utilizadores previamente adicionados ao sistema, tendo por base uma medida de semelhança. Contudo, em muitas aplicações *in the wild*, tais sistemas fariam mais sentido se operassem numa perspetiva de reconhecimento negativo: garantir com determinada confiança que um indivíduo do qual se desconhece a identidade não pertence a um conjunto de "pessoas de interesse", ao invés de o tentar identificar. Nessa base, o nosso último objetivo intermédio seria o estudo do estado da arte relativo ao paradigma de reconhecimento negativo.

De entre os objetivos a que nos propusemos apenas a possibilidade de estudar e levar a cabo o reconhecimento negativo não foi totalmente atingido, sendo apresentado como uma linha de trabalho futuro. Compreendemos contudo as vantagens práticas de um sistema com este tipo de funcionamento, em especial se o sistema se encontrar a funcionar de forma sub-reptícia, tal como estabelecido ao longo do nosso trabalho [2]. Ainda assim, optámos por priorizar a conceção de novos sistemas, igualmente desafiantes e aplicados a diferentes cenários (e.g.,

dispositivos móveis [3]).

Por último, o sistema biométrico proposto no final destes trabalhos doutorais, foi estruturado de forma a poder ser facilmente implementado, validado e otimizado.

Principais Contribuições

Segue uma breve descrição das contribuições resultantes do trabalho de pesquisa desenvolvido no âmbito deste programa doutoral, tal como ilustrado na Figura 1.3.

A primeira contribuição consiste numa avaliação preliminar sobre a capacidade discriminativa da íris quando adquirida no comprimento de onda visível, percebendo e quantificando as condições que permitem efetuar o reconhecimento biométrico com um determinado grau de confiança. Este estudo está descrito no capítulo 3, que consiste num artigo publicado nos proceedings do 6° *IEEE Workshop em Multimedia Information Processing and Retrieval* [4].

A segunda contribuição propõe um novo método de reconhecimento, baseado em técnicas substancialmente diferentes das usadas tradicionalmente. A análise o nível de correlação entre o resultado do método proposto e outras técnicas do estado-da-arte sugere que a fusão de ambas as estratégias de reconhecimento contribui para um aumento significativo da performance. Assim sendo, é um passo positivo em direção ao desenvolvimento de tipos de reconhecimento mais ambiciosos. Este estudo está descrito no capítulo 4, que constitui o artigo publicado no 116° volume da revista *Computer Vision and Image Understanding* [5].

A terceira contribuição oferece uma visão comparativa dos trabalhos de pesquisa mais relevantes no âmbito do reconhecimento periocular, sumariando os métodos desenvolvidos e enumerando os problemas identificados. Neste estudo consiste o capítulo 5, artigo publicado nos *proceedings* do *IEEE Symposium on Computational Intelligence in Biometrics and Identity Management – CIBIM 2013* [6].

A quarta contribuição fornece uma visão em maior detalhe sobre os métodos mais relevantes do reconhecimento periocular, com uma análise aprofundada das técnicas sub-adjacentes e uma análise comparativa dos resultados desses mesmos métodos contra uma base-de-dados comum. Começa com uma visão introdutória dos métodos e sistemas biométricos tradicionais, anatomia da região periocular e dificuldades existentes, e conclui com observações sobre os principais fatores de degradação e possíveis direções de pesquisa. Este estudo está descrito no capítulo 6, que consiste num artigo submetido para a revista *Artificial Intelligence Review*.

A quinta contribuição consiste num novo método de deteção dos cantos do olho, capaz de lidar com dados degradados, dando especial ênfase a aplicações em condições e ambientes não-controlados. As nossas experiências mostram que o método proposto supera os existentes na literatura tanto em imagens sem ruído como naquelas com degradação (desfoque, rotação e variação significativa de escala). Este estudo está descrito no capítulo 7, que consiste no artigo publicado nos *proceedings* da *IEEE International Conference on Biometrics – IJCB 2011* [7].

A sexta contribuição apresenta um novo método de segmentação (rotulagem) da região periocular, capaz de discriminar sete componentes principais: íris, esclera, pestanas, sobrance-

lhas, cabelo, pele e óculos. Este tipo de segmentação permite, entre outras coisas, melhorar a definição da região periocular e a estimativa da pose. Este estudo está presente no capítulo 8, que consiste num artigo publicado na *IEEE International Conference on Biometrics - IJCB 2014* [8].

A sétima contribuição propõe um novo método de reconhecimento não-cooperativo da íris, combinando diferentes sinais biométricos. Este método foi avaliado no desafio Noisy Iris Challenge Evaluation - Part II (NICE.II), onde demonstrou boa performance e terminou classificado em terceiro lugar de entre os métodos a concurso. Este estudo está descrito no capítulo 9, artigo publicado no 33° volume da revista *Pattern Recognition Letters* [9].

A oitava contribuição trata a questão do reconhecimento biométrico em dispositivos móveis, usando a íris e a região periocular como características biométricas. Anuncia ainda a disponibilização de uma nova base-de-dados da íris e região periocular, contendo imagens adquiridas com dez configurações diferentes de dispositivos móveis, juntamente com a máscara de segmentação da íris. Esta nova base-de-dados permite testar tanto métodos de segmentação como de reconhecimento da íris e da região periocular. Além disso, reporta também o resultado de uma técnica de calibração ajustada aos dispositivos móveis, capaz de compensar as diferentes perceções colorimétricas associadas a cada configuração. Este estudo está descrito no capítulo 10, artigo publicado na revista *Pattern Recognition Leters*.

A nona contribuição é centrada no reconhecimento biométrico em dados extremamente degradados. Nesta contribuição é anunciada a disponibilização de uma nova base-de-dados anotada, contendo informação de 101 participantes: fotos de registo de alta qualidade; imagens degradadas por diferentes fatores de ruído; e vídeos dos participantes adquiridos também em condições adversas. São ainda divulgados resultados de uma experiência conduzida simulando um cenário de *watchlist identification*: foi conduzido um inquérito online em que os participantes eram convidados a realizar o reconhecimento quer positivo quer negativo de um indivíduo não identificado, e cuja foto estava degradada, contra um grupo de identidades conhecidas. Juntamente com as respostas fornecidas, os participantes indicaram os motivos que justificavam as suas escolhas, permitindo-nos perceber quais as características mais frequentemente associadas ao processo de reconhecimento efetuado pelo ser humano. Por fim, os resultados sugerem que o reconhecimento negativo poderá ser uma alternativa mais interessante ao lidar com informação extremamente degradada. Este estudo está presente no capítulo 11, que consiste num artigo publicado na revista *IET Biometrics*.

A décima contribuição consiste num novo método de deteção de *landmarks*, capaz de identificar e localizar seis elementos-chave numa imagem contendo a cabeça de um indivíduo, independentemente do seu ângulo de aquisição ou da inclinação da cabeça. Além das vantagens do ponto de vista de compreensão da cena, um sistema de reconhecimento biométrico multi-modal pode também beneficiar deste tipo de método, na medida em que se torna possível decidir qual o classificador a usar ou, eventualmente, ajustar o peso de cada um dos classificadores individuais em função das regiões que estão efetivamente visíveis. Avaliando a performance do método em imagens que simulam o tipo de informação adquirida em ambientes de vídeo-vigilância, demonstrámos que este consegue lidar com diferentes ângulos de aquisição (cobrindo completamente todos os 360°) e apresentando também invariância ao ângulo de inclinação da cabeça. Este estudo está descrito no capítulo 12, que consiste num artigo submetido para a 10^{a} *International*

Conference on Biometrics.

A décima-primeira e última contribuição introduz o sistema de reconhecimento biométrico idealizado. Este sistema minimiza as restrições de operação de tal forma que permite levar a cabo o reconhecimento biométrico em cenários de vídeo-vigilância. Consiste num sistema completamente autónomo que combina deteção e *tracking* de pessoas na cena, com o reconhecimento biométrico. Este sistema está descrito no capítulo 13, e consiste num artigo submetido para a 10^{a} *International Conference on Biometrics*.

Estado da Arte

Esta secção começa por introduzir alguns conceitos de biometria, nomeadamente os antecedentes históricos, principais característica de um sistema biométrico e sinais biométricos. De seguida descreveremos os desafios associados a ambientes não-controlados, principais problemas e esforços desenvolvidos no âmbito do reconhecimento não-cooperativo. Analisamos de forma sumária as técnicas desenvolvidas para a íris mais adequadas a este tipo de ambientes, e introduzimos a região periocular como uma alternativa mais viável. Por fim, tecemos algumas considerações sobre questões éticas e de privacidade.

Biometria

O termo biometria teve origem no Grego "bio (vida)" + "metria (medida)", e o seu significado literal é "medir a vida". De uma forma mais objetiva, biometria refere-se à capacidade de reconhecer um sujeito com base num sub-conjunto das suas características físicas ou comportamentais. Comparado com outros meios de identificação pessoal (e.g., cartão de cidadão ou passwords), a utilização de sistemas biométricos traz inúmeras vantagens: as características biométricas não são perdidas, roubadas ou esquecidas; são relativamente difíceis de forjar; e não podem ser utilizadas por terceiros, garantindo a não-repudiação.

Antecedentes Históricos

Mesmo sem registos físicos, podemos com toda a segurança identificar o uso de características faciais como um dos mais antigos exemplos de utilização da biometria por parte do ser humano. Diariamente, mesmo sem que nos apercebamos, recorremos a informação facial para nos reconhecermos, tornando a face a característica biométrica mais utilizada em todo o mundo.

A utilização documentada da biometria data de 30000 a.C. altura em que o homem pré-histórico usou a impressão das suas mãos para *assinar* pinturas em cavernas. Cerca de 500 a.C., outra característica biométrica era usada pelos Babilónios, que registavam as suas transações comerciais juntamente com a impressão digital. Escritos de João de Barros (1496–1570) descrevem práticas mercantis chinesas que incluíam a utilização da impressão digital para firmar acordos comerciais, e a utilização das impressões das palmas das mãos e pés de crianças em papel para evitar a sua troca [10]. No antigo Egito (3100–332 a.C.) descritores físicos dos comerciantes eram utilizados para distinguir indivíduos de confiança. Nas escrituras canónicas existem também várias referências de personagens que se reconheciam através de diferentes características biométricas: no livro de Tobias 11:6 relata-se a utilização da forma de caminhar e da silhueta, *"[...] ela esperava o regresso de seu filho, quando viu ao longe que voltava reconheceu-o [...]"*;

e em Samuel 26:17 *"[...] Saul reconheceu a voz de David, e exclamou: Não é a tua voz que ouço, meu filho David? [...]"*; etc.

Mais recentemente, em 1858, William Herschel iniciou o registo da palma da mão de cada um dos trabalhadores civis na índia, para que pudesse verificar a suas identidades no dia de efetuar o pagamento [11]. Não obstante, o primeiro sistema biométrico real – "*anthropometrical signalment*" – foi criado em 1879 por Alphonse Bertillion, chefe da divisão de investigação criminal. Sabendo que os criminosos recorrentes usavam frequentemente disfarces ou forneciam informação falsa quando eram detidos pela polícia, Bertillion pretendeu estabelecer uma forma de identificação que não necessitasse desse nível de cooperação por parte dos indivíduos. Para isso, recorreu à medição de determinados atributos físicos registando-os para referência futura – Figura 2.1. Um total de onze medições era realizadas para cada indivíduo, entre as quais a altura em pé e sentado, a distância entre as pontas dos dedos, o tamanho da cabeça, etc. Esse sistema acabaria por ser utilizado por forças policiais um pouco por todo o mundo, ainda que falhas na sua conceção viessem a torna-lo obsoleto em 1903, uma vez que duas pessoas distintas podiam apresentar medidas semelhantes.

Em 1892 Francis Galton publica o primeiro estudo sobre as aplicações biométricas das impressões digitais, detalhando de que forma os pontos de minúcia podiam ser utilizados para o reconhecimento [12]. Cinco anos depois, em 1897, a utilização de impressões digitais foi introduzida nas forças policiais seguindo as especificações de Edward Richard Henry da Scothland Yard que, com a ajuda de Francis Galton e Azizul Haque incluiu diversas melhorias no sistema de indexação. A utilização dos pontos de minúcia proposta por Galton é uma técnica ainda atual.

Ao longo das últimas décadas temos assistido a um crescimento acentuado na área da biometria, pautado pelo aparecimento de novos sinais biométricos, técnicas e aplicações.

Principais Fases de um Sistema Biométrico

Ao desenhar um sistema biométrico, ou seja, um sistema de reconhecimento de padrões que aplicado à informação biométrica de um indivíduo determine a sua identidade, é necessário considerar quatro fases principais [13]: sensor; avaliação da qualidade e extração de características; comparação e tomada de decisão; e, por fim, base-de-dados.

Sensor

Responsável pela aquisição do sinal biométrico, é a interface entre o utilizador e o sistema. Dependendo da característica biométrica que se pretende adquirir, vários sensores podem ser escolhidos. Na medida em que a maior parte dos sinais utiliza informação visual, a maioria dos sistemas usa câmaras para realizar a sua aquisição.

Avaliação da qualidade e extração das características

Após adquirir o sinal biométrico, e assegurando que o processo foi conduzido corretamente e que a informação a ser processada tem qualidade suficiente, um conjunto de características discriminativas é extraído e codificado.

Comparação e tomada de decisão

Nesta fase, as características extraídas são comparadas com os *templates* armazenados

na base de dados, e tomada a decisão sobre a identidade do utilizador. De acordo com o modo de funcionamento, esta comparação pode ser 1:1 ou 1:N.

Base-de-dados

Este último módulo é onde toda a informação dos utilizadores registados no sistema é armazenada. Esta informação já se encontra codificada utilizando a técnica estabelecida no módulo de extração de características, para que possa ser diretamente comparada com a informação extraída de novos indivíduos que se apresentem ao sistema. Geralmente a base de dados é criada durante a fase de registo.

Modos de Funcionamento

O sistema de reconhecimento biométrico pode funcionar segundo um de três modos [14]: verificação, identificação, e identificação negativa (*screening*).

Modo de Verificação

No modo de verificação, o sistema tenta validar a identidade de um indivíduo em particular (Figura 2.2). Como o utilizador que tenta aceder ao sistema já reclama para si uma identidade conhecida, o sistema realiza apenas uma comparação 1:1 (um-para-um) de forma a produzir um resultado binário: ou a identidade do utilizador é verificada ou não. Este modo de funcionamento é utilizado no quotidiano em *logins* de computador, ATMs, etc.

Modo de Identificação

No modo de identificação o sistema vai comparar um determinado indivíduo contra todos os utilizadores previamente registados no sistema (Figura 2.3). Contrariamente ao modo de verificação, não existe uma identidade inicial que o utilizador reclame para si, pelo que as comparações são realizadas na forma 1:N (1-para-vários). Este modo de operação não pode ser aplicado a outros métodos de identificação tradicionais (e.g., *tokens* e passwords), só podendo ser alcançado pela biometria.

Modo de Identificação Negativa

O modo de reconhecimento negativo, também chamado de *screening* [14], pode ser visto como uma extensão ao modo de identificação. O objetivo é garantir que um determinado indivíduo não pertence a um sub-grupo de utilizadores registados no sistema, realizado um conjunto de comparações 1:N orientadas à exclusão (Figura 2.4). Este modo de funcionamento é particularmente útil na segurança de aeroportos, locais públicos, etc.

Classificação e Propriedades dos Sistemas Biométricos

Existe atualmente um vasto leque de sistemas biométricos em funcionamento. Para auxiliar à sua classificação, os sistemas biométricos podem ser agrupados de acordo com seis diferentes perspetivas de funcionamento [15]:

Ostensivo (*Overt*) vs. Sub-reptício (*Covert*)

Dependendo de se o indivíduo está ou não ciente de que o sistema se encontra em funcionamento, este pode ser classificado em ostensivo (*overt*) ou sub-reptício (*covert*) respetivamente. Sistemas biométricos cooperativos sempre funcionam de forma ostensiva, embora atualmente existam esforços da comunidade científica concentrados em atingir

um funcionamento o mais sub-reptício possível, resolvendo tanto os problemas associados com elevadas restrições de funcionamento, como aqueles que advêm de utilizadores não-cooperativos.

Com-habituação (*Habituated*) vs. Sem-habituação (*Non-habituated*)

Sistemas com os quais os utilizadores interajam com regularidade são chamados, após algum tempo, de "habituados". Embora seja relevante estabelecer qual o grau de envolvimento do utilizador com o sistema, o funcionamento ideal passa por uma interação natural entre os dois mesmo sem habituação.

Supervisionado (*Attended*) vs. Não-Supervisionado (*Non-attended*)

Um sistema biométrico que necessite de um supervisor que conduza o processo de reconhecimento é designado por supervisionado. Contudo, um sistema não-supervisionado é geralmente preferível, isto é, um sistema com o qual o utilizador possa interagir de forma autónoma, sem qualquer tipo de supervisão.

Ambiente Padrão (*Standard*) vs. Não-Padrão (*Non-standard environment*)

Sistemas biométricos instalados em condições controladas, geralmente em ambientes de interior, com iluminação regular, e que contem com a cooperação do utilizador, operam em "ambientes padrão". Quando o sistema está instalado em ambientes não-controlados, com condições de aquisição variáveis, é dito que se encontra a operar num ambiente "não-padrão".

Público (*Public*) vs. Privado (*Private*)

Esta perspetiva de funcionamento está relacionada com o tipo de utilizador que interage com o sistema. Se o sistema está instalado num ambiente empresarial, onde os utilizadores registados no sistema são empregados dessa mesma empresa, trata-se de um sistema privado. Quando os indivíduos apresentados ao sistema não têm qualquer filiação com a empresa, é chamado de sistema público.

Aberto (*Open*) vs. Fechado (*Closed*)

Se o modo como o sistema funciona, ou se a informação sobre a qual ele trabalha são desconhecidos ou proprietários, diz tratar-se de um sistema fechado. Caso contrário, trata-se de um sistema aberto.

No âmbito destes trabalhos doutorais, pretende-se criar um sistema aberto, que seja simultaneamente sub-reptício, não-habituado e não-supervisionado, com vista a ser instalado em ambientes públicos e não-padrão.

Ao desenvolver um sistema biométrico, e além dos requisitos relativos ao sinal ou característica biométrica, existem cinco propriedades que devem ser tidas em conta [15]:

Performance

Todos os fatores que influenciem quer a velocidade, quer a precisão do sistema devem ser contemplados, sendo escolhidos aqueles que conduzam aos níveis de performance exigidos pelo contexto ao qual irá ser aplicado;

Aceitação

É essencial que os utilizadores aos quais o sistema se destina estejam dispostos a aceitar a disponibilização dos dados biométricos necessários à operação do mesmo;

Evasão

O sistema deve ser relativamente robusto a tentativas de fraude para que não possa ser facilmente contornado;

Exceções

O sistema deve estar munido dos meios necessários a levar a cabo formas de reconhecimento alternativas (e.g., sistema multi-modal) no caso das características não poderem ser extraídas para um utilizador específico;

Custo

O custo associado ao desenvolvimento e manutenção do sistema deve ser adequado ao contexto a que irá ser aplicado.

Sinais Biométricos

A face e a íris encontram-se entre os sinais biométricos mais utilizados para fins de reconhecimento, e são juntamente com a impressão digital os mais referidos na literatura [16, 17]. Jain *et al*. [18] estabeleceram um conjunto de quatro requisitos aos quais uma determinada característica física ou comportamental deve obedecer para que possa ser utilizada como sinal biométrico:

Universalidade

Se um determinado indivíduo vai ser reconhecido através dessa característica, então ela tem de ser comum a todas as pessoas;

Discriminação

A característica deve ser de tal forma inequívoca que duas pessoas dificilmente sejam confundidas;

Performance

A característica deve permanecer estável por um período razoável de tempo;

Colectibilidade

Por fim, a característica tem de conseguir ser adquirida, e as suas propriedades extraídas de forma quantitativa.

De forma geral, todos e cada um dos sinais biométricos utilizados atualmente respeitam esses requisitos.

Apresentamos de seguida os sinais biométricos mais utilizados [14], agrupados por região corporal, tal como ilustrado na Figura 2.5.

A região da mão contém várias características exploradas pelo seu potencial biométrico: impressão digital, impressão da palma da mão, geometria da mão, padrões dos vasos sanguíneos, impressão dos nós dos dedos, etc.

Impressão digital

A pele dos nossos dedos contém pequenas irregularidades, onde cumes e sulcos formam um conjunto de padrões. A partir desses padrões pode ser extraído um conjunto de pontos de minúcia usado para o reconhecimento. Tal como referido anteriormente, este sinal biométrico tem vindo a ser utilizado à vários séculos com uma performance considerável.

Sendo uma característica bem aceite, fácil de usar e passível de ser adquirida recorrendo a dispositivos relativamente pequenos e baratos, encontra-se atualmente implementada em inúmeros ambientes, desde controlo de acesso em empresas até aos computadores pessoais.

Impressão da palma da mão

De forma análoga à impressão digital, o reconhecimento biométrico baseado na impressão da palma da mão assenta na existência de um conjunto de padrões formado por cumes e sulcos. Uma vez que a área da palma da mão é significativamente maior que a dos dedos, os padrões distintivos são também encontrados em maior número. Contudo, é necessário recorrer a um sensor maior para proceder à aquisição dos mesmos. Como nem toda a área da palma da mão entra em contacto com os objetos do dia-a-dia da mesma forma que a ponta dos dedos, este sinal biométrico perde algum interesse do ponto de vista da análise forense.

Geometria da mão

A geometria da mão também pode ser utilizada como sinal biométrico, através da medição do comprimento e largura da palma e dedos. Apresenta como vantagem uma forma de aquisição não-intrusiva, e não requerer equipamento tão dispendioso como outros sinais. Contudo, esta característica está longe de ser suficientemente discriminativa para que possa ser utilizada em larga escala, uma vez que muitas pessoas partilham as mesmas dimensões de mão.

Padrões dos vasos sanguíneos

Os padrões dos vasos sanguíneos são únicos para cada indivíduo. Nas mãos em particular, existe um complexo sistema de vasos que permite o reconhecimento biométrico. Devido à sua localização, tem uma boa aceitação por parte dos indivíduos, embora seja necessário que toda a mão esteja em contacto com o sensor enquanto é iluminada por radiação no comprimento de onda infravermelho, de forma a maximizar o contrate entre as veias e os músculos da mão.

Impressão dos nós dos dedos

A impressão dos nós dos dedos é recolhida da pele na parte de trás dos dedos. Embora a sua utilização não esteja tão generalizada, a textura desta região da pele é particularmente rica em informação, possuindo por isso boa capacidade discriminativa.

Na zona da cabeça encontram-se também várias regiões com características valiosas do reconhecimento biométrico: face, formato da orelha, dentição, etc. Embora a região periocular também esteja contida neste grupo, será tratada numa secção separada.

Face

O reconhecimento facial tornou-se uma das aplicações de maior sucesso na disciplina de processamento de imagem. Sendo não-intrusiva e permitindo aquisição sub-reptícia, é um sinal biométrico preferencial em relação a outros com melhores taxas de reconhecimento, para aplicações não-cooperativas. Vários sistemas comerciais de reconhecimento facial estão disponíveis, bem como inúmeras técnicas desenvolvidas tanto para imagens estáticas, como para vídeos, no comprimento de onda visível e no infravermelho. Estas técnicas baseiam-se quer na análise global de toda a região da face, quer na relação entre os vários atributos faciais, a sua localização e formato.

Formato da orelha

O formato da orelha também pode ser utilizado como característica biométrica, mais precisamente a estrutura da sua cartilagem. Os padrões por ela gerados podem ser adquiridos tanto no comprimento de onda visível e utilizando câmaras comuns, como com câmaras que trabalhem no infravermelho e permitam a recolha da assinatura térmica, ou ainda a sua impressão. O principal problema deste sinal biométrico está relacionado com o facto de exigir do utilizador um grau de cooperação significativo, uma vez que quando em ambientes não-controlados as orelhas facilmente se apresentarão cobertas. Além disso, o seu alinhamento com o sistema é difícil.

Dentição

A estrutura da dentição é única para cada indivíduo. Contudo, a aquisição da informação que lhe está associada consiste num processo altamente cooperativo e intrusivo. Por esse motivo, este tipo de reconhecimento é utilizado especialmente por equipas forenses em cenários onde não podem ser utilizados outros sinais biométricos.

A região ocular é uma das mais exploradas em biometria, não só por conter múltiplas características, mas também por potenciar sistemas com elevadas taxas de reconhecimento. Os sinais existentes na região ocular são a íris, retina, a própria região periocular, etc.

Íris

A íris humana apresenta uma morfogenia predominantemente aleatória, única para cada indivíduo, conduzindo por isso a elevadas taxas de reconhecimento. Existe por isso um elevado volume de pesquisa em torno deste sinal biométrico, tendo-se tornado rapidamente um dos mais populares. Sendo um órgão visível do exterior, os seus padrões podem ser adquiridos de forma menos intrusiva. Contudo, a sua performance está fortemente condicionada em ambientes não-ideais, devido ao seu tamanho reduzido e elevada mobilidade, tornando-se difícil de capturar a distâncias maiores sem a cooperação do utilizador.

Retina

A retina é a parte mais interna do olho, e os vasos sanguíneos que a atravessam podem ser utilizados para identificar inequivocamente um indivíduo. Embora seja um sinal biométrico altamente discriminativo, na medida em que não existem duas pessoas com a mesma configuração de vasos sanguíneos, a sua localização e tamanho tornam-no difícil de capturar. É necessário um nível elevado de cooperação por parte do utilizador, que tem de permanecer imóvel e com o olho alinhado com o sensor. Por outro lado, é extremamente difícil de forjar, e é adquirido com níveis muito baixos de ruído.

Periocular

A região periocular apresenta um bom compromisso entre a utilização de toda a face, ou apenas da íris. A informação desta região é facilmente adquirida, mesmo sem cooperação por parte do indivíduo, na medida em que não requer uma captura cooperativa e a curta distância. Como tal, a sua utilização como sinal biométrico tem vindo a emergir, constituindo hoje em dia uma forte alternativa para ambientes nãp-controlados.

Características biométricas que necessitam de sensores médicos para serem adquiridas são classificadas de *médico-químicas*: ADN, electro-cardiogram (ECG), odor, etc.

ADN

O ácido desoxirribonucleico (ADN) é a molécula responsável pela codificação de todas as

instruções genéticas, estando presente em cada célula do corpo humano. À exceção de gémeos idênticos, cada indivíduo tem um ADN único e que, como tal, pode ser usado para o identificar de forma inequívoca. Apesar da sua utilização estar generalizada junto de profissionais de ciência forense, o reconhecimento baseado no ADN ainda é altamente intrusivo e dispendioso em termos de tempo, e a sua aquisição e manuseamento têm de ser conduzidos com extremo cuidado para evitar contaminação. Além disso, existem várias questões de privacidade que se levantam, na medida em que a análise do ADN pode revelar a predisposição de uma pessoa a certas doenças.

Odor

O odor emanado por um indivíduo é também ele único. Os sistemas atuais para a utilização do odor como característica biométrica consistem em *arrays* de sensores químicos, cada um deles sensível a uma fragrância em particular. Ainda assim, esta característica é afetada por perfumes, desodorizantes, dietas alimentares, medicamentos, etc.

As características biométricas comportamentais enquadram-se num ramo distinto da biometria, na medida em que estabelecem a identidade de um determinado individuo ao analisar o modo como este se comporta, e não os seus atributos físicos ou fisiológicos. Os sinais comportamentais mais utilizados são a voz, assinatura, modo de andar, modo de digitar, etc.

Voz

A voz assume um papel importante no reconhecimento humano. Mesmo que não estejamos a ver uma determinada pessoa, conseguimos reconhece-la apenas por ouvir a sua voz. As características vocais são determinados por diversos fatores, tais como as cordas vocais e as cavidades bocal e nasal. Os atuais sistemas de autenticação baseados na voz não oferecem qualidade suficiente para aplicações de alta-segurança, uma vez que se trata de um sinal particularmente sujeito a interferências e difícil de capturar sem ruído mesmo em ambientes padrão. Além do mais, pode ser condicionado pelas emoções do utilizador, estado clínico (e.g., infeções relacionadas com a garganta), etc.

Assinatura

A utilização da assinatura para verificar a identidade de uma pessoa é um procedimento comum no mundo inteiro, estando de tal forma disseminado que é das modalidades de reconhecimento mais bem aceite. Contudo, não só a forma como um indivíduo assina varia com o tempo, como ainda é influenciada pelo seu estado emocional e requer a sua cooperação.

Modo de andar

O modo como uma pessoa anda (*gait*) é uma também ele uma característica comportamental que pode ser usada para o seu reconhecimento. É não-invasiva, pode ser adquirida à distância, e a maioria das técnicas existentes para a sua análise não necessitam de informação em alta-resolução. Pode por isso ser utilizada em dados capturados por câmaras de vídeo-vigilância e em locais públicos. Contudo, há vários fatores que podem influenciar o modo de andar de uma pessoa (e.g., calçado ou roupa desconfortável), levando também a alterações na performance do sistema.

Modo de digitar

A análise do modo como um utilizador interage com o teclado constitui também uma característica comportamental com aplicações no reconhecimento biométrico. De entre a informação que pode ser extraída deste sinal temos a velocidade de escrita, a cadência

entre cada uma das letras, e erros típicos de digitação. Tem como vantagem o facto de não ser necessário nenhum sensor dispendioso, e a possibilidade de ser adquirido de forma sub-reptícia. Contudo, a forma como um indivíduo digita é condicionada pelo seu estado de espírito, nível de relaxamento, ou tipo de teclado.

Finalmente, é importante referir as características de *soft-biometrics*, que embora não apresentem capacidades discriminativas suficientes para identificar um indivíduo, podem ser usadas em conjunto com outros sinais biométricos para melhorar as taxas de reconhecimento ou acelerar o processo de pesquisa (e.g., categorização dos dados). Como exemplo de *soft-biometrics* temos o género, etnia, altura, marcas na pele (e.g., tatuagens), cor do cabelo ou dos olhos, etc.

Avaliação da Performance

Por forma a avaliar a performance de um determinado sinal ou sistema biométrico, várias métricas podem ser utilizadas.

A decidabilidade d', inicialmente introduzida por John Daugman [19], quantifica a separabilidade entre comparações intra- e inter-classe, analisando a relação entre as suas médias μ e desvios padrão σ (2.1), dado que ambas sejam distribuições Gaussianas. Comparações intra-classe referem-se a comparações em que tanto a informação adquirida como o *template* ao qual está a ser comparada pertencem ao mesmo utilizador, e inter-classe quando pertencem a diferentes indivíduos.

Outra representação comum de performance em problemas de classificação binária é a curva Receiver Operating Characteristic (ROC). Este gráfico relaciona a sensibilidade (ou True Positive Rate (TPR)) de um sistema com o False Positive Rate (FPR) e, com base nele, é possível quantificar a performance também com base na Area Under Curve (AUC). Esta última métrica varia no intervalo $[0, 1]$, correspondendo a unidade ao cenário ideal em que todas as comparações genuínas foram obtidas sem falsos positivos. Estabelecendo o limiar de aceitação de tal forma a que o erro introduzido seja igual em ambas as classes, obtemos o Equal Error Rate (EER).

A sensibilidade (ou TPR), e especificidade (SPC) são dadas pelas equações (2.2) e (2.3) respetivamente, e relacionam a quantidade de respostas correctas com o total de resultados positivos (True Positives (TP) + False Negatives (FN)) e as repostas negativas (True Negatives (TN) + False Positives (FP)). A Accuracy (ACC) (2.4) é utilizada para expressar o rácio de comparações corretamente classificadas. O seu valor máximo é 1, que corresponde ao cenário ótimo em que todas as classes foram classificadas de forma correta. Para uma análise mais equilibrada em situações onde exista uma grande discrepância entre a quantidade de comparações genuínas e de impostores, pode ser utilizado o Mathews Correlation Coefficient (MCC) (2.5). Esta métrica varia no intervalo $[-1, 1]$, sendo a unidade o valor ótimo [20].

Ambientes Não-Controlados

Vários autores se debruçaram sobre os principais problemas encontrados em ambientes não-controlados. Fancourt *et al*. [21] concluiu ser possível adquirir informação da íris com qualidade suficiente em distâncias até 10 metros. Smith *et al*. [22] examinaram comparativamente a informação da íris capturada simultaneamente no infravermelho e no comprimento de onda visível,

abordando a possibilidade de combinar informação multi-espectral para aumentar as taxas de reconhecimento. Na nossa pesquisa [4] analisámos a usabilidade da íris no comprimento de onda visível, quantificando as condições que permitiriam uma maior confiança no reconhecimento. Concluímos que o nível de iluminação da cena tem um impacto significativo, não devendo ser inferior a $120cd/m^2$, sendo o tipo de iluminante em utilização de menor relevância. Ross *et al*. [23] centraram-se no problema do reconhecimento em imagens degradadas da íris, tendo considerado cinco fatores de ruído: 1) iluminação não-uniforme; 2) movimento; 3) desfoque; 4) desvio no olhar; e 5) deformações não-lineares. O aspeto chave do seu método é que a codificação baseada numa única característica não soluciona todas estas variações, tendo os autores proposto um processo de codificação e comparação baseado em três métodos distintos: 1) histogramas de gradientes orientados; 2) Scale-Invariant Feature Transform (SIFT); e 3) modelo probabilístico de deformação. A informação extraída de cada descritor é comparada de forma independente, e os resultados combinados no final recorrendo a uma regra clássica da soma. Experiências nas bases-de-dados FOCS e Face Recognition Grand Challenge (FRGC), encorajam o desenvolvimento de mais trabalhos com este tipo de abordagem híbrida [24]. Quanto aos diferentes fatores de ruído que influenciam o reconhecimento em cenários não-controlados, os mesmos encontram-se descritos e ilustrado em maior detalhe na secção 11.4.2 do capítulo 11, e Figura 11.3.

A eficiência de sistemas de reconhecimento baseados na face em ambientes não-controlados é influenciada por uma série de fatores [24]: a sua estrutura tridimensional provoca variações na sua aparência, essencialmente a pose do utilizador; regiões significativas da face estão geralmente sujeitas a oclusão (e.g., aquisição não ortogonal); a sua aparência é claramente afetada pelas expressões faciais; e pode ser facilmente disfarçada.

Por último, ambientes não-controlados podem não permitir adquirir de forma satisfatória nem a íris nem a face. Nesses casos, a utilização de informação na região circundante ao olho (i.e., região periocular) tem vindo a ganhar crescente interesse por parte da comunidade científica, representando um bom compromisso entre utilizar a face completa ou apenas a íris [24].

Dispositivos Móveis

A utilização de dispositivos móveis (e.g., telemóveis e *tablets*) tem vindo a crescer significativamente ao longo dos últimos anos, bem como as suas capacidades e aplicações. A implementação de tecnologias de reconhecimento biométrico neste tipo de *gadget* não só é desejada, como representa a disponibilização de soluções de reconhecimento ao utilizador comum e, consequentemente, em qualquer lado e em qualquer altura.

A maioria dos algoritmos desenhados para ambientes *in the wild* consegue funcionar sem problema em dispositivos de consumo. Contudo, e a par das questões já levantadas pelos ambientes não-controlados, um reconhecimento baseado na íris está, neste tipo de dispositivos móveis, associado a uma série de problemas: os telemóveis e *tablets* vêm equipados com uma vasta gama de sensores (i.e., câmaras) e lentes, introduzindo discrepâncias nas imagens adquiridas (e.g., distorções da cor); a acquisição dos sinais biométricos *on-the-go* por indivíduos "não-treinados" irá resultar em maiores variações de pose, expressão, ângulo de aquisição, escala e rotação (e.g., os utilizadores seguram os dispositivos móveis de forma distinta); o ambiente onde a imagem está a ser adquirida pode ter condições de iluminação insuficientes, ou mesmo ilumi-

nação exterior muito forte que pode provocar reflexos sobre uma parte significativa da íris; etc.

No capítulo 10 estão detalhados os esforços desenvolvidos ao longo destes trabalhos doutorais no sentido de minimizar estes problemas: descreve-se a aquisição de uma base de dados da íris e periocular, adquirida com dez configurações móveis distintas; detalha-se a utilização de um método de correção de cor adaptado a cada configuração móvel, de forma a compensar as diferentes perceções de cor que a ela são inerentes; e são aplicadas estratégias de reconhecimento da íris e da região periocular pro forma a atingir o reconhecimento biométrico em dispositivos móveis.

Fases de um Sistema de Reconhecimento *In the Wild*

Ao desenhar um sistema de reconhecimento destinado a ambientes *in the wild* é necessário complementar os quatro elementos-chave previamente mencionados, ficando o sistema dividido nas sete-fases principais ilustradas na Figura 2.6:

1. **Aquisição do Sinal Biométrico** A escolha do sensor de aquisição dependerá do sinal biométrico que se pretende adquirir. No caso específico de ambientes *in the wild*, deve recorrer-se a sensores que não necessitem de interação por parte do utilizador, na sua maioria câmaras. Outro fator a ter em conta, principalmente ao projetar sistemas de funcionamento em tempo real, é o compromisso entre a definição da informação adquirida e as taxas de aquisição.

2. **Avaliação da Qualidade** Mesmo o mais otimizado dos sensores nem sempre devolve informação considerada ideal, e que permita a extração de características com elevada confiança. Quantificar a qualidade da informação adquirida, descartando imagens que não se adequem às fases seguintes do processamento, permitirá diminuir tanto o tempo final de execução, como as taxas de erro do sistema.

3. **Melhoria do Sinal** Complementarmente à avaliação de qualidade, o sinal adquirido pelo sensor pode ser submetido a métodos de remoção de ruído, se forem conhecidos os fatores que provocaram a sua degradação. Ainda que o sinal original não possa ser restaurado, este passo pode ser usado para aumentar a performance do sistema, ou em situações em que os dados sejam adquiridos com qualidade especialmente reduzida.

4. **Deteção e Segmentação** O sinal biométrico necessita de ser devidamente localizado e segmentado, para que a extração de características possa ser levada a cabo de forma eficiente. A fase de segmentação permite também descartar fatores de ruído que tenham transitado das fases anteriores (e.g., oclusão da íris).

5. **Extração de características** Um conjunto representativo de características tem de ser extraído, de tal forma a que a informação discriminante de cada sujeito seja codificada. Em sistemas de tempo real, estas características devem ser extraídas tão rápido quanto possível, e codificadas minimizando o custo computacional. Durante a fase de registo, a informação recolhida dos utilizadores é armazenada na base-de-dados para que, mais tarde, estes possam ser identificados ou as suas identidades validadas.

6. **Comparação** As características são então comparadas com os *templates* da base-de-dados. Em sistemas de tempo real, não só a informação da base-de-dados deve estar otimizada

para um rápido acesso, como *soft-biomtrics* podem ser utilizadas para reduzir o espaço de procura.

7. **Tomada de decisão** Com base no resultado da fase de comparação (que podem ser vários, no caso de um sistema multi-modal), é tomada uma decisão final sobre a identidade do indivíduo. Eventualmente, o sistema pode não ser capaz de chegar a uma conclusão positiva sobre a identidade do utilizador, mas ainda assim ser capaz de reduzir o conjunto de possibilidades.

A Íris como Sinal Biométrico

A possibilidade de adquirir a informação da íris a uma determinada distância, e as elevadas taxas obtidas no reconhecimento cooperativo, tornam-na um possível sinal biométrico também para cenários não-controlados, justificando os esforços em "relaxar" as condições de aquisição [9, 25, 26].

A íris tem inúmeras provas dadas enquanto característica biométrica, sendo que os sistemas de reconhecimento nela baseados estão implementados para um funcionamento fortemente cooperativo: os utilizadores necessitam de permanecer imóveis e relativamente próximos ao dispositivo de aquisição, enquanto uma imagem da íris é adquirida no infravermelho-próximo (700 a 900 nm). O recurso a iluminação no infravermelho permite adquirir imagens de boa qualidade, evitando a maioria dos fatores de ruído associados à iluminação no comprimento de onda visível (e.g., reflexos). Contudo, ao passarmos para ambientes não-controlados em que a íris é adquirida a distâncias superiores e em que os utilizadores estão em movimento, necessitamos de câmaras com um *f*-number maior e menor tempo de exposição para que se obtenha um *depth-of-field* aceitável. Uma vez que existe uma relação direta entre esses valores e a quantidade de luz na cena, para que o sistema continuasse a operar no infravermelho seria necessário recorrer a iluminadores suficientemente potentes, sendo estes prejudiciais à visão humana uma vez que o organismo não tem mecanismos naturais de resposta (e.g., contração da pupila). A exequibilidade do reconhecimento da íris neste tipo de ambientes usando o comprimento de onda visível continua a ser um tema controverso, especialmente para íris com um elevado nível de pigmentação (o que representa a maior parte da população mundial).

Na fase de segmentação da íris, muitas técnicas baseiam-se numa correta parametrização da transformada de Hough. Junli et al. [27] desenvolveram uma técnica de *ellipse fitting* robusta a mapas de arestas ruidosos, especialmente adaptada a dados degradados. Para resolver o problema associado com a aquisição de imagens a maiores distâncias, Tan and Kimar [28] propuseram um método baseado no algoritmo *grow-cut* capaz de discriminar entre informação correspondente à íris e não. Os resultados são ainda refinados através da estimativa do centro da íris, do refinamento dos contornos, da remoção de pestanas e sombras, e da localização das pálpebras. Este método foi validado contra bases-de-dados bem conhecidas (UBIRIS.v2, FRGC e CASIA.v4 Diatance), apresentado um menor custo computacional em relação a estratégias análogas. Alonso-Fernandez e Bigun [29] segmentam a íris com base no algoritmo *Generalized Structure Tensor*: recorrendo a filtros complexos, os autores obtêm informação sobre a magnitude e orientação de cada pixel da fronteira, fornecendo informação adicional para uma melhor distinção dos contornos pertencentes aos limites da íris. Xinyu et al. [30] propõem um algoritmo

capaz de segmentar imagens da íris muito robusto a variações de escala (diâmetro da íris a variar entre 50 e 360 pixels). Começa por detectar um conjunto de contornos usando o detetor Canny, usando de seguida os componentes não ligados como nodos de um grafo. Por fim é utilizado um critério de corte para distinguir quais os contornos da íris mais prováveis.

Ao trabalhar em ambientes não-controlados, é de particular importância que o sistema possua um módulo de deteção de ruído para que se possam distinguir as sub-regiões da íris que se encontram ocluídas por outro tipo de informação (e.g., pálpebras, pestanas, reflexos). Desta forma, é possível atenuar o seu impacto na fase de codificação, diminuindo a taxa de falsas rejeições (FRR) [24]. A maioria dos autores que se debruçaram sobre este problema utilizou métodos baseados em regras e, por conseguinte, a sua eficácia pode ser questionada. Por outro lado, Li e Savvides [31] utilizaram *Gaussian Mixture Models* para modelar as distribuições de probabilidade das regiões livres de ruído, e das que estavam ocluídas. Como característica base os autores propuseram filtros de Gabor otimizados.

Determinar a qualidade da informação adquirida permite também, como já foi referido, melhorar a performance do sistema de reconhecimento biométrico. Zuo e Schmid [32] propuseram a determinação de três índices de qualidade: qualidade da amostra, confiança no resultado da comparação; e qualidade da amostra e *template*. Este algoritmo tem a vantagem de ser genérico e, por isso, adequado também a outras modalidades biométricas.

A textura da íris é caracterizada pela ditribuiçõa irregular de características locais, tais como *furrows*, criptas, rugas e *spots*. Como tal, as técnicas tradicionais de reconhecimento baseiam-se na textura, decompondo a informação da íris num conjunto de coeficientes numéricos correspondentes a orientações ou frequências dos padrões dominantes. Os sistemas de reconhecimento da íris implementados comercialmente baseiam-se no método pioneiro de Daugman [19], e conseguem alcançar uma performance excecional em dados adquiridos no infravermelho. Apesar de algumas melhorias [33], este processo de reconhecimento consiste na transposição da íris segmentada para um sistema de coordenadas pseudo-polar, por forma a obter invariância à escala e translação. Nesta versão normalizada da íris é então feita a convolução com um conjunto de filtros de Gabor, a múltiplas frequências e orientações, sendo o resultado quantificado num dos quadrantes, com a extração de dois bits para a informação da fase. Para a comparação de assinaturas é utilizada a distância de Hamming, sendo feitas comparações com diversas translações dos dados de forma a obter invariância à rotação. Outros métodos clássicos do reconhecimento da íris baseiam-se em *zero-crossing*, como Boles e Boashash [34] que usam *wavelets* 1D com diferentes resoluções em círculos concêntricos, ou na análise da textura da íris, como a técnica de Wildes [35] que codifica a informação com base numa pirâmide Laplaciana. Em todo o caso, qualquer uma destas técnicas pressupõe cooperação por parte do utilizador.

O reconhecimento em ambientes menos controlados ganhou maior relevância com o projeto *Iris-on-the-Move* [36], um sistema de aquisição de imagem que torna o reconhecimento da íris menos intrusivo para os utilizadores, e um bom exemplo de engenharia: imagens são adquiridas na gama de infravermelhos mais próxima do comprimento de onda visível, enquanto o indivíduo atravessa um portal de controlo de acesso deslocando-se à sua velocidade normal. A empresa *Honeywell Technologies* deu entrada no processo de patente para um sistema similar [37], também ele capaz de reconhecer pessoas utilizando capturando a íris a uma certa distância.

A Região Periocular

Ao olho e à sua área circundante dá-se o nome de região *periocular*. A sua utilização para efeitos de biometria foi inspirada na capacidade que os seres humano têm em se reconhecer entre si através do olhar. A utilidade desta região é particularmente notória em cenários menos controlados, quando o grau de confiança na informação adquirida seja menor, e como forma de compensar problemas introduzidos pelo meio envolvente ou indivíduos que não estejam dispostos a cooperar com o sistema. Outras aplicações interessantes da região periocular são cenários onde a face sofreu transformações (e.g., cirurgia plástica) [38, 39].

Adquirir imagens deste sinal biométrico não pressupõe a necessidade de ambientes controlados, nem da cooperação por parte do utilizador. É relativamente estável (quando comparado com a região facial) e raramente se encontra ocluído. Além disso, estando a íris localizada no seu interior, podem ambas ser adquiridas com uma única câmara e a sua informação combinada para melhorar a performance do reconhecimento [24].

Os capítulos 5 e 6 contêm uma análise dos métodos mais relevantes presentes na literatura, comparando as características das principais bases-de-dados existentes e apresentando resultados da sua aplicação, identificando os principais problemas associados ao seu funcionamento.

Preocupações Éticas e de Privacidade

A par dos desenvolvimentos no campo da biometria surgem preocupações éticas e de privacidade, especialmente em métodos e sistemas desenhado para um funcionamento sub-reptício. Se por um lado a informação biométrica é única, pessoal e intransmissível, e uma forma de autenticação forte, por outro lado se os dados biométricos de uma pessoa vierem a ser comprometidos não podem ser alterados como é o caso de uma palavra-chave ou código PIN. Em ambientes não-controlados os indivíduos podem ainda estar sujeitos ao processo de reconhecimento mesmo sem que se apercebam ou haja o seu consentimento formal, levantando ainda mais questões.

Do ponto de vista da legislação europeia, existe um conjunto de princípios a ter em conta quando se implementa um sistema biométrico num cenário real [40, 41]:

Princípio da proporcionalidade
: Deve ser analisado se a utilização dos dados biométricos é realmente necessária no contexto da aplicação, ou se podem ser utilizados para o mesmo fim outros métodos de autenticação que não recorram à biometria;

Risco de discriminação
: O sistema implementado não deve demonstrar falsos positivos, falsos negativos, ou falhas no registo de utilizadores do sistema;

Uso impróprio / Âmbito
: Os dados biométricos adquiridos não devem ser utilizados para outros fins além daqueles a que originalmente se destinam. Além disso os utilizadores devem conhecer quais os fins a que os seus dados se destinam, e consentir de forma explicita a sua utilização pelo sistema.

Obtenção sub-reptícia de dados biométricos e monitorização

Os utilizadores devem ser informados e estar cientes sempre que a sua informação biométrica esteja a ser recolhida. Para que um sistema funcione sem esse pressuposto é necessário que exista uma dispensa legal.

Proteção dos dados

Os dados biométricos devem ser adquiridos de tal forma que não permitam a extração de outro tipo de informação além da que será usada para a identificação (e.g., informação médica ou étnica), ou caso contrário deverá existir um consentimento informado.

Adicionalmente, um conjunto de medidas Privacy Enhancing Technologies (PET) foi tomado pelo European Committee for Standardization (CEN) através da iniciativa Information Society Standardization System (ISSS), visando a proteção e padronização da privacidade na Europa [42]. Exemplos dessas medidas são: a minimização da quantidade de dados em utilização, a sua encriptação e remoção logo que deixem de ser necessários; evitar a utilização de bases-de-dados centralizadas, que em certos países chegam mesmo a ser proibidas (e.g., Alemanha); etc.

Em todo o caso, as pessoas estão mais ou menos familiarizadas com a utilização de sistemas de vídeo-vigilância em locais públicos. Os Estados Unidos da América são um bom exemplo em como apesar do direito à privacidade estar instituído na constituição, os ataques de 11 de Setembro de 2001 alteraram de forma muito rápida a perceção civil do que seria a sua "expectativa razoável de privacidade" (*"reasonable expectation of privacy"*). A expectativa de privacidade das pessoas no que diz respeito a sinais biométricos que estejam normalmente visíveis é muito baixa, e a liberdade de escolha em não participar na capturada de dados pelo sistema nem sempre é razoável. Num aeroporto, por exemplo, uma pessoa não irá optar por não ser submetida aos procedimentos de segurança, sob pena de ser impedida de viajar [43]. Neste contexto, a utilização de mecanismos de reconhecimento negativo (*screening*) poderá ser uma alternativa mais bem aceite, uma vez que a informação armazenada na base de dados corresponde só a indivíduos "suspeitos", e a informação adquirida durante o funcionamento do sistema é descartada após a identificação negativa.

Em última análise, e independentemente das diretrizes e protocolos existentes, poucas são as garantias reais de que estes procedimentos sejam realmente postos em prática.

Principais Conclusões

Esta tese propõe-se a abordar o problema do reconhecimento biométrico em ambientes não-controlados, descrevendo o trabalho desenvolvido no sentido de implementar um sistema que funcione de forma completamente autónoma e não-supervisionada em ambientes adversos.

A estratégia seguida para abordar este problema dividiu-se em cinco passos: começamos por estudar a íris enquanto sinal biométrico, e em particular a sua usabilidade em cenários não-controlados, onde as restrições de aquisição são menores e, consequentemente, se lida com informação mais degradada; de seguida conduzimos uma análise mais detalhada sobre como a performance da íris enquanto sinal biométrico é condicionada pela luz no comprimento de onda visível, mais especificamente pelo tipo de iluminante em utilização, e pelo nível de luminância; estudámos posteriormente os sinais biométricos emergentes, e em particular a região periocular, uma vez que esta tem vindo a assumir um papel cada vez mais relevante na litera-

tura no que toca a ambientes não-ideais; os métodos de reconhecimento mais relevantes foram então implementados e testados, e construiu-se uma base de dados que replica os fatores de ruído e condições observadas em cenários não-cooperativos, ainda que de forma controlada e devidamente quantificada; por fim, pretendíamos estudar a possibilidade de recorrer ao reconhecimento negativo, convergindo assim para um cenário que julgamos mais adequado aos desafios atuais. Ainda que este último objetivo não tenha sido completamente alcançado, linhas complementares de pesquisa foram levadas a cabo e, em ultima analise, um sistema de reconhecimento biométrico foi proposto capaz de trabalhar autonomamente sobre cenários de vídeo-vigilância. A par com estes objetivos, várias contribuições científicas foram publicadas em revistas e conferências internacionais.

Durante a pesquisa realizada sobre a usabilidade da íris, avaliando o seu comportamento sob luz no comprimento de onda visível, foi possível observar como a aparência dos seus padrões é muito pouco condicionada pelo tipo de iluminante em utilização durante a sua aquisição, ainda que os níveis de luminância sejam de maior importância [4]. Tal facto permite avançar com maior confiança para a construção de um sistema biométrico baseado na íris capaz de trabalhar sob diferentes *tipos de luz*, quer natural quer artificial, desde que um nível adequado de luminância seja mantido, ou medidas adicionais sejam tomadas para contornar esta limitação. Foi também proposto um novo método de reconhecimento utilizando informação da íris, capturada no comprimento de onda visível, baseado na extração dos descritores de cor e forma presentes no MPEG-7 a partir de sub-regiões da íris definidas de forma autónoma. Esta nova técnica demonstrou não só níveis de performance próximos dos métodos do estado-da-arte, como também um baixo nível de correlação em relação aos mesmos, sugerindo que a sua eventual fusão poderá conduzir a um acréscimo adicional na performance [5].

No âmbito do reconhecimento periocular o estudo da literatura revelou que as mais recentes pesquisas se focam na análise da textura e na extração de pontos-chave, e que mesmo algoritmos menos complexos (e.g., técnicas baseadas em Local Binary Patterns (LBP)) resultam em níveis de performance consideráveis. A região periocular é particularmente favorável a ambientes não-ideais e não-cooperativos, quer quando utilizada por si só, quer quando combinada com a íris, e ainda que a informação contida nesta última não possa ser capturada de forma ótima. Testar os métodos do estado-da-arte contra uma base-de-dados comum permitiu-nos analisar comparativamente as suas performances, identificando os padrões que condicionam sistematicamente as suas capacidades [6, 44].

Tendo estabelecido a importância para aplicações biométricas da deteção dos cantos dos olhos em imagens da face (e.g., definição da região periocular), avaliámos os métodos presentes na literatura em imagens degradadas. Observando que as condições às quais os submetemos condicionavam de sobremaneira a sua performance, propusemos um novo método capaz de lidar com imagens capturadas *in the wild*. Comparativamente com as técnicas da literatura, o método por nós proposto revelou-se mais eficaz tanto em dados sem ruído, como em dados degradados (desfoque, rotação, diferenças de escala, etc.). De salientar ainda que essa eficácia foi alcançada sem perdas de eficiência computacional, fator importante quando se pretende construir um sistema biométrico capaz de funcionar em tempo real [7]. Também para uma melhor definição dos limites da região periocular, melhor estimativas da pose, e direção do olhar, foi proposta uma técnica de *labeling* capaz de discriminar sete componentes principais desta região. A abordagem seguida consistiu em duas etapas: inicialmente um grupo de classificadores locais recorre a des-

critores de textura para determinar, em cada pixel, a probabilidade de cada classe; de seguida, esta informação é combinada com restrições geométricas e informação sobre a forma de cada uma das regiões usando Markov Random Field (MRF) [8].

Com o objetivo de aumentar a fiabilidade do reconhecimento não-cooperativo baseado na íris sobre dados degradados, propusemos a fusão de diferentes técnicas de reconhecimento, de tal forma a que os problemas associados ao funcionamento no comprimento de onda visível fossem minimizados. A combinação da íris e da região periocular demonstrou melhorar a performance global do sistema, quer no modo de identificação quer de verificação, observação que foi corroborada com o terceiro lugar no desafio NICE.II [9].

Uma abordagem similar foi levada a cabo no Mobile Iris CHallenge Evaluation - Part I (MICHE I), desta vez tendo por base o reconhecimento biométrico em dispositivos móveis. Elaborámos e tornámos pública para a comunidade científica uma nova base-de-dados da íris e da região periocular – Cross-Sensor Iris and Periocular Dataset (CSIP) –, contendo imagens adquiridas em dez configurações móveis distintas e com oito fatores visíveis de ruído, juntamente com as respetivas máscaras de segmentação da íris. Esta base-de-dados permite avaliar técnicas de segmentação e de reconhecimento da íris e da região periocular. Foi-nos também possível identificar as disparidades cromáticas que alguns dispositivos introduziam nas imagens, pelo que propusemos a utilização de uma técnica de correção de cor para compensar as distorções inerentes a cada uma das configurações móveis. Os resultados mostraram esta abordagem conduz a resultados muito satisfatórios, especialmente em comparações *cross-sensor*. Demonstrámos ainda como a combinação de estratégias de reconhecimento da íris e do periocular podem ser usadas para superar os problemas associados à aquisição em dispositivos móveis. Adicionalmente, mostrámos como técnicas de codificação muito simples e de baixo custo computacional podem produzir performances consideráveis, o que se torna particularmente interessante no caso de aplicações desenvolvidas para plataformas móveis, onde as limitações computacionais são maiores [3].

Tendo em mente o objetivo final de desenvolver um sistema de reconhecimento capaz de lidar com informação adquirida em condições adversas, identificámos a necessidade de criar uma base-de-dados que agregasse múltiplos fatores de ruído de forma perfeitamente quantificada. Neste sentido, criámos a base-de-dados BioHDD, contendo informação de 101 participantes adquirida ao longo de várias sessões: fotos de registo de alta definição; um largo conjunto de imagens degradadas segundo dez fatores de ruído; sequências de vídeo com os participantes a percorrer um cenário não-uniforme. Levando a cabo um estudo online em que se simulava o processo de identificação no formato *watchlist* onde os participantes eram convidados a realizar tarefas de reconhecimento positivo e negativo, foi-nos possível identificar quais as características mais frequentemente associadas pelo ser humano ao processo de reconhecimento. Observámos a capacidade dos participantes em lidar com intensidades de iluminação adequadas e níveis moderados de oclusão, e o facto de terem atingido bons resultados quando na presença de imagens comprimidas ou de baixa-resolução sugere que são essencialmente usadas características globais. Como principais problemas identificámos situações em que as fotos dos participantes foram tiradas com maior inclinação da cabeça e níveis significativos de oclusão, o que pode ser particularmente relevante se pensarmos em indivíduos que tentem evitar a identificação olhando para longe da câmara ou cobrindo uma parte significativa da face. De entre os detalhes mais vezes identificados como condicionantes do reconhecimento temos informação sobre a forma e características holísticas, que por sua vez se revelaram também

as mais fiáveis. Este estudo permitiu-nos consolidar a escolha da região periocular enquanto característica biométrica principal, na medida em que as características localizadas na região central da face foram as que conduziram a maior taxa de acertos [2].

Tendo por base estas descobertas, e procurando um melhor entendimento das fotos adquiridas in the wild que contivessem a região da cabeça, propusemos um novo algoritmo para a deteção de *landmarks* [45] capaz de identificar e localizar seis elementos chave, entre os quais a região periocular. Este algoritmo, tendo sido testado contra um sub-grupo da base-de-dados BioHDD, demonstrou ter a capacidade de lidar com imagens adquiridas de múltiplos ângulos, apresentado igualmente robustez ao nível de inclinação da cabeça do utilizador.

Estes trabalhos de doutoramento culminaram com a conceção de um sistema híbrido de vídeo-vigilância com capacidade de reconhecimento biométrico: tendo por base num sistema de vídeo-vigilância capaz de detetar e fazer *tracking* de pessoas em ambientes *in the wild*, o mesmo foi complementado com mecanismos de análise de cenário e controle de uma câmara Pan-Tilt-Zoom (PTZ), adquirindo imagens mais detalhadas da região da cabeça dos indivíduos. Recorrendo à técnica de *landmarking* referida anteriormente, o sistema será capaz de ajustar o funcionamento dos diferentes módulos de reconhecimento às características biométricas que se encontrem efetivamente visíveis, decidindo quais usar e que peso dar a cada uma delas. Por fim, apresentamos uma *proof of concept* refletindo o funcionamento desse mesmo sistema num cenário de vídeo-vigilância [45].

Trabalho Futuro

A implementação de um sistema de reconhecimento biométrico, autónomo e completamente funcional, capaz de trabalhar de forma sub-reptícia *in the wild* e em condições adversas é ainda um trabalho em andamento. Embora o sistema já tenha sido desenhado na sua totalidade, alguns módulos encontram-se ainda em produção. Pretendemos levar a cabo uma completa validação do sistema, avaliando a sua performance em diferentes ambientes.

Embora dois dos métodos propostos tenham sido submetidos a desafios internacionais (NICE.II e MICHE I) e todas as contribuições tenham sido avaliadas pelos nossos pares, achamos que seria interessante alargar os testes realizados a diferentes bases-de-dados e de maior escala, para poder aumentar ainda mais a relevância estatística dos resultados apresentados.

Tal como referido na introdução, um dos objetivos definidos inicialmente para esta tese daria ainda lugar a uma pesquisa mais aprofundada, nomeadamente o estudo do modo de reconhecimento negativo enquanto alternativa mais adequada à realidade atual de determinadas aplicações biométricas. Ser capaz de garantir com elevado grau de confiança que um determinado indivíduo, do qual não sabemos a identidade, não pertence a uma determinada lista de "pessoas de interesse" é um objetivo muito tentador e de elevada aplicabilidade em segurança pública e de larga escala. Compreendendo todas estas vantagens, muitas das quais estão ainda relacionadas a questões de privacidade, pretendemos complementar o protótipo final com este modo de funcionamento.

Organização da Tese

Este documento encontra-se organizado em catorze capítulos. Os primeiro e último capítulos são dedicados respetivamente à introdução e conclusões e trabalho futuro. O segundo capítulo descreve os fundamentos da biometria, e os desafios associados a ambientes não-controlados. Oferece ainda detalhes sobre a usabilidade da íris neste tipo de cenário, uma visão geral dos métodos existentes e problemas identificados. Além disso, introduz a região periocular enquanto sinal biométrico com particular aplicabilidade em cenários de reconhecimento adversos. Por fim, são tecidas algumas considerações sobre questões éticas e de privacidade. Cada um dos restantes capítulos é constituído por um artigo, publicado ou submetido para apreciação junto de revistas ou conferências internacionais.

Abstract

Every human being is entitled, by his very nature, to a set of physiological and behavioral features that characterize him. The study of such features led to the development of a considerable amount of systems and applications, referred as biometric systems.

The usage of biometric systems has been significantly growing over the last years, particularly in the field of security: authentication, access control, criminal identification, etc. Being a high demanding sector, it is then natural that greater focus is placed on the biometric traits that are able to deliver high discrimination between subjects whilst being less prone to forgery. However, such constraints represent a significant impact on both system's usability and flexibility, requiring from the user a significant amount of cooperation. In this context, the iris is a primordial trait. The existing biometric recognition systems based on the iris follow the pioneer approach proposed by John Daugman, that proved itself as an excellent option for cooperative scenarios where images are acquired in the near-infrared spectrum.

However, not in every case user cooperation is expected and, when not, systems with such high acquisition constraints are of little or no use. Research is then focused on circumventing those issues, either by improving the existing methods or finding new and more fitting traits. On the later, the periocular region (i.e., the region surrounding the eye) is one of the most promising characteristics: it mimics a natural and spontaneous way of recognition employed by the human beings; has an advantageous localization in relation to the iris, making it easy to be simultaneously acquired; and has, as corroborated by the literature, a set of promising characteristics that can be used for recognition purposes.

The main objective of this doctoral work is then to either adapt or develop a novel biometric recognition system, suited for *in the wild* environments. Such systems should preferably use the periocular region as biometric trait, due to its flexibility and ease of acquisition in adverse conditions, and keep the operation constraints as low as possible. Subjects can be imaged at-a-distance, on-the-move, and under irregular lighting conditions, using cameras working in the visible wavelength.

To accomplish such goal, a set of intermediate milestones was established. At first, the iris was studied as biometric trait, paying particular attention to the techniques allowing its usage on *in the wild* scenarios. The effects of the visible wavelength light on iris performance for biometric purposes should not be disregarded and, as so, this factor was also studied. After rolling out iris usability as main distinctive feature, different emerging traits were analyzed, with special attention being paid to the periocular region. The most relevant methods were implemented and tested against the same dataset. Ultimately, multiple contributions were proposed and accepted by the scientific community, with applicability on different *in the wild* environments, the last of which is the proposal of an actual biometric system, working in real challenging conditions.

Keywords

biometrics; decision fusion; degraded data; error analysis; feature extraction; feature selection; global descriptors; image databases; iris biometrics; local descriptors; machine learning; multi-modal biometrics; ocular biometrics; performance assessment; periocular recognition; person identification; pose variation; privacy; recognition robustness; region of interest; supervised learning; texture analysis; unconstrained biometrics; visible wavelength spectrum; vision-based biometrics; visual surveillance.

Contents

List of Figures

List of Tables

Acronyms

AAM Active Appearance Models

ACC Accuracy

AUC Area Under Curve

CEN European Committee for Standardization

CIE Commission Internationale de l'Eclairage

CLM Constrained Local Models

CMC Cumulative Match Characteristic

CS candidate solutions

CSIP Cross-Sensor Iris and Periocular Dataset

DCT Discrete Cosine Transform

DEC Decidability

DNA deoxyribonucleic acid

DOG Difference of Gaussians

DWT Discrete Wavelet Transform

ECG electro-cardiogram

EER Equal Error Rate

EMFACS Emotional Facial Action Coding System

FACES Facial Expression Coding System

FACS Facial Action Coding System

FAST Facial Affect Scoring

FDA Fisher Discriminant Analysis

FERET Facial Recognition Technology

FP False Positives

FPR False Positive Rate

FN False Negatives

FPR False Positive Rate

FRGC Face Recognition Grand Challenge

FPS frames per second

GEC Genetic & Evolutionary Computing

GEFE Genetic & Evolutionary Feature Extraction

GLOH Gradient Location and Orientation Histogram

GOF goodness of fit

HD Hamming Distance

HMM Hidden Markov models

HOG Histogram of Oriented Gradients

HSV Hue-Saturation-Value

ISSS Information Society Standardization System

KCFA Kernel Correlation Feature Analysis

KLT Kanade-Lucas-Tomasi

kNN k-Nearest Neighbors

LBP Local Binary Patterns

LDA Linear Discriminant Analysis

LFDA Local Fisher Discriminant Analysis

LFW Labeled Faces in the Wild

LoG Laplacian of Gaussian

LPP Locally Preserving Projections

MBGC Multi Biometric Grand Challenge

MICHE I Mobile Iris CHallenge Evaluation - Part I

MCC Mathews Correlation Coefficient

MIS mismatch

MLP Multi Layer Perceptron

MLR multinomial logistic ridge regression

MOTA Multiple Object Tracking Accuracy

MOTP Multiple Object Tracking Precision

MRF Markov Random Field

mRMR minimum-Redundancy Maximum Relevance

NICE.I Noisy Iris Challenge Evaluation - Part 1

NICE.II Noisy Iris Challenge Evaluation - Part II

NIR near-infrared

NN Neural Networks

NTSC National Television System(s) Committee

PCA Principal Component Analysis

PET Privacy Enhancing Technologies

PIE Pose, Illumination and Expression

PTZ Pan-Tilt-Zoom

ROC Receiver Operating Characteristic

ROI Region of Interest

SIFT Scale-Invariant Feature Transform

SPC Specificity

SSGA Steady-State Genetic Algorithm

STR Signal to Noise Ratio

SURF Speed Up Robust Features

SVM Support Vector Machine

TP True Positives

TN True Negatives

TPR True Positive Rate

UBIPr UBI Periocular Recognition

UDP unsupervised discriminant projection

ULBP Uniform Local Binary Patterns

USA United States of America

VR Verification Rate

VSG2/5 Visual Stimulus Generator

VW Visible Wavelength

WLBP Walsh-Hadamard transform encoded Local Binary Patterns

X-TOOLSS eXploration Toolset for Optimization of Launch and Space Systems

Chapter 1

Introduction

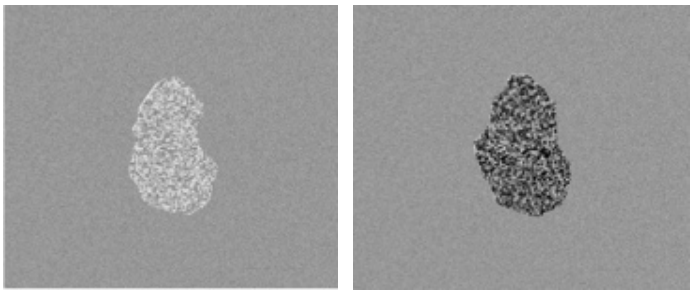
1.1 Overview

This thesis addresses the subject of biometric recognition in unconstrained environments. Being a rather challenging objective, we choose to adopt a divide and conquer approach. At start, iris usability for biometric recognition *in the wild*is analyzed, and more suited alternatives are sought. With the periocular region being established as the most promising biometric trait for the targeted scenarios, a more in depth study of that trait was conducted. This doctoral research adds up to multiple scientific contributions and, ultimately, a fully automated system is proposed. This chapter describes thesis focus and scope, followed by problem definition and objectives, the main contributions and thesis organization.

1.2 Thesis Focus and Scope

Societies' increasing concerns about both individual and global security has put biometrics as one of the most actives areas of research. This growth trend is visible on methods, techniques and applications, being deployed in a wide range of devices. Nowadays we can find biometrics being used in personal electronics, such as mobile devices and personal computers, but also and foremost on high-level access control systems, such as border access control. On this last group of applications and due to its highly demanding requirements, most of the research is focused on traits allowing higher user discriminability and lower forgery potential, at the cost of usability and system flexibility.

Achieving biometric recognition is in fact a pattern recognition problem, as we aim at recognizing a subject based on a set of his physiological or behavioral features. Pattern recognition is the scientific discipline responsible for processing raw data from certain objects, working out the information (i.e., patterns) needed to sort them into classes. Having been a theoretical statistical research field until 1960s, it was the evolution of computer systems that promptly



(a) Benign lesion (b) Malign lesion (cancer)

Figure 1.1: Illustration of image regions corresponding to two different classes [1].

demanded for a large amount of applications, making it a very active field within the machine learning domain, in particular on machine vision, data mining and knowledge discovery [1].

For illustration purposes, Figure 1.1 depicts two distinct classes from a classical pattern recognition problem: facing a depiction of a certain object, determine to which class it belongs. In this particular example, our objective would be to analyze a medical image containing an identified region and determine if it belongs to a benign lesion (class A) or a cancer (class B). For illustration purposes, lets assume we possess a database of such images, and the data from a particular class is similar.
In order to successfully classify a new instance of data, we first need to determine which features make the classes distinct. In this illustrative example, we can perceive from visual inspecting the images how pixel intensities significantly differ between classes. Thus, simply by computing the mean and standard deviation values of the pixel intensities, we can work out from Figure 1.2 how each class occupies a distinct area of the plot. Having such clear class distinction, a classification threshold can be unequivocally drawn (straight line).

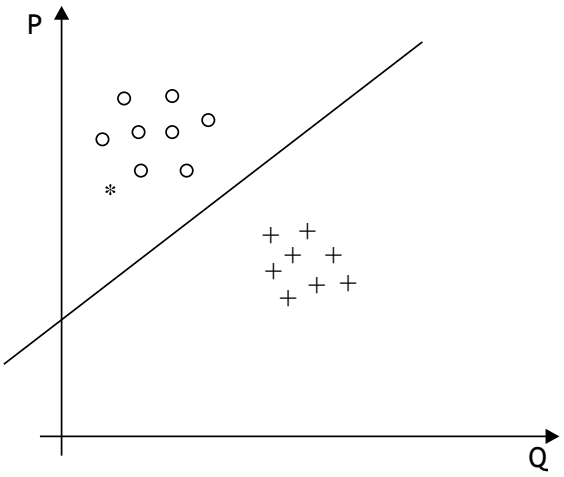


Figure 1.2: Mean values P and standard deviation Q for images belonging to classes A (o) and B (+). The observation marked with a star (*) corresponds to a new observation [1].

When *acquiring data* from a new observation, we carry on *feature extraction* the same way as we did before (P and Q). Placing that new observation on the plot (*) we can *classify* it as belonging to class A, as it is above the classification threshold. Those are the general steps to solve a pattern classification problem.

From the existing biometric traits, the commercially deployed biometric systems prefer those capable of lower error rates and, on this scope, the iris plays an important role. However, iris imaging is conditioned by strict acquisition protocols, mostly due to its reduced size and moving profile, thus not being a suitable trait for covert operation on non-cooperative scenarios. Contrary to the classical biometric systems that require high user cooperation, in *real world* applications that is not always expectable and, contrary to popular belief, no research efforts to date have produced a machine able to recognize human beings in such adverse conditions. In this context, the periocular region presents itself as an emerging biometric trait, with a growing number of publications on both its stand-alone potential and fusion possibility with more classical biometrics. In the later scenario, the iris is a particularly interesting fusion candidate, as both traits can be acquired simultaneously with a single sensor.

The scientific community is then facing two different roads: 1) lower the acquisition constraints present in the existing systems, adapting them for unconstrained operation; 2) explore the usability of new traits, best fit for the demands of these new working scenarios. This doctoral work merges those two goals by stressing the usability of existing traits (e.g., the iris) adapting them to the requirements of unconstrained operation, and fusing them with new and emerging traits, best fit for those scenarios from their conception (e.g., the periocular region). Our main purpose is to propose a biometric system for unconstrained environments, capable of dealing with at-a-distance data acquisition from moving subjects, under varying visible wavelength illumination settings. This motivation meets to another researching trend, as researchers are trying to combine biometrics with surveillance systems, aiming at the deployment of *hybrid* solutions capable of identifying *unfriendly* users *in the wild*.

1.3 Problem Definition and Research Objectives

This thesis addresses the problem of biometric recognition in unconstrained environments. Ultimately, the central objective of this doctoral program is to adapt or develop a biometric recognition system suited for unconstrained environments. Based on the conducted research, the periocular region is to be preferably used as main trait, due to its flexibility and ease of acquisition in non-ideal dynamic conditions.

To accomplish the proposed goal, a set of intermediate objectives was defined to better organize the required research work:

1. Study the iris as a biometric trait, with particular emphasis on the non-cooperative methods working over visible wavelength data. Such study should focus on the published iris recognition methods aimed at lowering the acquisition constraints, or work on degraded data.
2. The visible wavelength light introduces new noise factors with considerable impact on the recognition performance. We should then analyze iris reflectance whilst being illuminated by different illuminants (i.e., with different wavelengths), since this is a conditioning factor when working under the visible spectrum. Iris discriminatory properties should be analyzed over three axes: the type of illuminant on the scene; the luminance level of such illuminant; and the iris pigmentation level.
3. Study emerging biometric traits, paying particular interest to those who minimize the downside of non-ideal environments (e.g., the periocular region). Hence, the periocular region is an appropriate candidate for three reasons: 1) it is a natural way of recognition, used by the human being even without its awareness; 2) has a privileged location in relation to the iris, allowing simultaneous acquisition with a single sensor; 3) as corroborated by the literature, has a set of features that can be explored.
4. Implement and test the most relevant methods analyzed in the previous topic, having in mind that the evaluation of the implemented algorithms should be conducted over the same dataset. Moreover, we aim at gather a new dataset, with data acquired simultaneously over the near-infrared and visible wavelengths. Even though this new dataset is acquired on a controlled environment, it will introduce a set of noise factors that replicate those observed in unconstrained environments.

5. Most biometric systems attempt positive identification (or verification) against a gallery of enrolled users based on a (dis)similarity measure. In many *in the wild* applications however, biometric systems make more sense when used from the negative perspective: guarantee with enough confidence that an unknown subject does not belong to a gallery of "persons-of-interest", instead of attempting to identify him/her. On that basis our last intermediate objective is to study the state-of-the-art of the negative recognition paradigm.

From the proposed objectives, only the possibility of studying and performing negative recognition was not fully achieved, being presented as further work. We understand the practical advantages of deciding if a particular subject is not one of the users registered on the system, particularly in systems working covertly, which is further supported in our work [2]. However, we decided to prioritize the conception of equally challenging systems, with applications on different scenarios (e.g., mobile setups [3]).

Finally, the biometric system we aim at establishing should be developed in such a way that can be easily implemented, validated and debugged.

1.4 Main Contributions

This section briefly describes the contributions resulting from the research work developed in the scope of this doctoral program, as illustrated in Figure 1.3.

The first contribution consists of a preliminary assessment about the discriminating capability of the human iris when acquired under visible wavelength light, by perceiving and quantifying the conditions that enable iris recognition with enough confidence. This study is described in chapter 3, which consists of an article published in the proceedings of the *6th IEEE Workshop on Multimedia Information Processing and Retrieval* [4].

The second contribution proposes a new recognition scheme, based on techniques that are substantially different from those traditionally used. The minimal levels of linear correlation between the outputs produced by the proposed strategy and other state-of-the-art methods suggest that the fusion of both recognition schemes significantly improves performance, which is regarded as a positive step towards the development of extremely ambitious types of biometric recognition. This study is described in chapter 4, which consists of an article published in the 116th volume of *Computer Vision and Image Understanding* [5].

The third contribution provides a comparative overview of the most relevant research works in the scope of periocular recognition, summarizing the developed methods and enumerating the current issues. This study is described in chapter 5, which consists of an article published in the proceedings of the *IEEE Symposium on Computational Intelligence in Biometrics and Identity Management – CIBIM 2013* [6].

The fourth contribution delivers a more detailed comparative overview of the most relevant research on the scope of periocular recognition, with widen detail on the underlying techniques and a comprehensive analysis of the state-of-the-art results against a common dataset. It starts

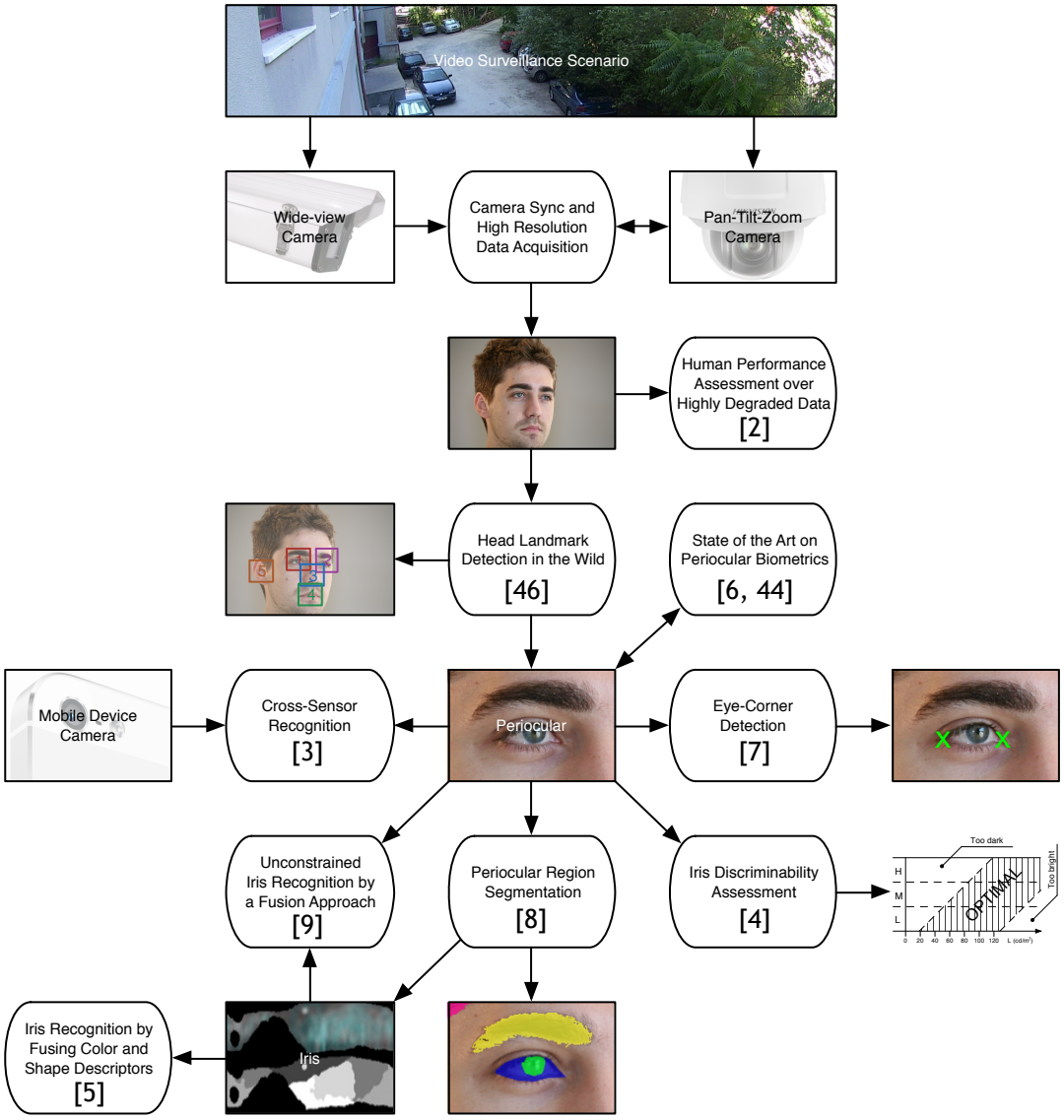


Figure 1.3: Illustrative overview of the main contributions of our research work. Some depicted methods were used as modules on the building of a fully automated surveillance system for human recognition purposes over video surveillance scenarios [45].

with an introductory three-fold framework, with relation to traditional biometric traits and systems, periocular anatomy and identified difficulties, and concludes with insights on the main degradation issues and directions for further improvements on this emerging trait. This study is described in chapter 6, which consists on a paper submitted for publication in the journal of *Artificial Intelligence Review*.

The fifth contribution provides an eye-corner detection method able to deal with degraded data, emphasizing robustness and applicability to real-world conditions. Our experiments show that the proposed method outperforms others in both noise-free and degraded data (blurred and rotated images and images with significant variations in scale). This study is described in chapter 7, which consists on an article published in the proceedings of the *IEEE International Conference on Biometrics – IJCB 2011* [7].

The sixth contribution provides a segmentation (labeling) method for the periocular region, able to discriminate between seven components in a single shot: iris, sclera, eyelashes, eyebrows, hair, skin and glasses. Having such a labeled image allows improving the definition of the periocular Region of Interest (ROI), with better pose and gaze estimation. This study is described in chapter 8, which consists on a paper published in the *IEEE International Conference on Biometrics – IJCB 2014* [8].

The seventh contribution proposes a novel strategy fusing different recognition approaches, describing how it contributes to a more reliable non-cooperative iris recognition by compensating for degraded images captured in less constrained acquisition setups and protocols, under visible wavelengths and varying lighting conditions. The proposed method was tested at the NICE.II contest, where its remarkable performance was corroborated by a third-place finish. This study is described in chapter 9, which consists on a paper published in the 33rd issue of *Pattern Recognition Letters* [9].

The eighth contribution focus on biometric recognition on mobile environments using the iris and periocular information as main traits. It announces the availability of an iris and periocular dataset containing images acquired with 10 different mobile setups, along with the corresponding iris segmentation data, allowing to evaluate both iris segmentation and recognition methods, and periocular recognition techniques as well. Furthermore, it reports the outcomes of device-specific calibration techniques that compensate for the different color perception inherent to each setup, and proposes the application of well-known iris and periocular recognition strategies, giving evidence on how they can be fused to overcome the issues associated with the mobile environments. This study is described in chapter 10, which consists on a paper published the journal *Pattern Recognition Letters*.

The ninth contribution also focuses on biometric recognition in extremely degraded data. The availability of an annotated dataset containing high quality mugshots of 101 subjects, and large sets of probes degraded extremely by ten different noise factors is announced. Furthermore, it reports the results of a mimicked watchlist identification scheme: an online survey was conducted, where participants were asked to perform positive and negative identification of probes against the enrolled identities. Along with their answers, volunteers had to provide the major reasons that sustained their responses. That enabled us to perceive the kind of features that are most frequently associated with successful / failed human identification processes, being

observed how humans greatly rely on shape information and holistic features. Finally, evidence is provided that the positive human identification on such extremely degraded data might be unreliable, whereas negative identification might constitute an interesting alternative for such cases. This study is described in chapter 11, which consists on a paper published in the journal *IET Biometrics*.

The tenth contribution consists of a novel landmark detection technique, able to identify six-key elements of the human head and pinpoint their location regardless of the image acquisition angle or head's pitch. Apart from the scene understanding point of view, on a multi-modal biometric system being able to tell if a particular landmark is visible or not would allow deciding which classifier to use, or adjust weights on a score-level fusion scheme. Evaluating the proposed technique against surveillance-like data showed that its ability to cope with images acquired over multiple angles covering a full 360° view, even when participants exhibited variations along the full head pitch range. This study is described in chapter 12, which consists on an article submitted for the 8^{th} *IAPR International Conference on Biometrics*.

The eleventh and last contribution is the introduction of a novel biometric system that effectively minimizes the operation restrictions, bringing biometric recognition to video surveillance scenarios. It consists on a fully automated surveillance system for recognition purposes, that extends a typical human detection and tracking machine by further enhanced it with a PTZ camera that delivers data with enough quality to perform biometric recognition. This study is described in chapter 13, which consists on an article submitted for the 8^{th} *IAPR International Conference on Biometrics*.

1.5 Thesis Organization

This thesis is organized in fourteen chapters. The first and last chapters are devoted to the introduction and conclusions and further work respectively. The second chapter introduces the basis of biometrics and the challenge of unconstrained scenarios, followed by insights on iris usability such environment. After an overview of the existing methods and identified issues, the periocular region is presented as a valuable trait for adverse environments. Each one of the remaining chapters is formed by an article, published or submitted for publication in international journals or indexed conferences.

Chapter 2

State-of-the-Art

2.1 Overview

This chapter introduces the basics of biometrics: its historical background, the main characteristics of a biometric system, and traits. We further describe the challenge of unconstrained scenarios, the related issues and the efforts towards non-cooperative recognition. We overview the iris recognition techniques best suited for those environments, and introduce the periocular region as a viable alternative. In the end, we describe some ethic and privacy concerns.

2.2 Biometrics

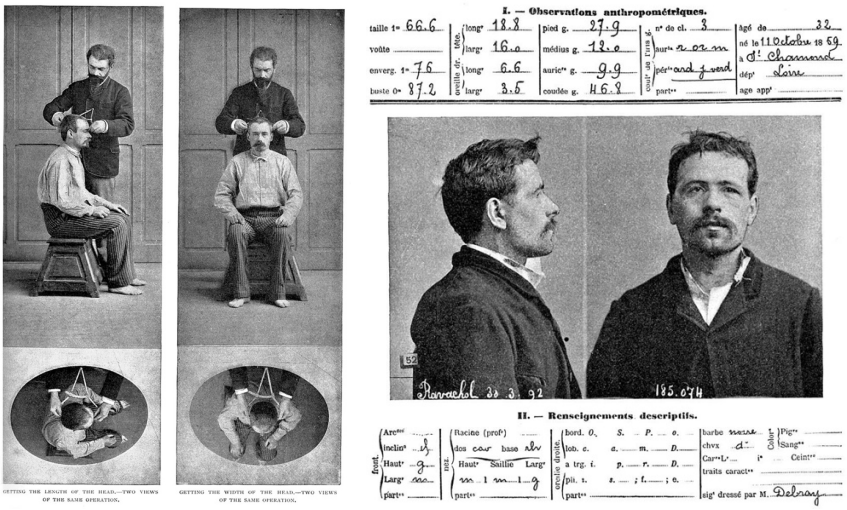
The term biometrics came from the Greek "bio" + "metrics", and literally means "to measure life". In more scientific therms, biometrics refers to the ability to recognize a subject based on a set of his / her physiological or behavioral features.

Compared to other means of proving a subject's identity as tokens (e.g., personal identity cards) or passwords (e.g., memorized PIN), the usage of a biometric system offers numerous advantages: traits cannot be lost, stolen or forgotten, are harder to forge, and cannot be used by third-parties thus assuring non-repudiation.

2.2.1 Historical Background

Even without records, we can safely point out the use of facial features as one of the oldest examples of biometrics actually employed by humans. Everyday and even without our awareness we all use facial information to recognize each other, making it the most widely used biometric trait.

Documented usage of biometrics date back to approximately 30,000 B.C. when prehistoric men used handprints to *sign* their paintings on caves. Around 500 B.C. another hand trait was used by Babylonians, who recorded their transactions in clay tables along with their fingerprints. Writings from João de Barros (1496–1570) also describe Chinese merchant practices including the usage of fingerprints to settle business agreements, and avoiding children being mistaken by imprinting their palm- and footprints on paper [10]. In ancient Egypt (3100–332 B.C.) physical descriptors were used to tell apart trusted traders in the market. In the canonical scriptures there are also several statements of characters performing biometric recognition using multiple traits: in the book of Tobit 11:6 gait and silhouette are used, "*[...] she saw him afar off, and presently perceived it was her son [...]*"; and in Samuel 26:17 "*[...] Saul recognized David's voice [...]*"; etc.



(a) The measuring procedure. (b) The "antropometric card".

Figure 2.1: Depiction of the Bertillion's system: illustration on how the measurements were taken; and the card where the information was stored[1].

More recently, in 1858, William Herschel started recoding handprints on every civil worker contract in India, so he can verify their identities on payday [11]. Nonetheless, the first true biometric system was set up in 1879 by the chief of the criminal identification division Alphonse Bertillion – the "*anthropometrical signalment*". It was known that recurring criminals often disguised themselves or provided false information when arrested. Bertillion's purpose was to establish a way of identify them even without their cooperation, by measuring invariant physical attributes and storing them for further reference – Figure 2.1. A total of eleven measurements were taken for each individual, including standing and sitting height, fingertip distance, size of the head, etc. This system became used by police forces all around the world, but had a flaw that eventually led to its face in 1903: two people can have the same measurements.

In 1892, Francis Galton publishes the first study on fingerprint biometrics, detailing how minutia points could be used for biometric recognition purposes [12]. Five years later, in 1897, the usage of fingerprint biometrics was introduced in the law enforcement, according to the specifications of Edward Richard Henry at the Scotland Yard, with the help of Francis Galton and Azizul Haque who also perfected the indexing system. They used the minutiae points present in the fingerprints, a technique that is still in use.

Over the last decades we have seen a dramatic growth in biometrics, as new traits, methods, techniques and applications emerged.

2.2.2 The Main Stages

When devising a biometric system, that is, a pattern analysis system working on biometric data to determine the identity of an individual, one must consider four main stages [13]: sensor module; quality assessment and feature extraction; matching and decision making; and database.

[1] Image source: `http://sherlockholmes.stanford.edu/print_issue3.html`

Sensor module

Responsible for acquiring the biometric trait, is the input interface between the individual and the system. One can choose from a wide variety of sensors, depending on which trait is being acquired, and the characteristics we want to capture. Since most traits use visual data, most systems use cameras for acquisition.

Quality assessment and feature extraction

Upon acquisition of the biometric trait, and assuring the process was conducted correctly and produced usable data, feature extraction is then conducted. The raw collected information is processed so that discriminatory features can be extracted and encoded.

Matching and decision making

At this stage, the features extracted from the acquired data are matched against the templates stored on the database. Depending on the operation mode, this can be a 1:1 or 1:N matching process, and the corresponding decision is taken.

System database

This last module is where all the information from the users enrolled in the system is stored. The stored information is already encoded, using the same technique established at the feature extraction module, so that it can be directly compared to new features extracted from further users. Usually the system database is built during an enrollment stage.

2.2.3 Operation Modes

The biometric system can operate in one of three distinct modes [14].

