



UNIVERSIDADE DA BEIRA INTERIOR
Engenharia

Reconhecimento Biométrico da Orelha
Análise de Factores de Qualidade de Imagem

Rui Gabriel Raposo

Tese para obtenção do Grau de Mestre em
Engenharia Informática
(2º ciclo de estudos)

Orientador: Prof. Doutor Hugo Proença

Covilhã, Outubro de 2012

Dedicatória

Em memória do meu pai.

Agradecimentos

Agradeço, especialmente, ao meu orientador, Professor Doutor Hugo Proença, por ter acreditado em mim, pelo seu conhecimento e pelo apoio incondicional em todos os momentos deste trabalho,
Um muito Obrigado!

Agradeço aos meus colegas de laboratório por toda a ajuda, pelos incentivos e amizade ao longo dos anos.

A todos os voluntários que disponibilizaram um pouco do seu tempo para a concretização deste trabalho, o meu agradecimento!

À minha família que acreditou sempre em mim e porque sem eles não era ninguém,
Muito Obrigado!

À minha namorada pelo apoio, compreensão e dedicação, em especial, ao longo deste processo.

Aos meus amigos, no geral, um muito obrigado pela amizade, por me terem ajudado a ultrapassar todos os obstáculos e por estarem sempre dispostos a ajudar.

A todos, muito Obrigado!

Resumo

Este trabalho incide na necessidade de avaliar o impacto dos fatores de qualidade da captura de imagens, nos sistemas de reconhecimento biométrico.

Com a captura da base de dados UBEAR [RHPP] foi constatado que o impacto da qualidade das imagens recolhidas, poderiam alterar bastante a percepção dos sinais biométricos (no caso concreto a forma da orelha). Uma vez que, cada vez mais os sistemas de reconhecimento biométricos se tentam aproximar de ambientes não controlados, temos um maior leque de alterações na qualidade das imagens captadas, pelo que esta análise se torna fundamental.

Nesta análise partindo das imagens de entrada, foram estudados os fatores que influenciam a qualidade, de forma a obter para cada um desses fatores, uma quantificação da qualidade de imagem. Tendo esta quantificação, avançou-se para o sistema de reconhecimento em si, que possibilitou-nos obter o resultados de comparações dos novos indivíduos, com os existentes na base de dados (pertencentes ao conjunto de treino).

Com estes valores, tanto da quantificação da qualidade de imagem, como dos resultados do reconhecimento biométrico, foi possível avaliar o impacto de cada um dos fatores. Primeiramente, analisando os fatores individualmente e só posteriormente avaliando-os como um todo.

Com este estudo, poderá ser fornecida uma indicação dos fatores que devem ser cuidadosamente ajustados para que assim possamos melhorar globalmente o reconhecimento biométrico.

Palavras-chave

Biometria, Reconhecimento de padrões, Reconhecimento Biométrico, Fatores de qualidade de imagem, avaliação da qualidade de imagem.

Índice

1	Introdução	1
1.1	Motivação e Objetivos	1
1.2	Estrutura	1
2	Estado da Arte	3
2.1	Reconhecimento de Padrões	3
2.2	Reconhecimento Biométrico	5
2.2.1	Sinais	6
2.2.2	Modos de Operação	9
2.2.3	Medidas de Desempenho	10
2.2.4	Reconhecimento Biométrico da Orelha	11
2.3	Base de Dados UBEAR	14
2.4	Image Quality Assessment	15
2.4.1	Critérios dos Sistemas de Avaliação de Imagens	16
2.4.2	Pose da Cabeça	18
2.4.3	Desfoque Ótico	22
2.4.4	Desfoque de Movimento	24
2.4.5	Fusão de Fatores de Qualidade de Imagem	25
3	Métodos Propostos	29
3.1	Pré Processamento das Imagens	29
3.1.1	Equalização de Histograma	29
3.2	Reconhecimento Biométrico	29
3.2.1	SIFT	29
3.2.2	Wavelets Diádicas	31
3.3	Factores Implementados	31
3.3.1	Pose	31
3.3.2	Luminosidade	31
3.3.3	Desfoque Ótico	32
3.3.4	Desfoque de Movimento	32
3.3.5	Oclusão	33
4	Resultados e Discussão	35
4.1	Calculo dos Pesos de Impacto de Cada Fator	35
4.1.1	Função de Distribuição Cumulativa	35
4.1.2	Método de Ajustamento de Curva	36
5	Conclusões	43
5.1	Resumo do Trabalho Realizado	43
5.2	Trabalho(s) Futuro(s)	44
A	Anexos	51
A.1	Artigo publicado na conferência CIBIM 2011	51