2.2.3.1 Verification Mode

In the verification mode the biometric system attempts to validate the identity of a particular individual (Figure 2.2). Since the user trying to gain access already claims an identity of his own, the system only performs a 1:1 (one-to-one) comparison to produce a binary result: either user identity is verified or it is not. This operation mode is used everyday on computer logins, ATMs, etc.

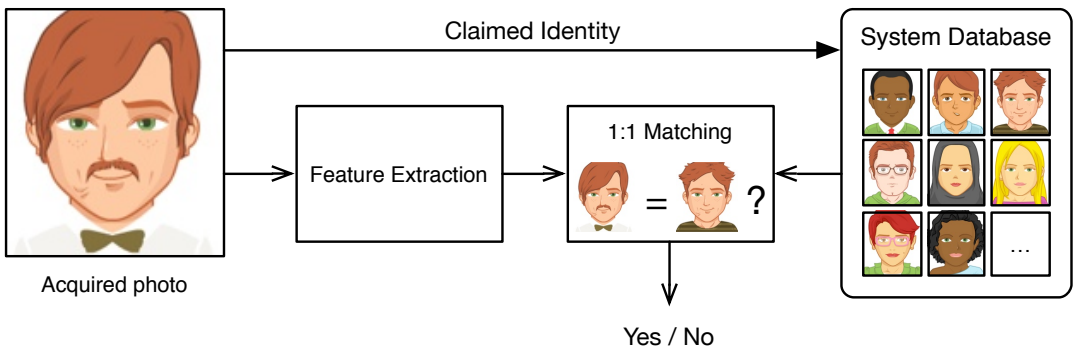


Figure 2.2: Block diagrams of a biometric system carrying out a verification task.

2.2.3.2 Identification Mode

In the identification mode the system tries to match a particular individual against all users previously enrolled on the database (Figure 2.3). Contrary to the verification mode, there is no identity claim to start with, and 1:N (one-to-many) comparisons need to be performed. This operation mode cannot be applied to traditional recognition methods (e.g., tokens and passwords) and can only be achieved through biometrics.

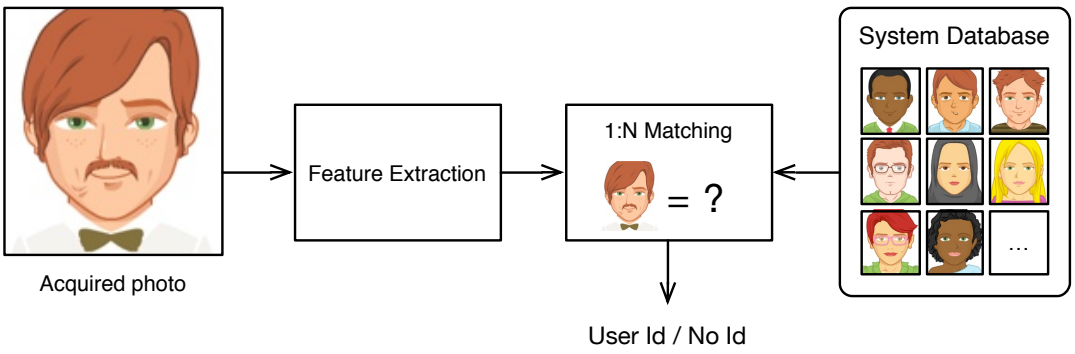


Figure 2.3: Block diagrams of a biometric system carrying out an identification task.

2.2.3.3 Negative-Identification Mode

This operation mode, also designated as *screening* [14], is an extension to the identification mode. Its purpose is to assure that a particular individual does not belong to a subset of users previously enrolled in the system, by performing 1:N exclusion-oriented comparisons (Figure 2.4). This watchlist-*like* operation mode is particularly useful for airport security, public places security, etc.

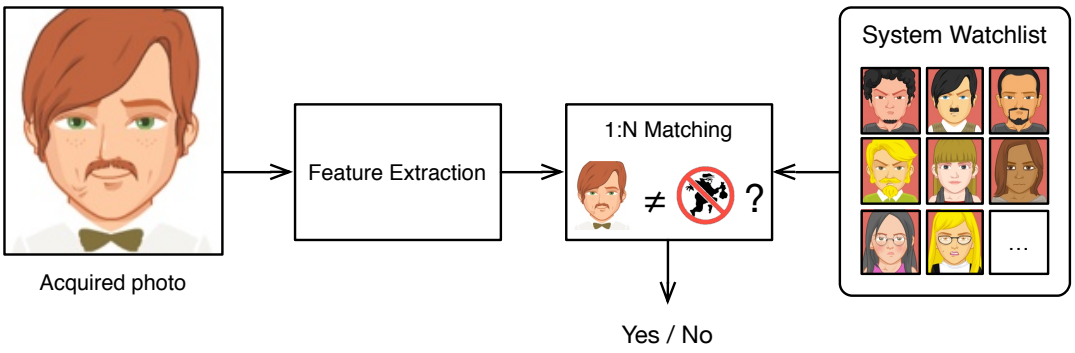


Figure 2.4: Block diagrams of a biometric system carrying out a screening task.

2.2.4 Classification and Properties of a Biometric System

There is a wide range of biometric systems operating nowadays. To better aid their classification, biometric systems can be grouped accordingly to six perspectives [15]:

Overt vs. Covert

Depending on the user being aware of system operation or not, the biometric system can be classified as *overt* or *covert*, respectively. Cooperative biometrics are always associated

with *overt* systems, but efforts are being put into the expansion of *covert* biometrics, circumventing the issues associated with both high-constraint setups and non-cooperative users.

Habituated vs. Non-habituated

Systems with which the users interact on a regular basis are called *habituated* (after some time). Although being relevant to establish the required degree of cooperation by the user, an ideal system would be that where individuals interact seamlessly even without habituation.

Attended vs. Non-attended

A biometric system requiring a supervisor to conduct the process is called *attended*. A *non-attended* system would be preferred, that is, a system the user can interact with alone, without any supervision.

Standard vs. Non-standard environment

Biometric systems operating under controlled conditions, usually indoor with regular illumination and cooperatively, are said to operate in a *standard environment*. If the system is deployed on uncontrolled scenarios where the acquisition conditions can vary, it is said to be on *non-standard environments*.

Public vs. Private

This perspective relates to the type of users interacting with the system. If the system is deployed in a enterprise environment, where the users enrolled are employees of the company, it is called a *private* system. When the individuals being present to the system are just customers, with no affiliation to the company, it is then a *public* system.

Open vs. Closed

When the way the system works, or the data it handles is unknown or proprietary, it is a *closed* system. Otherwise, it is an *open* system.

On the scope of this thesis we aim at establishing an open biometric system that is both covert, non-habituated and non-attended. It is to be deployed on a non-standard and public environment.

When developing a biometric system, and apart from the requirements that must be observed for the biometric trait (see section 2.2.5), there are five properties that must be regarded [15]:

Performance

All the factors that impact both performance (e.g., speed) and accuracy must be contemplated, being chosen the factors that lead to the accuracy and speed required by the context it is being used on;

Acceptability

It is essential that the individuals to whose the system is targeted at are willing to accept that the system can acquire that trait;

Circumvention

The system should not be easily bypassed by forgeries or impersonation attempts;

Exception Handling

The system should have means of accomplishing an alternative matching (e.g., multimodal), if the features can not be extracted from a particular user;

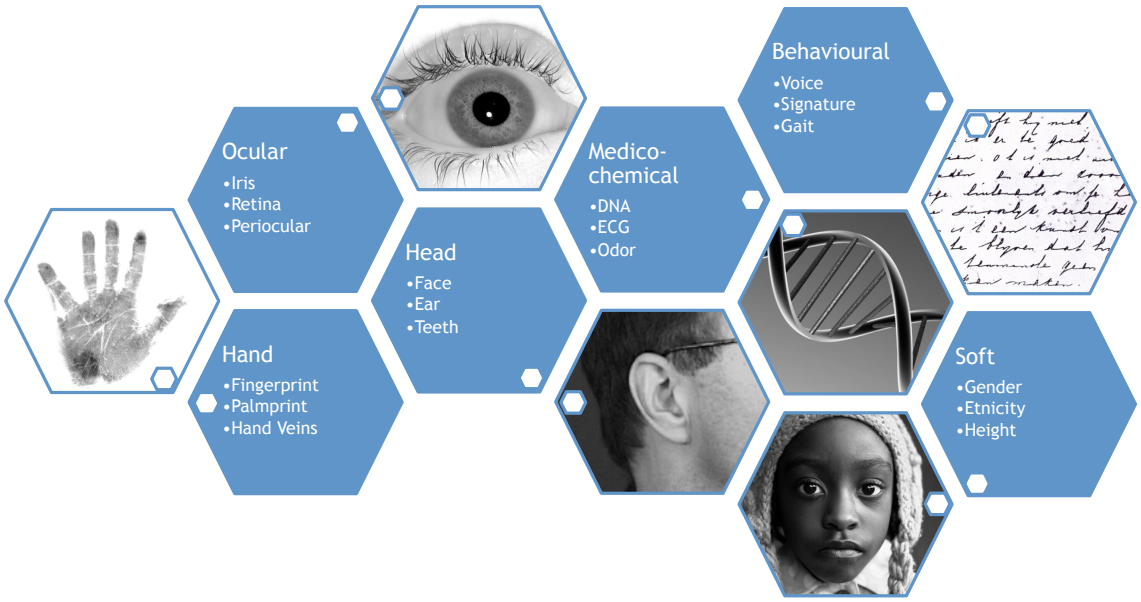


Figure 2.5: The main biometric modalities, and some examples of the traits they use.

System Cost

The cost associated with the development, deployment and maintenance of the system should be adequate for the context its being applied on.

2.2.5 The Biometric Traits

The face and the iris are among the most popular biometric traits used for recognition purposes and, along with fingerprint, the most reported in the literature [16, 17].

Jain *et al.* [18] established a set of four requirements for a particular physiological or behavioral characteristic, so it could be used as biometric trait. They are as follows:

Universality

If an individual is to be recognized using that characteristic, then it has to be common to every person;

Distinctiveness

It should be unambiguous enough so that two persons could hardly be misidentified;

Permanence

The characteristic should remain stable over a reasonable amount of time;

Collectability

The characteristic must be acquirable and its features extracted in a quantitatively way.

Accordingly, several traits are used nowadays, each one respecting those requirements up to a certain degree.

We now present the most used biometric traits [14], grouped by the body region they are located on, as illustrated in Figure 2.5. Some traits were not described, as even though some studies point these as suitable for biometric recognition further large scale studies are still required.

2.2.5.1 The Hand Region

The hand region contains several traits explored for their biometric potential: fingerprint, palmprint, hand geometry, vein pattern, finger knuckle print, etc.

Fingerprint
The skin in our fingers contains small irregularities, where ridges and furrows make a set of patterns. From those patterns, a set of minutiae points can be extracted, and used for biometric recognition. As said before this biometric trait has been used for several centuries (see section 2.2.1), with considerable good accuracy. Being a well accepted trait, easy to use and acquired with small and cheap devices, it is widely deployed in many scenarios, from access control in firms to personal computers.

Palmprint
Similar to the fingerprint, palmprint biometrics is based on skin ridge distribution along the palm of the hand. As the palm area is significantly higher than finger's, a lot more distinctive patterns can be observed. However, the palm of the hand does not become in contact with everyday surfaces as the fingerprint does, which can be a downside from the forensics point-of-view. As the same advantages and pitfalls of the fingerprints, plus a larger area thus also requiring a bigger sensor.

Hand geometry
The geometry of the hand can also be used as biometric trait, by measuring the length and width of the palm and fingers. As the advantages of being less intrusive and easy to acquire, and not requiring expensive equipment in comparison to the other traits. However, this is far from being a highly distinctive trait, as many people share the same hand dimensions.

Vein pattern
The vascular patterns of an individual are unique. In the hands in particular, there is a complex vein system allowing biometric recognition to be carried on. Due to its location, has good acceptance by the users, although requiring the whole hand to be in contact with the system whilst being illuminated by infrared light so that the contrast between the veins and the muscles is maximized.

Knuckle print
The knuckle print is extracted from the skin near the joints in the back of the fingers. Although not so commonly used, the finger knuckle skin is particular rich in texture information, thus holding high discriminative capabilities.

2.2.5.2 The Head

On the human head there are also several regions holding valuable cues for biometric recognition: face, ear shape, teeth, etc. The ocular region is also contained on the head, but since it gathers a lot of different traits and attention from the scientific community, will be treated as a separate section.

Face
The face recognition become one of the most successful applications of image analysis and understanding. Being non-intrusive and allowing cover acquisition, it became preferable over very reliable traits when aiming at less constrained subject recognition. Several commercial face recognition systems are now available, and a lot of techniques were developed for both still images and video, under the Visible Wavelength (VW) and infrared. Face recognition approaches are either based on a global analysis of the whole region as a set of pixel intensities, or the relation between facial attributes, their location and shape.

Ear shape
The shape of the ear can be used as a biometric trait or, more precisely, the structure of the cartilage it is made of. Its patterns can either be imaged on the VW with regular cameras, using far-infrared cameras to capture it thermal signature, or imprinting it in the form of *earprints*. The main issue is that all methods require user cooperation up to a certain degree, as *in the wild* they are easily occluded or poorly aligned for proper data to be acquired.

Teeth
The dental structure is unique for each individual. However, its acquisition evolves highly cooperative and intrusive processes. As so, teeth analysis for recognition purposes is used particularly in forensic scenarios where other traits can no longer be acquired.

2.2.5.3 The Ocular Region

The ocular region is one of the most explored in biometry, since not only it contains multiple traits, as they also lead to high accuracy systems. The traits from the ocular region used for biometric recognition are the iris, retina, sclera vein patterns, the periocular region itself, etc.

Iris
The iris has a predominantly randotypic morphogenesis, unique for each individual and allows very high recognition accuracy, which justifies the efforts being held on iris biometrics research and its quick ascent as one of the most popular biometric traits. Being a protected organ visible from the exterior, it can be acquired in a less intrusive way. However, iris performance as a biometric trait is severely impacted in non-ideal setups, and its relatively reduced size and moving profile make it difficult to image at-a-distance and without user cooperation (see Section 2.3.3).

Retina
The retina is the innermost part of the eye, and the blood vessels that pass through it can also be used to uniquely identify an individual. This trait has high distinctiveness, as there are no two persons with the same vein configuration, but its location ans size make it hard to acquire. A high level of cooperation is required, as the user needs to stand still and look through an eyepiece while being illuminated with an infrared beam. On the other side, its extremely hard to forge, and is acquired with low levels of noise.

Periocular
The periocular region represents a good trade-off between the whole face and the iris alone, and it is easy to acquire without user cooperation, not requiring a constrained close capturing.

As so, its use as a biometric trait has emerged and constitutes nowadays a strong alternative for less constrained environments (see section 2.3.4).

2.2.5.4 Medico-chemical

Traits requiring medical-grade sensors to be acquired were classified as *medico-chemical*: deoxyribonucleic acid (DNA), ECG, odor, heart sound, etc.

DNA
The DNA is a molecule contained in every cell of the human body that encodes our genetic instructions. With exception of identical twins, each person's DNA is unique and can be used to unequivocally recognize an individual. Despite its worldwide acceptance by forensic experts, DNA based biometrics are still highly intrusive and time consuming. The acquisition and handling need to be performed with extreme care to avoid contamination. Also, there are some privacy concerns, as DNA analysis can reveal a person's predisposition to certain diseases.

Odor
The odor emanated by an individual is also unique, and can be used to identify him. Existing odor acquiring sensors consist of arrays of chemical sensors, each one sensitive to a particular fragrance. Nonetheless, it is affected by deodorants, perfumes, diets, medicines, etc.

2.2.5.5 Behavioral

Behavioral biometrics are a rather different branch of biometrics since they establish a subject's identity by analyzing the way they behave rather than their physiological attributes. The major behavioral traits in use are: voice, signature, gait, keystroke dynamics, etc.

Voice
The human voice plays an important role on human recognition. Even without seeing the individual, we can identify it just by earing him speak. A person's voice is determined by several physiological factors, like the vocal tracts and the mouth and nasal cavities. The deployed voice authentication systems do not offer enough quality for high-security scenarios, as this trait is prone to interference and hard to acquire without noise even in standard environments. Plus, it can be conditioned by an individual's emotions or medical conditions (e.g., throat related infections).

Signature
Using a person signature to verify his identity is commonly used worldwide, and its so disseminated among users that it is one of the most accepted recognition modalities. However, not only the way an individual signs changes over time and it is influenced by his emotional condition, acquiring it requires substantial user cooperation.

Gait
The way a person walks (gait) is a behavioral trait that can be used for biometric recognition purposes. It is non-invasive, and can be acquired at-a-distance. The majority of the existing recognition methods do not require high resolution data, so they can run over security cameras

located at public locations. As there are some factors that affect the way people walk (e.g., awkward shoes or clothing) the performance of a gait recognition system can also be affected.

Keystroke
The analysis of the keystroke patterns of an individual can also be used to identify him. Among the features available are typing speed, the pace between different letters and typical typing errors. This trait has the advantage of using a regular keyboard as input sensor, not requiring rather expensive acquisition devices. However, the way an individual types on the keyboard lacks in permanence, being affected by the state of mind or the relaxation level. Plus, the same individual can have different performances when using different keyboards.

2.2.5.6 Soft-biometrics

Soft biometrics characteristics are not discriminatory enough to be used for subject identification. However, they can be used along with *hard*-biometrics to improve the recognition rate or speed up the system (e.g., data categorization). Examples of soft-biometrics are gender, ethnicity, height, skin marks (e.g., tattoos), hair / eye color, etc.

2.2.6 Performance Assessment

In order to evaluate the performance of a particular trait or biometric system, several statistical metrics can be used.

Decidability d' was introduced by Daugman [19], and quantifies *intra*- and *inter*-class separability by analyzing their mean μ and standard deviations σ (2.1), given that both distributions are Gaussian. For the sake of clarity, intra-class refers to the comparisons where both the acquired data and the matching template belong to the same person, and inter-class otherwise.

$$d' = \frac{||\mu_{\text{inter}} - \mu_{\text{intra}}||}{\sqrt{\frac{\sigma^2_{\text{inter}}}{2} + \sigma^2_{\text{intra}}}} \quad (2.1)$$

The ROC curve is also a common plot relating the sensitivity, or TPR with the FPR. From that plot, we can take the AUC to quantify how well pairwise comparisons were performed on a binary classification perspective. This metric varies in the $[0, 1]$ interval, being 1 the ideal scenario where all positive matches are ranked higher than the negatives. When setting the operating threshold that determines the accept / reject decision for an equal error on both classes, we obtain the EER.

Sensitivity, or TPR and Specificity (SPC) are given by (2.2) and (2.3) respectively, and relate the correct responses to the total of positive (TP + FN) and negative answers (TN + FP).

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2.2)$$

$$\text{SPC} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (2.3)$$

The ACC (2.4) is used to express the overall ratio of correctly classified matches. Its maximum value is one, expressing the optimal scenario where all classes have been correctly classified.

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}} \quad (2.4)$$

For a balanced analysis in situations where there is an high discrepancy between the amount of positive and negative matches, the MCC can be used (2.5). It varies in the $[-1, 1]$ interval, being one the optimal value [20].

$$\mathrm{MCC} = \frac{\mathrm{TP} \cdot \mathrm{TN} - \mathrm{TP} \cdot \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}} \quad (2.5)$$

2.3 Towards Non-Cooperative Scenarios

Several authors have stressed the main issues associated with non-cooperative environments. Fancourt *et al*. [21] concluded that is possible to acquire sufficiently high-quality iris data at distances up to 10 meters Smith *et al*. [22] comparatively examined iris data captured in both the near-infrared (NIR) and VW spectra, addressing the possibility of combining that multi-spectral data to improve recognition performance. In our work [4] we addressed iris usability for recognition purposes on the visible wavelength, quantifying the conditions that allow that process with enough confidence. We conclude on the significant impact of the luminance level that should be no lower that $120cd/m^2m$, in opposition to the illuminant being used that was less relevant. Ross *et al*. [23] addressed the problem of biometric recognition over degraded iris images, having authors considered five factors: 1) non-uniform illumination, 2) motion, 3) defocus blur, 4) off-axis gaze, and 5) nonlinear deformations. The key insight for the proposed method is that a single-feature encoding schema does not appropriately handle all these variations. Having that in mind, authors propose three feature extraction / matching strategies: 1) gradient orientation histograms; 2) scale invariant feature transforms; and a 3) probabilistic deformation model. The information extracted by each descriptor is independently matched, and results are combined at the score level using the classical sum-rule. Experiments on the FOCS and FRGC data sets encourage further work on this kind of hybrid techniques [24]. The noising factors are further described and illustrated at Chapter 11 Section 11.4.2 and Figure 11.3.

The effectiveness of face recognition systems is significantly decreased by several factors [24]:

1. its 3D structure introduces substantial differences in appearance, with respect to subject's pose;
2. large regions of the face are often occluded (e.g., non-orthogonal imaging);
3. its appearance is notoriously affected by facial expression;
4. can be easily disguised.

Ultimately, unconstrained scenarios could not allow the proper acquisition of the iris, nor the full facial picture. As so, growing attention has been paid to other potential traits. The use of information in the vicinity of the eye has been gaining particular interest and popularity, as it represents a good trade-off between the whole face and the iris alone [24].

2.3.1 Mobile Setups

The usage of mobile devices has substantially grown over the last years, along with their capabilities and applications. Deploying biometric technologies to such gadgets is not only quite desirable, but would also represent the ability to deliver *off-the-shelf* solutions for everyday consumers, extending biometric recognition everywhere, anytime, and to anyone.

Most of the algorithms designed for *in the wild* operation can run effortlessly on everyday electronics. However, when attempting to perform iris or periocular biometrics on mobile environments, several problems arise: the wide variety of camera sensors and lenses mobile phones and tablets come equipped with produce discrepancies in working images, as they are acquired with color distortions, at multiple resolutions, etc; on-the-go acquisition by potentially untrained subjects will result in demanding Pose, Illumination and Expression (PIE) changes (e.g., not all users hold their mobile devices at the same position), resulting in varying acquisition angles and scales, or rotated images; the acquisition environment can have poor or insufficient lighting, and uncontrolled outdoor daylight will most likely produce spectacle reflections over the iris region; etc.

Chapter 10 details our endeavors on tackling those issues: the acquisition of an iris and periocular database with 10 different mobile setups; reports on the application of a device-specific color calibration technique that compensate for the different color perception inherent to each setup; and the application of well-know iris and periocular recognition strategies to such data, evidencing how they can be fused to achieve biometric recognition over mobile setups.

2.3.2 Stages of Unconstrained Recognition Systems

When designing a system for unconstrained scenarios we can extend the previously mentioned four-step approach, devising a seven-step system as depicted in Figure 2.6:

1. **Trait acquisition** The acquisition sensor depends on which trait the system is going to deal with. In the specific case of unconstrained biometrics, we should rely on sensors that do not require user interaction, mostly cameras. Also, when aiming at real-time systems, we must consider the trade-off between the richness in detail of the acquired data and the acquisition rate.

2. **Quality assessment** Even the most tunned sensor setup does not always return optimal data from which reliable features can be extracted. Quantifying the quality of the acquired data, discarding unfit images will decrease both the overall processing time, and the system error rates.

3. **Signal enhancement** Complementary to quality assessment, the acquired signal can be enhanced if it is known which particular degradation factor is affecting it. Even if the original signal cannot be restored, this step is a good alternative to increase system performance, or to when few quality data is being acquired.

4. **Detection and Segmentation** The acquired trait needs to be properly located and segmented, so that feature extraction can be carried on effectively. A proper segmentation stage allows discarding noise factors not discarded during signal enhancement (e.g., eyelashes occluding the iris).

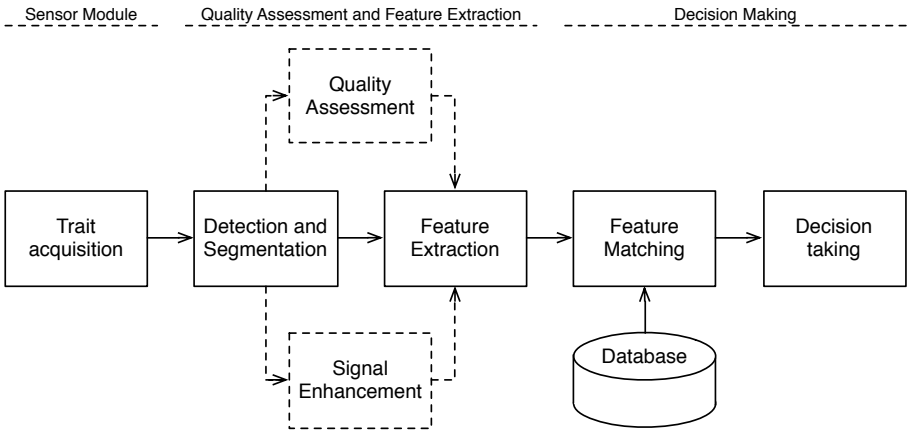


Figure 2.6: Illustration of the main steps and elements of a biometric recognition system.

5. **Feature extraction** A representative set of features needs to be extracted in such a way that discriminant user information is collected. On real-time systems, user features should be quickly extracted and encoded with minimum computational burden. During the enrollment stage, the gathered information is stored in the system database for later identification or verification of their identities.

6. **Feature matching** Features are then matched against the templates on the database. On real-time systems, not only the information on the database should be optimized for fast access, but also soft-biometrics are used to narrow the set of identities being matched.

7. **Decision making** Based on the results of feature matching stages (multiple possible results, if using multiple traits), a final decision about subject's identity is reached. Eventually, the system can be unable to positively identify a particular subject, but still be able to reduce the set of possible identities.

2.3.3 The iris as biometric trait

The data from the human iris can be acquired contactless at-a-distance, and it delivers great performance on constraint environments, making it a potential trait for unconstrained scenarios and justifying the efforts on "relaxing" its acquisition setup [9, 25, 26].

Iris effectiveness as a biometric trait has been proven, being widely deployed on quite constrained scenarios: subjects are required to stop-and-stare relatively close to the acquisition device, while their iris is imaged on the NIR slice of the electromagnetic spectrum (700 a 900 nm). NIR illumination enables the acquisition of good quality data, while avoiding main noise factors typically associated with the VW light imagery (e.g., reflections). Nonetheless, unconstrained scenarios involve iris acquisition at significantly larger distances and on moving targets, demanding cameras with simultaneously high f-numbers and short exposure times for an acceptable depth-of-field to be obtained. Since there is a direct relation between those values and the amount of the light on the scenery, to operate in such conditions the system would need strong NIR illumination, thus being hazardous as the human eye does not instinctively respond with its natural protection mechanisms (e.g., blinking and pupil contraction). The feasibility of iris recognition in VW conditions remains controversial, in particular for highly pigmented irises (the majority of world's population).

2.3.3.1 Iris Segmentation

Many iris segmentation techniques are based on Hough-transform parametrization.

Junli et al. [27] developed a ellipse fitting technique particularly adapted to degraded data due to its robustness to noisy edge-maps. The algorithm starts by selecting a subset of more accurate edge-points, which contribution is magnified by squaring the fitting residuals. Finally, solution is found using the computationally efficient interior-point methods.

To attenuate the issues inherent to images acquired at large distances, Tan and Kimar [28] propose a method based in the grow-cut algorithm that is able to discriminate between iris and non-iris data. Results are further refined through: iris center estimation, boundary refinement, pupil masking and refinement, eyelashes and shadow removal, and eyelid localization. The effectiveness of such approach was validated against well known datasets (UBIRIS.v2, FRGC and CASIA.v4 Distance), with lower computational burden than similar strategies.

Alonso-Fernandez and Bigun [29] perform iris segmentation based on the Generalized Structure Tensor algorithm: using complex filters authors are able to obtain both magniture and orientation information for each edge pixel, providing additional information to better discriminate between edges belonging to the iris boundaries.

Xinyu et al. [30] proposes an algorithm able to segment iris on images with very distinct resolution (iris diameter from 50 to 360 pixels). They start by detecting a set of edges using Canny detector, which non-connected components are used as nodes of a graph. They further use the normalized cuts criterion to discriminate between the most probable iris boundaries.

2.3.3.2 Noise Detection

When dealing with unconstrained environments it is particularly important to have a noise detection module, so that one can estimate the regions of the iris that are occluded by other types of information (e.g., eyelids, eyelashes, reflections). That will allow handling them prior to the feature encoding phase, otherwise increasing the false rejection rates [24].

Several authors addressed this problem, most of them with rule-based approaches and thus of questionable effectiveness. On that basis, Li and Savvides [31] used Gaussian Mixture Models to model probabilistic distributions of noise-free and noise regions of the irises, adjusting the number of Gaussians for a distribution in such way that those not supported by the observations were eliminated. Authors propose Gabor filters as basic features, optimized by a simulated annealing process.

2.3.3.3 Quality Assessment

Indexing the quality of the data being handled improves the performance of the biometric recognition system, as shown by Zuo and Schmid [32] who proposed three quality indexes:

1. quality of sample: by adaptively filtering the probe biometric data based on predicted values of Quality of Sample index (defined here as d-prime);

2. confidence in matching score: by adaptively filtering the matching scores based on predicted values of Confidence in matching Scores index (defined here as d-prime);

3. quality sample and template features: considering image quality measures as features for discriminating between genuine and impostor's matching scores.

The proposed algorithm has the advantage of being generic, and thus suitable for other biometric modalities.

2.3.3.4 Iris Recognition

The iris texture is characterized by the irregular distributions of local features such as furrows, crypts, freckles or spots. Hence, traditional approaches are texture-based and decompose the iris data into a set of numerical coefficients corresponding to specific orientations or frequencies of predominant patterns.

Commercially deployed iris recognition systems are based mainly on Daugman's pioneering approach [19], with extremely high performance over NIR data. Despite a few innovations [33] the process consists in the translation of the segmented iris to a pseudo-polar coordinate system, attaining invariance to both scale and translation. This normalized data is convolved with a set of Gabor filters, at multiple frequencies and orientations, and their output quantized into one of four quadrants, extracting two bits of phase information per convolution. For signature matching the factional Hamming distance is used, and several comparisons of shifted data are performed to achieve invariance to rotation.

Other classical iris recognition methods are based in zero-crossing, as Boles and Boashash [34] who use 1D wavelets over different resolutions of concentric circles, or iris textural analysis, as Wildes [35] whose encoding was based on the Laplacian pyramid. Nonetheless, any of those systems require high user cooperation.

The recognition in less constrained environments has been gaining relevance, being the *Iris-on-the-Move* project [36] a major example of engineering and image acquisition system that makes iris recognition less intrusive for subjects: NIR close-up iris images are acquired as subjects walk at normal speed through an access control point. *Honeywell Technologies* applied for a patent [37] on a similar system, which was also able to recognize irises at a distance.

2.3.4 The Periocular Region

The area in the vicinity of the eye is designated as *periocular* region. Its use for biometric purposes was inspired by human ability to recognize each other by his eyes. This region is particularly useful on less constrained scenarios, where image acquisition is unreliable, to avoid iris pattern spoofing, compensate for environmental adversities or cooperative subjects. The usage of periocular information has even proven itself to be of importance in scenarios where the face has been reshaped (e.g., plastic surgery), with interesting results [38, 39]. It does not require constrained close capturing or user cooperation, is relatively stable (when compared to the whole face), and rarely occluded. Also, having the iris in its middle, both can be acquired simultaneously with a single camera, and fused either at feature or score levels [24].

Chapters 5 and 6 provide a comprehensive summary of the most relevant research conducted in the scope of periocular recognition methods, comparing the main features of the publicly available datasets, and presenting state-of-the-art results and current issues on this topic.

2.4 Privacy and Ethical Concerns

Along with the developments on biometrics a lot of privacy and ethical concerns arise, specially on the field of covert biometrics. If by one side biometrics are unique and personal, and a strong way of authentication, on the other side if that information becomes compromised it cannot be changed like a regular password or PIN code. When advancing to unconstrained environments, further ethical problems arise, as individuals can engage on the biometric recognition process even without their acknowledgment or consent.

From the point of view of the European law, there are five principles that must be considered in order to deploy real-world biometric applications [40, 41]:

Proportionality principle

One should assess if the use of biometric traits is really necessary on the context of the application, or if other authentication methods can be applied to the same purpose, that do not require dealing with biometric or personal data.

Potential Risk of Discrimination

The deployed system should not evidence false positives or negatives, nor failure to enroll.

Improper Use/Scope Limitation/Function Creep

The gathered biometric data should not be put to other uses than the one it was originally acquired for. Furthermore, there must be an explicit consent of the user for his biometric data to be used on the system and to that purpose.

Possible covert obtaining of biometric data and monitoring

Users should be informed and aware that their data is being collected, or otherwise a legal provision must exist so that the collected data can be used.

Specific data warranting protection

The acquired biometric data should not give additional information other than the one used for identification purposes (e.g., health or ethnic data), or otherwise that must be covered by the explicit consent.

Furthermore, a set of PET measures to protect privacy were enforced by the CEN through the ISSS initiative on privacy standardization in Europe [42]. Examples of such measures include: the minimization of the amount of personal data in use, its proper encryption, and prompt deletion when no longer needed; avoid to use central databases, which are actually prohibited in some countries (e.g., Germany); etc.

Even so, people are more or less familiar with the usage of video surveillance systems on public places. The United States of America, where the right to privacy is also established by constitution and where the 9-11 attacks were a mind-turning event, is a good example on how the “reasonable expectation of privacy” can shift on a very short period. People’s “reasonable

expectation of privacy" regarding features that are exposed to the public are found to be considerably low, and the choice to do not be engaged on biometric procedures could not always be reasonable. At an airport, for instance, a person will not be able to travel if do not agree with engaging in a biometric recognition process [43]. In this context, the usage of a negative recognition approach (or screening) would be more easily accepted, as the data stored on the database is only from wrongdoing individuals and the gathered data should not be stored for further usage.

Ultimately, and regardless of the existing security directives and protocols, no or few guarantees are that those procedures will be followed or the system will not be put to misuse.

Chapter 3

Iris Recognition: Preliminary Assessment about the Discriminating Capacity of Visible Wavelength Data

3.1 Overview

This chapter consists of the following article:

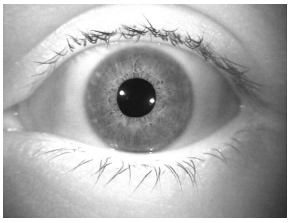
Iris Recognition: Preliminary Assessment about the Discriminating Capacity of Visible Wavelength Data
Gil Santos, Marco V. Bernardo, Hugo Proença and Paulo T. Fiadeiro
6^{th} *IEEE Workshop on Multimedia Information Processing and Retrieval - MIPR '10*, December 13-15, 2010, Taiwan.
ISBN: 978-0-7695-4217-1

Iris Recognition: Preliminary Assessment about the Discriminating Capacity of Visible Wavelength Data

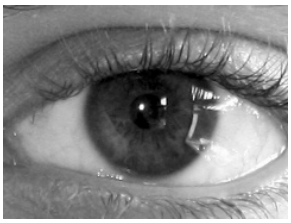
Gil Santos*, Marco V. Bernardo*, Hugo Proença* and Paulo T. Fiadeiro†
* *Dept. of Informatics, Instituto de Telecomunicações - Networks and Multimedia Group*
† *Dept. of Physics, Remote Sensing Unit - Optics, Optometry and Vision Sciences Group*
University of Beira Interior, Covilhã, Portugal
Email: gmelfe@ubi.pt, mbernardo@ubi.pt, hugomcp@di.ubi.pt, fiadeiro@ubi.pt

***Abstract*—The human iris supports contactless data acquisition and can be imaged covertly. These factors give raise to the possibility of performing biometric recognition procedure without subjects' knowledge and in uncontrolled data acquisition scenarios. The feasibility of this type of recognition has been receiving increasing attention, as is of particular interest in visual surveillance, computer forensics, threat assessment, and other security areas. In this paper we stress the role played by the spectrum of the visible light used in the acquisition process and assess the discriminating iris patterns that are likely to be acquired according to three factors: type of illuminant, it's luminance, and levels of iris pigmentation. Our goal is to perceive and quantify the conditions that appear to enable the biometric recognition process with enough confidence.**

***Keywords*-biometrics, iris recognition, visible light data, controlled standard illumination**



(a) NIR image, acquired under highly constrained conditions (WVU database [2]).



(b) VW image, acquired under less constrained conditions (UBIRIS database [3]).

Figure 1. Comparison between the appearance of NIR and VW images.

I. Introduction

Due to the effectiveness proven by the deployed iris recognition systems, the popularity of the iris as biometric trait has considerably grown over the last few years. A number of reasons justify this interest: it is a naturally protected internal organ visible from the exterior, it has a near circular and planar shape and its texture has a predominantly randotypic chaotic appearance. The typical scenarios where iris recognition systems were successfully deployed are quite constrained: subjects stop-and-stare relatively close to the acquisition device while their eyes are illuminated by a *near infrared* (NIR) light source that enables the acquisition of good quality data. Recently, several research initiatives have sought to increase capture distance and relax constraints on iris acquisition systems, making use of *visible wavelength* (VW) light imagery to covertly perform data acquisition (e.g. [1]), which broads the iris recognition applicability to forensic domains where the cooperation of the subjects is not expectable.

It is known that the VW light imagery engenders notorious differences in the appearance of the captured data when compared with the traditional NIR constrained setup (Figure 1). However, the acquisition of iris data from significantly larger distances and on moving targets demands simultaneously high f-numbers and very short exposure times for the optical system, in order to obtain acceptable depth-of-field values. These are in direct proportion with the amount of light required to proper imaging, which is a concern as excessively strong illumination can cause permanent eye damage. The American and European standards councils ([4] and [5]) proposed safe irradiance limits for NIR illumination of near $10\ mW/cm^2$. The NIR wavelength is particularly hazardous, because the eye does not instinctively respond with its natural mechanisms (aversion, blinking and pupil contraction).

The Eumelanin molecule is predominant in the human iris pigment [6] and has most of its radiative fluorescence under the VW light, which enables the capture of a much higher level of detail, but also of many more noisy artifacts: specular and diffuse reflections and shadows. Also, the spectral radiance of the iris in respect of the levels of its pigmentation varies much more significantly in the VW than in the NIR (Figure 2). These biological features usually engender acquired data with several other types of information within the iris rings: eyelashes, eyelids, specular and diffuse reflections obstruct portions of the iris texture and increase the challenges in performing accurate recognition.

The feasibility of the VW iris recognition remains controversial — specially for high pigmented irises that constitute the majority of the world's population — and fundamental research remains to be done. Hence, this paper gives — whenever possible — preliminary assessments about the

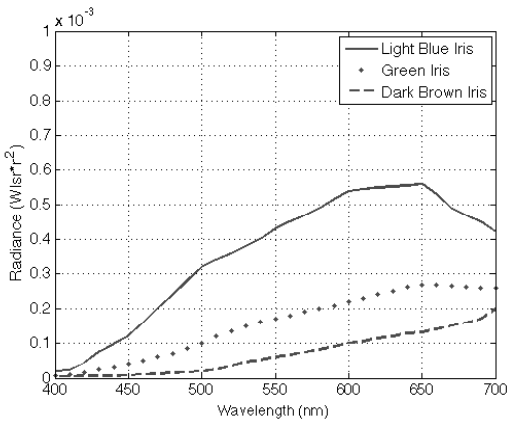


Figure 2. Spectral radiance of the human iris according to the levels of iris pigmentation [7].

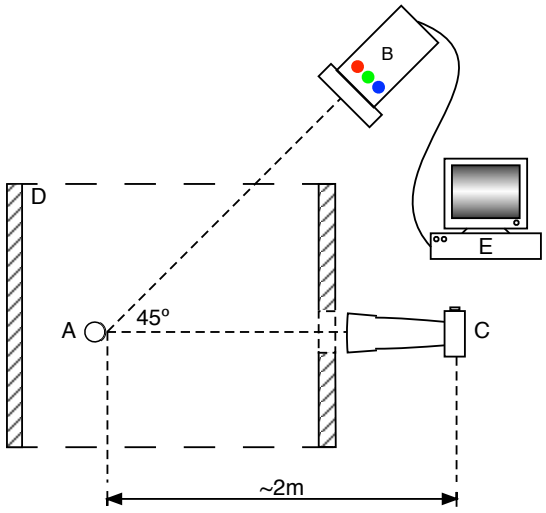


Figure 3. Acquisition setup. Subject (A) was positioned within an acquisition scene (D) and required to look forward, towards the camera (C) while illuminated at 45° by the RGB projector (B) controlled through a Visual Stimulus Generator (E).

amount of discriminating data able to be captured in such acquisition setup, regarding the spectrum and intensity of the used light and the levels of iris pigmentation. Also, we report the biological structures of the iris that are most likely to be used in discriminating between individuals in the VW acquisition scenario.

The remainder of this paper is organized as follows: Section II overviews mainstream iris recognition techniques and efforts towards the reduction of the involved constrains; Section III describes the image acquisition framework and setup used for the purposes of this paper; Section IV reports the performed experiments and discusses our results; Finally, Section V concludes and points further work.

II. Iris Recognition

Deployed iris recognition systems are mainly based on Daugman's pioneering approach [8], and have proven their effectiveness in relatively constrained scenarios where images are acquired in the NIR spectrum (700-900 nm). Regardless a few innovations [9], the process consists in the segmentation of the iris pupillary and limbic boundaries followed by the translation into a double dimensionless pseudo-polar coordinate system, that gives invariance to scale and translation. This normalized data is convolved with a set of Gabor filters at multiple frequencies and orientations and the corresponding output quantized to one of four quadrants, extracting two bits of phase information per convolution. The fractional Hamming distance is used to match iris signatures and several comparisons of shifted data are performed to achieve invariance to rotation.

The acquisition constraints for effective recognition have been motivating serious research efforts. The "Iris-on-the-move" project [10] should be emphasized: It is a major example of engineering an image acquisition system to make the recognition process less intrusive for subjects. The goal is to acquire NIR close-up iris images as a subject walks at normal speed through an access control point. Previously, Fancourt et al. [11] concluded that it is possible to acquire sufficiently high-quality images at a distance of up to 10 meters. Smith et al. [12] examined the iris information that could be captured in the NIR and VW spectra, addressing the possibility of using these multispectral data to improve recognition performance.

III. Image Acquisition Framework

The conceived acquisition setup was the one presented at Figure 3.

The subjects were placed in a dark acquisition scene, with their heads steady on a chin rest, and required do look forward towards the camera, which was placed at two meters and aligned with their right eyes. Varying the scene illuminant, produced by a *Barco RLM G5i Performer* (Barco Corporation, Belgian) RGB projector through a *Visual Stimulus Generator* (VSG) 2/5 (Cambridge Research Systems, UK), images were captured by the Canon EOS 5D camera using the configuration in Table I.

Previously, the VSG generated stimulus were verified and calibrated using a telespectroradiometer (PR-650 *Spectra-Colorimeter* ^{TM}- Photo Research, Inc., CA) and a reference white $BaSO_4$ sample placed on the chin rest. The maximum errors allowed were $0,002$ illuminant chromaticities in the *Commission Internationale de l'Eclairage* (CIE) 1931 (x, y) space and $1\ cd/m^2$ for luminance.

In order to mimic incandescent light, different phases of the daylight and fluorescent lamps, illuminants CIE A, D and F were picked as specified by the CIE 1931 standard colorimetric observer (2°) [13], [14]. Illuminants' luminance was also controlled, regulated in uniform steps of $20\ cd/m^2$,

Table I
DETAILS OF IMAGE ACQUISITION SETUP.

Image Acquisition Framework and Setup	
Camera = Canon EOS 5D	Format = tiff
Color Representation = sRGB	F-Number = f/5.6
Shutter Speed = 1/8 sec	ISO Speed = ISO-400
Focal Length = 400 mm	Metering Mode = Spot
Width = 4368 pixels	Height = 2912 pixels
Resolution = 240 dpi	Bit Depth = 24 bit
Details of the Manually Cropped Resultant Images	
Width = 800 pixels	Height = 600 pixels
Format = tiff	Resolution = 240 dpi

Table II
GENERATED ILLUMINANTS AND THEIR RESPECTIVE LUMINANCE. COMBINATIONS REPRESENTED BY × WERE NOT REPRODUCED AS THEY WERE OUT OF GAMUT ON THE RGB PROJECTOR.

	Luminance (cd/m^2)					
Illuminant	20	40	60	80	100	120
A	✓	✓	✓	×	×	×
D55	✓	✓	✓	✓	✓	×
D65	✓	✓	✓	✓	✓	✓
D75	✓	✓	✓	✓	✓	✓
F11	✓	✓	✓	✓	×	×

from 20 to 120 cd/m^2, producing the configurations shown in Table II.

According to this setup, 5 samples were consecutively acquired in each of the 24 scenarios (a specific illuminant at a certain luminance), delivering a total of 720 images. Such images came from 6 different subjects $\{S_1, ..., S_6\}$, equally divided into light, medium and heavily pigmented irises.

IV. EXPERIMENTS

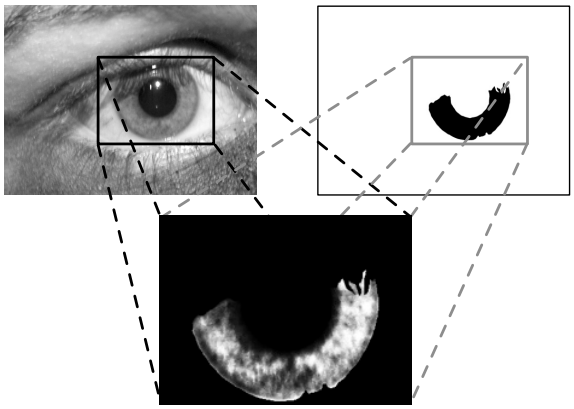
All irises were manually segmented (Figure 4a), avoiding that eventual errors were carried on to further experimental stages.

Both the pupillary and limbic iris boundaries were modeled as circles, and data was translated into a pseudo-polar dimensionless coordinate system (Figure 4b) in order to obtain invariance to scale and translation.

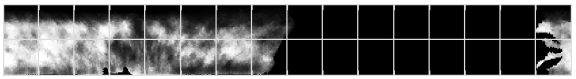
A. Visual Inspection

When comparing to the currently deployed iris recognition systems, there is a complexity gain in the VW image acquisition, as data become represented along three axis (usually Red, Green and Blue — RGB) instead of a single one.

However, our earliest analysis did not make use of this higher amount of information. Aiming at study the luminance bounds that — under visual inspection — enable the capturing of discriminating patterns in the iris data, we found appropriate to reduce data dimensionality to the luminance channel (Y) of the YCbCr colorspace [15]. Later, we performed an equalization of each image histogram, so that the iris structure could be better distinguished.



(a) Merging of original image with manual segmentation.



(b) Normalized iris on pseudo-polar coordinates.

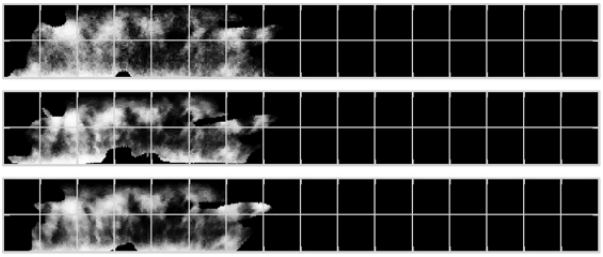
Figure 4. Illustration of the steps taken prior to visual inspection and color analysis.

We concluded that the easiness of detecting discriminating iris patterns varies proportionally to the illuminants' luminance level used in the acquisition, and is almost invariant to the type of used illuminant. For light pigmented irises, all the luminance levels on trial appeared to be propitious (Figure 5a). Oppositely, for highly pigmented ones, the detection of discriminating patterns has revealed as a much more difficult task, even using 120 cd/m^2 (Figure 5c). Medium pigmented irises have intermediate behavior, and luminance values higher than 60 cd/m^2 propitiate the capturing of discriminating iris patterns (Figure 5b). According to these observations, we found appropriate to define relatively large bounds for the amounts of light that enable the capturing of discriminating iris patterns, as given in Figure 6.

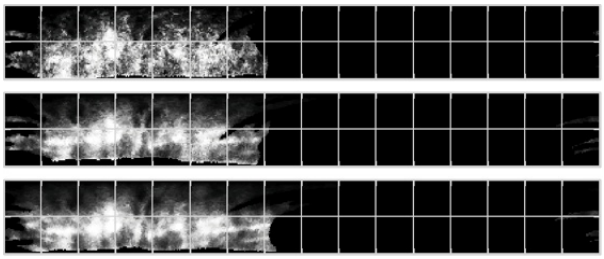
Also, it should be taken into account that the quantity of discriminating patterns able to be perceived under visual inspection is highly varying between different irises, although crypts and freckles were observed to be the most likely used to discriminate between individuals. For such, these bounds are regarded as rough initial values, that should be finely adjusted by further analysis.

Table III
CIELAB $\bar{\Delta}E^*_{ab}$ ON POINT-TO-POINT IMAGE COMPARISONS, IN DIFFERENT ILLUMINANTS FOR BOTH INTRA- (I) AND INTER- (II) CLASSES, AT 60 cd/m^2. VALUES ARE PRESENTED FOR AN 95% CONFIDENCE INTERVAL.

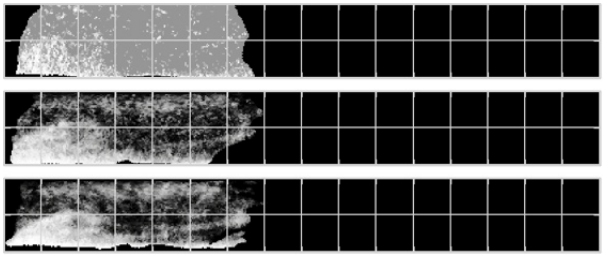
Illum.	Class	Heavy Pigmentation		Light Pigmentation		Medium Pigmentation	
		Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6
A	I	14,08 ± 4,01	14,15 ± 4,91	18,28 ± 4,57	17,71 ± 4,57	14,26 ± 4,20	14,19 ± 4,08
	II	22,06 ± 10,32	27,29 ± 11,17	19,11 ± 8,27	18,84 ± 8,15	23,46 ± 10,23	24,68 ± 9,95
D55	I	7,44 ± 4,91	8,03 ± 5,71	8,86 ± 7,84	9,08 ± 7,59	7,62 ± 6,15	8,22 ± 6,13
	II	17,51 ± 7,62	24,13 ± 9,35	27,87 ± 10,00	23,78 ± 9,79	15,42 ± 7,29	16,24 ± 7,37
D65	I	7,59 ± 5,79	8,10 ± 6,00	8,64 ± 7,90	8,55 ± 6,84	7,16 ± 5,93	7,16 ± 5,70
	II	17,74 ± 7,22	24,28 ± 8,89	28,21 ± 10,05	22,36 ± 9,28	15,49 ± 7,42	16,50 ± 7,57
D75	I	8,71 ± 6,27	8,83 ± 6,37	8,44 ± 7,61	8,72 ± 7,27	7,14 ± 5,82	7,25 ± 5,70
	II	17,84 ± 6,94	24,10 ± 8,66	27,77 ± 10,03	23,55 ± 9,76	15,53 ± 7,39	16,49 ± 7,54
F11	I	8,62 ± 2,64	8,46 ± 4,94	10,61 ± 1.09	10,12 ± 1,47	8,57 ± 1,72	8,43 ± 1,82
	II	18,77 ± 8,71	25,59 ± 10,23	21,98 ± 8,75	19,22 ± 7,87	18,27 ± 8,89	19,18 ± 8,76



(a) Light pigmented iris.



(b) Medium pigmented iris.



(c) Heavy pigmented iris.

Figure 5. Y channel histogram-equalized samples for different pigmentation levels, under CIE D65 illuminant, with 20, 60 and 120 cd/m^2 (top to bottom).

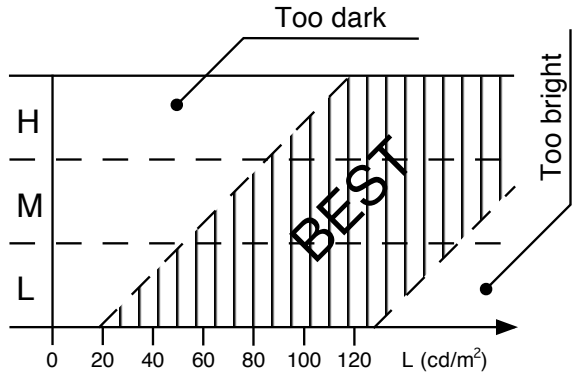


Figure 6. Best perceived luminance levels for high (H), medium (M) and low (L) pigmented irises.

B. Color Analysis

Normalized images were converted to the CIE 1976 ($L^*a^*b^*$) colorspace, also known as CIELAB, which is device independent, partially uniform and based on the human visual system, allowing absolute color representation according to the illuminant. We obtained the chromatic adaptation matrices using projector's CIE 1931 (x,y,Y) phosphor coordinates and RGB coordinates of the reference white for each illuminant. We selected the images captured at an intermediate luminance level (60 cd/m^2), and performed a pixel-to-pixel color difference (chromatic error) between image pairs I_1 and I_2 as Equation 1.

$$\Delta E^*_{ab} = \sqrt{(L^*_1 - L^*_2)^2 + (a^*_1 - a^*_2)^2 + (b^*_1 - b^*_2)^2} \quad (1)$$

Results were grouped into two distributions: intra-class for comparisons between the same eye and inter-class for different eyes, both captured in different illuminants. Inspecting those results, we observed that they fit the normal distributions detailed at Table III, also described through the *Receiver Operator Characteristic* (ROC) curves of Figure 7 and Table IV.

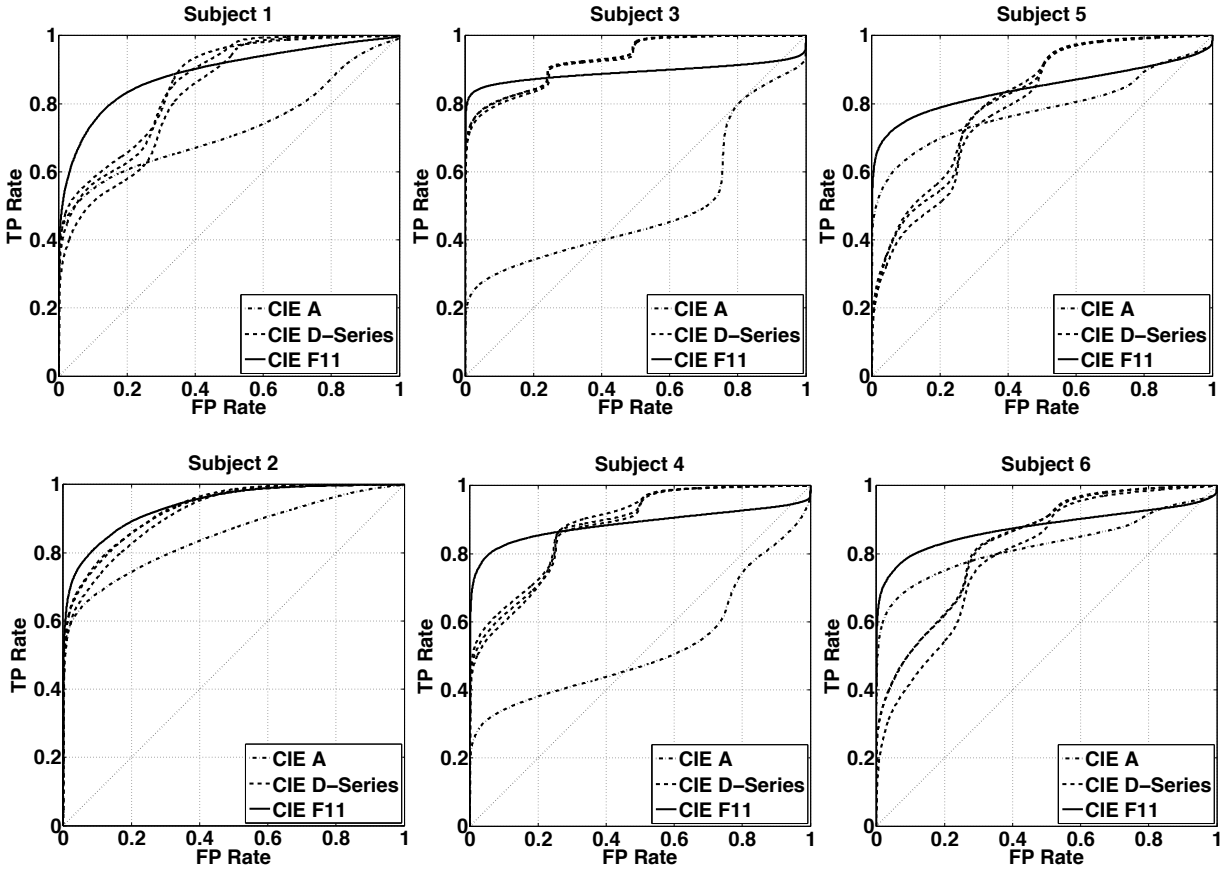


Figure 7. ROC Curves for all subjects, on different illuminants, at 60 cd/m^2.

Table IV
AREA UNDER ROC CURVE FOR ALL SUBJECTS ON DIFFERENT ILLUMINANTS, AT 60 cd/m^2.

	S 1	S 2	S 3	S 4	S 5	S 6
A	$0,724$	$0,850$	$0,501$	$0,527$	$0,781$	$0,824$
D55	$0,862$	$0,926$	$0,930$	$0,877$	$0,791$	$0,795$
D65	$0,855$	$0,930$	$0,934$	$0,880$	$0,808$	$0,834$
D75	$0,827$	$0,916$	$0,934$	$0,884$	$0,813$	$0,832$
F11	$0,890$	$0,940$	$0,892$	$0,886$	$0,845$	$0,876$

It can be seen that results obtained for the CIE D55, D65 and D75 illuminants were very similar. The poorest matches occurred when data captured with a CIE D is compared to data captured with a CIE A illuminant. The CIE F11 illuminant, with a *Correlated Color Temperature* (CCT) between the ones of CIE D-Series and CIE A, is the one with better overall performance.

The upper image of Figure 8 illustrates the pixel usage in the obtainance of the pixel-to-pixel color distances.

Here, the darkest region at the left part of the normalized image (which corresponds to the lower part of the cartesian data), is the one more frequently considered as noise-free. The brightest part by other side, commonly occluded, was

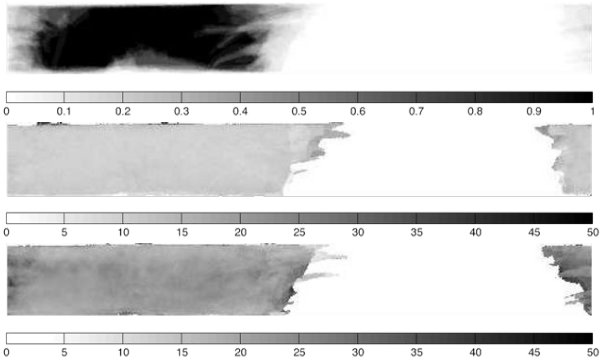


Figure 8. Pixel usage probability (top) and $\bar{\Delta}E^*_{ab}$ values per pixel for intra (middle) and inter (bottom) classes.

rarely used for comparison purposes. The middle and bottom images give the ΔE^*_{ab} average values ($\bar{\Delta}E^*_{ab}$), enabling us to perceive the relative contribution of each iris region to the intra-class (middle image) and inter-class (bottom image) distance values. Relatively homogeneous values were obtained in both cases, with exception to the pupillary region, where the average chromatic error values $\bar{\Delta}E^*_{ab}$ of

the inter-class comparisons tend to vanish close to the intra-class values. The obtained $\bar{\Delta}E^*_{ab}$ values were $9,82 \pm 6,36$ for the intra-class and $21,21 \pm 9,77$ for the inter-class comparisons, with a 95% confidence interval. Thus, even an extremely simple image matching technique is able to produce a clear separability between the two classical types of comparisons.

V. Conclusions and Further Work

Although preliminary, performed experiments allowed us to conclude that the appearance of the captured iris pattern is poorly conditioned by the type of VW illuminant used in the acquisition, in opposition to the levels of luminance, that play a much more significant role.

The obtained pixel-to-pixel ΔE^*_{ab} values appear to confirm that color information contained in VW data can be used to discriminate between different irises. Also, the CIE F11 illuminant should be preferably used in data enrollment, as it was the one that propitiated more compatibility between iris patterns acquired with all the remaining illuminants.

As further work, we plan to increase the statistical relevance of the described experiments — with both the inclusion of more test subjects and a subsequent higher range of pigmentation levels — and make use of more types of illuminants, so that CCT differences between them become more uniform and a larger area in the Planckian locus will be covered.

Acknowledgment

The authors are very grateful for the support given by the Optical Center of the University of Beira Interior, and also for the financial support given by "FCT-Fundação para a Ciência e Tecnologia" and "FEDER" in the scope of the PTDC/EIA/69106/2006 ("BIOREC: Non-Cooperative Biometric Recognition") and PTDC/EIA-EIA/103945/2008 ("NECOVID: Covert Negative Biometric Identification") research projects.

References

[1] C. Boyce, A. Ross, M. Monaco, L. Hornak and Xin Li. Multispectral Iris Analysis: A Preliminary Study. *Proceedings of Computer Vision and Pattern Recognition Workshop on Biometrics (CVPRW)*, New York, USA, June 2006.

[2] A. Ross, S. Crihalmeanu, L. Hornak, and S. Schuckers. A centralized web-enabled multimodal biometric database. In *Proceedings of the 2004 Biometric Consortium Conference (BCC)*, U.S.A, September 2004.

[3] H. Proença, S. Filipe, R. Santos, J. Oliveira and L. A. Alexandre The UBIRIS.v2: A Database of Visible Wavelength Iris Images Captured On-The-Move and At-A-Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 99*, 2009.

[4] American National Standards Institute American national standards for the safe use of lasers and LEDs used in optical fiber transmission systems, 1988. ANSI Z136.2.

[5] Commission International de l'Eclarirage Photobiological safety standards for lamps, 1999. Report of TC 6-38; CIE 134-3-99.

[6] P. Meredith and T. Sarna. The physical and chemical properties of eumelanin. *Pigment Cell Research, vol. 19*, pp. 572-594, 2006.

[7] F. Imai. Preliminary experiment for spectral reflectance estimation of human iris using a digital camera. Technical report, Munsell Color Science Laboratories, Rochester Institute of Technology, 2000.

[8] J. Daugman. How Iris Recognition Works. *IEEE Trans. CSVT 14(1)*, pp. 21-30, 1993.

[9] J. Daugman. New methods in iris recognition. *IEEE Trans. Systems, Man, Cybernetics B 37(5)*, pp. 1167-1175, 2007.

[10] J.R. Matey, D. Ackerman, J. Bergen, and M. Tinker. Iris Recognition in Less Constrained Environments. *Advances in Biometrics: Sensors, Algorithms and Systems*, pp. 107-131, Springer, Oct. 2007.

[11] C. Fancourt, L. Bogoni, K. Hanna, Y. Guo, R. Wildes, N. Takahashi, and U. Jain. Iris Recognition at a Distance. *Proc. 2005 IAPR Conf. Audio and Video Based Biometric Person Authentication*, pp. 1-13, July 2005.

[12] K. Smith, V.P. Pauca, A. Ross, T. Torgersen, and M. King. Extended Evaluation of Simulated Wavefront Coding Technology in Iris Recognition. *Proc. First IEEE Int'l Conf. Biometrics: Theory, Applications, and Systems*, pp. 1-7, Sept. 2007.

[13] CIE Colorimetry. *CIE Publication 15:2004*, (CIE, Vienna, 2004).

[14] T. Smith and J. Guild. The C.I.E. colorimetric standards and their use. *Transactions of the Optical Society 33 (3): pp. 73-134*, 1931-32.

[15] A. Poynton. A technical introduction to digital video. John Wiley & Sons, Inc, New York, USA, 2006.

Chapter 4

Fusing Color and Shape Descriptors in the Recognition of Degraded Iris Images Acquired at Visible Wavelengths

4.1 Overview

This chapter consists of the following article:

Fusing Color and Shape Descriptors in the Recognition of Degraded Iris Images Acquired at Visible Wavelengths
Hugo Proença and Gil Santos
Computer Vision and Image Understanding, 116, 167-178, 2012.

According to SCImago Journal & Country Rank, this journal's index[1] for the 2013 year are as follows:

Category	*Quartile*	*SJR*
Computer Vision and Pattern Recognition	Q1	1,462
Signal Processing	Q1	
Software	Q1	

[1]The SCImago Journal & Country Rank (SJR) indicator is a measure of journal's impact, influence or prestige. It expresses the average number of weighted citations received in the selected year by the documents published in the journal in the three previous years. http://www.scimagojr.com

Computer Vision and Image Understanding 116 (2012) 167–178

Contents lists available at SciVerse ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Fusing color and shape descriptors in the recognition of degraded iris images acquired at visible wavelengths

Hugo Proença *, Gil Santos

Department of Computer Science, IT – Instituto de Telecomunicações, SOCIA – Soft Computing and Image Analysis Group, University of Beira Interior, 6200-Covilhã, Portugal

ARTICLE INFO

Article history:
Received 13 May 2011
Accepted 24 October 2011
Available online 2 November 2011

Keywords:
Iris recognition
Biometrics
Security and surveillance
Forensics
Visible-light data

ABSTRACT

Despite the substantial research into the development of covert iris recognition technologies, no machine to date has been able to reliably perform recognition of human beings in *real-world* data. This limitation is especially evident in the application of such technology to large-scale identification scenarios, which demand extremely low error rates to avoid frequent false alarms. Most previously published works have used intensity data and performed multi-scale analysis to achieve recognition, obtaining encouraging performance values that are nevertheless far from desirable. This paper presents two key innovations. (1) A recognition scheme is proposed based on techniques that are substantially different from those traditionally used, starting with the dynamic partition of the noise-free iris into disjoint regions from which MPEG-7 color and shape descriptors are extracted. (2) The minimal levels of linear correlation between the outputs produced by the proposed strategy and other state-of-the-art techniques suggest that the fusion of both recognition techniques significantly improve performance, which is regarded as a positive step towards the development of extremely ambitious types of biometric recognition.

1. Introduction

Contrary to popular belief, no research effort to date has produced a machine able to covertly recognize human beings in *real-world* conditions. However, it is not difficult to anticipate the significant potential impact of such automation on the security and safety of modern societies (forensics and surveillance). Various research programs have pursued biometric recognition, and most regard the iris as the main biometric trait for three main reasons: (1) the iris is a naturally protected internal organ that is visible from the exterior; (2) the iris has a near-circular and planar shape that facilitates its segmentation and parameterization; and (3) its texture has a predominantly randotypic chaotic appearance that appears stable over the human lifetime and is unique for each individual.

The iris texture is characterized by the irregular distribution of local features such as furrows, crypts, freckles or spots. Hence, traditional approaches are texture-based and decompose the iris data into a set of numerical coefficients that correspond to specific orientations or frequencies of predominant patterns. The recognition in less controlled environments has been gaining relevance and was the focus of many recent proposals, among which the "Iris-on-the-move" project [1] should be highlighted: it is a major example of engineering an image acquisition system to make the recognition process less intrusive for subjects. The goal is to acquire near infra-red close-up iris images as a subject walks at normal speed through an access control point. *Honeywell Technologies* applied for a patent [2] on a very similar system, which was also able to recognize irises at a distance. Previously, Fancourt et al. [3] concluded that it is possible to acquire sufficiently high-quality images at a distance of up to 10 m.

However, recognition in *real-world* data presents many challenges to the pattern recognition process, such as using images acquired in the visible wavelength (VW) spectrum, at widely varying distances (4–8 m), in uncontrolled lighting conditions, on moving subjects and without their active participation in the acquisition process. It becomes relatively difficult to acquire data where the most discriminating iris patterns remain perceptible because the pigments of the human iris (brown-black Eumelanin (over 90%) and yellow-reddish Pheomelanin [4]) have most of their radiative fluorescence under visible light, which significantly varies with respect to the pigmentation levels of the subjects. Although previous technology evaluation initiatives by the authors [5,6] have empirically confirmed the possibility of recognizing human beings in VW *real-world* data, despite achieving error rates far from those obtained in constrained environments, state-of-the-art VW iris recognition methods have achieved decidability indexes of 2.5 at most. The approach that currently outperforms was developed by Tan et al. [7] and makes use of both iris and periocular data. Global color-based features and local ordinal measures were used to extract discriminating data from the iris region, later fused to periocular

* Corresponding author.
E-mail addresses: hugomcp@di.ubi.pt (H. Proença), gsantos@di.ubi.pt (G. Santos).

1077-3142/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2011.10.008

data extracted from texton representations. Finally, fusion was performed by the sum rule using the normalized scores generated for the different types of features. Wang et al. [8] used an adaptive boosting algorithm to build a strong iris classifier from a set of bi-dimensional Gabor-based features, each corresponding to a specific orientation and scale and operating locally. Given the fact that the pupillary boundary is especially difficult to segment in VW data, the authors later trained two distinct classifiers: one for irises deemed to be accurately segmented and another for cases in which the pupillary boundary is expected to be particularly hard to segment. Li et al. [9] used a novel weighted co-occurrence phase histogram to represent local textural features, which is claimed to model the distribution of both the phase angle of the image gradient and the spatial layout and overcomes the major weakness of the traditional histogram. A matching strategy based on the Bhattacharyya distance measures the goodness of match between irises. Marsico et al. [10] proposed the use of implicit equations to approximate both the pupillary and the limbic iris boundaries and to perform image normalization. They exploited local feature extraction techniques such as linear binary patterns and discriminable textons to extract information from vertical and horizontal bands of the normalized image. Although devised for near infrared data, Du et al. [11] aimed at robustness and used the SIFT transform and Gabor wavelets to extract iris features, which were used for local feature point description. Then two feature region maps were designed to locally and globally register the feature points, building a set of deformable iris subregions that take into account the pupil dilation/contraction and deformations due to off-angle data acquisition.

Having empirically analyzed the performance of the above described methods, this paper aims to propose a recognition strategy that uses techniques that are substantially different from those traditionally used in iris recognition, making use of *color* and *shape*. Color is a major visual feature in image and video analysis because color features are considered robust to viewing angle, translation, rotation and scale. Furthermore, for many situations, the shape of image objects often provides important clues for recognition, although shape is sensitive to geometric distortions. The proposed method begins by partitioning the iris into coherent regions in terms of space and color, using data self-organization techniques that tend to compensate for global changes in data. Our method then makes use of a set of well-known color and shape MPEG.7 descriptors to extract both global and local information from the iris data. According to the experiments performed, two types of conclusions are substantiated: (1) the proposed approach achieves performance close to the state-of-the-art methods, and (2) because the data encoding and matching techniques are radically different from the state-of-the-art approaches, the proposed method exhibits low levels of linear correlation with the outputs, which allows it to obtain significant improvements in performance when performing evidence fusion.

The remainder of this paper is organized as follows. Section 2 provides a description of the proposed recognition method. Section 3 provides and discusses the results obtained by our method and compares them to state-of-the-art techniques. The improvements obtained by fusion are highlighted. Finally, the overall conclusions are given in Section 4.

2. Proposed method

A cohesive perspective of the proposed recognition strategy is given in Fig. 1. A color constancy technique is used for regularization purposes, and data are normalized into a Polar coordinate system of constant dimensions, from which global MPEG.7 color descriptors are extracted. Next, a self-organizing data technique divides the noise-free iris data into spatially and color coherent regions that feed the local color and shape MPEG.7 descriptors. Finally, fusion is performed by means of a weighted sum rule.

2.1. Retinex

The original Retinex model was proposed by Land [12]. Its key insight is that color is determined by three independent retinal–cortical systems that use intensity information from different spectral regions of the input data. Each system determines a lightness quantity that is superimposed, yielding the output color for each point. As detailed by Provenzi et al. [13], given an image I, $|\gamma_{ki}|$ ordered chains of pixels can be obtained, starting at k and ending at i, where $|.|$ denotes cardinality. Let x_t and x_{t+1} be subsequent pixels of a chain. Let R_t be the ratio between the intensity of consecutive pixels in the chain, $R_{t_k} = \frac{I(x_t)}{I(x_{t+1})}$, with respect to each image channel. Lightness at position i is given by

$$L(i) = \frac{1}{N}\sum_{k=1}^{N}\prod_{t=1}^{|\gamma_{ki}|-1}\delta_k(R_{t_k}), \tag{1}$$

being δ_k given by

$$\delta_k(R_{t_k}) = \begin{cases} R_{t_k}, & \text{if } 0 < R_{t_k} \leqslant 1-\epsilon \\ 1, & \text{if } 1-\epsilon < R_{t_k} < 1+\epsilon \\ R_{t_k}, & \text{if } 1+\epsilon \leqslant R_{t_k} \leqslant \frac{1+\epsilon}{\alpha} \\ \frac{1}{\alpha}, & \text{if } R_{t_k} > \frac{1+\epsilon}{\alpha} \end{cases},$$

where $\epsilon > 0$ and $\alpha = \frac{1+\epsilon}{\prod_{m_k=0}^{t_k-1}\delta_k(R_{m_k})}$.

2.2. Noise-free iris segmentation

The segmentation of the noise-free iris data acquired in uncontrolled setups has motivated significant research efforts. He et al. [14] used a clustering-based scheme to roughly perform iris localization followed by an integro-differential constellation method for fine detection of each boundary, which not only accelerates the traditional integro-differential operator but also enhances its global convergence. Finally, parametric models were trained to deal with eyelids and eyelashes. Du et al. [15] used a high-pass filter to detect specular reflections inside the pupil and performed a coarse-to-fine segmentation scheme using a least-squares ellipse fitting strategy. A gradient-based technique detected noisy regions that corresponded to diffuse reflections inside the iris. Li et al. [16] used Viola and Jones' method to roughly detect eyes and normalized their region of interest by a K-means-based technique. These data fed the subsequent processing combining traditional iris segmentation methods with RANSAC-like techniques. Concerned about the computational requirements of previously published iris segmentation methods, Proença [17] considered the sclera the most easily distinguishable part of the eye in degraded VW images and fed a neural network with a feature set based in the local proportion of sclera in different directions, resulting in a process that runs in deterministically linear time with respect to the size of the image. Regarding all of the experiments described in this paper, it was observed that, although with noticeably higher computational requirements, the segmentation method of He et al. [14] outperforms the other strategies. Because we aim to obtain performance indicators that are as unbiased as possible, we chose to use this method as the basis for our recognition experiments. Fig. 2 gives examples of eye images and the corresponding noise-free iris segmentation masks, obtained by He et al.'s [14] method.

Parameterization of iris boundaries. Subsequent to segmentation, efficient parameterization of the iris boundaries that are *behind* occlusions was a key issue, especially regarding the normalization

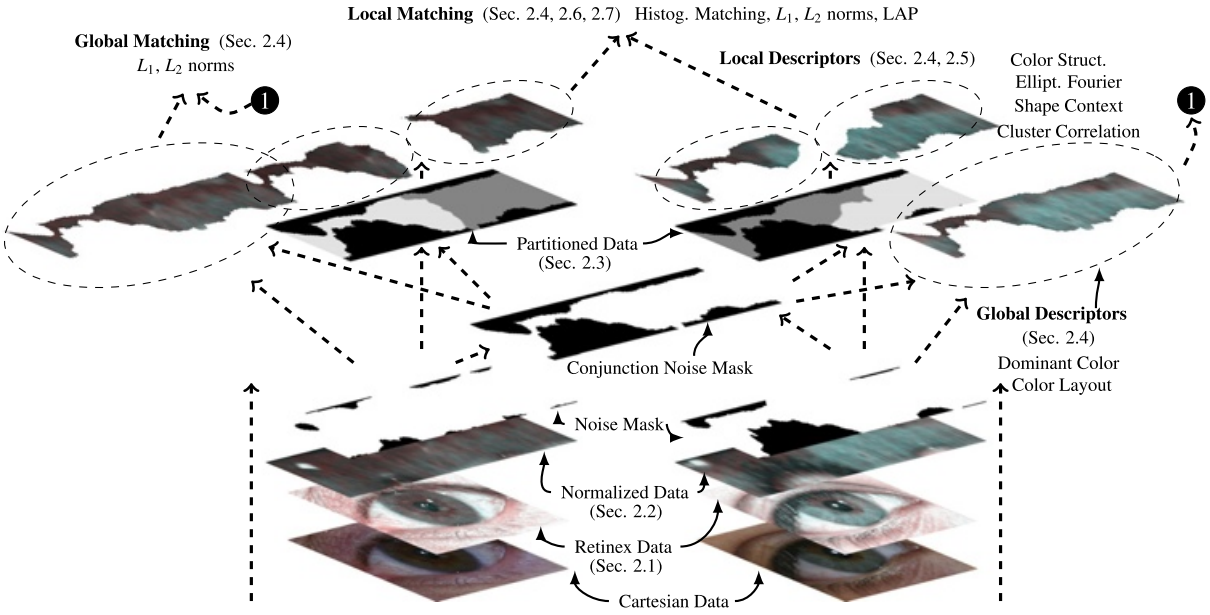
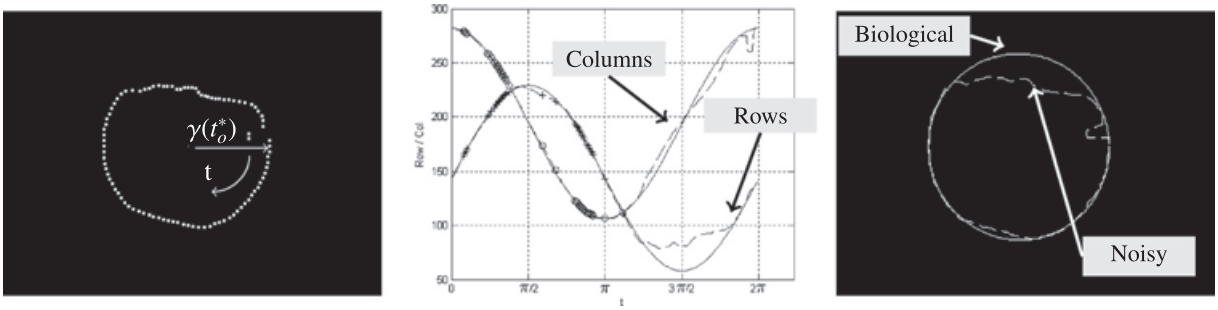


Fig. 1. Cohesive perspective of the proposed method, describing its major phases and the corresponding sections of this paper.

Fig. 2. Examples of degraded VW iris images and the corresponding noise-free segmentation masks obtained according to the method of He et al. [14]. The binary masks discriminate between the non-occluded pixels of the iris (white regions) and all of the remaining types of data (black regions).



(a) Segmented boundary (b) Fourier regression (c) Reconstructed boundary

Fig. 3. Parameterization of the biological iris boundaries. According to the values of (3), smoother regions with low energy of the cumulative angular descriptor (a) were deemed to belong to the biological boundaries (cross and circular data points of b) and used to reconstruct the deemed biological border through a regression of a Fourier series (c).

of the iris data into a pseudo-polar coordinate system of constant dimensions. As detailed in [18], this phase was divided into two steps: (1) discriminating between the boundary segments that correspond to biological iris borders and the boundary segments that delimit noisy regions and (2) reconstructing the full biological iris boundaries according to the former segments. The key insight in this step is that biological boundaries can be faithfully described by periodic signals, which justifies the use of Fourier series for such purposes. The cumulative angular function was used as a shape descriptor, defined as the amount of angular change from an arbitrary starting point:

$$\gamma(t) = \int_0^{\frac{Lt}{2\pi}} k(r)dr - k(0) + t \tag{2}$$

where $t \in [0, 2\pi]$ and $k(r)$ describe changes in direction at point t with respect to changes in arc length L. As illustrated in Fig. 3a and b, biological boundaries have smoother angular descriptor values with lower energy, which leads to the following objective function:

$$O(t) = \beta_0\gamma(t) + \sum_{i=1}^{2} \beta_i \frac{\partial^i \gamma(t)}{\partial t^i}, \tag{3}$$

where β_i were empirically obtained regularization constants. Arguments of the first quartile of $O(t^*)$ – t^* regularly spaced in [0, 2π] – were deemed to belong to the biological border and their coordinates (column and row), illustrated by the dot and cross data points of Fig. 3b. Finally, the reconstruction of the biological border

used the selected coordinates and was regarded as a nonlinear regression of a Fourier series of order r, with a fundamental frequency constrained to $\omega = 1$, which assures closure and completeness of the contour:

$$c(x) = \frac{a_0}{2} + \sum_{k=1}^{r}(a_k cos(x\omega k) + a_{r+k} sin(x\omega k)). \quad (4)$$

Using the deemed biological iris boundaries, in the next step we convert data into a pseudo-polar coordinate system of fixed dimensions, using the well known *Daugman rubber sheet* model [19].

2.3. Partitioning the iris into regions

Partitioning the iris into regions is one of the roots of the proposed recognition method and aims to divide the noise-free pixels of the iris into k disjoint sets $C = \{C_1, \ldots C_k\}$ such that elements within each C_i are as *homogenous* as possible in terms of both their position and their color. Considering the demands of a perceptually uniform color space and aiming to preserve the connectivity between pixels of each cluster, each element was represented by the feature set $f = \{r, |\frac{W}{2} - c|, L^*, a^*b^*\}$, with r and c the row and column coordinates with respect to the normalized iris image. W is the width of the normalized image, and L^*,a^*b^* are the color coordinates in the CIELAB color space, using a reference white provided by illuminant $D65$, 2nd observer, as described in [1]. In this 5D space, *distance* corresponds to the metric:

$$d(f_1, f_2) = \sqrt{\sum_i \omega_i (f_1^i - f_2^i)^2}, \quad (5)$$

where f^i denotes the ith feature of f. Using a partition-based clustering scheme (fuzzy c-means [20]), partitions were found by maximizing an objective function that considered both the within and between cluster variation:

$$J(C) = \sum_{i=1}^{k}\left(\sum_{j=1}^{k} d(C_i, C_j) - \sum_{j \in C_i} d(C_i, f_j)\right) \quad (6)$$

being $d(C_i,C_j)$ the sum of L_2 distances (5) between every combination of elements of C_i and C_j and $d(C_i, f_j)$ the sum of L_2 distances (5) between every element of C_i and the feature point f_j.

Fig. 4 illustrates how the typical appearance of the generated clusters would vary with respect to different ω_i values. Black pixels denote regions that were classified as *noisy* by the segmentation method and, as such, were not considered in the clustering process. The remaining intensities represent the clusters assigned to each pixel of the normalized iris data when privileging the (ω_1, ω_2) weights (associated with spatial features) and the (ω_3, ω_4, ω_5) (associated with color features) (bottom left image). The image at the bottom illustrates clusters generated for the *optimal* weight values $\widehat{\omega_i}$, which constitutes a trade-off between space and color, as follows:

$$\widehat{\omega_i} = \arg\min_{\omega_i} \sum_k \sum_j \alpha\phi_k(\omega_i, \{C_{kj}\}) + (L_{kj}^2 - 4\pi A_{kj}), \quad (7)$$

where L_{kj} and A_{kj} represent the perimeter and the area of the region delimited by the jth cluster of the kth image (C_{kj}), ϕ_k corresponds to the total of connected components in that cluster [21] and α is a regularization term that was empirically found according to the training data set of k images ($\alpha = 10^{1.5}$, $k = 100$ in our experiments). To account for the dynamic conditions that propitiate occlusions in different regions of the irises, the clustering process was performed using as a noise-mask the conjunction of the noise-masks of images to be matched.

[1] http://www.csse.uwa.edu.au/du/Software/graphics/xyz2lab.m

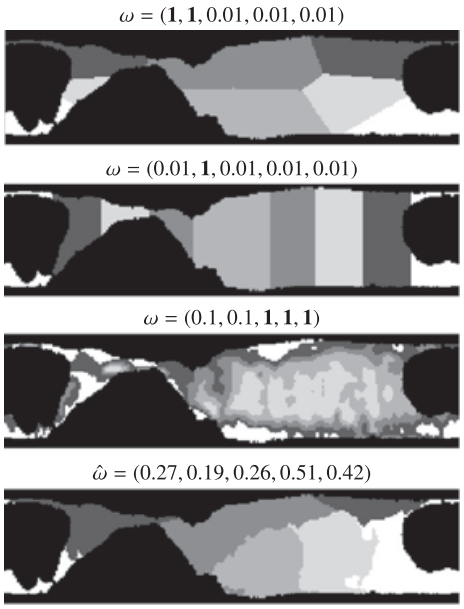


Fig. 4. Comparison between the regions resulting from different clustering processes with respect to weights given to each feature. The image at the top resulted from large weights (denoted by bold font) for spatial features, whereas in the case of the second image at the top, a large weight was given exclusively for one of those spatial features (column). The second image at the bottom resulted from low weight values for spatial features, and clusters were formed, accounting for the color values. The weights used in the case of the image at the bottom were obtained by (7); this type of cluster is used in all subsequent processing phases.

2.4. Color descriptors

Most of the MPEG.7 descriptors have compression/reconstruction purposes and – consequently – tend to focus in the lowest frequency components of signals. When compared to other biometric traits, one of the most interesting features of the iris is that most of its discriminating information lies in the lowest and middle-low frequency components. For such, these descriptors would intuitively be useful for iris recognition purposes, which constituted the main key insight for their utilization in this work.

2.4.1. Dominant color descriptor

The dominant color descriptor summarizes the image content by extracting the most *important* colors in an image or region, naturally perceived as the most frequent. Let $I = \{\vec{x}\}, \vec{x} = (x_1^j, x_2^j, x_3^j)$ be a $r \times c$ image represented in the CIELAB color space, known to more closely fit the Euclidean difference between colors and the visual perception of color difference. Let k be the number of colors to extract from the image. Aiming to obtain deterministic results, the k geometric centroids (s_i) were used as initial values of the centers of clusters:

$$s_i = \frac{i\max\{\vec{x}\} + (k+1-i)\min\{\vec{x}\}}{k+1}, \ \mathrm{i} = \{1, \ldots, \mathrm{k}\}. \quad (8)$$

The coordinates of these centroids were updated according to the generalized Lloyd algorithm, minimizing the objective function:

$$\arg_s \min \sum_{s=1}^{k} \sum_{\vec{x} \in s_i} \|\vec{x} - s_i\|^2. \quad (9)$$

The iterative procedure continues until the values of s_i at successive steps do not differ more than a positive value that acts as stopping criterium ($0 < \epsilon \ll 1$), i.e., $\|s_i^t - s_i^{t+1}\| < \epsilon$.

2.4.2. Color layout descriptor

This descriptor extracts the spatial distribution of the most representative colors of visual signals according to their position on a grid superimposed on data. Let I be an image of size $r \times c$ expressed in the YCbCr color space [22]. First, the most representative colors in each $n \times n$ region are obtained, as described in Section 2.4.1, which yields a $r/n \times c/n$ array that was transformed using the 1D Discrete Cosine Transform type-II, performed first along the rows and then along the image columns, yielding a set of 2D components:

$$X_{i,j} = \sum_{n=0}^{c-1} \sum_{m=0}^{r-1} I(n,m) \cos\left[\frac{\pi}{r}(m+0.5)j\right] \cos\left[\frac{\pi}{c}(n+0.5)i\right].$$

Finally, the $X_{\{.\}}$ coefficients were zigzag ordered [23], resulting in a vector $\vec{v}$ of $(r \times c)/n^2$ real components. Matching between vectors $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$ was performed according to the L_2 norm of the vector $\overrightarrow{v_1} - \overrightarrow{v_2} = \sqrt{\sum_{i=1}^{(r\times c)/n^2} (v_1(i) - v_2(i))^2}$.

2.4.3. Color structure descriptor

The color structure descriptor [22] generalizes a simple color histogram and uses a structuring element that moves across data, defining a neighborhood where the dominant color values are analyzed and counted for each bin. In our experiments, we used a rectangular structuring element with $\frac{1}{12}$ of the image width and height. Because this descriptor is very similar to an image histogram, the same L_1 based matching functions were used in matching.

2.4.4. Scalable color descriptor

Scalable color descriptors [22] are global descriptors mainly used for image-to-image matching. The process starts by extracting a color histogram of k bins in the HSV color space, where the hue component is usually quantized to a larger number of bins compared to saturation and value layers. Such a histogram feeds a dyadic decomposition process based in a pair of Haar wavelet transforms. Let I be an image represented in the HSV color space and let h be the corresponding normalized histogram with k bins. The convolution between h and the low-pass kernel of the Haar transform is equivalent to summing pairs of adjacent bins, whereas the high frequency components are obtained by the difference in adjacent bins. Such decomposition is repeated n times, using at each iteration the lower frequency components previously obtained. The default matching function is based on the L_1 metric (i.e., the sum of the absolute differences between corresponding elements): $S = \sum_{i=1}^{k} |h_A[i] - h_B[i]|$, where k denotes the number of extracted coefficients.

2.5. Order statistics of dominant colors

Let $\vec{c}$ be the dominant color of the noise-free iris portion of I. Let $\overrightarrow{c_i}$ be the dominant color of each cluster C_i and $d : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ the L_2 norm of the vector $\overrightarrow{c_i} - \vec{c}$. For any pair $(\overrightarrow{c_i}, \overrightarrow{c_j})$, we define a pseudometric $d^*(\overrightarrow{c_i}, \overrightarrow{c_j})$ given by

$$d^*(\overrightarrow{c_i}, \overrightarrow{c_j}) = |d(\overrightarrow{c_i}, \vec{c}) - d(\overrightarrow{c_j}, \vec{c})|.$$

Similarly, a binary relation $\succ$ on $\mathbb{R}^3 \times \mathbb{R}^3$ is defined by

$$\overrightarrow{c_i} \succ \overrightarrow{c_j} \iff d(\overrightarrow{c_i}, \vec{c}) > d(\overrightarrow{c_j}, \vec{c}) \tag{10}$$

. The rationale behind $\succ$ is to consider as *greater* colors those that are more distant from the dominant color of the iris. From this definition, it is straightforward to infer that $\succ$ is irreflexive, asymmetric and transitive, which is particularly useful for our purposes. Let $X = \{\overrightarrow{x_1}, \ldots, \overrightarrow{x_n}\}$ be a random variable that represents the distance between the dominant colors inside each cluster and the whole iris. According to the elementary theory of rank tests and using (10), the kth order statistic $\overrightarrow{x_{(k)}}$ of a statistical sample $\{\overrightarrow{x_1}, \ldots, \overrightarrow{x_n}\}$ is equal to its kth smallest value. Let $\overrightarrow{x_{(1)}}, \ldots, \overrightarrow{x_{(n)}}$ be the order statistics of a set of independent observations, that is, $\overrightarrow{x_{(1)}} < \overrightarrow{x_{(2)}} < \ldots < \overrightarrow{x_{(n)}}$. Assuming that $\overrightarrow{x_i}$ is mutually independent, the distribution function of $\overrightarrow{x_{(k)}}$ is equal to [24]:

$$F_{(k)}(y) = P(\overrightarrow{x_{(k)}} \leqslant y) = \sum_{i=k}^{n} \binom{n}{i} [F(y)]^i [1 - F(y)]^{n-i} \tag{11}$$

being $F(y)$ the cumulative distribution function of X. For a pair of images, having two vectors with the kth, ($k = \{1, \ldots, n\}$) order statistics of $\overrightarrow{x_{(k)}}$, matching was performed according to the L_1 metric, i.e., the sum of the absolute differences between corresponding elements. As described in the experiments section, the kth order statistics have evident discriminating information between individuals and was often one of the features automatically selected for the classification stage.

2.6. Linear assignment problem

The normalization of the iris data into a polar coordinate system propitiates invariance to translation and scale of the original data, but not to rotation, which appears as differences in translation of the normalized data and of the resultant clusters. For this, we used an automated method that seeks the maximal similarity between clusters, independent of their position in the normalized data, which was handled by a linear assignment strategy. Let $G = (U,V;E)$ be a bipartite graph with a separable set of vertices U and V ($|U| = |V| = n$) and a set of edges $E = \{e_{ij}\}$, such that e_{ij} denotes an edge from the ith vertex of U to the jth vertex of V. Let $c(e_{ij})$ denote the cost of the edge $c(e_{ij})$, such that $c(e_{ij}) \geqslant 0, \forall i,j \in \{1, \ldots n\}$. The linear assignment problem aims to find E^*, a subset of E that satisfies the following properties: (1) the accumulated cost of its edges is minimal, and (2) each vertex of U and V appear exactly once in E^*. Let $\phi(i,j)_{\{.\}} : \mathbb{N} \times \mathbb{N} \to \{0,1\}$ be an indicator function, such that $\phi(i,j) = \mathbb{I}_{\{e_{ij} \in E^*\}}$. The *optimal* correspondence between elements of U and V is given by

$$\begin{aligned} &\min \sum_{i=1}^{n} \sum_{j=1}^{n} \phi(i,j) c(e_{ij}) \\ &\text{s.t.} \quad \sum_{i=1}^{n} \phi(i,j) = 1, \forall i \in \{1, \ldots, n\} \\ &\quad \sum_{j=1}^{n} \phi(i,j) = 1, \forall j \in \{1, \ldots, n\} \end{aligned} \tag{12}$$

Due to computational concerns, the problem was regarded as a shortest augmenting path algorithm with an implementation of the Dijkstra's shortest path method, which is known to run in time $O(n^3)$. Details can be found in the work of Jonker and Volgenant [25]. In practical terms, when matching two clustered iris images, the relative position of each cluster center is regarded as a vertex and included respectively in U and V. The cost (c) of edges E corresponds to the Euclidean distance between elements of U and V, which complies the above formalization.

2.7. Histogram matching

In every phase of our method where the distance between histograms had to be obtained, several possibilities were tested, and the results were evaluated in a training data set. The best results were obtained with the cross-bin Quadratic-Chi distance histogram proposed by Pele and Werman [26]: let h_1 and h_2 be two non-negative bounded histograms, and let $A = [a_{ij}]$ be a non-negative and symmetric bib-similarity matrix, such that $a_{ii} \geqslant a_{ij}, \forall j \neq i$.

The Quadratic-Chi histogram distance is given by (13), where $h^i_{\{.\}}$ denotes the histogram value at position i, and m is a regularization factor (the best results were obtained in our experiments with 0.9).

$$QC(h_1,h_2)=\sqrt{\sum_{ij}\left(\frac{h_1^i-h_2^i}{\left(\sum_c(h_1^c+h_2^c)A_{ci}\right)^m}\right)\left(\frac{h_1^j-h_2^j}{\left(\sum_c(h_1^c+h_2^c)A_{cj}\right)^m}\right)A_{ij}}. \tag{13}$$

2.8. Shape context descriptor

Proposed by Belongie et al. [27], this descriptor provides an efficient way to measure the similarity between shapes, represented by a set of contour points $\{p_i\}$. For each p_i, we extract a histogram h_i of the relative coordinates of the remaining p_j points ($i \neq j$) with respect to p_i and represented in a log-polar coordinate system. Each h_i histogram is defined as the *shape context* of p_i and is used in all subsequent processing. Let p_1 and p_2 be boundary points of two shapes that are to be matched. The cost of matching p_1 with p_2 uses the χ^2 statistic:

$$C_{12}=\frac{1}{2}\sum_{k=1}^{K}\frac{(h_1(k)-h_2(k))^2}{h_1(k)+h_2(k)}, \tag{14}$$

where $h_1(k)$ and $h_2(k)$ denote the kth bin of the histograms of p_1 and p_2. The set of all costs C_{ij} between all pairs of points of two shapes is regarded as the cost matrix of a bipartite graph-matching problem and was solved as described in Section 2.6. As illustrated in Fig. 5, this descriptor is an efficient way to extract discriminating information about the shape of the regions resulting from the data partitioning phase and is used as a soft biometric measure in the recognition process.

2.9. Robustness to data variation factors

The basic premise of the proposed method is that the uniqueness of each iris texture determines that pixels are grouped in a specific way for each iris and compose clusters that are specific in terms of their positions and shapes, although these clusters cannot be expected to provide enough information for strong biometric recognition. Fig. 6 illustrates such discriminating ability, showing the clusters that result from two different heavily pigmented irises. Here, the existence of four predominantly

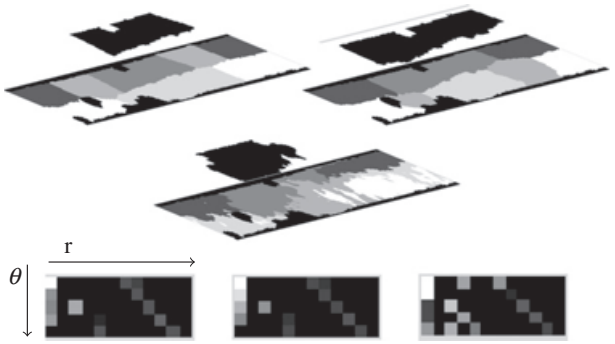


Fig. 5. Illustration of the shape descriptor used to characterize each iris region. The upper row shows two similar shapes, from which shape context descriptors were extracted. The image at the center has a significantly different shape. Images in the bottom row illustrate the corresponding shape descriptors at point $L/4$, with L being the length of the contour and starting in the upper left pixel. Note the similarity between the far left and the centered descriptor and their dissimilarity to the far right image.

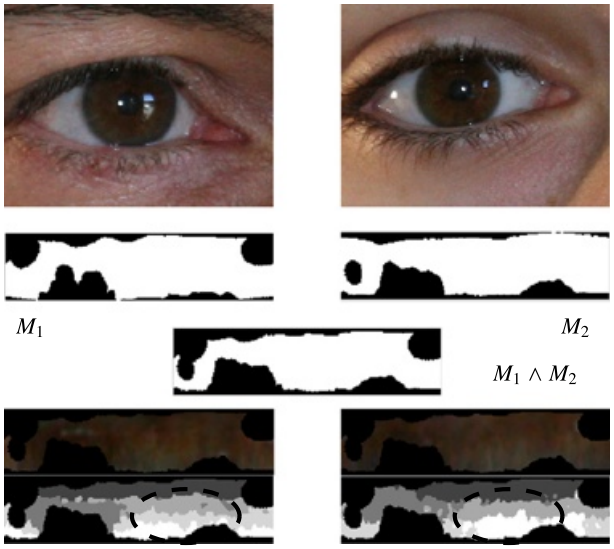


Fig. 6. Clusters generated for two different heavy pigmented irises, where local contrast inside the iris ring is hardly perceived by a human observer. Even so, the appearance of the resulting group is evidently different. Images are "C_1_S1_I12.tiff" and "C_101_S1_I10.tiff" of the UBIRIS.v2 data set.

horizontal clusters in the left image is in opposition to the right image and is particularly evident in the regions delimited by the dashed ellipses.

Due to the dynamics of the acquisition setup, it is expected that the unoccluded regions of the iris will vary, which will affect the clustering results. This was overcome by obtaining the *conjunction noise-mask* of the pair of images to be matched (illustrated in Fig. 7), yielding two properties: (1) multiple biometric signatures are possible to extract from each image, depending on the other image that it will be matched against and (2) privacy concerns about the recognition process because it is required that the raw iris data and the corresponding noise-mask be stored in the database instead of the biometric signature.

Furthermore, it is important that the positions, sizes and shapes of regions are not subject to sudden or extreme changes as a result of the dynamics of the acquisition setup. Fig. 7 illustrates two images from the same eye acquired from different distances (9 and 4 m). It can be seen that clusters remained relatively stable, essentially due to translation into the polar coordinate system and to the known property of invariance to color perception, as a result of moderate changes in scale.

The acquisition of a small moving target as the iris at relatively large and varying distances propitiates very different levels of

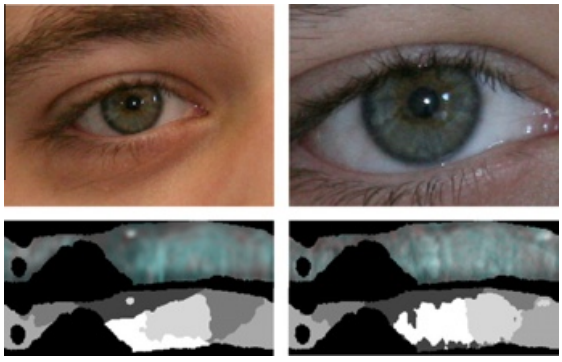


Fig. 7. Robustness to changes in scale. Images are "C_111_S1_I4.tiff" and "C_111_S1_I13.tiff" of the UBIRIS.v2 data set.

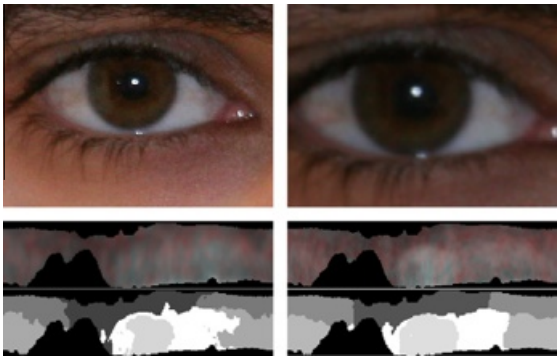


Fig. 8. Robustness to defocused data. Images are "C_183_S2_I10.tiff" and "C_183_S2_I13.tiff" (defocused by a Gaussian kernel of $\sigma = 1.4$) of the UBIRIS.v2 data set.

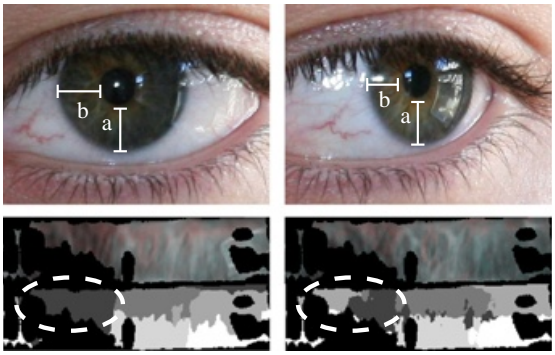


Fig. 10. Robustness to off-angle image acquisition. Images are "C_24_S1_I13.tiff" and "C_24_S1_I15.tiff" of the UBIRIS.v2 data set.

image focus. Fig. 8 illustrates such variations. Although the similarity between the clustered images is evident, we observe that the shape of the clusters often becomes smoother in defocused data. If the defocus is exaggerated, the clustering process tends to augment the relevance given to spatial features, resulting in clusters with more regular shape.

Rotations in the original Cartesian space directly correspond to translations in the Polar coordinate system. However, significant changes in rotation are not expected due to the natural and biologically determined position of the head with respect to the neck and shoulders of stand-up subjects. Fig. 9 illustrates the behavior of the clustering process for a pair of images of the same eye where one of them was artificially rotated by $\frac{\pi}{6}$ (a value that is beyond the expected rotations). The relative position of clusters was shifted approximately $\frac{1}{12}$ of the width of the polar image. In this case, shapes remain roughly constant and the position of *corresponding* clusters varies significantly, which was handled by the Linear Assignment process described in Section 2.6, which finds the optimal correspondence between clusters according to their shape.

Off-angle images are of special interest because gaze is known to be a primary source of error in traditional recognition strategies, particularly when circular iris parameterization techniques introduce differences in the phase of the normalized data and the bias phase-based in encoding/matching methods. The translation into the polar coordinate system implies that the data are sampled at different rates with respect to the length of the iris ring at each angle ($a \gg b$ in the right image of Fig. 10, but $a \simeq b$ in the left image), which does not significantly affect the color perception of the resultant data. This relationship was observed even in cases where exaggerated deviations occlude portions of the iris. Furthermore, this figure gives a typical failure situation motivated by iris segmentation inaccuracies: the region delimited by the dashed ellipse in the right figure should have been classified as *noise* (corresponds to the upper part of the iris, partially occluded by eyelashes) but was erroneously considered for the clustering process and induced substantial differences in the resultant clusters near that region.

Lighting variations are due to the type of illuminants or to the amount of light in the environment and constitute a problematic factor, especially for local variations. The upper and middle row images of Fig. 11 were acquired from the same eye under substantially different lighting conditions but were mostly compensated by the *Retinex* process described in Section 2.1 (compensated images are shown in the central column). Even so, higher variability in the shapes of the resulting clusters was observed, as highlighted by the regions delimited by the dashed horizontal ellipses. Finally, local lighting variations were observed to be the most problematic factor and to significantly bias the clustering process. Images at the bottom row illustrate such types of variations and, as highlighted by the diagonal dashed ellipses, the Retinex algorithm was not able to handle such variations, and the resulting clusters varied significantly.

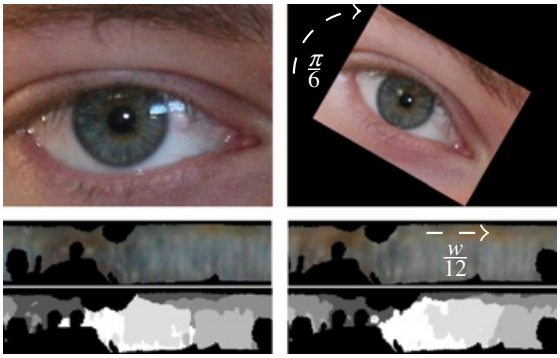


Fig. 9. Robustness to changes in rotation. Images are "C_171_S1_I10.tiff" and "C_171_S2_I10.tiff" (rotated by $\frac{\pi}{6}$) of the UBIRIS.v2 data set.

3. Experiments

According to the review of other VW iris recognition methods given in Section 1 and to the performance that we empirically observed, four methods were selected to be used as comparison terms of our proposal: Tan et al. [7], Wang et al. [8] and Marsico et al. [10] were the outperforming methods of a recently performed contest about VW iris recognition, and simultaneously exhibited the lowest levels of linear correlation. Finally, even though the approach of Du et al. [11] was devised for NIR data, it was selected for contextualization purposes, in order to assess the adaptability of NIR-based approaches to VW data. All these methods are our own implementations, validated by comparing the performance described by authors (in the NICE:II data sets) and ours.

3.1. Feature selection

In a training set of 1000 images used by the participants of the NICE:II contest (available at[2]), the discriminating ability of a large set of features was assessed, testing different values for the number of clusters (between two and seven) and for the most relevant parameters of the described encoding strategies, yielding a total of 112 features. Fig. 12 gives the probability density functions and the corresponding cumulative density functions of the ten most discriminating features, selected based on mutual information and the

[2] http://nice2.di.ubi.pt

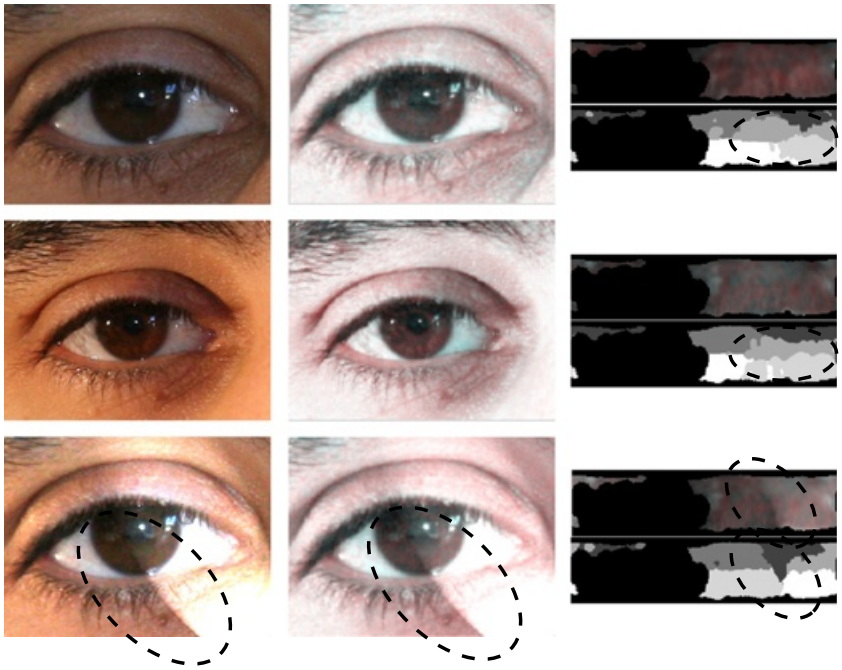


Fig. 11. Robustness to global and nonuniform lighting changes. Images are "C_137_S1_I10.tiff" (top) and "C_137_S1_I7.tiff" (middle) and "C_137_S1_I10.tiff" (bottom, with a directional artificial light effect) of the UBIRIS.v2 data set.

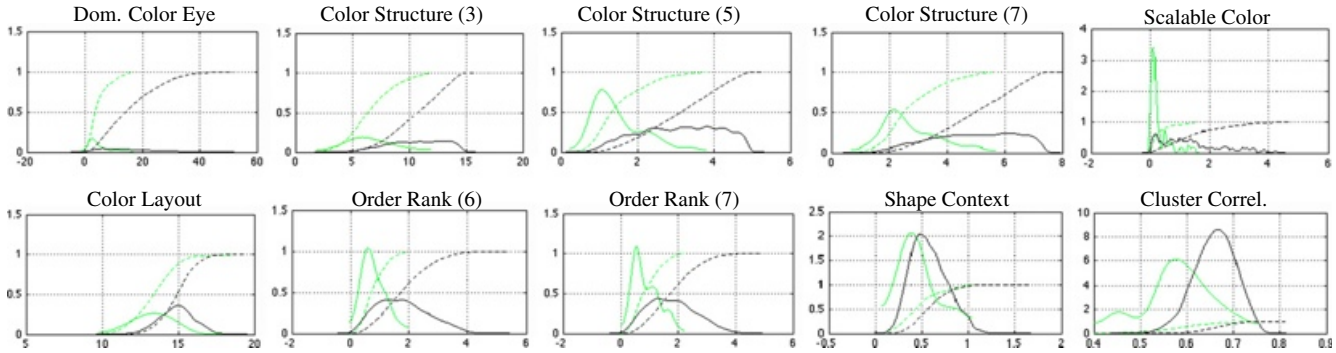


Fig. 12. Probability density (continuous lines) and cumulative density functions (dashed lines) of the features selected for the biometric recognition process. The *non-match* comparisons are represented by darker lines and the *match* comparisons by lighter lines.

criteria of maximum dependency, maximum relevance and minimum redundancy, as proposed by Peng et al. [28]. We considered two sets of observations in a k-dimensional space, one for *match* and other for *non-match* comparisons, assumed to be independent and identically distributed. The probability functions f were estimated by Gaussian-based kernel density estimators, as proposed by Botev et al. [29]:

$$\hat{f}(x;t) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi t}}e^{-(x-d_i)^2/(2t)}, \tag{15}$$

where the bandwidth t was determined by the analysis of the mean integrated square error. Having assessed performance in this training set, near maximal performance was observed when selecting more than thirty features, linearly combined to maximize performance in that data set. Further, subsequent recognition experiments were made when using this classifier.

3.2. Verification mode

fig. 13 compares the ROC curves obtained by the proposed method and other methods selected for comparison, where each

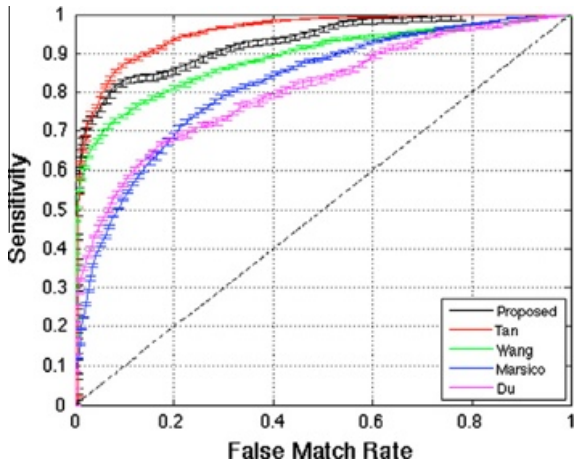


Fig. 13. Comparison between the receiver operating characteristic curves of the proposed methods and others used for contextualization purposes.

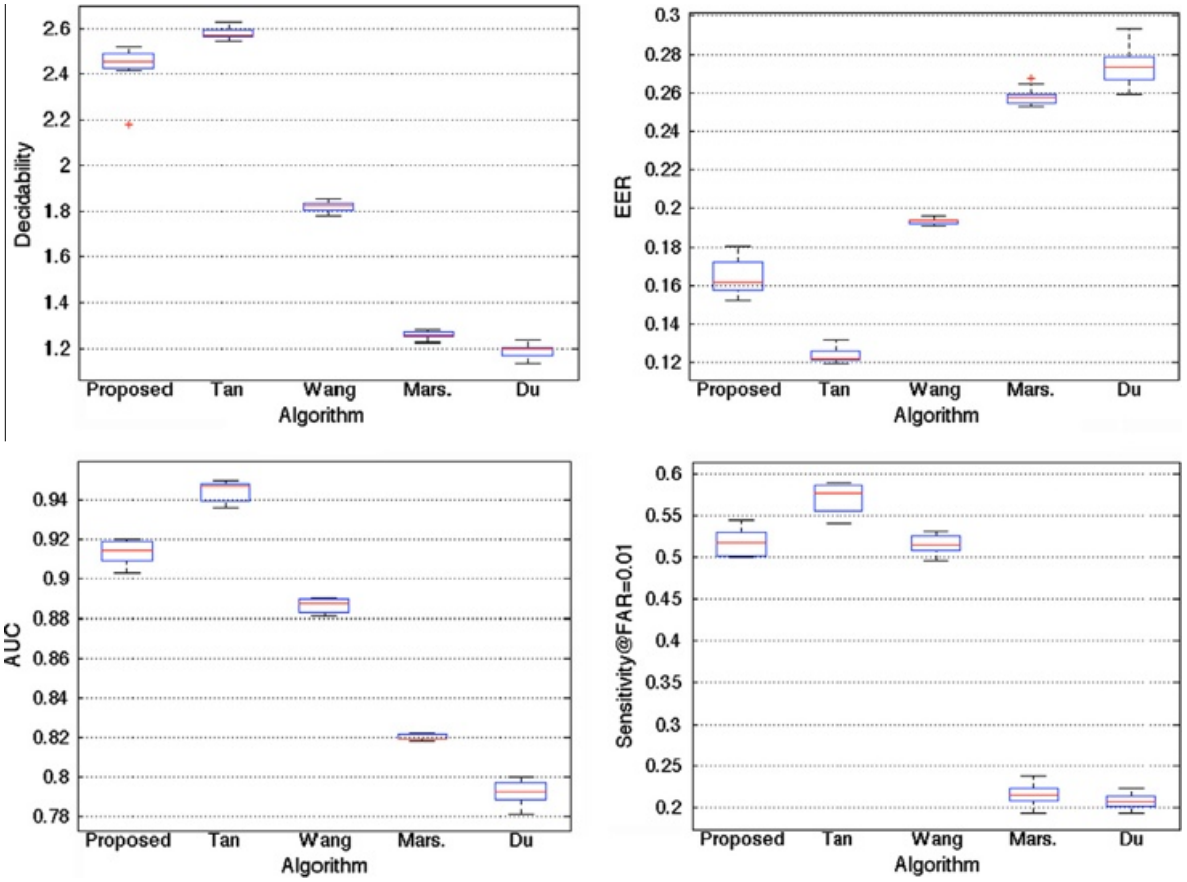


Fig. 14. Comparison between the performance measures obtained by the tested recognition strategies. Each column represents one recognition approach, where the corresponding median value is represented by the horizontal line through the middle of each box. The top and bottom of the boxes denote the first and third quartile of the observations. Outliers appear as dot data points.

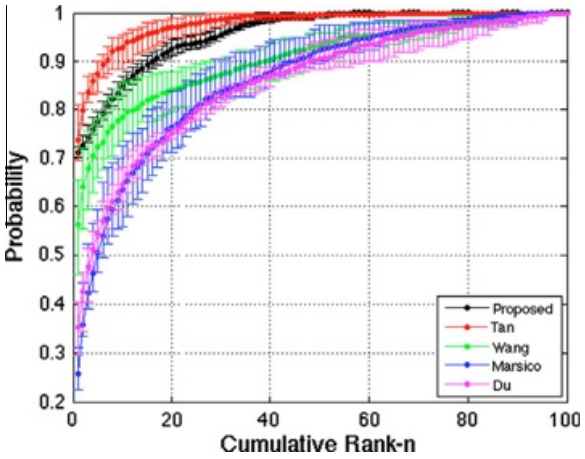


Fig. 15. Average cumulative rank n curves obtained by the proposed method and others used as comparison terms. The bottom and top horizontal lines around the data series denote the worst and best values obtained.

data point is surrounded by two horizontal bars that denote the best and worst values obtained at that operating point. The algorithm of Tan et al. [7] outperformed others at most operating points, whereas the proposed strategy usually performs better than that of Wang et al. [8]. The performance of Du et al. [11] and Marsico et al. [10] often intersect.

Another comprehensive comparison between the error rates obtained by these methods is given in Fig. 14, which shows the

Table 1
Pearson's sample correlation coefficients between the tested recognition methods and ours. Values are given with the corresponding 95% confidence intervals.

	Proposed	Tan	Wang	Marsico	Du
Proposed	1.00	–	–	–	–
Tan	0.38 ± 0.016	1.00	–	–	–
Wang	0.33 ± 0.017	0.56 ± 0.013	1.00	–	–
Marsico	0.37 ± 0.016	0.56 ± 0.013	0.41 ± 0.016	1.00	–
Du	0.32 ± 0.017	0.42 ± 0.016	0.33 ± 0.017	0.30 ± 0.017	1.00

decidability index (d') given by $d' = \frac{|\mu_E - \mu_I|}{\sqrt{\frac{\sigma_I^2 + \sigma_E^2}{2}}}$, where $\mu_I = \frac{\sum_i d_i^I}{k}$ and $\mu_E = \frac{\sum_i d_i^E}{m}$ are the means of the match/non-match distributions, and $\sigma_I = \frac{\sum_i (d_i^I - \mu_I)^2}{k-1}$ and $\sigma_E = \frac{\sum_i (d_i^E - \mu_E)^2}{m-1}$ are their standard deviations. The approximated equal error rate (*EER*), the area under the curve (*AUC*) and the average sensitivity when operating at a FAR of approximately 0.01 were also obtained. The results are expressed in terms of boxplots, showing the median of the observed performance range (horizontal solid line) and the first and third quartile values of the observations (top and bottom of the box marks). The upper and lower whiskers are denoted by the horizontal lines outside each box, and the outliers are denoted by dot points.

3.3. Identification mode

assuming a closed universe model, we tested the effectiveness of each method when trying to answer the following question:

Table 2
Best results obtained by classification ensembles, according to the number of fused methods.

# Fused	Methods	Rule	Decid. d' ($\Delta d'$)
2	{Proposed, Tan}	*	2.848(+0.2629)
3	{Proposed, Tan, Marsico}	*	2.860(+0.2831)
4	{Proposed, Tan, Wang, Marsico}	*	2.738(+0.1609)
5	{Proposed, Tan, Wang, Marsico, Du}	*	2.605(+0.0280)

"Is the correct identity among the best k matches?" This type of performance measure is usually expressed by means of rank and cumulative rank histograms, where ranks appear in the horizontal axis and probabilities in the vertical one. Let $\mathbb{T} = \{T_1, \ldots, T_t\}$ be the set of gallery images such that $i \neq j \Rightarrow id(T_i) \neq id(T_j)$ and $\mathbb{S} = \{S_1, \ldots, S_s\}$ the set of samples that are to be compared against $\mathbb{T}$. Each S_i constitutes a query that is matched against all elements of $\mathbb{T}$, yielding a set of $D = \{d_{i1}, \ldots, d_{it}\}$ dissimilarity scores, where d_{ij} denotes the dissimilarity between the ith sample and the jth template. Let $D' = \{d'_{i1}, \ldots, d'_{it}\}$ be the ordered version of D, such that $d'_{i1} \leqslant d'_{i2} \leqslant \ldots \leqslant d'_{it}$. S_i is said to have rank k if the score d_{ij} is in the kth position of D' and $id(S_i) = id(T_j)$. The probability of having rank k $P(rank\text{-}k)$ is estimated by the ratio between the number of sample queries with rank k and the total number of queries. Accordingly, the probability of cumulative rank k can be calculated as $\sum_{i=1}^{k} P(\text{rank} - i)$. Fig. 15 gives the probability distribution of the cumulative rank values obtained for $t = 100$, representing the identification performance obtained.

3.4. Correlation and fusion

The statistical correlation between the outputs given by our method and others used as comparison terms was analyzed to determine whether performance could be improved by fusing several of them. It was assumed that any eventual dependence between scores would be linear, which justifies the use of the Pearson's correlation coefficient to analyze the strength of these dependences. Table 1 gives the correlation r of 10,000 responses given by each biometric system, where $r(X,Y) = \frac{1}{n-1}\sum_i \frac{X_i-\overline{X}}{\sigma_X}\frac{Y_i-\overline{Y}}{\sigma_Y}$, where X_i and Y_i denote the system outputs, $\overline{X}, \overline{Y}$ are the sample means and σ_X, σ_Y the standard deviations.

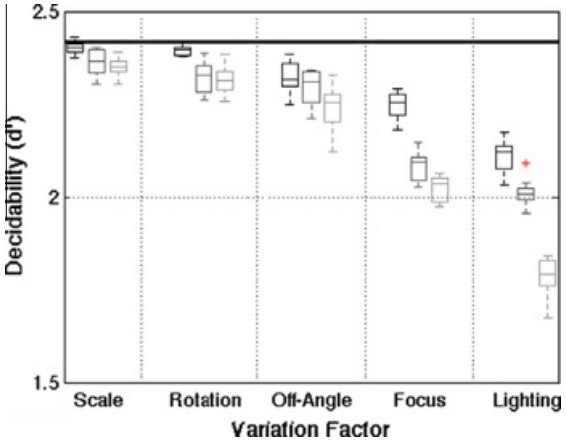


Fig. 17. Degradation in recognition performance, expressed in terms of the decidability values, with respect to variations in different factors. The results are expressed in terms of boxplots, showing the median of the observed performance range (horizontal solid line) and the first and third quartile values of the observations (top and bottom of the box marks). The upper and lower whiskers are denoted by the horizontal lines outside of each box, and the outliers are denoted by dot points.

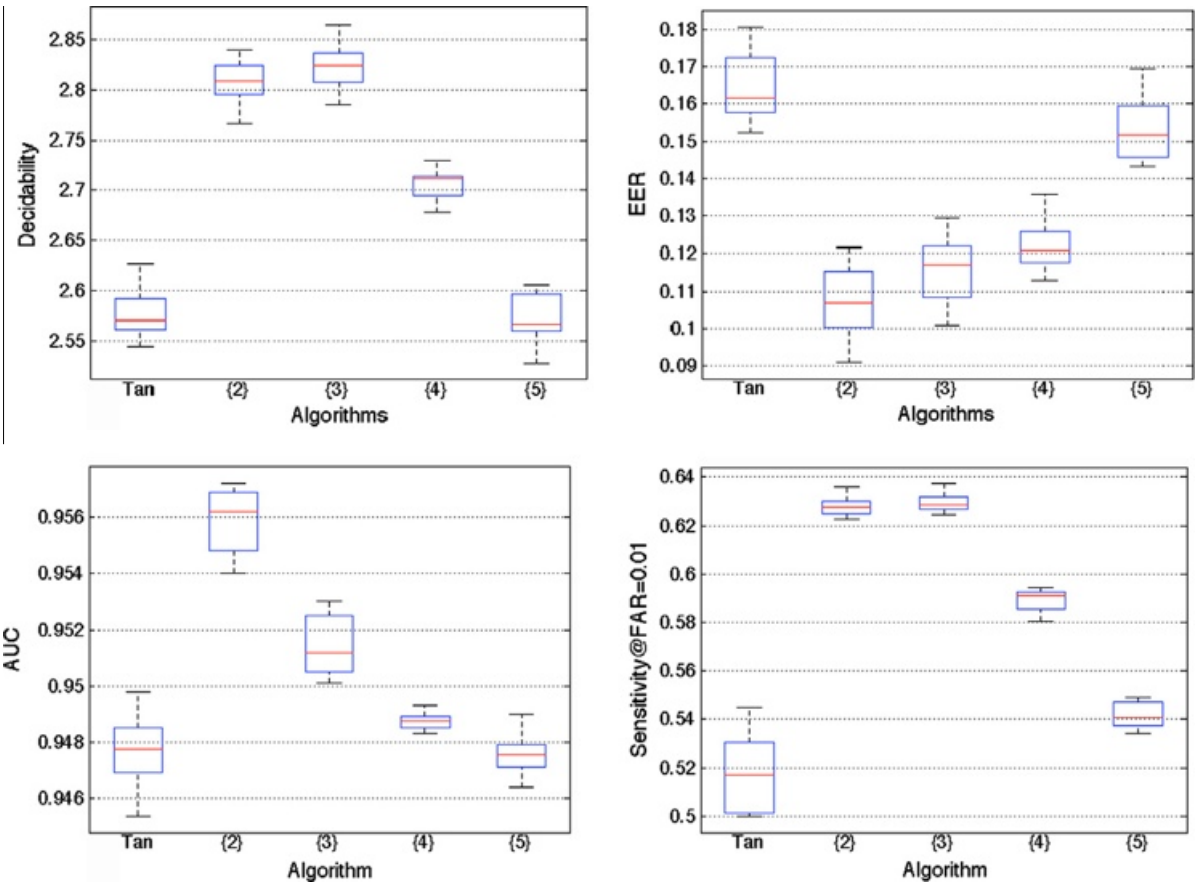


Fig. 16. Comparison between the performance obtained by the best classification ensembles composed of 2–5 recognition methods.

Using the theoretical framework developed by Kittler et al. [30], we tested all of the combinations of the recognition methods described above according to the usual fusion rules: product (*), sum (+), min (m) and max (M). Without any assumption of the prior probabilities, the posterior probability that a pattern $\overrightarrow{x_i}$ belongs to class w_j was obtained by

$$P(w_j|\overrightarrow{x_i}) = \frac{P(\overrightarrow{x_i}|w_j)}{\sum_s P(\overrightarrow{x_i}|w_s)}. \quad (16)$$

An input pattern is assigned to class w_c if $w_c = \arg_j \max \phi P(w_j|\overrightarrow{x_i})$, where ϕ denotes the combination rule. Table 2 lists the best classification ensembles obtained according to the number of fused experts (column #Fused) and the best combination rule observed. Also, a comparison between the performance of such ensembles is given in Fig. 16.

3.5. Degradations in performance

As a summary, Fig. 17 reports the degradation in performance of the proposed method with respect to each of the factors discussed previously. We show the boxplots of the decidability values obtained for data sets with increasingly higher levels of variation (from darker to lighter boxplots). For every factor where the UBIRIS.v2 database has enough images to perform statistically relevant experiments (scale and off-angle), we compared the results obtained in a *homogenous* subset of the data (represented by the horizontal line) and in degraded data sets. For the remaining factors (blur, rotation and lighting), variations were introduced artificially, resulting in different versions of the same data sets, each with different amounts of variation. It can be confirmed that the proposed strategy behaves robustly to changes in scale and rotation and moderately degrades for off-angle and defocused data. The most problematic case was observed for changes in lighting conditions, especially for non-global lighting changes, where performance has degraded substantially.

4. Conclusions

Current state-of-the-art methods to perform iris recognition in VW *real-world* data achieve encouraging performance values that are, however, still far from the demands of the applications of this technology in large-scale identification scenarios. Having analyzed the typical strategies of these approaches, the key innovations of this paper can be summarized in two terms: (1) we propose a recognition scheme based on autonomously defined sub-regions of the iris from which MPEG-7 color and shape descriptors are extracted, achieving performance close to the best-known techniques, and (2) minimal levels of linear correlation between the outputs given by the proposed strategy and state-of-the-art techniques were observed, which suggests that the fusion of evidence between these techniques improved performance. The progress described here is regarded as a positive step towards the development of an extremely ambitious type of biometric recognition.

Regarding further directions of the work given in this paper, some issues can be enumerated: (1) analyze how different color contrast levels and sensor quality would affect the recognition accuracy. Regarding this issue, it is expected that such changes are mainly handled by the retinex phase (several contrast enhancement retinex-based methods are reported in the literature), and should yield different weights (7) for each feature used in the data partition process; and (2) a more objective assessment about the conditions in the environments that enable this type of recognition with enough confidence (specification of the type of illuminants, amount of light and angles of incidence). We are currently working on both these issues.

Acknowledgments

The financial support given by "FCT-Fundação para a Ciência e Tecnologia" and "FEDER" in the scope of the PTDC/EIA/103945/2008 research project "NECOVID: Negative Covert Biometric Recognition" is acknowledged.

References

[1] J.R. Matey, D. Ackerman, J. Bergen, M. Tinker, Iris recognition in less constrained environments, in: Springer Advances in Biometrics: Sensors, Algorithms and Systems, 2007, pp. 107–131.
[2] Honeywell International Inc. A Distance Iris Recognition, 2007. United States Patent 20070036397.
[3] C. Fancourt, L. Bogoni, K. Hanna, Y. Guo, R. Wildes, N. Takahashi, U. Jain, Iris recognition at a distance, in: Proceedings of the 2005 IAPR Conference on Audio and Video Based Biometric Person Authentication, USA, July 2005, pp. 1–13.
[4] P. Meredith, T. Sarna, The physical and chemical properties of eumelanin, Pigment Cell Res. 19 (2006) 572–594.
[5] H. Proença, L.A. Alexandre (Eds.), Special Issue on the Segmentation of Visible Wavelength Iris Images Acquired At-a-Distance and On-the-Move, Image and Vision Computing, vol. 28, no. 2, 2010.
[6] H. Proença, L.A. Alexandre (Eds.), Special Issue on Signatures Encoding and Matching of Segmented Noisy Iris Images, Pattern Recognition Letters, Special Issue on the Recognition of Visible Wavelength Iris Images Acquired On-The-Move and At-A-Distance, in press.
[7] T. Tan, X. Zhang, Z. Sun, H. Zhang. Noisy Iris Image Matching by Using Multiple Cues, Pattern Recognition Letters, Special Issue on the Recognition of Visible Wavelength Iris Images Acquired On-The-Move and At-A-Distance, in press.
[8] Q. Wang, X. Zhang, M. Li, X. Dong, Q. Zhou, Y. Yin, Adaboost and Multi-orientation 2D Gabor-based accurate noisy iris recognition, Pattern Recognition Letters, Special Issue on the Recognition of Visible Wavelength Iris Images Acquired On-The-Move and At-A-Distance, in press.
[9] Weighted Co-occurrence Phase Histogram for Iris Recognition. Pattern Recognition Letters, Special Issue on the Recognition of Visible Wavelength Iris Images Acquired On-The-Move and At-A-Distance, in press.
[10] M. Marsico, M. Nappi, D. Riccio. Iris Recognition in Non-ideal Imaging Conditions. Pattern Recognition Letters, Special Issue on the Recognition of Visible Wavelength Iris Images Acquired On-The-Move and At-A-Distance, in press.
[11] Y. Du, C. Belcher, Z. Zhou, Scale Invariant Gabor Descriptor-Based Noncooperative Iris Recognition, EURASIP J. Adv. Signal Process. 2010 (2010). ID 936512.
[12] E.H. Land, The retinex, Am. Sci. (52) (1964) 247–264.
[13] E. Provenzi, L. Carli, A. Rizzi, Mathematical definition and analysis of the Retinex algorithm, J. Opt. Soc. Am.: A 22 (12) (2005) 2613–2621.
[14] Z. He, T. Tan, Z. Sun, X. Qiu, Towards accurate and fast iris segmentation for iris biometrics, IEEE Trans. Pattern Anal. Mach. Intellig. 31 (9) (2009) 1617–1632.
[15] Y. Du, E. Arslanturk, Z. Zhou, C. Belcher, Video-based non-cooperative iris image segmentation, IEEE Trans. Syst. Man Cybern. – Part B: Cybern. 41 (1) (2011) 64–74.
[16] P. Li, X. Liu, L. Xiao, Q. Song, Robust and accurate iris segmentation in very noisy iris images, Image Vision Comput. 28 (2) (2010) 246–253.
[17] H. Proença, Iris recognition: on the segmentation of degraded images acquired in the visible wavelength, IEEE Trans. Pattern Anal. Mach. Intellig. 32 (8) (2010) 1502–1516.
[18] H. Proença, Quality assessment of degraded iris images acquired in the visible wavelength, IEEE Trans. Informat. Forensics Security 6 (1) (2011) 82–95.
[19] J. Daugman, New methods in iris recognition, IEEE Trans. Syst. Man Cybern. – Part B: Cybern. 37 (5) (2007) 1167–1175.
[20] J.C. Bezdek, J.M. Keller, R. Krishnapuram, N.R. Pal, Fuzzy models and algorithms for pattern recognition and image processing, IEEE Trans. Informat. Forensics Security 6 (1) (2011) 82–95.
[21] M. Dillencourt, H. Samet, M. Tamminen, A General Approach to Connected-component Labeling for Arbitrary Image Representations, Springer, NY, 1999.
[22] L. Cieplinski, MPEG-7 color descriptors and their applications, Lecture Notes Comput. Sci. 2124/2001 (2001) 11–20.
[23] W. Pennebaker, J. Mitchell, JPEG Still Image Data Compression Standard, Van Nostrand Reinhold, New York, USA, 1993.
[24] J. Hajek, Z. Sidak, Theory of Rank Tests, 1st ed., Academic Press, 2000.
[25] R. Jonker, A. Volgenant, A Shortest Augmenting Path Algorithm for Dense and Sparse Linear Assignment Problems Computing 38 (1987) 325–340.
[26] O. Pele, M. Werman, The quadratic-chi histogram distance family, in: Proceedings of the European Conference on Computer Vision, vol. 2, 2010, pp. 749–762.

[27] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Trans. Pattern Anal. Mach. Intellig. 24 (24) (2002) 509–522.
[28] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intellig. 27 (8) (2005) 1226–1238.
[29] Z.I. Botev, J.F. Grotowski, D.P. Kroese, Kernel density estimation via diffusion, Ann. Stat. 38 (5) (2010) 2916–2957.
[30] J. Kittler, M. Hatef, R. Duin, J. Matas, On combing classifiers, IEEE Trans. Pattern Anal. Mach. Intellig. 20 (3) (1998) 226–239.

Chapter 5

Periocular Biometrics: An Emerging Technology for Unconstrained Cenarios

5.1 Overview

This chapter consists of the following article:

Periocular Biometrics: An Emerging Technology for Unconstrained Scenarios
Gil Santos and Hugo Proença
IEEE Symposium on Computational Intelligence in Biometrics and Identity Management – CIBIM '13, April 16-19, 2013, Singapore.
ISBN: 978-1-4673-5879-8/13

Periocular Biometrics: An Emerging Technology for Unconstrained Scenarios

Gil Santos and Hugo Proença
IT - Instituto de Telecomunicações
Universidade da Beira Interior
Covilhã, Portugal
Email: gmelfe@ubi.pt; hugomcp@di.ubi.pt

***Abstract*—The periocular region has recently emerged as a promising trait for unconstrained biometric recognition, specially on cases where neither the iris and a full facial picture can be obtained. Previous studies concluded that the regions in the vicinity of the human eye - the periocular region- have surprisingly high discriminating ability between individuals, are relatively permanent and easily acquired at large distances. Hence, growing attention has been paid to periocular recognition methods, on the performance levels they are able to achieve, and on the correlation of the responses given by other. This work overviews the most relevant research works in the scope of periocular recognition: summarizes the developed methods, and enumerates the current issues, providing a comparative overview. For contextualization, a brief overview of the biometric field is also given.**

I. Introduction

Due to increasing concerns on security and safety of modern societies, biometrics has emerged in the last decade as a major domain of knowledge and has been motivating significant research efforts. Considering the outstanding levels of performance that currently deployed biometric systems achieve, the interest now in putted in the development of systems able to work in uncontrolled acquisition environments, which significantly increases the challenges on reliable recognition. In this setup, alternatives are sought [1] by improving the existing algorithms, by using multi-modal systems or exploring new traits. Despite a broad variety of traits that has been researched, the classical traits to perform *at-a-distance* recognition are the face and the iris.

The face is the most widely used biometric trait. Everyday and even without noticing it, we all use facial information to recognize each other. Not only that, it become one of the most successful applications of image analysis and understanding. Being non-intrusive and allowing cover acquisition, it became preferable over very reliable traits like the iris or fingerprint when aiming at less constrained subject recognition. Several commercial face recognition systems are now available, and a lot of techniques were developed [2] for both still images and video. Face recognition approaches are either based on a global analysis of the whole region as a set of pixel intensities, or the relation between facial attributes, their location and shape.

The iris texture has a predominantly randotypic morphogenesis unique for each individual and allows very high recognition accuracy, which justifies the efforts being held on iris biometrics research [3] and its quick ascent as one of the most popular biometric traits. While most of the commercially deployed iris recognition systems work with constrained near-infrared (NIR) data that favors perception of its patterns whilst reducing the number of noise factors associated, literature on extending this biometrics usability to "relaxed" visible wavelength (VW) setups has broaden [4]–[6]. However, iris performance as a biometric trait is severely impacted in non-ideal setups, and its relatively reduced size and moving profile make it difficult to image at-a-distance and without user cooperation.

The periocular region represents a trade-off between the whole face and the iris alone. Containing the eye and its immediate vicinity, it covers eyelids and eyelashes, nearby skin area and eyebrows. Its use as a biometric trait has emerged, constituting nowadays a strong alternative for less constrained environments, when image acquisition is not reliable, and to avoid spoofing of the iris patterns. It is easy to acquire without user cooperation and does not require a constrained close capturing. Also, this region is not so affected by the aging process as other facial regions are, as for instance the mouth and cheek whose skin become loosened over time.

Periocular biometrics can be used alone or complementary to iris recognition, considering that the use of multiple traits might be specially important to compensate for the adversity of the environments and uncooperative subjects. Most times, the responses of periocular methods are fused at the score level to the corresponding iris scores, due to their spatial proximity and to the fact that a single camera might be able to acquire both traits. Being relatively stable and rarely occluded, it's particularly useful when the subject is wearing a mask or otherwise only exposing their eyes. Although this paper is focused on recognition, periocular biometrics as proven to be suitable for other purposes as well (*e.g.* [7], [8]).

In terms of features of the periocular region, they can be divided into two levels, as suggested by [9]: the first level comprise the eyelids, eye folds, and eye corners; and the second level comprises the skin texture, wrinkles, color and pores. Analysis of those features can be carried on based on their geometry, texture or color.

As described by Park et al. [10], the problems that arise from periocular recognition can be summarized as follows:
Imaging: What would be the optimal spectrum band for periocular biometrics? Is VW, more advantageous on covert biometrics, fit for this trait?
Region definition What are the actual "boundaries" of the periocular region? Should iris, sclera or the eyebrows be

included or masked/cropped?
Encoding Which features would better describe and discriminate this region? How reliable would they be when relaxing imaging conditions?
Matching What's the best matching scheme for those features? Will coarse classification be of any use?
Fusion What would be the benefit on fusing periocular with other traits? Which ones, and how to fuse them?

The remainder of this paper is organized as follows: Sections II and III overview the recognition systems and existing datasets; Section IV comparatively details the relevant methods developed on periocular recognition; and finally Section V present some final considerations.

II. Biometric System

The importance of the biometric authentication system must not be disregarded, as it will be the responsible for carrying the whole process, from the data acquisition, to feature extraction, and matching against the database. Therefore, designing a system that adapts to its application scenario is most important. In a general way, a recognition system is composed of four modules [11]:

1) Sensor Module: A wide variety of sensors are available, depending on which biometric trait we are going to work with. Since most of the biometric traits consist on visual data, cameras will be used for acquisition. On real-time systems, the balance between the richness on detail of the acquired data and the acquisition rate is essential, and therefore choosing a proper camera also is. This module is strictly related with the first step of recognition systems (trait acquisition) and is where the trade-off between the quality of the gathered data and user cooperation is set.

2) Quality assessment and Feature extraction: Even with an optimal sensor setup, not always the acquired data is suited for feature extraction. Therefore, its quality is usually assessed, and the image discarded if no minimum requirements are met, thus saving time in additional processing. The trait needs to be properly located and segmented (specially useful to gather preferably "good" data), and then encoded as feature templates.

3) Matching and decision-making: In this module, features are matched against the templates in the database, thus deciding either to be in the presence of a genuine or impostor comparison.

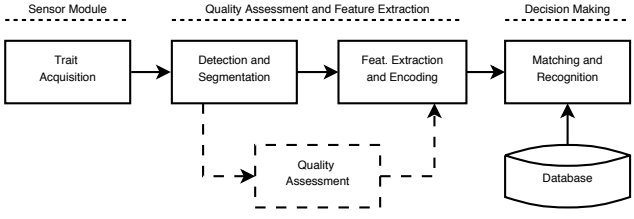


Fig. 1. General steps and elements of biometric recognition systems.

4) System database: This module consists on the repository of user biometrics and other identification information, which is acquired during the enrollment stage, and used for later identification or verification of users' identity.

III. Datasets

Only a few public datasets were designed for the development of periocular recognition methods. Instead, face and iris databases are generally used for that purpose. The most commonly used databases for the evaluation periocular methods are now introduced[1], and their specifications summarized at Table I.

A. FERET

The Facial Recognition Technology (FERET) database [12] was designed as a standard for developing face recognition methods, and acquired at George Mason University over 11 sessions and a three years period (1993 to 1996). Initially released as low resolution (256×384 pixels) grayscale data, years later a high-resolution color version was also disclosed. A total of 14051 images were gathered from 1199 different subjects. Image acquisition protocol contemplates a semi-controlled environment, with strict expression, pose and illumination changes.

B. FRGC

Collected at the University of Notre Dame, the Face Recognition Grand Challenge (FRGC) database [13] consists of high resolution ($\approx 1200 \times 1400$ pixels) color still images, captured on both controlled and uncontrolled environments. The controlled subset was captured on a studio under uniform illumination, where subjects were required to stand still while looking straight at the camera and essay neutral and smiling expressions. As for the uncontrolled acquisition, images were shoot in different scenarios, disregarding both background and illumination. Data is split into a training partition of 12776 images from 275 subjects, and a testing partition of 24042 images from 466 subjects, 6 images per session for each subject in both partitions. Illumination is not regular, as the illumination bursts for a short period of time, and main noise factors are observable (eye blink, motion blur, occlusions, reflections). Acquired data is stored on 2048×2048, 15 frames per second (fps) AVI files, where iris spatial extension is about 120 pixels [14].

[1] Although not so common, the FC-NET database will be included by its relevant facial aging characteristics.

(a) FERET (b) FRGC (c) MBGC (d) UBIRIS2 (e) FG-NET

Fig. 2. Sample images from the commonly used datasets on evaluating periocular algorithms. Except from (d), data has been cropped for illustration purposes.

TABLE I. OVERVIEW OF DATABASE SPECIFICATIONS. VARYING ELEMENTS ARE DISTANCE (D), EXPRESSION (E), ILLUMINATION (I), OCCLUSION (O) AND POSE (P).

Name	Images	Subj.	Dimensions	Variations
FERET	14051	1199	512×768	E, I, P.
FRGC	36818	741	$\approx 1200 \times 1400$	E, I.
MBGC	149 AVI	114	2048×2048	D, E, I, O, P.
UBIRIS.v2	11102	261	800×600	D, O, I.
UBIPr	10950	261	Multiple	D, I, O, P.
FG-NET	1002	82	$\approx 400 \times 500$	D, E, I, P.

C. UBIRIS.v2

The UBIRIS.v2 is a unconstrained iris database [15], captured on the VW from moving subjects, at different distances and challenging illumination conditions, simulating realistic acquisition issues with the associated noise factors. Data for both eyes is separately available, as well as the surrounding periocular data, thus being prone to stress not only robust iris related methods for the visible spectrum, but periocular ones and their fusion as well. The 11102 acquired images represent a total of 261 subjects, from different ages and ethnicities.

D. UBIPr

This newly created UBI Periocular Recognition (UBIPr) database, by Padole and Proença [16], represent a renewed effort to advance periocular biometric research, providing new means of evaluating robust methods, at "higher levels of heterogeneity".

In opposition the most common datasets used for periocular method evaluation, noise factors were actually introduced through acquisition setup: varying acquisition distance, irregular illumination, pose and occlusion. In addition, database manual annotation include ROI and essential landmarks.

Dimensions vary, accordingly to the acquiring distance, between 501×401 (8m) and 1001×801 (4m).

E. FG-NET

FG-NET is a facial aging database with around one thousand images from 82 subjects, 0 to 69 years old. Captured with different acquisition setups and many years apart, subjects have clear changes in illumination, pose and expression. Images are 400×500 pixels in size, captured on VW, and for each one a 68 facial landmark points annotation is also provided.

IV. RELEVANT RESEARCH

In this section we will detail the relevant research on periocular biometric recognition, providing at Table III a summarized overview over the described methods and reported results.

A. Park et al. [10], [17]

Park et al. pioneering approach [17] explored the recognition capabilities of the periocular region. Feature extraction is divided in two approaches: local and global, as information concerns local regions, or is extracted from the whole image (or, in this case, several region of interest (ROI)).

For global feature extraction images are properly aligned using iris center and radius as reference. Although authors acknowledge eye corners to be more fit for such task [10], they claim that such points cannot be reliably determined. Then, two well-known distribution-based descriptors are employed, namely Histogram of Oriented Gradients (HOG) [18] and Local Binary Patterns (LBP) [19], [20]. Values are computed for a given ROI independently, and then quantized into 8-bin histograms. The ROI are contiguous squares, where the side equals in length the iris radius, forming a 7 by 5 grid centered on the iris. Those histograms, combining shape and texture information, are merged into a single-dimension array, easily matchable to an identical one (from another image) simply by computing the Euclidean distance.

As for the local features, Scale-Invariant Feature Transform (SIFT) [21] allowed the detection of a set of key-points, encoded with their surrounding pixels information, and compared against their counterparts from the testing image. SIFT offers invariance to translation, scaling and rotation.

Tests were conducted over a "small" (899 images, 30 subjects, 2 sessions) database of frontal periocular images, acquired in the VW. Although face matching achieving 100% rank-1 recognition accuracy, the reported recognition for periocular range from 62.5% when using HOG features, to 80.8% when fusing them with SIFT results. Curiously, combining the three descriptors didn't overcome those results, although joint performance was very close: 80%.

On their later work [10], authors went further on stressing periocular applicability for biometric recognition, analyzing the impact of diverse factors over performance: eyebrow inclusion or disguising, automatic segmentation, side information, iris and sclera masking and expression variation.

As expected, results highlighted eyebrow information importance, being more significant over SIFT where improvements reached almost 19%. Nonetheless, the eyebrow inclusion is more favorable over manual segmentation, as its performance degraded when using automatic segmentation through OpenCV, which was not observed on "eyebrow-less" data. Facial side information, on the other hand, can be considered almost irrelevant, since performance variation from both to same side matching didn't go behind 1% except for SIFT on 2 of the 48 test setups.

Changes in subjects' expression significantly lowered the performance of LBP and HOG, although on SIFT, more robust to distortions, a slightly increase was registered. Masking the iris and the entire eye also caused performance to decrease, this time being SIFT the more disfavored. Top accuracy for single classifiers was 79.49%, achieved through SIFT on unmasked periocular images, manually segmented with the eyebrow, when compared to an image captured from the same side and expression. As reported in their prior paper [17], score level fusion didn't represent a significant performance improvement.

The authors also simulated periocular recognition over non-ideal conditions, performing four simple tests: result comparison against recognition with partial (occluded) facial and periocular images; conducting cosmetic changes on the eyebrows; template aging; and perspective variations. For the first step, they used FaceVACS[2] face recognition system, whose 99.77% recognition accuracy on "clear" face images, dropped

[2]FaceVACS SDK available at: http://www.cognitec-systems.de

to 39.55% simply by occluding the lower region. Occluding the periocular region is also an element of concern, since relatively low occlusions lead to significant decay on performance. Without score fusing the feature encoding methods, 10%, 20% and 30% periocular occlusion led to accuracies no greater than 25.97%, 20.51% and 10.12% respectively (all with SIFT).

On eyebrow modifications, the TAAZ[3] tool was used to simulate eyebrow makeover, producing a decay from 7.5% on LBP to 10% on the other descriptors. The tests regarding pose effect were the ones with greater impact over periocular recognition accuracy, specially when using SIFT. Apart from frontal images, subjects shoot with $15°$ and $30°$ rotation of the head, produced a 35% and 45% decay on this method's accuracy, respectively. Finally, another concern the authors rise is the apparent tendency of the periocular region not to be stable over relatively small amounts of time. Images captured 3 months apart from each other have up to 15% less accuracy, and about 30% on only half an year.

As further work, multi-spectral analysis is suggested, along with improvements on the alignment and matching methods. Fusion with iris or face recognition is also not discarded.

B. Miller et al. [22], [23]

Miller et al. [22] analyze periocular skin texture using Uniform Local Binary Patterns (ULBP) alone, with some deeper insights on each region's impact on the recognition process. The ULBP, as it name states, is an LBP-based method, with "improved rotation invariance with uniform patterns and finer quantization of the angular space" [24].

At a first stage, the periocular region is cropped proportionally to the distance between the eyes, and scaled to 100×160 pixels. Then, a 7 by 4 grid of square ROIs is defined, centered on the eye, and iris and sclera texture effects are eliminated overlapping an elliptical neutral mask to the image. Each ROI's histogram is normalized, and ULBP calculated using an 8-pixel neighborhood. As such neighborhood produces 59 different possible results, 59-bin histograms are populated with the result count, and then merged to produce a single-dimension array as the final periocular signature. Manhattan distance is used for subject identification against the database.

Experiments were conducted on subsets of the FRGC and FERET databases, for the left and right eyes separately and both eyes together. Recognition rates were around 84% and 71% for each eye individually, and 90 and 74% for both eyes together, on FRGC and FERET respectively.

Further to this work, Miller et al. [23] conducted deeper analysis on image quality impact over periocular local texture based recognition, namely changing blur, resolution and illumination, while comparing the results with similar experiments conducted with the entire face.

As preprocessing, the periocular region was cropped from the FRGC database in proportion to the distance between the eyes, and then resized to a square region with 251 pixel long sides. Upon grayscale conversion, image histogram is equalized and the eye is masked. Texture is then encoded using LBP over a regular block division of the image, and values used to populate an histogram, similarly to other periocular approaches.

Image blurring was achieved through Gaussian filter convolution, and results showed that even though face being far less affected by small amounts of blur than periocular, this last trait is slightly better at high blur levels. As for resolution, images were down-sampled up to 40% its original size, and behavior was similar to the one of blurred images.

Illumination variation was not simulated, since the FRGC database already contains both controlled and uncontrolled acquired images. The low accuracy verified when matching pairs of images captured on uncontrolled setups suggest that local appearance approaches like LBP are not suited for irregular lighting conditions.

Finally, information differences from one color channel to the others were also analyzed. Conclusions show the green channel as the more discriminant, with accuracy levels $\approx 23\%$ higher than for the red channel (which is presented as the less discriminant). In fact, when fusing scores from all three channels, the red contribution only lowers the overall performance. Blue channel has similar texture information as the green one.

In a general way, periocular was proven to outperform face recognition in the stressed setups.

Further work includes conducting the same tests for different classification methods, possibly adapting Support Vector Machines (SVM) usage as suggested by Savvides et al. [25].

C. Adams et al. [26]

Adams et al. extended Miller's work [22], proposing the usage of a Genetic & Evolutionary Computing (GEC) method to optimize the original feature set.

The first stage of feature extraction was conducted as described by Miller et al. [22], and on the second stage the Genetic & Evolutionary Feature Extraction (GEFE) chosen was the Steady-State Genetic Algorithm (SSGA), as implemented by the NASA's eXploration Toolset for Optimization of Launch and Space Systems (X-TOOLSS)[4].

Reported results were about 86% accuracy for either eye on the FRGC database, and 80% on similar experiments for the FERET. Best results were obtained when using both eyes: 85% and 92% for those same datasets.

The usage of GEC represented an improvement of at least 10%, and only $49 \approx 52\%$ of the initial features were used. Nevertheless, the selected algorithm was not proven to be the optimal for that specific periocular features.

D. Juefei-Xu et al. [27], [28]

Inspired by the work of Park et al. [17], the authors decided to expand their experiments to less ideal imaging environments, evaluating the performance of different feature schemes over the FRGC database [27].

In addition to LBP and SIFT, both local and global feature extraction schemes were stressed: Walsh masks [29], Law's masks [30], DCT [31], DWT [32] Force Fields [33],

[3]Free virtual makeover took, available at http://www.taaz.com

[4]http://nxt.ncat.edu/

SURF [34], Gabor Filters [35] and Laplacian of Gaussian (LoG). The LBP itself was tested while applied over some of the other methods (Table II). For matching, different distance metrics were tested: Normalized Cosine, Euclidian and Manhattan.

TABLE II. RANK-1 ACCURACY FOR LBP FUSION WITH OTHER METHODS [27].

Fused methods	Accuracy
LBP + LBP	42.5%
Walsh Masks + LBP	52.9%
Laws' Masks + LBP	51.3%
Discrete Cosine Transform (DCT) + LBP	53.1%
Discrete Wavelet Transform (DWT) + LBP	53.2%
Force Field Transform + LBP	41.7%
Gabor Filters + LBP	12.8%
LoG Filters + LBP	30.9%

Experiments shown that best results were registered when using local descriptors, and the post-application of LBP was translated in a performance boost (Table II). Top accuracy of 53.2% was attained with DWT + LBP, followed closely when combining this last one with DCT (53.1%) and Walsh or Laws' Masks (52.9% and 51.3%).

Worst outcomes were registered for SIFT and Speed Up Robust Features (SURF), with a Verification Rate (VR) no greater than 1%, possibly due to the low resolution of the images.

Later on [28], they addressed the aging effect on periocular recognition, reported to be an issue by several authors (e.g. Park et al. [10]), even at relatively small time lapses (months). This important issue is not trivial, as modeling the aging process would require large datasets, and the decoding of its dependence on external factors, as ethnicity, gender, etc. The authors method was developed and validated on images from the FG-NET database, taken years apart at different acquisition setups, thus also dealing with illumination, pose and expression issues.

Their method starts by preprocessing the periocular region: pose is corrected through Active Appearance Models (AAM), illumination is dealt with anisotropic diffusion model, and region is normalized using the landmark points provided with the database. Next step is feature extraction using Walsh-Hadamard transform encoded LBP (WLBP), followed by unsupervised discriminant projection (UDP) [36] application that boosted results to very high performance levels.

Results show UDP to give better accuracy than Principal Component Analysis (PCA) and Locally Preserving Projections (LPP) by up to 40%. As for WLBP, results were 15% better than raw pixel intensity matching, and pose correction resulted in a 20% improvement. Finally, the proposed method for the tested images resulted in a complete 100% identification accuracy.

E. Bharadwaj et al. [37]

Bharadwaj et al. propose the combination of a global matcher (GIST) with ULBP for periocular recognition over VW uncooperative images from UBIRIS.v2 database.

The GIST algorithm consists on combining five perceptual dimensions, usually associated with scene description [38]: naturalness,openness, roughness, expansion and ruggedness.

When computing the global GIST descriptor, and to achieve local contrast normalization, the image is preprocessed with Fourier transform. Then, the spatial envelope is computed using a set of Gabor filters (4 scales $\times$ 8 orientations, producing a 1536 element GIST descriptor).

The ULBP is computed over the original image, sliced into 64 patches (producing a 64×256 descriptor).

For both descriptors, matching is computed using χ^2 distance, and min-max normalized results from both eyes are fused simply by using a weighted sum.

Results showed that GIST overperformed ULBP, with Rank-1 accuracy around 62% for the regions separately, and 70.82% for their fusion. The ULBP performance was around 53%, and 63.77% when fusing both region results. When combining both descriptors, accuracy was boosted to 73.65%.

F. Woodard et al. [9], [14]

In their work, Woodard et al. [9] aimed at evaluating periocular performance, thus determining its usability as a biometric trait over NIR and VW data. Their analysis is focused only on second level features (texture and color).

As pre-processing, periocular slice of images is cropped, and an elliptical mask overlapped to the iris and sclera region for "unbiased" periocular analysis. Cropped color images from the FRGC are scaled down to 100×160, while the periocular NIR frames from the Multi Biometric Grand Challenge (MBGC) are 601×601 pixel.

Texture features were encoded the same way for both databases, through LBP computation over a ROI grid, which was then quantized into histograms. As for the color information on FRGC images, it was encoded using color histograms for red and green channels. On this database, score level fusion was used to combine texture and color results. Matching was achieved using Manhattan distance for LBP and Bhattacharya distance for color histograms.

Results suggest texture information to be more discriminant that color, and score fusion only slightly improves overall performance. As a comparison term, reported texture based accuracy was around 90% and 88% on the VW, and 81% and 87% on NIR for the left and right periocular regions respectively.

On their later work, Woodard et al. [14] make use of the periocular region texture information to improve iris data reliability, aiming at overcoming the difficulties when dealing with non-ideal imaging.

Tests were conducted over MBGC that, although being a NIR database, is a challenging one for iris recognition due to at-a-distance in-motion subjects and illumination variations. Frames were treated as described above, with texture measured computing LBP the same way. Iris processing was as of Daugman's [39], except for the segmentation that was manually performed to avoid further errors. Both methods' results were then normalized using min-max scheme, and combined by a simple weighted sum.

Results demonstrate iris' poor accuracy ($10.1\% \approx 13.8\%$) to benefit from fusing with periocular results, raising rank-1 to 96.5%.

G. Padole and Proença [16]

Padole and Proença also stressed how noise factors deteriorate periocular recognition, using natural images where those factors were included by the acquisition framework instead of simulating them: pose variation, distance of the subject, pigmentation and occlusion.

Inspired by the work of Park et al. [17], they used the same feature extraction techniques, except that ROI center was computed with relation to eye-corners instead of iris center. This new alignment method led to most significant improvements, specially since in unconstrained biometrics gaze variations are more prone to happening.

On score level fusion, linear an non-linear methods were also tested: logistic regression [40] and Multi Layer Perceptron (MLP) respectively. Although the last one reported to lead to slightly better results, difference was not significant.

For the stressed covariates, interesting conclusions were reached. Results shown that closer acquired distances didn't led to better performance, and neither did very large ones. Worst results were obtained for images acquired at 4 m, and though highest stressed distance was 8 m, top performance was obtained at 7 m. Not surprisingly, pose variation impact on performance was in inverse proportion: higher tilting angle result in lower accuracy values. Same as for the occlusion.

Finally, iris pigmentation was reported to also impact periocular recognition performance, specially on heavily pigmented ones which lead to lower accuracy. Best results were obtained for medium pigmented irides.

Another interesting discovery was that subject gender affects recognition rates. More precisely, female subjects are easily identified using periocular biometrics than male ones.

H. Hollingsworth et al. [41], [42]

The human ability to use contextual information and to "disregard" most of noise factors adapting itself to surrounding conditions is outstanding, making it a harder task for machines to mimic. In fact, recognition algorithms should not try to just mimic the human perception system, but to understand its way of working, and then seek alternate strategies to tackle the same issues.

Hollingsworth et al. understood existing methods to have overlapped that step. Having that in mind, they [41] established parallelisms between human perception and automatic recognition systems, identifying which ocular elements humans find more useful for periocular recognition.

On their essay, 640×480 NIR images were acquired from 120 subjects using an iris camera (LG2200), and the iris was completely masked to avoid biased answers. Only periocular from eyes' tight vicinity is visible, with some features used by other methods partially hidden (e.g. eyebrows). 80 pairs of images were presented to 25 human observers, who were asked to tell if they belong to the "same person" or "different people", and how "certain" they were. Further to that, the observer had to individually rate each one of the features' helpfulness, in a three level scale. Results showed eyelashes to be the most helpful periocular feature, closely followed by the *medial canthus* and the eye shape. The observers based themselves on eyelash clusters, density, direction, length and intensity. To the human observers, skin was actually the less useful. Average human accuracy was 92%.

On their later work [42], similar tests with human observers were widen to the VW band, with a more extent study on new factors. The algorithms suggested by Park et al. [17] were also implemented for periocular performance comparison, and irides were evaluated using the IrisBEE biometric system from ICE [43].

Trial data was also widen to 210 subjects, imaged on the same controlled fashion with a setup as above, and on the VW using a Canon D80 camera. The amount of observers also increased to 56, to whom 140 pairs of images were presented for each one of the four sets of experiments built: NIR and VW, periocular and iris images. Test subjects could then rank their certainty of a positive match in a 5 level scale, and for the periocular images they had to specify how helpful individual features were ("eye shape", "tear duct", "outer corner", "eyelashes", "skin", "eyebrow", "eyelid", "color", "blood vessels" and "other").

Human NIR periocular recognition accuracy dropped to 78.8%, probably due to the different pairing system and limited observation time, and VW performance was set on 88.4%. Machine results were similar, within a 1% difference on overall accuracy. The features identified as fit for periocular NIR region were similar to the ones at [41], but for VW data changes occurred: blood vessels, skin and eye shape were reported to be more helpful than eyelashes.

When acquiring data on VW band, differences on acquired skin details are perceptible. Also with the LG2200 camera illumination, being designed for iris recognition, usually causes skin saturation. As so, VW band was found to be preferable for periocular recognition tasks.

Human perception of iris features is greater on NIR images, leading to 85.6% accuracy against 79.3% on VW. However, and unlike periocular, machines recognition was 13% better, on average, than human observers, with 100% and 90.7% accuracy for those same bands.

V. Conclusions

The interest on the periocular region as a biometric trait has justifiably increased over the last years, considering the pioneer approach of Park et al. [17] a starting point. Subsequently, even simple algorithms led to fair performance levels, and the surprisingly good response of LBP based methods (like ULBP and WLBP) is noteworthy.

The recently developed methods focus mainly on texture analysis and keypoint extraction. Periocular is currently regarded as specially suitable for unconstrained and uncooperative scenarios, where iris cannot be properly imaged and neither a full facial picture can be obtained. Also, results favoring VW periocular over NIR also show its fitness for more relaxed setups and for its use based on conventional surveillance cameras.

TABLE III. OVERVIEW OF THE MOST RELEVANT PERIOCULAR RECOGNITION METHODS.

Approach	Features	Extract	Classifier	Dataset	Accuracy
Park et al. [17]	Shape, Texture, Key-Points	HOG, LBP, SIFT	Euclidean distance, SIFT matcher	899 VW images, 30 subjects, 2 sessions	HOG: 62.5%, LBP: 70.0%, SIFT: 74.2%, Best: 80.8%
Miller et al. [22]	Texture	ULBP	Manhattan distance	FRGC, FERET	FRGC: 89.8%, FERET: 74.1%.
Adams et al. [26]	Texture	LBP +GEFE	Manhattan distance	FRGC, FERET	FRGC: 92.2%, FERET: 85.1%.
Woodard et al. [9]	Color, Texture	RG color histogram, LBP	Bhattacharya, Manhattan distance	FRGC, MBGC	Left VW peri: 90% Right VW peri: 88% Left NIR peri: 81% Right NIR peri: 87%
Woodard et al. [14]	Texture	Daugman's irisCode, LBP	Hamming distance, Manhattan distance	MBGC	Left Iris: 13.8% Left Peri: 92.5% Fusion: 96.5% Right Iris: 10.1% Right Peri: 88.7% Fusion: 92.4%
Juefei-Xu et al. [27]	Texture, Key-Points	Walsh Masks, Laws' Masks, DCT, DWT, Force Field Transform, Gabor Filters, LBP, SIFT, SURF,	Cosine distance, Euclidean distance, Manhattan distance	FRGC	DWT+LBP: 53.2% DCT+LBP: 53.1% Walsh+LBP: 52.9% Laws'+LBP: 51.3% ...
Juefei-Xu et al. [28]	Texture	WLBP+UDP	Cosine distance	FG-NET	100%
Bharadwaj et al. [37]	Naturalness, Openness, Roughness, Expansion, Ruggedness, Texture	GIST, ULBP	χ^2 distance	UBIRIS.v2	GIST: 70.82% ULBP: 63.77% Fusion: 73.65%
Hollingsworth et al. [41]	Human	Human	Human	NIR images, 120 subject	92%
Hollingsworth et al. [42]	Human	Human	Human	NIR and VW, 210 subjects	NIR Peri: 78.8% VW Peri: 88.4% NIR Iris: 85.6% VW Iris: 79.3%

However, some issues remain to be properly addressed, specially the about poses, occlusions and aging. Regarding the latter, extending Juefei-Xu et al [28] work to different scenarios should be considered.

The work of Hollingsworth et al. [41], [42] on human perception suggests that eye shape constitutes a powerful ally to the skin analysis methods on both spectral bands, thus making us rethink periocular recognition, possibly taking a leap away the overused texture methods. Eyelashes are also pointed as a good indicator, specially for NIR, keeping in mind that images differ from the "traditionally" used periocular images and the close capturing of the data could have biased the results. Those issues should be addressed in further work, as well as a more complete and uniform study of existent methods' performance over the UBIPr dataset.

ACKNOWLEDGMENT

The financial support given by "FCT - Fundação para a Ciência e Tecnologia"and "FEDER"in the scope of the PTDC/EIA/103945/2008 research project "NECOVID: Negative Covert Biometric Recognition"is acknowledge. Also, the support given by IT - Instituto de Telecomunicações in the scope of the "NOISYRIS"research project is acknowledge too.

REFERENCES

[1] K. Ricanek, M. Savvides, D. Woodard, and G. Dozier, "Unconstrained biometric identification: Emerging technologies," *Computer*, vol. 43, no. 2, pp. 56 –62, February 2010.

[2] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, December 2003.

[3] K. Bowyer, K. Hollingsworth, and P. Flynn, "Image understanding for iris biometrics: A survey," *Comput. Vis. Image Underst.*, vol. 110, no. 2, pp. 281–307, May 2008.

[4] K. Y. Shin, G. P. Nam, D. S. Jeong, D. H. Cho, B. J. Kang, K. R. Park, and J. Kim, "New iris recognition method for noisy iris images," *Pattern Recognition Letters*, vol. 33, no. 8, pp. 991 – 999, 2012.

[5] T. Tan, X. Zhang, Z. Sun, and H. Zhang, "Noisy iris image matching by using multiple cues," *Pattern Recognition Letters*, vol. 33, no. 8, pp. 970 – 977, 2012, ¡ce:title¿Noisy Iris Challenge Evaluation II - Recognition

of Visible Wavelength Iris Images Captured At-a-distance and On-the-move¡/ce:title¿.

[6] G. Santos and E. Hoyle, "A fusion approach to unconstrained iris recognition," *Pattern Recognition Letters*, vol. 33, no. 8, pp. 984 – 990, 2012.

[7] J. Lyle, P. Miller, S. Pundlik, and D. Woodard, "Soft biometric classification using periocular region features," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, September 2010, pp. 1 –7.

[8] J. Merkow, B. Jou, and M. Savvides, "An exploration of gender identification using only the periocular region," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, September 2010, pp. 1 –5.

[9] D. Woodard, S. Pundlik, J. Lyle, and P. Miller, "Periocular region appearance cues for biometric identification," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, June 2010, pp. 162 –169.

[10] U. Park, R. Jillela, A. Ross, and A. Jain, "Periocular biometrics in the visible spectrum," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 1, pp. 96 –106, 2011.

[11] A. Jain, P. Flynn, and A. Ross, Eds., *Handbook of biometrics*. Springer, 2008.

[12] P. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The feret evaluation methodology for face-recognition algorithms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 10, pp. 1090 – 1104, October 2000.

[13] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, June 2005, pp. 947 – 954 vol. 1.

[14] D. Woodard, S. Pundlik, P. Miller, R. Jillela, and A. Ross, "On the fusion of periocular and iris biometrics in non-ideal imagery," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, August 2010, pp. 201 –204.

[15] H. Proença, S. Filipe, R. Santos, J. Oliveira, and L. Alexandre, "The ubiris.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 8, pp. 1529 –1535, August 2010.

[16] C. Padole and H. Proença, "Periocular recognition: Analysis of performance degradation factors," in *Biometrics (ICB), 2012 5th IAPR International Conference on*, April 2012, pp. 439 –445.

[17] U. Park, A. Ross, and A. Jain, "Periocular biometrics in the visible spectrum: A feasibility study," in *Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on*, September 2009, pp. 1 – 6.

[18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *In Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.

[19] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in *Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision amp; Image Processing., Proceedings of the 12th IAPR International Conference on*, vol. 1, October 1994, pp. 582 –585 vol.1.

[20] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, January 1996.

[21] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, November 2004.

[22] P. Miller, A. Rawls, S. Pundlik, and D. Woodard, "Personal identification using periocular skin texture," in *Proceedings of the 2010 ACM Symposium on Applied Computing*, ser. SAC '10. New York, NY, USA: ACM, 2010, pp. 1496–1500.

[23] P. Miller, J. Lyle, S. Pundlik, and D. Woodard, "Performance evaluation of local appearance based periocular recognition," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, September 2010, pp. 1 –6.

[24] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971 –987, July 2002.

[25] M. Savvides, R. Abiantun, J. Heo, S. Park, C. Xie, and B. Vijayakumar, "Partial holistic face recognition on frgc-ii data using support vector machine," in *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on*, June 2006, p. 48.

[26] J. Adams, D. Woodard, G. Dozier, P. Miller, K. Bryant, and G. Glenn, "Genetic-based type ii feature extraction for periocular biometric recognition: Less is more," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, August 2010, pp. 205 –208.

[27] F. Juefei-Xu, M. Cha, J. Heyman, S. Venugopalan, R. Abiantun, and M. Savvides, "Robust local binary pattern feature sets for periocular biometric identification," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, sept. 2010, pp. 1 –8.

[28] F. Juefei-Xu, K. Luu, M. Savvides, T. Bui, and C. Suen, "Investigating age invariant face recognition based on periocular biometrics," in *Biometrics (IJCB), 2011 International Joint Conference on*, October 2011, pp. 1 –7.

[29] T. Beer, "Walsh transforms," *American Journal of Physics*, vol. 49, no. 5, pp. 466–472, 1981.

[30] K. Laws, "Rapid texture identification," in *Proc. SPIE Conf. Image Processing for Missile Guidance*, 1980, pp. 376–381.

[31] N. Ahmed, T. Natarajan, and K. Rao, "Discrete cosine transform," *Computers, IEEE Transactions on*, vol. C-23, no. 1, pp. 90 –93, January 1974.

[32] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 7, pp. 674 –693, July 1989.

[33] D. Hurley, M. Nixon, and J. Carter, "A new force field transform for ear and face recognition," in *Image Processing, 2000. Proceedings. 2000 International Conference on*, vol. 1, 2000, pp. 25 –28.

[34] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[35] D. Clausi and M. Jernigan, "Towards a novel approach for texture segmentation of sar sea ice imagery," in *26th International Symposium on Remote Sensing of Environment and 18th Annual Symposium of the Canadian Remote Sensing Society*, Vancouver, BC, Canada, 1996, pp. 257–261.

[36] J. Yang, D. Zhang, J. yu Yang, and B. Niu, "Globally maximizing, locally minimizing: Unsupervised discriminant projection with applications to face and palm biometrics," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 4, pp. 650 –664, April 2007.

[37] S. Bharadwaj, H. Bhatt, M. Vatsa, and R. Singh, "Periocular biometrics: When iris recognition fails," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, September 2010, pp. 1 –6.

[38] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, pp. 145–175, 2001.

[39] J. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 15, no. 11, pp. 1148 –1161, November 1993.

[40] D. Hosmer and S. Lemeshow, *Applied logistic regression (Wiley Series in probability and statistics)*. Wiley-Interscience Publication, September 2000.

[41] K. Hollingsworth, K. Bowyer, and P. Flynn, "Identifying useful features for recognition in near-infrared periocular images," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, September 2010, pp. 1 –8.

[42] K. Hollingsworth, S. Darnell, P. Miller, D. Woodard, K. Bowyer, and P. Flynn, "Human and machine performance on periocular biometrics under near-infrared light and visible light," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 2, pp. 588 –601, April 2012.

[43] P. Phillips, W. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe, "Frvt 2006 and ice 2006 large-scale experimental results," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 5, pp. 831 –846, May 2010.

Chapter 6

On Periocular Biometrics: A Comprehensive Outline

6.1 Overview

This chapter consists of the following article:

On Periocular Biometrics: A Comprehensive Outline
Gil Santos and Hugo Proença
Artificial Intelligence Review, submitted for consideration, 2013.

According to SCImago Journal & Country Rank, this journal's index[1] for the 2013 year are as follows:

Category	*Quartile*	*SJR*
Artificial Intelligence	Q2	1,242
Language and Linguistics	Q1	
Linguistics and Language	Q1	

[1]The SCImago Journal & Country Rank (SJR) indicator is a measure of journal's impact, influence or prestige. It expresses the average number of weighted citations received in the selected year by the documents published in the journal in the three previous years. http://www.scimagojr.com

On Periocular Biometrics

A Comprehensive Outline

Gil Santos · Hugo Proença

Abstract The usage of ocular data for recognition purposes has recently emerged, being actually one of the most promising traits for unconstrained scenarios, or when neither the iris nor the full face can be properly imaged. Supported by the literature is its high discriminability, relative stability and acquisition simplicity, making the periocular region – the region in close vicinity of the human eye – a good trade-off between the whole face and the iris alone. Furthermore, being easily acquired covertly without requiring constrained close capturing, it is an effective alternative when user cooperation is not expectable. This article offers the following contributions: 1) introductory three-fold framework on periocular biometrics, with the relation to the classical biometric traits and systems, the anatomy of the periocular region, and identified difficulties; 2) comparative overview of the most relevant research on the scope of periocular recognition, with widen detail on the underlying techniques; 3) state-of-the-art results against a common dataset; 4) comprehensive analysis of those results, using well-known evaluation metrics; 5) analysis of the biometric menagerie underlying each method, with insights about the main degradation issues; and, finally, 6) directions for further improvements on this technology.

We acknowledge the financial support provided by *FCT - Fundação para a Ciência e Tecnologia* through the research grant SFRH/BD/80182/2011, and the funding from 'FEDER - QREN - Type 4.1 - *Formação Avançada*', subsidized by the European Social Fund and by Portuguese funds through 'MCTES'.

IT - Instituto de Telecomunicações
Universidade da Beira Interior
Covilhã, Portugal
E-mail: gmelfe@ubi.pt; hugomcp@di.ubi.pt

Keywords Biometrics · Ocular Biometrics · Periocular Recognition · Recognition Robustness · Unconstrained Biometrics · Vision Based Biometrics

1 Introduction

The concept of biometrics has been evolving along with modern societies' increasing concern on individual and global security. From personal computers to border access control, everyone wish for securing their identity, their assets and, primarily, their homeland. In order to achieve such safety, the ability to accurately identify subjects based on their biometric features, either biological or compartmental, is essential.

Biometric systems rely on both the accurate *extraction* of individuals' distinctive features and their adequate *encoding*, so that the essential information can be preserved. Those requisites were traditionally assured by high constraining acquisition setups, with subject cooperation being a key element. However, when acquisition constraints are lowered or subject cooperation is not expectable, recognition become more challenging and alternatives are sought [38], either by improving the existing algorithms, resorting to multi-modal biometric systems, or exploring new traits that could better fit this new reality.

1.1 A leap from the traditional traits

Since biometrics emerged as a science, researchers have established biometric recognition with a wide variety of traits. From the most used traits, two of them are particularly related to the spring of periocular biometrics: the face and the iris.

(a) heavy weather clothing (b) balaclava (c) niqab (d) medical mask (e) helmet

Fig. 1 Illustration of situations where, due to subjects wearing severe headgear, the periocular region is the most advisable trait .

The face is presumably the most common and widely used biometric trait. Everyday and even without noticing it we all use facial information to recognize each other. Not only that, it became one of the most successful applications of image analysis and understanding. Several face recognition systems are commercially deployed, with a lot of techniques accessible for both still images and video [52]. Those methods are based either on global analysis of the whole region (as a set of pixel intensities), or in the relation between facial key-elements, their location and shape. However, their effectiveness is conditioned by several factors: its 3D structure lead to substantial differences in appearance accordingly to subject's pose; large portions are often occluded on non-orthogonal data acquisition; facial expressions affect their appearance; and it is particularly easy to disguise.

In opposition, the iris is known to deliver extreme recognition accuracy in exchange for high acquisition constraints. The high recognition performance attained with the iris is due to its predominantly randotypic morphogenesis unique for each individual. As so, a considerable amount of efforts are being held on iris biometrics [9], that quickly ascended as one of the most popular biometric traits. While most of the commercially deployed iris recognition systems work with constrained Near-Infrared (NIR) data, favoring the perception of its patters whilst reducing the quantity of inherent noise factors, literature on extending this trait usability to "relaxed" visible wavelength (VW) setups has broadened [42,45,39]. However, iris performance is severely impacted in non-ideal setups, as its relatively reduced size and moving profile make it difficult to image at-a-distance and without user cooperation. Accordingly, growing attention has been paid to other traits with potential application on unconstrained biometric recognition.

The periocular region, being a central figure of the face and representing a good trade-off between the whole face and the iris alone, has been receiving increased attention from the scientific community. Literature shows it to be a strong alternative on less constrained biometrics, when image acquisition is otherwise unreliable, or complementary to avoid iris pattern spoofing. It is relatively stable, rarely occluded, and easy to acquire covertly without requiring constrained close capturing, being an effective alternative when user cooperation is not expectable – Figure 1. Its proximity with the iris is definitely a surplus, as they can be imaged simultaneously with a single camera and fused (at feature or score levels) compensating for environmental adversities and uncooperative subjects.

1.2 Anatomy of the ocular region

Facial appearance is determined by both the superficial features of the skin, and the concavities and convexities conferred by the underlying bones and muscles. The periocular region in particular comprises many anatomic features and landmarks fit for recognition purposes – Figure 2.

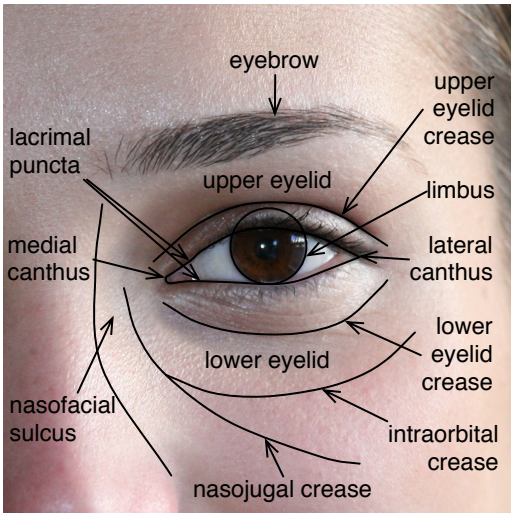


Fig. 2 Anatomic features in the vicinity of the eye.

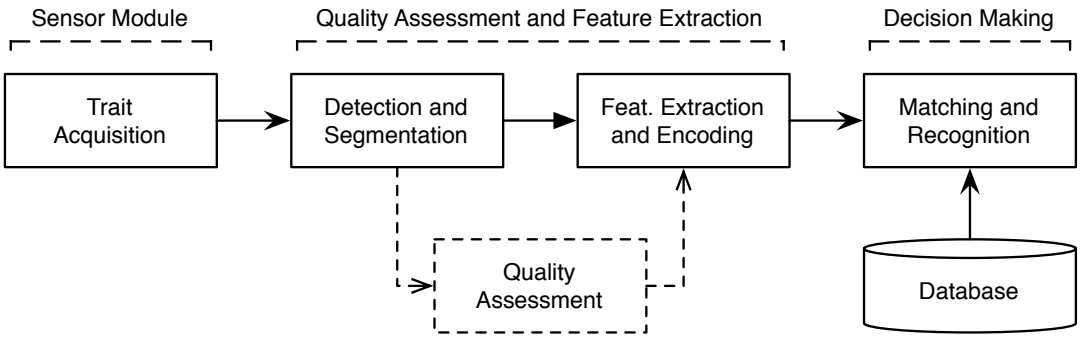


Fig. 3 General steps and elements of biometric recognition systems.

Centered on the eye, which is located on the orbital aperture, the periocular region has its creases and sulcus decided essentially by four bones: 1) the frontal bone, ending with the supra-orbital process where the eyebrow is located and which affects its appearance; 2) the nasal bone, defining the upper part of the nose; 3) the lacrimal bone, that forms the cavity for the tear gland; and 4) the zygomatic bone, also known as cheek bone.

As for the superficial features, and besides the skin texture and landmarks, we have eyebrows, eyelids and eyelashes. Eyebrows constitute the foundation for eyelids, and are straighter on men and more arched on women. Eyebrow thickness changes accordingly to ethnicity and, with the aging process, their orientation and height also change. Concerning the eyelids, their contours depend on gender, ethnic group and age, and dimension intervals are defined in previous studies [44]. We must remark that not only this region is not so affected by the aging process as other facial regions (e.g., mouth and cheek skin that become loosened over time), but even when the anatomic features of the face has been reshaped (e.g., plastic surgery), periocular data usage for recognition purposes remains advantageous [19,8].

Even considering the richness of ocular elements, the features actually being used on periocular biometrics algorithms are quite simple and can be divided into two levels, as suggested by Woodard et al. [46]. The first level comprises eyelids, eye folds, and eye corners, and the second level comprises skin texture, wrinkles, color and pores. This simplicity might be due to the relative novelty of the field: having passed only a couple of years since the first relevant study on periocular recognition, the earliest recognition algorithms firstly employed classical techniques in the computer vision domain-of-knowledge, before attempting more sophisticated and specific methods.