Lista de Figuras

2.1	Exemplos de abordagens baseadas na geometria da orelha.	14
2.2	Localização da Triangular Fossa e da Antitragus [CBSV03].	14
2.3	Imagens da base de dados UBEAR e a sua máscara binária correspondente [RHPP].	15
2.4	Diagrama de um sistema de avaliação da qualidade de imagem de RR [WB06]. . .	17
2.5	Máscaras de LGP [LSTC10].	17
2.6	Nomenclatura dos tipos de rotação da cabeça [Fit01].	18
2.7	Diagrama de blocos do processo de estimativa do Desfoque de Movimento [KZSC10].	25
3.1	Exemplos de níveis de luminosidade muito diferenciada antes e depois da fase de pré processamento.	30
3.2	Histograma da densidade relativa ao fator Pose dos dados utilizados.	31
3.3	Histograma da densidade relativa ao fator Luminosidade dos dados utilizados. . . .	32
3.4	Histograma da densidade relativa ao fator Desfoque Ótico dos dados utilizados. . .	33
3.5	Histograma da densidade relativa ao fator desfoque de movimento dos dados utilizados.	33
3.6	Exemplo de um ouvido e o correspondente descritor angular.	34
3.7	Histograma da densidade relativa ao fator Oclusão dos dados utilizados.	34
4.1	Exemplo da função de probabilidade cumulada.	35
4.2	Tabela resumo dos coeficientes encontrados para cada fator individualmente. . . .	37
4.3	Tabela resumo de erro dos coeficientes encontrados para cada fator individualmente.	37
4.4	Fitting resultante do fator de Desfoque Ótico usando o método de wavelets diádicas.	37
4.5	Fitting resultante do fator de Desfoque Ótico usando o método de SIFT	38
4.6	Fitting resultante do fator de Luminosidade usando o método de wavelets diádicas.	38
4.7	Fitting resultante do fator de Luminosidade usando o método de SIFT	39
4.8	Fitting resultante do fator de Desfoque de Movimento usando o método de wavelets diádicas.	39
4.9	Fitting resultante do fator de Desfoque de Movimento usando o método de SIFT .	40
4.10	Fitting resultante do fator de Oclusão usando o método de wavelets diádicas. . . .	40
4.11	Fitting resultante do fator de Oclusão usando o método de SIFT	41
4.12	Fitting resultante do fator de Pose usando o método de wavelets diádicas.	41
4.13	Fitting resultante do fator de Pose usando o método de SIFT	42

Lista de Tabelas

2.1	Descrição de áreas de utilização das técnicas de reconhecimento de padrões [JDM00].	4
2.2	Matriz de confusão para o retorno do sistema de classificação [LS07].	10
2.3	Framework usado na construção da base de dados UBEAR.	15
2.4	Kernel de convolução de 8x8 de Daugman.	24
4.1	Resultados obtidos pelo cálculo da correlação de Pearson	39
4.2	Tabela resultante do calculo dos coeficientes com todos fatores.	42
4.3	Tabela dos valores residuais no ajuste realizado pela função lsqcurvefit.	42

Lista de Acrónimos

AAM	Active Apperance Model
ADN	Ácido Desoxirribonucleico
ASM	Acive Shape Model
AUC	Area Under Curve
BIOREC	Non-Cooperative Multimodal Biometric Recognition
CDF	Cumulative Distribution Function
CIBIM	Computational Intelligence in Biometrics and Identity Management
DET	Detection Error Tradeoff
DNA	Deoxyribonucleic Acid
DOF	degrees of freedom
DS	Dempster-Shafer
EGM	Elastic Graph Matching
FN	False Negative
FP	False Positive
FR	Full-Reference
FSLDA	Foley'Sammon linear discriminant analysis
GC	Global Context
GFD	Generic Fourier Descriptor
HVS	Human Vision System
ICA	Imperialist Competitive Algorithm
ICP	Iterated Closest Point
IQM	Image Quality Measure
KLDA	Kernelized Linear Discriminant Analysis
KNN	K Nearest Neighbor
KPCA	Kernel Principal Component Analysis
LABSSFE	long axis based shape and structural feature extraction
LEA	Locally Embedded Analysis
LGP	local gradient patterns
LLM	Locally Linear Map
MIS	Modified Image Spectrum
MLP	Multilayer Perceptron
MSE	Mean Square Error
MTF	modulation transfer function
NDMIS	normalized differences of the modified image spectrum
NR	No-Reference
PCA	Principal Component Analysis
PR	Precision-Recall
QA	Quality Assessment
ROC	Receiver Operating Characteristic
RR	Reduced-Reference
SIFT	Scale Invariant Feature Transform
SFS	Sequential Forward Selection
SOCIA Lab	Soft Computing and Image Analysis Laboratory
SVM	Support vector machine
SVR	Support Vector Regressors
TN	True Negative
TP	True Positive
UBI	Universidade da Beira Interior