1.3 Structure of a Biometric System

The importance of a biometric authentication system as a whole must not be disregarded, as it will be the responsible for carrying the whole process from data acquisition to feature extraction and matching against the database. Thus, designing a system that adapts to its application scenario is of the most importance. A recognition system comprises four core modules [18] – Figure 3:

1. **Sensor Module** A wide variety of sensors is available, depending on which biometric trait we are going to work with. Since most of biometric traits consist on visual information, cameras are more likely to be used on data acquisition. On real-time systems (RTS) the balance between the richness on detail of the acquired data and the acquisition rate is as essential as choosing a proper camera. This module is strictly related with the first step of recognition systems (trait acquisition), and is where the trade-off between the quality of gathered data and user cooperation is set.
2. **Quality assessment and Feature extraction** Even with an optimal sensor setup, not always the acquired data is suited for feature extraction. It is therefore important to access its quality, discarding images where minimum requirements are not met, thus saving further processing time. The trait needs to be properly located and segmented (specially useful to gather preferably "good" data), and then encoded as feature templates.
3. **Matching and decision-making** In this module features are matched against the templates on the database, thus deciding to be facing either a genuine or an impostor comparison.
4. **System database** The system database module consists on the repository of user biometrics acquired during the enrollment stage and used for further identification or verification of users' identity.

Since the use of the periocular region as a biometric trait is relatively recent, there are a set of identified pitfalls regarding its use for recognition purposes. These were grouped into five topics, based on criteria suggested by Park et al. [32]. Right from the **trait acquisition** stage, it is yet to be determined the optimal imaging spectrum for periocular biometrics. While former research on ocular biometrics usually prefers near-infrared data, expectation aim towards the visible wavelength where unconstrained recognition is favored. It is moreover expected that the fusion from data acquired at different wavelengths, yielding multispectral information, would result in relevant advantages. The second concern is related with the actual **boundaries** of the periocular region. Although the literature shows the inclusion of some traces (e.g., the eyebrows, iris or sclera) to improve overall performance, researchers sometimes disagree on whether those elements should rather be masked or cropped to avoid biased results. Even for the region itself, only recently an *optimal* periocular region of interest (ROI) by balancing minimal template size with maximal recognition accuracy was proposed by Bakshi and Majhi [4], through an insightful study on the impact of its proportions to both the recognition performance and computational cost. At the **feature extraction and encoding** stage, new questions arise, as it is not yet settled which features are the most representative when aiming at discriminating this region. Furthermore, the heterogeneity of the components in the periocular region may suggest that more elaborate feature schemes are required to describe such different types of information. Subsequently, a fit **matching** scheme should be determined, taking into account the techniques most suitable to handle data variations inherent to the less controlled acquisition process and how to optimally handle the variations in the traditional data variation factors. On the fusion of periocular biometrics with other traits, we must consider that even if the use of multiple traits might be important to compensate for acquisition adversities, and iris being a fit candidate for score level fusion during periocular recognition, the way of maximize the outcome of this (or other) association is yet to be clearly established.

1.4 Datasets

Another known adversity inherent to the novelty of the use of the periocular region as biometric trait is that only a few public databases are available specifically designed for periocular method development. Hence, face and iris datasets are usually used for that purpose, being the most relevant illustrated on Figure 4 and detailed through Table 1. For each dataset here introduced we detail the amount of images and subjects, image dimensions and main variability factors, essential in evaluating the robustness of recognition algorithms. The heterogeneity present on the datasets make them a good mimic of non-cooperative operation conditions.

FERET[1]
: The Facial Recognition Technology (FERET) database [35] was designed as a standard for developing face recognition methods. It was acquired at George Manson University over eleven sessions and a three year period, and initially released as low resolution (256×384 pixel) grayscale data. Only later a *high*-resolution color version was disclosed. It contains a total of 14051 images, gathered from 1199 different subjects within a semi-controlled acquisition protocol with strict expression, pose and illumination changes.

FRGC[2]
: Collected at the university of Notre Dame, the Face Recognition Grand Challenge (FRGC) database [34] consists of high resolution ($\sim 1200 \times 1400$ pixel) color images, captured on both controlled and uncontrolled environments. On the controlled acquisition scenario uniform studio-grade illumination was used and subjects requested to stand still, look straight to the camera and essay sequentially both neutral and smiling expressions. As for the uncontrolled acquisition, images were shot at different scenarios disregarding both background and illumination.

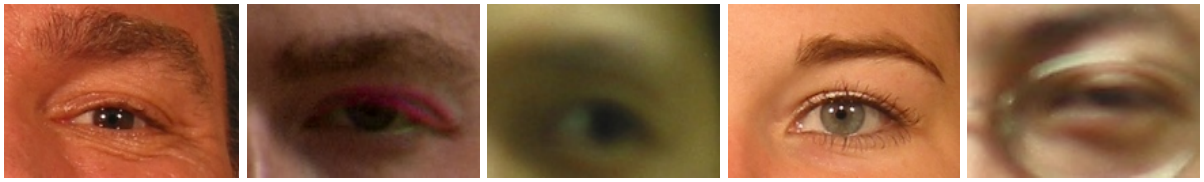
MBGC[3]
: The Multi Biometric Grand Challenge dataset consists on 149 videos, acquired under the NIR wavelength and introducing several noise factors: non-uniform illumination, eye-blink, motion blur, occlusions and reflexions. Videos were shot at 15 frames

[1] `http://www.nist.gov/itl/iad/ig/colorferet.cfm/`
[2] `http://face.nist.gov/frgc`
[3] `http://face.nist.gov/mbgc/`

Table 1 Specification of the most commonly used databases. Except for the MBGC, all datasets contain VW data. Variation abbreviations refer to Distance (D), Expression (E), Illumination (I), Occlusion (O) and Pose (P).

Name	Images	Subjects	Dimensions	Variations
FERET	14051	1199	512×768	E, I, P.
FRGC	36818	741	$\sim 1200 \times 1400$	E, I.
MBGC	149 AVI	114	2048×2048	D, E, I, O, P.
UBIRIS.v2	11102	261	800×600	D, O, I.
UBIPr	10950	261	Multiple	D, I, O, P.
FG-NET	1002	82	$\sim 400 \times 500$	D, E, I, P.

(a) FERET



(b) FRGC

(c) MBGC

(d) UBIRIS.v2

(e) FG-NET

Fig. 4 Sample images from the most commonly used datasets. Except from (d), data was cropped for illustration purposes.

per second (fps), on 2048 × 2048 pixel frames, representing an iris spatial extension of 120 pixel [47].

UBIRIS.v2[4]

The UBIRIS.v2 [37] is an unconstrained iris database captured on the VW from moving subjects, at different distances and challenging illumination conditions, thus simulating unconstrained acquisition and the inherent noise factors. Data from both eyes is available separately, along with the surrounding periocular data, thus allowing to test both periocular methods and their fusion with iris recognition techniques.

UBIPr[5]

More recently, the UBI Periocular Recognition dataset by Padole and Proença [31] represents an effort to advance of periocular biometric research, providing data to evaluate robust methods at "higher levels of heterogeneity". Noise factors were also introduced on the acquisition setup: varying acquisition distance, irregular illumination, pose and occlusion. Image dimensions vary accordingly to acquisition

[4] http://iris.di.ubi.pt/ubiris2.html

[5] http://socia-lab.di.ubi.pt/~ubipr

distance, ranging from 501 × 401 pixels at 8m to 1001 × 801 pixels at 4m. Manual database annotation is also provided, more specifically the ROI and essential landmarks.

FG-NET[6]

The FG-NET is a facial aging database that, although not so commonly used, is relevant due to the facial aging features: it contains around one thousand images from 82 subjects, up to 69 years old. Captured with different acquisition setups and many years apart, it is clear how subjects were shot under very irregular illumination, pose and expression conditions. Images are 400 × 500 pixels in size, captured on the VW, and for each one a 68 landmark points annotation is provided.

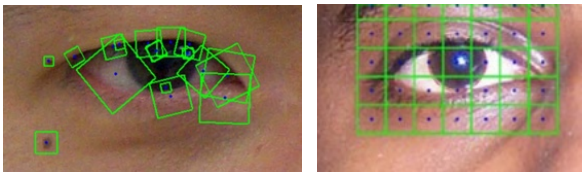
Considering that the collection of data for biometric experiments is particularly hard due to both security and privacy concerns and the substantial amount of required participants, Cardoso et al. [10] recently developed an algorithm for synthesizing degraded ocular images [7]. They described a stochastic method able to generate a practically infinite number of iris images, simulating acquisition under uncontrolled conditions. The images generated using their technique display eight noise factors: optical defocus, motion blur, iris occlusions, gaze, pose, distance, levels of iris pigmentation and lighting conditions.

The work presented in this paper extends the previous published work on periocular biometrics [40], by introducing a more detailed description and a side-by-side analysis of periocular methods on a fixed dataset. The remainder of this article is organized in the following manner: Section 2 comparatively details the relevant methods developed on periocular recognition; Section 3 presents and comparatively discusses the results of the implemented methods; and finally Section 4 states the final considerations.

2 The Most Relevant Recognition Algorithms

This section we summarize the most relevant techniques published in the scope of biometric recognition using information from the periocular area.

The pioneer approach on periocular biometrics dates back to Park et al. [33], who proposed a twofold feature extraction based on local and global features, as information relates to local regions, or the whole image – Figure 5. Global feature extraction starts by image alignment using as reference the iris coordinates and its dimensions, followed by the computation of two well-known distribution-based descriptors, Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP), over several contiguous square ROI forming a 7 × 5 grid (Figure 5(b)). Both descriptors are sequentially computed and quantized for each ROI into 8-bin histograms, storing both shape and texture information in a global single-dimension array. Although authors identify the eye-corners as better reference points for image alignment [32], they claim that they cannot be reliably determined.



(a) Local Features (b) Global Features

Fig. 5 Illustration of the a) extracted local features and b) ROI for global feature extraction (adapted from [33]).

LBP [27,28] works in a quite simple yet efficient fashion, measuring pixel intensity changes in a determined neighborhood P of radius R. Taking the difference T of the P intensities I to the central pixel, $LBP_{P,R}$ can be easily computed through (1), where s denotes the signal function. The HOG descriptor [12] is also well-known for its wide applications in computer vision. It simply computes gradient orientation by filtering the image with two kernels: $[-1, 0, 1]$ and $[-1, 0, 1]^T$.

$$LBP_{P,R} = \sum_{p=0}^{P-1} T_p.2^p \tag{1a}$$

$$T = s(I_0 - I_c), ..., s(I_{P-1} - I_c) \tag{1b}$$

$$s(a-b) = \begin{cases} 1, & \text{if } a \geq b \\ 0, & otherwise \end{cases} \tag{1c}$$

The local features are represented by a set of keypoints, and their surrounding information extracted using Scale-Invariant Feature Transform (SIFT) [23]. Keypoints are detected in scale space using a Difference of Gaussians (DOG) function, and features extracted from their bounding boxes (scale proportional) based on the gradient magnitude and orientation – Figure 6. The usage of SIFT offers invariance to translation, scaling and rotation.

[6] http://sting.cycollege.ac.cy/~alanitis/fgnetaging/index.htm

[7] http://iris.di.ubi.pt/NOISYRIS

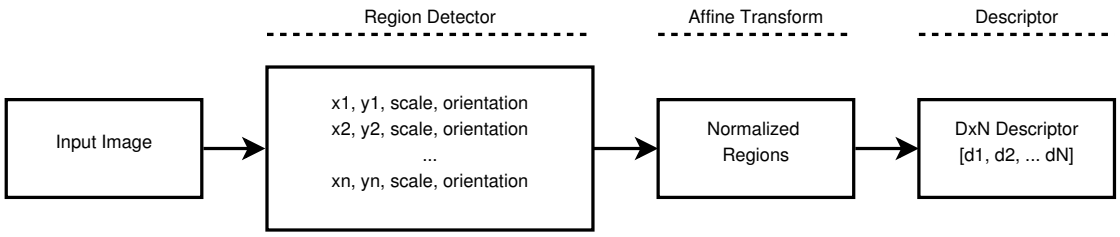


Fig. 6 Block diagram of the SIFT steps (adapted from [20]).

At the matching stage both arrays (H_a and H_b) containing the global features are compared using the Euclidean distance, simply by applying the Pythagorean formula (2) over their elements. The key-points representing the local features are matched using their geometrical alignment to their counterparts on the testing image.

$$d(H_a, H_b) = \sqrt{\sum_{i=1}^{n} (H_{ai} - H_{bi})^2} \tag{2}$$

The authors conducted their tests over a "small" (899 images, 30 subjects, 2 sessions) database of frontal periocular images acquired in the visible wavelength of the electromagnetic spectrum. The reported performances range from 62.5% when using the HOG features alone, to 80.8% when fusing them with the results from SIFT. Curiously, combining all the three features didn't led to further improvements on those results, with the joint performance being set at 80%. For that same dataset, authors report that Rank-1 recognition accuracy when using the whole face was 100%.

On their later work, Park et al. [32] went further on stressing periocular recognition by analyzing the impact of several factors: eyebrow inclusion or disguising, automatic segmentation, side information, iris and sclera masking and expression variation. Their results showed that although automatic OpenCV segmentation exhibits better performance on "eyebrow-less" data, its inclusion in the periocular region improves the SIFT results in almost 19%. Information about the side of the face is almost irrelevant, with performance variations of about 1%. From the stressed variations facial expressions was the one with the most significant impact over periocular recognition performance, except for SIFT, which is more robust to distortion. On the other side, the SIFT descriptor revealed itself handicapped when facing iris or sclera occlusions. Top accuracy for the classifiers singlehanded was 79.49%, achieved using SIFT on manually segmented and unmasked data, including the eyebrow, when compared to images taken from the same side of the face and the same facial expression. Compared to their previous work (i.e., [33]) score level fusion did not result in significant improvements.

As recognition over non-ideal situations was also a concern, authors compared their results with FaceVACS[8] face recognition system marks – 99.77% recognition accuracy on "clear" facial images. Significant performance drops were registered, for instance, with occlusions (about 60% when occluding the lower part of the face – Figure 7(a)), and even for small occlusions on the periocular area. Without resorting to score-level fusion, the encoding methods led to accuracy values no greater than 25.97%, 20.51% and 10.12% respectively for 10%, 20% and 30% of periocular occlusion. Eyebrow modification was also subject for testing (Figure 7(b)), using the TAAZ[9] tool to simulate makeover. The registered decay on performance was 7.5% for LBP and 10% for other descriptors. On subjects facing at 15◦ to

[8] FaceVACS SDK available at `http://www.cognitec-systems.de`

[9] Free virtual makeover took, available at `http://www.taaz.com`

(a) Facial occlusion

(b) Eyebrow makeover

Fig. 7 Illustration of the non-ideal conditions simulated by Park et al: a) facial occlusion and b) eyebrow makeover (adapted from [32]).

Table 2 Average accuracy degradation under different factors tested by Park et al. [32].

Factor	Accuracy avg. diff.
Eyebrow inclusion	+10.66%
Automatic segmentation	−0.06%
Expression change	−3.36%
Same side matching	+0.33%
Iris masking	−3.65%
Eye masking	−15.95%
10% periocular occlusion	−48.67%
Different sessions	−25.00%

30◦ (i.e., head rotation), 35% to 45% performance deterioration was registered, being SIFT the most affected descriptor. Finally, authors pointed out an additional issue associated with the periocular region – its lack of stability over time. Images captured three months apart from each other appear to perform 15% worst, and 30% when captured with half-year gap. Results from those experiments [32] are summarized on Table 2. Authors suggest several further improvements: better alignment and matching methods; multi-spectral analysis; and the possibility of fusion with iris (or face) recognition methods.

Miller et al. [26] analyzed the periocular skin texture by applying the Uniform Local Binary Patterns (ULBP) method, and providing further insights on each region's impact on the recognition process. This LBP-based approach is meant to achieve "improved rotation invariance with uniform patterns and finer quantization of the angular space" [29]: from the 2^P possible binary patterns returned from regular LBP computation (1) over a P neighborhood, a uniformity measure U can be calculated representing the number of bitwise changes in that pattern (3a). Using a uniformity value of 2, users employed the ULBP operator using equation (3b).

$$U(LBP_{P,R}) = |s(I_{P-1} - I_c) - s(I_0 - I_c)| + \sum_{p=1}^{P+1} |s(I_p - I_c) - s(I_{p-1} - I_c)| \tag{3a}$$

$$LBP^{u2}_{P,R} = \begin{cases} \sum_{p=0}^{P-1} s(I_p - I_c), & \text{if } U(LBP_{P,R}) \leq 2 \\ P+1, & \text{otherwise} \end{cases} \tag{3b}$$

Similarly to the previous approach, the periocular region was cropped proportionally to the intra-eye distance, scaled to 100×160 pixels and divided into a 7×4 ROI grid. To avoid that the iris and the sclera information could possibly influencing the results, an elliptical neutral mask was overlapped to the periocular image. After histogram normalization the ULBP is computed for each region, using an 8-pixel neighborhood and producing 59 possible results used to populate the histogram for the periocular signature array. Finally, the Manhattan distance (4) is used for comparison at the matching stage.

$$d(H_a, H_b) = \sum_{i=1}^{n} |H_{ai} - H_{bi}| \tag{4}$$

Experiments conducted on images from the FRGC and FERET datasets, for both eyes separately and combined, reported 84% and 71% and 90% and 74% recognition rates respectively.

Miller et al. [25] later work addressed the impact of image quality over three axes: blur, resolution and illumination. Image was preprocessed using a similar periocular crop and resizing (251×251 pixels), grayscale conversion, histogram equalization and eye masking, but instead of the ULBP a base LBP was used. When blurring the data by convolving it with a Gaussian filter, the periocular recognition outperformed that of the entire face in particular for high levels of blur. A similar conclusion was reached upon down-sampling the data to 40% of its original size. As for uncontrolled illumination conditions (already introduced on the dataset, i.e., not simulated), performance degrades to low levels since local approaches (e.g., LBP) are not suited for irregular lighting conditions. The authors also compared the discriminant capabilities of each color channel, concluding that the green channel is the one leading to higher differentiation (23% higher accuracy than the red channel), and encodes the texture information in a very similar way as the blue channel. In general, authors concluded that performance achieved on the periocular region was higher than when using the whole face, and suggest the usage of different classification methods as Support Vector Machines (SVM) [41].

Adams et al. [2] extended Miller's work [26], by using a Genetic & Evolutionary Computing (GEC) method to optimize the original feature set, namely the Steady-

```
Procedure GEC{
        t=0\;
        Initialize Pop(t)\; /* Initial Population */
        Evaluate Pop(t);
        while (While Not Done){
                dad = Select_Parent(Pop(t)); /* Dad */
                mom = Select_Parent(Pop(t)); /* Mom */
                offspring = Create_Offspring(mom,dad);
                Evaluate(offspring);
                Pop(t+1) = Replace(worst,offspring);
                t = t + 1;
        }
}
```

Fig. 8 Pseudo-code example of a GEC [1].

State Genetic Algorithm (SSGA) algorithm implemented by eXploration Toolset for Optimization of Launch and Space Systems (X-TOOLSS)[10]. The GEC algorithm comprehends the process of creating problem solvers, having as basis the simulated evolution. Starting from a randomly selected set of candidate solutions (CS), the goodness of fit (GOF) of an evaluation function is determined to select two parents and create offspring that replaces the worst members of the current population – Figure 8. Authors reported 86% accuracy for either eye on the FRGC database, and 80% over the FERET data, and top results of 85% and 92% when using both eyes. Using only $49 \sim 52\%$ of the original features improved on, at least, 10%. Nonetheless, the chosen algorithm was not proven to be the optimal for that specific periocular features.

Inspired by Park et al. [33] work, Juefei-Xu et al [20] expanded their experiments to less ideal imaging environments, conducting performance analysis for different feature schemes over the FRGC dataset. At the encoding stage, and in addition to the LBP and SIFT techniques, both local and global feature extraction schemes were tested: Walsh masks, Law's masks, DCT, DWT Force Fields and SURF; Discrete Wavelet Transform (DWT), Gabor Filters and Laplacian of Gaussian (LoG). For the matching stage, the impact of different distance metrics was also tested: Normalized Cosine (5), Euclidian (2) and Manhattan (4).

$$d(H_a, H_b) = \frac{\sum_{i=1}^{n} H_{ai} \times H_{bi}}{\sqrt{\sum_{i=1}^{n}(H_{ai})^2} \times \sqrt{\sum_{i=1}^{n}(H_{bi})^2}} \tag{5}$$

Walsh masks [6] are a set of filters based on the Walsh function (6), whose convolution with the image captures its binary characteristics in terms of contrast. Sampling each function at the integer points, 5 vectors (7) of size 5 are produced and, when combined in between themselves, 25 base images are obtained.

$$W_{2j+q}(t) = (-1)^{\lfloor \frac{2}{j} \rfloor + q}[W_j(2t) + (-1)^{j+q} W_j(2t-1)] \tag{6}$$

where $\lfloor \frac{2}{j} \rfloor$ is the integer part of $j/2$, and q is either 0 or 1.

$$\begin{cases} W_0^T = (1,1,1,1,1) \\ W_1^T = (-1,-1,-1,1,1) \\ W_2^T = (-1,-1,1,1,-1) \\ W_3^T = (1,1,-1,1,-1) \\ W_4^T = (1,-1,1,1,-1) \end{cases} \tag{7}$$

Laws filter masks [22] were pioneer techniques of texture identification through filtering. They consist on using a bank of 25 filters, five in each dimension, accordingly to the masks from (8) and assuming the shape of level (L), edge (E), spot (S), wave (W) and ripple (R).

$$L_5 = [1,4,6,4,1] \tag{8a}$$
$$E_5 = [-1,-2,0,2,1] \tag{8b}$$
$$S_5 = [-1,0,2,0,-1] \tag{8c}$$
$$W_5 = [-1,2,0,-2,1] \tag{8d}$$
$$R_5 = [1,-4,6,-4,1] \tag{8e}$$

Discrete Cosine Transform (DCT) [3] is a commonly used image encoding technique with a good balance between performance and computational cost. In this particular case authors employed the 2-D DCT, transforming each intensity $I_{x,y}$ to the frequency domain $D_{u,v}$ using (9b) over windows of size N.

$$C_u = \begin{cases} 1/\sqrt{2} & \text{if } u = 0 \\ 1 & \text{otherwise} \end{cases} \tag{9a}$$

$$\mathrm{D}_{u,v} = \frac{1}{4} C_u C_v \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} I_{x,y} \cos\left(u\pi \frac{2x-1}{2N}\right) \cos\left(v\pi \frac{2y-1}{2N}\right) \tag{9b}$$

C_u and C_v are computed likewise.

The Force Field Transform [17], based on gravitational force field, assumes that any pair of pixels maintains attraction between themselves. Every pixel generates a field that affects another pixel, at position vector r_j and through the force F_i (10a). The resulting force at a given pixel will then be the sum of all forces (10b).

$$F_i(r_j) = I(r_i) \frac{r_i - r_j}{|r_i - r_j|^3} \tag{10a}$$

$$F(r_j) = \sum_{i=0, \neq j}^{N-1} F_i(r_j) = \sum_{i=0, \neq j}^{N-1} \left(I(r_i) \frac{r_i - r_j}{|r_i - r_j|^3} \right) \tag{10b}$$

As for DWT [24], it was computed using the wavelet at equation 11.

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \leq I_{x,y} < \frac{1}{2} \\ -1 & \text{if } \frac{1}{2} \leq I_{x,y} < 1 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

Gabor filters were also used to encode texture features [11] (12) with the following parameters: wavelength $\lambda = 8$, orientations $\theta = \{0°, 45°, 90°, 135°\}$, phase

[10] `http://nxt.ncat.edu/`

offset $\psi = \{\pi/2, -\pi/2\}$, sigma $\sigma = 4.5$ and ellipticity $\gamma = 1$.

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos(2\pi\frac{x'}{\lambda}+\psi) \tag{12a}$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{12b}$$

The purpose of the Laplacian of Gaussian is to detect edges, and can be described as a function of Gaussian standard deviation σ and the radial distance ρ (13).

$$-\frac{1}{\pi\sigma^4}\left(1 - \frac{\rho^2}{2\sigma^2}\right)e^{-\rho^2/2\sigma^2} \tag{13}$$

The main objective of Speed Up Robust Features[11] [5] is to provide a faster alternative for SIFT on detecting and describing local features.

The LBP itself was fused with other methods, yielding the results given in Table 3.

Authors show local descriptors to register better results, with the post-application of LBP translated into a significant performance boost. Although top accuracy being registered for DWT + LBP (53.2%), results were very similar when DCT and Walsh or Laws' masks were used. SIFT and Speed Up Robust Features (SURF) verification rate was surprisingly low (< 1%), mostly due to low image resolution.

On a later work, Juefei-Xu et al. [21] addressed the aging effect on periocular recognition previously identified as an issue (e.g., Park et al. [32]) of no trivial resolution due to its influence from external factors, such as ethnicity, gender, etc. Their approach starts by performing two type of corrections: pose, through Active Appearance Models (AAM), and illumination, through anisotropic diffusion model. The periocular region was normalized from the provided landmark points and features encoded using Walsh-Hadamard transform encoded LBP (WLBP). On a final stage the unsupervised discriminant projection (UDP) technique [51] was used to boost results to very high performance levels, using both global and local information to minimize the local scatter matrix S_L (14a) while maximizing the non-local S_N (14b) using the adjacency matrix A produced by (14c) based on a set of x_N classified instances.

$$S_L = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} A_{i,j}(x_i - x_j)(x_i - x_j)^T \tag{14a}$$

$$S_N = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(1 - A_{i,j})(x_i - x_j)(x_i - x_j)^T \tag{14b}$$

$$A_{ij} = \begin{cases} 1 & \text{if } x_i, x_j \text{ are mutually KNN} \\ 0 & \text{otherwise} \end{cases} \tag{14c}$$

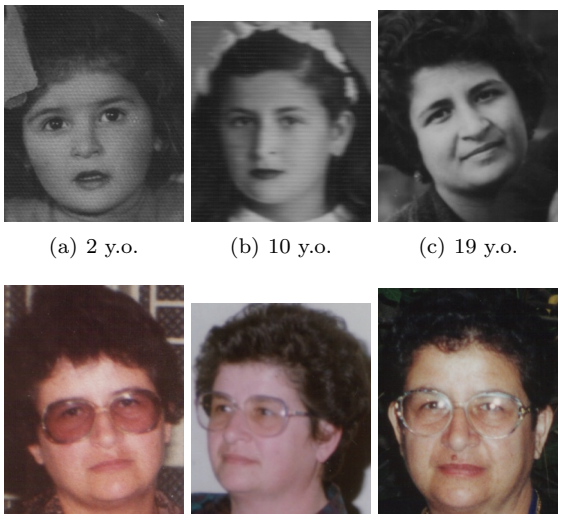
This method was tested on the FG-NET database, with images taken years apart at different acquisition setups (non-uniform illumination, pose and expression) – Figure 9. The reported results show 20% of performance improvement, with WLBP performing 15% better than raw pixel intensity matching. UDP also delivers better accuracy (up to 40%) than Principal Component Analysis (PCA) or Locally Preserving Projections (LPP). Authors report that all stages combined result in 100% identification accuracy.

Bharadwaj et al. [7] research on periocular biometrics was focused specifically on unconstrained VW captured data (UBIRIS.v2 dataset). The authors tackled the question combining ULBP with a global matcher – GIST – what consists in the combination of five scene descriptors [30]: *naturalness* (i.e., how vertical and horizontal edges are distributed); *openness* (i.e., the presence or lack of reference points); *roughness* (i.e., size of the largest prominent object); *expansion* (i.e., depth of the space gradient); and *ruggedness* (i.e., deviation from the horizontal, assessed by contour orientation). ULBP was computed over 64 patches of the original image. Prior to the GIST analysis, local contrast normalization was achieved with Fourier transform and the special envelope computed using a set of χ^2 distance (15) and min-max normalized results from both eyes are fused by a weighted sum. GIST gave better performance than ULBP, with 70.82% against 63.77 respec-

[11] Matlab interface to OpenCV SURF package available at `http://www.maths.lth.se/matematiklth/personal/petter/surfmex.php`

Table 3 Rank-1 accuracy obtained when fusing LBP with the other methods [20].

Fused methods	Accuracy
LBP + LBP	42.5%
Walsh Masks + LBP	52.9%
Laws' Masks + LBP	51.3%
DCT + LBP	53.1%
DWT + LBP	53.2%
Force Field Transform + LBP	41.7%
Gabor Filters + LBP	12.8%
LoG Filters + LBP	30.9%



(a) 2 y.o. (b) 10 y.o. (c) 19 y.o.

(d) 35 y.o. (e) 45 y.o. (f) 42 y.o.

Fig. 9 Example of aging subject from the FG-NET database, at ages 2 (a), 10 (b), 19 (c), 35 (d), 45 (e) and 52 (f) years old (y.o.).

tively, and fusing both results led to 73.65% Rank-1 accuracy.

$$d(H_a, H_b) = \frac{1}{2}\sum_{i=1}^{n}\frac{(H_{ai} - H_{bi})^2}{(H_{ai} + H_{bi})} \quad (15)$$

To determine which slice of the electromagnetic spectrum would better favor the periocular recognition, Woodard et al. [46] conducted a comparative analysis of second-level features on both visible (FRGC) and NIR (Multi Biometric Grand Challenge (MBGC)) data. As preprocessing, and to avoid biased results, an elliptical mask was used to remove iris and sclera information. Then LBP was computed over a ROI grid (on both datasets), and color information was extracted from both the red and green channels' histograms and fused at score-level (on VW data). At the matching stage, the LBP histograms were matched using Manhattan distance, with Bhattacharya distance being chosen for color histograms. Results suggest texture information to be more discriminant, with only a slight improvement registered after the fusion. As for the electromagnetic spectrum, visible wavelength data delivered better results ($88 \sim 90\%$ accuracy) than NIR ($81 \sim 87\%$).

On their later work, Woodard et al. [47] assessed how periocular texture information could improve iris data reliability, thus overcoming the difficulties associated with non-ideal imaging. Tests were conducted over the MBGC dataset, that despite containing NIR data, represents a challenging scenario for iris recognition. Periocular information was handled the same way as before (i.e., [46]), and iris processing was achieved followed Daugman's approach [13], except with manual segmentation. Information from both traits was combined at score-level using a simple weighted sum after min-max normalization. Their work shows how iris' low performance on such difficult data benefits from the fusion with periocular information, raising Rank-1 accuracy in over 80% to 95.6%. Both Woodard et al. studies [46,47] were unified and extended, providing a closer insight to their previous results [48].

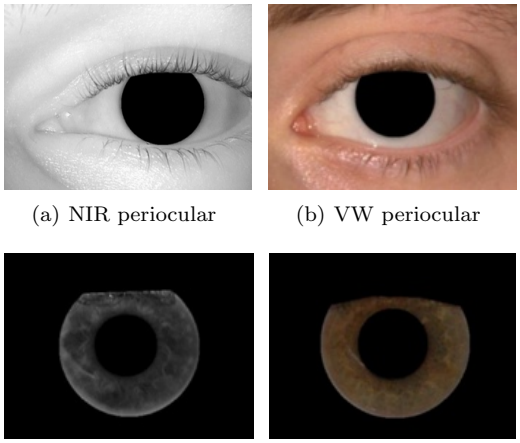
The effect of periocular information on improving face recognition performance was stressed by Jillela & Ross [19]. In their study, authors aimed at improving commercially deployed face recognition software identification performance against subjects who have been submitted to plastic surgery. Authors applied SIFT and LBP over the periocular region, as proposed by Park et al. [33], except for this last descriptor that was computed on all color channels independently and averaged at score-level. Tests were conducted over a plastic surgery database [43] consisting of images downloaded from plastic surgery information websites, with considerable changes in resolution, scale and expression. Results showed periocular information to boost face recognition software Rank-1 accuracy by 2%, attaining a top performance of 87.4%.

On stressing noise factors' impact on periocular recognition, Padole & Proença [31] conducted a series of tests on images with four inherent variations: subjects' pose, distance to the camera ($4m$ to $8m$), iris pigmentation and occlusion. Periocular information was analyzed as suggested by Park et al. [33], with some minor variations: the ROI definition that was based on eye-corner position instead of iris center, which led to most significant improvements since unconstrained biometrics favor gaze variations; and for the fusion stage authors tested both logistic regression and Multi Layer Perceptron (MLP).

The logistic regression model [16] is a weight fitting methodology that works as a single-output neural network with a logistic-activation function trained under log loss (16), relating output weights β_i with *the odds* of a positive match ($p/(1-p)$).

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{LBP} + \beta_2 x_{HOG} + \beta_3 x_{SIFT} \quad (16)$$

Interestingly, closer acquiring distances didn't led to better performance, as worst results came from comparisons between subjects imaged at 4 meters. From the tested distances, authors found the "optimal" one to be $7m$. Not so surprising was pose variation impact on recognition, with higher tilting angles resulting in



(a) NIR periocular (b) VW periocular

(c) NIR iris (d) VW iris

Fig. 10 Sample NIR periocular images used by Hollingsworth et al. (adapted from [14,15]).

lower accuracy values. Similar observations were found for the occlusion trials. Iris pigmentation was also reported to impact periocular recognition performance, with darker eyes leading to poorer results and medium pigmented irides the best ones. As for gender, female were found to be more easily identified through their periocular features.

The Human ability to use contextual information and "disregard" most of the noise factors, adapting itself to surrounding conditions is outstanding, marking it a hard task for machines to mimic. In fact, when designed recognition algorithms a logic path would be to figure out its way of working, seeking alternate strategies to tackle the same issues. On that subject, Hollingsworth et al. [14] conducted a study aimed at identifying which ocular elements humans find more useful for the periocular recognition task. Using an iris camera, authors acquired NIR data from 120 subjects with the periocular region closer to the eye visible and only few features missing (e.g., incomplete eyebrows) – Figure 10(a). In order to avoid biased responses, iris was completely masked with a circular patch. Eighty pairs of images were presented to 25 human observers, who were asked to tell apart which pairs belonged to the same or different subjects, indicating their degree of certainty and individually rating each feature's helpfulness in a three level scale. Results pointed eyelashes to be the most helpful periocular feature, closely followed by the medial canthus and the eye shape. Participants based their responses on eyelash clusters, density, direction, length and intensity. To the inquired observers, skin was actually the less useful. Average human accuracy on such setup was 92%.

In order to extend that analysis to the visible spectrum, to new factors and to a wider dataset, a second study was conducted by Hollingsworth et al. [15]. The human observers' performance was compared to both periocular (Park et al. [33]) and iris (IrisBEE biometric system from ICE [36]) recognition algorithms. Imaging 210 subjects on a controlled environment, 140 pairs of images were presented to 56 observers for each one of four setups: NIR and VW, periocular and iris data. Test subjects could then rank their certainty on a five level scale, specifying how helpful individual features were (*eye shape*, *tear duct*, *outer corner*, *eyelashes*, *skin*, *eyebrow*, *eyelid*, *color*, *blood vessels* and *other*). Due to the different pairing system and limited observation time, NIR accuracy dropped to 78.8%, and it was set on 88.4% for the visible wavelength. Machine performance was similar, with 1% difference on overall accuracy. As for the feature discrimination capacity, results for NIR data were similar to the previous ones [14], with some differences on the visible spectrum where blood vessels, skin and eye shape were reported to be more helpful than eyelashes. Skin details were in fact more perceptible on visible wavelength data, as NIR camera illumination caused frequent skin saturation. In general, the visible light was found to be preferable for periocular recognition tasks. Human perception of iris features is greater on NIR images, with 85.6% accuracy against 79.3% on the visible wavelength. However, and unlike periocular, machine performance was 13% better than humans', with 100% and 90.7% accuracy for those same bands.

3 Empirical Evaluation of Algorithms

A summarized overview over the described methods and reported results is provided at Table 4, along the type of features extracted, the classification scheme, and the dataset used by each author. As we can see, the methods from the literature are focused mainly on texture analysis and key-point extraction. Even simple algorithms (e.g., LBP based) lead to fair performance levels, and can be improved with further refinements, optimization or score-level fusion. Reported results also suggest periocular fitness for unconstrained setups, with VW prevailing over NIR. Nonetheless, the heterogeneity between test data renders methods' relative performance in-between themselves difficult to assess.

To achieve a clearer performance insight, algorithms should be tested over the same data, with results analyzed side-by-side. As most of the literature reports results against FRGC, we choose it as the most fit candidate for the evaluation stage. A total of 6225 im-

Table 4 Overview of the most relevant periocular research available on the literature.

Approach	Features	Extract	Classifier	Dataset	Reported Accuracy
Park et al. [33]	Shape, Texture, Key-Points	HOG. LBP, SIFT	Euclidean distance, SIFT matcher	899 VW img., 30 subject, 2 sessions	HOG: 62.5% LBP: 70.0% SIFT: 74.2% Best: 80.8%
Miller et al. [26]	Texture	ULBP	Manhattan distance	FRGC, FERET	FRGC: 89.8% FERET: 74.1%
Adams et al. [1]	Texture	LBP+GEFE	Manhattan distance	FRGC, FERET	FRGC: 92.2% FERET: 85.1%
Woodard et al. [46]	Color, Texture	RG color hist, LBP	Bhattacharya, Manhattan distance	FRGC, MBGC	Left VW peri: 90% Right VW peri: 88% Left NIR peri: 81% Right NIR peri: 87%
Woodard et al. [47]	Texture	Daugman's iriscode, LBP	Hamming distance, Manhattan distance	MBGC	Left Iris: 13.8% Left Peri: 92.5% Fusion: 96.5% Right Iris: 10.1% Right Peri: 88.7% Fusion: 92.4%
Juefei-Xu et al. [20]	Texture, Key-Points	Walsh Masks, Laws' Masks, DCT, DWT, Force Fields, Gabor Filters, LBP, SIFT, SURF	Cosine distance, Euclidean distance, Manhattan distance.	FRGC	DWT+LBP: 53.2% DCT+LBP: 53.1% Walsh+LBP: 52.9% Laws'+LBP: 51.3% ...
Juefei-Xu et al. [21]	Texture	WLBP+UDP	Cosine distance	FG-NET	100%
Bharadwaj et al. [7]	Naturalness, Openness, Roughness, Expansion, Ruggedness, Texture	GIST, ULBP	χ^2 distance	UBIRIS.v2	GIST: 70.82% ULBP: 63.77% Fusion: 73.65%
Hollingsworth et al. [14]	Human	Human	Human	NIR img, 120 subject	NIR peri: 92%
Hollingsworth et al. [15]	Human	Human	Human	NIR &VW, 210 subject	NIR Peri: 78.8% VW Peri: 88.4% NIR Iris: 85.6% VW Iris: 79.3%

ages were selected, with the right-side periocular region manually cropped to avoid further errors, resulting in over 250 thousand matching trials with a 1 : 2 intra-inter-class ratio. The iris segmentation required by Woodard et al. method [47] was also manually conducted. We reproduced the presented algorithms as close as possible, choosing *omitted* parameters in such way that overall performance was maximized. At score-level fusion, weights were optimized using logistic regression with 10-fold cross-validation. The results of that evaluation are the ones at Table 5. As some papers reported results from multiple setups, values may differ from the ones at Table 4, as we now choose to display those best fitting the testing conditions.

For the tested methods, registered performances are quite similar with Rank-1 accuracy around 97%. Having Park et al. [33] pioneering approach as comparison term, we can see that improvements introduced by subsequent algorithms rely on three different factors: different image preprocessing and ROI definition, more robust procedures, and bringing in new techniques. Starting at this first topic, we can easily observe how changing image pre-processing and ROI definition lead to performance discrepancies for the same descriptor over the exact same dataset and comparisons. ULBP, for instance, displays a rather good performance on Miller et al, a better one on Woodard et al. [46], and a slightly worse on Bharadwaj et al.. Another example is Park et al. LBP not being over-performed by ULBP from methods [25] and [7]. Apart from that, when applied under similar preprocessing ULBP preforms better that regular LBP (Woodard et al. [47] *vs* Woodard et al. [46]). From the introduced descriptors GIST is the most noteworthy for its surprisingly high performance, being the feature with higher Area Under ROC Curve (AUC) and lower Equal Error Rate (EER). As we can see on the last two methods, classification accuracy don't always agree with AUC and EER about the best classifier. On those situations, we found these last two metrics to be more reliable indicators.

On methods using multiple descriptors, relative individual performances are comparable to the ones authors reported, except for two particular cases. The first one is when color information is used (Woodard et al. [46]), as we didn't manage to reproduce such scores even though the same database was used. Even obtaining better accuracy for ULBP, fusing it with the color descriptors didn't led to significant improvements, as score-level fusion optimization technique (logistic regression) didn't give color information enough weight to make itself representative. The second one is for iris based recognition, where we obtained higher performance levels even with a VW dataset. We must have in mind that the MBGC dataset was originally chosen by authors to test their approach because of its challenging conditions for iris biometrics.

Analyzing Pearson's linear correlation coefficients between techniques – Table 6, we are able to understand their pairwise dependence. This measure (ρ) relates two methods (X, Y), by analyzing the relation between their covariance (cov) and standard deviations (σ), returning values in the $[-1:1]$ interval according to (17). Total positive correlation is represented by 1, negative correlation -1, and the absence of correlation is then 0.

$$\rho_{(X,Y)} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \tag{17}$$

As anticipated, high correlation values were registered between LBP based descriptors, particularly the ones used on both Woodard et al. methods and Miller's [26]. On Park et al. pioneer approach, LBP correlation with HOG was also significant. In fact, with exception of color and iris correlation values were generally high, being those the two less-dependent routines (the ones with lower average absolute correlation coefficients). The third one was SIFT. GIST descriptor, the one with better individual performance, was actually found to be significantly correlated with most of the tested techniques.

3.1 Biometric Menagerie

To better understand subjects' influence on the tested methods, we analyzed them individually in terms of genuine/impostor distance, classifying them accordingly to the four biometric menagerie classes suggested by Yager & Dunstone [49] – Figure 11(f): doves, chameleons, phantoms and worms.

Let us define two regions for the genuine distance distribution, G_L and G_H, containing the subjects below the first quartile or over the third quartile respectively. If we define two similar regions (I_L and I_H) for the impostor distance distribution, subjects s can be classified as one of the *animal* families accordingly to (18) [50]. From Figure 11, we can see how certain subjects systematically fork from the sheep group for most of the tested periocular recognition methods.

$$\begin{cases} \text{Dove}, & \text{if } s \subset G_L \cap I_H \\ \text{Chameleon}, & \text{if } s \subset G_L \cap I_L \\ \text{Phantom}, & \text{if } s \subset G_H \cap I_H \\ \text{Worm}, & \text{if } s \subset G_H \cap I_L \end{cases} \tag{18}$$

Doves are the most favorable subjects, as they produce low distances for intra-class comparisons and high distances for inter-class. This *animal* family is an extension of the herd, and the optimal group for any recognition system as it does not produce verification error. As we can see from Figure 11, there are only two samples who could be identified as *doves*: subjects A, imaged with a slight head rotation (Figure 12(a)), and subject B, with a significant occlusion caused by her hair (Figure 12(b)), those presumably being the causes of such good response. As so, we were not able to generalize which features (or conditions) best describe those subjects, being the optimal for each method operation.

Chameleons are subjects who produce low distances on both classes matches, thus being easily misidentified as they always appear similar to others. Their false-reject rate is low, but false-accept is extremely high. The graphics at Figure 11 show us a considerable concentration of *chameleons*, who we assume to be the individuals with more generic features, as subjects C and D (Figure 12(c) and 12(d)) who systematically display that behavior for all methods.

Phantoms are associated with high matching distances on both intra- and inter-class comparisons. Contrary to *chameleons*, they are associated with high false-reject rate, and low false-accept. As we can see from the graphical representation, subjects on the *phantom*-region are more disperse, and a larger set of subjects recurrently belonging to this class was identified – subjects E to H. Sample elements of this family are depicted at Figures 12(e) to 12(h). Subjects identified as *phantoms* were enrolled under non-uniform lightening, or with significant changes in their appearance – subjects being captured with and without glasses.

Contrary to *doves*, *Worms* are the most problematic subjects in a biometric system. They behave in the worst possible way, yielding high distances on genuine matches and low distances on impostor ones. As pointed by Yager and Dunstone [49], in real biometric applications this group does not exist, as it would represent a significant flaw in the matching algorithm.

Table 5 Tested periocular recognition methods performance indicators: Area Under ROC Curve (AUC), Equal Error Rate (EER), Computed (CA) and Reported Accuracy (RA) and Original testing dataset.

Approach	**Features**	**AUC**	**EER**	**CA**(%)	**RA**(%)	**Dataset**
Park et al. [33]	LBP	0.87	0.20	96.64	70.00	899 images,
	HOG	0.85	0.22	94.32	62.50	30 subjects,
	SIFT	0.86	0.19	96.38	74.20	2 sessions,
	Fusion	**0.89**	**0.17**	**97.16**	80.80	visible wav.
Miller et al. [26]	ULBP	**0.85**	**0.21**	**96.64**	89.90	FRGC
Woodard et al. [46]	ULBP	**0.87**	**0.18**	97.16	83.80	FRGC
	Color	0.63	0.40	35.66	74.20	
	Fusion	0.86	0.19	**97.42**	87.10	
Woodard et al. [47]	LBP	0.85	0.20	**96.90**	88.70	MBGC
	Iriscode	0.78	0.28	75.45	10.10	
	Fusion	**0.86**	**0.19**	96.64	92.40	
Bharadwaj et al. [7]	ULBP	0.78	0.27	96.38	54.30	UBIRIS.v2
	GIST	0.90	0.17	**97.42**	63.34	
	Fusion	**0.92**	**0.15**	95.61	73.65	

Table 6 Linear correlation coefficient matrix for the tested periocular recognition methods. Results refer to LBP (L), HOG (H), SIFT (S), Fusion (F), ULBP (U), Color (C), Iriscore (I) and GIST (G).

		[33]				[26]	[46]			[47]			[7]		
		L	H	S	F	U	U	C	F	L	I	F	U	G	F
[33]	L	1													
	H	0.82	1												
	S	-0.60	-0.61	1											
	F	0.81	0.85	-0.93	1										
[26]	U	0.84	0.72	-0.63	0.77	1									
[46]	U	0.80	0.71	-0.65	0.77	0.96	1								
	C	0.20	0.20	-0.18	0.21	0.20	0.20	1							
	F	0.79	0.71	-0.65	0.77	0.95	0.99	0.36	1						
[47]	L	0.83	0.73	-0.63	0.78	0.97	0.95	0.18	0.93	1					
	I	0.43	0.45	-0.49	0.53	0.45	0.47	0.10	0.46	0.46	1				
	F	0.79	0.73	-0.67	0.79	0.90	0.89	0.17	0.88	0.93	0.76	1			
[7]	U	0.60	0.45	-0.43	0.52	0.87	0.81	0.11	0.79	0.86	0.30	0.75	1		
	G	0.71	0.72	-0.59	0.73	0.71	0.71	0.21	0.71	0.73	0.48	0.73	0.50	1	
	F	0.64	0.68	-0.55	0.68	0.57	0.59	0.21	0.59	0.60	0.45	0.63	0.31	0.98	1

Consistently, the *worms* were unlikely to be found on the tested methods.

4 Conclusions

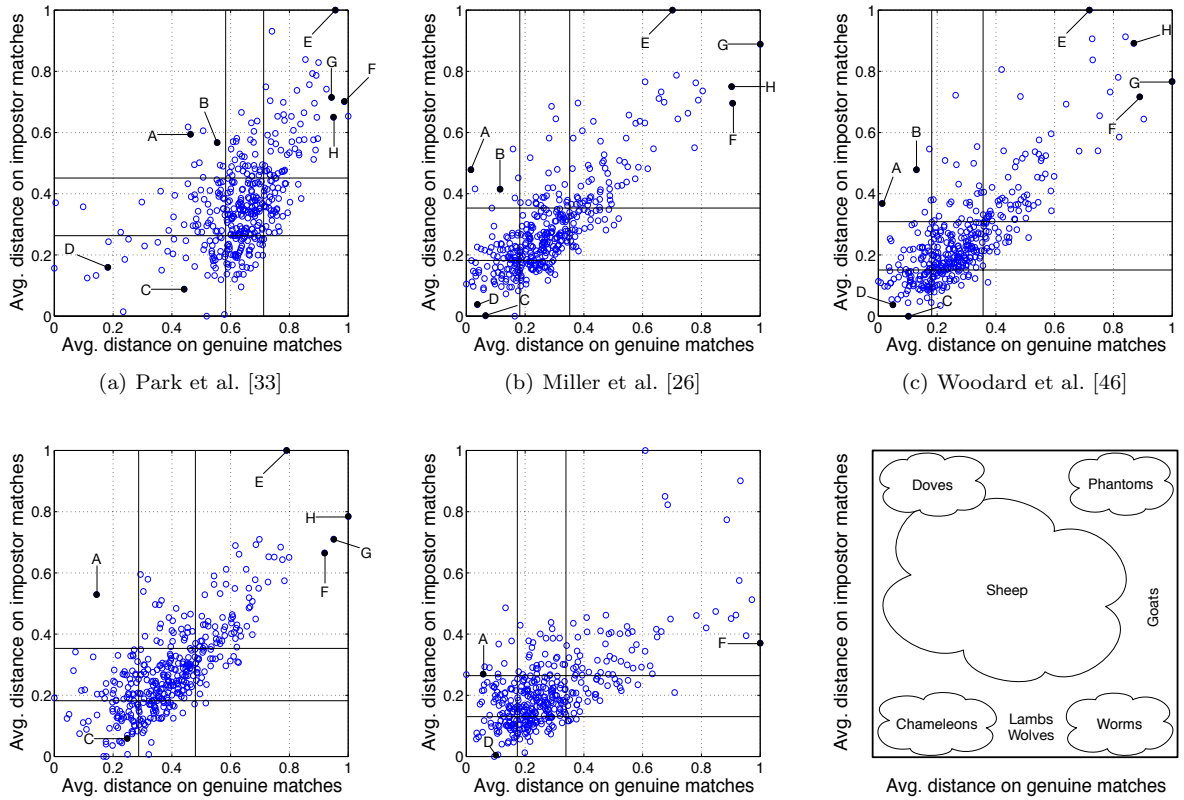
This study addressed how information in the vicinity of the eye (periocular region) can be used to perform biometric recognition, as the interest on this new biometric trait has justifiably increased over the last years.

We identified the pioneer approach of Park et al. [33] as the starting point, and how simple algorithms led to fair recognition accuracies, being noteworthy the surprisingly good response of LBP based methods.

The recently developed methods focus mainly on texture analysis and key-point extraction, and present periocular as a fit biometric trait specially for unconstrained and uncooperative scenarios, where iris cannot be properly imaged and neither a full facial picture can be obtained. Results favoring VW periocular over NIR also show its fitness for more relaxed setups, and for its use based on conventional surveillance cameras.

However, some issues were identified to considerably impact system accuracy and should be addressed, specially pose, minor occlusion, illumination and aging. Regarding the later, extending Juefei-Xu et al [21] work to different scenarios should be considered.

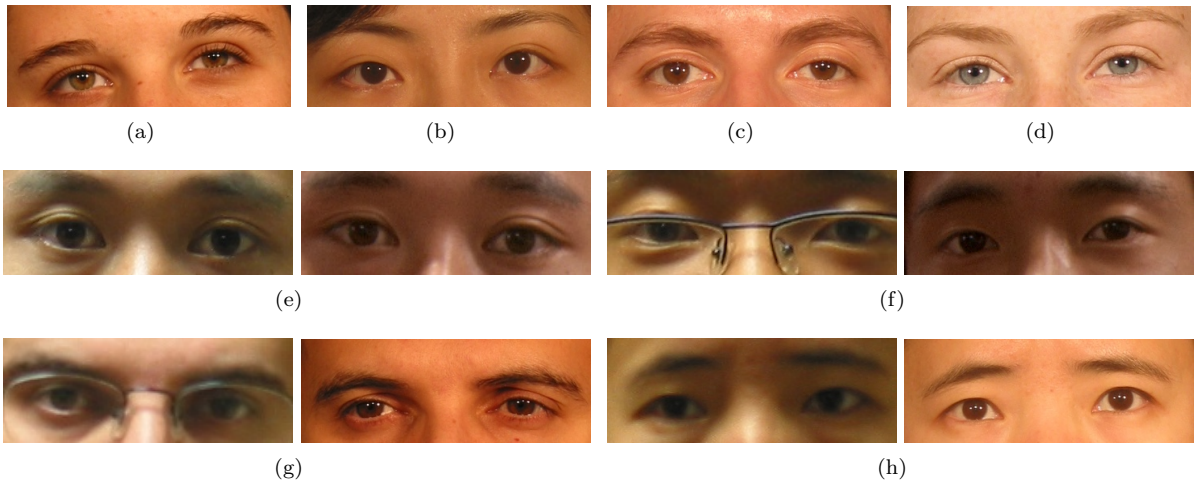
The work of Hollingsworth et al. [14,15] on human perception suggest that eye shape constitutes a powerful ally to skin analysis methods on both spectral bands, thus making us look at the periocular recognition task from a different perspective, and where a leap away the



(a) Park et al. [33] (b) Miller et al. [26] (c) Woodard et al. [46]

(d) Woodard et al. [47] (e) Bharadwaj et al. [7] (f) Reference chart

Fig. 11 Zoo plots for the tested periocular algorithms: a) Park et al. [33], b) Miller et al. [26], c) Woodard et al. [46], d) Woodard et al. [47] and e) Bharadwaj et al. [7]. Black lines represent the first and third quartiles for genuine and impostor distance distributions. For illustration of the localization of each family, a reference chart f) is also displayed (adapted from [49]).



(a) (b) (c) (d)

(e) (f)

(g) (h)

Fig. 12 Sample images from database subjects potentially belonging to each one of the identified *animal* families: doves (a and b), chameleons (c and d) and phantoms (e to h).

overused texture methods could be advantageous. Eyelashes are also point as a good indicator, specially for NIR, but we must keep in mind that used images differ from the "traditionally" used periocular images and the close capturing of the data could have biased the results.

References

1. Adams, J., Woodard, D., Dozier, G., Miller, P., Bryant, K., Glenn, G.: Genetic-based type ii feature extraction for periocular biometric recognition: Less is more. In: Pattern Recognition (ICPR), 2010 20th International Conference on, pp. 205 –208 (2010). DOI 10.1109/ICPR.2010.59
2. Adams, J., Woodard, D.L., Dozier, G.V., Miller, P.E., Glenn, G., Bryant, K.S.: Gefe: genetic & evolutionary feature extraction for periocular-based biometric recognition. In: ACM Southeast Regional Conference'10, pp. 44–44 (2010)
3. Ahmed, N., Natarajan, T., Rao, K.: Discrete cosine transform. Computers, IEEE Transactions on **C-23**(1), 90 –93 (1974). DOI 10.1109/T-C.1974.223784
4. Bakshi, S., Sa, P., Majhi, B.: Optimized periocular template selection for human recognition. BioMed Research International **2013**, 14 (2013). DOI 10.1155/2013/481431
5. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Comput. Vis. Image Underst. **110**(3), 346–359 (2008). DOI 10.1016/j.cviu.2007.09.014. URL `http://dx.doi.org/10.1016/j.cviu.2007.09.014`
6. Beer, T.: Walsh transforms. American Journal of Physics **49**(5), 466–472 (1981). DOI 10.1119/1.12714. URL `http://link.aip.org/link/?AJP/49/466/1`
7. Bharadwaj, S., Bhatt, H., Vatsa, M., Singh, R.: Periocular biometrics: When iris recognition fails. In: Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, pp. 1 –6 (2010). DOI 10.1109/BTAS.2010.5634498
8. Bhatt, H., Bharadwaj, S., Singh, R., Vatsa, M.: Recognizing surgically altered face images using multiobjective evolutionary algorithm. Information Forensics and Security, IEEE Transactions on **8**(1), 89–100 (2013). DOI 10.1109/TIFS.2012.2223684
9. Bowyer, K., Hollingsworth, K., Flynn, P.: Image understanding for iris biometrics: A survey. Comput. Vis. Image Underst. **110**(2), 281–307 (2008). DOI 10.1016/j.cviu.2007.08.005. URL `http://dx.doi.org/10.1016/j.cviu.2007.08.005`
10. Cardoso, L., Barbosa, A., Silva, F., Pinheiro, A., Proença, H.: Iris biometrics: Synthesis of degraded ocular images. Information Forensics and Security, IEEE Transactions on **8**(7), 1115–1125 (2013). DOI 10.1109/TIFS.2013.2262942
11. Clausi, D., Jernigan, M.: Towards a novel approach for texture segmentation of sar sea ice imagery. In: 26th International Symposium on Remote Sensing of Environment and 18th Annual Symposium of the Canadian Remote Sensing Society, pp. 257–261. Vancouver, BC, Canada (1996)
12. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: In CVPR, pp. 886–893 (2005)
13. Daugman, J.: High confidence visual recognition of persons by a test of statistical independence. Pattern Analysis and Machine Intelligence, IEEE Transactions on **15**(11), 1148 –1161 (1993). DOI 10.1109/34.244676
14. Hollingsworth, K., Bowyer, K., Flynn, P.: Identifying useful features for recognition in near-infrared periocular images. In: Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, pp. 1 –8 (2010). DOI 10.1109/BTAS.2010.5634529
15. Hollingsworth, K., Darnell, S., Miller, P., Woodard, D., Bowyer, K., Flynn, P.: Human and machine performance on periocular biometrics under near-infrared light and visible light. Information Forensics and Security, IEEE Transactions on **7**(2), 588 –601 (2012). DOI 10.1109/TIFS.2011.2173932
16. Hosmer, D., Lemeshow, S.: Applied logistic regression (Wiley Series in probability and statistics). Wiley-Interscience Publication (2000)
17. Hurley, D., Nixon, M., Carter, J.: A new force field transform for ear and face recognition. In: Image Processing, 2000. Proceedings. 2000 International Conference on, vol. 1, pp. 25 –28 (2000). DOI 10.1109/ICIP.2000.900883
18. Jain, A., Flynn, P., Ross, A. (eds.): Handbook of biometrics. Springer (2008)
19. Jillela, R., Ross, A.: Mitigating effects of plastic surgery: Fusing face and ocular biometrics. In: Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on (2012)
20. Juefei-Xu, F., Cha, M., Heyman, J., Venugopalan, S., Abiantun, R., Savvides, M.: Robust local binary pattern feature sets for periocular biometric identification. In: Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, pp. 1 –8 (2010). DOI 10.1109/BTAS.2010.5634504
21. Juefei-Xu, F., Luu, K., Savvides, M., Bui, T., Suen, C.: Investigating age invariant face recognition based on periocular biometrics. In: Biometrics (IJCB), 2011 International Joint Conference on, pp. 1 –7 (2011). DOI 10.1109/IJCB.2011.6117600
22. Laws, K.: Rapid texture identification. In: Proc. SPIE Conf. Image Processing for Missile Guidance, pp. 376–381 (1980). DOI 10.1117/12.959169
23. Lowe, D.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision **60**(2), 91–110 (2004). DOI 10.1023/B:VISI.0000029664.99615.94. URL `http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94`
24. Mallat, S.: A theory for multiresolution signal decomposition: the wavelet representation. Pattern Analysis and Machine Intelligence, IEEE Transactions on **11**(7), 674 –693 (1989). DOI 10.1109/34.192463
25. Miller, P., Lyle, J., Pundlik, S., Woodard, D.: Performance evaluation of local appearance based periocular recognition. In: Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, pp. 1 –6 (2010). DOI 10.1109/BTAS.2010.5634536
26. Miller, P., Rawls, A., Pundlik, S., Woodard, D.: Personal identification using periocular skin texture. In: Proceedings of the 2010 ACM Symposium on Applied Computing, SAC '10, pp. 1496–1500. ACM, New York, NY, USA (2010). DOI 10.1145/1774088.1774408. URL `http://doi.acm.org/10.1145/1774088.1774408`
27. Ojala, T., Pietikainen, M., Harwood, D.: Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In: Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision amp; Image Processing., Proceedings of the 12th

IAPR International Conference on, vol. 1, pp. 582 –585 vol.1 (1994). DOI 10.1109/ICPR.1994.576366
28. Ojala, T., Pietikainen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recognition **29**(1), 51–59 (1996). DOI 10.1016/0031-3203(95)00067-4. URL http://dx.doi.org/10.1016/0031-3203(95)00067-4
29. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Analysis and Machine Intelligence, IEEE Transactions on **24**(7), 971 –987 (2002). DOI 10.1109/TPAMI.2002.1017623
30. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. International Journal of Computer Vision **42**, 145–175 (2001)
31. Padole, C., Proença, H.: Periocular recognition: Analysis of performance degradation factors. In: Biometrics (ICB), 2012 5th IAPR International Conference on, pp. 439 –445 (2012). DOI 10.1109/ICB.2012.6199790
32. Park, U., Jillela, R., Ross, A., Jain, A.: Periocular biometrics in the visible spectrum. Information Forensics and Security, IEEE Transactions on **6**(1), 96 –106 (2011). DOI 10.1109/TIFS.2010.2096810
33. Park, U., Ross, A., Jain, A.: Periocular biometrics in the visible spectrum: A feasibility study. In: Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on, pp. 1 – 6 (2009). DOI 10.1109/BTAS.2009.5339068
34. Phillips, P., Flynn, P., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1, pp. 947 – 954 vol. 1 (2005). DOI 10.1109/CVPR.2005.268
35. Phillips, P., Moon, H., Rizvi, S., Rauss, P.: The feret evaluation methodology for face-recognition algorithms. Pattern Analysis and Machine Intelligence, IEEE Transactions on **22**(10), 1090 – 1104 (2000). DOI 10.1109/34.879790
36. Phillips, P., Scruggs, W., O'Toole, A., Flynn, P., Bowyer, K., Schott, C., Sharpe, M.: Frvt 2006 and ice 2006 large-scale experimental results. Pattern Analysis and Machine Intelligence, IEEE Transactions on **32**(5), 831 –846 (2010). DOI 10.1109/TPAMI.2009.59
37. Proença, H., Filipe, S., Santos, R., Oliveira, J., Alexandre, L.: The ubiris.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. Pattern Analysis and Machine Intelligence, IEEE Transactions on **32**(8), 1529 –1535 (2010). DOI 10.1109/TPAMI.2009.66
38. Ricanek, K., Savvides, M., Woodard, D., Dozier, G.: Unconstrained biometric identification: Emerging technologies. Computer **43**(2), 56 –62 (2010). DOI 10.1109/MC.2010.55
39. Santos, G., Hoyle, E.: A fusion approach to unconstrained iris recognition. Pattern Recognition Letters **33**(8), 984 – 990 (2012). DOI 10.1016/j.patrec.2011.08.017. URL http://www.sciencedirect.com/science/article/pii/S0167865511002686
40. Santos, G., Proença, H.: Periocular biometrics: An emerging technology for unconstrained scenarios. In: Proceedings of the IEEE Symposium on Computational Intelligence in Biometrics and Identity Management – CIBIM 2013, pp. 14–21 (2013)
41. Savvides, M., Abiantun, R., Heo, J., Park, S., Xie, C., Vijayakumar, B.: Partial holistic face recognition on frgc-ii data using support vector machine. In: Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on, p. 48 (2006). DOI 10.1109/CVPRW.2006.153
42. Shin, K.Y., Nam, G.P., Jeong, D.S., Cho, D.H., Kang, B.J., Park, K.R., Kim, J.: New iris recognition method for noisy iris images. Pattern Recognition Letters **33**(8), 991 – 999 (2012). DOI 10.1016/j.patrec.2011.08.016. URL http://www.sciencedirect.com/science/article/pii/S0167865511002674
43. Singh, R., Vatsa, M., Bhatt, H., Bharadwaj, S., Noore, A., Nooreyezdan, S.: Plastic surgery: A new dimension to face recognition. Information Forensics and Security, IEEE Transactions on **5**(3), 441–448 (2010). DOI 10.1109/TIFS.2010.2054083
44. Tan, K., Oh, S.R., Priel, A., Korn, B., Kikkawa, D.: Surgical anatomy of the forehead, eyelids, and midface for the aesthetic surgeon. In: G.G. Massry MD, M.R. Murphy MD, B. Azizzadeh MD (eds.) Master Techniques in Blepharoplasty and Periorbital Rejuvenation, pp. 11–24. Springer New York (2011). DOI 10.1007/978-1-4614-0067-7_2
45. Tan, T., Zhang, X., Sun, Z., Zhang, H.: Noisy iris image matching by using multiple cues. Pattern Recognition Letters **33**(8), 970 – 977 (2012). DOI 10.1016/j.patrec.2011.08.009. URL http://www.sciencedirect.com/science/article/pii/S0167865511002601. ¡ce:title¿Noisy Iris Challenge Evaluation II - Recognition of Visible Wavelength Iris Images Captured At-a-distance and On-the-move¡/ce:title¿
46. Woodard, D., Pundlik, S., Lyle, J., Miller, P.: Periocular region appearance cues for biometric identification. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pp. 162 –169 (2010). DOI 10.1109/CVPRW.2010.5544621
47. Woodard, D., Pundlik, S., Miller, P., Jillela, R., Ross, A.: On the fusion of periocular and iris biometrics in non-ideal imagery. In: Pattern Recognition (ICPR), 2010 20th International Conference on, pp. 201 –204 (2010). DOI 10.1109/ICPR.2010.58
48. Woodard, D., Pundlik, S., Miller, P., Lyle, J.: Appearance-based periocular features in the context of face and non-ideal iris recognition. Signal, Image and Video Processing **5**, 443–455 (2011). DOI 10.1007/s11760-011-0248-2. URL http://dx.doi.org/10.1007/s11760-011-0248-2
49. Yager, N., Dunstone, T.: Worms, chameleons, phantoms and doves: New additions to the biometric menagerie. In: Automatic Identification Advanced Technologies, 2007 IEEE Workshop on, pp. 1–6 (2007). DOI 10.1109/AUTOID.2007.380583
50. Yager, N., Dunstone, T.: The biometric menagerie. Pattern Analysis and Machine Intelligence, IEEE Transactions on **32**(2), 220–230 (2010). DOI 10.1109/TPAMI.2008.291
51. Yang, J., Zhang, D., yu Yang, J., Niu, B.: Globally maximizing, locally minimizing: Unsupervised discriminant projection with applications to face and palm biometrics. Pattern Analysis and Machine Intelligence, IEEE Transactions on **29**(4), 650 –664 (2007). DOI 10.1109/TPAMI.2007.1008
52. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. ACM Comput. Surv. **35**(4), 399–458 (2003). DOI 10.1145/954339.954342. URL http://doi.acm.org/10.1145/954339.954342

Chapter 7

A Robust Eye-Corner Detection Method for *Real-World* Data

7.1 Overview

This chapter consists of the following article:

A Robust Eye-Corner Detection Method for *Real-World* Data
Gil Santos and Hugo Proença
IEEE International Joint Conference on Biometrics – IJCB '11, October 11-13, 2011, Washington DC, USA.
DOI: 10.1109/IJCB.2011.6117596

This paper was also chosen for publication in the *IEEE Biometrics Compendium*.

A Robust Eye-Corner Detection Method for *Real-World* Data

Gil Santos and Hugo Proença

***Abstract*— Corner detection has motivated a great deal of research and is particularly important in a variety of tasks related to computer vision, acting as a basis for further stages. In particular, the detection of eye-corners in facial images is important in applications in biometric systems and assisted-driving systems. We empirically evaluated the state-of-the-art of eye-corner detection proposals and found that they achieve satisfactory results only when dealing with high-quality data. Hence, in this paper, we describe an eye-corner detection method that emphasizes robustness, i.e., its ability to deal with degraded data, and applicability to *real-world* conditions. Our experiments show that the proposed method outperforms others in both noise-free and degraded data (blurred and rotated images and images with significant variations in scale), which is a major achievement.**

I. INTRODUCTION

A corner is defined by the intersection of at least two edges. For decades, it was believed that most primitives of the human visual system were based on the detection of such points, which have well-defined positions. Corner detection is known to have particular relevance in computer vision, as it is often used as a starting point for other image recognition processes. Hence, various corner detection strategies have been emphasized in previous investigations of image segmentation, tracking, recognition and motion detection systems.

In this paper, we are particularly interested in the detection of both the temporal and nasal eye-corners in facial images. Eye-corners constitute relevant points of interest, and the ability to accurately pinpoint them is of great value in areas, such as biometrics, and applications, such as driving assistance systems. In biometrics, an emerging type of recognition is called *periocular*, based on human recognition by using data collected from around the eyes. The periocular region is particularly useful when the quality of data reduces the efficacy of other recognition strategies, such as with uncooperative subjects, when using visible light imagery or when acquiring data from moving subjects at a distance (e.g., [12], [10], [9], [15]).

Among all of the points of interest that can be extracted from the periocular region, we highlight eye-corners – the intersections between the upper and lower eyelids – because the position of eye corners does not vary with different facial expressions, levels of eye closure, gaze, eyelashes or makeup. After reviewing the state-of-the-art research on eye-corner detection, we concluded that published methods lack robustness and were developed to operate successfully only with high-quality data. We empirically determined that the performance of these approaches tends to significantly deteriorate with *real-world* data of significantly higher heterogeneity. Hence, this work proposes an eye-corner detection method suitable for imperfect environments, such as uneven lighting conditions and rotated or blurred data, with substantial differences in scale and levels of eye closure. Our method uses a periocular image as input, segments the iris and the sclera and defines a region of interest from which candidate points are extracted. Then, multiple features are linearly combined in an objective function whose optimization determines the pair of points that constitute the nasal and temporal eye-corners.

Department of Computer Science
IT - Instituto de Telecomunicações
University of Beira Interior, Covilhã, Portugal
gmelfe@ubi.pt hugomcp@di.ubi.pt

A. Related Works

Several approaches for the detection of eye-corners can be found in the literature. Harris and Stephens [6] proposed a general purpose corner detection method, which is often used in the specific case in which eye-corners with satisfactory results with high-quality data are available. Zheng *et* al. [17] estimated an initial region of interest from integral projections and located eye-corners according to a bank of Gabor-based filters, convolved at five different scales and orientations, from which averaged outputs yielded the final detection kernel. A more in-depth description of this strategy can be found in [18]. Khosravi and Safabakhsh [7] localized eye-corners in gray data, starting from the center of the iris and selecting two points on its scleric boundary at symmetric angles. Next, they found points on the eyelids according to local differences in brightness and used four masks to define motion direction. Xu *et al.* [16] used the approach of Harris and Stephens to select candidate points and then parsed them, combining semantic features using logistic regression. However, this method relies on image edges, which are difficult to obtain in unconstrained acquisition environments. Haiying and Gouping [5] proposed the weighting of Harris's response function with the variance projection function, achieving a more robust system for frontal images with no significant lighting variations or rotation. The variance projection function itself was proposed for similar purposes by Feng and Yuen [4]. More recently, Erdogmus and Dugelay [3] proposed a method that achieves good results on frontal images but also heavily relies on edge detection, and eye-corners result from the interception of polynomial functions fitted to these edges.

The remainder of this paper is organized as follows: in Section II we describe our methods in detail; Section III

presents our experiments and discusses the obtained results, with an emphasis on the robustness factor. Finally, Section IV presents our conclusions.

II. PROPOSED METHOD

A. Iris Segmentation and the Definition of the Region of Interest

(a) Eye image (b) Iris segmentation mask

Fig. 1. Data used as the input in applying our method.

As illustrated in figure 1, our method uses a periocular image as the input, and the first step is to obtain the corresponding noise-free iris binary segmentation mask. This mask discriminates between the noise-free regions of the iris and all of the remaining data and was obtained as described by Tan *et al.* [14]. This method has been shown to be effective with *real-world* data. In addition, this iris segmentation algorithm was selected because it outperformed in the NICE.I contest [1]. The segmented iris data are represented by the black regions of figure 1(b) and contain holes that correspond to the pupil and occluded iris regions. These holes were removed by zeroing out all of the regions that were unreachable when filling out the background from the edges of the image, as described in [13].

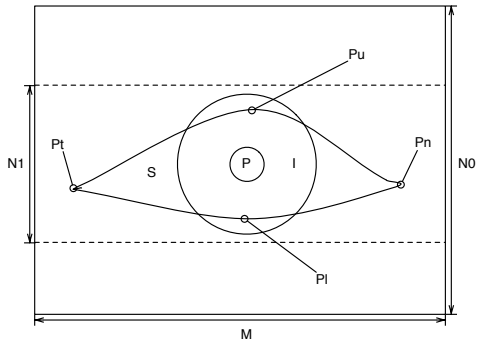


Fig. 2. An illustration of the regions of the eye involved in our work. P, I and S correspond to the pupil, iris and sclera, respectively. P_n and P_t are the nasal and temporal eye-corners, respectively. P_u and P_l are the vertical extremes of the region segmented as the iris and were used by the proposed method.

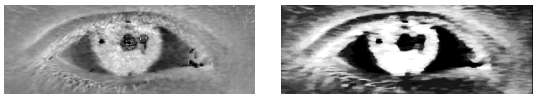
Next, we defined a region of interest (ROI) from which subsequent processing would be completed. This region is illustrated in figure 2 and was obtained by cropping the input image and the segmentation mask, avoiding unnecessary regions, such as the eyebrow and the skin underneath the eye. With an input image of dimensions $M \times N_0$, this yields regions of dimensions $M \times N_1$, according to horizontal projection techniques. This procedure ensures that the ROI is composed of the extreme coordinates of pixels belonging to the iris (P_u and P_l of figure 2):

$$\begin{aligned} y_u &= \max(y_p) \\ y_l &= \min(y_p) \end{aligned} \tag{1}$$

where y_p are the row coordinates of all of the pixels that belong to the iris.

B. Sclera Segmentation

The localization of regions that correspond to the sclera inside the ROI is critical to our method, as both eye corners should be adjacent to the sclera. In addition, pixels belonging to the human sclera have particularly low levels of saturation, which is illustrated by figure 3. The left image gives the saturation channel of the HSV colorspace (figure 3(a)), and the right image shows the result of the convolution with a unidimensional horizontal median filter [8] for eyelash attenuation, followed by data quantization and histogram equalization (figure 3(b)). This example illustrates that the sclera became more homogenous and had evidently lower intensities, enabling their classification using empirically adjusted thresholds.

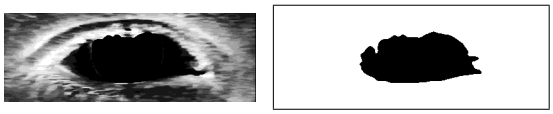


(a) Saturation values (b) 12 level histogram equalization

Fig. 3. Sclera enhancement.

C. Eye Contour Approximation

Once the iris and sclera were segmented, the next stage involved approximating the contours of the eyelids. This was performed in two steps: 1) a morphological dilation of the iris segmentation mask with a horizontal structuring element, which horizontally *expands* the iris regions, and 2) a point-by-point multiplication between the dilated and the enhanced data illustrated in figure 3(b), as described by Caselles [2]. We obtained an image similar to that illustrated in figure 4(b) and whose boundary constitutes a close approximation of the contours of the eyelids.



(a) Starting image (b) Result

Fig. 4. The eyelid contour determined corresponds to the boundary of the region indicated by black pixels.

[1] NICE.I: Noisy Iris Challenge Evaluation - Part I `http://nice1.di.ubi.pt`

D. The Generation of Eye-Corner Candidates

This stage involved the generation of a set of candidate points for the positions of the eye-corners, which was performed by using the approach of Harris and Stephens [6]. However, because of the high probability of producing too many false positives, this detector was exclusively applied inside the nasal (R_n) and temporal (R_t) regions, cropped from the extremes of the major axis of the sclera mask, as illustrated in figure 5.

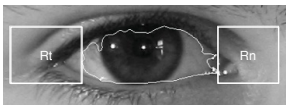


Fig. 5. An approximation of the eyelid contour (white snake) and the regions from which corner candidates are extracted (represented by white rectangles).

E. Feature Set

This stage involved finding the appropriate features to discriminate between the set of corner candidates. We also wanted to ensure that such a feature set would be robust in response to differences in translation, rotation, scale, affine-transformation and blurred data. In all subsequent descriptions, we consider $\{c_i\}_{i=1}^{n}$, $c_i = (x_i, y_i)$ to be the set of eye-corner candidates.

a) Harris Pixel Weight $H(P_c)$: Because all candidates were generated according to the Harris and Stephens method, it is straightforward to include the corresponding score in the proposed feature set. This score is given by

$$H = |M| - k\ \mathrm{tr}(M)^2 \tag{2}$$

where $|.|$ denotes the matrix determinant, tr(.) is the trace of a matrix and M is the Hessian matrix obtained from a blurred version of the original data:

$$M(x,y) = \begin{bmatrix} G_u^2(x,y) & G_{uv}(x,y) \\ G_{uv}(x,y) & G_v^2(x,y) \end{bmatrix}$$

where $G(x,y) = I(x,y) \otimes h(x,y)$, with $h(x,y) = \frac{1}{2\pi}\exp\left(\frac{x^2+y^2}{2}\right)$ and $\otimes$ denotes convolution.

b) Internal Angles: Let $B = \{b_i\}_{i=1}^{k}$, $b_i = (x_i, y_i)$ be the set of pixels belonging to the eyelid boundary obtained as described in section II-C. An ellipse fitted to B points is parameterized as follows:

$$E = (x_e, y_e) + Q(\gamma).\begin{bmatrix} A.\cos(\sigma) \\ B.\sin(\sigma) \end{bmatrix} \tag{3}$$

where (x_e, y_e) is the central point of the ellipse, $Q(\gamma)$ is a rotation matrix and A and B are the lengths of the major and minor axes, respectively. Two sets of pixels located along the opposite directions of the ellipse's minor axis are given by

$$b_l = \left(x_e - \cos\left(\gamma - \frac{\pi}{2}\right).B, y_e - \sin\left(\gamma - \frac{\pi}{2}\right).B\right) \tag{4a}$$

$$b_u = \left(x_e + \cos\left(\gamma - \frac{\pi}{2}\right).B, y_e + \sin\left(\gamma - \frac{\pi}{2}\right).B\right) \tag{4b}$$

For every candidate point c_i, two vectors $\overrightarrow{u} = c_i - b_u$ and $\overrightarrow{u} = c_i - b_l$ were obtained , and their internal angle $\theta(c_i, E)$ is given by

$$\theta_1(c_i, E) = \arccos\left(\frac{\langle u, v\rangle}{||u||.||v||}\right) \tag{5}$$

where $\langle u, v\rangle$ is the dot product between u and v, and $||\cdot||$ denotes the norm of a vector.

Let m_1 be the slope of the ellipse's major axis and m_2 be the slope of the line connecting (x_e, y_e) and the candidate point c_i:

$$m_2 = \frac{y_e - y_i}{x_e - x_i} \tag{6}$$

Their internal angle measures the agreement between the directions of the ellipse's major axis and the straight line that passes through the candidate point and the center of the ellipse:

$$\alpha_2(c_i, E) = \arctan\left(\frac{m_2 - m_1}{1 + m_1.m_2}\right) \tag{7}$$

Finally, because we are interested in pairs of eye corners, we found it useful to obtain a feature that relates any two candidates as a pair rather than scoring them independently. Let c_{i1} and c_{i2} be two corner candidates, one from the temporal and the other from the nasal region, and let l_{12} be the line that passes through both points. If the plausibility of both candidates is high, the direction of l_{12} should be similar to that of the major axis of the previously defined ellipse E. Thus, according to (7), we obtained the internal angle between these vectors ($\alpha_3(c_{i1}, c_{i2}, E)$).

c) Positions in ROIs: A complementary feature measures the relative position of each candidate in the ROIs, i.e., the proportion of pixels inside the ROI that are above each candidate. This feature is given by

$$p(c_i, R) = \frac{\sum_{i=x_i}^{N_1}\sum_{j=1}^{M}\mathbb{I}_{\{(i,j)\in R\}}}{\sum_{i=1}^{N_1}\sum_{j=1}^{M}\mathbb{I}_{\{(i,j)\in R\}}} \tag{8}$$

where $\mathbb{I}_{\{.\}}$ is an indicator function.

d) Relative Distances: This type of feature considers the distance between each candidate point c_i and the center of the ellipse:

$$d_1(c_i, E) = \frac{\sqrt{(x_i - x_e)^2 + (y_i - y_e)^2}}{A} \tag{9}$$

where (x_e, y_e) denotes the coordinates of the center of the ellipse and A the length of the ellipse's major axis to compensate for the imbalance between acquisition distance and eye size.

Let $\overrightarrow{v_a}$ be a vector with the same direction of the major axis of the ellipse and $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$ be the antipodal points of the ellipse. Let $p_{tan} = (x_{tan}, y_{tan})$ be a point tangential to the ellipse that belongs to a line that passes through c_i:

$$x_{tan} = x_1 + u(x_2 - x_1)$$
$$y_{tan} = y_1 + u(y_2 - y_1) \quad (10)$$

where u given by:

$$u = \frac{(x_c - x_1)(x_2 - x_1) + (y_c - y_1)(y_2 - y_1)}{||p_2 - p_1||^2}$$

The Euclidean distance between p_{tan} and each candidate c_i ($d_2(c_i, p_{tan})$) was also added to the feature set.

e) The Intersection of Interpolating Polynomials: The nasal and temporal eye corners can be regarded as the intersections between the upper and lower eyelids. Because of this, we parameterized two lines, each corresponding to one eyelid. The intersections t of both polynomials are illustrated by figure 6 and provide a rough estimate of the nasal and temporal eye corners. Based on our observations of the typical shape of eyelids, we used second and third degree polynomials to fit the contours of the upper and lower eyelids, respectively. Thus, the Euclidean distance between each candidate and the interception point of the corresponding ROI ($d_3(c_i, t)$) also acts as a measure of goodness for that candidate.

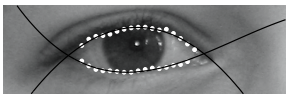


Fig. 6. Interpolating second (upper eyelid) and third (lower eyelid) degree polynomials. The interception points of both polynomials constitute an accurate approximation of the eye corners.

F. Objective Function

According to the description given in section II-E, the proposed feature set is composed of seven features: F=$\{h(c_i), \theta_1(c_i, E), \theta_2(c_i, E), p(c_i, R), d_1(c_i, E), d_2(c_i, E), d_2(c_i, p_{tan}), and d_3(c_i, t)\}$, which should be fused to produce the final score. With two sets of corner candidates (nasal and temporal), the final score for every pair of nasal c_n and temporal c_t candidates is given by the weighted sum of these features:

$$\Gamma(c_t, c_n) = \sum_{i=1}^{7} \beta_i \ f_i + \sum_{j=8}^{14} \beta_j \ f_{j-7} \quad (11)$$

where $\{\beta_1, \ldots, \beta_{14}\}$ are regularization terms adjusted to maximize performance in a training set. This optimization procedure was carried out by linear regression, and these terms were adjusted to minimize the mean squared error between the predicted values and the ground-truth data using the Akaike criterion [1]:

$$J(c_t, c_n) = (\Gamma(c_t, c_n) - g(c_t, c_n))^2 \quad (12)$$

where $g(c_t, c_n))$ is the sum of the Euclidean distances between the coordinates of the candidates and the ground-truth data.

Regularization coefficients were estimated on a sub-set of frontal images, resampled in a ten-fold cross-validation.

III. Experiments

A. Datasets

The performance of the proposed method was assessed on right-eye images of the UBIRIS.v2 database [11]. The images have dimensions of 400×300 pixels and were acquired from moving subjects in visible wavelengths at different distances and under varying lighting conditions. Additionally, the quality of the images was degraded by different factors, such as blur, motion, rotation and gaze. To check the reduction in the performance of the proposed method with respect to each factor, five dataset configurations were used and are illustrated in figure 7:

- *Frontal* – includes 300 images with the subjects' gazes aligned toward the camera;
- *Deviated Gaze* – 200 images in which the subjects' heads were deviated;
- *Blur* – images with an artificially made 50-pixel-length motion blur in the $\pi/4$ direction;
- *Clockwise rotation (CR)* – images artificially rotated by $\pi/8$ clockwise;
- *Counter-clockwise rotation (CCR)* – the same as the previous but with a counter-clockwise rotation.

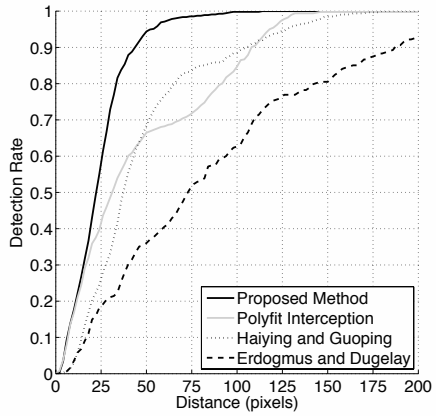
For the *Blur*, *CR* and *CCR* experiments, the images selected from the UBIRIS.v2 database were not enough, and variations were artificially made by image processing software, starting from the *frontal* subset. For all images, the data were ground-truthed manually by different experts in order to reduce subjectivity.

(a) Deviated Gaze (b) Motion Blur (c) Clockwise Rotation

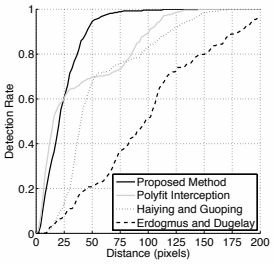
Fig. 7. Sample images from the different datasets.

B. Results

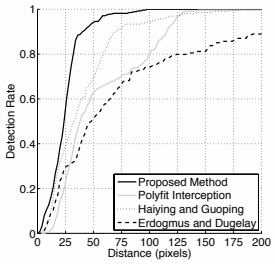
Based on the analysis of previously published research, the type of data used in this research and the results reported by the authors, we compared the performance of our method to the strategies employed by Haiying and Guoping [5] and Erdogmus and Dugelay [3]. The methods we compare ourselves to were implemented on the scope of this work and, although designed for different databases, were the ones best fitting our purposes. In addition, because we found that one of the proposed features (the intersection of polynomials) constitutes a strong estimator even when used alone, we also included this feature in our comparisons (Polyfit I.). All of the error values provided in this section correspond to the Euclidean distance between the estimated location of the eye-corners and the *true* location obtained by a manual annotation of all the images in our datasets.



(a) Overall Results



(b) Temporal Region



(c) Nasal Region

Fig. 8. Detection rate for frontal images.

Figure 8 gives the results obtained for frontal images, which is the data subset that in appearance most closely resembles the type of data the other methods are concerned with. Figure 8(a) provides the global detection rates, and figures 8(b) and 8(c) specify the results obtained for the temporal and nasal eye-corners. The horizontal axes denote the error values, and the vertical axes illustrate the proportion of images with such error values. From the analysis, it is evident that the proposed approach clearly outperformed previously reported strategies in the frontal images. When the analysis was performed separately for the nasal and temporal corners and for the temporal region, the polynomial interpolation interception was more accurate than the Erdogmus and Dugelay method, and in most cases, it showed performance similar o the proposed method. Regarding the nasal corners, we observed that all three methods behave similarly for small error values, whereas our proposal is notably better for moderate and large error values (larger than 25 pixels).

For the sake of clarity, figure 9 compares the boxplots of the error values observed for the proposed method and the methods used for comparison in the temporal (black bars) and nasal (gray bars) corners. The median of the observed performance range (horizontal solid lines) and the first and third quartile values of the observations (top and bottom of the box marks) are shown. The upper and lower whiskers are denoted by the horizontal lines outside of each box, and the outliers are denoted by dot points. This plots highlights the

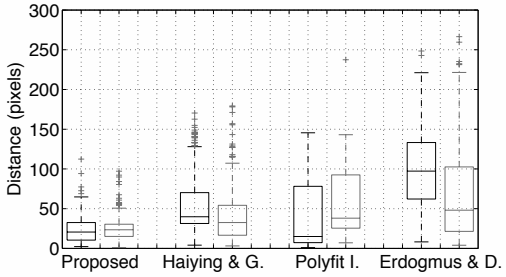
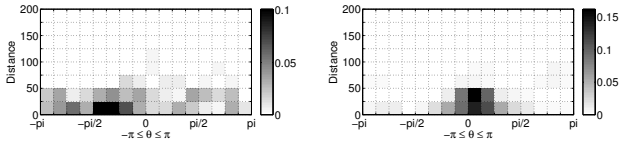


Fig. 9. The distances between the predicted corners and the true locations on frontal images. Black and gray represent the temporal and nasal regions, respectively.

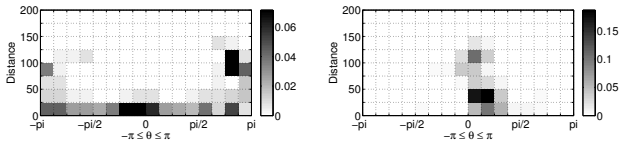
methods that are generally more efficient in detecting the nasal eye-corner, with the exception of the Erdogmus and Dugelay strategy. Again, the proposed method outperformed the previous methods.

C. Analysis of Bias

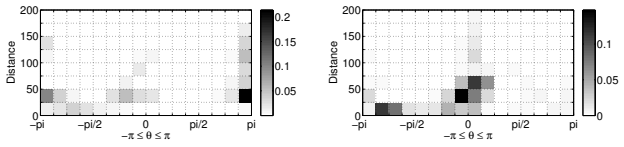
To analyze the errors that are predominant in the outputs of each method, for each case, we obtained a vector $\overrightarrow{v} = (m, \theta)$, where m is the Euclidean distance between the estimated (x_e, y_e) and true (x_t, y_t) corner position, and θ is the arctangent of $(x_e - x_t, y_e - y_t)$. The relative



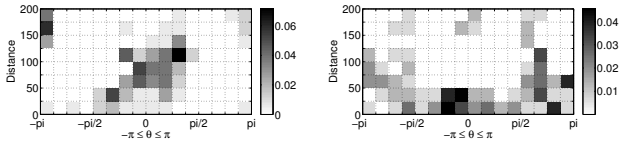
(a) Proposed Method



(b) Polyfit Interception



(c) Haiying and Guoping



(d) Erdogmus and Dugeley

Fig. 10. The relative frequencies of the observed deviations between the predicted and true positions of eye-corners. The left and right images represent the temporal and nasal corners, respectively.

frequency of these values is illustrated in figure 10, where the horizontal axis denotes the angle, and the vertical axis denotes magnitude. Deviations from the proposed method and from the polynomial interpolation interceptions are homogeneously distributed in all directions, slightly skewed toward the $[0, \frac{\pi}{2}]$ interval. Considering that our datasets are composed exclusively of right-eye images, the estimates tend to be biased northeast of the true eye-corners. On the nasal region, the prediction tends to be closer to the center of the eye than the true location. This fact is especially evident for the estimates using the Haiying and Guoping method. With the Erdogmus and Dugeley approach, temporal deviations were observed more rarely, with a slight predominance to the right of the true corner. Whereas the other methods seem to have a clear bias toward the center of the face in the nasal region, deviations were spread in all directions with the Erdogmus and Dugeley approach. This atypical behavior shown by the Erdogmus and Dugeley method in both regions probably results from the fact that, as this method is heavily dependent on edge detection, it is also considerably affected by data degradation. Notably, such distributions of deviations are in concordance with the observed correlation values, where a higher similarity between the proposed method, the interception of the polynomials and the Haying and Guoping methods was observed.

D. Robustness to Variations in the Data

Robustness is a key requirement for the proposed method, and we aimed to assess the decrease in performance when the quality of the data was degraded by different factors. In this analysis, we decided to exclusively compare the results obtained by the proposed method with those obtained using the Haiying and Guoping approach, as the latter is considered a state-of-the-art approach, and its performance was closest to ours. Figure 11 summarizes the obtained error values in the dataset, where the images were substanctially degraded as a result of the corresponding factor. The black boxplots denote the results of our method and the gray bars those determined by the Haiying and Guoping method. The analysis demonstrates the higher stability of the performance of our method across the different datasets, as the average error values are steady and remain under 50 pixels. The performance of the Haiying and Guoping method, in contrast, notably diminished when handling rotated iris data, simultaneously increasing its variance.

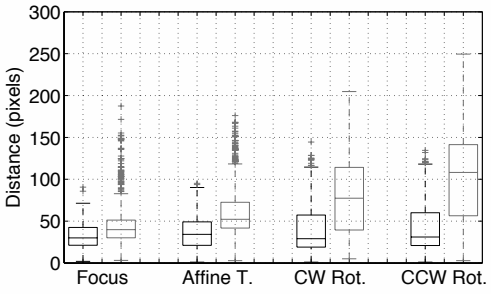
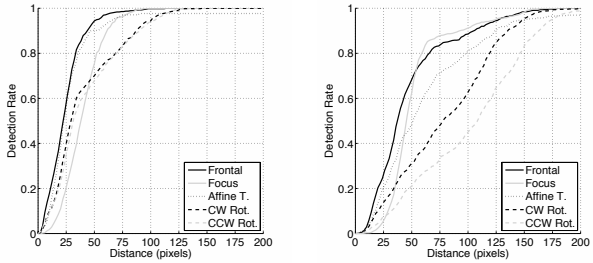


Fig. 11. The distance from the different methods' outputs to the actual eye-corners on frontal images. Black and gray represent the proposed method and Haiying and Guoping's method, respectively.

Figure 12 highlights these decreases in performance and provides the detection rates with respect to the error value (in pixels). Here, the higher slope of our method's performance plots for small errors is especially evident, which may indicate that large errors in the estimates are quite unlikely, as opposed to the values observed for the other strategy.



(a) Proposed Method (b) Haiying and Guoping

Fig. 12. Detection rate as a function of the distance for all image variations.

1) Blur: Acquiring sharp data in less controlled acquisition environments is an important issue, as slight movements of subjects often correspond to severely blurred data, a result of small depth-of-focus ranges. Thus, the ability to handle blurred data is a desirable property of any robust corner detection method. Our method only slightly decreased in performance, whereas Haiying and Guoping's performed better in some circumstances (distances from 55 to 130 pixel present an higher detection rate), with blurred data than with the focused images. The minor degradation in performance of our proposal occurred during the stage that defines the ROIs, as illustrated in figure 13; the edges become less prominent in blurred data, the region growing process stops at different iterations and consequently, the candidate search areas are also different. This, coupled with the fact that the blur also degrades the performance of the method used for the extraction of the candidates, led to a worse outcome in our proposed method.

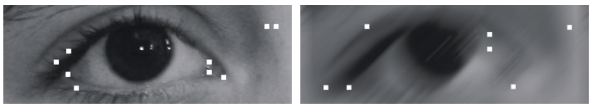


Fig. 13. Extraction of candidate points in frontal image and in the corresponding blurred version.

2) Deviated Gaze: Gaze is another important factor in less controlled acquisition environments, as it is expected that most of the time, a subject's head and eyes will not be aligned with the camera. In this case, our method behaves robustly, which was regarded as extremely positive and may indicate good performance with this type of data. There was a typical case in which our method performed better than the others: when the images had a visible background or notable facial

elements (e.g., the nasal bone). Figure 14 illustrates such cases and highlights the robustness of the proposed method for deviations in gaze.

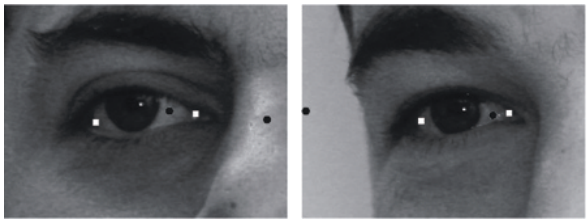
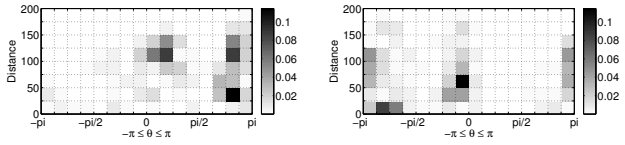
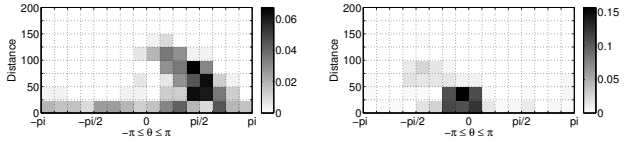


Fig. 14. An illustration of the results typically obtained in gaze-deviated images. White squares and black circles represent the outputs of our method and Haiying and Guoping's method, respectively.

3) Rotation: Rotation is another case of special interest, and significant rotations in data are expected as a result of different types of movements in an uncontrolled acquisition scene. Again, our method showed a much more robust behavior than the approach of Haiying and Guoping, which had a significantly diminished performance. We believe that this was the result of the vertical and horizontal variance projection functions that produce different results in rotated data and, consequently, bias further processing. This is highlighted by figure 15(a), in which a visible predominant bias in the opposite direction of the rotation can be seen. This is in opposition to our method, as illustrated in figure 15(b), in which a different behavior for each corner was observed: in the nasal corner, vectors counteract the direction, but angle changes are minimal. For the temporal corner, the prediction tends to follow the rotation with a larger angle variation.



(a) Haiying and Guoping



(b) Proposed Method

Fig. 15. The relative frequencies of the deviations in clockwise rotated data. The images on the left and right images are of the temporal and nasal regions, respectively.

IV. CONCLUSIONS

Several researchers are working on eye-corner detection, and the performances of different proposed methods have been found to significantly diminish in response to degraded data acquired under less controlled conditions. These shortcomings led us to propose a new method for the detection of eye-corners in periocular images that simulate *real-world* data. We compared the results obtained by our proposal to other state-of-the-art methods and concluded that our method consistently outperformed these methods, both when operating with noise-free and with degraded data (rotated, blurred, affine-transformed and with significant differences in scale). Finally, these improvements were obtained without significant increases in the computational demands of the task, which is a significant asset, considering the real-time demands that eye-corner detection techniques typically impose.

REFERENCES

[1] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, December 1974.

[2] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, February 1997.

[3] N. Erdogmus and J.-L. Dugelay. An efficient iris and eye corners extraction method. In *Proceedings of the 2010 joint IAPR international conference on Structural, syntactic, and statistical pattern recognition*, SSPR&SPR'10, pages 549–558, Berlin, Heidelberg, 2010. Springer-Verlag.

[4] G. C. Feng and P. C. Yuen. Variance projection function and its application to eye detection for human face recognition. *Pattern Recognition Letters*, 19(9):899 – 906, 1998.

[5] X. Haiying and Y. Guoping. A novel method for eye corner detection based on weighted variance projection function. In *CISP '09: Proceedings of the 2nd International Congress on Image and Signal Processing*, pages 1 –4, October 2009.

[6] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the Fourth Alvey Vision Conference*, pages 147–151, Manchester, 1988.

[7] M. H. Khosravi and R. Safabakhsh. Human eye sclera detection and tracking using a modified time-adaptive self-organizing map. *Pattern Recognition*, 41:2571–2593, 2008.

[8] J. S. Lim. *Two-Dimensional Signal and Image Processing*, pages 469–476. Prentice Hall, 1990.

[9] P. E. Miller, A. W. Rawls, S. J. Pundlik, and D. L. Woodard. Personal identification using periocular skin texture. In *SAC '10: Proceedings of the 2010 ACM Symposium on Applied Computing*, pages 1496–1500, New York, NY, USA, 2010. ACM.

[10] U. Park, A. Ross, and A. Jain. Periocular biometrics in the visible spectrum: A feasibility study. In *BTAS '09: Proceedings of the IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–6, September 2009.

[11] H. Proença, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre. The ubiris.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:1529–1535, 2010.

[12] M. Savvides, K. R. Jr., D. L. Woodard, and G. Dozier. Unconstrained biometric identification: Emerging technologies. *Computer*, 43:56–62, 2010.

[13] P. Soille. *Morphological Image Analysis: Principles and Applications*, pages 173–174. Springer-Verlag, 1999.

[14] T. Tan, Z. He, and Z. Sun. Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. *Image and Vision Computing*, 28(2):223–230, 2010.

[15] D. Woodard, S. Pundlik, P. Miller, R. Jillela, and A. Ross. On the fusion of periocular and iris biometrics in non-ideal imagery. In *ICPR 2010: Proceedings of the 20th International Conference on Pattern Recognition*, pages 201–204, August 2010.

[16] C. Xu, Y. Zheng, and Z. Wang. Semantic feature extraction for accurate eye corner detection. In *ICPR 2008: Proceedings of the 19th International Conference on Pattern Recognition*, pages 1–4, December 2008.

[17] Z. Zheng, J. Yang, M. Wang, and Y. Wang. A novel method for eye features extraction. In *Proceedings of the International Conference on Computational and Information Sciences, Lecture Notes on Computer Science, vol. 3314*, pages 1002–1007, 2004.

[18] Z. Zheng, J. Yang, and L. Yang. Semantic feature extraction for accurate eye corner detection. *Pattern Recognition Letters*, 26:2252–2261, 2005.

Chapter 8

Segmenting the Periocular Region using a Hierarchical Graphical Model Fed by Texture / Shape Information and Geometrical Constraints

8.1 Overview

This chapter consists of the following article:

Segmenting the Periocular Region using a Hierarchical Graphical Model Fed by Texture / Shape Information and Geometrical Constraints
Hugo Proença, João C. Neves and Gil Santos
IEEE International Joint Conference on Biometrics – IJCB '14, September 19 - October 2, 2014, Clearwater, Florida, USA

Segmenting the Periocular Region using a Hierarchical Graphical Model Fed by Texture / Shape Information and Geometrical Constraints

Hugo Proença, João C. Neves and Gil Santos
IT - Instituto de Telecomunicações
University of Beira Interior, Portugal
{hugomcp,jcneves, gsantos}@di.ubi.pt

Abstract

Using the periocular region for biometric recognition is an interesting possibility: this area of the human body is highly discriminative among subjects and relatively stable in appearance. In this paper, the main idea is that improved solutions for defining the periocular region-of-interest and better pose / gaze estimates can be obtained by segmenting (labelling) all the components in the periocular vicinity. Accordingly, we describe an integrated algorithm for labelling the periocular region, that uses a unique model to discriminate between seven components in a single-shot*: iris, sclera, eyelashes, eyebrows, hair, skin and glasses. Our solution fuses texture / shape descriptors and geometrical constraints to feed a two-layered graphical model (Markov Random Field), which energy minimization provides a robust solution against uncontrolled lighting conditions and variations in subjects pose and gaze.*

1. Introduction

Motivated by the pioneering work of Park *et* al. [14], the concept of periocular recognition has been gaining relevance in the biometrics literature, particularly for uncontrolled data acquisition setups. For such cases, the idea is that - apart the iris - additional discriminating information can be obtained from the skin and sclera textures, and the shape of eyelids, eyelashes and eyebrows.

Most of the relevant periocular recognition algorithms work in a *holistic* way, i.e., they define a region-of-interest (ROI) around the eye and apply a feature encoding strategy independently of the biological component at each position. The exceptions (e.g., [17] and [6]) regard the iris and the sclera components, for which specific feature encoding / matching algorithms are used. This observation leads that some components (e.g., hair or glasses) might be erroneously taken into account and bias the recognition process.

The automatic labelling (segmentation) of the components in the periocular region has - at least - two obvious advantages: it enables to define better ROIs and conducts to more accurate estimates of subjects' pose and gaze. Hence, this paper describes an image labelling algorithm for the periocular region that discriminates between seven components (iris, sclera, eyelashes, eyebrows, hair, skin and glasses), according to a model composed of two phases:

1. seven non-linear classifiers running at the pixel level are inferred from a training set, and provide the posterior probabilities for each image position and class of interest. Each classifier (neural network) is specialized in detecting one component and receives local statistics (texture and shape descriptors) from the input data;

2. the posteriors based on data local *appearance* are combined with geometric constraints and components' adjacency priors, to feed a hierarchical Markov Random Field (MRF), composed of a *pixel* and a *component* layer. MRFs are a classical tool for various computer vision problems, from image segmentation (e.g., [10]), image registration (e.g., [8]) to object recognition (e.g., [5]). Among other advantages, they provide non-causal models with isotropic behavior and faithfully model a broad range of local dependencies. The model proposed in this paper inherits some insights from previous works that used shape priors to constraint the final model (e.g., [3]) and multiple layered MRFs (e.g., [19]).

To illustrate the usefulness of the proposed algorithm, we compare the effectiveness of the Park *et* al.'s [14] recognition method, when using the ROI as originally described and according to an improved version, that considers the center of mass of the cornea as reference point (less sensitive to gaze) and avoids that hair and glasses inside the ROI are considered in feature encoding / matching. The observed improvements in performance anticipate other benefits that can be attained by labelling the periocular

region before recognition: pose / gaze estimates based in the labelled data and development of component-specific feature encoding / matching strategies.

The remainder of this paper is organized as follows: Section 2 summarizes the most relevant periocular recognition algorithms. Section 3 provides a description of the proposed model. Section 4 regards the empirical evaluation and the corresponding results. Finally, the conclusions are given in Section 5.

2. Periocular Recognition: Literature Review

The first work in this field was published in 2009, due to Park *et* al. [14]. They characterised the periocular region by local binary patterns (LBP), histograms of oriented gradients (HOG) and scale-invariant feature transforms (SIFT), fused at the score level. Subsequently, the same authors [13] described additional factors that affect performance, including segmentation inaccuracies, partial occlusions and pose. Woodard *et* al. [20] observed that fusing the responses from periocular and iris recognition modules improves performance with respect to each system considered individually. Bharadwaj *et* al. [4] fused a global descriptor based on five perceptual dimensions (image naturalness, openness, roughness, expansion and ruggedness) to circular LBPs. The Chi-square distances from both types of features were finally fused at the score level. Ross *et* al. [16] handled challenging deformed samples, using probabilistic deformation models and maximum-a-posteriori estimation filters. Also concerned about robustness, Woodard *et* al. [21] represented the skin texture and color using separate features, that were fused in the final stage of the processing chain. Tan *et* al. [18] proposed a method that got the best performance in the *NICE: Noisy Iris Challenge Evaluation*[1]. contest. This method is actually a periocular recognition algorithm: texton histograms and semantic rules encode information from the surroundings of the eye, while ordinal measures and color histograms encode the iris data. Oh *et* al. [9] combined sclera and periocular features: directional periocular features were extracted by structured random projections, complemented by a binary representation of the sclera. Tan and Kumar [17] fused iris information (encoded by Log-Gabor filters) to an over-complete representation of the periocular region (LBP, GIST, HOG and Leung-Malik Filters). Both representations were matched independently and fused at the score level.

3. Proposed Method

As Fig. 1 illustrates, the proposed MRF is composed of two layers: one works at the *pixel* level, with a bijection between each image pixel and a vertex in the MRF. The second layer regards the major *components* in the periocular vicinity, with six vertices representing the eyebrows, irises and corneas from both sides of the face. The insight behind this structure is that the pixels layer mainly regards the data appearance, while the components layer represents the geometrical constraints in the problem and assures that the generated solutions are biologically plausible.

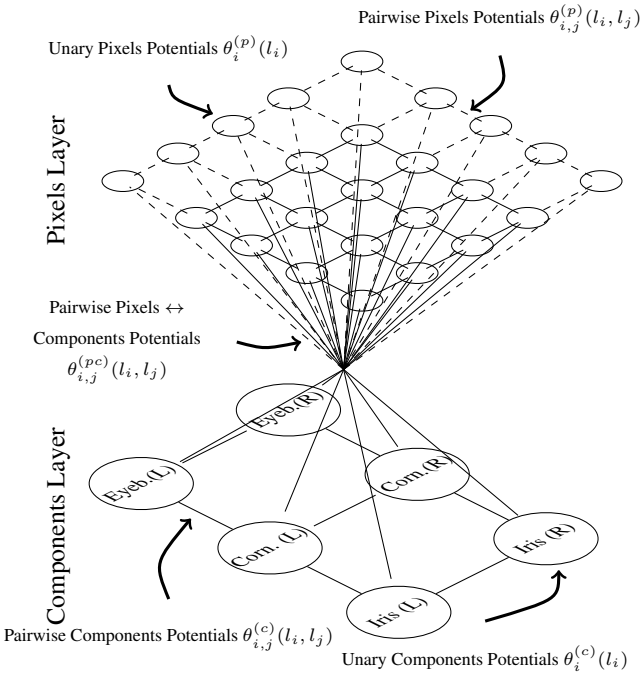


Figure 1. Structure of the MRF that segments the periocular region.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph representing a MRF, composed of a set of t_v vertices $\mathcal{V}$, linked by t_e edges $\mathcal{E}$. Let t_p be the number of vertices in the *pixels* layer and let t_c be the number of vertices in the *components* layer, such that $t_v = t_p + t_c$. Let $\mathfrak{C}(x, y)$ denote the biological component at position (x, y) of an image and $\mathfrak{T}_j$ be the component's *type* of the jth component node: either 'iris', 'cornea' or 'eyebrow'.

The MRF is a representation of a discrete latent random variable $\boldsymbol{L} = \{L_i\}, \forall i \in \mathcal{V}$, where each element L_i takes one value l_i from a set of labels. Let $\boldsymbol{l} = \{l_1, \ldots, l_{t_p}, l_{t_p+1}, \ldots, l_{t_p+t_c}\}$ be one configuration of the MRF. In our model, every component node is directly connected to each pixel node and the pixel nodes are connected to their horizontal / vertical neighbors (4-connections). Also, the edges between component nodes correspond to geometrical / biological constraints in the periocular region: the nodes representing both irises, corneas and eyebrows are connected, as do the iris, cornea and eyebrow nodes of the

[1] http://nice2.di.ubi.pt/

same side of the face. Note that the proposed model does not use high-order potentials. Even though there is a point in Fig. 1 that joins multiple edges, it actually represents overlapped pairwise connections between one component and one pixel vertex.

The energy of a configuration $\boldsymbol{l}$ of the MRF is the sum of the unary $\theta_i(l_i)$ and pairwise $\theta_{i,j}(l_i, l_j)$ potentials:

$$E(\boldsymbol{l}) = \sum_{i \in \mathcal{V}} \theta_i(l_i) + \sum_{(i,j) \in \mathcal{E}} \theta_{i,j}(l_i, l_j). \quad (1)$$

According to this formulation, labelling an image is equivalent to infer the random variables in the MRF by minimizing its energy:

$$\hat{\boldsymbol{l}} = \arg\min_{\boldsymbol{l}} E(\boldsymbol{l}), \quad (2)$$

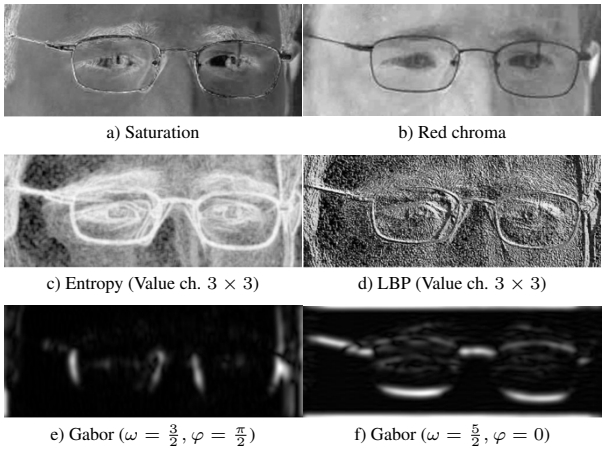
where $\{\hat{l}_1, \ldots, \hat{l}_{t_p}\}$ are the labels of the pixels and $\{\hat{l}_{t_p+1}, \ldots, \hat{l}_{t_p+t_c}\}$ specify the components' parameterizations. In this paper, the MRF was optimized according to the Loopy Belief Propagation [7] algorithm. Even though it is not guaranteed to converge to global minimums on loopy non-submodular graphs (such as our MRF), we concluded that the algorithm provides visually pleasant solutions most of the times. As future work, we plan to evaluate the effectiveness of our model according to more sophisticated energy minimization algorithms (e.g., sequential tree-reweighed message passing [11]).

3.1. Feature Extraction

Previous works reported that the hue and saturation channels of the HSV color space are particularly powerful to detect the sclera [15], whereas the red / blue chroma values provide good separability between the skin and non-skin pixels [1]. Also, the iris color triplets are typically distant from the remaining periocular components and there is a higher amount of information in patches of the eyebrows and hair regions than in the remaining components. Accordingly, a feature set at the pixel level is extracted, composed of 34 elements (Fig. 2): {red, green and blue channels (RGB); hue, saturation and value channels (HSV); red and blue chroma (yCbCr); LBP and entropy in the value channel}, all averaged in square patches of side {3, 5, 7} around the central pixel. Also, the convolution between the value channel and a set of Gabor kernels $\boldsymbol{G}$ complements the feature set:

$$\boldsymbol{G}[x, y, \omega, \varphi, \sigma] = \exp\left[\frac{-x^2 - y^2}{\sigma^2}\right] \exp[2\pi\omega i \Phi] \quad (3)$$

being $\Phi = x\cos(\varphi) + y\sin(\varphi)$, ω the spatial frequency, φ the orientation and σ the standard deviation of an isotropic Gaussian kernel ($\omega \in \{\frac{3}{2}, \frac{5}{2}\}, \varphi \in \{0, \frac{\pi}{2}\}, \sigma = 0.65\omega$).



a) Saturation b) Red chroma

c) Entropy (Value ch. 3×3) d) LBP (Value ch. 3×3)

e) Gabor ($\omega = \frac{3}{2}, \varphi = \frac{\pi}{2}$) f) Gabor ($\omega = \frac{5}{2}, \varphi = 0$)

Figure 2. Illustration of the discriminating power of the features extracted, for the seven classes considered in this paper.

3.2. Unary Potentials

Let $\gamma : \mathbb{N}^2 \rightarrow \mathbb{R}^{34}$ be the feature extraction function, that for each image pixel (x, y) returns a feature vector $\gamma(x, y) \in \mathbb{R}^{34}$. Let $\Gamma = [\gamma(x_1, y_1), \ldots, \gamma(x_n, y_n)]^T$ be a $n \times 34$ matrix extracted from a training set, that is used to learn seven non-linear binary classification models, each one specialized in detecting a component (class) $\omega_i \in$ {Iris, Sclera, Eyebrows, Eyelashes, Hair, Skin, Glasses}. Let $\eta_i : \mathbb{R}^{34} \rightarrow [0, 1]$ be the response of the ith non-linear model, used to obtain the likelihood of class ω_i: $p\Big(\eta_i\big(\gamma(x, y)\big)|\omega_i\Big)$. According to the Bayes rule, assuming equal priors, the posterior probability functions are given by:

$$P\Big(\omega_i|\eta_i\big(\gamma(x, y)\big)\Big) = \frac{P\Big(\eta_i\big(\gamma(x, y)\big)|\omega_i\Big)}{\sum_{j=1}^{7} P\Big(\eta_j\big(\gamma(x, y)\big)|\omega_j\Big)}. \quad (4)$$

The unary potentials of each vertex in the pixels layer are defined as $\theta_i^{(p)}(l_i) = 1 - p\Big(\omega_i|\eta_i\big(\gamma(x, y)\big)\Big)$.

Each label in the components layer represents a parameterisation of an ellipse (found by the Random Elliptical Hough Transform (REHT)) [2] that roughly models the eyebrows, corneal or iris regions. Starting from images labelled by the index of the maximum posterior probability $I_m(x, y) = \arg\max_j p\Big(\omega_j|\eta_j\big(\gamma(x, y)\big)\Big)$ (upper image in Fig. 3), a binary version per component can be obtained (bottom images in Fig. 3):

$$I_{m_i}(x, y) = \begin{cases} 1 & \text{, if } I_m(x, y) = i \\ 0 & \text{, otherwise} \end{cases} \quad (5)$$

The output of the REHT algorithm in $I_{m_i}(x,y)$ gives the unary potential of the component vertices: $\theta_i^{(c)}(l_i) = -\log\big(\kappa(i)\big), \forall i \in t_{p+1},\ldots,t_{p+c}$, being $\kappa(i)$ the votes returned by the REHT for the ith ellipse parameterisation.

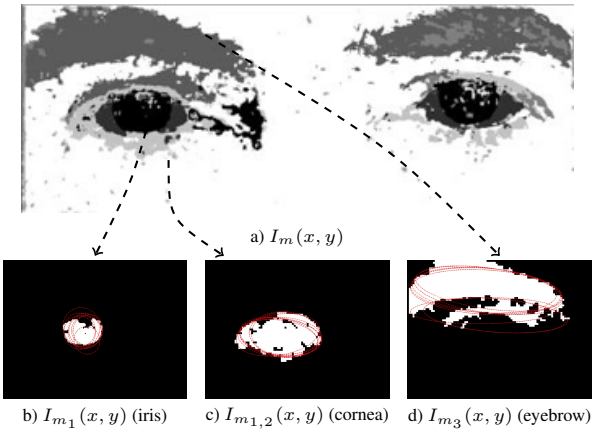


Figure 3. (Upper row) Example of an image labelled by the maximum of the posteriors given by the classification models $\eta_i\big(\gamma(x,y)\big)$. The red ellipses in the bottom images represent the parameterisations returned by the REHT algorithm for the left iris, cornea and eyebrow.

3.3. Pairwise Potentials

There are three types of pairwise potentials in our model: 1) between two pixel nodes; 2) between two component nodes; and 3) between a pixel and a component. The pairwise potential between pixel nodes spatially adjacent $\theta_{i,j}^{(p)}(l_i,l_j)$ is defined as the prior probability of observing labels l_i, l_j in adjacent positions of a training set (e.g., it is much more probable that an "eyebrow" pixel is adjacent to a "skin" pixel than to an "iris" one):

$$\theta_{i,j}^{(p)}(l_i,l_j) = \frac{1}{\alpha_0 + P(\mathfrak{C}(x',y') = \omega_i, \mathfrak{C}(x,y) = \omega_j)}, \quad (6)$$

where $P(.,.)$ is the joint probability, (x',y') and (x,y) are 4adjacent positions and $\alpha_0 \in \mathbb{R}^+$ avoids infinite costs (likewise, all α_i terms below are regularization terms).

The pairwise potentials between component nodes consider the geometrical constraints in the periocular area, i.e., enforce that the irises are inside the cornea, and below the eyebrows. Also, both irises, corneas and eyebrows should have similar vertical coordinate and similar size. Let $(x_i, y_i, a_i, b_i, \varphi_i)$ be the ith parameterisation of an ellipse, being (x_i, y_i) the ellipse centre, (a_i, b_i) its major / minor axes and φ_i the rotation. For pairs of nodes of the same type ($\mathfrak{T}_i = \mathfrak{T}_j$), similar vertical coordinates and similar sizes are privileged:

$$\theta_{i,j}^{(c1)}(l_i,l_j) = \alpha_1|y_i - y_j| + \alpha_2|a_i + b_i - a_j - b_j|. \quad (8)$$

For edges connecting the cornea (ith node) and the eyebrow (jth node) we privilege similar horizontal coordinates and locations having the eyebrow above the cornea:

$$\theta_{i,j}^{(c2)}(l_i,l_j) = \alpha_3|x_i - x_j| + \alpha_4 \max\{0, y_i - y_j\}. \quad (9)$$

Regarding the iris / cornea pairwise potentials, we penalize parameterizations with portions of the iris outside the cornea:

$$\theta_{i,j}^{(c3)}(l_i,l_j) = \alpha_5\Big(1 - \frac{\sum_{x_i}\sum_{y_i}\psi(x_i,y_i,x_j,y_j,a_j,b_j,\varphi_j)}{\sum_{x_i}\sum_{y_i}1}\Big), \quad (10)$$

being (x_i, y_i) a pixel labelled as iris and $\psi(x_i,y_i,x_j,y_j,a_j,b_j,\varphi_j)$ an indicator function that verifies if that position is inside the ellipse defined by the jth parameterisation (7). Overall, the pairwise potentials in the components layer are defined as:

$$\theta_{i,j}^{(c)}(l_i,l_j) = \sum_{k=1}^{3}\theta_{i,j}^{(ck)}(l_i,l_j). \quad (11)$$

Lastly, the pairwise potentials between pixels and components enforce that pixels inside a component parameterisation are predominantly labelled by the value that corresponds to that type of node, whereas pixels outside that parameterisation should have label different of the component's type. Let (x_{jk}, y_{jk}) be the coordinates of the ellipse defined by the jth parameterization. The pairwise cost between the ith pixel node and the jth component node is given by:

$$\theta_{i,j}^{(pc)}(l_i,l_j) = \begin{cases} \min_k ||(x_i,y_i) - (x_{jk},y_{jk})||_2, \text{if } l_i \in \mathfrak{T}_j \\ \quad \text{and } \psi(x_i,y_i,x_j,y_j,a_j,b_j,\varphi_j) = 0 \\ 0, \text{if } l_i \notin \mathfrak{T}_j \\ \quad \text{and } \psi(x_i,y_i,x_j,y_j,a_j,b_j,\varphi_j) = 0 \\ 0, \text{if } l_i \in \mathfrak{T}_j \\ \quad \text{and } \psi(x_i,y_i,x_j,y_j,a_j,b_j,\varphi_j) = 1 \\ \max_k ||(x_i,y_i) - (x_{jk},y_{jk})||_2, \text{if } l_i \notin \mathfrak{T}_j \\ \quad \text{and } \psi(x_i,y_i,x_j,y_j,a_j,b_j,\varphi_j) = 1 \end{cases}, \quad (12)$$

where $||.||$ is the Euclidean distance.

4. Experiments

Our experiments were carried out in a data set composed of 5,551 visible-light images (with resolution 800 $\times$

$$\psi(x,y,x_i,y_i,a_i,b_i,\varphi_i)=\begin{cases}1 & \text{, if } \frac{\left(\cos(\varphi_i)(x-x_i)+\sin(\varphi_i)(y-y_i)\right)^2}{a_i^2}+\frac{\left(\sin(\varphi_i)(x-x_i)+\cos(\varphi_i)(y-y_i)\right)^2}{b_i^2}\leq 1\\ 0 & \text{, otherwise}\end{cases} \quad (7)$$

300) containing the periocular regions from both sides of the face. These images were the source for the UBIRIS.v2 dataset: they were collected in indoor unconstrained lighting environments and feature significant variations in scale, subjects' pose and gaze. For learning / evaluation purposes, 200 images were manually labelled, covering the seven classes we aim to deal with. This set was divided into two disjoint parts: 1) one used to learn the classification models and to estimate the prior unary / pairwise costs of the MRF; and 2) the complementary part served for quantitative performance evaluation.

To obtain the seven classification models, we used feed-forward neural networks with three layers and $\{34:17:1\}$ topology, with *tan-sigmoid* transfer functions in the input and hidden layers and linear transfer functions in the output layer. The learning sets were always balanced (random sampling) and the Resilient Back-propagation algorithm used to learn the classifiers. Regarding the MRF optimization, every image was resized to 200×75 pixels, i.e., $t_p = 15,000$ in our MRFs. Also, $\boldsymbol{\alpha} = \{0.01, 1, 2, 10, 10\}$.

4.1. Segmentation Performance

Fig. 4 illustrates the results typically attained by the proposed model. Their visual coherence is evident, where regions labelled as hair appear in pink, eyebrows in yellow, irises in green, eyelashes in black, sclera in blue and glasses in blueberry color. Also, solutions were biologically plausible in the large majority of the cases, for various hairstyles, and different subjects poses / gazes. A particularly interesting performance was observed for glasses, where the algorithm attained remarkable results for various types of frames. This was probably due to the fact that glasses were the unique non-biological component among the classes considered, which might had increased their dissimilarity with respect to the remaining components.

In opposition, the most concerning cases happened when the eyebrows and the hair were overlapped (bottom-right image in Fig. 4) . Also, for heavily deviated gazes, the sclera was sometimes under-segmented (typically, by non-detecting the less visible side). In opposition, eyelashes tended to be over-segmented, with isolated eyelashes being grouped in large eyelash regions, which might be due to excessive pairwise cost for observing different labels in adjacent positions of the pixels layer.

It should be noted that α_i were found in an empirical and independent way, i.e., no exhaustive evaluation of combined configurations was carried out, nor any parameter optimization algorithm was used, which also points for the robustness of the proposed model against sub-optimal parameterizations. Table 1 gives the error rates per class, when considering exclusively the first phase of our model (maximum of the posterior probabilities, column "NN") and the full processing chain (MRF optimization, column "MRF"). In this table, FP stands for the false positives rate, whereas FN refers to the false negatives rate. In all cases, it is evident that the MRF substantially lowered the labeling error rates, essentially by imposing smoother responses and constraining the range of biologically acceptable solutions.

Labeling Error	NN (%)		MRF (%)	
Component	FP	FN	FP	FN
Iris	1.12 ± 0.29	9.06 ± 1.80	0.17 ± 0.03	2.61 ± 0.51
Sclera	1.61 ± 0.49	5.17 ± 0.83	0.19 ± 0.03	3.60 ± 0.82
Eyebrows	2.20 ± 0.40	6.93 ± 0.95	0.79 ± 0.28	2.25 ± 0.46
Eyelashes	1.47 ± 0.38	5.12 ± 1.13	0.93 ± 0.23	0.62 ± 0.53
Hair	3.16 ± 0.56	6.74 ± 1.27	1.26 ± 0.30	3.09 ± 0.88
Skin	4.10 ± 1.03	4.09 ± 0.69	2.63 ± 0.43	3.86 ± 1.01
Glasses	1.08 ± 0.22	5.03 ± 1.45	0.06 ± 0.01	0.60 ± 0.09

Table 1. Average pixel labelling errors per component, when considering exclusively the $\arg\max_j p\Big(\omega_j|\eta_j\left(\gamma(x,y)\right)\Big)$ value (NN column) and with the proposed MRF model (MRF column).

As the machine learning algorithm described in this paper is *supervised*, it is important to perceive its variations in performance with respect to the amount of learning data used to create the classification models and the prior unary / pairwise potentials. To this end, performance was compared while varying the number of images used in learning, and keeping constant the number of images used in performance evaluation (to assure comparable bias / variance scores). Figure 5 expresses the results: the horizontal axis gives the number of learning images used and the vertical axis is the corresponding pixel classification error, with the corresponding 95% confidence intervals. We observed that when more than 35 images were used in learning, the pixel classification errors tend to converge. This is evident in terms of the absolute error values and of the narrowness of the confidence intervals.

Figure 4. Examples of the segmented periocular regions. "Hair" class is represented by the pink color, "Eyebrows" appear in yellow, "Iris" in green, "Sclera" in blue, "Glasses" in blueberry and "Eyelashes" in gray. Pixels classified as "Skin" are transparent.

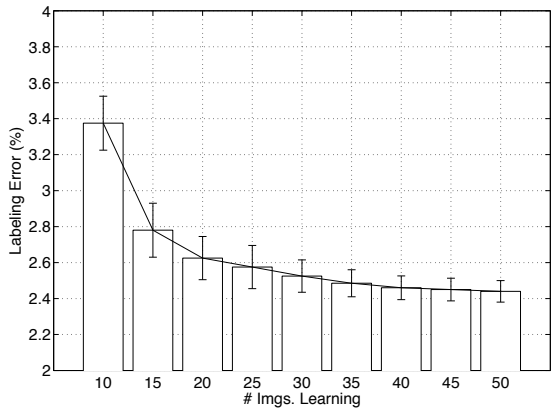


Figure 5. Variations in labelling errors with respect to the number of images used in the learning phase of the algorithm.

4.2. Periocular Biometrics Performance

To exemplify the usefulness of periocular segmentation algorithms, one *all-against-all* matching experiment was designed, using the method of Park *et* al. [13] and two different strategies to define the ROI: as baseline, the iris center was the unique reference for the ROI (upper-left image in Fig. 6). Next, according to the labels provided by the MRF, the center of mass of the cornea was used to define the ROI, which is obviously less sensitive to changes in gaze. Also, regions labelled as hair and glasses were disregarded from the recognition phase, considering that they likely suffer of significant variations among samples of a subject (upper-right image in Fig. 6). The Receiver Operating Characteristic curves for both variants are compared in the bottom plot of Fig. 6 and turn evident the benefits attained due to data segmentation (Equal error rate of 0.128 for the classical ROIs and 0.095 for the improved ROIs configuration). The improvements were substantial in all regions of the performance space, having at some operating points increased the system sensitivity over 10%. It should be stressed that no particular concerns were taken in optimizing the recognition method for the used data set, meaning that the focus was putted much more in the performance gap between both recognition schemes than in the recognition errors in absolute values, which are out of the scope of this paper.

5. Conclusions and Further Work

In this paper we have proposed an algorithm for *one-shot* labelling of all the components in the periocular region: iris, sclera, eyelashes, eyebrows, hair, skin and glasses. Our solution is composed of two major phases: 1) a group of local classification models gives the posterior probabilities for each pixel and class considered; 2) this *appearance*-based information is fused to geometrical constraints and shape priors to feed a two-layered MRF. One layer represents *pixels*, and analyzes the local data appearance while enforcing smoothness of the solutions. The second layer represents *components*, and assures that solutions are biologically plausible. By minimizing the MRF energy, the label of each pixel is found, yielding solutions that are robust against changes in scale, subjects' pose and gaze and dynamic lighting conditions.

As further directions for this work, our efforts are focused in estimate gaze / pose from the labelled data, in order to compensate for deviations before the recognition process.

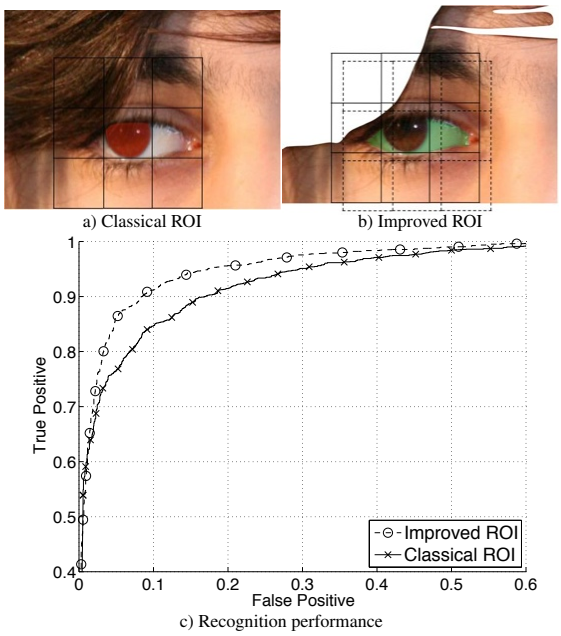


Figure 6. Improvements in periocular recognition performance due to the semantic categorization (labeling) of each pixel in the periocular region.

References

[1] A. Albiol, L. Torres and E. Delp. Optimum color spaces for skin detection. In Proceedings of the *International Conference on Image Processing*, vol. 1, pag. 122–124, 2001.

[2] C. Basca, M. Talos and R. Brad. Randomized Hough Transform for Ellipse Detection with Result Clustering. in Proceedings of the *The International Conference on Computer as a Tool EUROCON*, vol. 2, pag. 1397–1400, 2005.

[3] A. Besbes, N. Komodakis, G. Langs and N. Paragios. Shape Priors and Discrete MRFs for Knowledge-based Segmentation. in Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition*, pag. 1295–1302, 2009.

[4] S. Bharadwaj, H. Bhatt, M. Vatsa and R. Singh. Periocular biometrics: When iris recognition fails. In Proceedings of the 4^{th} *IEEE International Conference on Biometrics: Theory Applications and Systems*, pag. 1–6, 2010.

[5] B. Caputo, S. Bouattour and H. Niemann. Robust appearance-based object recognition using a fully connected Markov random field In Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition*, vol., 3, pag. 565–568, 2002.

[6] S. Crihalmeanu and A. Ross. Multispectral scleral patterns for ocular biometric recognition. *Pattern Recognition Letters*, vol. 33, no. 14, pag. 1860–1869, 2012.

[7] P. Felzenszwalb and D. Huttenlocher. Efficient Belief Propagation for Early Vision. *International Journal of Computer Vision*, vol. 70, no. 1, pag. 41–54, 2006.

[8] B. Glocker, D. Zikic, N. Komodakis, N. Paragios and N. Navab. Linear Image Registration Through MRF Optimization. In Proceedings of the *IEEE International Symposium on Biomedical Imaging*, pag. 422–425, 2009.

[9] K. Oh, B-S. Oh, K-A. Toh, W-Y. Yau and H-L. Eng. Combining sclera and periocular features for multi-modal identity verification. *Neurocomputing*, vol. 128, pag. 185–198, 2014.

[10] Z. Kato and T.C. Pong. A Markov random field image segmentation model for textured images. *Image and Vision Computing*, vol. 24, pag. 1103–1114, 2006.

[11] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pag. 1568–1583, 2006.

[12] T. Ojala, M. PietikŁinen and D. Harwood. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In Proceedings of the 12^{th} *IAPR International Conference on Pattern Recognition*, pag. 582–585, 1994.

[13] U. Park, R. Jillela, A. Ross and A. Jain. Periocular Biometrics in the Visible Spectrum. *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pag. 96–106, 2011.

[14] U. Park, A. Ross and A. Jain. Periocular biometrics in the visible spectrum: A feasibility study. In Proceedings of the *IEEE* 3^{rd} *International Conference on Biometrics: Theory, Applications, and Systems*, pag. 1–6, 2009.

[15] H. Proença. Iris Recognition: On the Segmentation of Degraded Images Acquired in the Visible Wavelength. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pag. 1502–1516, 2010.

[16] U. Park, A. Ross and A. Jain. Matching highly non-ideal ocular images: an information fusion approach. In Proceedings of the *IEEE* 5^{th} *International Conference on Biometrics*, pag. 446–453, 2012.

[17] C-W. Tan and A. Kumar. Towards Online Iris and Periocular Recognition Under Relaxed Imaging Constraints. *IEEE Transactions on Image Processing*, vol. 22, no. 10, pag. 3751–3765, 2013.

[18] T. Tan, X. Zhang, Z. Sun and H. Zhang. Noisy iris image matching by using multiple cues. *Pattern Recognition Letters*, vol. 33, pag. 970–977, 2012.

[19] C. Wang, M. Gorce and N. Paragios. Segmentation, Ordering and Multi-Object Tracking using Graphical Models. In Proceedings of the *in* 12^{th} *International Conference on Computer Vision*, pag. 747–754, 2009.

[20] D. Woodard, S. Pundlik, P. Miller, R. Jillela and A. Ross. On the fusion of periocular and iris biometrics in non-ideal imagery. In Proceedings of the *in* 20^{th} *International Conference on Pattern Recognition*, pag. 201–204, 2010.

[21] D. Woodard, S. Pundlik, P. Miller and J. Lyle. Appearance-based periocular features in the context of face and non-ideal iris recognition. *Signal, Image and Video Processing*, vol. 5, pag. 443–455, 2011.

Chapter 9

A Fusion Approach to Unconstrained Iris Recognition

9.1 Overview

This chapter consists of the following article:

A Fusion Approach to Unconstrained Iris Recognition
Gil Santos and Edmundo Hoyle
Pattern Recognition Letters, 33(8), 984-990, June 2012.
DOI: 10.1016/j.patrec.2011.08.017

According to SCImago Journal & Country Rank, this journal's index[1] for the 2013 year are as follows:

Category	*Quartile*	*SJR*
Computer Vision and Pattern Recognition	Q1	0,995
Artificial Intelligence	Q2	
Signal Processing	Q1	
Software	Q1	

[1]The SCImago Journal & Country Rank (SJR) indicator is a measure of journal's impact, influence or prestige. It expresses the average number of weighted citations received in the selected year by the documents published in the journal in the three previous years. `http://www.scimagojr.com`

Pattern Recognition Letters 33 (2012) 984–990

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

A fusion approach to unconstrained iris recognition

Gil Santos [a,*], Edmundo Hoyle [b]

[a] Department of Computer Science, IT – Instituto de Telecomunicações, University of Beira Interior, Covilhã, Portugal
[b] Department of Electrical Engineering, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

ARTICLE INFO

Article history:
Available online 7 September 2011

Keywords:
Nonideal iris images
Iris recognition
Unconstrained biometrics
Decision fusion

ABSTRACT

As biometrics has evolved, the iris has remained a preferred trait because its uniqueness, lifetime stability and regular shape contribute to good segmentation and recognition performance. However, commercially deployed systems are characterized by strong acquisition constraints based on active subject cooperation, which is not always achievable or even reasonable for extensive deployment in everyday scenarios. Research on new techniques has been focused on lowering these constraints without significantly impacting performance while increasing system usability, and new approaches have rapidly emerged. Here we propose a novel fusion of different recognition approaches and describe how it can contribute to more reliable noncooperative iris recognition by compensating for degraded images captured in less constrained acquisition setups and protocols under visible wavelengths and varying lighting conditions. The proposed method was tested at the NICE.II (Noisy Iris Challenge Evaluation – Part 2) contest, and its performance was corroborated by a third-place finish.

1. Introduction

The use of the iris as main biometric trait has emerged as one of the most recommended methods due not only to the possibility of noncontact data acquisition and to its circular and planar shape that facilitates detection, segmentation and compensation for off-angle capture but also for its predominately randotypic appearance. Although these factors contribute to high effectiveness in the currently deployed iris-recognition systems, their typical scenarios are quite constrained: subjects stop and stare relatively close to the acquisition device while their eyes are illuminated by a near-infrared light source, enabling the acquisition of high-quality data. As reported in the study conducted by *Aton Origin* for the United Kingdom Passport Service,[1] imaging constraints are a major obstacle for the mass implementation of iris-based biometric systems. Notably, several researchers are currently working on minimizing the constraints associated with this process, in a way often referred to as noncooperative iris recognition, referring to several factors that can make iris images nonideal, such as at-a-distance imagery, on-the-move subjects, and high dynamic lighting variations.

In this study, we stress multiple recognition techniques, each one based on a different rationale and exploiting different properties of the eye region. Furthermore, we show how their fusion can increase the robustness to the degraded data typically captured in unconstrained acquisition setups.

The recognition techniques used in our proposition can be divide in two main categories. In one approach, we use wavelet-based iris-feature-extraction methods, complemented with a zero-crossing representation (Hoyle et al., 2010, 2009) and the analysis of iriscode-matching bit distribution (Santos and Proença, 2010). Complementarily, we expanded the extraction of features to the ocular region outside the iris, as recent studies (Savvides et al., 2010; Miller et al., 2010; Park et al., 2009) have suggested using these data, which appear to be a middle ground between iris and face biometrics and incorporates some advantages of each.

The performance of the fusion method we propose is highlighted by its third-place finish at the NICE.II (Noisy Iris Challenge Evaluation – Part 2), an international contest involving almost seventy participants worldwide.

The remainder of this paper is structured as follows: Section 2 describes the steps for iris-boundary localization and normalization, feature extraction and matching for the different approaches, and how their outputs are joined; Section 3 details the experimental process followed by a discussion of the obtained results; finally, Section 4 states the conclusions.

2. Proposed methodology

This section describes the five steps of our approach: iris-boundary detection, iris normalization, feature extraction, matching and decision ensemble (as schematized in Fig. 1). Furthermore, for feature extraction and matching, five recognition techniques are detailed.

* Corresponding author. Tel.: +351 92 683 24 68.
E-mail addresses: gmelfe@ubi.pt (G. Santos), edhoyle@pads.ufrj.br (E. Hoyle).

[1] http://www.ips.gov.uk/cps/rde/xchg/ips_live/hs.xsl/publications.htm.

0167-8655/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2011.08.017

G. Santos, E. Hoyle / Pattern Recognition Letters 33 (2012) 984–990

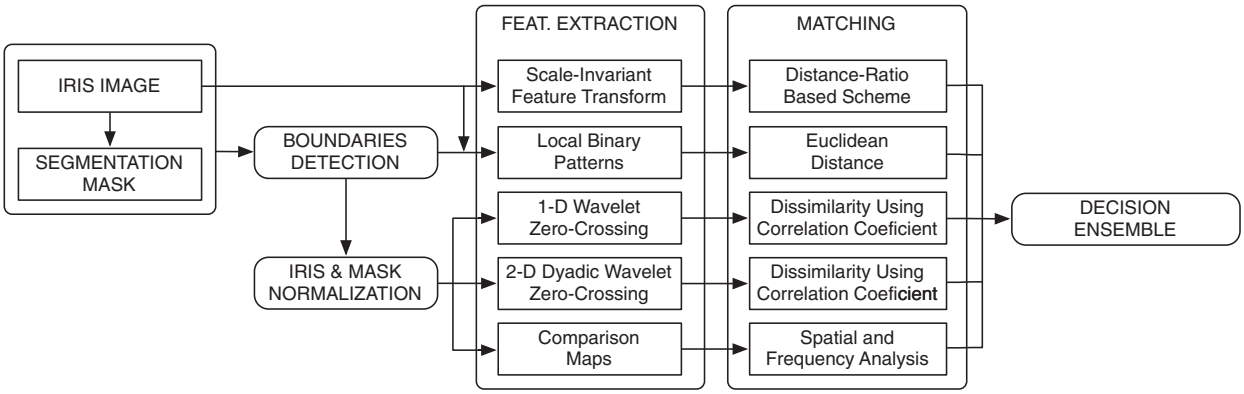


Fig. 1. Proposed methodology.

2.1. Iris boundaries detection

The first task was to locate the circles that best approximate iris and pupil boundaries, a necessity in the majority of methods used for this work. To accomplish this, we utilized a binary mask representing only parts containing iris information, created using the method proposed by Tan et al. (2010), winner of the NICE.I contest.

The steps taken in boundary approximation (Fig. 2(h)) were as follows:

- A contour is extracted from the segmentation mask Fig. 2(b), created with Tan et al. method (Tan et al., 2010). A pixel is part of such contour if it is nonzero, and connected to at least one zero-valued pixel.
- From the contour Fig. 2(c) of the segmentation mask Fig. 2(b), a Hough transform (Ballard, 1981) is applied to obtain the circle best fitting the iris Fig. 2(d).
- Convert the eye image Fig. 2(a) to grayscale and enhance it through histogram equalization Fig. 2(e).
- To the enhanced image Fig. 2(e), a Canny edge detection (Canny, 1986) is applied inside the circular region Fig. 2(f) concentric with the iris and 2/3 its radius, producing the edges shown in subFig. 2(g).
- Finally, a Hough transform is used on the resulting edge map Fig. 2(g) to obtain the circle that best fits the pupil.

Although this method produces good iris-boundary approximations, the estimated pupil limits sometimes diverge from ideal contours (*e.g.* Fig. 3). The main reason for this occurrence is poor lighting conditions when imaging heavily pigmented irises, which results in a low contrast ratio between the iris and the pupil.

2.2. Iris normalization

The iris-normalization process aims to obtain invariance with respect to size, position and pupil dilatation in the segmented iris region, which is accomplished by assigning each pixel to a pair of real coordinates (r,θ) over the double dimensionless pseudopolar coordinate system. For this purpose, we proceeded with the rubber-sheet model originally proposed by Daugman (2004).

$$I(x(r,\theta),y(r,\theta)) \rightarrow I(r,\theta) \tag{1}$$

$$\begin{aligned} x(r,\theta) &= (1-r)x_p(\theta) + rx_s(\theta) \\ y(r,\theta) &= (1-r)y_p(\theta) + ry_s(\theta) \end{aligned} \tag{2}$$

where r and θ denote the radius and the angle, respectively, and $x(r,\theta)$ and $y(r,\theta)$ are defined as linear combinations of both the set of pupillary boundary points $(x_p(\theta),y_p(\theta))$ and the set of limbus boundary points along the outer perimeter of the iris $(x_s(\theta),y_s(\theta))$ bordering the sclera.

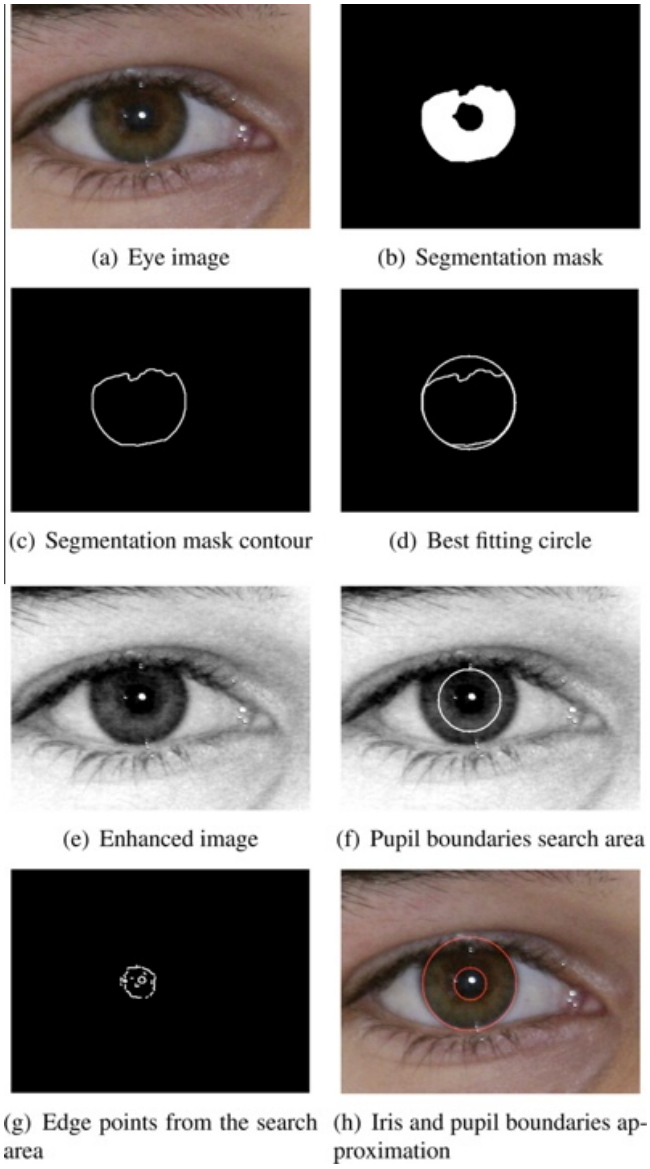


Fig. 2. Illustration of the steps taken during the segmentation stage.

Eqs. (1) and (2) give a transformation similar to that depicted in Fig. 4: subfigure (a) is the normalized iris image; subfigure (b)

Fig. 3. Illustration of unsuccessful inner boundary detection.

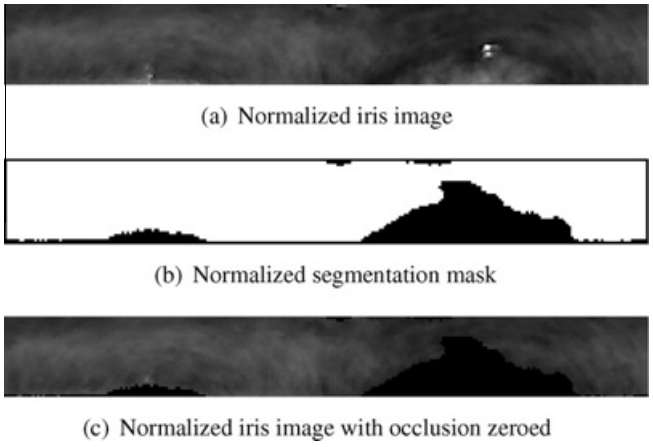


Fig. 4. Normalized images. Iris data are represented in grayscale.

represents the normalization of its mask (occlusions being the black region); and subfigure (c) is the normalized iris image where the occlusion has been zeroed. In either case, no interpolation was used, being chosen the nearest pixel to fill eventual gaps.

2.3. Feature extraction

Feature extraction and representation varies according to the employed method, as detailed herein.

2.3.1. 1-D wavelet zero-crossing representation

The representation method applied here is an extension of the Boles method (Boles and Boashash, 1998; Boles, 1997). Other studies (Hoyle et al., 2010, 2009) have shown that the proposed extension significantly improves the recognition performance.

The starting point for iris representation is the pixel-intensity data for the normalized iris image. In this representation, two normalized images are analyzed – with and without zeroed occlusion – as shown in the Fig. 4. Each row of the normalized images forms a vector which is later treated as a single-period sample of a one-dimensional periodic signal. A 1-D Gaussian wavelet transform (Daubechies, 1992) is applied to each row vector and decomposed into different resolution levels. The zero-crossing representation is then calculated for each row and resolution level. Zero crossings

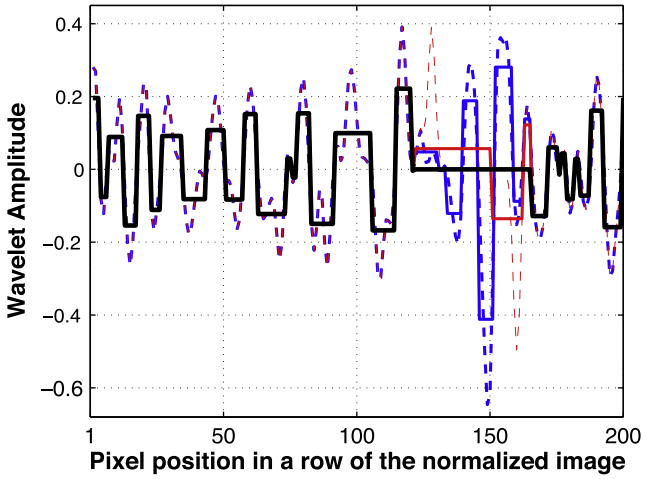


Fig. 5. Wavelet and zero-crossing representations.

occur where wavelet signals have abrupt changes in signal amplitude. Once the zero-crossings have been located, the average value between each two consecutive zero-crossing points in the wavelet output is computed.

Illustrated in Fig. 5 are wavelets for one resolution level (blue[2] and red dashed lines) and the respective zero-crossing representation (blue and red solid lines) from the same row of both the normalized iris (blue) and zeroed image (red), as well as the final zero-crossing representation (black line). As shown here, these wavelets (and their respective zero-crossing representations) differ at the regions where occlusion has been identified (the zeroed image). The final representation (black solid line in Fig. 5) is produced by starting from the zero-crossing representation (solid blue line) of the normalized iris image (Fig. 4(a)) and zeroing where it differs from the image (red solid line) for which occlusion was considered (Fig. 4(c)).

The values from the black solid line used in the iris representation were extracted from a 200 × 16 pixels normalized iris image, decomposed into three resolution levels (2, 3 and 4) for each row, resulting in a matrix of 48 rows by 200 columns.

2.3.2. 2-D dyadic wavelet zero-crossing representation

The earlier representation method used a 1-D Gaussian wavelet transform for each row of the normalized iris image. Here, a 2-D Daubechies dyadic wavelet transform (Daubechies, 1992) is applied instead.

To extract features from the normalized iris image, it is first convolved with a 2-D dyadic wavelet low-pass filter, minus the estimated value of both normalized iris images (Figs. 4(a) and (c)).

The resulting information matrices are then processed using the same technique for each row as detailed in Section 2.3.1, to obtain the final zero-crossing representation (Fig. 5 – black line) and represent the iris in a 200 × 16 matrix.

2.3.3. Periocular

New trends in biometrics (Park et al., 2009; Woodard et al., 2010) suggest the use of periocular information as an important addition in noncooperative biometric recognition, as information derived from this area is less prone to degradation in visible wavelengths than other traits (e.g., the iris). Representing a tradeoff between facial and iris recognition techniques, this method has the advantage of not requiring any additional equipment, as usually

[2] For interpretation of color in Figs. 4 and 5, the reader is referred to the web version of this article.

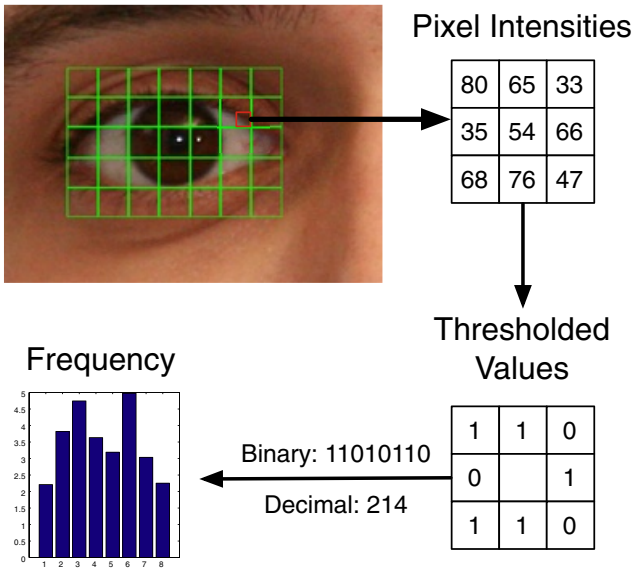


Fig. 6. Steps for LBP feature extraction.

such information is not discarded in iris databases. This technique is also less vulnerable to problems resulting from a lack of proper illumination or low-resolution acquisition, motion blur and varying imaging distances.

For the purpose of this work, we used the simple yet effective analysis suggested by Park et al. (2009).

Distribution-based descriptors. The iris location and size being known, images were aligned and normalized for both scale and translation as a set of regions of interest (Fig. 6) were defined according to those parameters. As shown here, iris size is proportional to the sides of each square region, and the central one is concentric with the iris itself.

Local Binary Pattern Pietikainen (2005) descriptors were then extracted, as depicted in Fig. 6. Using pixel intensities in a square window iterated over the entire region of interest, the difference between the central pixel and its eight neighbors was computed and its signal used to produce a binary result (thresholded values). Converting those results to decimal, values from each region were then quantized into eight-bin histograms, which upon concatenation produce the complete 280-feature array (35 regions $\times$ 8 bins per region).

Scale-Invariant Feature Transform. Differing from the previous method, where features were only extracted from the region closest to the eye, the Scale-Invariant Feature Transform (SIFT) (Lowe, 2004) was applied to all available data, here seeking salient regions (e.g., facial marks). SIFT is one of the most popular descriptors for image point matching, as it can achieve invariance to scale and rotation and is also robust to affine distortion. The method is based on the extraction of key points represented by vectors containing scale, orientation and location information. To achieve those results, a publicly available SIFT implementation[3] was used, and its parameters optimized based on tests performed on the training dataset.

2.3.4. Comparison maps

This approach (Santos and Proença, 2010) can be regarded as an extension to the widely known Daugman method (Daugman, 2004), which is the most widely acknowledged, with great acceptance over the scientific community.

This method begins with the detection and segmentation of the iris. For our approach, we used the procedures detailed above for the iris-boundary detection and normalization, except for the normalized iris sizes, which were 450×64 pixels for both the iris (Fig. 4(a)) and the noise mask (Fig. 4(b)).

Later, features were extracted through the convolution of the normalized data with a bank of 2-D Gabor wavelets, followed by a quantization stage that produced a binary *iriscode*, in which every complex-valued bit $h_{\{Re,Im\}}$ depends on the sign of the 2-D integral.

We decided on the use of a very small yet optimized wavelet bank, for which performance was optimized using the training data. For such optimization, we parameterized the wavelets cycling through a range of scales, orientations and frequencies we found fit, searching for the configuration that maximized the decidability (13).

2.4. Matching

In this section, the matching process is described for each one of the feature-extraction methods.

2.4.1. 1-D and 2-D wavelet zero-crossing representation

To compute the dissimilarity between two irises, their zero-crossing representations are compared. Boles (1997) proposed four functions to measure the dissimilarity between the signals. In this work, we used the dissimilarity measure defined by Eq. (3).

$$d_{lm}(f,g) = 1 - \frac{\sum_{e=0}^{E-1} Z_l f(e) \cdot Z_l g(e+m)}{\|Z_l f\| \|Z_l g\|} \tag{3}$$

In the above equation, $d_{lm}(f,g)$ denotes the dissimilarity of irises f and g associated with the lth row of their representation matrices for a displacement m, the vectors $Z_l f$ and $Z_l g$ are the lth row of the zero-crossing representations of irises f and g, respectively, E is the number of elements of $Z_l f$ and $Z_l g$ and m, $e \in [0, E-1]$. The symbol $\|\cdot\|$ denotes the vector-norm operation. Note that $d_{lm}(f,g)$ is equal to 1 minus the correlation coefficient between $Z_l f(e)$ and $Z_l g(e)$. Thus the dissimilarity $d_{lm}(f,g)$ may take values between 0 and 2, whereby 0 corresponds to a perfect match.

Eq. (3) is computed for each row of the representation matrices and determine which mean is taken as the dissimilarity (D_m) between irises f and g for a given value of m.

This work proposes the use of a weighted mean rather than a simple mean, whereby the weights are given by the number of nonzeroed values in $Z_l f(n)$ and $Z_l g(n)$ according to:

$$D_m = \frac{\sum_{l=1}^{L} d_{lm}(f,g) \times K_l}{\sum_{l=1}^{L} K_l} \tag{4}$$

where $d_{lm}(f,g)$ is given by (3) and K_l is the number of nonzeroed values in the lth row of the zero-crossing representations of both images.

It is important to notice that m in Eq. (3) represents the shifts of the second signal. Varying m in (3) from 0 to $E-1$ yields E dissimilarity values (D_m). The overall dissimilarity D between irises f and g is given by:

$$D = \min(D_m) \tag{5}$$

2.4.2. Periocular

From periocular analysis, two types of results were produced.

To compute the matching between two feature vectors u and v with n elements produced by the distribution-based descriptor, we used a Euclidean distance (6):

[3] VLFeat open-source library http://www.vlfeat.org/.

(a) Iriscode match sample

(b) Iriscode match with high concordance (de limited region)

Fig. 7. Illustration of two iriscode matching results. Black pixels express concordant bits in the correspondent biometric signatures.

$$d(u,v) = \sqrt{\sum_{i=1}^{n}(u_i - v_i)^2} \tag{6}$$

As for the features extracted by the SIFT, the distance-ratio-based matching scheme (Lowe, 2004) was applied.

2.4.3. Comparison maps

With two binary codes (*codeA* and *codeB*) and the corresponding segmentation masks (*maskA* and *maskB*), the *Hamming* distance (8) is applied as comparison measure.

$$c = (codeA \otimes codeB) \cap maskA \cap maskB \tag{7}$$

$$HD = \frac{\|c\|}{\|maskA \cap maskB\|} \tag{8}$$

where $\otimes$ is the logical XOR operation, $\cap$ is the logical AND and c is the "comparison map".

Instead of using the Hamming Distance alone (which is simply the ratio of concordant iriscode bits) as a single comparison measure, the resulting "comparison maps" (Fig. 7(a) and Eq. (7)) from the similarity between iriscodes are then subjected to both spatial and frequency-domain analysis in a search for high-concordance areas (Fig. 7(b)).

Spatial-domain analysis. For the spatial-domain analysis, we proceeded with a set of convolutions with Haar-based wavelets of different sizes, which allowed us to ascertain the concordance level of regions with different sizes.

Let c be a *comparison map* of $M \times N$ dimensions. Let h be a Haar-based mother wavelet with size $s \times s$. The similarity r in local regions of c is given by:

$$r_s = h_s * c, \quad s = \{2k\}, \quad k = 2,3,\ldots,16 \tag{9}$$

where $*$ denotes the bidimensional convolution and r_s has the same dimensions of c.

Let $\omega_s = \max\{r_s(i,j)\}$, $i = 1, 2, \ldots, N$; $j = 1, 2, \ldots, M$.

Let H be the 25-bin histogram of $r_{\frac{\psi}{2}}$ where ψ is the maximum size of the Haar-wavelet, such that $H = \{h_1, h_2, \ldots, h_{25}\}$.

Using ω_i and h_i values, features were produced and used as detailed in the Classification subsection.

Frequency-domain analysis. For the frequency-domain analysis, the Fourier transform F of the comparison map c of $M \times N$ dimensions was computed as follows:

$$F(u,v) = \frac{1}{N}\sum_{x=0}^{M}\sum_{y=0}^{N} c(x,y)e^{-j2\pi(ux/M+vy/N)} \tag{10}$$

where j is the square root of -1 and e denotes the natural exponent.

The results were then regularly windowed in sixteen subregions, and statistical features were extracted from each region. For the central part, where the most relevant information lies, we considered a $P \times N$ window centered in the $P \times M$ matrix that contains the noticeable central shape such that $P = 2M/8$. Ten features F_i are then extracted, representing the distribution of an evenly spaced ten-bin histogram:

$$T_i = \min(A) + i\frac{\Delta A}{10} \tag{11a}$$

$$F_i = \sum_{m=1}^{P}\sum_{n=1}^{N} sgn(A_{(m,n)} - T_i) \tag{11b}$$

with $\Delta A = \max(A) - \min(A)$ and $i = 1, 2, \ldots, 10$.

Classification. Combining the best features (according to their individual decidability) and performing a dimensionality reduction through Local Fisher Discriminant Analysis (Sugiyama, 2006), a logistic regression (Hosmer and Lemeshow, 2000) was used to describe the function that eventually produced the final result for this method.

2.5. Decision ensemble

With several outputs coming from the different representation methodologies, a logistic regression model (Agresti, 2002; Cantor, 2002; Hosmer and Lemeshow, 2000) was used to describe the relationship between them and a final response. This weight fitting methodology efficiency was verified on identical situations, with multiple classifiers of different accuracies (Monwar and Gavrilova, 2008, 2009; Santos and Proença, 2010).

The way this logistic regression works is equivalent to a single-output neural network with a logistic-activation function trained under log loss; this model is described by Eq. (12):

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_5 x_5 \tag{12}$$

where the fraction $p/(1-p)$ is called *the odds* of a positive match, that is, the ratio between that probability and its complementary. The β_i value is the weight relating the outputs x_i from the previously described methods to the odds.

3. Analysis of results

To assess the performance of the proposed method, experiments were conducted using 1,000 iris images from the UBIRIS.v2 (Proença et al., 2010) database used for the NICE.II[4] contest, and their respective segmentation masks. Although this contest was based only on identification mode (performance was ranked through the decidability measure), our experiments were carried out in two modes: verification mode (one-to-one matching) and identification mode (one-to-many matching).

In verification mode, we selected the well-known receiver-operating characteristic curves (ROC), the area under curve (AUC), the equal-error rate (EER) and the decidability (Daugman and Williams, 1996) index, given by Eq. (13):

$$d' = \frac{|\mu_{inter} - \mu_{intra}|}{\sqrt{\frac{\sigma^2_{inter} + \sigma^2_{intra}}{2}}} \tag{13}$$

where μ_{inter} and μ_{intra} denotes the means of the interclass and intra-class comparisons and σ_{inter} and σ_{intra} are the respective standard deviations.

The ROC curve is a graphical plot of the sensitivity, or true positive rate vs. false positive rate. The AUC can be perceived as a measure based on pairwise comparisons between classifications of two classes. With a perfect ranking, all positive examples are ranked higher than the negative ones and the area equal to 1. Any deviation from this ranking decreases the AUC. The EER of a

[4] NICE.II - http://www.nice2.di.ubi.pt.

Table 1
Recognition rates of each test.

	DEC	EER (%)	AUC
LBP	0.99	31.87	0.76
SIFT	0.87	32.09	0.74
1-D wavelet	1.44	23.12	0.85
2-D wavelet	1.29	25.04	0.82
Comparison maps	1.27	24.99	0.82
Fusion	1.74	18.48	0.90

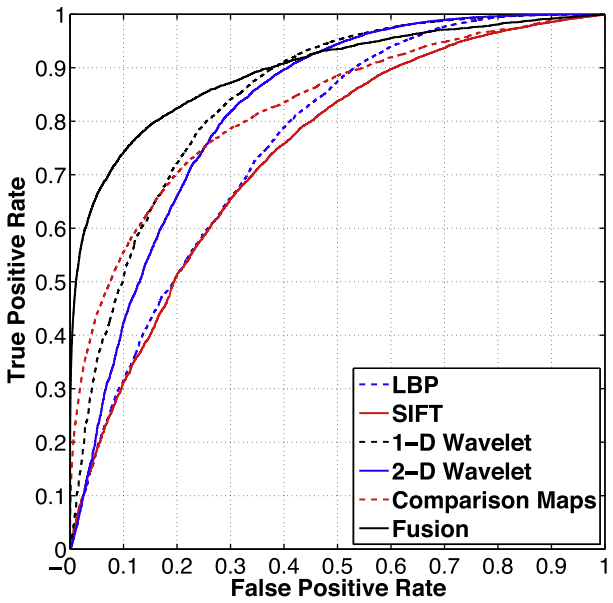


Fig. 8. ROC curves for all matchers and their fusion.

verification system means that the operating threshold for the accept/reject decision is adjusted so that the probability of false acceptance and false rejection becomes equal.

In identification mode, where a subject is matched against a database in a 1:*N* way, a good performance assessment is the Cumulative Match Characteristic (CMC), as it shows the identification probability against the *N* closest candidates.

The parameters specified in the method description were tuned for best performance; here we chose those with maximal decidability indices, *i.e.*, those that maximize the average distance between distributions obtained for the two classical types of biometric comparisons: data extracted from the same (*intraclass*) and different eyes (*interclass*).

When applying the described methods independently on the training dataset, we obtained the results presented in Table 1 and Figs. 8 and 9.

As shown in Table 1, with respect to decidability (which was the criterion under consideration for the NICE.II contest), the best individual results were with the 1-D and 2-D Wavelet methods, with decidability (DEC) values of 1.44 and 1.29, respectively, closely followed by comparison maps at 1.27. The same observation is valid for the AUC values, whereas for EER the comparison maps slightly outperform the 2-D wavelet. Periocular features, despite low individual performance, proved to be of great help when fused with the other methodologies. In fact, inspection of the CMC plot (Fig. 9), where the separability between intra- and interclass distributions is not as pronounced, shows that LBP is the best of all individual methods, with a 56.4% rank-1 cumulative accuracy *versus* the 41.9% of the 1-D wavelet, beaten only by the fusion, with a 74.3% rank-1 cumulative accuracy.

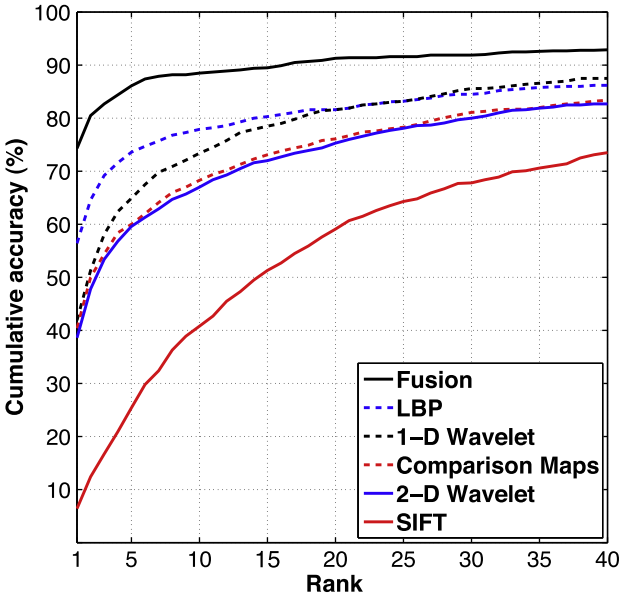


Fig. 9. CMC curves for all matchers and their fusion.

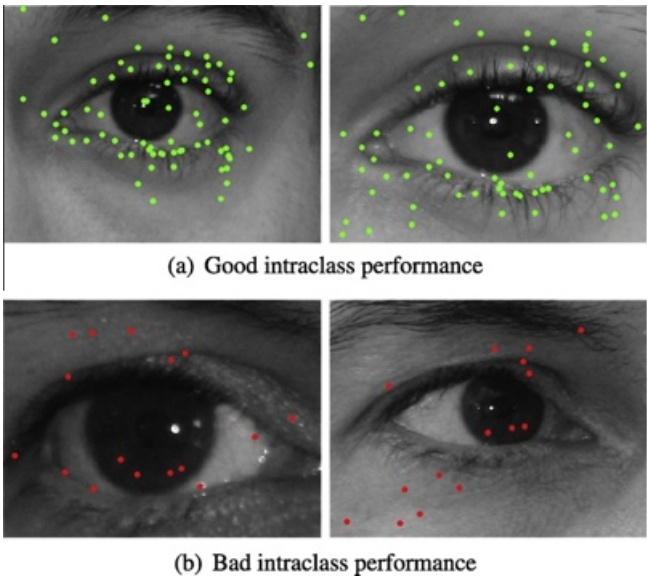


Fig. 10. SIFT performance examples in intraclass comparisons.

Fusing all the methods enhanced decidability to 1.74, representing an improvement of 20.8% over the best individual method. Improvements in identification performance were even more significant, as rank-1 was raised to 31.7%.

From these results, we can see that the entire method performance cannot be accessed by a single operational mode. We thus infer that, although some approaches improve recognition capabilities in verification scenarios, and some others work well for identification mode, their fusion produces more suitable outcomes, demonstrating the effectiveness of our method in both cases.

As the SIFT method uses more area for feature extraction than the others, it is more likely to be affected by strong variations in imaging conditions (*e.g.* pose or illumination; see Fig. 10(b)), thus producing globally unsatisfactory results. However, its good performance in some particular cases (*e.g.*, Fig. 10(a)) led us to include it, as its use improved the overall fused decidability by 4.5%.

4. Conclusions

In this study, we presented a novel fusion of different recognition approaches to address the issue of noncooperative iris recognition using nonideal visible-wavelength images captured in an unconstrained environment.

We tested several different autonomous approaches; their individual performances were evaluated in identification and verification modes and then the methods were fused, resulting in improved accuracy. We also showed that combining features extracted from the iris region itself with periocular information improves the overall performance in both recognition modalities.

The robustness of our approach was corroborated by independent evaluation in the NICE.II iris-recognition contest, where our method placed third rank among almost seventy participants from all over the world.

Acknowledgments

We acknowledge the financial support provided by "FCT-Fundação para a Ciência e Tecnologia" and "FEDER" in the scope of the PTDC/EIA/69106/2006 "BIOREC: Non-Cooperative Biometric Recognition" and PTDC/EIA-EIA/103945/2008 "NECOVID: Negative Covert Biometric Identification" research projects.

References

Agresti, A., 2002. Categorical Data Analysis (Wiley Series in Probability and Statistics). Wiley Interscience, Hoboken, NJ.

Ballard, D., 1981. Generalizing the Hough transform to detect arbitrary shapes. Pattern Recognition 13 (2), 111–122.

Boles, W., 1997. A security system based on human iris identification using wavelet transform. In: KES'97: Proc. First Internat. Conf. on Knowledge-Based Intelligent Electronic Systems, vol. 2, pp. 533–541.

Boles, W., Boashash, B., 1998. A human identification technique using images of the iris and wavelet transform. IEEE Trans. Signal Process. 46 (4), 1185–1188.

Canny, J., 1986. A computational approach to edge detection. IEEE Trans. Pattern Anal. Machine Intell. PAMI-8 (6), 679–698.

Cantor, A.B.M., 2002. Understanding logistic regression. Evidence-based Oncology 3 (2), 52–53.

Daubechies, I., 1992. Ten Lectures on Wavelets. SIAM.

Daugman, J.G., 2004. How iris recognition works. IEEE Trans. Circuits Systems Video Technol. 14 (1), 21–30.

Daugman, J., Williams, G., 1996. A proposed standard for biometric decidability. In: Proceedings of the CardTech/SecureTech Conf., pp. 223–234.

Hosmer, D.W., Lemeshow, S., 2000. Applied logistic regression (Wiley Series in probability and statistics). Wiley-Interscience Publication.

Hoyle, E., Feitosa, R., Petraglia, A., 2009. Iris recognition using one-dimensional signal analysis. In: Proc. 8th Internat. Seminar on Electrical Metrology.

Hoyle, E., Feitosa, R., Petraglia, A., 2010. VIII Semetro Book, EDUFCG: Editora da Universidade Federal de Campina Grande, 2010, pp. 141–161 (Chapter. 8: Robust Iris Segmentation for Biometric Recognition).

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Internat. J. Comput. Vision 60, 91–110.

Miller, P.E., Rawls, A.W., Pundlik, S.J., Woodard, D.L., 2010. Personal identification using periocular skin texture. In: SAC'10: Proc. 2010 ACM Symp. on Applied Computing. ACM, New York, NY, USA, pp. 1496–1500.

Monwar, M., Gavrilova, M., 2008. FES: A system for combining face, ear and signature biometrics using rank level fusion. In: ITNG 2008: Fifth Internat. Conf. on Information Technology: New Generations, pp. 922–927.

Monwar, M., Gavrilova, M., 2009. Multimodal biometric system using rank-level fusion approach. IEEE Trans. Systems Man Cybernet. Part B: Cybernetics 39 (4), 867–878.

Park, U., Ross, A., Jain, A., 2009. Periocular biometrics in the visible spectrum: A feasibility study. In: BTAS'09: Proc. IEEE 3rd Internat. Conf. on Biometrics: Theory, Applications, and Systems, 2009, pp. 1–6.

Pietikainen, M., 2005. Image analysis with local binary patterns. In: Kalviainen, H., Parkkinen, J., Kaarna, A. (Eds.), Image Analysis, Lecture Notes in Computer Science, vol. 3540. Springer, Berlin/Heidelberg, pp. 115–118.

Proença, H., Filipe, S., Santos, R., Oliveira, J., Alexandre, L.A., 2010. The ubiris.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. IEEE Trans. Pattern Anal. Machine Intell. 32, 1529–1535.

Santos, G., Proença, H., 2010. Iris recognition: Analyzing the distribution of iriscodes concordant bits. In: CISP 2010: Proc. 3rd Internat. Congress on Image and Signal Processing, vol. 4, pp. 1873–1877.

Savvides, M., Ricanek Jr., K., Woodard, D.L., Dozier, G., 2010. Unconstrained biometric identification: Emerging technologies. Computer 43, 56–62.

Sugiyama, M., 2006. Local fisher discriminant analysis for supervised dimensionality reduction. In: ICML'06: Proc. 23rd Internat. Conf. on Machine Learning. ACM Press, New York, NY, USA, pp. 905–912.

Tan, T., He, Z., Sun, Z., 2010. Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. Image Vision Comput. 28 (2), 223–230.

Woodard, D., Pundlik, S., Miller, P., Jillela, R., Ross, A., 2010. On the fusion of periocular and iris biometrics in non-ideal imagery. In: ICPR 2010: Proc. 20th Internat. Conf. on Pattern Recognition, pp. 201–204.

Chapter 10

Fusing Iris and Periocular Information for Cross-sensor Recognition

10.1 Overview

This chapter consists of the following article:

Fusing Iris and Periocular Information for Cross-sensor Recognition
Gil Santos, Emanuel Grancho, Marco V. Bernardo and Paulo T. Fiadeiro
Pattern Recognition Letters, accepted for publication subject to minor revision, 2014.

According to SCImago Journal & Country Rank, this journal's index[1] for the 2013 year are as follows:

Category	*Quartile*	*SJR*
Computer Vision and Pattern Recognition	Q1	0,995
Artificial Intelligence	Q2	
Signal Processing	Q1	
Software	Q1	

[1]The SCImago Journal & Country Rank (SJR) indicator is a measure of journal's impact, influence or prestige. It expresses the average number of weighted citations received in the selected year by the documents published in the journal in the three previous years. `http://www.scimagojr.com`

Pattern Recognition Letters
journal homepage: www.elsevier.com

Fusing iris and periocular information for cross-sensor recognition

Gil Santos[a,**], Emanuel Grancho[a], Marco V. Bernardo[a,b], Paulo T. Fiadeiro[b]

University of Beira Interior, Covilhã, Portugal

[a] *Department of Computer Science, IT - Instituto de Telecomunicações*
[b] *Department of Physics, Remote Sensing Unit - Optics, Optometry and Vision Sciences Group*

ABSTRACT

Over the last years the usage of mobile devices has substantially grown, along with their capabilities and applications. Extending biometric technologies to such gadgets is quite desirable, as it would represent the ability to perform biometric recognition virtually anytime, anywhere, and by everyone. This paper focus on biometric recognition on mobile environments using the iris and periocular information as main traits, and its main contributions are three-fold: 1) announce the availability of an iris and periocular dataset containing images acquired with 10 different mobile setups, along with the corresponding iris segmentation data. Such dataset allows to evaluate both iris segmentation and recognition methods, as well as periocular recognition techniques; 2) report the outcomes of device-specific calibration techniques that compensate for the different color perception inherent to each setup; 3) propose the application of well-known iris and periocular recognition strategies, based on classical encoding and matching techniques, giving evidence on how they can be fused to overcome the issues associated with mobile environments.

1. Introduction

The evolution of biometric systems over the last years is notorious, with the appearance of new traits and algorithms and the refinement of the existing ones. At the same time that the acquisition constraints are being lowered favoring *in-the-wild* operation, efforts are being put into delivering *off-the-shelf* solutions for everyday consumers, so that biometric systems can run easily on everyday electronics. Mobile devices in particular are preferable targets, as they comprise all the necessary components to carry the whole process, from trait acquisition to the final decision.

From the existing traits, the face and the iris are present in the literature among the most popular (along with the fingerprint) (Bowyer et al., 2008; Zhao et al., 2003). Iris usage as main biometric trait has remained stable despite the evolution of biometrics in the last years. Being a naturally protected organ, visible from the exterior and allowing contact-less acquisition, its circular and planar shape that favors detection and segmentation, and its predominantly randotypic appearance that assures high recognition effectiveness. There are, however, certain scenarios where the iris cannot be properly imaged, and where the complementary use of other ocular information is regarded as a good way to compensate for unreliable iris acquisition – periocular biometrics.

Particular useful on unconstrained scenarios, the periocular region does not require constrained capturing or complex imaging systems, being fairly easy for a mobile user to operate a periocular identification application. The grounds for periocular recognition came from human intrinsic ability to recognize someone just by looking at his/her eyes, which are known to provide substantial amounts of discriminant information whilst remaining relatively stable over large periods of time. Periocular biometrics analyze not only iris structure, but also other surrounding features, such as the shape of eyelids, eyelash distributions, or sclera and skin texture information. At last, both the iris and the periocular region are imaged simultaneously with a single camera.

1.1. Contextualization: Iris biometrics

The commercially deployed iris recognition systems are mainly based on Daugman (1993) pioneering approach, with great effectiveness in relative constrained scenarios, and with data acquired in the near-infrared (NIR) slice of the electromagnetic spectrum (700-900 nm). Even that a few innovations were introduced later on (Daugman, 2007), the process consists in a three stage approach: 1) the segmentation of the iris

** Corresponding author: Tel.: +351-275-242081; Fax : +351-275-319899;
e-mail: gmelfe@ubi.pt (Gil Santos)

boundaries (both pupillary and limbic) followed by the translation into a double dimensionless pseudo-polar coordinate system to achieve invariance to scale and translation; 2) the convolution of this normalized data with a set of Gabor filters at multiple frequencies and orientations and the corresponding output quantized to one of four quadrants, extracting two bits of phase information per convolution; 3) matching of the iris signatures using the fractional Hamming distance, with several comparisons of shifted data to achieve invariance to rotation.

In addition to Daugman's phase-based approach other iris recognition variants were introduced, mainly zero-crossing and texture-analysis methods: Boles and Boashash (1998) computed the zero-crossing representation of a 1D wavelet at different resolutions of concentric circles, and Wildes (1997) proposed the characterization of the iris texture through a Laplacian pyramid with four different levels.

Efforts on "relaxing" the acquisition setup are also registered, being the "iris-on-the-move" project (Matey et al., 2006) a major example on engineering a less intrusive system for subjects: its goal is to acquire near-field NIR iris images as the subjects walk through an access control point.

1.2. Contextualization: Periocular biometrics

The usage of the periocular region as a biometric trait has emerged over the last years (Santos and Proença, 2013). The first relevant studies on periocular biometrics can be traced back to Park et al. (2009) and their pioneering approach, fusing both local and global features from the ocular area. On global feature extraction images were aligned using iris center as anchoring point, and a 7×5 region of interest (ROI) grid defined around it. Scale invariance was achieved using iris radius as side length for the ROI. Those patches were then encoded applying two well known distribution-based descriptors, Local Binary Patterns (LBP) (Ojala et al., 1994) and Histogram of Oriented Gradients (HOG) (Dalal and Triggs, 2005), quantized into 8-bin histograms. Merging those histograms into a single-dimension array containing both texture and shape information, matching was carried off simply by computing an Euclidean distance. For the local analysis, authors employed Scale-Invariant Feature Transform (SIFT) (Lowe, 2004), allowing sets of key-points to be extracted, encoded with their surroundings, and matched, while providing translation, scaling and rotation invariance. Reported performance was fairly good, showing periocular fitness for recognition purposes, and further analysis was held on noise factors impact on performance (Park et al., 2011).

Inspiring by their work other approaches arose, either by improving Park *et al.* approach, or by introducing new perspectives. Miller et al. (2010) presented an analysis also focused on periocular skin texture, taking advantage of Uniform Local Binary Patterns (ULBP) to achieve "improved rotation invariance with uniform patterns and finer quantization of the angular space" (Ojala et al., 2002). Later on, their work was extended by Adams et al. (2010), who proposed using Genetic & Evolutionary Computing (GEC) to optimize feature set. Juefei-Xu et al. (2010) stressed many local and global feature extraction techniques (Walsh (Beer, 1981) and Laws' masks (Laws, 1980), Discrete Cosine Transform (DCT) (Ahmed et al., 1974), Discrete Wavelet Transform (DWT) (Mallat, 1989), Force Fields (Hurley et al., 2000), Speed-Up Robust Transform (SURF) (Bay et al., 2008), Gabor filters (Clausi and Jernigan, 1996) and Laplacian of Gaussian (LoG)), and on their later work (Juefei-Xu et al., 2011) efforts were made to compensate aging degradation effects on periocular performance. The possibility of score-level fusion with other biometric traits was also addressed (e.g. iris (Woodard et al., 2010)).

Bharadwaj et al. (2010) proposed the fusion of ULBP with five perceptual dimensions, usually applied as scene descriptors: naturallness, openness, roughness, expansion and ruggedness – GIST (Oliva and Torralba, 2001). Images were preprocessed with Fourier transform for local contrast normalization, and then a spatial envelope computed with a set of Gabor filters (4 scales $\times$ 8 orientations). On the final stage, χ^2 distance was used to match the feature arrays, and results fused with a weighted sum. This approach was validated against UBIRIS.v2 data (Proença et al., 2010), simulating realistic unconstrained acquisition setups.

1.3. The Mobile Constraints

When attempting to perform iris or periocular biometrics on mobile environments, several problems arise: the wide variety of camera sensors and lenses mobile phones and tablets come equipped with produce discrepancies in working images, as they are acquired with color distortions, at multiple resolutions, etc.; *on-the-go* acquisition by potentially untrained subjects will result in demanding Pose, Illumination and Expression (PIE) changes, as not all users hold their mobile devices at the same position, resulting in varying acquisition angles and scales, or rotated images; the acquisition environment can have poor or insufficient lighting, and uncontrolled outdoor daylight will most likely produce spectacle reflections over the iris region; etc.