Capítulo 1

Introdução

O conceito de biometria está muito presente no nosso dia a dia, inclusive utilizamos os recursos biométricos sem a percepção da tecnologia que envolve o próprio processo. O reconhecimento biométrico é o recurso disponível na tecnologia da informação e que pode identificar pessoas e a partir vários tipos sinais. A palavra biometria provem de duas outras "Bio"(vida) e "metria"(medida). O reconhecimento é feito através de algoritmos e de sensores que comparam os modelos, com os dados capturados de um dado sujeito (tais como, a iris, a palma da mão, a impressão digital, etc.). Atualmente o reconhecimento biométrico é utilizado por um vasto leque de aplicações, tais como empresas, polícia, instituições bancárias, entre outros. Este tipo de sistemas são altamente confiáveis, servindo assim como forma de evitar fraudes, violação de senhas, etc.

Uma parte importante da biometria passa não só pelos efeitos do reconhecimento mas também pela qualidade do input que é fornecido ao sistema de reconhecimento.

Para que este sistema os sistemas biométricos sejam confiáveis, um dos pontos que temos de analisar é o nível de qualidade dos dados capturados. A qualidade de imagem pode ser alterada por inúmeros fatores, alguns destes iremos analisar o seu impacto ao longo deste trabalho. Mais concretamente este trabalho irá incidir no impacto que a qualidade de imagem tem no reconhecimento biométrico da orelha.

1.1 Motivação e Objetivos

Sendo a qualidade da imagem um ponto preponderante para que o reconhecimento tenha sucesso, o principal objetivo deste trabalho prende-se com a avaliação do impacto dos fatores de qualidade numa imagem. Este trabalho partiu da necessidade que houve de sucessivos ajustes quando realizados os testes para capturada da base de dados disponibilizada à comunidade científica [RHPP]. Numa primeira fase serão analisados cada um dos fatores separadamente e, de seguida, analisados os fatores como um todo, podendo assim retirar as devidas lações.

1.2 Estrutura

Este trabalho encontra-se organizado da forma que passamos a descrever. No Capítulo 1 (Introdução) foi feita uma introdução ao tema, frisando vários aspectos relacionados. No Capítulo 2 (Estado da Arte) são apresentados alguns métodos existentes tanto para o reconhecimento biométrico, como para o métodos de avaliação da qualidade de imagem. No Capítulo 3 (Métodos Propostos) são apresentados métodos desenvolvidos para a análise do impacto da qualidade da imagem no reconhecimento, bem como as implementações utilizadas descritas no Estado da Arte. No Capítulo 5 (Resultados e Discussão) são apresentados os resultados obtidos, usando os vários métodos de reconhecimento e discussão sobre o seu significado. No último capítulo, apresentamos algumas conclusões/argumentos sobre a realização desta dissertação bem como descrevemos um pouco o que poderá ser acrescentado a este trabalho na tentativa da obtenção de ainda melhores resultados ou validação dos mesmos.

Capítulo 2

Estado da Arte

2.1 Reconhecimento de Padrões

O reconhecimento, a descrição, a classificação e o agrupar automático de padrões representam um conjunto de problemas importantes para uma grande variedade de áreas como a biologia, a psicologia, a medicina, o marketing, a visão computacional, a inteligência artificial e os sensores remotos [JDM00]. Um padrão é definido por Watanabe [Wat85] como uma entidade vagamente definida à qual pode ser atribuído um nome. Podemos classificar como padrão uma imagem de uma impressão digital, uma face humana, ou um sinal de um discurso. Dado um padrão, o seu reconhecimento / classificação pode consistir numa das seguintes fases: classificação supervisionada na qual um padrão é identificado como um membro de uma classe predefinida; classificação não supervisionada na qual o padrão é atribuído a uma classe até então desconhecida. O problema de reconhecimento é desta forma colocado como uma tarefa de classificação ou categorização, onde as classes são definidas em função do programador do sistema (na classificação supervisionada), ou são aprendidos com base na similaridade dos padrões (na classificação não supervisionada).