The remainder of this paper is organized as follows: Section 2 describes the Cross-Sensor Iris and Periocular Dataset (CSIP) database, detailing the acquisition conditions, enrolled participants and perceived noise factors; Section 3 presents the proposed methodology, with details on the four main stages: image normalization with device-specific color calibration, iris and periocular feature encoding and matching, and score-level fusion; Section 4 contains a thorough analysis of the results obtained by using the proposed methodology; finally, Section 5 states some final considerations, along with further lines of work.

2. The Cross-Sensor Iris and Periocular Dataset

The main objective of the CSIP database was to gather images from a representative group of participants, acquired over cross-sensor setups and varying acquisition scenarios, thus mimicking the conditions faced on mobile application scenarios. Along with the data acquired with different mobile devices, an iris segmentation mask is also provided, allowing assessing the performance of both iris and periocular segmentation and recognition algorithms on mobile environments.

Table 1: Details of the devices and setups used during the CSIP dataset acquisition.

Device	A		B			C			D	
Manufacturer	Sony Ericsson		Apple			ThL			Huawei	
Model	Xperia Arc S		iPhone 4			W200			U8510	
O.S.	Android 2.3.4		iOS 7.1			Android 4.2.1			Android 4.3.3	
Camera	Rear		Frontal	Rear		Frontal	Rear		Frontal	Rear
Resolution	3264×2448		640×480	2592×1936		2592×1920	3264×2448		640×480	2048×1536
Flash	No	Yes	No	No	Yes	No	No	Yes	No	No
Setup ID	AR0	AR1	BF0	BR0	BR1	CF0	CR0	CR1	DF0	DR0

2.1. The Imaging Setup

Considering the heterogeneity of camera sensor/lens setups consumer mobile devices can deliver, a total of 10 different setups were used during the dataset acquisition stage: four different devices, some of them with frontal and rear cameras and LED flash (Table 1). Each participant was imaged at all the considered setups.

Aiming at mimicking the variability of noise factors associated with *on-the-go* recognition, participants were not imaged at a single particular location, but on multiple sites, *as they were*, with artificial, natural and mixed illumination conditions. As we can see from Figure 1, there is a substantial difference between each acquisition setup and surrounding conditions, even when the same setup was used to capture images from different subjects. From visual inspection, eight different noise factors are distinguishable, and can affect the biometric recognition process: multiple scales; chromatic distortions; image rotation; poor lighting; off-angle acquisition; out-of-focus images; deviated gaze; and iris obstructions (including reflexions).

The images were acquired through the standard camera application on mobile phone devices, using default settings for both focus and white-balance. The corresponding files were stored at JPEG format, with the highest possible quality and resolution. A total of 50 participants were enrolled, all Caucasian and most of them male (82%), with ages comprehended between 21 and 62 years old (31.18 ± 9.93). All the participants gave informed consent about the experiment.

2.2. Iris segmentation masks

For each periocular image acquired by the mobile devices, a binary iris segmentation mask is provided with the CSIP dataset. Those masks were automatically obtained using the state-of-the-art iris segmentation approach proposed by Tan et al. (2010). That approach is particularly suitable for uncontrolled acquisition conditions, which has been corroborated by the first place achieved at the Noisy Iris Challenge Evaluation - Part 1 (NICE.I)[1].

At a first stage, a small ROI containing a rough estimate of the iris location is defined. This ROI is determined using a cascade object detector based on Viola and Jones (2001) algorithm, trained for the detection of the right eye using Haar features to encode details (Castrillón et al., 2007). A reflexion removal process is applied, followed by an eight-neighbor connection approach for clustering, and based on the degree of similarity between each pixel and the previously established heterogeneous regions, a region set is established accordingly to the degree of similarity between their elements. In order to label the different clusters, semantic refinements are applied. Several semantic priors like orientation and shape of each region are used to determine the iris correspondent cluster.

Further to that, iris pupillary and limbic boundaries are estimated using an integrodifferential-constellation: based on Daugman (2007) integrodifferential, a constellation is built from several integrodifferential rings of increasing radii, minimizing the initial method's tendency to output local optimal

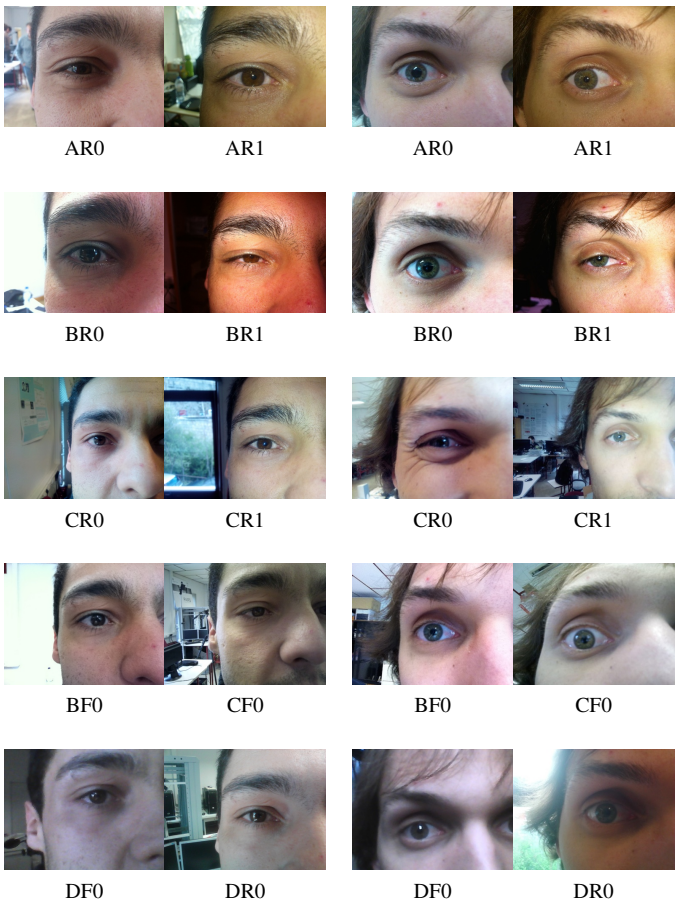


Figure 1: Dataset pictures acquired from two participant at all different setups. Images in the left belong to the first participant, and images in the right belong to the second participant.

[1] http://nice1.di.ubi.pt/

solutions. Possible localization inaccuracies are detected and eliminated based on a threshold, estimated by the intersection of two consecutive annular rings intensity distributions. After iris boundaries are established, eyelid localization and shadow subtraction are performed in order to reduce noise and occlusion. The presence of eyelashes is minimized through 1-D filtering, and edge-detection is applied to find the edge points corresponding to eyelids. Using those edge-points, the localization of the eyelids is estimated using both an upper and lower statistically established curvature model. Ultimately, eyelash and shadow subtraction take into account their darker appearance when compared to the iris itself. The optimization of the classification threshold is obtained from the analysis of the intensity histogram of small homogeneous regions, on iris and shadows noise, and eyelash regions are removed. This technique is explained in more detail at Tan et al. (2010).

2.3. Dataset Availability

The complete CSIP dataset is public and freely available for academic and research purposes[2]. Researchers are granted access to: 1) 2004 images, acquired from 50 subjects at 10 different setups; and 2) the corresponding 2004 binary iris segmentation masks.

3. Proposed Methodology

In this section we describe the four main steps of our approach (Figure 2): the normalization stage, with device-specific color correction and iris boundaries estimation for coordinate conversion and periocular ROI definition; feature encoding, with information from both the iris and the periocular region; feature matching; and score-level fusion.

3.1. Image Normalization

The first stage, image normalization, will allow to compensate for some of the noise factors identified in the dataset: chromatic distortions, varying scales and off-angle acquisition.

3.1.1. Device-Specific Color Correction

Having an uniform calibrated output for each sensor that minimizes the discrepancy to colors as they really appear can be of particular value in mobile scenarios, as a wide range of sensor/lens setups are available.

The access to a reference image captured at a known illuminant allows to estimate the color adaptation matrix that compensate for the inaccurate color representation introduced by each sensor. That adaptation matrix encodes the optimized color channels combination to approximate color information in the acquired image from the ones originally observed in the scene. This section describes the device-specific color correction technique.

A Macbeth ColorChecker® Color Rendition Chart was placed in a dark acquisition scene, illuminated by a standard illuminant produced by a *Barco RLM G5i Performer* (Barco Corporation, Belgium) RGB projector driven by a *Visual Stimulus Generator (VSG2/5)* (Cambridge Research Systems, United Kingdom). In order to mimic standard open-air conditions, the Commission Internationale de l'Eclairage (CIE) D65 illuminant was chosen, as specified by the CIE standard colorimetric observer (2°) (on Illumination, 2004; Smith and Guild, 1931). Illuminants' luminance was regulated at 100 cd/m^2.

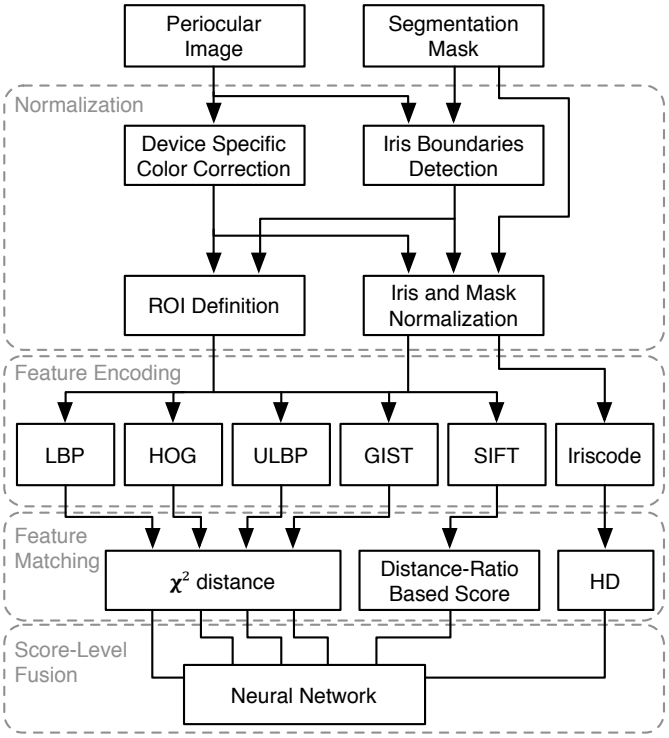


Figure 2: Diagram illustrating the four stages of the proposed methodology.

Previously, the VSG2/5 generated stimulus were verified and calibrated using a telespectroradiometer (PR-650 *SpectraColorimeter*™- Photo Research, Inc., CA) and a white reference Spectralon® target (Labsphere, Inc., NH). The maximum errors allowed were 0,002 illuminant chromaticities in the CIE 1931 color space and 1 cd/m^2 for luminance. A set of images of the color charts was then captured at all setups (mobile devices), using the standard camera application at default settings.

To obtain the estimate for the color correction matrix we applied the methodology introduced by Wolf (2003), specially designed for digital imaging systems. Knowing the ground-truth red (R), green (G) and blue (B) coordinates for the 24 color samples from the color chart under the D65 illuminant, let us summarize it in a 24×3 matrix O (1). Then, from the color chart photo acquired with the mobile device, we populate a similar matrix P with the RGB coordinates for the same 24 color samples.

$$O = \begin{bmatrix} O\_R_1 & O\_G_1 & O\_B_1 \\ O\_R_2 & O\_G_2 & O\_B_2 \\ \dots & \dots & \dots \\ O\_R_{24} & O\_G_{24} & O\_B_{24} \end{bmatrix} \quad (1)$$

The initial estimate for the adaptation matrix A, that converts the device acquired colors to an approximation $\hat{O}$ of the original

[2] http://csip.di.ubi.pt

ones, was then found using a least-squares solution (2) where $\underline{1}$ is a 24 positions column vector initialized with ones.

$$\begin{aligned} O \approx \hat{O} = [\underline{1}\ P]A \Leftrightarrow \\ \Leftrightarrow A = ([\underline{1}\ P]^T[\underline{1}\ P])^{-1}[\underline{1}\ P]^T\ O \end{aligned} \tag{2}$$

Further optimization of the adaptation matrix was achieved by applying the following four steps iteratively, until convergence up to the fourth decimal place: 1) compute a cost vector $\underline{C}$ based on the Euclidean distance $\underline{E}$ to the ground-truth color information (3), where ϵ is the relative weight for misfit points; 2) normalize $\underline{C}$ for unity norm, and compute $\underline{C}^2$; 3) generate an empty 24×24 matrix C^2, and populate its diagonal with $\underline{C}^2$; 4) recompute the adaptation matrix, using cost-weighted least-squares fitting (4).

$$\underline{C} = \frac{1}{\underline{E} + \epsilon} \tag{3}$$

$$A = ([\underline{1}\ P]^T C^2 [\underline{1}\ P])^{-1}[\underline{1}\ P]^T C^2 O \tag{4}$$

An adaptation matrix was computed for each acquisition setup (mobile device camera). Prior to the feature extraction stages each one of the dataset images was color corrected, accordingly to its acquisition device and setup, using the corresponding adaptation matrix followed by a non-linear transform.

3.1.2. Iris Boundaries Detection

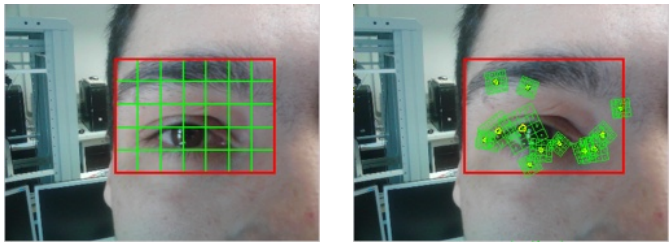
Accurately determining the iris boundaries is a requirement for the following steps, iris and segmentation mask conversion to a pseudo-polar coordinate system and periocular ROI definition), as that will allow to achieve image alignment and scale invariance.

To determine the iris boundaries, the information from both the device acquired image and the binary segmentation mask were combined using a three step approach (Santos and Hoyle, 2012): a) a Hough transform (Ballard and Brown, 1982) is fit to the binary mask boundaries, determining the circle best fitting iris limbic contours; b) a smaller circular ROI is defined of the acquired image, centered in the previously located limbic circle and with 2/3 its radius. Such region is converted to grayscale, its histogram equalized, and an edge map extracted using a Canny edge detector (Canny, 1986); c) a second circle fit to the resulting edge map using another Hough transform, thus approximating the pupillary boundaries.

3.1.3. Iris and Segmentation Mask Normalization

Knowing the iris boundaries, each iris pixel I was assigned to a pair of real coordinates over a double dimensionless pseudopolar coordinate system (5). We followed the rubber-sheet model originally proposed by Daugman (2004) (6), where r and θ are the radius and angle respectively, $x(r,\theta)$ and $y(r,\theta)$ linear combinations of both the set of pupilar boundary points $(x_p(\theta), y_p(\theta))$ and the set of boundary points $(x_s(\theta), y_s(\theta))$ bordering the sclera.

$$I(x(r,\theta), y(r,\theta)) \Rightarrow I(r,\theta) \tag{5}$$



(a) Grid of individual patches (b) SIFT detected features

Figure 3: Illustration of the ROI defined for the global periocular analysis (red), the set of patches used on the distribution-based analysis (a), and SIFT detected features (b).

$$\begin{aligned} x(r,\theta) = (1-r)x_p(\theta) + rx_s(\theta) \\ y(r,\theta) = (1-r)y_p(\theta) + ry_s(\theta) \end{aligned} \tag{6}$$

3.1.4. ROI Definition

To carry on with the periocular analysis, a ROI is defined based on the known iris spatial location (x_i, y_i) and radius (r_i). That ROI is composed by 35 square patches, forming a 7×5 grid, where each patch has an area equivalent to $1.4r_i^2$ (Figure 3).

3.2. Feature Encoding and Matching

At the feature encoding and matching stages, information from two different biometric traits were handled as described below: iris and periocular. On the methods designed to work with single channel images, RGB values were converted to grayscale using a weighted sum (7) prior to feature extraction. The weights in equation (7) are the ones used as standard in National Television System(s) Committee (NTSC) colorspace conversion for computing the effective luminance of a pixel.

$$I(x,y) = 0.2989\mathrm{R}(x,y) + 0.5870\mathrm{G}(x,y) + 0.1140\mathrm{B}(x,y) \tag{7}$$

3.2.1. Periocular Feature Analysis

The periocular analysis here proposed was inspired on the works of Park et al. (2009) and Bharadwaj et al. (2010). In the previously defined ROI, two types of analysis were used: and a distribution-based analysis of every patch, and a global analysis of the whole region.

The distribution-based analysis consists in the computation of three well-known descriptors: HOG, LBP and ULBP. Each descriptor is computed sequentially over each patch and quantized into histograms, forming a global 1-D array where shape and texture information is stored. The HOG descriptor (Dalal and Triggs, 2005), widely applied on computer vision, computes the gradient orientation by filtering the image with two kernels: $[-1, 0, 1]$ and $[-1, 0, 1]^T$. The LBP (Ojala et al., 1994) also works in a quite simple yet efficient fashion: pixel intensity changes from an 8-neighbor region to its central pixel are quantized (8) having the sign of their intensities' difference (9) as reference. $I_{x,y}$ denotes the intensity of the original image at position (x, y), and I_n the intensity of a neighbor pixel.

$$\mathrm{LBP}_{x,y} = \sum_{n=0}^{7} \mathrm{sgn}(I_n - I_{x,y})\, 2^n \tag{8}$$

$$\mathrm{sgn}(I_n - I_{x,y}) = \begin{cases} 1, & \text{if } I_n \geq I_{x,y} \\ 0, & otherwise. \end{cases} \tag{9}$$

The ULBP descriptor differs from the LBP as it achieves "improved rotation invariance with uniform patterns and finer quantization of the angular space" (Ojala et al., 2002). Instead of the 2^n possible binary patterns outputted from the regular LBP over a 8-neighbor region, a uniformity measure U is calculated representing the number of bitwise changes in that same pattern (10). This measure can only assume 59 distinct values.

$$\begin{aligned} U(\mathrm{LBP}_{x,y} = {} & |\mathrm{sgn}(I_7 - I_{x,y}) - \mathrm{sgn}(I_0 - I_{x,y})| + \\ & + \sum_{n=1}^{7} |\mathrm{sgn}(I_n - I_{x,y}) - \mathrm{sgn}(I_{n-1} - I_{x,y})| \end{aligned} \tag{10}$$

At the matching stage, the histogram arrays of size N containing the extracted information were compared through χ^2 distance (11).

$$\chi^2_{(\mathrm{hist}_A, \mathrm{hist}b)} = \frac{1}{2} \sum_{n=1}^{N} \frac{(\mathrm{histA}_n - \mathrm{histB}_n)^2}{\mathrm{histA}_n + \mathrm{histB}_n} \tag{11}$$

On the global analysis, feature extraction techniques were applied not to each individual patch, but to the whole ROI. The applied descriptors were SIFT, and GIST. At first, set of key-points and their surrounding information is extracted using SIFT (Lowe, 2004), known to deliver invariance to translation, scale and rotation. SIFT key-points detection relies on a Difference of Gaussians (DOG) function, and features are extracted for their neighborhood based on gradient magnitude and orientation (Figure 3). At the matching stage, their geometrical alignment is used. Finally, a set of five scene descriptors were used (GIST) as proposed by Oliva and Torralba (2001): *naturalness*, that quantifies vertical and horizontal edge distribution); *openness*, as the presence or lack of reference points; *roughness*, the size of the largest prominent object; *expansion*, the depth of the space gradient; and *ruggedness*, a quantification of contour orientation that assesses the deviation from the horizontal. The GIST descriptor was extracted from each color channel individually, and at the matching stage a χ^2 distance (11) was used upon min-max normalization.

3.2.2. Iris Feature Analysis

The iris information was encoded based on Daugman (1993) approach: iris features were extracted convolving iris data in the pseudopolar coordinate system with a bank of 2-D Gabor wavelets, followed by a quantization stage that produced a binary *iriscode* accordingly to the sign of the 2-D integral. To the purpose of iris identification on mobile environments, we choose to use a very small yet optimized wavelet bank. During filter optimization a smaller representative subset of images was used, and filter parameters cycled through a range of scales, orientations and frequencies, fit for our environment. Chosen configurations were the ones that maximized decidability (13).

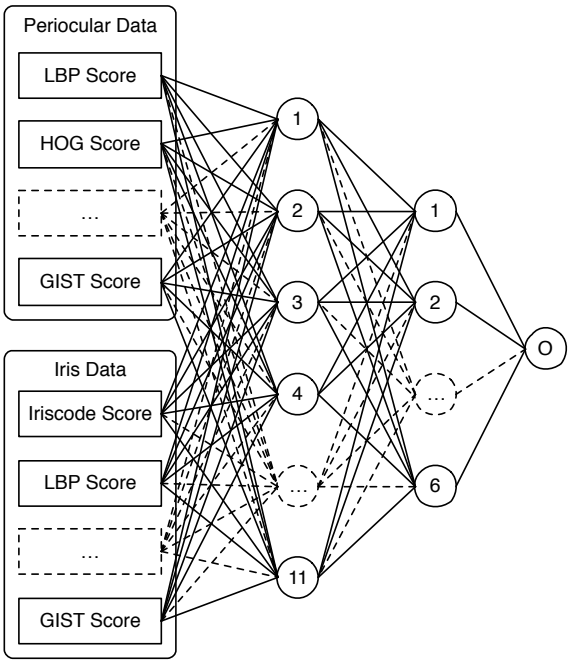


Figure 4: Illustration of the NN architecture used at the score fusion stage. Each circle represents a neuron of the network, and depicted input scores come from the feature matching stage.

At the matching stage, the similarity between two binary codes of size N representing the two irises being compared is assessed through a simple Hamming Distance (HD) (12).

$$\mathrm{HD} = \frac{1}{N} \sum_{n=1}^{N} \mathrm{codeA}_n \otimes \mathrm{codeB}_n \tag{12}$$

Further to that, and as Daugman's technique was developed to deal with iris images captured in controlled settings, the same techniques used to encode periocular data were also applied on the normalized iris region as well.

3.3. Score-level fusion

With several scores resulting from the different encoding/matching methodologies, an Artificial Neural Network (NN) was trained to fuse them into a final recognition score. NN-based methods have been widely applied on classification problems, for their learning abilities and good generalization.

For the purpose of this work a two hidden layers NN was trained with back-propagation (Figure 4). The architecture of the NN was as follows: the first hidden layer had eleven neurons, the same number of scores resulting from the matching stage; the second hidden layer had six neurons; and the final (output) layer with just one, since we are dealing with a binary classification problem. Once again, a smaller data partition was used at the training stage, and was not included on the test phase.

Table 2: Individual performance metrics for each recognition method and trait, along with the ones from iris, periocular and global fusion. Performance metrics are Decidability (DEC), Area Under Curve (AUC) and Equal Error Rate (EER). Top scores are marked *bold*.

	Trait →	*Periocular*						*Iris*							Global
	Method →	LBP	HOG	SIFT	ULBP	GIST	Fusion	Iriscode	LBP	HOG	SIFT	ULBP	GIST	Fusion	Fusion
No color correction	DEC	0.989	0.969	0.716	1.272	1.859	2.164	0.674	0.289	0.515	0.324	0.324	0.320	0.835	2.295
	AUC	0.764	0.751	0.715	0.816	0.915	0.923	0.684	0.588	0.641	0.583	0.589	0.615	0.717	0.932
	EER	0.308	0.315	0.348	0.261	0.166	0.159	0.366	0.443	0.401	0.443	0.440	0.418	0.344	0.148
Histogram equalization	DEC	0.986	0.860	0.668	1.267	1.841	2.101	0.616	0.246	0.371	0.293	0.353	0.199	0.753	2.215
	AUC	0.763	0.725	0.696	0.815	0.910	0.917	0.669	0.581	0.601	0.576	0.600	0.582	0.696	0.925
	EER	0.309	0.340	0.365	0.262	0.172	0.165	0.374	0.446	0.433	0.445	0.433	0.442	0.361	0.155
Device specific correction	DEC	0.989	1.009	0.731	1.270	1.889	2.215	0.639	0.173	0.482	0.347	0.266	0.230	0.809	**2.331**
	AUC	0.766	0.761	0.720	0.817	0.919	0.927	0.675	0.578	0.637	0.590	0.584	0.593	0.711	**0,934**
	EER	0.305	0.308	0.343	0.259	0.163	0.155	0.372	0.450	0.402	0.437	0.444	0.434	0.349	**0,145**

4. Results and Discussion

To assess our method performance, a total of 121.245 random matches were generated, between images from any two acquisition setups, being the inter- to intra-class comparisons ratio 2:1. Three performance measures were used: DEC, AUC and EER. Decidability d' was first introduced by Daugman (1993), and quantifies *intra-* and *inter*-class separability by relating their mean μ and standard deviation σ values.

$$d' = \frac{\|\mu_{\text{inter}} - \mu_{\text{intra}}\|}{\sqrt{\frac{\sigma^2_{\text{inter}} + \sigma^2_{\text{intra}}}{2}}} \tag{13}$$

The Receiver Operating Characteristic (ROC) curve relates the sensitivity, or true positive rate (TPR) with the false positive rate (FPR). Based in that plot, the AUC can be perceived as a quantification of how well pairwise comparisons are performed on a binary classification problem. On the perfect scenario, all positive matches are ranked higher than the negative ones, and the AUC equals one. Finally, setting the operating threshold for the accept/reject decision so that the probability of false acceptance equals the probability of false rejection, we obtain the EER.

The performance registered for every feature encoding and matching technique, *per* trait behavior and global fusion outcome is registered on Table 2. As we can see top results are registered for global fusion, over device-specific color corrected images, with a Decidability of 2.331, and an AUC of 0.934. On quantifying the performance improvement introduced by the sensor-specific color correction technique, we applied our proposed methodology to all matches with three variations in the normalization stage: without performing any color-correction; with the proposed color correction; and, for comparison purposes, with simple histogram equalization over all channels of the working image. As we can see, the proposed color correction technique improves the performance over all the stressed periocular recognition approaches, as well as on the final score resulting from the fusion of all methods. It was a much better approach than the commonly used histogram equalization, whose application actually worsened five of the six periocular approaches, as well as the score-fusion output. Even so, the improvement produced by the sensor-specific correction was not so expressive as initially expected after visually inspecting the color corrected images. A possible explanation is that some of the used feature encoding methods were designed to work over single-channel images, thus not implicitly taking into account some of the chromatic features that could have been lost during grayscale conversion.

Examining *per* trait performance, and paying attention to the values obtained over color-corrected images, we can see how the combined information extracted from the whole periocular region is far more discriminant than the iris, in the mobile application scenario. That is particularly visible in the ROC curves at Figure 5, where we can see that the plot corresponding to the periocular fusion almost overlaps the plot corresponding to the global fusion, being the area between them of only 0.007. In fact, we can't say that periocular analysis does not take iris features into account, as it was not removed nor overlapped prior to the encoding stage.

Reviewing the individual performance of each one of the methods that constitute the proposed periocular analysis (Figure 6a), we can see how GIST descriptors is the approach with

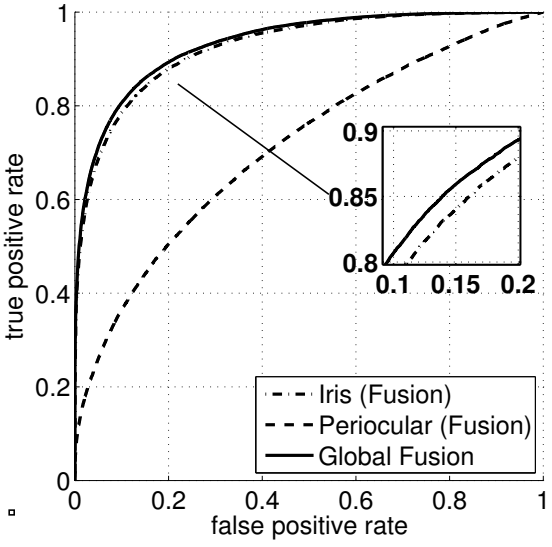
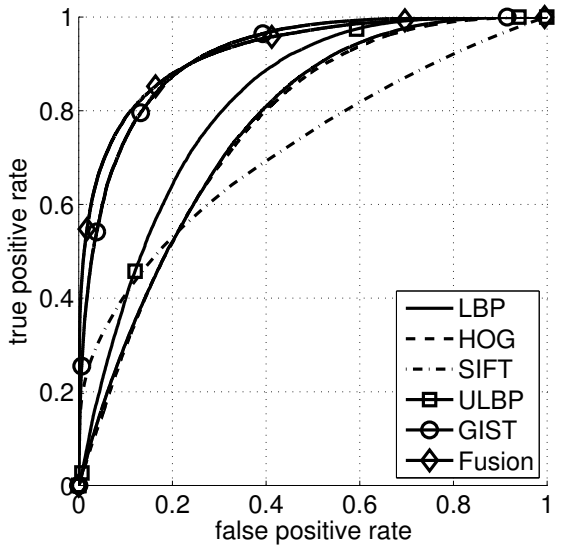


Figure 5: Receiver Operating Characteristic curves for the score-level fusion of the stressed iris recognition methods, the periocular recognition methods, and the global fusion.

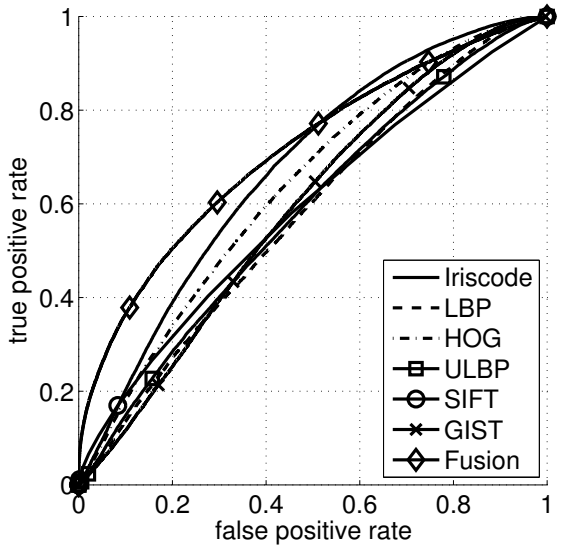
Table 3: Method fusion performance, after color correction, for each acquisition setup.

		AR0	AR1	BF0	BR0	BR1	CF0	CR0	CR1	DF0	DR0
No color correction	DEC	2.481	2.392	2.141	2.456	2.137	2.045	2.153	2.459	2.083	2.423
	AUC	0.940	0.941	0.917	0.940	0.933	0.917	0.922	0.939	0.916	0.938
	EER	0.135	0.135	0.158	0.138	0.149	0.169	0.162	0.138	0.167	0.139
Device specific correction	DEC	2.497	2.446	2.174	2.485	2.164	2.034	2.201	2.561	2.099	2.501
	EER	0.941	0.943	0.921	0.942	0.935	0.917	0.927	0.945	0.917	0.943
	AUC	0.134	0.132	0.156	0.138	0.145	0.165	0.155	0.130	0.166	0.131

highest benefits, followed by ULBP. LBP and HOG have very similar performance, and SIFT was the descriptor with lowest AUC, even though it was able to achieve higher sensitivity with lower FPR for more restrictive thresholds than most of the stressed periocular methods. Since we are aiming at performing biometric recognition in mobile devices, known to have more resources constraints than regular computers, that can be regarded as a good indicator: since SIFT is more computationally expensive than the other tested methods, we can choose not to include it with less impact on the overall performance. Even if only GIST were used, with its five scene descriptors being easily and quickly computed, we could still get an AUC of 0.919. Nonetheless, it is remarkable how such simple feature encoding techniques produce relatively good scores, considering the constraints associated with the mobile working conditions, specially the deterioration of the acquired images. Attending at the same methods' performance over iris data alone (Figure 6b), we can observe that they are not so good at discriminate its features, being the individual method with best performance Daugman's iriscode analysis. We must have in mind that the CSIP acquisition setups didn't favor the capturing of iris details.



(a) Periocular Recognition



(b) Iris Recognition

Figure 6: Receiver Operating Characteristic curves for the stressed periocular (a) and iris (b) recognition methods and their fusion.

Table 3 reports on the recognition performance for when using images acquired at the different imaging setups, with and without performing color correction. Those values were obtained selecting from the total of generated matches the ones where at least one of the images was enrolled at that specific setup. We can see how color correction considerably improves the decidability values on the top performing devices. As we can see, top performances are achieved over images taken with rear cameras, usually without using the device flash. In fact, choosing to use the built-in flash tends to result in performance degradation, even that color correction impact in performance was greater when using images acquired with the flash light on. As for frontal cameras, they do not seem as fit for mobile biometrics as rear ones. That can be particularly tough if the intuit of the application is to verify the phone's user identity, since its fairly more easily obtain a good self-captured image using the frontal camera. Despite the frontal cameras having significantly less resolution that the rear ones, it does not seem to be a relation between that fact and their lack of performance. Device D, for instance, is the device with lower rear sensor resolution and its performance is almost identical to the other setups.

5. Final Considerations

This paper introduces the Cross-Sensor Iris and Periocular Dataset (CSIP) dataset, containing images acquired under ten different mobile setups, with eight visible noise factors. Such particularities make it suitable to evaluate iris and periocular

recognition methods on heterogeneous mobile conditions, and the distribution of each image iris segmentation masks also allow to stress iris segmentation methods on those same conditions.

Further, we identify the chromatic disparity introduced by some devices, proposing the usage of a sensor-specific color correction technique. Results shown that top results were obtained after color correction. Being aware that application scenarios where the biometric recognition process is conducted with images acquired on a single device can deliver better results, we aim at achieving higher cross-sensor performance. That will allow to attain higher confidence on matches between images acquired with very distinct mobile setups, or even on comparisons against a previously stored dataset acquired with other (or multiple) devices.

We proposed the fusion of iris and periocular information to achieve reliable biometric identification in mobile setups, and observed how simple feature encoding techniques deliver considerably good performance. That is particularly convenient when aiming at conduct the whole recognition performance on mobile devices with higher computational constraints, as the top performing methods indeed had considerably low computational cost. Ultimately, and if aiming at reducing even more the computational cost on mobile environments, using only the GIST classifier can be an option.

Finally, results point out that, for the tested setups, high image resolution is not an essential requisite to mobile biometrics, and rear cameras are best suited for periocular recognition, preferable without flash.

5.1. Further Work

At a further stage, authors plan to expand the CSIP dataset with a more significant amount of participants, and a wider range of acquisition setups. We theorize that widening the dataset to further devices and participants, and applying the proposed recognition technique, would emphasize the contribution of both color correction and the usage of iris features.

Another interesting line of work will be to conduct further tests with different iris and periocular recognition methods, specially the ones that explicitly rely on color data, comparing the cross-sensor performance with the performance registered for other existing datasets.

Stressing different color correction techniques could also be interesting, despite the one we applied in this paper having the advantage of being computational efficient and easy to apply, as long as the correction matrix for the camera sensor is known.

Acknowledgments

The authors would like to acknowledge the financial support provided by *FCT - Fundação para a Ciência e Tecnologia* through the research grant SFRH/BD/80182/2011, and the funding from 'FEDER - QREN - Type 4.1 - Formação Avançada', co-founded by the European Social Fund and by national funds through Portuguese 'MEC - *Ministério da Educação e Ciência*'. This work was also supported by the RSU - Remote Sensing Unit through 'PEst-OE-FIS/UI0524/2014' with funds provided by the FCT.

References

Adams, J., Woodard, D., Dozier, G., Miller, P., Bryant, K., Glenn, G., 2010. Genetic-based type ii feature extraction for periocular biometric recognition: Less is more, in: Pattern Recognition (ICPR), 2010 20th International Conference on, pp. 205 –208. doi:10.1109/ICPR.2010.59.

Ahmed, N., Natarajan, T., Rao, K., 1974. Discrete cosine transform. Computers, IEEE Transactions on C-23, 90 –93. doi:10.1109/T-C.1974.223784.

Ballard, D., Brown, C., 1982. Computer Vision. Englewood Cliffs, NJ: Prentice-Hall, E.U.A.

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-up robust features (surf). Comput. Vis. Image Underst. 110, 346–359. URL: http://dx.doi.org/10.1016/j.cviu.2007.09.014, doi:10.1016/j.cviu.2007.09.014.

Beer, T., 1981. Walsh transforms. American Journal of Physics 49, 466–472. URL: http://link.aip.org/link/?AJP/49/466/1, doi:10.1119/1.12714.

Bharadwaj, S., Bhatt, H., Vatsa, M., Singh, R., 2010. Periocular biometrics: When iris recognition fails, in: Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, pp. 1 –6. doi:10.1109/BTAS.2010.5634498.

Boles, W., Boashash, B., 1998. A human identification technique using images of the iris and wavelet transform. IEEE Transactions on Signal Processing 46, 1185–1188.

Bowyer, K., Hollingsworth, K., Flynn, P., 2008. Image understanding for iris biometrics: A survey. Comput. Vis. Image Underst. 110, 281–307. URL: http://dx.doi.org/10.1016/j.cviu.2007.08.005, doi:10.1016/j.cviu.2007.08.005.

Canny, J., 1986. A computational approach to edge detection. IEEE Trans. on Pattern Analysis and Machine Intelligence , 679–698.

Castrillón, M., Dénis, O., Guerra, C., Hernández, M., 2007. Encara2: Real-time detection of multiple faces at different resolutions in video streams. Journal of Visual Communication and Image Representation 18, 130–140.

Clausi, D., Jernigan, M., 1996. Towards a novel approach for texture segmentation of sar sea ice imagery, in: 26th International Symposium on Remote Sensing of Environment and 18th Annual Symposium of the Canadian Remote Sensing Society, Vancouver, BC, Canada. pp. 257–261.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 886–893.

Daugman, J., 1993. High confidence visual recognition of persons by a test of statistical independence. Pattern Analysis and Machine Intelligence, IEEE Transactions on 15, 1148 –1161. doi:10.1109/34.244676.

Daugman, J., 2004. How iris recognition works. Circuits and Systems for Video Technology, IEEE Transactions on 14, 21–30. doi:10.1109/TCSVT.2003.818350.

Daugman, J., 2007. New methods in iris recognition. IEEE Trans. Systems, Man, Cybernetics B 37, 1167–1175.

Hurley, D., Nixon, M., Carter, J., 2000. A new force field transform for ear and face recognition, in: Image Processing, 2000. Proceedings. 2000 International Conference on, pp. 25 –28. doi:10.1109/ICIP.2000.900883.

on Illumination, I.C., 2004. Colorimetry. CIE technical report, Commission internationale de l'Eclairage, CIE Central Bureau.

Juefei-Xu, F., Cha, M., Heyman, J., Venugopalan, S., Abiantun, R., Savvides, M., 2010. Robust local binary pattern feature sets for periocular biometric identification, in: Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, pp. 1 –8. doi:10.1109/BTAS.2010.5634504.

Juefei-Xu, F., Luu, K., Savvides, M., Bui, T., Suen, C., 2011. Investigating age invariant face recognition based on periocular biometrics, in: Biometrics (IJCB), 2011 International Joint Conference on, pp. 1 –7. doi:10.1109/IJCB.2011.6117600.

Laws, K., 1980. Rapid texture identification, in: Proc. SPIE Conf. Image Processing for Missile Guidance, pp. 376–381. doi:10.1117/12.959169.

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60, 91–110. URL: http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94, doi:10.1023/B:VISI.0000029664.99615.94.

Mallat, S., 1989. A theory for multiresolution signal decomposition: the wavelet representation. Pattern Analysis and Machine Intelligence, IEEE Transactions on 11, 674 –693. doi:10.1109/34.192463.

Matey, J., Naroditsky, O., Hanna, K., Kolczynski, R., LoIacono, D., Mangru, S., Tinker, M., Zappia, T., Zhao, W., 2006. Iris on the move: Acquisition of

images for iris recognition in less constrained environments, in: Proceedings of the IEEE, pp. 1936–1947. doi:10.1109/JPROC.2006.884091.

Miller, P., Rawls, A., Pundlik, S., Woodard, D., 2010. Personal identification using periocular skin texture, in: Proceedings of the 2010 ACM Symposium on Applied Computing, ACM, New York, NY, USA. pp. 1496–1500. URL: http://doi.acm.org/10.1145/1774088.1774408, doi:10.1145/1774088.1774408.

Ojala, T., Pietikainen, M., Harwood, D., 1994. Performance evaluation of texture measures with classification based on kullback discrimination of distributions, in: Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision amp; Image Processing., Proceedings of the 12th IAPR International Conference on, pp. 582 –585 vol.1. doi:10.1109/ICPR.1994.576366.

Ojala, T., Pietikainen, M., Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Analysis and Machine Intelligence, IEEE Transactions on 24, 971 –987. doi:10.1109/TPAMI.2002.1017623.

Oliva, A., Torralba, A., 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. International Journal of Computer Vision 42, 145–175.

Park, U., Jillela, R., Ross, A., Jain, A., 2011. Periocular biometrics in the visible spectrum. Information Forensics and Security, IEEE Transactions on 6, 96 –106. doi:10.1109/TIFS.2010.2096810.

Park, U., Ross, A., Jain, A., 2009. Periocular biometrics in the visible spectrum: A feasibility study, in: Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on, pp. 1 – 6. doi:10.1109/BTAS.2009.5339068.

Proença, H., Filipe, S., Santos, R., Oliveira, J., Alexandre, L., 2010. The ubiris.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. Pattern Analysis and Machine Intelligence, IEEE Transactions on 32, 1529 –1535. doi:10.1109/TPAMI.2009.66.

Santos, G., Hoyle, E., 2012. A fusion approach to unconstrained iris recognition. Pattern Recognition Letters 33, 984 – 990. URL: http://www.sciencedirect.com/science/article/pii/S0167865511002686, doi:10.1016/j.patrec.2011.08.017.

Santos, G., Proença, H., 2013. Periocular biometrics: An emerging technology for unconstrained scenarios, in: Proceedings of the IEEE Symposium on Computational Intelligence in Biometrics and Identity Management – CIBIM 2013, pp. 14–21.

Smith, T., Guild, J., 1931. The C.I.E. colorimetric standards and their use. Transactions of the Optical Society 33, 73–134.

Tan, T., He, Z., Sun, Z., 2010. Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. Image and Vision Computing 28, 223 – 230. doi:10.1016/j.imavis.2009.05.008.

Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features, in: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 511–518.

Wildes, R., 1997. Iris recognition: an emerging biometric technology, pp. 1348–1363.

Wolf, S., 2003. Color Correction Matrix for Digital Still and Video Imaging Systems. Technical Report TM-04-406. NTIA - National Telecommunications and Information Administration. USA.

Woodard, D., Pundlik, S., Miller, P., Jillela, R., Ross, A., 2010. On the fusion of periocular and iris biometrics in non-ideal imagery, in: Pattern Recognition (ICPR), 2010 20th International Conference on, pp. 201 –204. doi:10.1109/ICPR.2010.58.

Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A., 2003. Face recognition: A literature survey. ACM Comput. Surv. 35, 399–458. URL: http://doi.acm.org/10.1145/954339.954342, doi:10.1145/954339.954342.

Chapter 11

BioHDD: a Dataset for Biometric Identification on Heavily Degraded Data

11.1 Overview

This chapter consists of the following article:

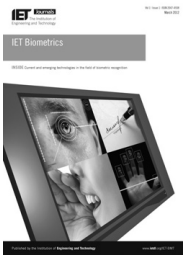
BioHDD: a Dataset for Biometric Identification on Heavily Degraded Data
Gil Santos, Paulo T. Fiadeiro and Hugo Proença
IET Biometrics, 2014.
DOI: 10.1049/iet-bmt.2014.0045

According to SCImago Journal & Country Rank, this journal's index[1] for the 2013 year are as follows:

Category	*Quartile*	*SJR*
Computer Vision and Pattern Recognition	Q3	0,250
Signal Processing	Q3	
Software	Q3	

[1]The SCImago Journal & Country Rank (SJR) indicator is a measure of journal's impact, influence or prestige. It expresses the average number of weighted citations received in the selected year by the documents published in the journal in the three previous years. http://www.scimagojr.com

Published in IET Biometrics
Received on 15th July 2014
Revised on 17th September 2014
Accepted on 23rd September 2014
doi: 10.1049/iet-bmt.2014.0045



ISSN 2047-4938

BioHDD: a dataset for studying biometric identification on heavily degraded data

Gil Santos[1], *Paulo T. Fiadeiro*[2], *Hugo Proença*[1]

[1]*Department of Computer Science, IT – Instituto de Telecomunicações, University of Beira Interior, Covilhã, Portugal*
[2]*Department of Physics, Remote Sensing Unit – Optics, Optometry and Vision Sciences Group, University of Beira Interior, Covilhã, Portugal*
E-mail: gmelfe@ubi.pt

Abstract: Substantial efforts have been put into bridging the gap between biometrics and visual surveillance, in order to develop automata able to recognise human beings 'in the wild'. This study focuses on biometric recognition in extremely degraded data, and its main contributions are three-fold: (1) announce the availability of an annotated dataset that contains high quality mugshots of 101 subjects, and large sets of probes degraded extremely by 10 different noise factors; (2) report the results of a mimicked watchlist identification scheme: an online survey was conducted, where participants were asked to perform positive and negative identification of probes against the enrolled identities. Along with their answers, volunteers had to provide the major reasons that sustained their responses, which enabled the authors to perceive the kind of features that are most frequently associated with successful/failed human identification processes. As main conclusions, the authors observed that humans rely greatly on shape information and holistic features. Otherwise, colour and texture-based features are almost disregarded by humans; (3) finally, the authors give evidence that the positive human identification on such extremely degraded data might be unreliable, whereas negative identification might constitute an interesting alternative for such cases.

1 Introduction

The evolution of the concept of biometrics over the last decades is linked with societies' increasing concerns about both individual and global security. From personal computers to border access control everyone aims at securing their identities, their assets and, primarily, their homeland. Such safety relies on the ability to accurately identify subjects based on biometric features, either biological or behavioural.

Biometric systems rely on the accurate 'extraction' of individuals' distinctive features and their proper 'encoding', so that the essential information is preserved. Those requisites are traditionally assured by high acquisition constraints, with the subject cooperation being a key-element. When moving to unconstrained scenarios, those acquisition constraints are lowered and subject cooperation is not expectable. Recognition became more challenging and alternatives are sought [1, 2], either by: (1) improving the existing algorithms; (2) resorting to multi-modal biometric systems; or (3) exploring new traits could better cope with this new reality. Despite those efforts, no system yet exists capable of effectively dealing with all the issues introduced by biometrics 'in the wild'. In fact, even biometric systems able to cope with less constrained conditions (e.g. Iris-on-The-Move project [3]) still lack an ideal level of user abstraction.

Visual surveillance is a very active field in computer vision, with a lot more applications other than biometrics 'per se' [4]. Existing automatic surveillance systems are rather focused on activity recognition (e.g. W^4 project [5]), and not many projects are prepared to deal with surveillance scenarios from a watchlist approach (e.g. Kamgar-Parsi *et al.* [6]). Furthermore, none of the latter works from the negative identification perspective.

Most biometric systems attempt positive identification (or verification) against a gallery of enrolled users based on a (dis)similarity measure. In many 'in the wild' applications however, biometric systems make more sense when used from the negative perspective: guarantee with enough confidence that an unknown subject does not belong to a gallery of 'persons-of-interest', instead of attempting to identify him. On that basis, facing a watchlist scenario one can aim at spotting a distinctive feature on the probe subject, and exclude those who neither share that feature, nor any of its possible transformations. Moreover, even if we do not have enough distinctive features to support a positive recognition (e.g. because of the quality of acquired images) we can still perform reliable negative recognition.

1.1 Contextualisation: facial biometrics

The everyday use of facial cues includes recognising our peers or unveiling their state of mind, which happens seamlessly and unawarely. Is then easy to place face as the most common and widely used biometric trait, and one of the most successful applications of image analysis and understanding. Several face recognition systems are

commercially deployed and a lot of techniques accessible [7], working on both still and video images. Algorithms are based either on the global analysis of the whole image, or on the relation between facial elements, their localisation and shape. In either case, their effectiveness is conditioned by several factors, which become even more evident 'in the wild': its three-dimensional structure lead to substantial differences in appearance, accordingly to the subject's pose; large portions are often occluded on non-orthogonal data acquisition; facial expressions affect their appearance; and its particularly easy to disguise.

Analysing the human ability to recognise each other, researchers can identify the more reliable cues, valuable for the develop well-grounded recognition methods. Previous studies report interesting findings when exploring the human ability to identify faces (e.g. Sinha *et al.* [8]), encouraging further researching on understand how people cope with 'in the wild' circumstances. In this study we do not aim at mimicking the identification process taking place in human vision, but rather to provide useful insights for further research on this topic. We analyse the noise factors' impact on human identification performance, identifying the features people recall as basis for their judgement.

1.2 Contextualisation: similar datasets

Publicly available datasets exist for both video surveillance [4] and face biometrics [9] research, acquired under less constrained conditions. Although a much higher extent of databases is available, five significant datasets must be mentioned, which contain a more significant amount of pie changes: FERET [10], CMU-PIE [11], CAS-PEAL [12], Multi-PIE [13] and LFW [14] (Table 1). The latter two datasets are presumably the most completes, each one by its own reasons: the Multi-PIE provides facial images from 337 subjects, imaged over four sessions under 15 pose and 19 illumination variations, along with high-resolution registration photos; the Labeled Faces in the Wild (LFW) dataset contains a larger amount of images and subjects, 13 233 and 5749, respectively, at completely 'in-the-wild' conditions, and thus without uniformity among subjects. Although not being an extensive listing of the existing datasets, the ones we present are the most directly comparable to the one we are now establishing.

In this paper we introduce a newly created dataset of heavily degraded facial images, where the 'in the wild' conditions associated with visual surveillance systems are closely simulated. Full 360° illumination and pose variations are introduced, among with other realistic noise factors at different reasonable levels, along with ground-truth information for research validation. Despite containing a lower amount of participants when compared to the existing databases, this new dataset contains a wider range of pose and illumination variations, uniform and comparable for all subjects.

The remainder of this paper is organised as follows: Section 2 describes the BioHDD dataset, detailing the acquisition framework, enrolled participants and introduced noise factors; Section 3 presents the experimental method used in our study, with a thorough analysis of its results; finally, Section 4 states some final considerations, along with further lines of work.

Table 1 Overview of the most relevant and public available face recognition datasets with pie variations, with comparison to our working dataset

Dataset	Subjects	Sessions	Pose	Illumination	Expression
FERET	1199	2	20	2	2
CMU-PIE	68	1	13	43	4
CAS-PEAL	1040	2	21	15	6
multi-PIE	337	4	15	19	6
LFW	5749	?	?	?	?
BioHDD	101	2	24	72	1

Values marked with '?' can not be determined because of the nature of the dataset

2 BioHDD dataset

The main objective of the BioHDD database was to gather images from a significant group of individuals, ranging from clear frontal shots to heavily degraded facial images, enabling to assess the feasibility of biometric recognition 'in the wild'.

2.1 Imaging framework and setup

The imaging framework was installed in a closed lounge without uncontrolled lightening sources. Participants were illuminated with a single 800 W halogen projector, and a white cloth was used as image background to avoid contextual interferences. The acquisition process consisted of three acquisition stages: registration, still image acquisition and video acquisition.

At the registration stage three reference facial images of high-quality were acquired from each participant (frontal, left- and right-hand side – Fig. 1). The acquisition device gathered information from the visible wavelength slice of the electromagnetic spectrum, with the light source directly above it. Participants were asked to essay a neutral expression and look forward, aided by three fixation points, so that all observers were facing the same direction during this stage.

On a second stage images were acquired 'simultaneously' on both NIR and VIS, while introducing four variations: illumination angle and intensity, subject revolution and head-tilt – Fig. 3, columns 1–4. Changes on the illumination angle were achieved with the halogen projector shifting on 45° steps (Fig. 2, *A–H*), while participants kept facing the acquisition device. Additionally, participants were asked to face eight fixation points evenly distant from each other, introducing subject revolutions in full 360°. For all variations, participants were imaged facing forward and tilting their head up and down, while simulating illumination intensity changes using the acquisition device exposure settings.

At a final stage, subjects walked trough a corridor with non-uniform illumination conditions whilst captured by a VIS greyscale camera placed on a upper level. As we can

Fig. 1 *Example of images acquired used as gallery data: left-hand side, frontal and right mugshots*

doi: 10.1049/iet-bmt.2014.0045

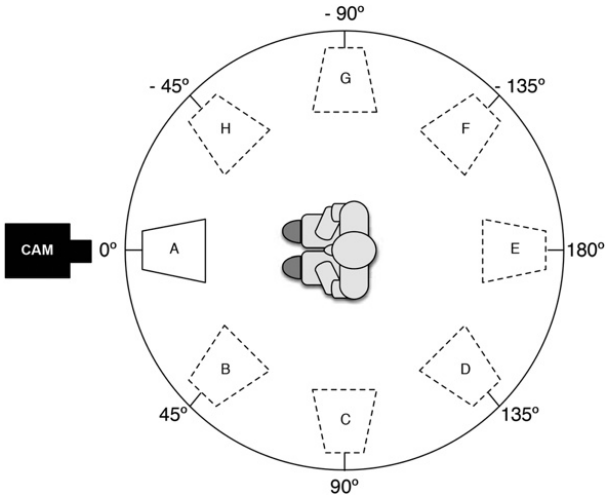


Fig. 2 *Schematic perspective of the image acquisition framework (over-top view)*

For illumination changes, the light source alternate on positions A to H with participants facing the camera
For rotation changes, participants were asked to align themselves with the different reference points while the camera and light source remained aligned at the initial position

see from the samples at Fig. 4, surveillance-like data acquisition was closely simulated.

Table 2 presents a complete hardware specification.

Data was gathered on two acquisition sessions with a minimum of two weeks apart. On the first acquisition session participants wearing glasses were required to remove them, and the ones with longer hair were asked to tie it. Likewise, videos acquired during that same session had participants looking at a fixation point while walking. To increase variability, on the second session such constraints were not applied. No modifications were introduced on the hardware setup or location. Attendance to both sessions was around 88%, representing a total of 101 participants. As described on Table 3, 66 male subjects and 35 female subjects were enrolled, most of them Caucasian. For normalisation purposes, acquired images were manually cropped to 600 × 600 px, while keeping the face centred. Registration images from Stage 1 were stored with 2, 048 × 2, 048 px.

2.2 Heavily degraded data

Not all noise factors associated with recognition 'in the wild' were introduced during the acquisition stage. As so, additional image degradation procedures were carried on.

A total of ten noise factors were identified and grouped in three different sets: (1) 'real' noise factors introduced with the imaging setup; (2) 'simulated' noise factors that although not introduced at the acquisition stage are related with the imaging process; and (3) noise factors associated with data storage and transmission. Each noise agent comprises different levels (L_i), as illustrated on Fig. 3, and their presence follows the reasoning we now describe.

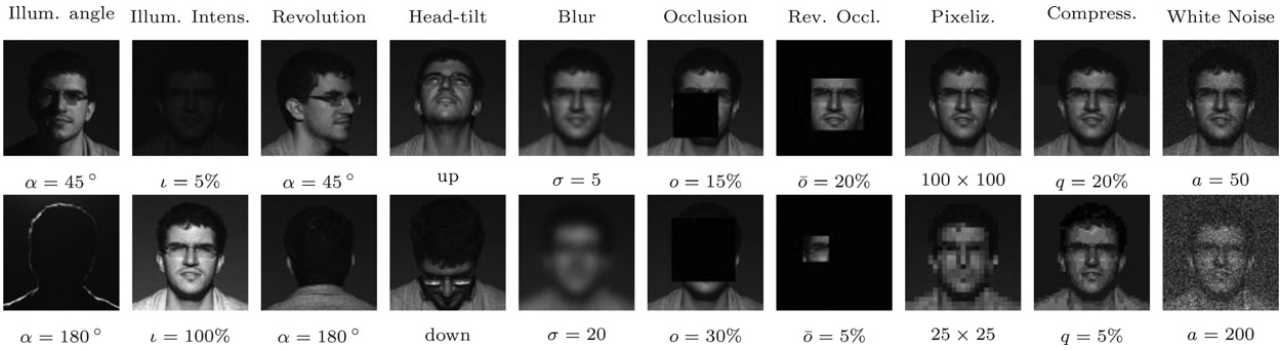


Fig. 3 *Examples of the types of image degradation factors in the BioHDD dataset*

From left- to right-hand side: illumination angle and intensity, subject revolution, head-tilt, blur, occlusion and reverse occlusion, pixelisation, compression and white noise
The top row corresponds to the first noise level L_1, and the bottom row to the maximum noise level L_{max}
On illumination intensity and head-tilt, both images represent L_1, since their difficulty is similar
Although only VIS data is depicted, each image has its NIR counterpart

Fig. 4 *Samples from the video acquisition stage*

Frames were cropped for illustration purposes

Table 2 Details of the BioHDD acquisition devices, image and video settings

	Registration	Image Acq.	Video Acq.
camera	Canon EOS 5D	JAI AD-080GE	Stingray F504-B
lens	Canon EF 100-400	NIKKOR 55-80	HR F1.4/8 mm
spectrum	visible	visible + NIR	visible
color space	RGB	RGB + greyscale	greyscale
channel depth	8bit	8bit	8bit
frame size	4368 × 2912 px	984 × 768 px	1224 × 1028 px
cropped size	2048 × 2048 px	600 × 600 px	–
format	PNG	PNG	AVI
frame-rate	–	–	15 fps

2.2.1 Real noise factors: As previously described, this set of noises was directly introduced at the acquisition stage. When working in unconstrained scenarios optimal illumination cannot be assumed. Along with the images captured at the 'best' conditions (with average exposure and having the light source directly over the acquisition device), data was also captured varying the 'lightening angles' and the 'illumination intensities' (low lighting and over-exposure). The chosen angles cover all 360° degrees (at 45, ° steps), and intensity changed from 5 to 100%. To cover a higher amount of poses, subject 'revolution' was also introduced over eight angles (similarly to illumination) and 'head-tilting' in two, with participants facing up and down. Those choices were based on the reasoning that individuals trying to avoid detection are most likely to be facing the ground or away from any visible cameras.

2.2.2 Simulated noise factors: To mimic acquisition issues as the ones associated with inappropriate lens settings, poor focus, subject movement etc, four levels of 'blur' were simulated applying Gaussian filters with standard deviation raging from $\sigma_{L_1} = 5$ to $\sigma_{L_4} = 20$.

Face occlusion was simulated by overlapping a black patch to the original image, covering $o_{L_1} = 15\%$ to $o_{L_1} = 30\%$. A different flavour of occlusion where only a small portion of the image is left-hand side visible, $\bar{o}_{L_1} = 20\%$ to $\bar{o}_{L_4} = 5\%$, was also simulated. This noise factor can also be related with the use of certain headgear (e.g. balaclava).

In certain scenarios we observe that the used devices are of low or insufficient spatial resolution, or post-processing censorship is applied to avoid detection of a particular subject or distinctive feature that is intended to remain anonymous. This 'pixelisation' effect was obtained by downscaling the original photo: $\text{size}_{L_1} = 100 \times 100\,\text{px}$ to $\text{size}_{L_4} = 25 \times 25\,\text{px}$.

2.2.3 Storage/transmission related noise factors: Finally, 'compression' degradation found on systems that rely on digital storage or broadcasting was simulated using a standard JPEG algorithm. Quality ranged from $q_{L_1} = 20\%$ to $q_{L_4} = 5\%$. Based on the same reasoning, inherent to data storage on photographic film or broadcasted through analogue channels, 'white noise' was simulated.

To generate probe images I_P, one transformation from each set T_1, T_2, T_3 was randomly selected, and the corresponding noise factor was applied to the original image at a random level k, l, m, respectively (1). Noise application was sequential, with the last noise transformation T_3 being applied upon T_2 result, denoted by ° and T_2 being applied over T_1 output. Sample probe images obtained using this fusing technique are illustrated at Fig. 5

$$\begin{aligned} I_P &= (T_3(L_m) \circ T_2(L_l) \circ T_1(L_k))(I) \\ &= T_3(L_m)(T_2(L_l)(T_1(L_k)(I))) \end{aligned} \quad (1)$$

Table 3 Details of the BioHDD subjects that offered themselves as volunteers to both imaging sessions

Gender	Male	65%	Age	[0, 20]	10.89%
	Female	35%		[21, 25]	44.55%
Origins	European	95%		[26, 30]	15.84%
	African	4%		[31, 35]	9.90%
	Asian	1%		[35, 99]	18.81%

2.3 Dataset availability

The complete BioHDD dataset is public and freely available for academic and research purposes [http://biohdd.di.ubi.pt]. Researchers are granted access to: (1) 606 registration images; (2) 27 270 probe samples with the variations introduced during the acquisitions stage; (3) 27 270 similar images on the NIR spectrum; (4) 2500 probe images with combined noises; (5) 202 greyscale videos with surveillance like data from each participant. Further probe images can be generated 'on-demand', and all data comes with ground-truth information about the associated noise levels.

3 Experiments and discussion

3.1 Experimental method

Our goals to study the human ability to identify their peers on heavily degraded data were: (1) identify the noise factors whose avoidance would be preferable, by associating each one of them with a specific impact on human identification performance; (2) pin down the regions identified as part of the process and, if possible, even specific features; (3) illustrate how negative recognition might still be reliable 'in the wild', where the positive approach is unattainable. To do so, a web-page was built with a custom participation interface mimicking a watchlist recognition scenario – Fig. 6.

For this experiment, a total of 200 000 trials were generated, combining 2500 probe images and 2500 galleries. At the begging of each test the interface was populated with a random trial, with 3/4 probability of the gallery containing the subject on the probe image. Each participant was asked to do one of three actions, for each gallery identity shown

1. mark it as green if they feel that the identity on the mugshot corresponds to the probe image (positive identification);
2. mark it red if they are certain that the identity on the mugshot does not correspond to the query image (negative identification);
3. leave it blank, in case of uncertainty.

In the case of identification, participants were asked to fill the appropriate text-box to justify their answer. No time restriction was set for image examination, and upon finishing a new test was loaded. Each participant was free to take as many tests as he wanted. The experiment ended after one month, collecting a total of 3650 participations from 45 different countries. A total of 17 438 identifications and 1422 justifications were obtained.

doi: 10.1049/iet-bmt.2014.0045

Fig. 5 *Sample trial images with different levels of noise combined*

3.2 Results and discussion

Although including a third class for 'no decision' in our testing interface, we simplified our problem to a binary one by analysing the answers where participants were sure enough of their answer to give a specific identification (either positive or negative). To assess the identification performance, four well-known statistical measures were used: sensitivity (or true positive rate [TPR]), specificity (SPC), accuracy (ACC) and Matthews correlation coefficient (MCC). TPR and SPC, are given by (2) and (3), respectively, and weight the correct responses by the total of positive (true positives (TP) + false negatives (FN)) and negative (true negatives (TN) + false positives (FP)) answers

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

$$\mathrm{SPC} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{3}$$

The accuracy gives us the overall ratio of correctly classified matches, 1 being the optimal value where all instances have been correctly classified. For a balanced analysis we used MCC, which takes into account the high discrepancy between the amount of positive and negative matches. It can be regarded as a correlation coefficient between participants' answers and the correct identification, and its output ranges in the $[-1, 1]$ interval, where 1 the optimal value [15]

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}} \tag{4}$$

$$\mathrm{MCC} = \frac{\mathrm{TP} \cdot \mathrm{TN} - \mathrm{FP} \cdot \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}} \tag{5}$$

In Fig. 7 we can see from the true positive rate and specificity probability density functions, computed for all subjects on the dataset. TPR is clearly more prone to variations, as positive samples are more difficult to be found in the experimental

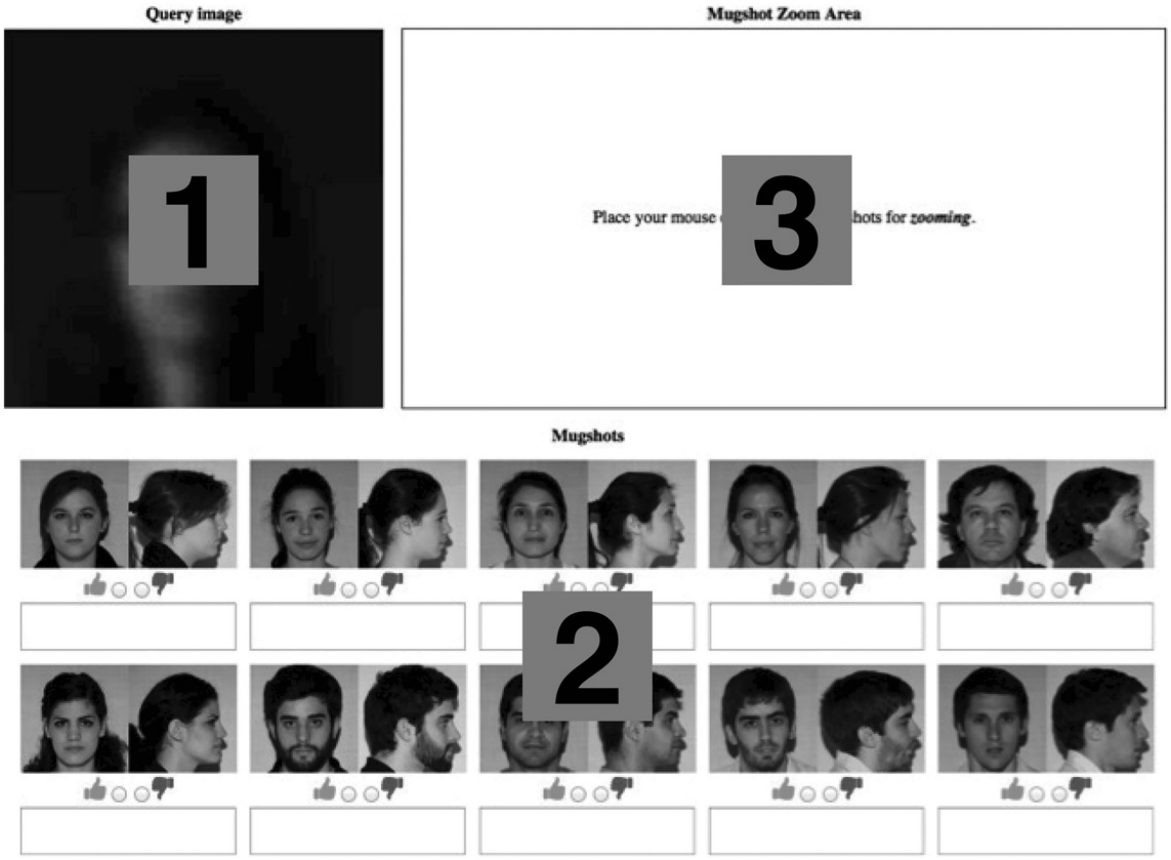


Fig. 6 *Web interface of the conducted survey, with three major panels: (1) a probe sample from an unknown identity; (2) a set of 10 profile / frontal mugshots, representing the gallery dataset; (3) zoomed-up perspective of each gallery sample, populated on mouse-over on region 2*

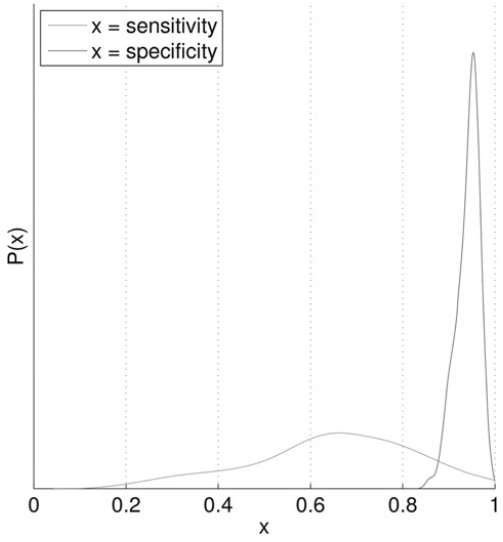


Fig. 7 *Per-subject sensitivity and specificity probability density functions*

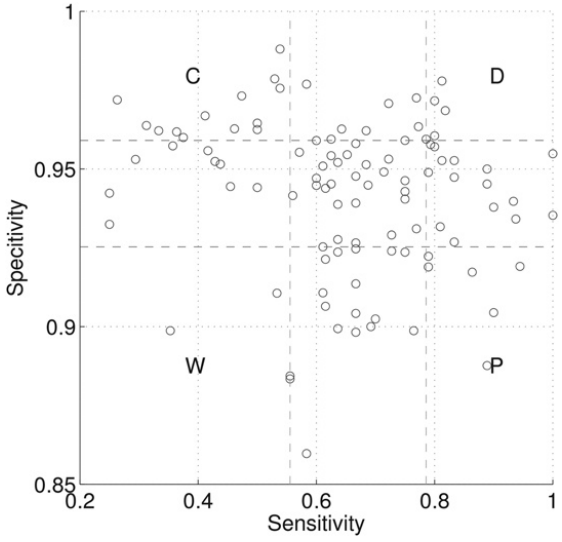


Fig. 8 *Zoo plot for the overall user performance*

Dashed lines represent the first and third quartiles for sensitivity and specificity distributions

The identities on the 'P','D', 'C' and 'W' regions are more likely to assume dove (D), chameleon (C), phantom (P) and worm-like (W) behaviour

setup. We should have in mind that one out of four trials could not lead to TP, since the subject from the probe image is not in the gallery. To stress the possible relationship between false positives and the impossibility of making a positive match, implying participants had attempted identification either way, we performed a paired-sample Student's t-test: For all the n subjects in the database, let us define the fall-out distributions d_1 for trials where positive matches were possible, and d_2 when not. Let us then consider the null hypothesis H_0 where the difference D between d_1 and d_2 follows a normal distribution with mean equal to zero and unknown variance, tested through (6)

$$t = \frac{\overline{X}_D - \mu_0}{s_D/\sqrt{n}} \tag{6}$$

where $\overline{X}_D$ and s_D are D average and standard deviation values, and μ_0 the mean for the d_1 distribution. Experimental data returned a $p$value of 1.81×10^{-14}, thus rejecting the null hypothesis: the distributions are significantly different, indicating that most participants indeed attempted to perform a positive match, even when it was not possible.

Plotting each one of the subjects in the dataset as a function of the TPR and SPC, we can understand their individual propensity to correct identification – Fig. 8. Furthermore, we can group them in four biometric menagerie classes as suggested by Yager & Dunstone [16]: doves, chameleons, phantoms and worms. 'Doves' are the most favourable subjects and the optimal group for any recognition system, as they do not produce verification error. High values are observed for both TPR and SPC. 'Chameleons' are subjects who are easily misidentified as they always appear similar to others, their specificity is high, but true positive rate is extremely low. 'Phantoms', in opposition to chameleons, are associated with low SPC and high TPR. 'Worms', contrary to doves, are the most critical subjects in a biometric system. They behave in the worst possible way, yielding low true positive rate and specificity. At a central location we have 'the herd', where the most common users ('sheep') are located.

To define the limits for each class, we start by defining two regions for the true positive rate, TPR_{Q1} and TPR_{Q3}, containing the subjects below the first quartile and over the third quartile, respectively. If we define two similar regions (SPC_{Q1} and SPC_{Q3}) for the specificity, a subject s is said to assume a particular behaviour according to (7) [17]

$$\begin{cases} \text{Dove}, & \text{if } s \subset \text{TPR}_{Q3} \cap \text{SPC}_{Q3} \\ \text{Chameleon}, & \text{if } s \subset \text{TPR}_{Q1} \cap \text{SPC}_{Q3} \\ \text{Phantom}, & \text{if } s \subset \text{TPR}_{Q3} \cap \text{SPC}_{Q1} \\ \text{Worm}, & \text{if } s \subset \text{TPR}_{Q1} \cap \text{SPC}_{Q1} \end{cases} \tag{7}$$

As we can see on Table 4, even with the degradation introduced in the probe images participants were able to correctly match 92% of the instances they were presented with. To assess the effect of each noise on that performance level, we computed the same metrics for when removing each one of them. Additionally, we analysed how each menagerie class relocated as a consequence of a specific noise, as follows.

Take an initial point $A(\text{TPR}_a, \text{SCP}_a)$ representing the global recognition capabilities of an individual on the dataset, and a point $B(\text{TPR}_b, \text{SPC}_b)$ computed likewise for when a specific noise is removed. We can then compute the global distance to the optimal point $O(1, 1)$ as (2), and the distance upon noise removal d_b likewise.

$$d_a = \sqrt{(1 - \text{TPR}_a)^2 + (1 - \text{SPC}_a)^2} \tag{8}$$

Finally, the individual optimisation produced by noise removal can be accessed through $\zeta \rightarrow [-1, 1]$, where -1 represents the worst case scenario and 1 the best improvement possible. Zero means no performance change

$$\zeta = \frac{d_a - d_b}{d_a + d_b} \tag{9}$$

Assessing the average $\zeta-$ values on each one of the zoo-plot regions, we obtain the values at Table 5.

doi: 10.1049/iet-bmt.2014.0045

Table 4 Overall sensitivity (TPR), specificity (SPC), accuracy (ACC) and MCC values and the same statistics for when a noise factor is removed

	TPR	SPC	ACC	MCC
overall	0.657	0.941	0.918	0.547
illumination angle	0.682	0.944	0.922	0.573
illumination intensity	0.633	0.938	0.913	0.518
revolution	0.641	0.941	0.916	0.537
head-tilting	0.671	0.941	0.919	0.558
Gaussian Blur	0.670	0.943	0.920	0.560
occlusion	0.641	0.939	0.914	0.532
rev. occlusion	0.675	0.942	0.920	0.558
pixelisation	0.641	0.941	0.917	0.537
compression	0.618	0.937	0.911	0.506
white noise	0.688	0.945	0.923	0.580

When a noise factor is withdrawn, one would expect the optimisation to always be positive. However, both analyses show only four noise factors that led to significant improvements. The most conditioning element is the introduction of the 'white noise' associated with analogue channels, and as the opposite tendency is observed for digital 'compression', we can conclude that digital channels should be used. The second considerable constraint is the 'illumination angle': when the subject being identified was not frontally lit, participants exhibited higher error rates. On the other side, variations on the 'lightening levels' were not relevant, as participants were able to accommodate to both under- and over-exposure. We also observed their ability to cope with 'occlusion' up to a certain degree, and only when a portion of the face was visible ('reverse occlusion') their performance started to degrade. Finally, participants' performance was also significantly conditioned by 'head-tilting'. This last observation is of special importance: as mentioned before, individuals trying to avoid detection are most likely to be facing the ground or away from any visible cameras. Along with illumination intensity and occlusion, some other noise factors' removal did not led to improvements in performance: 'revolution', indicating that useful features can also be derived from the side of the head, and are actively used in human identification; and 'pixelisation', that along with 'compression' lead us to infer that global features are preferred over local and more detailed ones.

When performing a 'per' species analysis, its perceptible how sensitivity tens to decrease at an higher rate than specificity increases. That explains the negative ζ-values for species located over the TPR' third quartile (doves and phantoms), associated with a convergence to 'the herd'. The class that benefits the most from noise removal is 'Worms', with improvements over three times greater than those observed for 'Sheep'.

Table 6 Probability (%), sensitivity TPR, specificity (SPC), accuracy (ACC) and MCC values for feature category usage on recognition justifications

	(%)	TPR	SPC	ACC	MCC
shape	49.64	0.87	0.88	0.88	0.62
color	6.05	0.71	0.92	0.90	0.52
texture	0.51	1.00	0.50	0.71	0.55

As above stated, a set of justifications for each of the responses given by the volunteers of our on-line survey we're collected. These answers are an important source of information to perceive the type of features predominantly used by humans in identification tasks, as well as to relate the usability of each feature to the degree of success in the corresponding identification. Hence, the responses were grouped by the type of feature they mention and the facial region, as detailed in Tables 6 and 7.

On a 'per' category analysis (Table 6), we can see how almost half the justifications mention shape related features, making it the most commonly used feature type. Colour related features are much less used (6.04%), skin and hair colour being the most significant ones. This is a considerable difference, even considering that the dataset consists mainly of young European participants. Attending to the accuracy levels alone, one could be biased into considering the latter to be a better feature.

To take into account both the high specificity value and the difference in class sizes MCC was also analysed. This measure weights the importance of TPR and SPC by the size of each class, shows shape to be not only the most used feature type, but also the more reliable on both positive and negative identification. Finally, the number of participants that used textural information is almost residual (0.51), and usually refers to freckles and another skin signs, tattoos and jewellery.

In Table 7 we summarise the 'per' region analysis. As we can see, when looking to justify the identifications they make participants use holistic features on almost 2/3 of the justifications, with two most relevant cues: perception on probe subject gender, and a broad analysis of head's shape. From that, special attention is paid to top regions, which can intuitively be related to a higher amount of detail, as more elements are present. Actually, if we analyse the weighted accuracy average per region we can see how topmost areas are indeed less deceiving than lower ones, which is explained by the high volume of negative identifications based on hairstyle. Hair related features played an important

Table 5 Average ζ-values for all zoo-plot regions ($\times 10^{-2}$) upon noise removal

	Doves	Chamel.	Phant.	Worms	Sheep	Average
illum. angle	6.66	3.78	−3.26	11.90	3.01	2.79
illum. intens.	4.48	−5.32	−6.19	−8.89	−6.82	−3.00
revolution	−17.27	−0.37	8.47	1.23	−2.37	−2.51
head-tilting	4.58	0.70	−0.61	−0.13	4.64	2.17
Gaussian Blur	−0.81	−0.88	4.26	−1.79	5.76	0.92
occlusion	5.26	−0.96	−6.41	−3.02	−6.73	−1.06
rev. occlusion	1.84	0.82	2.90	3.84	4.91	2.69
pixelisation	−13.49	−0.03	−12.04	7.86	−3.93	−4.38
compression	13.61	−0.83	−19.47	−14.57	−3.99	−3.98
white noise	−17.74	5.74	10.99	13.49	2.71	3.33
average	−1.29	0.27	−2.14	0.99	0.28	

Table 7 Probability (%), sensitivity (TPR), specificity (SPC), accuracy (ACC) and MCC values for feature usage as recognition justification

	(%)		(%)	TPR	SPC	ACC	MCC
holistic	64.29	gender	35.86	0.50	0.99	0.98	0.38
		age	2.33	0.75	0.96	0.94	0.71
		face/head	22.89	0.86	0.93	0.92	0.68
		skin	3.21	0.60	0.90	0.86	0.43
upper face	25.73	hair	23.03	0.74	0.89	0.87	0.54
		forehead	2.70	0.86	0.93	0.92	0.75
mid face	24.28	eyebrows	4.66	0.85	0.78	0.80	0.53
		eyes	3.94	0.91	0.79	0.81	0.59
		glasses	2.26	0.33	1.00	0.94	0.56
		ears	3.43	0.93	0.67	0.74	0.54
		nose	9.99	0.96	0.85	0.88	0.71
lower face	16.11	cheeks	0.36	1.00	0.50	0.60	0.41
		beard/mustache	4.52	1.00	0.89	0.90	0.64
		mouth	2.26	0.89	0.82	0.84	0.66
		chin/jaw	6.05	0.86	0.88	0.88	0.53
		neck	2.92	0.80	0.71	0.73	0.36
other	2.33	shoulders	0.15	1.00	0.00	0.50	—
		clothes	2.19	0.50	0.86	0.83	0.24

Features are grouped per type/region and sub-region.

role as decision factor, being mentioned in almost 1/4 of the answers. Allusions to the forehead were also mostly related to hair-to-skin boundaries, and if we group them as 'upper face' we cover 25.73% of the answers. The second most used region was 'mid-face', whose observation aided on justifying 24.28% of the identifications. Here, periocular information was the most used (10.86%), closely followed by the nose information. From the lower face, the most mentioned feature is a mix of the chin/jaw shape and the texture (the presence of facial hair).

When balancing positive and negative identifications through MCC, we can see how the mid-face is the less deceiving area. For the holistic features, age was the most effective recognition factor. As most of the database participants are young adults (academic students), the ones older than them (academic staff) are easily spotted.

3.3 Positive against negative identification

The degree to which we can rely on positive identification changes significantly when the decision environment degrades. To illustrate a poor decision environment, we computed entropy η as the single feature for subject identification over images acquired at the first three levels of illumination angle and subject revolution. Let I be an image in this set and x_i a pixel intensity level on the [0, 255] interval. Using histogram counts to estimate its relative frequency $P(x_i)$, the global image entropy is given by (10). Attending to probability densities (Fig. 9), we can verify how that constitutes a poor decision environment for any Bayesian classifier to perform positive identification, as functions overlap

$$\eta(I) = \sum_i P(x_i) \log_2 P(x_i) \tag{10}$$

Yet, assuming a null hypothesis H_0 corresponding to the genuine matches, and H_1 to the impostors, we can use the Neyman-Pearson statistical test [18] to optimise the classification decision in function of a threshold λ (11)

$$S = \begin{cases} 0, & \text{if } P(S|H_1) > \lambda P(S|H_0) \\ 1, & \text{if } P(S|H_0) \leq \lambda P(S|H_1) \end{cases} \tag{11}$$

Class density distributions $P(S/H_0)$ was estimated through $\hat{i}$ for positive identification (12), and $\hat{\hat{i}}$ on the negative approach (13), from class predictions ω_i

$$\hat{i} = \arg_i \max P(\omega_i|\eta) \tag{12}$$

$$\hat{\hat{i}} = \arg_i \min P(\omega_i|\eta) \tag{13}$$

Computing Bayes error rate (14) for both identification modes at varying λs, we obtain the receiver operating characteristic (ROC) curve at Fig. 10. This graphic we illustrate the performance of both identification modes by plotting the true positive rate against the false positive rate for various λ-values. A point closer to the origin (0, 0) corresponds to an higher λ-value and, consequently, a more restrictive system. We can see that relaxing the parameter λ makes true positives increase at an higher rate on negative identification than on positive identification. In the latter, true positive never gets over 0.02, which is understandable since we are using image entropy as the single feature. Nonetheless, we can see how such a poor decision environment built form a single feature, which do not provide enough information to attain positive identification,

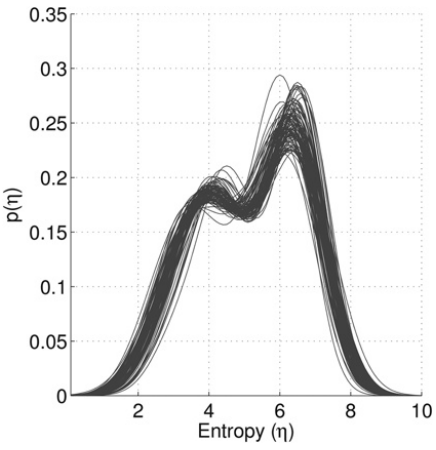


Fig. 9 *Probability density function for entropy values (η) on all subjects on the dataset*

doi: 10.1049/iet-bmt.2014.0045

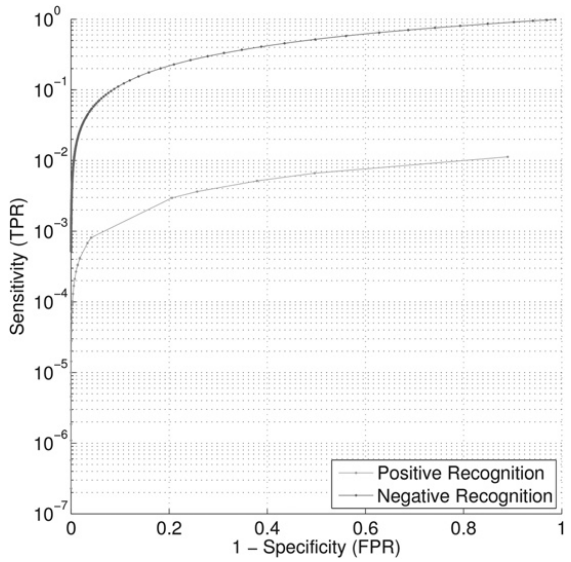


Fig. 10 *ROC curve for positive and negative identifications, using Neyman-Pearson criterium with different λ values*

still allows reliable negative identification

$$P(\text{error}|S) = \sum_{\omega_i \neq \omega_{\max}} \int_{\eta \in H_i} P(\eta|\omega_i) p(\omega_i)\, \mathrm{d}\eta \qquad (14)$$

4 Conclusions

This paper introduced the BioHDD, a new multi-session dataset of heavily images, with two singularities that turn it suitable for evaluating biometric recognition methods in extremely degraded data: (1) it contains a set of profile and frontal mugshots from 101 subjects, simulating good quality enrolment data; (2) it contains large sets of probes degraded under combinations of ten types of noise factors, resulting in images that are extremely hard to classify.

Further, we conducted an extensive on-line survey on the BioHDD data. Participants were asked to positively/negatively identify probes against the enrolled identities, along with a description of the major features used in their responses. The analysis of identification performance showed that humans have no issues cooping with inadequate illumination intensity and moderate levels of occlusions. Also, a notable ability to cope with low-resolution and compressed images was observed, suggesting that humans mostly rely on global features for identification tasks. On the other side, probes with subjects looking straight up or down and higher levels of occlusion were found to be stressful elements. That is probably the most concerning issue, as subjects trying to avoid detection 'in the wild' are more likely to be caring headgear or facing down, away from visible cameras.

A second level analysis was carried out on the justifications that participants gave for their responses: we concluded that high-frequency information, although not latent to the identification process, is taken into account when looking for specific attributes than can support their decisions. In both cases, shape related cues were the most accounted for, and also the more reliable. On the other side, texture information was rarely indicated as a decisive element. Holistic features, although not the more reliable ones, were also used on most justifications. From the identified features, the more reliable were the ones located on the mid-face: periocular features, the nose and the ears.

As further lines of work, authors plan to: (1) extend the acquisition setup in order to make it even more complete at mimicking 'in the wild' conditions (e.g. complement it with different light source angles); (2) expand the BioHDD dataset with a larger amount of participants, increasing even more the statistical significance of the dataset. Such improvements will be made available at the database website.

5 Acknowledgments

The authors would like to acknowledge the financial support provided by *FCT – Fundação para a Ciência e Tecnologia* through the research grant SFRH/BD/80182/2011, and the funding from 'FEDER – QREN – Type 4.1 – *Formação Avançada*', subsidised by the European Social Fund and by Portuguese funds through 'MCTES'.

6 References

1 Jain, A.K., Pankanti, S., Prabhakar, S., Hong, L., Ross, A.: 'Biometrics: a grand challenge'. Proc. 17th Int. Conf. on Pattern Recognition (ICPR), 2004, vol. 2, pp. 935–942
2 Ricanek, K., Savvides, M., Woodard, D.L., Dozier, G.: 'Unconstrained biometric identification: Emerging technologies', *Computer*, 2010, **43**, (2), pp. 56–62
3 Matey, J.R., Naroditsky, O., Hanna, K., *et al.*: 'Iris on the move: acquisition of images for iris recognition in less constrained environments'. Proc. IEEE, 2006, vol. 94, pp. 1936–1947
4 Hu, W., Tan, T., Wang, L., Maybank, S.: 'A survey on visual surveillance of object motion and behaviors', *IEEE Trans. Syst. Man Cybern. C, Appl. Rev.* 2004, **34**, (3), pp. 334–352
5 Haritaoglu, I., Harwood, D., Davis, L.S.: 'W4: real-time surveillance of people and their activities', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, **22**, (8), pp. 809–830
6 Kamgar-Parsi, B., Lawson, W., Kamgar-Parsi, B.: 'Toward development of a face recognition system for watchlist surveillance', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011, **33**, (10), pp. 1925–1933
7 Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: 'Face recognition: a literature survey', *ACM Comput. Surv.*, 2003, **35**, (4), pp. 399–458
8 Sinha, P., Balas, B., Ostrovsky, Y., Russell, R.: 'Face recognition by humans: nineteen results all computer vision researchers should know about', *Proc. IEEE*, 2006, **94**, (11), pp. 1948–1962
9 Gross, R.: 'Face databases' (Springer Verlag, 2005)
10 Phillips, P.J., Wechsler, H., Huang, J.S., Rauss, P.J.: 'The FERET database and evaluation procedure for face recognition algorithms', *Image Vis. Comput.*, 1998, **16**, (5), pp. 295–306
11 Sim, T., Baker, S., Bsat, M.: 'The CMU pose, illumination and expression database', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2003, **25**, (12), pp. 1615–1618
12 Gao, W., Cao, B., Shan, S., Zhou, D., Zhang, X., Zhao, D.: 'The cas-peal large-scale chinese face database and baseline evaluations'. Technical Report JDL-TR-04-FR-001, ICT-ISVISION, Chinese Academy of Sciences, May 2004
13 Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: 'Multi-pie'. FG '08. Eighth IEEE Int. Conf. on Automatic Face Gesture Recognition, September 2008, pp. 1–8
14 Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: 'Labeled faces in the wild: a database for studying face recognition in unconstrained environments'. Technical Report 07-49, University of Massachusetts, Amherst, October 2007
15 Matthews, B.W.: 'Comparison of the predicted and the observed secondary structure of t4 phage lysozyme', 1975, **405**, pp. 442–451
16 Yager, N., Dunstone, T.: 'Worms, chameleons, phantoms and doves: new additions to the biometric menagerie'. IEEE Workshop on Automatic Identification Advanced Technologies, 2007, pp. 1–6
17 Yager, N., Dunstone, T.: 'The biometric menagerie', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2010, **32**, (2), pp. 220–230
18 Neyman, J., Pearson, E.S.: 'On the problem of the most efficient tests of statistical hypotheses', *Philos. Trans. R. Soc. London. A, Containing Pap. Math. Phys. Charact.*, 1933, **231**, pp. 289–337

Chapter 12

A Dual-Step Approach to Head Landmark Detection In The Wild

12.1 Overview

This chapter consists of the following article:

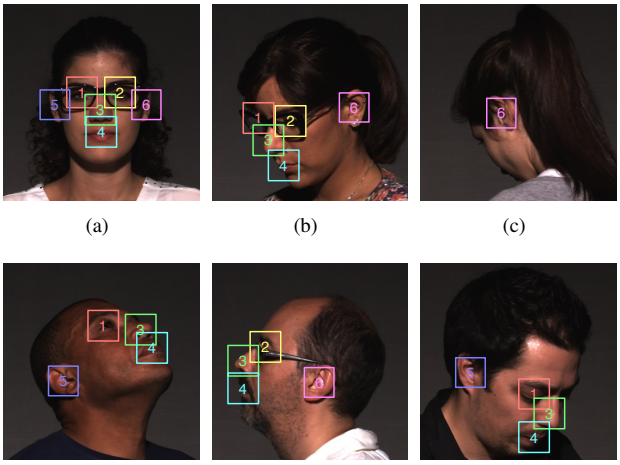
A Dual-Step Approach to Head Landmark Detection In The Wild Gil Santos and Hugo Proença Submitted for the 8^{th} *IAPR International Conference on Biometrics - ICB-2015*, May 19-22, 2015, Phuket, Thailand

A Dual-Step Approach to Head Landmark Detection In The Wild

Anonymous ICB 2015 submission

Abstract

The correct determination of facial and head landmarks is of significant value in multiple computer vision domains. In this paper a novel landmark detection approach is proposed, capable of identifying the presence of six key-elements of the human head and pinpoint their location regardless of the image acquisition angle or head's pitch. For this purpose, a dual-step approach is used: 1) a pixel-level statistical analysis assigns each image pixel into one of seven primary components of a typical head photo-shot; 2) head landmarks are detected by fusing HOG-based features extracted from both the acquired image and the output from the first stage. Finally, performance evaluation of the proposed method is carried out over surveillance-like data, where subjects' heads are imaged over multiple angles covering a full 360 degree view and participants exhibiting full head pitch range.



(a) (b) (c)

(d) (e) (f)

Figure 1. Illustration of the optimal output for our method. The detected landmarks are 1) right eye, 2) left eye, 3) nose, 4) mouth, 5) right ear, and 6) left ear.

1. Introduction

The human face is of the most importance for recognition, interaction and communication purposes. Being able to accurately detect facial landmarks is a challenging goal with many applications in the field of computer vision, such as head pose estimation, expression analysis and face alignment for biometric recognition.

The pioneering research focused on facial landmarks used elastic graphs matching for the detection of facial images in cluttered scenes [11], and compensate for large rotations in depth while carrying on face recognition [16]. Later on the Active Appearance Models (AAM) algorithm was introduced [3], which is actively used nowadays on face matching and tracking, with further improvements to learn appearance variations of a set of templates [4]. More recently, different approaches address the issues inherent to unconstrained scenarios, either by using tree-structured models to effectively handle deformation, combining face detection, pose estimation and landmark localization [19], SIFT-based face part detectors [1], or regression-based models to achieve face alignment [2].

From the point of view of unconstrained biometrics, where efforts are being held into extending robust recognition to *in the wild* scenarios, landmark detection is commonly used in early stages in order to achieve proper face alignment. Techniques developed for such purpose seek to find the correct location of the eye's center and corners, eyebrow corners, nose tip, mouth corners, etc. However, an effective head landmark detection technique can be put to further uses. Being able to identify which are the visible landmarks can help decide on the best recognition technique to use. In fact, being able to tell to which extent a particular biometric trait is reliable or not, based on its visibility, can be more interesting that having a close estimation of the head's pose, or conducting facial alignment. To this purpose, the present work aims at detecting six landmarks in particular: the left and right eyes, with the corresponding periocular region; the left and right ears; the nose; and the mouth.

The head landmarking problem is particularly challenging over unconstrained, noisy or degraded data, as not only the head is a moving imaging target, as it rotates over three axes: pitch, roll and yaw. Furthermore, on most video-surveillance scenarios it can be imaged from any direction, and most of the proposed facial landmark detection methods require a frontal or near-frontal imaging of the subject

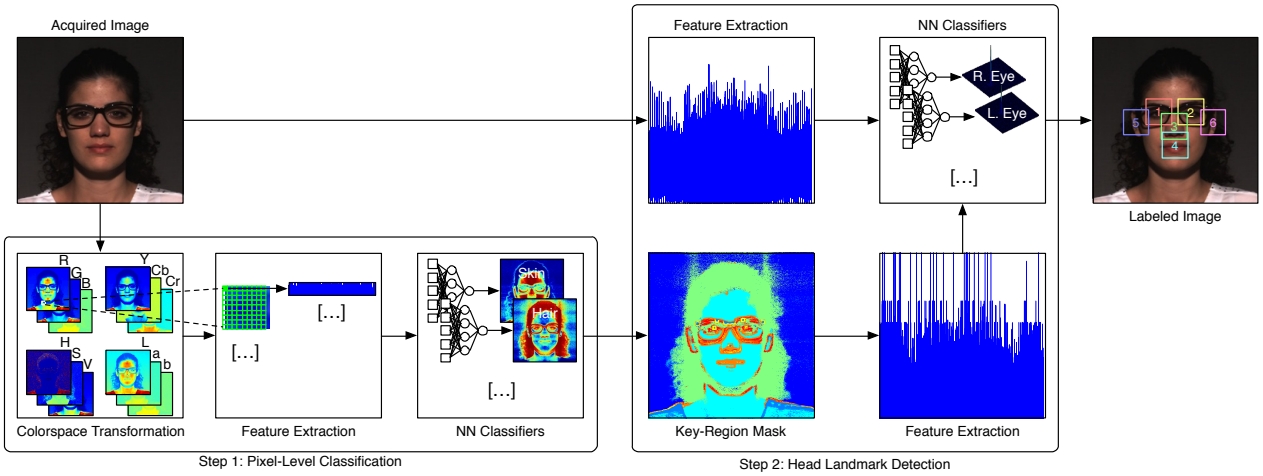


Figure 2. Illustrative diagram of the proposed method. Apart from the acquired and labeled photos, images were scaled and mapped to a different color-map for better perception.

to properly function. The technique presented in this paper allows identifying which ones of the six main head landmarks are visible (*i.e.* eyes, nose, mouth, and ears) and their location in the head image, regardless of the imaging angle and head-tilting – Figure 1. To achieve that goal, a two-step approach is followed. At first, a pixel-level classification assigns each one of the image pixels into one of seven key-regions representing the main elements of a head photo-shoot: skin, hair, eyes, mouth, clothes, accessories and background. To do so, a set of Neural-Networks (NNs) was used, working over a group of color and textural features. On a second stage, a Histogram of Oriented Gradients (HOG) analysis combining information from both the acquired image and the newly obtained pixel-level classification map is fed to another set of NNs, resulting in the final landmark detection output.

The remainder of this paper is organized as follows: Section 2 describes the proposed method, with details on both its pixel-level classification and landmark detection steps; Section 3 details the experimental procedure, starting by reporting the utilized dataset, and followed by the method results. Finally, Section 4 contains some final considerations, and further lines of work.

2. Proposed Method

The landmark detection technique proposed in this paper consists in two main steps – Figure 2: 1) pixel-level classification between seven predominant regions; 2) landmark detection combining both the acquired image and the pixel-level classification.

For the purpose of this work the problem of head detection is assumed to be already solved, as literature provides a considerable number of techniques [17, 18], among which the widely implemented Viola and Jones detector [15].

2.1. Pixel-level Classification

At the pixel-level classification stage, each one of the pixels is individually classified into one of seven key-regions found to be the predominant ones in head images: skin, hair, eyes, mouth, clothes, accessories and, for the remaining pixels, background. *Eye* pixels include the visible elements of the ocular globe (*i.e.* sclera, iris and pupil), and mouth pixels comprise the lips and, occasionally, other elements from the human mouth made visible due to subject's expression (*e.g.* teeth). For such purpose, a set of features was extracted and fed to a NN for classification. The selected features were chosen for their simplicity, and computation ease, aiming at establishing a fast technique able to work on real time.

As some identified key-regions were visually more distinct over different color-spaces, four different color representations were used for feature extraction: RGB, Hue-Saturation-Value (HSV), YCbCr and CIE Lab. The intensity values in the RGB color-space were directly red from the database images, and used as reference to compute the remainder. The HSV color-space was firstly introduced by Joblove and Greenberg [9] to represent color in a more intuitive and perceptually relevant manner, by mapping the color values to a cylinder: the "hue" value corresponds to the angle around the cylinder axis; "saturation" the distance from that same axis; and finally the "value" is the height, representing the perceived luminance having saturation as reference. YCbCr is a color-space more commonly used in digital video systems, and its three components means respectively the luminance (Y) and the blue-difference chroma (Cb) and red-difference chroma (Cr). Conversion was attained based on ITU-R recommendation BT.601 [8]. The Commission Internationale de l'Eclairage (CIE) 1976 ($L^*a^*b^*$) color space, also known as CIE Lab, is

partially uniform representation, based on the human visual system and allowing absolute color characterization. Since proper chromatic adaptation matrices were not disclosed along with the used database, the transform was based on the ITU-R recommendation BT.709 [7], using as reference the CIE D65 illuminant [14].

For each pixel, and besides its own intensity, a set of features were computed for each one of the color representations as follows. At first, the entropy was used to encode the image texture using its randomness [6]: let I be an image from the dataset, and l_i a pixel intensity level on the $[0.255]$ interval; using histogram counts to estimate its relative frequency $P(l_i)$, the entropy value η is given through (1).

$$\eta = \sum_{i=0}^{255} P(l_i) \log_2 P(l_i) \tag{1}$$

This metric was computed for 8-, 35- and 224-neighbor windows. For the same neighborhoods of N pixels x_i were also computed the average μ, standard deviation σ and range (maximum value - minimum value), asymmetry α (2) and kurtosis k (3).

$$\alpha = \frac{1}{N} \frac{\sum_{i=1}^{N} (x_i - \mu)^3}{\sigma} \tag{2}$$

$$k = \frac{P(x_i - \mu)^4}{\sigma^4} \tag{3}$$

Finally, two more well known distribution-based descriptors were used HOG and Uniform Local Binary Patterns (ULBP). The HOG descriptor [5], widely applied on computer vision, computes the gradient orientation by filtering the image with two kernels: $[-1, 0, 1]$ and $[-1, 0, 1]^T$. The ULBP descriptor differs from the regular Local Binary Patterns (LBP) as it achieves "improved rotation invariance with uniform patterns and finer quantization of the angular space" [12]. Instead of the 2^n possible binary patterns outputted from the regular LBP over an 8-neighbor region, a uniformity measure U is calculated representing the number of bitwise changes in that same pattern (4). This measure can only assume 59 distinct values.

$$\begin{aligned} \text{ULBP} = &|\text{sgn}(x_7 - x_i) - \text{sgn}(x_0 - x_i)| + \\ &+ \sum_{n=1}^{7} |\text{sgn}(x_n - x_i) - \text{sgn}(x_{n-1} - x_i)| \end{aligned} \tag{4}$$

As we can see, from this analysis many features were extracted, resulting in a classification problem with very large dimensionality. As so, we choose to conduct dimensionality reduction using the Principal Component Analysis (PCA) technique [10], prior to the training of the NNs. NN-based methods have been widely applied on classification problems, for their learning abilities and good generalization. For the purpose of this work, a NN was trained to identify each one of the identified key-regions. The architecture of each NN was as follows: the hidden layer had fifteen neurons, half the number of features resulting from the dimensionality reduction stage, and the final (output) layer had a single neuron, since we are dealing with binary classification problems. At the classification phase of this stage, each pixel is mapped to the key-region whose NN outputted the greatest value in terms of cumulative density.

2.2. Landmark Detection

The second-stage combines textural information from both the acquired image and the pixel-level classification output to produce the final decision about the presence of head-landmarks and their location.

At first, the acquired image is divided into multiple overlapping spatial blocks. Those blocks constitute the set of detection windows that will be matched against each one of the head landmarks. Then, the information from each one of those windows is encoded, along with the correspondent output from the pixel-level classification stage, simply by using the HOG-based encoding technique detailed on the previous section.The single use of HOG-based feature encoding was preferred over more complex techniques, so that this procedure could to be carried over a large amount of detection windows with a lower impact on the computational cost.

Finally, those features fusing information from both the acquired image and the output from the previous stage is fed to a set NNs, previously trained to identify each one of the landmarks. For this second stage, each NN has two hidden layers: a first hidden layer with 100 neurons, and a second one with ten. Once again, and since each NN was trained to fit a single landmark, we are facing a binary classification problem and a single output neuron was used. The output of each NNs was analyzed against its cumulative density function to produce the final detection decision.

3. Experimental Results

To the best of our knowledge and despite the literature cited in the introductory section, from the published research none is aimed at the same goals our method is.

3.1. Dataset

For the purpose of this work, we found that the most fit dataset was the BioHDD [13]. The BioHDD is a database built to test biometric methods against extremely degraded data, gathering head images from over one hundred participants in a full 360° view. This dataset provides images acquired in both the near-infrared (NIR) and visible wavelength (VW), but only the latter were used.

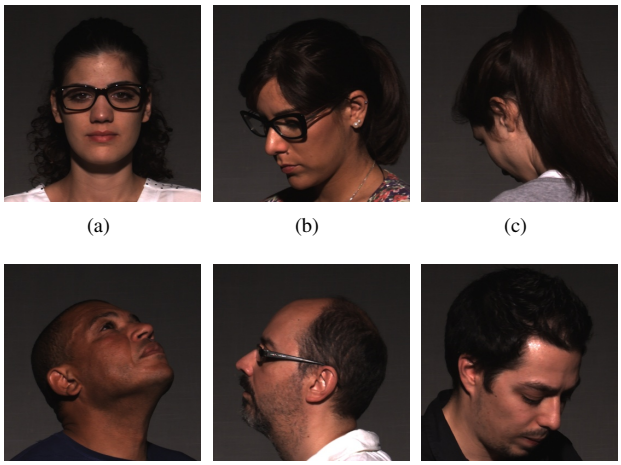
Table 1. Confusion matrix for the results obtained at the pixel-level classification stage. Performance values for the three top-performing classifiers are outlined in bold.

Actual Class \ Predicted Class	Skin	Hair	Eyes	Mouth	Clothes	Accessoires	Background	Total	Sensitivity
Skin	1 886 583	141 688	31 298	361 551	115 075	39 893	53 905	2 629 993	**0.717**
Hair	370 676	5 261 177	27 794	208 666	246 236	168 670	242 134	6 525 353	**0.806**
Eyes	78	406	2 046	1 820	863	1 372	20	6 605	0.310
Mouth	6 584	2 096	739	21 005	1 801	1 752	365	34 342	0.612
Clothes	47 362	208 181	10 057	78 840	423 049	53 885	74 181	895 555	0.472
Accessoires	591	17 356	3 702	6 244	24 125	21 505	2 144	75 667	0.284
Background	21 554	334 266	3 407	45 338	147 108	21 554	7 259 258	7 832 485	**0.927**
Total	2 333 428	5 965 170	79 043	723 464	958 257	308 631	7 632 007	18 000 000	
Precision	**0.809**	**0.882**	0.026	0.029	0.441	0.070	**0.951**		

Although ten different types of noise are present in the dataset, we have chosen a subset where variations occur over two axes: 1) imaging angle – subjects' heads were imaged from different angles, covering all $360°$ degrees at $45°$ steps; 2) head-tilting – for each one of the imaging angles, three variations were observable, with subjects were facing forward, or looking up or down. As we can see from Figure 3, there is a considerable variability among the used data. For the sake of clarification, the $0°$ angle corresponds to the subject being imaged frontally. For illustrative purposes, sample images from the BioHDD subset used in our experiments are present at Figure 3.

3.2. Evaluation Metrics

To quantify the performance of the proposed method, the following metrics were used: for the first stage (pixel-level classification), the confusion matrix were computed, along with the sensitivity and precision for each key-region; for the second stage (and global method performance) performance was assessed through Receiver Operating Characteristic (ROC), Area Under Curve (AUC) and Equal-Error Rate (EER).



(a) (b) (c)

(d) (e) (f)

Figure 3. Sample images from the BioHDD subset used in our experiments: subjects' head imaged from different angles, with participants facing forward, looking up or down.

The confusion matrix summarizes the performance of the classifiers, counting the matches that were observed between each one of the actual classes (rows) and the predicted values (columns). That allows assessing the amount of positive pixels that were actually classified as such (sensitivity) and identify which classes are more commonly mistaken. The precision, on the other side, gives us the proportion of positively identified classes that actually were genuine.

The ROC curve relates the sensitivity, or True Positive Rate (TPR) with the False Positive Rate (FPR). Based in that plot, the AUC can be perceived as a quantification of how well pairwise comparisons are performed on a binary classification problem. On the perfect scenario, all positive matches are ranked higher and the negative ones, and the AUC equals one. Finally, setting the acceptance threshold so that the probability of false acceptance equals the probability of false rejection, we obtain the EER.

During the experimental procedure all results were verified using ten-fold cross-validation: 90% of the data was randomly selected for training, being the testing conducted on the remaining 10%. This procedure was repeated ten times.

3.3. Pixel-Level Classification

To train the classifiers for this stage, a subset of fifty images from the BioHDD database was used. Although only fifty images were used, each one has 600×600 pixels representing a total of eighteen million potential instances spread between the seven classes.

In order to establish the ground-truth, images were manually segmented, assigning each pixel to one of the established key-regions. However, as the number of positive and negative instances of each key-class were significantly unbalanced in particular for the smaller key-regions (*e.g.* eyes), we have chosen to use all positive instances for training, and randomly choose double that amount of negative

instances.

For each selected pixel, features were then encoded as detailed in Section 2.1. The NN classifiers results over the 10-fold cross-validation are present at Table 1 and illustrated at Figure 4. For a better perception of the achieved results we choose to display the same images that were previously used to illustrate the dataset (Figure 3).

Analyzing the sensitivity values for each class, we can see how that performance metrics appears to be directly proportional to the total of pixels of each key-region, with exception of the clothes and accessories. That leads to theorize that the low performance of some classifiers were possibly conditioned by an insufficient number of training samples, as the ones with better performance were also the ones with a larger training set. As for the two exceptions, visually inspecting samples from those classes we observe that they display a much more significant variability in textural information between the dataset participants. Without making any considerations about this classification stage impact over the final method's performance, there are three key-regions satisfactorily classified over this step: background, hair and skin, which is further corroborated by the results illustrated at Figure 4.

Attending to the precision of each classifier, we can observe how the eyes, mouth and accessories are effectively problematic classes, as less that 10% of the pixels actually classified as such were genuine positives. For the pixels classified as eyes, 40% actually belong to the skin region, and 35% are hair. A possible explanation is the fact that the eye region is relatively small, and the skin and hair are its direct neighbors. As some features were extracted over relatively large widows, that could have misled the classifiers. A similar conclusion can be drawn for the mouth and accessories regions.

There is also a significant confusion between the clothes, background and hair regions. That is perceptible not only by analyzing the misclassified pixel count, but also by visually inspecting the resulting region maps at Figure 4. The BioHDD dataset, although intended to simulate unconstrained conditions, were acquired on a controlled environment, where the background of each image was a regular white sheet of fabric. As so, those regions can be assumed to have enough similarities to mislead the classifiers.

Even if not representing an optimal classification solution for all the regions *per se*, the output from the pixel-level classification stage represents a significant contribution to the overall performance of our proposed method, as shown on the next section.

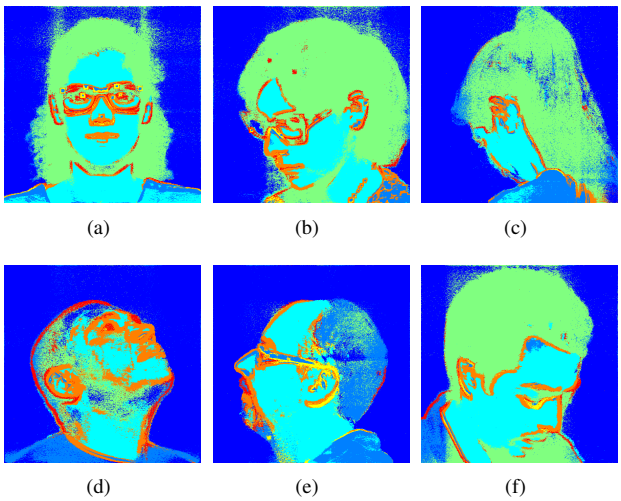


Figure 4. Sample output from the pixel-level classification stage. For a better perception of the achieved results, the images being shown are the same images previously used for dataset illustration (Figure 3).

3.4. Landmark Detection

To conduct the second and final stage of our experiments, we randomly picked over a thousand different images from the selected BioHDD subset (excluding the ones used for pixel-level classification training and performance assessment). Each one of those images was then manually annotated to provide ground-truth information about the visible landmarks and their precise locations. To delimit the landmarks a square 120×120 pixel region was used and, regardless of the imaging angle and head pitch, that region was always unrotated and centered on the landmark. That dimension was selected based on experimentation and attending to the dataset characteristics, aiming at fitting the whole landmark without including unnecessary surroundings.

To train each one of the classifiers, one genuine landmark and two impostors were extracted from each image. Those two negative samples correspond to non-overlapping regions of the same dimensions of the landmarks, randomly located in the same database image. At the evaluation stage, a detected landmark is considered a true positive if the real landmark center is located within the detection window. Likewise, any landmark left undetected is considered to be a false negative. Once again, experiments were conducted over a 10-fold cross-validation. For comparison purposes, along with the feature-level fusion from the acquired image with the output from the pixel-level classification stage, different classifiers were trained using the features from each one of those sources separately. The performance assessments are present at Figure 5 and Table 3. The impact of the imaging angle and head pitch over the feature-level fusion classifier are presented at Tables 2 and 4 respectively.

As we can see all landmarks exhibit relatively good detection performance, considering the demanding characteristics of the dataset: once again, we are stressing the pro-

Table 2. Detection performance of the proposed method, accordingly to the subject imaging angle. The displayed performance metrics are Area Under Curve (AUC) and Equal-Error Rate (EER). Cells with empty values represent angles where that landmark was neither detected nor visible. Top values for each landmark are outlined in bold.

Imaging Angle →	0°		±45°		±90°		±135°		180°	
	AUC	EER	AUC	EER	AUC	EER	AUC	EER	AUC	EER
Right Eye	**0.948**	**0.114**	0.865	0.183	0.782	0.264	0.801	0.268	–	–
Left Eye	**0.913**	**0.160**	0.854	0.207	0.799	0.260	0.838	0.245	–	–
Nose	**0.850**	**0.206**	0.841	0.223	0.816	0.239	0.828	0.222	–	–
Mouth	**0.919**	**0.132**	0.872	0.172	0.806	0.254	0.306	0.695	–	–
Right Ear	0.839	0.239	0.935	0.139	0.920	0.156	**0.947**	**0.099**	0.826	0.235
Left Ear	0.831	0.243	0.907	0.154	**0.929**	0.156	0.920	**0.114**	0.732	0.313

posed method against images where the subject could have been imaged from a wide variety of angles on a full $360°$ range and with variable head pitch.

At Table 3 we can observe how the features extracted from the subject as it was imaged appear to be more discriminative than when using only the features from the pixel-level classification output. The only exceptions are the eye landmarks. Remarkably, the eye region was one of the key-elements of the image that presented significant classification issues during the first stage. Nonetheless, feature-level fusion is always preferable: as we can see from the ROC curves the fusion classifier is able to obtain higher true positive rates (sensitivity) with lower false positive rates (leftmost part of the each plot). Top performed is observed for the ear landmarks, with the right ear achieving an AUC of 0.910 and 0.155 EER. Right after that, the eyes landmarks attained AUC of 0.870 and 0.858, being the third and fourth top performing landmarks. Finally, the nose and mouth were the landmarks with lower performance: 0.835 and 0.853 for AUC, and 0.228 and 0.203 EER respectively.

Analyzing the landmark detection performance as function of the imaging angle (Table 2), we can observe that frontal imaging of the subject lead to better performance on detecting the right and left eyes, nose and mouth landmarks. At that imaging angle, the average of the AUC values observed for the eyes landmarks was 0.931, followed by the mouth with 0.919 and the nose with 0.850. The detection performance for the ears landmarks achieved even higher performance values, with an average top AUC of 0.938. However, those two landmarks performed better as the subject was imaged over higher rotation angles (*i.e.* sideways). When the subject is completely tuned by $180°$, facing away from the camera, and despite four of the six head landmarks not being visible, the ear landmarks are still detected with relatively good performance. On those latter conditions, 0.826 AUC was observed for the right ear, with an EER of 0.235. Attending the facial landmark detection techniques in the literature, we can consider that from the stressed imaging amplitude range the most adverse to be located between $\pm 135°$ and $180°$. As we saw, the proposed method is able to cope with such conditions with satisfactory performance.

Finally, from Table 4 we can see how our method is relatively stable to variations on head's pitch, with the detected landmarks registering standard deviations not higher than 0.026 for the AUC and 0.031 for the EER among all the stressed pitch ranges.

Table 3. Detection performance for each individual landmark, when features were used from the acquired image, from the pixel-level classification stage output, and from their fusion. The displayed performance metrics are Area Under Curve (AUC) and Equal-Error Rate (EER). Top results for each landmark are outlined in bold.

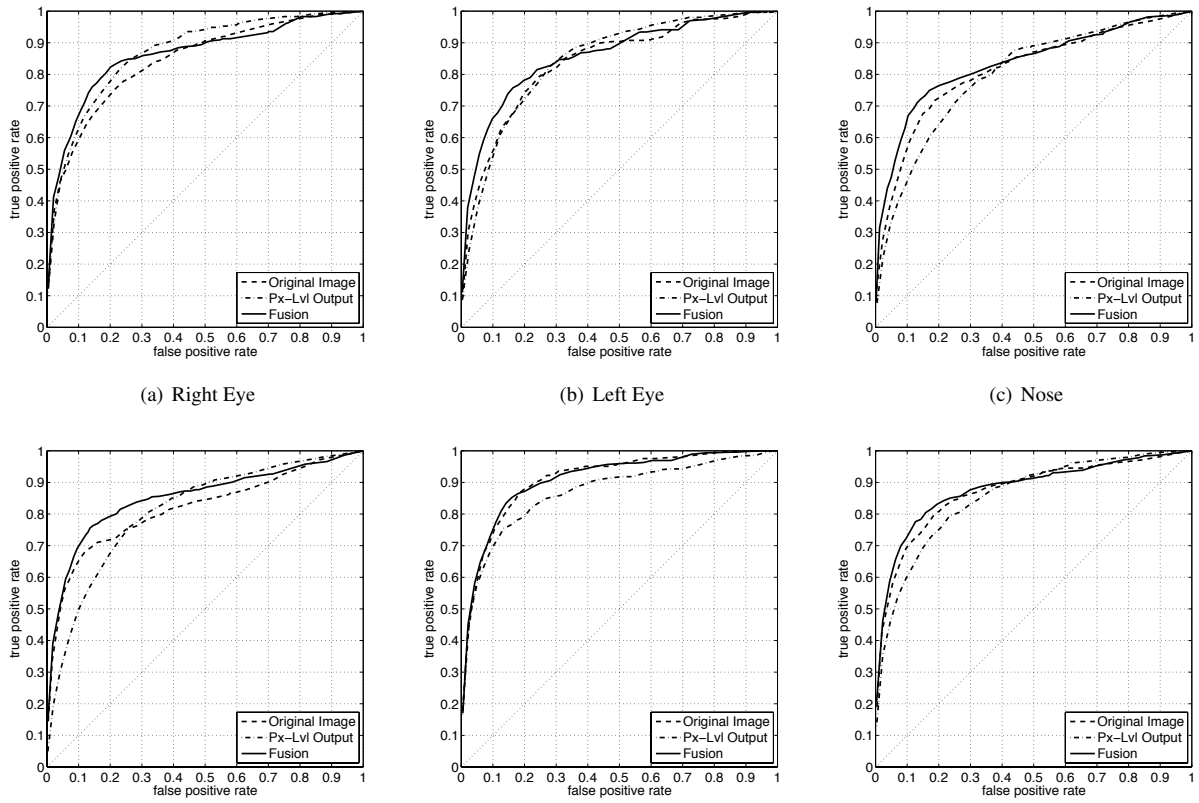
	Acquired Image		Px-Lvl Output		Fusion	
	AUC	EER	AUC	EER	AUC	EER
Right Eye	0.843	0.228	**0.870**	0.210	0.863	**0.190**
Left Eye	0.839	0.229	0.845	0.235	**0.858**	**0.214**
Nose	0.815	0.246	0.799	0.271	**0.835**	**0.228**
Mouth	0.819	0.249	0.811	0.254	**0.853**	**0.203**
Right Ear	**0.910**	0.161	0.870	0.204	0.909	**0.155**
Left Ear	0.872	0.195	0.855	0.224	**0.881**	**0.182**

4. Final Considerations

In this paper we propose a novel algorithm for head landmark detection, capable of identifying and pinpoint their location of six key-elements of the human head: right and left eyes, nose, mouth, and right and left ears. The presented technique solution is composed of two major phases: 1) a group of locally extracted features classifies each im-

Table 4. Detection performance of the proposed method, accordingly to the subjects' head pitch. The displayed performance metrics are Area Under Curve (AUC) and Equal-Error Rate (EER).

Pitch →	Facing Forward		Looking Up		Looking Down	
	AUC	EER	AUC	EER	AUC	EER
Right Eye	0.857	0.195	0.891	0.159	0.844	0.211
Left Eye	0.831	0.240	0.883	0.179	0.865	0.218
Nose	0.802	0.254	0.844	0.215	0.855	0.206
Mouth	0.838	0.231	0.854	0.191	0.867	0.173
Right Ear	0.915	0.152	0.910	0.148	0.899	0.174
Left Ear	0.869	0.185	0.904	0.146	0.868	0.188



(a) Right Eye (b) Left Eye (c) Nose

(d) Mouth (e) Right Ear (f) Left Ear

Figure 5. Receiver Operating Characteristic (ROC) curves for the detection of each individual landmark, when features were used from the acquired image, from the pixel-level (Px-Lvl) classification stage output, and from their fusion.

age pixel into one of the seven main-regions of a head photo-shot (skin, hair, eyes, mouth, clothes, accessories and background); 2) HOG features are extracted from both the acquired image and the pixel-level classification stage for landmark detection. When tested against a database of degraded data that simulates surveillance-like conditions, our method delivered relatively good performance regardless of the subject's imaging angle and his head pitch.

As future lines of work, further validation of the proposed method is intended, not only against the full range of noise factors provided by the BioHDD dataset, but different and more widely used databases as well. Another step would be the extension of the proposed method to head pose estimation, using as information the coordinates of the detected landmarks.

References

[1] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar. Localizing parts of faces using a consensus of exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 35(12):2930–2940, December 2013.

[2] X. Cao, Y. Wei, F. Wen, and J. Sun. Face alignment by explicit shape regression. *International Journal of Computer Vision*, 107(2):177–190, 2014.

[3] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In H. Burkhardt and B. Neumann, editors, *Computer Vision - ECCV'98*, volume 1407 of *Lecture Notes in Computer Science*, pages 484–498. Springer Berlin Heidelberg, 1998.

[4] D. Cristinacce and T. Cootes. Feature detection and tracking with constrained local models. In *Proceedings of the British Machine Vision Conference*, pages 929–938, Edinburgh, September 2006.

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893, June 2005.

[6] R. Gonzalez, R. Woods, and S. Eddins. *Digital Image Processing Using MATLAB*. Prentice-Hall, New Jersey, 2003.

[7] ITU-R. Parameter values for the hdtv standards for production and international programme exchange. Recommendation ITU-R BT.709, Radiocommunication Sector of International Telecommunicatio Union, April 2002.

[8] ITU-R. Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios. Recommendation ITU-R BT.601-7, Radiocommunication Sector of International Telecommunicatio Union, March 2011.

[9] G. Joblove and D. Greenberg. Color spaces for computer graphics. *Computer Graphics*, 12(3):20–25, August 1978.

[10] I. Jolliffe. *Principal Component Analysis*. Springer, 2nd edition, 2002.

[11] T. Leung, M. Burl, and P. Perona. Finding faces in cluttered scenes using random labeled graph matching. In *Proceedings of the Fifth International Conference on Computer Vision*, pages 637–644, 1995.

[12] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971 –987, July 2002.

[13] G. Santos, P. Fiadeiro, and H. Proença. Biohdd: a dataset for studying biometric identification on heavily degraded data. *IET Biometrics*, 2015.

[14] T. Smith and J. Guild. The C.I.E. colorimetric standards and their use. *Transactions of the Optical Society*, 33(3):73–134, 1931.

[15] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 2:137–154, 2004.

[16] L. Wiskott, J. Fellous, N. Kuiger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, July 1997.

[17] M.-H. Yang, D. Kriegman, and N. Ahuja. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24:34–38, January 2002.

[18] C. Zhang and Z. Zhang. *Face Detection and Adaptation*. Morgan & Claypool, 1 edition, 2010.

[19] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2879–2886, June 2012.

Chapter 13

Quis-Campi: Extending *In The Wild* Biometric Recognition to Surveillance Environments

13.1 Overview

This chapter consists of the following article:

Quis-Campi: Extending *In The Wild* Biometric Recognition to Surveillance Environments
Gil Santos, João C. Neves, Sílvio Filipe, Emanuel Grancho, Silvio Barra and Fabio Narducci
Submitted for the 8^{th} *IAPR International Conference on Biometrics – ICB-2015*, May 19-22, 2015, Phuket, Thailand

Quis-Campi: Extending *In The Wild* Biometric Recognition to Surveillance Environments

Anonymous ICB 2015 submission

Abstract

Efforts in biometrics are being held into extending robust recognition techniques to in the wild scenarios. Nonetheless, and despite being a very attractive goal, human identification in the surveillance context remains an open problem. In this paper main we introduce a novel biometric system – Quis-Campi – that effectively bridges the gap between surveillance and biometric recognition while having a minimum amount of operation restrictions. We propose a fully automated surveillance system for human recognition purposes, attained by combining human detection and tracking, further enhanced by a PTZ camera that delivers data with enough quality to perform biometric recognition. Along with the system concept, implementation details for both hardware and software modules are provided, as well as preliminary results over a real scenario.

1. Introduction

Biometrics is one of the most active fields in the area of computer vision, which is justified by our societies' increasing concern about security. Biometric systems significantly rely on the accurate extraction of individuals' distinctive features, which is conditioned by the acquisition environment and constraints. As such, the most reliable systems are deployed on controlled scenarios and count on subject cooperation. On the other side, surveillance cameras are widely deployed and can constitute a good source of input for biometric systems. Filling the gap between biometrics and visual surveillance is quite a desirable goal, allowing to produce *automata* capable of recognizing human beings *in the wild*, without their cooperation and, possibly, even without their awareness.

When moving to *in the wild* scenarios the acquisition constraints are substantially lowered and, most of the time, subject cooperation is not even expectable. In order to deal with such challenging conditions alternatives are sought over three axes [6]: 1) improve the existing algorithms so they can handle more degraded data; 2) resort to multimodal biometric systems so that the usage of multiple traits can compensate for their lack of "quality"; 3) explore new biometric traits that could better cope with this new reality. Despite the recent efforts, no system yet exists capable of dealing effectively with all the issues introduced by *in the wild* biometrics, and even those systems able to cope with less constrained conditions (e.g. the Iris On The Move project [11]) still lack an ideal level of user abstraction.

Despite the several applications of visual surveillance, most of existing surveillance systems are focused on activity recognition (e.g. W^4 project [5]), and not that many of them are prepared to handle surveillance scenarios from a watchlist approach (e.g. Kamgar-Parsi *et al.* [8]). In this paper we present a novel biometric recognition system, designed to work covertly in a non-habituated and non-attended fashion, over non-standard environments. Our main goal is to conceive a system that links together both biometrics and visual surveillance, being able to conduct biometric recognition over typical surveillance scenarios, with the minimum possible amount of operation restrictions.

The remainder of this paper is organized as follows: in Section 2 we detail the three layers of the recognition system, its operation premises and devised modules; in Section 3 we present the stressed techniques for each module, along with preliminary results of our system over a real surveillance scenario and, finally, Section 4 states some final considerations.

2. The Recognition System

The optimal recognition system would operate on any environment, thus minimizing the amount of operation restrictions. Since we are trying to bridge biometrics with the visual surveillance, we have developed our system in a typical surveillance scenario: a parking lot (Figure 3(a)). Such scenario is particularly harsh for recognition purposes, for a number of reasons: 1) it is a non-standard environment, with irregular lighting that changes not only during the day, but also accordingly to weather conditions, reflections, etc.; 2) complex background regions and the varying resolution of humans poses increasing challenges for both detection and recognition phases. 3) subjects can come from any direction, and they are rarely facing the camera which is typically placed on an upper position. Furthermore, the system

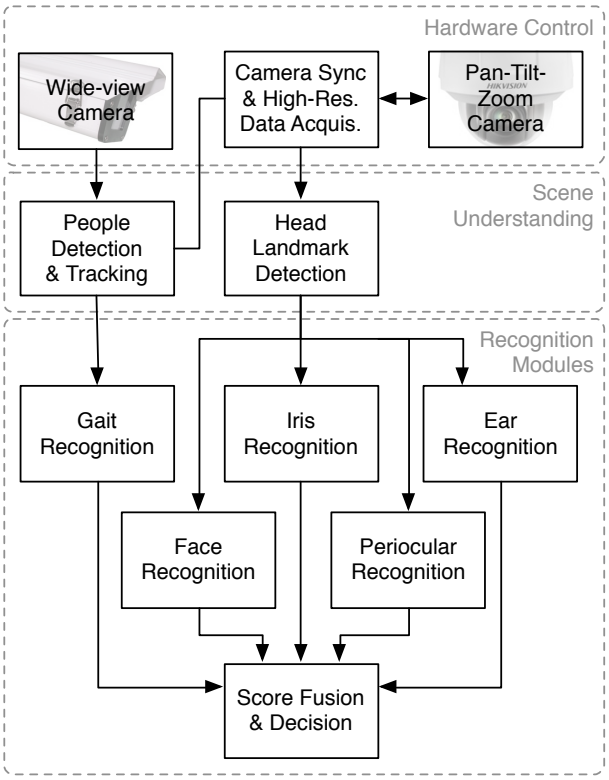
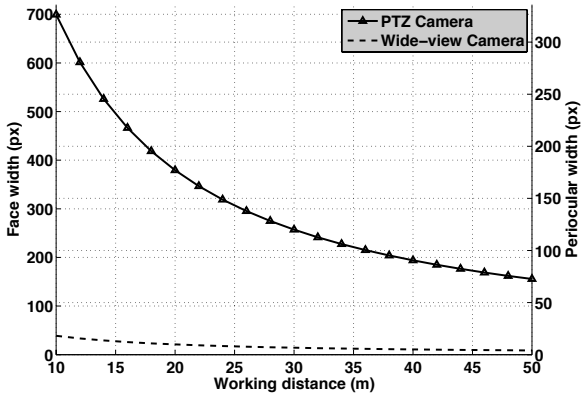
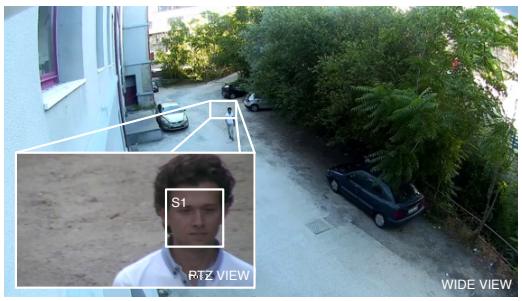


Figure 1. Working diagram of the proposed system, and the three-layer architecture: hardware control (top), scene understanding (middle) and recognition modules (bottom).



(a) Trait visibility *vs* distance



(b) Illustration of the acquired data

Figure 2. Visible face and periocular width, in pixels, as function of the system's working distance (a), and illustration of the acquired data for both cameras (b).

should run in real-time. That being said, the recognition system we propose is devised over three main layers (Figure 1): hardware control, scene understanding, and recognition modules.

2.1. Hardware Control

To mimic a surveillance scenario, a wide-view camera was mounted on the exterior wall of a building, at a first-floor level (approximately 5m above the ground), and pointing towards a parking lot. Although this kind of camera offers a more complete overview of the scenery, it does not provide enough quality for recognition methods to work at the distances the driveway ranges from (15 to 35 meters) – Figure 2(a).

To provide recognition methods with reasonable quality data, a PTZ camera was added to the system. This way, pointing and zooming over a specific region allows acquiring a detailed view of detected subjects. As we can see from Figure 2(a), there is a substantial difference in the usable face and periocular pixel area between the two cameras as a function of the working distance. A Canon VB-H710F was used as the wide-view camera, and a Hikvision DS-2DE5286-AEL as the PTZ camera, with a framerate of 30 frames per second (fps) at maximum resolution. However, the independence between cameras demands a camera synchronization module able to map coordinates from the wide-view camera referential to the Pan-Tilt-Zoom (PTZ) coordinate system, allowing the acquisition of the high-resolution view of a portion of the scene (Figure 2(b)).

2.2. Scene Understanding

At the scene understanding layer, the system has two main modules: 1) people detection and tracking; and 2) facial landmark detection.

The first module is responsible for locating the people as they enter the scene, and tracking them until they are no longer visible. It takes as input the video feed from the wide-view camera, and has three main steps: background subtraction, upper-body detection and tracking, illustrated in Figure 3.

Further to that, the second module of the scene understanding layer is applied to that closer view, identifying which facial landmarks are visible, thus deciding the weight of each recognition module. Being able to describe which facial traits are visible and where, is far more important that actually getting a close estimation of the head's pose, as we can tell to which extent the trait is reliable or not.

(a) Wide-view Feed (b) Background subtract. (c) People tracking

Figure 3. Illustration of the preliminary results obtained by the people detection and tracking module: a) sample image acquired with the wide-view camera; b) foreground regions attained by background subtraction; c) people tracking module results.

2.3. Recognition Modules

For recognition purposes the proposed system relies on a multi-modal biometric approach that combines face, iris, periocular, ear shape and gait information.

The face is not only one of the most common and widely used biometric trait, but also one of the most successful applications of image analysis and understanding. Several face recognition systems are commercially deployed, and a lot of techniques accessible [18] for both still and video images. However, as stated by Bledsoe [2], the "great variability in head rotation and tilt, lighting intensity and angle, facial expression and aging" make face recognition an extremely hard challenge. Since in surveillance scenarios the impact of such factors is even more significant, it is mandatory to rely on robust approaches (e.g. [17]).

The ocular region is one of the most explored in biometry. Iris in particular is a very popular biometric trait, delivering very high recognition accuracy under controlled environments. Although iris performance as a biometric trait being severely impacted in non-ideal setups, due essentially to its reduced size and moving profile, researchers are putting efforts in overcome those limitations. The periocular region represents a good trade-off between the whole face and the iris, being easy to acquire without user cooperation, and not requiring a constrained close capturing. As such, is one of the strongest candidates for the purposes of our system.

The shape of the ear can also be used as biometric trait, as the structure of its cartilage is unique for each individual. Despite all ear recognition methods traditionally require some degree of user cooperation, if proper alignment estimation can be established and the ear imaged with few or no occlusion, it can be used as biometric trait *in the wild*.

Gait is the only trait that will be imaged from the wide-view cam. Acquiring data about way a person walks is non-invasive, and can be done at-a-distance. The majority of the gait recognition methods in the literature do not require high-resolution data, so they can run over surveillance camera data.

3. Experimental Results

This section details the stressed techniques for each module, along with preliminary results over the selected surveillance scenario.

3.1. People Detection and Tracking

At a first stage, several state-of-the-art background subtraction techniques were evaluated. Visually inspecting their performance on our surveillance scenario, we found SOBS [10] and Mixture of Gaussians [14] to be the most robust ones over our testing data. Using the output from the background subtraction technique, we filtered the regions most consistent with human presence exploiting a simple upper body detector distributed with OpenCV using Haar feature-based cascade classifiers [16].

After dealing with detection, the tracking phase is initialized in order to find the correspondence between the same subjects in consecutive frames. Different features such as motion, appearance and shape can be exploited for that purpose. Considering the specifications of our scenario, we have chosen to use motion and appearance features, whereas shape was not considered due to high variations caused by viewpoint. We observed that although maintaining their exterior looking while passing through the scene, dynamic lighting and shadow interference perturb persons' appearance. On the contrary, persons move at constant speed, providing higher confidence on motion features. Using the omega-shape (head and shoulder region) as the primary source of key-points, the Kanade-

Table 1. Tracking performance in our surveillance scenario, when using KLT. Performance metrics are Multiple Object Tracking Accuracy (MOTA), Multiple Object Tracking Precision (MOTP), True Positive Rate (TPR), False Positive Rate (FPR) and mismatch (MIS).

Scenario	MOTA	MOTP	TPR	FPR	MIS
S1	0.940	0.600	0.970	0.030	0
S2	0.800	0.590	0.900	0.100	0
S3	0.745	0.336	0.862	0.138	0
S4	0.589	0.288	0.792	0.202	3

(a) Wide-view key-points (b) PTZ key-points (c) Key-point match

Figure 4. Key-point detection and alignment between the two cameras, wide-view (a) and PTZ (b), prior to geometric transform estimation.

Lucas-Tomasi (KLT) algorithm [13] tracks the initial set of features accordingly to motion and appearance constraints. Since some features may be lost during the process, re-initialization of the features is ensured by the detection phase. Figure 3(c) exemplifies the result of KLT tracking in our scenario. The KLT algorithm was preferred since it assumes that a set of discriminant points of the object move with a constant speed and maintain a constant appearance. Based on the set of previous locations provided by the tracking module, a Kalman filter [7] is used to provide a coarse estimation of the future position.

To assess the reliability of the proposed method for tracking, we considered four simple different scenarios with increasing level of difficulties: *S1*- a single person is moving away from the camera, the background subtraction mask is of high quality (i.e. absent of noise), and there are no significant changes in lighting conditions; *S2*- a single person is moving away from the camera at a higher speed, the background subtraction result contains some noise, and there are no significant changes in lighting conditions; *S3*- a single person walks towards the camera, the is significant noise in the background subtraction mask due to wind and significant lighting changes; *S4*- three persons are moving away from the camera, and despite the background subtraction technique outputs little noise, one of the subjects crosses the path of the other two. Results are present in Table 1.

To evaluate tracking performance, the CLEAR MOT metrics [9] were used: MOTA, MOTP, TPR, FPR and MIS. As we can see, the first scenario (*S1*, that represents the best case) produces high levels of accuracy (MOTA) and precision (MOTP). The FPR is so low that is negligible in this first experimental trial (the high quality mask from background subtraction leads to a very precise tracking). Regarding scenarios *S2* and *S3*, all MOTA, FPR and TPR confirm encouraging levels of performance of the tracking algorithm. The significant loss of precision obtained in *S3* comparatively to *S2* is strongly related to the distance that the subject enters the scene. At long working distances, the number of pixels in the scene that represents a person is very small. This condition leads to a failure of the upper body detector, which is the main cause of the loss of precision. In the most challenging scenario (*S4*) the FPR increases, along with some mismatches (mainly related to the path of one person that crosses the path of other two). Nonetheless, we can assert that the tracking method achieves good level of performances.

3.2. Camera Synchronization

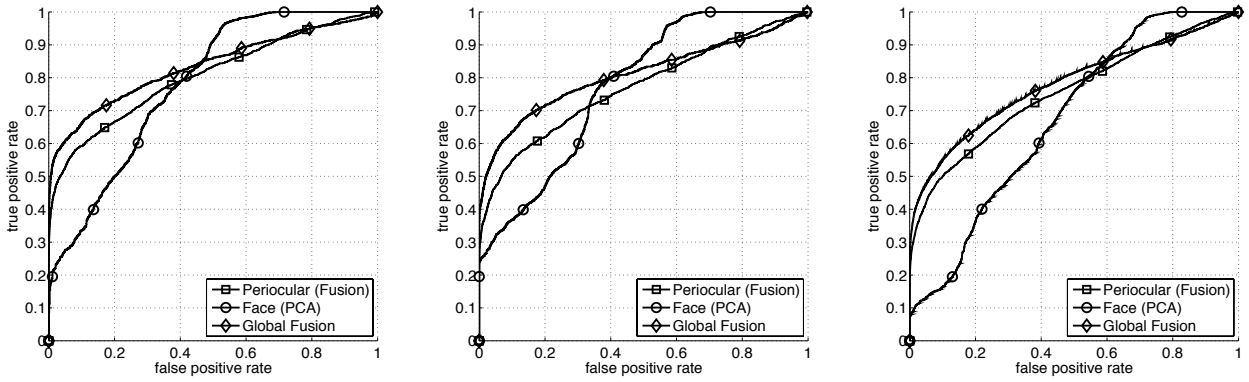
In order to acquire a closer view of the people being tracked, his position needs to be converted from the wide-view camera referential, to the PTZ one. An approximation for this conversion was achieved by mapping key-points between the two views (wide's and PTZ's), and estimating a 2-D geometric transform – Figure 4. For key-point detection, feature encoding and matching, the Speeded-Up Robust Features (SURF) algorithm was used.

3.3. Biometric Recognition

To have a preliminary assessment about the recognition performance of our system, a number of participants were imaged between distances 15 to 35 meters, using the PTZ camera. These working distances ensure regions with widths between 500 px and 200 px for the face, and approximately 220 px to 100 px for the periocular region. Facial region was determined using a cascade object detector based on Viola and Jones algorithm [16], and facial features encoded using the Principal Component Analysis (PCA) approach [15]. Prior to encoding the periocular features, a second Region of Interest (ROI) containing the periocular region was defined using also a Viola and Jones based cascade object detector, trained for the detection of the right eye using Haar features to encode the details [3]. Upon that region, five different descriptors were extracted, based on the works of Park *et al.* [12] and Bharadwaj *et al.* [1]: Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), Scale-Invariant Feature Transform (SIFT), Uniform Local Binary Patterns (ULBP) and GIST. The HOG, LBP and ULBP descriptors deliver a distribution-based analysis, and were computed over 35 non-overlapping patches of the periocular ROI, evenly distributed on a 7×5 grid. Each descriptor was computed sequentially, forming a global 1-D array storing both shape and texture information. Finally, two score-level fusion were also stressed: one combining the scores from the individual periocular recognition meth-

Table 2. Performance for each one of the stressed methods, traits and working distances. Metrics are Decidability (DEC), Area Under Curve (AUC) and Equal Error Rate (EER).

Trait →		Periocular						Face	Global
Method →		LBP	HOG	SIFT	ULBP	GIST	Fusion	PCA	Fusion
15m - 25m	DEC	0.802	0.699	0.404	1.090	0.918	1.162	1.171	1.407
	AUC	0.753	0.703	0.617	0.786	0.772	0.805	0.779	0.835
	EER	0.302	0.358	0.416	0.281	0.304	0.287	0.307	0.246
25m - 35m	DEC	0.677	0.641	0.341	0.972	0.808	1.033	1.173	1.267
	AUC	0.697	0.674	0.598	0.744	0.755	0.771	0.772	0.810
	EER	0.376	0.380	0.431	0.334	0.321	0.303	0.328	0.254
15m - 35m	DEC	0.529	0.520	0.310	0.830	0.747	0.891	0.676	1.025
	AUC	0.663	0.640	0.591	0.710	0.721	0.754	0.674	0.779
	EER	0.396	0.409	0.435	0.360	0.348	0.317	0.395	0.293



(a) 15m - 25m (b) 25m - 35m (c) 15m - 35m

Figure 5. Receiver Operating Characteristic (ROC) curves for the periocular recognition, face recognition and global fusion, at different working distances.

ods; and a second one combining them with the PCA results. Score fusion was achieved training Neural-Network (NN) with two hidden layers using back-propagation. NN based methods are widely applied on classification problems, for their learning abilities and good generalization capabilities. The architecture of the used NN consisted on a first hidden layer with the number of neurons equaling the number of scores to be fused, and a second hidden layer of three neurons. The final (output) layer had one neuron, since we were dealing with a binary classification problem. NNs were trained with a smaller partition of the data, not included on the test phase.

Three metrics were used to assess recognition modules' performance: DEC [4], AUC and EER. The evaluation of the stressed feature encoding techniques for the different working distances and traits is registered at Table 2. For a better interpretation of their performance, the ROC curves are also presented in Figure 5. Results refer to a total of 69960 comparisons, performed on a 1:N fashion.

As we can see from Table 2, top recognition performance was attained at closer working distances (15m to 25m), with an AUC of 0.835. However, widening the working range to the whole driveway (15m to 35m), a considerable good performance is still achieved (AUC = 0.779). We must have in mind that results come from a fully automated system, operating on an adverse surveillance scenario. Furthermore, matches were not performed against a separate dataset of good registration images, but between different PTZ images acquired during system operation.

As for the differences between the different stressed traits, the periocular region seems to be less affected by changes in distance, although further facial recognition techniques should be stressed. Also from the ROC curves at Figure 5, we can see how the PCA applied to the face alone delivers lower true positive rate while introducing higher amounts of false positives, when compared to the fusion of methods operating on the periocular region. Nonetheless, fusing that information with the periocular methods scores produces a considerable improvement on the latter. Thus, if considering deploying a more restrictive system with higher security constraints, the face trait should not be used alone, but can be a powerful ally to further improve its final out-

come.

4. Final Considerations

In this paper we present the concept of a fully automated surveillance and biometric recognition system, able to complement human detection and tracking with biometric recognition over *in the wild* surveillance environments. Although further state-of-the-art techniques can be stressed for each module, we give evidence on the feasibility of such system, providing both tracking performance and biometric recognition results over a real surveillance scenario.

4.1. Further Work

Although a functional system is presented, further work should be considered over three axes: 1) a larger dataset should be acquired, not only with a larger number of subjects going through the scene, but also with the system running over different environments (e.g. indoor lounge); 2) some modules are still to be developed, that would increase the recognition performance even further (e.g. head landmark detector); 3) additional state-of-the-art techniques should be tested for each module, and results cross-validated over the different scenarios. In particular, different face recognition techniques should be stressed, along with ear shape and iris biometrics and gait recognition.

References

[1] S. Bharadwaj, H. Bhatt, M. Vatsa, and R. Singh. Periocular biometrics: When iris recognition fails. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1 –6, September 2010.

[2] W. W. Bledsoe. The model method in facial recognition. Technical Report PRI 15, Panoramic Research, Inc., Palo Alto, California, Palo Alto, California, 1964.

[3] M. Castrillón, O. Dénis, C. Guerra, and M. Hernández. Encara2: Real-time detection of multiple faces at different resolutions in video streams. *Journal of Visual Communication and Image Representation*, 18(2):130–140, 2007.

[4] J. Daugman. High confidence visual recognition of persons by a test of statistical independence. *Pattern Analysis and Machine Intelligence*, 15(11):1148 –1161, November 1993.

[5] I. Haritaoglu, D. Harwood, and L. Davis. W4: Real-time surveillance of people and their activities. *Pattern Analysis and Machine Intelligence*, 22(8):809–830, August 2000.

[6] A. Jain, S. Pankanti, S. Prabhakar, L. Hong, and A. Ross. Biometrics: A grand challenge. In *Pattern Recognition (ICPR), 2004 Proc. of the 17th International Conference on*, volume 2, pages 935–942, 2004.

[7] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Trans. of the ASME âĂŞ Journal of Basic Engineering*, (82 (Series D)):35–45, 1960.

[8] B. Kamgar-Parsi, W. Lawson, and B. Kamgar-Parsi. Toward development of a face recognition system for watchlist surveillance. *Pattern Analysis and Machine Intelligence*, 33(10):1925–1933, October 2011.

[9] B. Keni and S. Rainer. Evaluating multiple object tracking performance: the clear mot metrics. *EURASIP Journal on Image and Video Processing*, 2008.

[10] L. Maddalena and A. Petrosino. A self-organizing approach to background subtraction for visual surveillance applications. *Image Processing, IEEE Trans.*, 17(7):1168–1177, July 2008.

[11] J. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D. LoIacono, S. Mangru, M. Tinker, T. Zappia, and W. Zhao. Iris on the move: Acquisition of images for iris recognition in less constrained environments. In *Proc. of the IEEE*, volume 94, pages 1936–1947, 2006.

[12] U. Park, A. Ross, and A. Jain. Periocular biometrics in the visible spectrum: A feasibility study. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on*, pages 1 – 6, September 2009.

[13] J. Shi and C. Tomasi. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proc. CVPR'94., 1994 IEEE Computer Society Conference on*, pages 593–600. IEEE, 1994.

[14] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*, volume 2, page 252, 1999.

[15] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proc. CVPR '91., IEEE Computer Society Conference on*, pages 586–591, June 1991.

[16] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 511–518, 2001.

[17] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma. Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *Pattern Analysis and Machine Intelligence*, 34(2):372–386, February 2012.

[18] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2000.

Chapter 14

Conclusions and Future Work

14.1 Overview

This chapter presents the conclusions resulting from the research work conducted in the scope of this doctoral thesis. Furthermore, additional topics of research that can be addressed in further studies are also discussed.

14.2 Final Considerations

This thesis purpose is to address the subject of biometric recognition in unconstrained environments, describing the research work developed towards a non-habituated, non-attended and non-standard covert system targeted at public environments.

In order to devise a way to build a biometric system able to work under such challenging conditions, the research work was divided in five steps. 1) We started by studying the iris as biometric trait, in particular its usability for challenging *in the wild* scenarios with low acquisition constraints and over highly degraded data. 2) Next, we conducted a more in-depth analysis on how iris performance is conditioned by visible wavelength light. This analysis was carried on over two axes: the type of illuminant being used in the scene, and the level of luminance. 3) Further to that, emerging biometric traits were also studied, in particular the periocular region. Advances in research show it to be a fit candidate to handle the downsides of non-ideal environments. 4) We then implemented the most relevant periocular methods, testing them over a fixed dataset. A new dataset was also built, with a number of noise factors that replicate those observed in unconstrained scenarios, in a quantified and controlled fashion. 5) At last, studying the possibility of negative recognition usage to better met the requirements of *in the wild* scenarios was also intended. Although this last goal was not fully met, additional lines of work were conducted and, ultimately, a fully automated biometric system was devised able to work over video-surveillance scenarios delivering biometric recognition.

Alongside with these objectives a set of scientific contributions were achieved and published in international journals and indexed conferences.

14.3 Contributions and Achievements

Whilst studying the iris usability by assessing its behavior under visible wavelength light, we observed how the appearance of its patterns is poorly conditioned by the type of illuminant being used during the acquisition process, even though the luminance level was of the most importance [4]. That evidence should support the build of a non-standard biometric system

based on the iris, able to work under different *types of light*, both natural or artificial, as long as a fit luminance level is assured or additional measures taken to circumvent this limitation. A new recognition scheme for VW data was also proposed, based on the extraction of MPEG-7 color and shape descriptors from autonomously defined iris sub-regions. This novel technique shows not only performance levels close to those of the best-known techniques, but also low levels of linear correlation in relation to them, suggesting that they can be fused for further performance improvements [5].

On the scope of periocular biometrics, our assessment about the literature showed how recent research focus mainly on texture analysis and key-point extraction, with even simple algorithms (e.g., LBP based techniques) leading to fair performance levels. Periocular is regarded as particularly suited for unconstrained and uncooperative scenarios, either by being used alone or combined with the iris, even though the later can not be properly imaged under such constraints. Testing the state-of-the-art methods against a common dataset allowed us to comparatively access their performance and identify the relevant patterns that systematically condition their capacities [6, 44].

Having identified the detection of eye-corners in facial images of great importance in biometric applications (e.g., periocular ROI definition), state-of-the-art methods for eye-corner detection were empirically evaluated against low-quality data. We found that their performance is significantly diminished in such conditions, leading to the proposition of a novel eye-corner detection method able to deal with degraded *in the wild* images. Comparing our method results against state-of-the-art techniques we could see how it outperformed both on noise-free and degraded data (blurred, rotated, with differences in scale, etc.). These improvements were achieved without loss of computational efficiency, which is essential when aiming at deploying a real-time biometric system [7]. Also, on the scope of better defining the periocular ROI with better pose / gaze estimates, a component labeling method was developed with the capacity to discriminate between seven key-elements using a two-step approach: at the first stage, a group of local classifiers use texture descriptors to compute the pixel-level probability of each class; on a second stage this information is fused with geometrical constraints and shape information to feed a two-layered MRF [8].

Aiming at increase the reliability of non-cooperative iris recognition over the degraded data we encounter *in the wild*, we proposed a novel fusion of different recognition approaches in such a way that the pitfalls of less constrained acquisition setups under visible wavelength lighting could be minimized. The proposed fusion of iris and periocular features was shown to improve the overall performance on both identification and verification modes, conclusion that was further corroborated by a third place on the Noisy Iris Challenge Evaluation - Part II [9].

A similar attempt was conducted at the Mobile Iris CHallenge Evaluation - Part I, aimed this time at mobile setups. We have built and made available for the scientific community a novel iris and periocular database, the CSIP, containing images acquired at ten different mobile setups and eight visible noise factors, along with ground truth data for the iris segmentation. Such dataset allows evaluating both iris segmentation and recognition techniques, and periocular recognition methods. We also identified the chromatic disparity introduced by some devices, and proposed the usage of a color correction technique to compensate for the color distortions inherent to each one of the setups. Results showed this approach leading to top results, spe-

cially on cross-sensor comparisons. At the recognition stage, we showed how the score-level fusion of well-known iris and periocular recognition strategies can be used to overcome the issues associated with mobile setups. Furthermore, we showed how even simple feature encoding techniques, with low computational cost, deliver considerably good performance, which is particularly interesting if the application is to be deployed on mobile devices with higher computational constraints [3].

Setting our efforts into deploying a fully functional system able to deal with extremely harsh data, we identified the need for a dataset that gathered multiple noise factors on a precisely quantified manner. As so, we built the BioHDD dataset, containing multi-session information from 101 subjects: high quality registration mugshots; large sets of extremely degraded probes with ten different noise factors; and video sequences with the subjects walking through a non-standard hallway. Conducting an online survey, mimicking a watchlist identification scenario where participants were asked to perform both positive and negative recognition tasks, allowed us to perceive which features humans most frequently associate with successful and failed identifications. We observed that, for the humans, to deal with inadequate illumination intensity and moderate levels of occlusion is no issue. The good performance observed over low-resolution and compressed images also suggests a significant usage of global features. As main issues we point out subjects looking up or down from the camera level, and high-levels of occlusion, which can be a significant limitation to the system as subjects trying to avoid detection are most likely to be facing away from visible cameras or carry headgear. Shape information and holistic cues were both the most accounted for and the most reliable ones. This research paper allowed us to further support the usage of the periocular region, as features located on the mid-face were found to be the more reliable ones [2].

Following these research findings and aiming at establishing a way to better understand the contents of a head photo acquired by the recognition system *in the wild*, we proposed a novel algorithm for head landmark detection [45] able to identify and pinpoint the location of six key-elements of the human head, among witch is the periocular region. This algorithm was tested against a subset of the BioHDD database and was found to work with relatively good performance, being invariant to the image acquisition angle on a full 360° view of the subject and to changes in users' head pitch.

At last, the research work of this doctoral program culminated with the concept of a fully automated surveillance and biometric recognition system. Based on of a video surveillance system able to detect and track human subjects *in the wild*, its scene understanding layer was enhanced to control a PTZ camera allowing the acquisition of head-shots of the tracked individuals. Using the previously mentioned head landmarking technique, the multi-modal biometric part of the system is able to identify the visible traits and chose which recognition module to use. Ultimately, evidence is given on the feasibility of such system, with a proof-of-concept being tested on a real surveillance scenario [45].

14.4 Further Work

The deployment of a fully functional biometric system *in the wild* able to work covertly in harsh conditions is still a work in progress. Although the system has already been fully devised, some

modules are still under production. We intent to completely validate and debug it, carrying a full evaluation of its performance under different environments.

Even though two of the proposed recognition methods have been independently evaluated on international contests (NICE.II and MICHE I) and all contributions were peer-evaluated, we still think that it would be interesting to perform their evaluation against different and larger datasets, so that the statistical relevance of the results can be even higher.

As stated in the introductory section, one of the objectives of this doctoral research could have started a more in-depth research, namely the negative recognition approach as a fit alternative for *real-world* scenarios. Being able to guarantee with enough confidence that an unknown subject does not belong to a given watchlist of "persons-of-interest" is a most tempting goal for nowadays security demands. Understanding its practical advantages, most of which being privacy related, we intend to complement the final prototype with this mode of operation.

Bibliography

[1] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. Californa, USA: Academic Press, 2009. xv, xlvii, 1, 2

[2] G. Santos, P. Fiadeiro, and H. Proença, "Biohdd: a dataset for studying biometric identification on heavily degraded data," *IET Biometrics*, 2015. xvii, xxxvii, 4, 5, 153

[3] G. Santos, E. Grancho, M. Bernardo, and P. Fiadeiro, "Fusing iris and periocular information for cross-sensor recognition," *Pattern Recognition Letters*, 2014. xviii, xxxvi, 4, 5, 153

[4] G. Santos, M. Bernardo, P. Fiadeiro, and H. Proença, "Iris recognition: Preliminary assessment about the discriminating capacity of visible wavelength data," in *Proceedings of the 6th IEEE Workshop on Multimedia Information Processing and Retrieval - MIPR '10*, Taiwan, 2010, pp. 324–329. xviii, xxix, xxxv, 4, 5, 19, 151

[5] H. Proença and G. Santos, "Fusing color and shape descriptors in the recognition of degraded iris images at visible wavelengths," *Computer Vision and Image Understanding*, vol. 116, pp. 168–178, 2012. xviii, xxxv, 4, 5, 152

[6] G. Santos and H. Proença, "Periocular biometrics: An emerging technology for unconstrained scenarios," in *Proceedings of the IEEE Symposium on Computational Intelligence in Biometrics and Identity Management - CIBIM 2013*, April 2013, pp. 14–21. xviii, xxxv, 4, 5, 152

[7] G. Santos and H. Proença, "A robust eye-corner detection method for real-world data," in *Biometrics (IJCB), 2011 International Joint Conference on*, 2011, pp. 1–7. xviii, xxxv, 5, 6, 152

[8] H. Proença, J. Neves, and G. Santos, "Segmenting the periocular region using a hierarchical graphical model fed by texture / shape information and geometrical constraints," September 2014. xix, xxxvi, 5, 6, 152

[9] G. Santos and E. Hoyle, "A fusion approach to unconstrained iris recognition," *Pattern Recognition Letters*, vol. 33, no. 8, pp. 984 – 990, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865511002686 xix, xxxi, xxxvi, 5, 6, 21, 152

[10] S. Garfinkel, *Database Nation: The Death of Privacy in the 21st Century*. O'Reilly Media, 2000. xx, 9

[11] P. Komarinsky, *Automated Fingerprint Identification Systems (AFIS)*, 1st ed. Academic Press, 2005. xxi, 10

[12] F. Galton, *Finger Prints*. London: Macmillan and Co., 1892. xxi, 10

[13] A. Jain, P. Flynn, and A. Ross, Eds., *Handbook of biometrics*. Springer, 2008. xxi, 10

[14] J. Undar, W. Seng, and A. Abbasi, "A review of biometric technology along with trends and prospects," *Pattern Recognition*, no. 47, pp. 2673–2688, 2014. xxii, xxiv, 11, 12, 14

[15] D. Dessimoz and J. Richiardi, "MBioID multimodal biometrics for identity documents," Université de Lausanne, Lausanne, Research Report PFS 341-08.05, June 2006. xxii, xxiii, 12, 13

[16] K. Bowyer, K. Hollingsworth, and P. Flynn, "Image understanding for iris biometrics: A survey," *Comput. Vis. Image Underst.*, vol. 110, no. 2, pp. 281–307, May 2008. [Online]. Available: http://dx.doi.org/10.1016/j.cviu.2007.08.005 xxiv, 14

[17] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2000. xxiv, 14

[18] A. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, no. 14, pp. 4–20, 2004. xxiv, 14

[19] J. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 15, no. 11, pp. 1148 –1161, November 1993. xxviii, xxxii, 18, 23

[20] B. Matthews, "Comparison of the predicted and the observed secondary structure of t4 phage lysozyme," *Biochim. Biophys. Acta*, vol. 405, pp. 442–451, 1975. xxviii, 19

[21] C. Fancourt, L. Bogoni, K. Hanna, Y. Guo, R. Wildes, N. Takahashi, and U. Jain, "Iris recognition at a distance," July 2005, pp. 1–13. xxviii, 19

[22] K. Smith, V. Pauca, A. Ross, T. Torgersen, and M. King, "Extended evaluation of simulated wavefront coding technology in iris recognition," in *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, September 2007, pp. 1–7. xxviii, 19

[23] A. Ross, R. Jillela, J. Smereka, V. Boddeti, B. Kumar, R. Barnard, H. Xiaofei, P. Pauca, and R. Plemmons, "Matching highly non-ideal ocular images: An information fusion approach," in *Proceedings of the 2013 5th IAPR International Conference on Biometrics (ICB)*, April 2012, pp. 446–453. xxix, 19

[24] H. Proença, G. Santos, and J. Neves, "Using ocular data for unconstrained biometric recognition," in *Face Recognition in Adverse Conditions*, M. Marsico, M. Nappi, and M. Tistarelli, Eds. IGI Global, 2013. xxix, xxxii, xxxiii, 19, 22, 23

[25] K. Y. Shin, G. P. Nam, D. S. Jeong, D. H. Cho, B. J. Kang, K. R. Park, and J. Kim, "New iris recognition method for noisy iris images," *Pattern Recognition Letters*, vol. 33, no. 8, pp. 991 – 999, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865511002674 xxxi, 21

[26] T. Tan, X. Zhang, Z. Sun, and H. Zhang, "Noisy iris image matching by using multiple cues," *Pattern Recognition Letters*, vol. 33, no. 8, pp. 970 – 977, 2012, <ce:title>Noisy Iris Challenge Evaluation II - Recognition of Visible Wavelength Iris Images Captured At-a-distance and On-the-move</ce:title>. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865511002601 xxxi, 21

[27] L. Junli, Z. Miaohua, L. Ding, Z. Xianju, O. Ojowu, Z. Kexin, L. Zhan, and L. Han, "Robust ellipse fitting based on sparse combination of data points," *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2207–2218, 2013. xxxi, 22

[28] C. Tan and A. Kumar, “Efficient iris segmentation using grow-cut algorithm for remotely acquired iris images,” in *Proceedings of the 2013 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems*, September 2012, pp. 99–104. xxxi, 22

[29] F. Alonso-Fernandez and J. Bigun, “Iris boundaries segmentation using generalized structure tensor: A study on the effects of image degradation,” pp. 426–431, September 2012. xxxi, 22

[30] H. Xinyu, F. Bo, T. Changpeng, A. Tokuta, and Y. Ruigang, “Robust varying-resolution iris recognition,” in *Proceedings of the 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems*, September 2012, pp. 47–54. xxxi, 22

[31] Y.-H. Li and M. Savvides, “An automatic iris occlusion estimation method based on high dimensional density estimation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, pp. 784–796, 2012. xxxii, 22

[32] J. Zuo and N. Schmid, “Adaptive quality-based performance prediction and boosting for iris authentication: Methodology and illustration,” *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 6, pp. 1051–1060, 2013. xxxii, 22

[33] J. Daugman, “New methods in iris recognition,” *IEEE Trans. Systems, Man, Cybernetics B*, vol. 37, no. 5, pp. 1167–1175, 2007. xxxii, 23

[34] W. Boles and B. Boashash, “A human identification technique using images of the iris and wavelet transform,” *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 1185–1188, April 1998. xxxii, 23

[35] R. Wildes, “Iris recognition: an emerging biometric technology,” vol. 85, no. 9, September 1997, pp. 1348–1363. xxxii, 23

[36] J. Matey, D. Ackerman, J. Bergen, and M. Tinker, “Iris recognition in less constrained environments,” in *Advances in Biometrics*, N. Ratha and V. Govindaraju, Eds. Springer London, 2008, pp. 107–131. xxxii, 23

[37] Honeywell International Inc,, “A distance iris recognition,” 2007, united States Patent 20070036397. xxxii, 23

[38] R. Jillela and A. Ross, “Mitigating effects of plastic surgery: Fusing face and ocular biometrics,” in *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, 2012. xxxiii, 23

[39] H. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, “Recognizing surgically altered face images using multiobjective evolutionary algorithm,” *Information Forensics and Security, IEEE Transactions on*, vol. 8, no. 1, pp. 89–100, 2013. xxxiii, 23

[40] BWG, “Biometric security concerns,” UK Government Biometrics Working Group, United Kingdom, Tech. Rep., September 2003. xxxiii, 24

[41] CESG, “Ms06 privacy issues.” [Online]. Available: https://www.cesg.gov.uk/policyguidance/biometrics/Pages/MS06-Privacy-Issues.aspx xxxiii, 24

[42] CEN, “Initiative on privacy standardization in europe,” European Comitte for Standardization, Brussels, Final Report, February 2002. xxxiv, 24

[43] K. Bowyer, “Face recognition technology and the security versus privacy tradeoff,” *IEEE Technology and Society*, pp. 9–20, Spring 2004. xxxiv, 25

[44] G. Santos and H. Proença, “On periocular biometrics, a comprehensive outine,” submitted for publication at the Artificial Intelligence Review journal. xxxv, 5, 152

[45] G. Santos, J. Neves, S. Filipe, E. Grancho, S. Barra, and F. Narducci, “Quis-campi: Extending in the wild biometric recognition to surveillance environments,” submitted to the 8th IAPR International Conference on Biometrics – ICB-2015. xxxvii, xlvii, 5, 153

[46] G. Santos and H. Proença, “A dual-step approach to head landmark detection in the wild,” submitted to the 8th IAPR International Conference on Biometrics – ICB-2015. 5

Appendix A

Robust Periocular Recognition by Fusing Sparse Representation of Color and Geometry Information

A.1 Overview

This appendix consists of the following article:

Robust Periocular Recognition by Fusing Sparse Representation of Color and Geometry Information
Juan C. Moreno, V. B. Surya Prasath, Gil Santos and Hugo Proença
Journal of Signal Processing Systems, submitted for consideration, 2014.

Robust periocular recognition by fusing sparse representations of color and geometry information

Juan C. Moreno · V. B. Surya Prasath · Gil Santos · Hugo Proença ·

Abstract In this paper, we propose a re-weighted elastic net (REN) model for biometric recognition. The new model is applied to data separated into geometric and color spatial components. The geometric information is extracted using a fast cartoon - texture decomposition model based on a dual formulation of the total variation norm allowing us to carry information about the overall geometry of images. Color components are defined using linear and nonlinear color spaces, namely the red-green-blue (RGB), chromaticity-brightness (CB) and hue-saturation-value (HSV). Next, according to a Bayesian fusion-scheme, sparse representations for classification purposes are obtained. The scheme is numerically solved using a gradient projection (GP) algorithm. In the empirical validation of the proposed model, we have chosen the *periocular region*, which is an emerging trait known for its robustness against low quality data. Our results were obtained in the publicly available UBIRIS.v2 data set and show consistent improvements in recognition effectiveness when compared to related state-of-the-art techniques.

Keywords Sparse Representation · Periocular Recognition · Total Variation · Elastic Net Regularization · Color · Texture Decomposition.

J. C. Moreno · Gil Santos · H. Proença
IT-Instituto de Telecomunicações
Department of Computer Science
University of Beira Interior, Portugal
E-mail: {jcmb,gmelfe,hugomcp}@ubi.pt

V. B. S. Prasath
Department of Computer Science
University of Missouri-Columbia, MO 65211 USA
E-mail: prasaths@missouri.edu

1 INTRODUCTION

Biometrics attempts to recognize human beings according to their physical or behavioral features [17]. In the past, various traits were used for biometric recognition, out of which *iris* and *face* are the most popular [35,41,19,29]. The use of the periocular region is found to be useful on unconstrained scenarios [38]. The exploration of the periocular region as a biometric trait started with Park *et al.*'s pioneering approach [33], who performed local and global feature extraction. Images were aligned to take advantage of iris location, in order to define a 7×5 region of interest (ROI) grid. Patches were encoded by applying two well known distribution-based descriptors, local binary patters (LBP) [26] and histogram of oriented gradients (HOG) [10], quantized into 8-bin histograms. Finally, they merged all histograms into a single-dimension array containing both texture and shape information, and matching was carried out based on the Euclidean distance. For the local analysis, authors employed Scale-Invariant Feature Transform (SIFT) [24]. The reported performance was fairly good, showing periocular fitness for recognition purposes, and further analysis was held on noise factors impact on performance [31].

Recently, various extensions and improvements based on Park et al work [33] has been carried out. Miller *et al.* [25] presented an analysis which focused on periocular skin texture, taking advantage of uniform local binary patters (ULBP) [27] to achieve improved rotation invariance with uniform patterns and finer quantization of the angular space. Their work was extended by Adams *et al.* [1], who proposed using Genetic & Evolutionary Computing (GEC) to optimize feature set. Juefei-Xu *et al.* [20] used multiple local and global feature extraction techniques such as Walsh transforms and Laws' masks, discrete cosine transform (DCT), discrete wavelet transform (DWT), Force Fields, Speed Up Robust Transform (SURF), Gabor filters and Laplacian of Gaussian (LoG). In their later work [21] efforts were made to compensate aging degradation effects on periocular performance. The possibility of score level fusion with other biometric traits was also addressed, for example in iris recognition [47]. Bharadwaj *et al.* [2] proposed the fusion of ULBP with five perceptual dimensions, usually applied as scene descriptors: naturalness, openness, roughness, expansion and ruggedness – GIST [28]. In their approach the images were pre-processed with with Fourier transform for local contrast normalization, and then a spacial envelope computed with a set of Gabor filters (4 scales $\times$ 8 orientations). On the final stage, χ^2 distance was used to match the feature arrays, and results fused with a weighted sum.

Based on the pioneering work of Wright *et al.* [48], the sparse representation theory is emerging as a popular method in the biometrics fields and is considered specially suitable to handle degraded data acquired under uncontrolled acquisition protocols [34,40]. A query image is first sparsely coded over the template images, and then the classification is performed. Sparse Representation based Classification (SRC) is robust to occlusion, illumination and noise, and achieves excellent performance.

1.1 Sparse Representation

Model selection in high-dimensional problems has been gaining interest in the statistical signal processing community [11,4]. Using convex optimization models, the main problem is recovering a sparse solution $\hat{\mathbf{x}} \in \mathbb{R}^n$ of an underdetermined system of the form $\mathbf{y} = A\mathbf{x}^*$, given a vector $\mathbf{y} \in \mathbb{R}^m$ and a matrix $A \in \mathbb{R}^{m \times n}$. There is a special interest in signal recovery when the number of predictors are much larger than the number of observations (n$\gg$ m). A direct solution to the problem is to select a signal whose measurements are equal to those of $\mathbf{x}^*$, with smaller sparsity by solving a minimization problem based on the ℓ^0-norm:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subj. to } A\mathbf{x} = \mathbf{y}, \tag{1}$$

($\|\mathbf{x}\|_0 = \#\{i : \ x_i \neq 0\}$), being a direct approach to seek the sparsest solution. Problem (1) is proved to be NP-hard and difficult to approximate since it involves non-convex minimization [5]. An alternative method is to relax the problem (1) by means of the ℓ^1-norm ($\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$). Hence problem (1) can be replaced by the following ℓ^1-minimization problem:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subj. to } A\mathbf{x} = \mathbf{y},$$

which can be solved by standard linear programming methods [9]. In practice, signals are rarely exactly sparse, and may often be corrupted by noise. Under noise, the new problem is to reconstruct a sparse signal $\mathbf{y} = A\mathbf{x}^* + \boldsymbol{\kappa}$, where $\boldsymbol{\kappa} \in \mathbb{R}^m$ is white Gaussian noise with zero mean and variance σ^2. In this case the associated ℓ^1-minimization problem adopts the form:

$$\min_{\mathbf{x}} \left\{ \tau \|\mathbf{x}\|_1 + \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \right\}, \tag{2}$$

where τ is a nonnegative parameter and $\|\cdot\|_2$ denotes the ℓ^2-norm ($\|\mathbf{x}\|_2 = \left(\sum_{i=1}^n x_i^2\right)^{\frac{1}{2}}$). The convex minimization problem (2) is known as the least absolute value shrinkage and selection operator (LASSO) [43].

Although sparsity of representation seems to be well established by means of the LASSO approach, some limitations were remarked by Hastie *et al.* [51]. LASSO model tends to select at most m variables before it saturates and in case predictors are highly correlated, LASSO usually selects one variable from a group, ignoring others. In order to overcome these difficulties, Hastie *et al.* [51] proposed the elastic net (EN) model as a new regulation technique for outperforming LASSO in terms of prediction accuracy. The elastic net is characterized by the presence of ridge regression term (ℓ^2-norm) and it is defined by the following convex minimization problem:

$$\min_{\mathbf{x}} \left\{ \tau_1 \|\mathbf{x}\|_1 + \tau_2 \|\mathbf{x}\|_2^2 + \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \right\}, \tag{3}$$

where τ_1 and τ_2 are non-negative parameters. An improvement for the EN model was proposed in [52] where a combination of the ℓ^2-penalty and an

adaptive version of the ℓ^1-norm have been implemented by considering the minimization problem

$$\min_{\mathbf{x}} \left\{ \tau_1 \sum_{i=1}^{n} \omega_i |x_i| + \tau_2 \|\mathbf{x}\|_2^2 + \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \right\}, \tag{4}$$

where the adaptive weights are computed using a solution given by the EN minimization problem (3). If we let the solution of EN to be $\hat{\mathbf{x}}(EN)$, then the weights are given by the equation $\omega_i = 1/(|\hat{x}_i(EN)| + (1/m))^{\vartheta}$ where ϑ is a positive constant. A variant of the above model was proposed in [16] by incorporating the adaptive weight matrix W in the ℓ^2-penalty term:

$$\min_{\mathbf{x}} \left\{ \tau_1 \sum_{i=1}^{n} \omega_i |x_i| + \tau_2 \sum_{i=1}^{n} \omega_i^2 x_i^2 + \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \right\}. \tag{5}$$

In this paper we use a re-weighted elastic net regularization model for periocular recognition application.

1.2 Summary of Contributions

The main contribution of this paper is to propose a re-weighted elastic net (REN) regularization model, that enhances the sparsity of the solutions found. The proposed REN model is a regularization and variable selection method that enjoys sparsity of representation, particularly when the number of predictors are much larger than the number of observations. The weights are computed such that larger weights will encourage small coordinates by means of the ℓ^1-norm, and smaller weights will encourage large coordinates due to the ℓ^2-norm. Our model differs from the schemes in [52] and [16] (see equations (4) and (5) above), since the ℓ^1 and ℓ^2 terms are automatically balanced by weights which are continuously updated using $\omega_i = 1/(|\hat{x}_i| + \epsilon)$ with ϵ a positive parameter [7]. We also provide a concise proof of the existence of a solution for the proposed model as well as its accuracy property.

A complete presentation of the numerical implementation of the REN model using a gradient projection (GP) method [14], seeking sparse representations along certain gradient directions is described in this paper. We use a reformulation of the REN model as a quadratic programming (QP) problem. As a main application of our model, we consider the periocular recognition problem. The periocular region has been regarded as a trade-off between using the entire face or only the iris in biometrics. Periocular region is particularly suitable for recognition under visible wavelength light and uncontrolled acquisition conditions [32,46,30].

We enhance periocular recognition through the sparsity-seeking property of our REN model over different periocular sectors, which are then fused according to a Bayesian decision based scheme. The main idea is to benefit from the information from each sector, which should contribute in overall recognition robustness. Two different domains are considered for this purpose: (1)

Fig. 1 Examples of periocular images of different subjects and varying gazes, containing the *corneal*, *eyebrows* and *skin* regions.

geometry and (2) *color*. Full geometry information is accessed by decomposing a given image into their cartoon - texture components by means of a dual formulation of the weighted total variation (TV) scheme [37]. For color, a key contribution is the use of nonlinear features such as chromaticity and hue components, which are thought to improve image geometry information according to human perception [22]. Our methodology is inspired by two related works: 1) Wright *et al.* [48], which introduced the concept of *sparse representation* for *classification* (SRC) purposes; and 2) Pillai *et al.* [34], that used a SRC model for disjoint sectors of the iris and fused results at the score level, according to a confidence score estimated from each sector.

Our experiments are carried out in periocular images of the UBIRIS.v2 data set [36]: images were acquired at visible wavelengths, from 4 to 8 meters away from the subjects and uncontrolled acquisition conditions. Varying gazes, poses and amounts of occlusions (due to glasses and reflections) are evident in this data set and makes the recognition task harder, see Figure 1. The results obtained using our model allowed us to conclude about consistent increases in performance when compared to the classical SRC model and other important approaches (e.g., Wright *et al.* [48] and Pillai *et al.* [34]). Also, it should be stressed that such increase in performance were obtained without a significant overload in the computational burden of the recognition process.

The rest of the paper is organized as follows. Section 2 summarizes the most relevant in the scope of this work concerning penalized feature selection for sparse representation. The re-weighted elastic net (REN) model is introduced together with statistical motivation ensuring high prediction rates. An algorithm based on gradient projection (GP) for the REN model is also introduced. Section 3 describes the different geometrical information extracted from periocular images for performing recognition based on cartoon - texture and chromaticity features in a total variation framework. Section 4 describes the experimental validation procedure carried out together with remarkable comparisons. Finally, Section 5 concludes the paper.

2 The Reweighted Elastic Net model for Classification Model

2.1 The LASSO Model for Recognition

We first briefly describe the sparse representation based classification framework which is a precursor to our REN based approach. Having a set of labeled training samples (n_i samples from the ith subject), they are arranged as columns of a matrix $A(i) = [\mathbf{v}_{i,1}, \ldots, \mathbf{v}_{i,n_i}] \in \mathbb{R}^{m \times n_i}$. A dictionary results from the concatenation of all samples of all classes:

$$A = [A(1), \ldots, A(k)] = [\mathbf{v}_{1,1}, \cdots, \mathbf{v}_{1,n_1} | \ldots | \mathbf{v}_{k,1}, \cdots, \mathbf{v}_{k,n_k}].$$

The key insight is that any probe $\mathbf{y}$ can be expressed as a linear combination of elements of A. As the data acquisition process often induces noisy samples, it turns out to be practical to make use of the LASSO model. In this case it is assumed that the observation model has the form $\mathbf{y} = A\mathbf{x}^* + \boldsymbol{\kappa}$.

Classification is based on the observation that high values of the coefficients in the solution $\hat{\mathbf{x}}$ are associated with the columns of A of a single class, corresponding to the identity of the probe. A residual score per class $\mathbb{1}_i : \mathbb{R}^n \to \mathbb{R}^n$ is defined: $\hat{\mathbf{x}} \to \mathbb{1}_i(\hat{\mathbf{x}})$, where $\mathbb{1}_i$ is a indicator function that set the values of all coefficients to 0, except those associated to the ith class. Over this setting, the probe $\mathbf{y}$ is then reconstructed by $\hat{\mathbf{y}}_i = A\mathbb{1}_i(\hat{\mathbf{x}})$, and the minimal reconstruction error deemed to correspond to the identity of the probe, between $\mathbf{y}$ and $\hat{\mathbf{y}}_i$:

$$\mathrm{id}(\mathbf{y}) = \arg\min_i r_i(\mathbf{y}),$$

with $r_i(\mathbf{y}) = \|\mathbf{y} - \hat{\mathbf{y}}_i\|_2$.

In [48] a sparsity concentration index (SCI) is used to accept/reject the response given by the LASSO model. The SCI of a coefficient vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ corresponds to:

$$SCI(\hat{\mathbf{x}}) = \frac{\frac{k \max_i \|\mathbb{1}_i(\hat{\mathbf{x}})\|_1}{\|\hat{\mathbf{x}}\|_1} - 1}{k - 1} \in [0, 1].$$

If SCI($\hat{\mathbf{x}}$) ≈ 1, the computed signal $\hat{\mathbf{x}}$ is considered to be acceptably represented by samples from a single class. Otherwise, if SCI($\hat{\mathbf{x}}$) ≈ 0 the sparse coefficients spread evenly across all classes and a reliable identity for that probe cannot be given.

The recognition model proposed by Pillai *et al.* [34] obtains separate sparse representations from disjoint regions of an image and fusing them by considering a quality index from each region. Let L be the number of classes with labels $\{c_i\}_{i=1}^{L}$. A probe $\mathbf{y}$ is divided into sectors, each one described by the SRC algorithm. SCI values are obtained over each sector, allowing to reject those with quality bellow a threshold. Let $\{d\}_i$ represent the class labels of the

retained sectors, and $\mathbb{P}(d_i|c)$ be the probability that the i-th sector returns a label d_i, when the true class is c:

$$\mathbb{P}(d_i|c) = \begin{cases} \dfrac{t_1^{SCI(d_i)}}{t_1^{SCI(d_i)} + (L-1)t_2^{SCI(d_i)}} & \text{if } d_i = c, \\ \dfrac{t_2^{SCI(d_i)}}{t_1^{SCI(d_i)} + (L-1)t_2^{SCI(d_i)}} & \text{if } d_i \neq c, \end{cases}$$

being t_1 and t_2 constants such that $0 > t_1 > t_2 > 1$. According to a maximum a posteriori (MAP) estimate of the class label, the response corresponds to the class having the highest accumulated SCI:

$$\tilde{c} = arg \max_{c \in \mathbf{C}} \frac{\sum_{j=1}^{L} SCI(d_j)\delta(d_j = c)}{\sum_{j=1}^{L} SCI(d_j)}.$$

2.2 The Re-weighted Elastic Net (REN) Method

The proposed REN model is a sparsity of representation approach balances the LASSO shrinkage term (ℓ^1-norm) and the strengths of the quadratic regularization (ℓ^2-norm) coefficients by the following minimization problem:

$$\min_{\mathbf{x}} \left\{ \sum_{i=1}^{n} \omega_i |x_i| + \sum_{i=1}^{n} (1-\omega_i)^2 x_i^2 + \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \right\}, \tag{6}$$

where $\omega_1, \ldots, \omega_n$ are positive weights taking values in $(0, 1)$. The REN-penalty $\sum_{i=1}^{n} \omega_i |x_i| + \sum_{i=1}^{n} (1-\omega_i)^2 x_i^2$ is strictly convex and it is a compromise between the ridge regression penalty and the LASSO. The convex combination in the REN-penalty term is natural in the sense that both the ℓ^1 and ℓ^2 norms are balanced by weights controlling the amount of sparsity versus smoothness expected from the minimization scheme. As in [7], the weights are chosen such that they are inversely related to the computed signal according to the equation $\omega_i = 1/(|\hat{x}_i| + \epsilon)$ with ϵ a positive parameter. Under this setting, large weights w_i will encourage small coordinates with respect to the REN-penalty term, whereas small weights imply big coordinates with respect to the REN-penalty term, respectively. Then, it is seen that the new model combines simultaneously a continuous shrinkage and an automatic variable selection approach. We next consider the existence of solution and the sign recovery property of the REN model.

Next we describe an algorithm for the REN model allowing us to directly deal with the case $n \gg m$. It turns out that our REN model can be expressed as a quadratic program (QP), thus allowing us to apply a gradient projection approach to perform the sparse reconstruction.

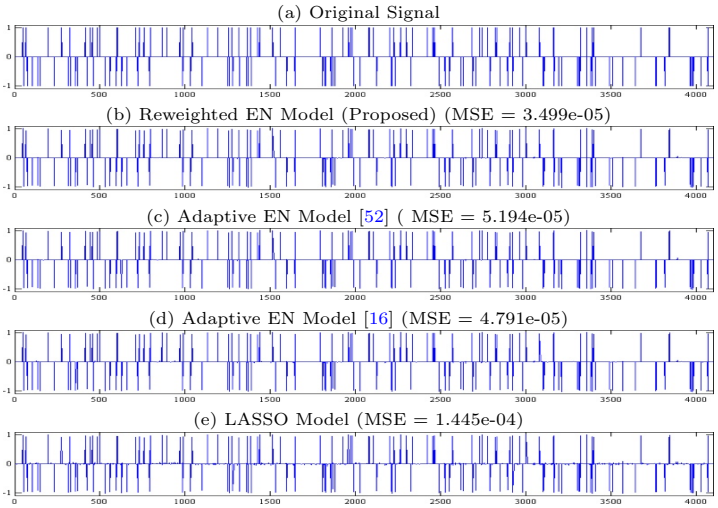


Fig. 2 Sparse signal reconstruction with EN and LASSO models. (a) Sparse signal of Length $n = 4096$ with $k = 1024$ observations. (b)-(e) Response signals computed with the proposed reweighted elastic net, [52], [16] and LASSO, respectively.