Um detetor de objetos opera classificando imagens baseando-se em valores com características simples [VJ01]. Estas características, ao contrário do que seria mais comum pensar, vão para além da análise direta dos pixels, visto que as características podem agir para codificar o conhecimento num conjunto finito de dados de treino. Assim, o desempenho de sistema baseado em recursos opera substancialmente mais rápido quando comparado com um sistema baseado na análise dos pixels.

Como podemos verificar na tabela 2.1, existem diversos problemas aos quais é possível o uso métodos de reconhecimento de padrões, que auxiliam a resolução dos mesmos. Dependendo do tipo de objetos a classificar (ou descrever/reconhecer), o sistema usa algum (ou alguns) dos seguintes métodos ou abordagens:

- Abordagem estatística - metodologia clássica, historicamente mais antiga, denominada por "Teoria da Decisão". Assume que as características das classes se regem por determinados modelos probabilísticos;
- Abordagem sintática - procura descrever a estrutura dos padrões usando inter-relações de características básicas denominadas primitivas;
- Abordagem neuronal - laboração tipo "caixa negra" que procura determinar um mapeamento ótimo, entre entradas e saídas inspirando-se em modelos de neurónios do cérebro;
- Abordagem difusa - tem em conta o grau de incerteza por vezes inerente a características e a classificações, que usa a Teoria dos Conjuntos Difusos para modelar esse grau de incerteza.

O modo de operação de um sistema de reconhecimento de padrões pode ser descrito em três fases:

- Extração de características dos objetos a classificar (ou a descrever);

Domínio do Problema	Área de Aplicação	Padrão de entrada	Classes do padrão
Bioinformática	Análise de sequência	ADN/sequência de proteínas	Tipos de genes / padrões conhecidos
Mineração de dados	Procura de padrões significativos	Pontos de um espaço multidimensional	Clusters compactos e bem separados
Classificação de documentos	Pesquisa na Internet	Documentos de texto	Categorias semânticas (ex. negócios, desporto, etc.)
Análise de imagens de documentos	Máquina de leitura para cegos	Imagem de documentos	Caracteres, palavras alfanuméricas
Automatização industrial	Inspeção da impressão de circuitos nas placas	Intensidade ou alcance da imagem	Natureza defeituosa / não defeituosa do produto
Extração em bases de dados Multimédia	Pesquisa na Internet	Vídeo clip	Gêneros de vídeo (ex. ação, dialogo, etc.)
Reconhecimento Biométrico	Identificação pessoal	Face, íris, impressão digital	Autorização de acessos controlados
Sensores remotos	Previsão de rendimento de culturas	Imagens multiespectral	Utilização de categorias de solos, padrões de crescimento das culturas
Reconhecimento de discurso	Inquéritos telefónicos sem operador	Formas de onda da fala	Palavras faladas

Tabela 2.1: Descrição de áreas de utilização das técnicas de reconhecimento de padrões [JDM00].

- Seleção das características mais discriminativas;
- Construção de um classificador (ou descritor).

A fase inicial (de extração de características dos objetos) atua de forma a recolher tudo o que se destaca no domínio da análise que em conjunto formam um dado padrão. Anteriormente a esta fase de extração das características há uma outra, que consiste em pré-processar os dados. O intuito desta fase prévia é o de evitar casos como dados incompletos, ruidosos, e ausência de valores. Para esta fase de pré-processamento faz-se uma limpeza dos dados, para preencher dados em falta, atenuar ou remover os valores discrepantes, bem como uma normalização e agregação de dados similares. Como principais métodos de seleção de características temos o PCA e o ICA.

A seleção das características para além de eleger as características mais relevantes, serve também para evitar a maldição da dimensionalidade. A referida maldição pode ocorrer quando o número de características é arbitrariamente grande ou a distribuição estatística das classes é completamente conhecida, bem como quando a probabilidade de erro de classificação de uma regra de decisão não aumenta com o aumento do número de características consideradas. Todos os classificadores amplamente utilizados, inclusive redes neuronais multicamadas com retroalimentação (feed-forward), sofrem da maldição da dimensionalidade. As abordagens mais conhecidas desta seleção são a clássica pesquisa exaustiva, a pesquisa em árvore e a SFS que seleciona o atributo simples que melhor caracteriza o objeto e, em seguida, adiciona uma característica de cada vez que, em combinação com as anteriores selecionadas, maximiza a função de critério.