2.3 Numerical Implementation

The algorithm that alternates between the computed signal and redefining the weights is as follows:

1. Choose initial weights $w_i = 1/2$, $i = 1, \ldots, n$.
2. Find the solution $\hat{\mathbf{x}}$ of the problem

$$\min_{\mathbf{x}} \|W\mathbf{x}\|_1 + \|(1-W)\mathbf{x}\|_2^2 + \frac{1}{2}\|\mathbf{y} - A\mathbf{x}\|_2^2, \tag{7}$$

3. Update the weights: for each $i = 1, \cdots, n$,

$$w_i = \frac{1}{|\hat{x}_i| + \epsilon},$$

   where ϵ is a positive stability parameter.
4. Terminate on convergence or when a specific number of iterations is reached. Otherwise, go to step 2.

Note that our REN problem in (7) can also be expressed as a quadratic program [15], by splitting the variable $\mathbf{x}$ into its positive and negative parts. That is, $\mathbf{x} = \mathbf{x}_+ - \mathbf{x}_-$, where $\mathbf{x}_+$ and $\mathbf{x}_-$ are the vectors that collect the positive and negative coefficients of $\mathbf{x}$, respectively. Then, we handle the minimization problem,

$$\min_{\mathbf{z}} \left\{ Q(\mathbf{z}) = \mathbf{c}^T\mathbf{z} + \mathbf{z}^T B\mathbf{z} \right\}, \tag{8}$$

where $\mathbf{z} = [\mathbf{x}_+, \mathbf{x}_-]^T$, $\mathbf{w}_n = [\omega_1, \ldots, \omega_n]^T$, $\mathbf{c} = \mathbf{w}_{2n} + [-A^T\mathbf{y}; A^T\mathbf{y}]^T$ and $B = \frac{1}{2}B_1 + B_2$ with

$$B_1 = \begin{pmatrix} A^TA & -A^TA \\ -A^TA & A^TA \end{pmatrix}, \qquad B_2 = \begin{pmatrix} (1-W)^2 & -(1-W)^2 \\ -(1-W)^2 & (1-W)^2 \end{pmatrix}.$$

The minimization problem (8) can then be solve using the Barzilai-Borwein Gradient Projection Algorithm [39]. Under this approach the iterative equation is given by,

$$\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} - \zeta^{(k)}\nu^{(k)},$$

where $\zeta^{(k)}$ is the step size computed as

$$\zeta^{(k)} = \left(\mathbf{z}^{(k)} - \alpha^{(k)}\nabla Q\left(\mathbf{z}^{(k)}\right)\right)_+ - \mathbf{z}^{(k)},$$

with

$$\alpha^{(k+1)} = \begin{cases} mid\left\{\alpha_{min}, \dfrac{\left\|\zeta^{(k)}\right\|^2}{\left(\zeta^{(k)}\right)^T B\zeta^{(k)}}, \alpha_{max}\right\}, & \text{if } \left(\zeta^{(k)}\right)^T B\zeta^{(k)} \neq 0 \\ \alpha_{max}, & \text{otherwise.} \end{cases}$$

The operator *mid* is the define as the middle value of three scalar arguments and α_{min} and α_{max} are two given parameters. The parameter ν take the form

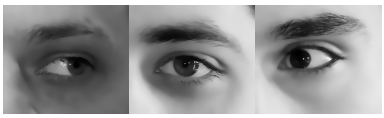
$$\nu^{(k)} = \begin{cases} mid\left\{0, \dfrac{\left(\zeta^{(k)}\right)^T \nabla Q\left(\mathbf{z}^{(k)}\right)}{\left(\zeta^{(k)}\right)^T B\zeta^{(k)}}, 1\right\}, & \text{if } \left(\zeta^{(k)}\right)^T B\zeta^{(k)} \neq 0, \\ 1, & \text{otherwise.} \end{cases}$$

The performance of the REN minimization along with comparisons is shown is Figure 2 for a sparse signal. We want to reconstruct a length-n sparse signal (in the canonical basis) from m observations, with $m \ll n$. The matrix $A_{m\times n}$ is build with independent samples of a standard Gaussian distribution and by ortho-normalizing the rows, while the original signal $\mathbf{x}^*$ contains 160 randomly placed $\pm spikes$ and the observation is defined as $\mathbf{y} = A\mathbf{x}^* + \boldsymbol{\kappa}$ with $\boldsymbol{\kappa}$ a Gaussian noise of variance $\sigma^2 = 10^{-4}$. The reconstruction of the original signal over the REN minimization problem produces a much lower mean squared error (MSE $= (1/n)\|\hat{\mathbf{x}} - \mathbf{x}^*\|$ with $\hat{\mathbf{x}}$ been an estimate of $\mathbf{x}^*$) equal to 3.499×10^{-05}, while the MSE given by the adaptive elastic model proposed in [16], [52] and LASSO are 5.194×10^{-05}, 4.791×10^{-05} and 1.445×10^{-04} respectively. Therefore, the proposed REN approach does an excellent job at locating the spikes.

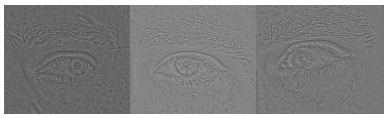
Remark 1 The iterative reweighted algorithm falls in the general class of Majorization - Minimization (MM) algorithms [23]. An interesting example of separable iterative reweighing for sparse solutions is presented in [7] where the selection

$$w_i^{(k+1)} \to \frac{1}{\left|x_i^{(k+1)}\right| + \epsilon} \tag{9}$$

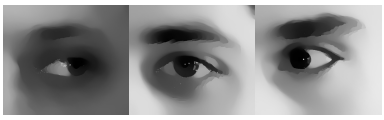
(a) Grayscale periocular images



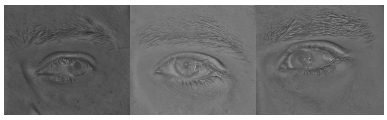
(b) Cartoon Component with 80 iterations



(c) Texture Component with 80 iterations



(d) Cartoon Component with 400 iterations



(e) Texture Component with 400 iterations

Fig. 3 Cartoon - Texture component for grayscale periocular images using a weighted TV model (3.1). (a) Grayscale periocular images. (b)-(c) Cartoon - Texture decomposition with 80 iterations. (d)-(e) Cartoon - Texture decomposition with 400 iterations.

is suggested. Here ϵ is generally chosen as a fixed, application-dependent constant. In the noiseless case, it is demonstrated based on [13] that this amounts to iteratively solving

$$\min_{\mathbf{x}} \sum_{i=1}^{n} \log\left(x_i + \epsilon\right), \quad \text{subj. to } A\mathbf{x} = \mathbf{y}, \tag{10}$$

and convergence to a local minimum or saddle point is guaranteed. In [45] the reweighting

$$w_i^{(k+1)} \to \frac{1}{\left(x_i^{(k+1)}\right)^2 + \epsilon \left|x_i^{(k+1)}\right|}, \tag{11}$$

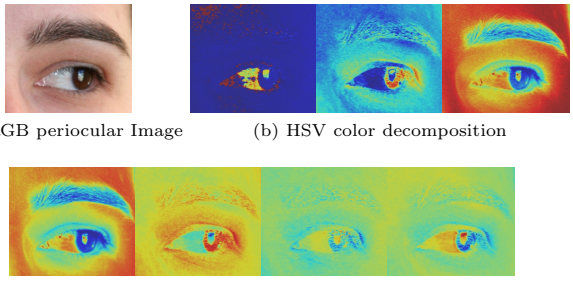
is also considered together with the case $\epsilon \to 0$. Related with the sparse solution of the model

$$\min_{\mathbf{x}} \left\{ \tau \|\mathbf{x}\|^2 + \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 \right\}, \tag{12}$$

the reweighting

$$w_i^{(k+1)} \to \frac{1}{\left(x_i^{(k+1)}\right)^2 + \epsilon^{(k+1)}} \tag{13}$$

is implemented in [8], where $\epsilon^{(k+1)} \geq 0$ is regularization factor the is reduced to zero as k becomes large.



(a) RGB periocular Image (b) HSV color decomposition

(c) CB color decomposition

Fig. 4 Different color decomposition for a given periocular image. (a) RGB color periocular image. (b) HSV color decomposition. (c) CB color decomposition.

3 Geometric and Color Spaces for Image Decomposition

3.1 Cartoon + Texture (CT) Space

The periocular images contain cartoon (smooth) and texture parts (small scale oscillations) which can be obtained using the total variation (TV) [37] model effectively. In this setting, the grayscale version of a periocular image is divided into two components representing the geometrical and texture parts. The TV based decomposition model is defined as an energy minimization problem,

$$\min_u \left\{ E_{TV}^{L^1}(u) = \int_\Omega g(\mathbf{x})|\nabla u|\,\mathrm{d}\mathbf{x} + \lambda \int_\Omega |u - I|\,\mathrm{d}\mathbf{x} \right\}$$

where I is the input grayscale image, and $g(\mathbf{x}) = \frac{1}{1+K|\nabla I|^2}$ is an edge indicator type function. Following [3] we use a splitting with an auxiliary variable v to obtain the following relaxed minimization,

$$\min_{u,v} \left\{ \tilde{E}_{TV}^{L^1}(u,v) = \int_\Omega g(\mathbf{x})|\nabla u|\,\mathrm{d}\mathbf{x} + \frac{1}{2\theta}\int_\Omega (u+v-I)^2\,\mathrm{d}\mathbf{x} + \lambda \int_\Omega |v|\,\mathrm{d}\mathbf{x} \right\} \quad (14)$$

After a solution u is computed, it is expected to get the representation $I \approx u+v$, where the function u represents the geometric cartoon part, the function v contains texture information, and the function g represent edges. The minimization (14) is achieved by solving the following alternating sub-problems based on the dual minimization technique:

1. Fixing v, the minimization problem in u is:

$$\min_u \left\{ \int_\Omega g(\mathbf{x})|\nabla u|\,\mathrm{d}\mathbf{x} + \frac{1}{2\theta}\|u+v-I\|^2_{L^2(\Omega)} \right\}. \quad (15)$$

The solution of (15) is given by $u = v - \theta \mathrm{div}\,\mathbf{p}$ where $\mathbf{p} = (p_1, p_2)$ satisfies $g(\mathbf{x})\nabla(\theta\,\mathrm{div}\,\mathbf{p} - (I-v)) - |\nabla(\theta \mathrm{div}\,\mathbf{p} - (I-v))|\mathbf{p} = 0$, solved using a fixed

point method: $\mathbf{p}^0 = 0$ and iteratively

$$\mathbf{p}^{n+1} = \frac{\mathbf{p}^n + \delta t \nabla(\mathrm{div}(\mathbf{p}^n) - (I - v)/\theta)}{1 + \frac{\delta t}{g(\mathbf{x})}|\nabla(\mathrm{div}(\mathbf{p}^n) - (I - v)/\theta)|}.$$

2. Fixing u, the minimization problem in v is:

$$\min_v \left\{ \frac{1}{2\theta}\|u + v - I\|^2_{L^2(\Omega)} + \lambda\|v\|_{L^1(\Omega)} \right\},$$

and the solution is found as

$$v = \begin{cases} I - u - \theta\lambda & \text{if } I - u \geq \theta\lambda, \\ I - u + \theta\lambda & \text{if } I - u \leq -\theta\lambda, \\ 0 & \text{if } |I - u| \leq \theta\lambda. \end{cases}$$

Figure 3 illustrates cartoon - texture decomposition of three grayscale periocular images for different iterations. As the number of iterations we notice that the cartoon component becomes smoother and texture component picks up more oscillations.

3.2 Color Spaces

For color periocular images we can obtain intensity and chromaticity decomposition which exploits color information. In computer vision there has been increasing interest in non flat image features that live on curved manifolds which are well suited for edge detection and enhancement in color and multichannel images [42]. The flatness concept is related to functions taking all possible values in an open set in a linear space. The chromaticity feature of color images is an example of non-flat features. Given a color periocular image $\mathbf{I} : \Omega \to \mathbb{R}^3$, the RGB representation is defined by a vector with three components $\mathbf{I} = (I_1, I_2, I_3)$. From the RGB color space, the chromaticity-brightness (CB) model arises by decomposing into the brightness component $\mathbf{B} : \Omega \to \mathbb{R}$ computed as $\mathbf{B} = |\mathbf{I}|$ and chromaticity components $\mathbf{C} = (C_1, C_2, C_3) : \Omega \to \mathbb{S}^2$ (where $\mathbb{S}^2$ is the unit sphere in $\mathbb{R}^3$) is computed by $C_i = I_i/\mathbf{B}$. We also make use of the Hue-Saturation-Value (HSV) color space commonly used since it is believed to be more natural and is related to human perception [49]. Figure 4 illustrates CB decomposition, and HSV color space conversions of a given RGB periocular image. In our experiments we compare grayscale CT decomposition and CB, RGB and HSV color space based decompositions along with the proposed REN model.

Signal $\hat{\mathbf{x}}$	Optimal Signal Recovery	Non-Optimal Signal Recovery
$\text{SCI}(\hat{\mathbf{x}}) > \beta \rightarrow$ Positive	True Positive (TP)	False Positive (FP)
$\text{SCI}(\hat{\mathbf{x}}) \leq \beta \rightarrow$ Negative	False Negative (FN)	True Negative (TN)

Table 1 Types of errors, according to the SCI value and the sparse signal reconstruction following Wright *et al.* [48] and Pillai *et al.* [34] models.

Mehtod	Feature	sens.	far	acc.	thres.	AUC	EER
REN (Proposed)	Grayscale (SRC)	90.05	8.55	90.99	0.1553	0.9643	0.0904
	Texture (SRC)	92.10	1.89	92.40	0.0756	0.9756	0.0589
	CT (Fusion)	99.90	7.18	98.77	0.1641	<u>0.9994</u>	<u>0.0018</u>
	CB (Fusion)	99.82	7.37	98.49	0.2333	0.9992	0.0061
	RGB (Fusion)	99.83	4.11	99.31	0.1670	0.9990	0.0020
	HSV (Fusion)	99.83	2.13	99.57	0.1832	0.9991	0.0019
Wright2009	Grayscale (SRC)	84.70	9.59	85.14	0.05642	0.9307	0.1529
Park2009	LBP	80.70	9.99	86.90	0.7468	0.9189	0.1553
	HOG	69.29	9.99	83.11	0.6421	0.8656	0.2088
	SIFT	86.00	9.36	88.96	0.0477	0.9453	0.1232
	Fusion	90.58	9.99	90.21	0.1052	0.9564	0.0954
Bharadwaj2010	GIST	75.56	9.99	85.20	0.7623	0.8927	0.1846
	ULBP	85.82	9.99	88.61	0.8673	0.9259	0.1311
	Fusion	83.96	9.99	88.00	0.8008	0.9235	0.1386

Table 2 AUC and EER values, as well as the best sensitivity for far $\leq 10\%$ for left side periocular images. The Underline fonts indicate the best model observed.

4 Experiments and Discussion

4.1 Performance Measures

Images were down-sampled to 10×9 pixels and stored in "png" format. The resulting sensitivity and specificity values were considered, obtaining the Receiver Operating Characteristic curves (ROC). In this case, given a signal $\hat{\mathbf{x}}$, if $\text{SCI}(\hat{\mathbf{x}}) > \beta$, the classifier outputs a positive response (P), otherwise a negative (N) result. For a fixed β, the sensitivity corresponds to the proportion of signals correctly detected by the SRC algorithm, whereas specificity counts the proportion for which the corresponding SCI values are bellow β, where β is an accepted threshold value.

$$\text{sensitivity} = \frac{\#\text{TP}}{\#\text{TP} + \#\text{ FN}} \quad \text{and} \quad \text{specificity} = \frac{\#\text{ TN}}{\#\text{ TN} + \#\text{FP}},$$

where TP, FP, TN and FN correspond to the True Positive, False Positive, True Negative and False Negative, respectively. Table 1 summarizes these notions, combining the different classes of periocular signals and their relation with the classifier induced by the minimal reconstruction error and the accumulated SCI value. The overall accuracy is given by:

$$\text{accuracy} = \frac{\#\text{ TN} + \#\text{ TP}}{\#\text{ TN} + \#\text{ FP} \ + \#\text{ TP} + \#\text{ FN}}.$$

In a ROC plot, the optimal recognition method would yield a point in the upper-left corner, corresponding to full sensitivity (no false negatives) and full

specificity (no false positives). The statistical correlation between the outputs given by each channels considered in our method was also assessed. Considering that eventual dependences will be linear, the Pearson's sample correlation was used for that purpose. Given a pair of samples, the correlation coefficient is given by:

$$r(\hat{\mathbf{x}}^{(1)}, \hat{\mathbf{x}}^{(2)}) = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{\hat{x}_i^{(1)} - \bar{\mathbf{x}}^{(1)}}{\sigma_{\hat{\mathbf{x}}^{(1)}}} \right) \left(\frac{\hat{x}_i^{(2)} - \bar{\mathbf{x}}^{(2)}}{\sigma_{\hat{\mathbf{x}}^{(2)}}} \right),$$

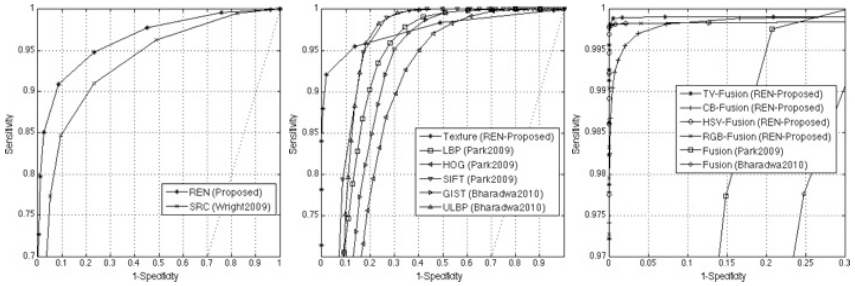
where $\hat{x}_i^{(1)}$, $\hat{x}_i^{(2)}$ denote the systems outputs, $\bar{\mathbf{x}}^{(1)}$, $\bar{\mathbf{x}}^{(2)}$ are the sample means and $\sigma_{\hat{\mathbf{x}}^{(1)}}$, $\sigma_{\hat{\mathbf{x}}^{(2)}}$ the standard deviations.

4.2 Results

For our first experiment, we focus on only left side periocular images. Six samples from 150 different subjects were used, such that one image per class was randomly chosen as probe and the remaining five samples included in the dictionary. Experiments were repeated, changing the image used as probe (per subject). Hence, 100 dictionaries with dimension 90×750 were considered, each one tested in 150 probe samples.

Results are summarized in Table 2 in terms of true and false positive rates where the best sensitivity (sens.) and corresponding accuracy (acc.) for far (=1-specificity) $\leq 10\%$ have been computed for various schemes and models studied here. The proposed reweighed elastic net demonstrates to be superior than the original SRC approach over grayscale impulses. In this case the area under the curve (AUC) and the equal error rate (EER) are equal to 0.9643 and 0.0904 for our model, against 0.9307 and 0.1529 produced by the original SRC model. The proposed models approximates more to the *optimal performance* point (complement of specificity = 0, sensitivity = 1). For the REN approach applied to the grayscale and the texture components alone the minimal distance from the ROC values to the $(0, 1)$ point was of 0.1511 and 0.0812 respectively, while the value 0.1805 was observed for the classical SRC model. In relation to other image representation components, the minimal distance from the ROC values to the $(0, 1)$ point was of 0.0022 0.0106, 0.0023, 0.0019 for the CT, CB, RGB and HSV spaces.

Comparisons have been carried out by implementing the well known models of Park *et al.* [33] and Bharadwaj *et al.* [2]. Even both models make full use of local and global periocular information to perform recognition, they have shown not to improve better than our approach. In our experiments, we have compared the AUC as well as the EER values in the case REN model approach uses the texture periocular components as feature extraction, against those features used in Park *et al.* [33] and Bharadwaj *et al.* [2]. For the comparison models the highest AUC is equal to 0.9564 and lowest EER is equal to 0.0954 when applying their fusion techniques. Meanwhile, using the texture information provided by the cartoon - texture space, our model got the values



(a) ROC curves - Left Side (b) ROC curves - Left Side (c) ROC curves - Left Side

Fig. 5 ROC curves for periocular images recognition. (a) ROC curves for the original REN approach and the SRC model and the REN approach. (b) ROC curves for the REN approach applied to the texture components together with different features extrated by Park *et al.* [33] and Bharadwaj *et al.* [2]. (c) ROC curves for the REN model applied to the proposed fusion over the different geometry and color spaces, as well as the fusion implemented in Park *et al.* [33] and Bharadwaj *et al.* [2].

	Grayscale	Texture	CT	CB	RGB	HSV
Grayscale	1	0.7173	0.5331	0.4272	0.7139	0.6230
Texture	-	1	0.6041	0.6285	0.9776	0.9043
CT	-	-	1	0.2206	0.6134	0.6146
CB	-	-	-	1	0.6146	0.7213
RGB	-	-	-	-	1	0.9180
HSV	-	-	-	-	-	1

Table 3 Pearson's sample correlation coefficients between the left side responses given by the recognition algorithms using the REN model with various components studied here.

0.9756 and 0.0589 for the AUC and EER, respectively. Our fusion method using different spaces completely describing the geometry and color periocular feature have also shown to reached great statistical values in comparison to those values got it by Park *et al.* [33] and Bharadwaj *et al.* [2] approaches. In this case, the highest AUC and the lowest EER values are given by the CT space with values 0.9994 and 0.0018, see Figure 5.

As it can be observed from Table 3, the proposed REN model applied to grayscale in texture setting are in high correlation when compared to the signals recovered in the CT, RGB and HSV spaces. The result is due to the high accuracy rates achieved over these image representations. Similarly the signals recovered in the CB space are in low correlation with the signals lying in the grayscale setting and the CT space, and in high correlation with the signals computed over texture domain alone. This is because the chromaticity components lying in the unit sphere $\mathbb{S}^2$ have the advantage of depicting nonlinear features in different directions and therefore both strong and weak edges are distributed and represented along chromaticity components. Also, it should be noted the strong correlation between the outputs given by the fusion model when using exclusively color components. This is also can be explained, as the skin region comprises a large majority of the periocular region (see Figure 4). It is particularly interesting to observe that the positive (and small)

correlation values between the signals are obtained when using different color spaces representation, pointing for a complementarity that might contribute for the outperforming results of the method proposed in this paper. Although the CT space produces good recognition rates, its computed signals are in low correlation with respect to other signals over different domains, owing to the fact that CT space is given by geometric information in case of cartoon component, whereas weak and strong edges describe texture components, see Figure 3.

5 Conclusions

This paper describes a novel re-weighted elastic net (REN) model that improves the sparsity of representations in periocular regions which is an emerging biometric trait with high potential to handle data acquired under uncontrolled conditions. From this perspective, we have fused multiple sparse representations, associated with various spaces from different domains in geometry and color, which allow us to faithfully handle distortions in periocular images such as blur and occlusions. Our experiments were carried out in the highly challenging images of the UBIRIS.v2 dataset, and allowed us to observe consistent improvements in performance, when compared to the classical sparse representation model, and state-of-the-art periocular recognition algorithms. In addition, theoretical existence results have been proved for the REN minimization problem, mainly emphasizing our approach is good in the sense it performs as well as if the true underlying model were given in advance. As far as numerical approximation is concerned, the REN model is expressed as a quadratic programing (QP) expediting the implementation of the proposed gradient projection (GP) algorithm and providing good results. Evaluating the proposed methodology on other biometric traits (iris, face, etc.) and on different databases are our future works.

A Existence of Solution

We state necessary and sufficient conditions for the existence of a solution for the proposed model (6). We follow the notations used in [44,18]. Suppose that $A_i = (A_{1i}, \ldots, A_{mi})^T$, $i = 1, \cdots, n$ are the linear independent predictors and $\mathbf{y} = (y_1, \cdots, y_m)^T$ is the response vector. Let $A = [A_1, \cdots, A_n]$ be the predictor matrix. In terms of ℓ^1 and ℓ^2 norms, we rewrite the minimization problem in (6) as,

$$\min_{\mathbf{x}} \left\{ m\|W\mathbf{x}\|_1 + \frac{m}{2}\|(1-W)\mathbf{x}\|_2^2 + \frac{1}{2}\|\mathbf{y} - A\mathbf{x}\|_2^2 \right\}. \tag{16}$$

Let us denote by $\mathbf{x}^*$ and $\hat{\mathbf{x}}$ the real and estimated solution of (16) respectively. Given $\mathcal{I} = supp(\mathbf{x}^*) = \{i : x_i^* \neq 0\}$, we define the block-wise form matrix

$$A_{\mathcal{I},\mathcal{I}^c} = \frac{1}{m} \begin{pmatrix} A_{\mathcal{I}}^T A_{\mathcal{I}} & A_{\mathcal{I}}^T A_{\mathcal{I}^c} \\ A_{\mathcal{I}^c}^T A_{\mathcal{I}} & A_{\mathcal{I}^c}^T A_{\mathcal{I}^c} \end{pmatrix},$$

where $A_{\mathcal{I}}$ ($A_{\mathcal{I}^c}$) is a $m \times \#\mathcal{I}$ ($m \times \#\mathcal{I}^c$) matrix formed by concatenating the columns $\{A_i : i \in \mathcal{I}\}$ ($\{A_i : i \in \mathcal{I}^c\}$) and $A_{\mathcal{I}}^T A_{\mathcal{I}}$ is assumed to be invertible.

First we assume that there exist $\hat{\mathbf{x}} \in \mathbb{R}^n$ satisfying (16) and $sign(\hat{\mathbf{x}}) = sign(\mathbf{x}^*)$. Lets define $\mathbf{b} = W_{\mathcal{I}} sign(\mathbf{x}^*_{\mathcal{I}})$ together with the set,

$$\mathcal{D} = \left\{ \mathbf{d} \in \mathbb{R}^n : \begin{cases} d_i = b_i, & \text{for } \hat{x}_i \neq 0 \\ |d_i| \leq w_i, & \text{otherwise} \end{cases} \right\}.$$

From the Kauush-Kuhn-Tucker (KKT) conditions we obtain

$$\begin{cases} A_i^T(\mathbf{y} - A\hat{\mathbf{x}}) - m(1-w_i)^2 \hat{x}_i = m w_i sign(x_i^*), & \text{if } \hat{x}_i \neq 0 \\ \left| A_i^T (\mathbf{y} - A\hat{\mathbf{x}}) \right| \leq m w_i, & \text{otherwise} \end{cases}$$

which can be rewritten as,

$$A_i^T A(\hat{\mathbf{x}} - \mathbf{x}^*) - A_i^T \boldsymbol{\kappa} + m(1-w_i)^2 \hat{x}_i + m d_i = 0, \tag{17}$$

for some $\mathbf{d} \in \mathcal{D}$ with components d_i, $i = 1, \ldots, n$. By substituting the equality $\mathbf{y} = A\mathbf{x}^* + \boldsymbol{\kappa}$. From the above Eqn. (17) the following two equations arise:

$$A_{\mathcal{I}}^T A_{\mathcal{I}}(\hat{\mathbf{x}}_{\mathcal{I}} - \mathbf{x}^*) - \frac{A_{\mathcal{I}}^T \boldsymbol{\kappa}}{m} + (1-W)^2 \hat{\mathbf{x}}_{\mathcal{I}} = -\mathbf{b}, \tag{18}$$

$$A_{\mathcal{I}^c}^T A_{\mathcal{I}}(\hat{\mathbf{x}}_{\mathcal{I}} - \mathbf{x}^*) - \frac{A_{\mathcal{I}^c}^T \boldsymbol{\kappa}}{m} = -\mathbf{d}_{\mathcal{I}^c}. \tag{19}$$

Solving for $\mathbf{x}_{\mathcal{I}}$ in (18) and replacing in (19) to get $\mathbf{b}$ in terms of $\mathbf{x}_{\mathcal{I}}$ leave us with

$$\hat{\mathbf{x}}_{\mathcal{I}} = \left(A_{\mathcal{I}}^T A_{\mathcal{I}} + (1-W)^2 \right)^{-1} \left(A_{\mathcal{I}}^T A_{\mathcal{I}} \mathbf{x}^*_{\mathcal{I}} + \frac{A_{\mathcal{I}} \boldsymbol{\kappa}}{m} - \mathbf{b} \right), \tag{20}$$

$$A_{\mathcal{I}^c}^T A_{\mathcal{I}} \left(\left(A_{\mathcal{I}}^T A_{\mathcal{I}} + (1-W)^2 \right)^{-1} \left(A_{\mathcal{I}}^T A_{\mathcal{I}} \mathbf{x}^*_{\mathcal{I}} + \frac{A_{\mathcal{I}}^T \boldsymbol{\kappa}}{m} - \mathbf{b} \right) - \mathbf{x}^*_{\mathcal{I}} \right) - \frac{A_{\mathcal{I}^c}^T \boldsymbol{\kappa}}{m} = -\mathbf{b}. \tag{21}$$

From (20) and (21), we finally get the next two equations:

$$sign\left(\left(A_{\mathcal{I}}^T A_{\mathcal{I}} + (1-W)^2 \right)^{-1} \left(A_{\mathcal{I}}^T A_{\mathcal{I}} \mathbf{x}^*_{\mathcal{I}} + \frac{A_{\mathcal{I}}^T \boldsymbol{\kappa}}{m} - \mathbf{b} \right) \right) = sign(\mathbf{x}^*_{\mathcal{I}}) \tag{22}$$

and

$$\left| A_i^T A_{\mathcal{I}} \left(\left(A_{\mathcal{I}}^T A_{\mathcal{I}} + (1-W)^2 \right)^{-1} \left(A_{\mathcal{I}}^T A_{\mathcal{I}} \mathbf{x}^*_{\mathcal{I}} + \frac{A_{\mathcal{I}}^T \boldsymbol{\kappa}}{m} - \mathbf{b} \right) - \mathbf{x}^*_{\mathcal{I}} \right) - \frac{A_i^T \boldsymbol{\kappa}}{m} \right| \leq w_i, \tag{23}$$

for $i \in \mathcal{I}^c$.

Now, let us assume that equations (22) and (23) both hold. It will be proved there exist $\hat{\mathbf{x}} \in \mathbb{R}^n$ satisfying $sing(\hat{\mathbf{x}}) = sign(\mathbf{x}^*)$. Setting $\hat{\mathbf{x}} \in \mathbb{R}^n$ satisfying $\hat{\mathbf{x}}_{\mathcal{I}^c} = \mathbf{x}^*_{\mathcal{I}^c} = 0$ and

$$\mathbf{x}_{\mathcal{I}} = \left(A_{\mathcal{I}}^T A_{\mathcal{I}} + (1-W)^2 \right)^{-1} \left(A_{\mathcal{I}}^T A_{\mathcal{I}} \mathbf{x}^*_{\mathcal{I}} + \frac{A_{\mathcal{I}}^T \boldsymbol{\kappa}}{m} - \mathbf{b} \right),$$

which guarantees the equality $sign(\hat{\mathbf{x}}_{\mathcal{I}}) = sign(\mathbf{x}^*_{\mathcal{I}})$ due to (22). In the same manner, we define $\mathbf{d} \in \mathbb{R}^n$ satisfying $\mathbf{d}_{\mathcal{I}} = \mathbf{b}$ and

$$\mathbf{d}_{\mathcal{I}^c} = -\left(A_{\mathcal{I}^c}^T A_{\mathcal{I}} \left(\left(A_{\mathcal{I}}^T A_{\mathcal{I}} + (1-W)^2 \right)^{-1} \left(A_{\mathcal{I}}^T A_{\mathcal{I}} \mathbf{x}^*_{\mathcal{I}} + \frac{A_{\mathcal{I}}^T \boldsymbol{\kappa}}{m} - \mathbf{b} \right) - \mathbf{x}^*_{\mathcal{I}} \right) - \frac{A_{\mathcal{I}^c}^T \boldsymbol{\kappa}}{m} \right),$$

implying from (23) the inequality $|d_i| \leq w_i$ for $i \in \mathcal{I}^c$ and therefore $\mathbf{d} \in \mathcal{D}$. From previous, we have found a point a point $\hat{\mathbf{x}} \in \mathbb{R}^n$ and $\mathbf{d} \in \mathcal{D}$ satisfying (18) and (19) respectively or equivalently (17). Moreover, we also have the equality $sign(\hat{\mathbf{x}}) = sign(\mathbf{x}^*)$. Under these assertions we can prove the sign recovery property of our model as illustrated next.

B Sign Recovery Property

Under some regularity conditions on the proposed REN model, we intend to give an estimation for which the event $sign(\hat{\mathbf{x}}) = sign(\mathbf{x}^*)$ is true. Following similar notations in [52, 50], we intend to prove that our model enjoys the following probabilistic property:

$$Pr\left(\min_{i\in\mathcal{I}} |\hat{x}_i| > 0\right) \to 1. \tag{24}$$

For theoretical analysis purposes, the problem (6) is written as

$$\min_{\mathbf{x}} \left\{ \|W\mathbf{x}\|_1 + \|(1-W)\mathbf{x}\|_2^2 + \|\mathbf{y} - A\mathbf{x}\|_2^2 \right\}.$$

The following regularity conditions are also assumed:

1. Denoting with $\Lambda_{\min}(S)$ and $\Lambda_{\max}(S)$ the minimum and maximum eigenvalues of a symmetric matrix S, we assume the following inequalities hold:

$$\theta_1 \leq \Lambda_{\min}\left(\frac{1}{m}A^TA\right) \leq \Lambda_{\max}\left(\frac{1}{m}A^TA\right) \leq \theta_2,$$

   where θ_1 and θ_2 are two positive constants.
2. $\lim_{m\to\infty} \dfrac{\log(n)}{\log(m)} = \nu$ for some $0 \leq \nu < 1$
3. $\lim_{m\to\infty} \sqrt{\dfrac{m}{n}} \dfrac{1}{\max_{i\in\mathcal{I}} w_i} = \infty.$

Let

$$\tilde{\mathbf{x}} = arg\min_{\mathbf{x}} \left\{ \|\mathbf{y} - A\mathbf{x}\|_2^2 + \|(1-W)\mathbf{x}\|_2^2 \right\}. \tag{25}$$

By using the definitions of $\hat{\mathbf{x}}$ and $\tilde{\mathbf{x}}$, the next two inequalities arise

$$\|\mathbf{y} - A\tilde{\mathbf{x}}\|_2^2 + \|(1-W)\,\tilde{\mathbf{x}}\|_2^2 \geq \|\mathbf{y} - A\tilde{\mathbf{x}}\|_2^2 + \|(1-W)\,\tilde{\mathbf{x}}\|_2^2 \tag{26}$$

and

$$\|\mathbf{y} - A\tilde{\mathbf{x}}\|_2^2 + \|(1-W)\,\tilde{\mathbf{x}}\|_2^2 + \sum_{i=1}^{n} w_i|\tilde{x}_i| \geq \|\mathbf{y} - A\hat{\mathbf{x}}\|_2^2 + \|(1-W)\,\hat{\mathbf{x}}\|_2^2 + \sum_{i=1}^{n} w_i|\hat{x}_i|. \tag{27}$$

The combination of equations (26) and (27) give

$$\begin{aligned}\sum_{i=1}^{n} w_i(|\tilde{x}_i| - |\hat{x}_i|) &\geq \|\mathbf{y} - A\hat{\mathbf{x}}\|_2^2 + \|(1-W)\hat{\mathbf{x}}\|_2^2 - \|\mathbf{y} - A\tilde{\mathbf{x}}\|_2^2 - \|(1-W)\tilde{\mathbf{x}}\|_2^2 \\ &= (\hat{\mathbf{x}} - \tilde{\mathbf{x}})^T \left(A^TA + (1-W)^2\right)(\hat{\mathbf{x}} - \tilde{\mathbf{x}})\end{aligned} \tag{28}$$

On the other hand

$$\sum_{i=1}^{n} w_i\,(|\tilde{x}_i| - |\hat{x}_i|) \leq \sum_{i=1}^{n} w_i\,|\tilde{x}_i - \hat{x}_i| \leq \sqrt{\sum_{i=1}^{n} w_i^2}\,\|\tilde{\mathbf{x}} - \hat{\mathbf{x}}\|_2 \tag{29}$$

By combining equations (28) and (29) we get

$$\begin{aligned}\Lambda_{min}\left(\left(A^TA\right) + (1-W)^2\right)\|\hat{\mathbf{x}} - \tilde{\mathbf{x}}\|_2^2 &\leq (\hat{\mathbf{x}} - \tilde{\mathbf{x}})^T \left(A^TA + (1-W)^2\right)(\hat{\mathbf{x}} - \tilde{\mathbf{x}}) \\ &\leq \sqrt{\sum_{i=1}^{n} w_i^2}\,\|\tilde{\mathbf{x}} - \hat{\mathbf{x}}\|_2\end{aligned}$$

which together with the identity

$$0 \leq \theta_1 \leq \Lambda_{min}\left(A^T A\right) \leq \Lambda_{min}\left(\left(A^T A\right) + (1-W)^2\right)$$

allow us to prove

$$\|\hat{\mathbf{x}} - \tilde{\mathbf{x}}\|_2 \leq \frac{\sqrt{\sum_{i=1}^n w_i^2}}{\Lambda_{min}\left(A^T A\right)}, \tag{30}$$

Let us notice that

$$\begin{aligned} E\left(\|\tilde{\mathbf{x}} - \mathbf{x}^*\|_2^2\right) &= E\left(-\left(A^T A + (1-W)^2\right)^{-1}(1-W)^2\,\mathbf{x}^* + \left(A^T A + (1-W)^2\right)^{-1} A^T \boldsymbol{\kappa}\right) \\ &\leq 2\frac{\|(1-W)\mathbf{x}^*\|_2^2 + n\Lambda_{\max}\left(A^T A\right)\sigma^2}{\Lambda_{\min}\left(A^T A\right)} \end{aligned} \tag{31}$$

From equations (30) and (31) we conclude that

$$\begin{aligned} E\left(\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2\right) &\leq 2\left(E\left(\|\tilde{\mathbf{x}} - \mathbf{x}^*\|_2^2\right) - E\left(\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2^2\right)\right) \\ &\leq 4\frac{\|(1-W)\,\mathbf{x}^*\|_2^2 + n\Lambda_{\max}(A^T A)\sigma^2 + E\left(\sum_{i=1}^n w_i^2\right)}{\Lambda_{\min}\left(A^T A\right)}. \end{aligned} \tag{32}$$

Let $\eta = \min_{i\in\mathcal{I}} |x_i^*|$ and $\hat{\eta} = \max_{i\in\mathcal{I}} w_i$. Because of (30),

$$\|\hat{\mathbf{x}}_{\mathcal{I}} - \tilde{\mathbf{x}}_{\mathcal{I}}\|_2^2 \leq \frac{\sqrt{n}\hat{\eta}}{\theta_1 m}.$$

Then

$$\min_{i\in\mathcal{I}} |x_i^*| > \min_{i\in\mathcal{I}} |\tilde{x}_i| - \frac{\sqrt{n}\hat{\eta}}{\theta_1 m} > \min_{i\in\mathcal{I}} |\hat{x}_i| - \|\tilde{\mathbf{x}}_{\mathcal{I}} - \mathbf{x}_{\mathcal{I}}^*\|_2 - \frac{\sqrt{n}\hat{\eta}}{\theta_1 m}. \tag{33}$$

Now, we notice that

$$\frac{\sqrt{n}\hat{\eta}}{\theta_1 m} = O\left(\frac{1}{\sqrt{n}}\right)\left(\sqrt{\frac{n}{m}}\eta^{-1}\right)(\hat{\eta}\eta).$$

Since

$$\begin{aligned} E\left((\hat{\eta}\eta)^2\right) &\leq 2\eta^2 + 2\eta^2 E\left((\hat{\eta}-\eta)^2\right) \leq 2\eta^2 + 2\eta^2 E\left(\|\hat{\mathbf{x}} - \mathbf{x}^*\|^2\right) \\ &\leq 2\eta^2 + 8\eta^2\frac{\left\|(1-W)^2\,\mathbf{x}^*\right\|_2^2 + \theta_2 nm\sigma^2 + E\left(\sum_{i=1}^n w_i^2\right)}{\theta_1 m} \end{aligned}$$

and $\eta^2 m/n \to \infty$ as long as $m \to \infty$, it follows that

$$\frac{\sqrt{n}\hat{\eta}^{-1}}{\theta_1 m} = o\left(\frac{1}{\sqrt{n}}\right) O_{Pr}(1). \tag{34}$$

By using (32), we derive

$$E\left(\left\|\hat{\tilde{\mathbf{x}}}_{\mathcal{I}} - \mathbf{x}_{\mathcal{I}}^*\right\|_2^2\right) \leq 4\frac{\|\,(1-W)^2\,\mathbf{x}^*\|_2 + \theta_2 nm\sigma^2}{(\theta_1 m)^2} = \sqrt{\frac{n}{m}} O_{Pr}(1). \tag{35}$$

Substituting (34) and (35) in (33) allow us to conclude that

$$\min_{i\in\mathcal{I}} |x_i^*| > \eta - \sqrt{\frac{n}{m}} O_{Pr}(1) - o\left(\frac{1}{\sqrt{n}}\right) O_{Pr}(1).$$

Then (24) holds.

Remark 2 There is special interest in applying the REN model in the case the data satisfies the condition $n \gg m$. For the LASSO model it was suggested in [6] to make use of the Dantzig selector which can achieve the ideal estimation up to a $log(n)$ factor. In [12] a performing of the Dantzig selector called the Sure Independence Screening (SIS) was introduced in order to reduce the ultra-high dimensionality. We remark that the SIS technique can be combined with the REN model (6) for dealing the case $n \gg m$. Then previous computations can be still applied to reach the sign recovery property.

References

1. J. Adams, D.L. Woodard, G. Dozier, P. Miller, K. Bryant, and G. Glenn. Genetic-based type ii feature extraction for periocular biometric recognition: Less is more. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 205 –208, August 2010.
2. S. Bharadwaj, H.S. Bhatt, M. Vatsa, and R. Singh. Periocular biometrics: When iris recognition fails. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1 –6, September 2010.
3. X. Bresson, S. Esedoglu, P. Vandergheynst, J. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28(2):151–167, 2007.
4. E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.
5. E. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.
6. E. Candes and T. Tao. The dantzig selector: statistical estimation when p is much larger than n. *The Annals of Statistics*, 35(6):2392–2404, 2007.
7. E. Candès, M. Wakin, and Stephen P. Boyd. Enhancing sparsity by reweighted ℓ^1- minimization. *Journal of Fourier Analysis and Applications*, 14(5):877–905, 2008.
8. R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. In *33rd International Conference on Acoustics, Speech, and Signal Processing*, 2008.
9. S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998.
10. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *In CVPR*, pages 886–893, 2005.
11. D. Donoho. For most large underdetermined systems of equations, the minimal ℓ^1-norm near-solution approximates the sparsest near-solution. *Communications on Pure and Applied Mathematics*, 59(7):907–934, 2006.
12. J. Fan and J. Lv. Sure indepence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5):849–911, 2008.
13. M. Fazel, H. Hindi, and S. Boyd. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. In *Proceedings of the American Control Conference*, pages 2156–2162, 2003.
14. M. Figueiredo, R. Nowak, and S. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problem. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):586–597, 2007.
15. J. J. Fuchs. Multipath time-delay detection and stimation. *IEEE Transactions on Signal Processing*, 47(1):237–243, 1999.
16. D. Hong and F. Zhang. Weigthed elastic net model for mass spectrometry image processing. *Mathematical Modelling of Natural Phenomena*, 5(3):115–133, 2010.
17. A. K. Jain, P. Flynn, and A. Ross (Eds). *Handbook of biometrics*. Springer-Verlag, New York, USA, 2007.
18. J. Jia and B. Yu. On model salection consistency of the elastic net when $p \gg n$. *Statistica Sinica*, 20:595–611, 2010.
19. R. Jiang, D. Crookes, and N. Lie. Face recognition in global harmonic subspace. *IEEE Transactions on Information Forensics and Security*, 5(3):416–424, 2010.
20. F. Juefei-Xu, M. Cha, J.L. Heyman, S. Venugopalan, R. Abiantun, and M. Savvides. Robust local binary pattern feature sets for periocular biometric identification. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1 –8, sept. 2010.
21. F. Juefei-Xu, K. Luu, M. Savvides, T.D. Bui, and C.Y. Suen. Investigating age invariant face recognition based on periocular biometrics. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1 –7, October 2011.
22. S. Kang and R. March. Variational models for image colorization via chromaticity and brightness decomposition. *IEEE Transactions on Image Processing*, 16(9):2251–2261, 2007.

23. K. Lange. *Optimization*. Springer Text in Stadistic. Springer, New York, USA, 2004.
24. D.G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, November 2004.
25. P.E. Miller, A.W. Rawls, S.J. Pundlik, and D.L. Woodard. Personal identification using periocular skin texture. In *Proceedings of the 2010 ACM Symposium on Applied Computing*, SAC '10, pages 1496–1500, New York, NY, USA, 2010. ACM.
26. T. Ojala, M. Pietikainen, and D. Harwood. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision amp; Image Processing., Proceedings of the 12th IAPR International Conference on*, volume 1, pages 582 –585 vol.1, October 1994.
27. T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971 –987, July 2002.
28. A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42:145–175, 2001.
29. U. Park and A. K. Jain. Face matching and retrieval using soft biometrics. *IEEE Transactions on Information Forensics and Security*, 5(3):406–415, 2010.
30. U. Park, R. R. Jillela, A. Ross, and A. K. Jain. Periocular biometircs in the visible spectrum. *IEEE Transactions on Information Forensics and Security*, 6(1):96–106, 2011.
31. U. Park, R.R. Jillela, A. Ross, and A.K. Jain. Periocular biometrics in the visible spectrum. *Information Forensics and Security, IEEE Transactions on*, 6(1):96 –106, 2011.
32. U. Park, A. Ross, and A. K. Jain. Periocular bimetrics in the visible spectrum: A feasibility study. In *IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 153–158, Virginia, USA, 2009.
33. U. Park, A. Ross, and A.K. Jain. Periocular biometrics in the visible spectrum: A feasibility study. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on*, pages 1 – 6, September 2009.
34. J. K. Pillai, V. M. Patel, R. Chellappa, and N. K. Ratha. Secure and robust iris recognition using random projections and sparse representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(9):1877–1893, 2011.
35. H. Proença and L. Alexandre. Toward covert iris biometric recognition: experimental results from the NICE contests. *IEEE Transactions on Information Forensics and Security*, 7(2):798–808, 2012.
36. H. Proena, S. Filipe, R. Santos, J. Oliveira, and L.A. Alexandre. The ubiris.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(8):1529 –1535, August 2010.
37. L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1–4):259–268, 1992.
38. G. Santos and H. Proena. Periocular biometrics: An emerging technology for unconstrained scenarios. In *Proceedings of the IEEE Symposium on Computational Intelligence in Biometrics and Identity Management – CIBIM 2013*, pages 14–21, April 2013.
39. T. Serafini, G. Zanghirati, and L. Zanni. Gradient projection methods for large quadratic programs and applications in training support vector machines. *Optimization Methods and Software*, 20(2-3):353–378, 2003.
40. S. Shekhar, V. M. Patel, N. M. Nasrabadi, and R. Chellappa. Joint sparse representation for robust multimodal biometrics recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(99), 2013.
41. R. Sznitman and B. Jedynak. Active testing for face detection and localization. *IEEE Transactions on Pattern Analysis and Machine Inteligence*, 32(10):1914–1920, 2010.
42. B. Tang, G. Sapiro, and V. Caselles. Color image enhancement via chromaticity diffusion. *IEEE Transactions on Image Processing*, 10(5):701–707, 2001.
43. R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, 58(1):267–288, 1996.

44. M. Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ^1-constrained quadratic programming (lasso). *IEEE Transactions on Information Theory*, 55(5):2183–2202, 2009.
45. D. Wipf and S. Nagarajan. Iterative reweighted ℓ^1 and ℓ^2 methods for finding sparse solutions. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):317–329, 2010.
46. D. L. Woodard, S. Pundlik, P. Miller, R. Jillela, and A. Ross. On the fusion of periocular and iris biometrics in non-ideal imagery. In *IEEE International Conference on Pattern Recognition (ICPR)*, pages 201–204, Istanbul, Turkey, 2010.
47. D.L. Woodard, S. Pundlik, P. Miller, R. Jillela, and A. Ross. On the fusion of periocular and iris biometrics in non-ideal imagery. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 201 –204, August 2010.
48. J. Wright, A. Y. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.
49. G. Wyszecki and W. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulas*. Wiley, New York, NY, USA, 1982.
50. H. Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006.
51. H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67(2):301–320, 2005.
52. H. Zou and H. Zhang. On the adaptive elastic-net with a diverging number of parameters. *The Annals of Statistics*, 37(4):1733–1751, 2009.

Appendix B

Iris Recognition: Analyzing the Distribution of the Iriscodes Concordant Bits

B.1 Overview

This appendix consists of the following article:

Iris Recognition: Analyzing the Distribution of the Iriscodes Concordant Bits
Gil Santos and Hugo Proença
IEEE Proceedings of the 3rd International Congress on Image and Signal Processing - CISP 2010, October 16-18, 2010, Yantai, China.

Iris Recognition: Analyzing the Distribution of the Iriscodes Concordant Bits

Gil Santos and Hugo Proença
Department of Computer Science
Instituto de Telecomunicações - Networks and Multimedia Group
University of Beira Interior, Covilhã, Portugal
Email: gmelfe@ubi.pt, hugomcp@di.ubi.pt

***Abstract*—The growth in practical applications for iris biometrics has been accompanied by relevant developments in the underlying algorithms and techniques. Efforts are being made to minimize the tradeoff between the recognition error rates and data quality, acquired in the visible wavelength, in less controlled environments, over simplified acquisition protocols and varying lighting conditions. This paper presents an approach that can be regarded as an extension to the widely known Daugman's method. Its basis is the analysis of the distribution of the concordant bits when matching iriscodes on both the spatial and frequency domains. Our experiments show that this method is able to improve the recognition performance over images captured in less constrained acquisition setups and protocols. Such conclusion was drawn upon trials conducted for multiple datasets.**

I. Introduction

The use of the iris as main biometric trait is emerging as one of the most recommended, due not only to the possibility of contactless data acquisition and to its circular and planar shape that makes easy the detection, segmentation and compensation for off-angle capturing, but also for its predominately randotypic appearance. Although these factors contribute to the high effectiveness of the deployed iris recognition systems, their typical scenarios are quite constrained: subjects stop-and-stare relatively close to the acquisition device, while their eyes are illuminated by a near-infrared light source, enabling the acquisition of good quality data. Remarkably, several researchers are trying to minimize the constraints associated with this process, in a way often referred as *non-cooperative iris recognition*.

Traditional iris recognition methods are based on the statistical Pattern Recognition paradigm and regard the biometric signatures as points of hyper-dimensional spaces. Here, a match occurs when the distance between two signatures is lower than a threshold. However, dealing with degraded data might lead to huge deformations of the feature space and significant increases of the error rates. In this paper we propose a method that accounts for the spatial and frequency analysis of the bits that are in agreement when comparing two biometric signatures (iriscodes). The goal is to increase the robustness to degraded data, captured in unconstrained acquisition setups. The Daugman's approach, widely known for it's low error rates and commercially deployed in iris recognition systems worldwide, has proven to perform well in different types of images and, therefore, will be the basis of our work and our comparison term.

The remaining of this paper has the following structure: section II overviews the iris recognition process, namely the less constrained acquisition setup and the Daugmans's approach; section III describes the proposed method; section IV describes the used datasets and discusses the obtained results; finally, section V states the conclusions.

II. Iris Recognition

The iris recognition process starts with the segmentation of the iris ring. Further, data is transformed into a double dimensionless polar coordinate system, through the *Daugman's Rubber Sheet* process. Regarding the feature extraction stage, existing approaches can be roughly divided into three variants: phase-based [1], zero-crossing [2] and texture-analysis methods [3]. Dauman [1] used multi-scale quadrature wavelets to extract texture phase-based information and obtain an iris signature with 2048 binary components. Boles and Boashash [2] computed the zero-crossing representation of a 1D wavelet at different resolutions of concentric circles. Wildes [3] proposed the characterization of the iris texture through a Laplacian pyramid with four different levels. Finally, in the feature comparison stage, a numeric dissimilarity value is produced, which determines the subjects identity. Here, it is usual to apply different distance metrics (Hamming [1], Euclidian [4] or weighted Euclidian [5]), or methods based on signal correlation [3].

The accuracy of the deployed iris recognition systems is remarkable, as reported by the study conduced by Daugman [6] and three other independent evaluations [7]–[9]. Nevertheless, recent publications emphasize the significance of some iriscode bits [10], aiming at improving by either masking less consistent bits [11] or condensing high discriminatory information regions [12]. However, we stress that the claimed effectiveness is conditioned to the acquisition of good quality images, captured in stop-and-stare interfaces at reduced imaging distances. In less constrained conditions, where a trade-off between data acquisition constrains and recognition accuracy is inevitable, the challenge is to maximally increase flexibility in three axes: subjects position and movement, imaging distances and lighting conditions. The main problem

is the appearance of other noise factors [13] (Subsection IV-A), that represent a substantial issue. As before stated, this area receives growing interests from the research community and constituted the scope of several publications [14]–[16].

A. Daugman's Approach

The Daugman's approach [17] to perform iris recognition is the most widely acknowledged, with great acceptance over the scientific community. Apart from being the unique implemented in commercially deployed systems, it usually acts as comparison term for alternative proposals. His method starts by the detection and segmentation of the iris. Later, the normalization of the segmented region is performed and features are extracted through the convolution of the normalized data with a bank of 2D Gabor Wavelets (1), followed by a quantization stage that produces a binary *iriscode*. This code is used in the matching stage, that applies the *Hamming* distance (3) as comparison measure.

$$h_{\{Re,Im\}} = sgn_{\{Re,Im\}} \int_{\rho}\int_{\phi} I(\rho,\phi) e^{iw(\theta_0-\phi)} . \\ .e^{-(r_0-\rho)^2/\alpha^2} e^{-(\theta_0-\phi)^2/\beta^2} \rho d\rho d\phi \quad (1)$$

$$c = (codeA \otimes codeB) \cap maskA \cap maskB \quad (2)$$

$$HD = \frac{\|c\|}{\|maskA \cap maskB\|} \quad (3)$$

where $\otimes$ is the logical XOR operation and $\cap$ the logical AND.

III. Proposed Methdod

The similarity measure used by Daugman at the matching stage simply gives the ratio of concordant iriscode bits over the whole iris and does not take into account their spatial and frequency distributions. In this paper, such analysis is performed, hoping that the location of the concordant bits and how they spread in the iriscode can provide useful information in the discrimination between match and non-match comparisons.

(a) Iriscodes match with regular distributed concordant bits

(b) Iriscodes match with an high concordance region (delimited by the light gray rectangle)

Fig. 1. Illustration of two iriscode matching results. Black pixels express concordant bits in the corresponding biometric signatures.

Figure 1 illustrates two comparisons between iriscodes, from now on called "*comparison maps*" (2), where black pixels denote concordant bits of the corresponding biometric signatures. Although their Hamming Distance is the same (50% of concordant bits) we claim that - intuitively - the *comparison map* from Figure 1(b) has a much higher probability of being an intra-class comparison. This is due to the (delimited) region that gives an area where both iriscodes have high concordance.

A. Spatial Domain Analysis

To ascertain the level of concordance in regions of different sizes, we performed a set of convolutions with Haar-based wavelets of different sizes. Let c be a *comparison map* of $M \times N$ dimensions. Let h be a Haar-based mother wavelet with size $s \times s$. The similarity r in local regions of c is given by:

$$r_s = h_s * c \qquad , s = \{2k\},\ k = 2,3,...,16 \quad (4)$$

where $*$ denotes the bi-dimensional convolution and r_s has the same dimensions of c.

Let $\omega_s = \max\{r_s(i,j)\},\ i = 1,2,...,N;\ j = 1,2,...,M$.

Let H be the 25-bin histogram of $r_{\frac{\psi}{2}}$ where ψ is the maximum size of the Haar-wavelet, such that $H = \{h_1, h_2, ..., h_{25}\}$. All the ω_i and h_i values where used as fetures for the further stages (sub-section IV-B) and give the proportion of concordant bits across regions of different sizes of the comparison map.

B. Frequency Domain Analysis

Together with the above described analysis we also analyzed the frequencies spread of the concordant bits. The rationale is that matching between inter-class iriscodes should give a distribution close to white-noise of concordant bits. Oppositely, an intra-class comparison should present a higher amount of low frequency components, according to the same key insight given in the previous section.

At this stage, two modifications to the Daugman's approach were performed: remotion of the signal-wise binary conversion step; and replacement of the XOR operation by the difference between coefficients. When applying a Fourier transform to both the binary and the differential *comparison maps*, we found that the later produces more discriminating results, which is easily justified by its higher amount of information.

Let c be a *comparison map* of $M \times N$ dimensions. The 2D Fourier transform F is given by:

$$F(u,v) = \frac{1}{N} \sum_{x=0}^{M} \sum_{y=0}^{N} c(x,y) e^{-j2\pi(ux/M + vy/N)} \quad (5)$$

where j is the square root of -1 and e denotes the natural exponent.

Results were decomposed into sixteen sub-regions, regularly distributed in small windows, and a set of attributes was ascertained: minimum and maximum values, average, standard deviation and local entropy. Since the central shape of F (where the lower frequencies lie) might contain important information which could not be properly processed by this windowing, another method was used to extract specific features from this area.

Let A be a $P \times N$ window, centered in the $P \times M$ matrix that contains the noticeable central shape such that $P = 2M/8$. Ten features F_i are then extracted as explained in the initial

part of section (6b), representing the distribution of evenly spaced 10-bins histogram:

$$T_i = \min(A) + i\frac{\Delta A}{10} \tag{6a}$$

$$F_i = \sum_{m=1}^{P}\sum_{n=1}^{N} \operatorname{sgn}\left(A_{(m,n)} - T_i\right) \tag{6b}$$

with $\Delta A = \max(A) - \min(A)$ and $i = 1, 2, ..., 10$.

IV. Experiments

Thus, the required parameters of the Gabor wavelets (1) were tuned for best performance, being chosen those with maximal decidability index (7); *i.e.* maximizing the distance between the distributions obtained for the two classical types of biometric comparisons: between signatures extracted from the same (*intra-class*) and different eyes (*inter-class*).

$$d' = \frac{|\mu_{inter} - \mu_{intra}|}{\sqrt{\frac{\sigma^2_{inter} + \sigma^2_{intra}}{2}}} \tag{7}$$

where μ_{inter} and μ_{intra} denote the means of the inter- and intra-class comparisons and σ_{inter} and σ_{intra} the respective standard deviations.

Regarding iris segmentation, all images were manually segmented, avoiding that segmentation errors corrupt the obtained results. A central and contiguous region was extracted from the normalized image, free of eyelid and eyelash occlusions, and used for the extraction of the iriscode. This option was taken to empower the spatial and frequency domain analysis.

A. Datasets

Two different datasets were used in our experiments: UBIRIS [18] and UBIRIS.v2 [19]. These are noisy datasets [13] with the following factors that degrade the quality of the data:

- **Out-of-Focus** - caused by subject movement allied to imaging systems limitations (namely in the depth-of-field, poor lightning/exposure ratio);
- **Off-Angle** - subject head and eye rotation or lack of alignment;
- **Rotation** - tilt of the head, despite of the subject being or not facing the camera;
- **Motion blur** - blurred iris images caused by eyelid movement;
- **Obstructions** - various types of blocking objects can be found, being the most commons: eyelids, eyelashes, glasses and contact lenses;
- **Reflections** - generally strong reflections caused by light sources or weak ones introduced by surroundings;
- **Partial Iris** - images appear where iris is not completely visible;
- **Out-of-iris** - images where iris is not present at all, either because fully occlusions or the eye not being present in the frame.

The higher range of acquisition distances enables the capturing of data at different scales and should make the results more visible.

Four dataset configurations were employed:

- UBIRIS.v2 – the first one is made of 500 images from UBIRIS.v2 without any kind of particular selection;
- UBIRIS.v2 Frontal – a second arrangement consists of 175 images, also from UBIRIS.v2, captured with the subject looking at camera's direction;
- UBIRIS.v2 Frontal Close – the third setup is composed by 100 images from the same database, with the subject also looking at the camera, but at relatively closer distances (4 to 6 meters);
- UBIRIS – the last setup include 500 images from UBIRIS.

The number of comparisons c_n is given from the number of irises n in the database through (8), from which about 4.5% are intra-class comparisons. In every case, we selected a group of images that we believe to represent each dataset.

$$c_n = \sum_{i=1}^{n-1} i \tag{8}$$

B. Feature Selection and Classification

Upon trials using different feature selection and dimensionality reduction techniques, carried out on frontally captured irises, we decided to apply Logistic Regression (LR) to the best 125 features, sorted using Peng *et al*. method [20] - *minimum-Redundancy, Maximum Relevance*, and then mapped to 15 dimensions through *Local Fisher Discriminant Analysis* [21]. By conducting our earliest tests in an UBIRIS.v2 sub-set of frontal images, we avoided problems associated with others noise factors (as gaze look), which might require specific corrections.

C. Results and Discussion

When applied to the different datasets, the comparison of our method and of the Daugman's gave the results contained in tables I and II.

TABLE I
LOGISTIC REGRESSION RESULTS FOR DIFFERENT DATASET CONFIGURATIONS. "HD" REPRESENTS DAGUMAN'S APPROACH PERFORMANCE, AND "125 FEAT" REFERS TO OUR APPROACH. AUC STANDS FOR AREA UNDER ROC CURVE AND CA FOR CLASSIFICATION ACCURACY

	HD		125 Feat	
	AUC	CA	AUC	CA
UBIRIS.v2	0.7315	0.9574	0.7598	0.9589
UBIRIS.v2 Frontal	0.8499	0.9582	0.8562	0.9590
UBIRIS.v2 Frontal Close	0.8740	0.9632	0.8897	0.9643
UBIRIS	0.9865	0.9868	0.9932	0.9897

Starting by frontal UBIRIS.v2 images, the subset our method was initially projected on, and attending to *Area Under ROC Curve* (AUC) assessment, we can observe an apparently residual increment of 1%. However, the AUC differs from Daugman's approach from 1.57% on close-captured images

to 2.83% on images without restrictions of any kind, which is a more significative improvement. For the first version of UBIRIS, where Daugman's approach has a good performance considering acquisition conditions, our method once again presents enhancements of almost 1% (Figure 2).

For the *Classification Accuracy* (CA), more permeable to class unbalancing, the most notable boost occurs for UBIRIS, as for the second version of this database advancements are proportional to the ones of AUC.

TABLE II
CONFUSION MATRICES REPRESENTING PROPORTIONS OF TRUE FOR DIFFERENT DATASET CONFIGURATIONS. "HD" REPRESENTS DAUGMAN'S APPROACH PERFORMANCE, AND "125 FEAT" REFERS TO OUR APPROACH.

	Real	Predicted: HD 0	HD 1	Predicted: 125 Feat 0	125 Feat 1
UBIRIS.v2	0	1.000	0.000	1.000	0.000
	1	0.991	0.009	0.950	0.050
UBIRIS.v2 Frontal	0	0.997	0.003	0.997	0.003
	1	0.858	0.142	0.841	0.159
UBIRIS.v2 Frontal Close	0	0.997	0.003	0.996	0.004
	1	0.797	0.203	0.744	0.256
UBIRIS	0	0.998	0.002	0.998	0.002
	1	0.262	0.738	0.194	0.806

Table II allows us to interpret the results in a more perspicuous way. Having in mind the priority given to lower the *False Accept Rate* (FAR) as much as possible, is at the *False Reject Rate* (FRR) that the improvements due to our method can be better observed. Without jeopardizing the FAR, FRR has a drop of about 1.7% for the subset where our method was schemed, when compared to the information provided by HD alone. Analyzing the other two UBIRIS.v2 datasets, this drop became more suggestive reaching 4.1% to 5.3%, and 6.8% on UBIRIS.

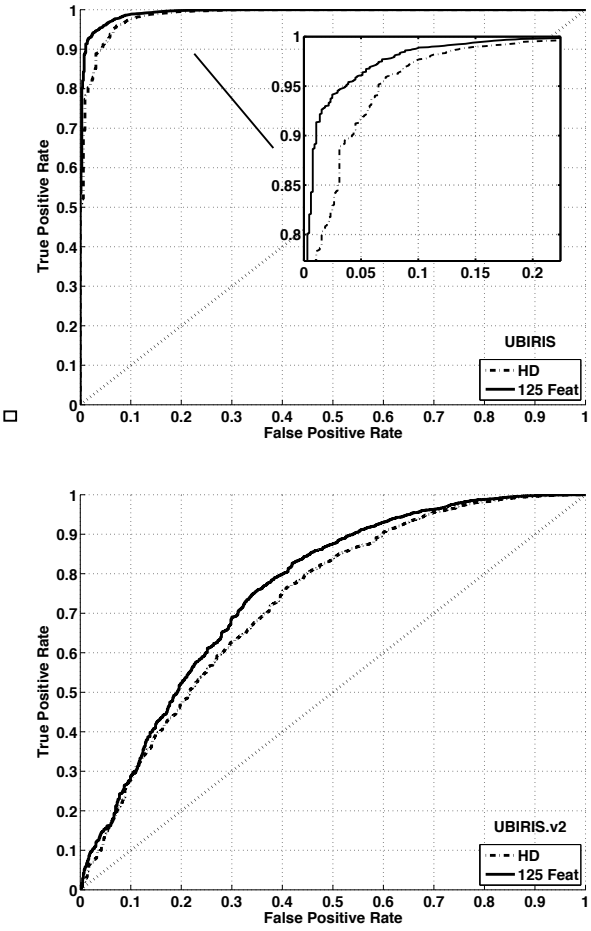


Fig. 2. ROC curves for UBIRIS and UBIRIS.v2 respectively. Inside curves represent Daugman's method (HD) result and outer ones refer to the proposed method (125 Feat).

V. CONCLUSIONS

Unconstrained image acquisition setups and protocols lead to the appearance of degraded data that significantly increases the challenges in performing accurate iris biometric recognition. In this paper we assessed the spatial and frequency distributions of the agreement bits resultant of the comparison between iriscodes having as main purpose the increase of the robustness to data acquired in less controlled conditions. Based in well-known feature extraction and data mining techniques, our method is to be used together with the traditional Daugman's approach and consistently contributed for an improvement in all experimented datasets.

ACKNOWLEDGMENT

We acknowledge the financial support given by "FCT-Fundação para a Ciência e Tecnologia" and "FEDER" in the scope of the PTDC/EIA-EIA/103945/2008 research project "NECOVID: Negative Covert Biometric Identification".

REFERENCES

[1] J. G. Daugman, "Phenotypic versus genotypic approaches to face recognition," in *Face Recognition: From Theory to Applications*. Heidelberg: Springer-Verlag, 1998, pp. 108–123.
[2] W. Boles and B. Boashash, "A human identification technique using images of the iris and wavelet transform," *Signal Processing, IEEE Transactions on*, vol. 46, no. 4, pp. 1185–1188, April 1998.
[3] R. P. Wildes, "Iris recognition: an emerging biometric technology," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1348–1363, September 1997.
[4] Y. Huang, S. Luo, and E. Chen, "An efficient iris recognition system," in *Proceedings of the First International Conference on Machine Learning and Cybernetics*, China, November 2002, pp. 450–454.
[5] L. Ma, T. Tan, Y. Wang, and D. Zhang, "Efficient iris recognition by characterizing key local variations," *Image Processing, IEEE Transactions on*, vol. 13, no. 6, pp. 739–750, June 2004.
[6] J. G. Daugman, "New methods in iris recognition," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 37, no. 5, pp. 1167–1175, October 2007.
[7] J. L. Cambier, "Iridian large database performance," Iridian Technologies, Tech. Rep., 2007, http://iridiantech.com.

[8] "Independent test of iris recognition technology," International Biometric Group, Tech. Rep., 2005, http://www.biometricgroup.com.
[9] T. Mansfiel, G. Kelly, D. Chandler, and J.Kane, "Biometric product testing and final report, issue 1.0," 2001.
[10] K. P. Hollingsworth, K. W. Bowyer, and P. J. Flynn, "The best bits in an iris code," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 964–973, 2009.
[11] ——, "Using fragile bit coincidence to improve iris recognition," in *BTAS'09: Proceedings of the 3rd IEEE international conference on Biometrics: Theory, applications and systems*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 165–170.
[12] J. E. Gentile, N. Ratha, and J. Connell, "Slic: short-length iris codes," in *BTAS'09: Proceedings of the 3rd IEEE international conference on Biometrics: Theory, applications and systems*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 171–175.
[13] H. Proença and L. A. Alexandre, "The NICE.I: Noisy iris challenge evaluation - part i," in *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, September 2007, pp. 1–4.
[14] J. R. Matley, D. Ackerman, J. Bergen, and M. Tinker, "Iris recognition in less constrained environments," *Springer Advances in Biometrics: Sensors, Algorithms and Systems*, pp. 107–131, October 2007.
[15] C. Fancourt, L. Bogoni, K. Hanna, Y. Guo, R. Wildes, N. Takahashi, and U. Jain, "Iris recognition at a distance," in *Proceedings of the 2005 IAPR Conference on Audio and Video Based Biometric Person Authentication*, U.S.A., July 2005, pp. 1–13.
[16] R. Narayanswamy, G. Johnson, P. Silveira, and H. Wach, "Extending the imaging volume for biometric iris recognition," *Applied Optics, vol. 44, no. 5*, pp. 701–712, February 2005.
[17] J. G. Daugman, "How iris recognition works," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, no. 1, pp. 21–30, January 2004.
[18] H. Proença and L. A. Alexandre, "UBIRIS: A noisy iris image database," in *ICIAP*, September 2005, pp. 970–977.
[19] H. Proença, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre, "The UBIRIS.v2: A database of visible wavelength images captured on-the-move and at-a-distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 99, no. 1, 5555.
[20] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, August 2005.
[21] M. Sugiyama, "Local fisher discriminant analysis for supervised dimensionality reduction," in *ICML '06: Proceedings of the 23rd international conference on Machine learning*. New York, NY, USA: ACM Press, 2006, pp. 905–912.

Appendix C

Facial Expressions: Discriminability of Facial Regions and Relationship to Biometrics Recognition

C.1 Overview

This appendix consists of the following article:

Facial Expressions: Discriminability of Facial Regions and Relationship to Biometrics Recognition
Elisa Barroso, Gil Santos and Hugo Proença
IEEE Symposium on Computational Intelligence in Biometrics and Identity Management – CIBIM '13, 16-19 April 2013.
ISBN: 978-1-4673-5879-8/13

Facial Expressions: Discriminability of Facial Regions and Relationship to Biometrics Recognition

Elisa Barroso*, Gil Santos† and Hugo Proença‡
University of Beira Interior – Department of Computer Science
IT - Instituto de Telecomunicações†‡
6200 Covilhã – Portugal
ebarroso@di.ubi.pt*, gmelfe@ubi.pt†, hugomcp@di.ubi.pt‡

***Abstract*—Facial expressions result from movements of muscular action units, in response to internal emotion states or perceptions, and it has been shown that they decrease the performance of face-based biometric recognition techniques. This paper focuses in the recognition of facial expressions and has the following purposes: 1) confirm the suitability of using dense image descriptors widely known in biometrics research (*e.g.*, local binary patterns and histogram of oriented gradients) to recognize facial expressions; 2) compare the effectiveness attained when using different regions of the face to recognize expressions; 3) compare the effectiveness attained when the identity of subjects is known/unknown, before attempting to recognize their facial expressions.**

***Index Terms*—Facial Expressions, Biometric Recognition, Performance Analysis.**

I. Introduction

The recognition of facial expressions has been motivating growing research efforts in recent years and benefited from advances in machine learning, image processing, and human cognition domains. Facial expressions constitute responses to internal emotion states, intentions, or social environment. They may be intentional or without conscious control and are produced by the synergistic or co-operative action of various facial muscles, as illustrated and described in Figure 1. Another interesting property is their universality: Paul Ekman studied the nature of facial expression and concluded that all humans are able to identify enjoyment, surprise, sadness, anger, fear, disgust. Also, when a set of volunteers was asked to make facial expressions to depict various scenarios, they were unmistakable [1].

The recognition of facial expressions is used to study facial behavior and several observational coding systems for that purpose were previously proposed, such as the Facial Affect Scoring (FAST) [1], the Facial Action Coding System (FACS) [2], the Emotional Facial Action Coding System (EMFACS) [3] and Facial Expression Coding System (FACES) [2]. Most of these are based in six discrete emotions: happiness/joy, sadness, anger, fear, surprise and disgust. Also, methods due to Matias *et al.* [4], Matsumoto *et al.* [5] and Coan and Gottman [6] are used in infants to detect and track their facial affect behavior.

The recognition of facial expressions mainly evolves two types of techniques: dense appearance descriptors and statistical machine learning. In particular, local binary patterns

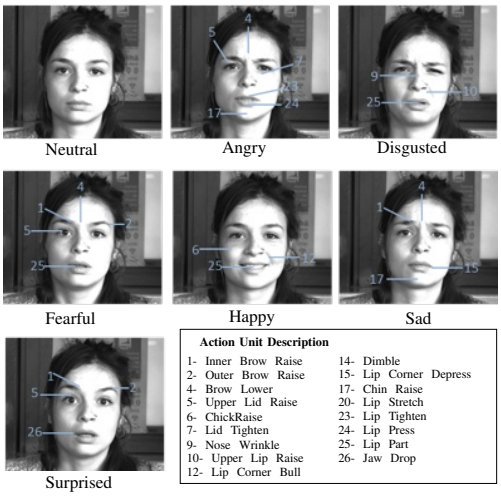


Figure 1. Targeted action units for the emotional expressions considered in this work, as suggested by Root and Stephens [7]

(LBP) [8], histogram of oriented gradients (HOG) [9] and scale invariant feature transform (SIFT) [10] were successfully applied to this problem. HOG [9] describes local object appearances and shapes by distribution of local intensity gradients or edge directions. LBPs [8] describe the pixels of an image by thresholding the neighborhood of each pixel with the value of the centre point and using these binary numbers to construct a label. SIFT [10] is a widely used local descriptor that starts by localizing keypoints with the local scale-space maxima of difference-of-Gaussian (DoG), and subsequently uses such keypoints as reference to generate a 3D histograms of gradient locations and orientations. Also, various classifiers were used, such as neural networks (NN) [11], support vector machines (SVM) [12], linear discriminant analysis (LDA) [13], K-nearest neighbors (KNN), multinomial logistic ridge regression (MLR) and Hidden Markov models (HMM) [14].

According to the above, this paper mainly focuses in the recognition of facial expressions, and the suitability of using different facial regions for that task. Our work plan was divided into three main phases: 1) we started by confirming the

suitability of fusing dense global and local image descriptors in the recognition of facial expressions; 2) we analyzed the effectiveness attained when using the *mouth*, the *periocular region*, the *whole face* and the *mouth* plus *periocular region* regions fused at feature level was compared; and 3) we assessed the improvements in performance that are due to knowing subjects identity before recognizing their facial expressions. To accomplish this plan, we start by manually defining the regions-of-interest, and then proceed for feature encoding according to the three feature extraction techniques. Then, for dimensionality reduction purposes, the principal components analysis (PCA) [13], [15] of data was carried out. Finally, feed-forward NN [11], [12] were used for classification purposes.

The remainder of this paper is organized as follows: a detailed description of the used dataset is given in Section II; Section III reports our experiments and discusses the results; and finally, Section IV presents the conclusions.

II. FaceExpressUBI Dataset

The *FaceExpressUBI* dataset was used as main data source for experiments. It contains $90,160$ color images acquired using a video camera, from 184 subjects (490 per subject), with resolution of 2056×2452 pixels. Each image is associated to a text/annotation file containing the coordinates for the face, periocular region, nose and mouth, respectively. Similarly to the majority of similar data sets, seven facial expressions were considered: happiness/joy, sadness, anger, fear, surprise and disgust plus the neutral expression. The dataset contains material from two imaging sessions, volunteers were 10 to 48 years of age, 35% female, 93% Caucasian European, 3% Latin-American, 3% African and 1% Asian. The number of participants wearing eyeglasses were 21 (12%). Furthermore, each expression was recorded during 5 seconds with a rate frame of 7 fps. For any given subject on the dataset, acquisition sessions were separated by at least two weeks. Also, from the first to the second session, location and orientation of the acquisition device and the artificial light sources were changed in order to increase the heterogeneity.

III. Experiments and Discussion

In our experiments $2,652$ images of the FaceExpressUBI dataset were used. They include seven facial expressions and were selected according to the evidence of the facial expressions they correspond to. According to the annotation files, the regions-of-interest that comprise the *mouth*, *periocular region* and *face* were cropped and normalized for a constant size using bi-cubic interpolation techniques. Then, due to the intrinsic properties of two of the feature encoding techniques used (LBPs and HOG), data was sub-divided into square patches, as detailed in Table I.

The cohesive perspective of our experiments is given in Figure 2. We used the LBP, HOG and SIFT descriptors to extract features from each region-of-interest (ROI). Then, PCA was used for dimensionality reduction and a feed forward neural network used for classification purposes. In this case, the problem was regarded as a binary classification task: for each pair of images regarding the same facial expression, a *positive* response from the neural network should be given; whereas pairs of images that regard different facial expressions should output a *negative* response.

Table I
Description of the pre-processing/size changes in the input images.

Anatomic Regions	Resize	Number of Blocks	Block Size
Mouth	45×54	5×6	9×9
Periocular	36×45	4×5	9×9
Face	54×54	6×6	9×9

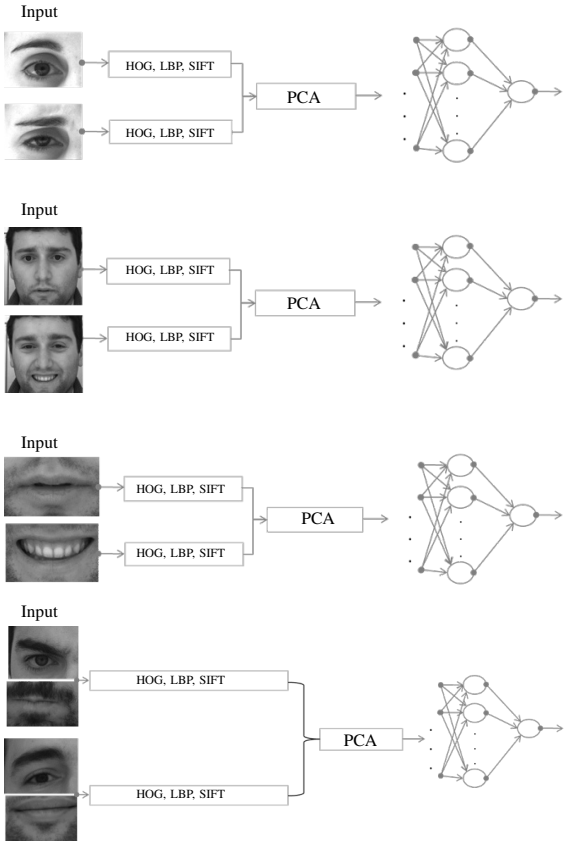
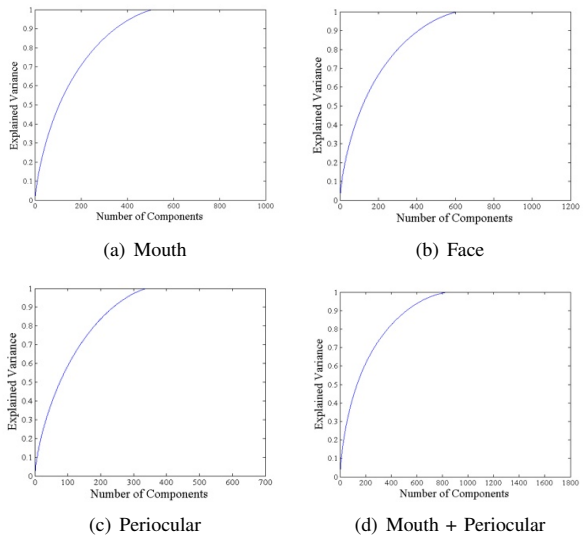


Figure 2. Cohesive perspective of our experiments, in order to assess the discriminating ability of each region of the face to recognize facial expressions: a) using exclusively the periocular region; b) using the whole face; c) using mouth; and d) using mouth and periocular region fused at the feature level.

The used feature encoding strategies projected each ROI into feature spaces of dimension: 961 for the mouth, 641 for the periocular region, $1,153$ for the face, and $1,602$ for mouth + periocular region. Then, as above stated, PCA was used for dimensionality reduction purposes, enabling projections to hyper-spaces of dimension 500, 330, 600, and 800 components respectively for the mouth, periocular, face, and mouth + periocular regions. The number of components used per region corresponds to the set that explained at least 98% of the

information in the initial set (Figure 3).



(a) Mouth (b) Face

(c) Periocular (d) Mouth + Periocular

Figure 3. Number of principal components selected for each region analyzed.

Finally, in the classification phase, data was divided into three disjoint subsets: training (70%), validation (15%) and test (15%). A set of feed-forward NN with three layers was created, using Levenberg-Marquardt back-propagation, and varying the number of neurons in the hidden layer (between 50% and 200% of the dimension of the feature space). As stoping criterium for the learning process, a maximum number of 50 validation checks was used. Due to the non-deterministic property of neural networks, the best observed configuration was repeated twenty times for every ROI and the median error rate taken.

In order to perceive the variance in performance when the identity of subjects is known / unknown, experiments were repeated in two different setups: 1) at first, only comparisons between facial expressions of the same subject were considered, corresponding to the setup where the identity of the subjects is known before attempting to recognize their facial expressions; 2) then, the identity constraint was relaxed, and comparison between facial expressions of different subjects were also considered.

A. Setup 1: Knowing Subjects' Identity

By selecting the data frames where facial expressions are most evident in our dataset, $33,306$ pair wise image comparisons were considered, from which $5,124$ were randomly sampled. This step was due to the computational burden of neural networks to learn in such high dimensional hyper-spaces and to the observation that results tend to maintain relatively stable when more than a few thousand of comparisons were used in the learning processes. Classes were balanced, meaning that the number of pair wise comparisons that regard the same facial expressions is equal to the number of comparisons that regard different facial expressions. Also, in order to perceive the discriminability of each facial region *per expression*, this experiment was repeated when considering all facial expressions together and each one separately, yielding seven variants of the experiment. At first, we attempted to recognize any facial expressions and then, exclusively attempted to recognize one at a time (among happy, anger, sad, surprise, disgust and fear). Results are given in Table II and the most interesting conclusions highlighted in bold font: the mouth outperformed in the recognition of the Happy expression, which is not too surprising due to the action units evolved in that expression. The whole face obtained the best error rates only twice: when considering all facial expressions, and for the Anger expression. Interestingly, the fusion at the feature level of mouth + periocular attained the best results most times (three). In opposition, a surprising observation was the low levels of performance attained by mouth + periocular when attempting to recognize all facial expressions, which was explained due to the sparsity of instances in the feature space of higher dimension when compared to the remaining ROIs. It should be stressed that in this experiment, only comparisons regarding facial expressions from the same subjects were selected, corresponding to the scenario where a biometric recognition system performs before the facial expression.

Table II
MEDIAN RECOGNITION RATES OBSERVED, WHEN ATTEMPTING TO RECOGNIZE ALL FACIAL EXPRESSIONS (ALL COLUMN) AND EACH ONE SEPARATELY. IN THIS CASE, THE IDENTITY OF THE SUBJECTS IS ASSUMED TO BE KNOWN BEFORE ATTEMPTING TO RECOGNIZE THEIR FACIAL EXPRESSIONS.

Region	All	Happy	Sad	Surprise	Fear	Anger	Disgust
Mouth	86.5	**95.5**	93.5	**95.1**	94.2	92.7	**94.4**
Periocular	90.0	94.8	93.9	91.8	88.7	91.1	89.8
Face	**90.1**	**94.9**	92.9	94.8	93.7	**94.7**	94.3
Mouth + Periocular	69.6	94.8	**94.0**	95.0	**94.3**	94.6	94.2

B. Setup 2: Unknowing Subjects Identity

This section regards an empirical setup similar to the described above, with the exception that this time the identity of subjects was not known, meaning that pair wise image comparisons between different subjects were also considered. In this case, starting from an initial number of $6,561,282$ pair wise comparisons, $5,124$ were randomly selected in order to obtain confidence intervals similar to the previous experiment. Table III gives the results, where the best recognition rate was obtained for the face region and happy expression (95%). Overall, a slight decrease in the effectiveness (around $3 \approx 4\%$) was observed when attempting to recognize facial expressions separately. The most notorious decreases in performance occurred when all facial expressions were considered, in some circumstances up to 50% of the performance observed for the knowing identity setup. This leaded us to conclude that biometric recognition techniques contribute for consistent improvements in the analysis of subjects facial expressions.

In summary, based on the observed error rates, we concluded that positive expressions (happy and surprise) are easier

to recognize than negative expressions (sad, anger, fear and disgust). Also, for most cases, the fusion at the feature level of both the mouth and periocular region did not contributed for consistent improvements in performance. Even though, using exclusively sub-parts of the face (as the mouth) led to performance levels similar to the attained when the whole face is considered.

Table III
MEDIAN RECOGNITION RATES OBSERVED, WHEN RECOGNIZING ALL FACIAL EXPRESSIONS (ALL COLUMN) AND EACH ONE SEPARATELY. IN THIS CASE, THE IDENTITY OF THE SUBJECTS IS NOT KNOWN WHEN ATTEMPTING TO RECOGNIZE THEIR FACIAL EXPRESSIONS.

Region	All	Happy	Sad	Surprise	Fear	Anger	Disgust
Mouth	61.4	94.3	89.9	92.1	88.3	88.0	89.1
Periocular	58.4	93.4	87.1	86.7	84.4	82.1	84.1
Face	63.1	**95.0**	91.5	**93.3**	89.7	90.1	89.3
Mouth + Periocular	**65.2**	92.3	**91.9**	91.0	**91.0**	**91.0**	**90.7**

IV. CONCLUSIONS

This paper mainly focused on two types of analysis: 1) we compared the discriminating ability of regions of the face to the attained by using the whole face; and 2) compared the results obtained when subjects identity is previously known, in opposition to unknown identities. A dataset of seven facial expressions was used and a set of ROIs cropped, comprising the whole face, the mouth and the periocular region. Then, LBPs, HOGs and SIFTs were used for feature encoding purposes and PCA for dimensionality reduction. Finally, for each pairs of images, a feed-forward NN binary discriminated between those that regard the same facial expression or not. This experimental setup was repeated in two different variants: at first, we assumed that the identity of subjects is previously known and only facial expressions that regard the same subject were considered; then, this constraint was relaxed and facial expressions from different subjects were also taken into account.

Accordingly, our main conclusions are: 1) LBP, HOG and SIFT are effective methods for feature encoding purposes in this specific scenario; 2) fusing the mouth and periocular regions at the feature level does not lead to performance improvements when compared to when each region is used separately; 3) the use of the whole face in the recognition of most facial expressions does not provide better results than using exclusively regions of the face, such as the mouth and periocular region (exceptions are the *happy* and *surprise*); and 4) by knowing subjects' identity, consistent improvements in recognizing their facial expressions are attained, giving support to the use of biometric recognition methods before attempting to recognize facial expressions.

ACKNOWLEDGMENT

The financial support given by "FCT - Fundação para a Ciência e Tecnologia"and "FEDER"in the scope of the PTDC/EIA/103945/2008 research project "NECOVID: Negative Covert Biometric Recognition"is acknowledge. Also, the support given by IT - Instituto de Telecomunicações in the scope of the "NOYSIRIS"research project is acknowledge too.

REFERENCES

[1] P. Ekman, "Facial Expressions," in *Handbook of Cohnition and Emotion*, T. Dalgleish and M. Power, Eds. San Francisco, California, USA: John Wiley & Sons, 1999, ch. 16, pp. 301–320.

[2] A. M. Kring and D. M. Sloan, "The Facial Expression Coding System (FACES): development, validation, and utility." *Psychological assessment*, vol. 19, no. 2, pp. 210–24, Jun. 2007.

[3] G. Koutlas, Anastasios (University of Ioannina and G. Fotiadis, Dimitrios (University of Ionnina, "Image Processing and Machine Learning Techniques for Facial Expression Recognition," in *Handbook of Reasearch on Advanced Techniques in Diagnostic Imaging and Biomedical Applications*. Book News, 2009, pp. 1–16.

[4] R. Matias, J. F. Conn, and S. Ross, "A Comparison of Two Systems That Code Infant Affective Expression," *Developmental Psychology - the American Psychological Association*, vol. 25, no. 4, pp. 483–489, 1989.

[5] D. Matsumoto, P. Ekman, and A. Fridlund, "Analyzing Nonverbal Behavior," in *Practical Guide to Using Video in the Behavioral Sciences*. New York: John Wiley & Sons, 1991, ch. 10, pp. 153–165.

[6] J. A. Coan and J. M. Gottman, "The Specific Affect Coding System (SPAFF)," in *Handbook of Emotion Elicitation and Assessment*, J. A. Coan and J. J. B. Allen, Eds. New York: Oxford University Press, 2007, ch. 16, pp. 106–123.

[7] A. A. Root and J. A. Stephens, "Organization of the central control of muscles of facial expression in man." *The Journal of physiology*, vol. 549, no. 1, pp. 289–98, May 2003.

[8] D. Huang and C. Shan, "Local Binary Patterns and Its Application to Facial Image Analysis: A Survey," *IEEE Transactions on Systems, Man and Cybernetics: Part C*, vol. 41, no. 6, pp. 765–781, 2011.

[9] M. Dahmane and J. Meunier, "Emotion recognition using dynamic grid-based HoG features," *Face and Gesture 2011*, pp. 884–888, Mar. 2011.

[10] H. Soyel and H. Demirel, "Facial expression recognition based on discriminative scale invariant feature transform," *Electronics Letters*, vol. 46, no. 5, p. 343, 2010.

[11] L. Ma and K. Khorasani, "Facial expression recognition using constructive feedforward neural networks," *IEEE Transactions on Systems, Man, and Cybernetics: Part B*, vol. 34, no. 3, pp. 1588–1595, 2004.

[12] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 172–187, 2007.

[13] G. Bai, W. Jia, and Y. Jin, "Facial Expression Recognition Based on Fusion Features of LBP and Gabor with LDA," *2009 2nd International Congress on Image and Signal Processing*, pp. 1–5, Oct. 2009.

[14] Y.-l. Tian, T. Kanade, and J. F. Cohn, "Facial expression analysis," *Handbook of Face Recognition*, vol. 3, no. 5, pp. 247–275, 2005.

[15] M. Uddin, J. Lee, and T.-S. Kim, "An enhanced independent component-based human facial expression recognition from video," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 4, pp. 2216–2224, Nov. 2009.