O classificador parte das características selecionadas obtidas através do conjunto de treino, e dado um novo conjunto (de teste) irá classificá-lo com base nos valores das características. As quatro abordagens mais conhecidas para o reconhecimento de padrões são baseadas em modelos, classificações estatísticas, correspondências sintéticas e estruturais e redes neuronais [JDM00]. Para além de poder haver interpretações diferentes de como classificar um método de um dado tipo, também

existem tentativas de desenvolvimento de sistemas híbridos como é disso exemplo, o trabalho de Fu [Fu86].

2.2 Reconhecimento Biométrico

A utilização da biometria teve o seu início em meados do século XIX, pelo chefe da polícia de Paris, que teve a ideia de realizar várias medições corporais para identificar criminosos [JRP04]. Rapidamente essa ideia foi acolhida pelas forças policiais, primeiro com o registo de impressões digitais numa base de dados (na verdade eram guardados num arquivo de cartão), sendo que mais tarde o processo evoluiu de forma a que de uma cena criminal seja possível retirar impressões digitais para serem comparadas com as existentes nas bases de dados. Para além do seu uso exaustivo pelas forças policiais para identificar criminosos, acessos restritos para empregados em trabalhos críticos, testes de paternidade, ciência forense, identificação de condenados e prisioneiros, houve incremento até aos dias de hoje, onde o principal objetivo é a sua utilização em aplicações comuns. Podemos classificar uma medição biológica como sendo uma qualquer característica fisiológica do ser humano e / ou comportamental que pode ser usado como uma característica biométrica, desde que satisfaça os seguintes requisitos:

- Universalidade: cada pessoa deve ter uma determinada característica;
- Distinção: duas pessoas devem ser suficientemente diferentes em termos da dita característica,
- Permanência: a característica deve ser suficientemente invariável (em relação ao critério de correspondência) durante um período de tempo,
- Coletável: a característica pode ser medida quantitativamente.

No entanto, num sistema biométrico prático (isto é, um sistema que emprega biometria para reconhecimento pessoal), há um número de outras questões que devem ser consideradas, incluindo:

- O desempenho, que se refere ao reconhecimento obtido através da precisão e da velocidade; os recursos necessários para atingir a precisão do reconhecimento desejado e a velocidade; bem como o funcionamento e os fatores ambientais que afetam a precisão e a velocidade;
- Aceitação, o que indica o grau em que as pessoas estão dispostas a aceitar o uso de um determinado identificador biométrico nas suas vidas diárias;
- Evasão, que reflete a facilidade com que o sistema pode ser enganado por métodos fraudulentos.

Um sistema prático biométrico deverá atender ao reconhecimento especificado de requisitos de precisão, velocidade e recursos. Poderá vir a ser prejudicial para o sistema caso não seja aceite pela população alvo, bem como a ausência de robustez perante métodos fraudulentos e ataques ao sistema. Um sistema biométrico é essencialmente um sistema de reconhecimento de padrões que opera através da aquisição de dados biométricos de uma pessoa, extraindo um conjunto de recursos a partir dos dados adquiridos, e comparando este conjunto de recursos com o modelo definido no banco de dados. Dependendo do contexto da aplicação, um sistema biométrico pode operar tanto no modo de verificação com no modo de identificação 2.2.2.

2.2.1 Sinais

Os sistemas biométricos podem basear-se em diversos sinais da biometria. Sendo que cada um deles tem os seus pontos fortes e fracos, a escolha entre eles será dependente da aplicação. Como não existe nenhum dado/sinal biométrico ótimo, não é possível que um único sinal atenda efectivamente às necessidades de todas as aplicações. E, assim sendo, essa escolha é feita em função do modo de funcionamento da aplicação e das propriedades das características biométricas. Segue-se uma pequena descrição dos sinais biométricos que são normalmente utilizados para as diferentes aplicações.

2.2.1.1 DNA

É um código unidimensional único para indivíduos, exceção de gémeos verdadeiros que terão DNA idênticos. É, contudo, a mais usada pela ciência forense para a identificação de pessoas. Não obstante de deter certas limitações, sendo elas:

- Contaminação e sensibilidade - é com relativa facilidade que podemos furtar um pedaço de ADN de outra pessoa sem que esta se aperceba;
- Problemas no que diz respeito ao reconhecimento automático em tempo real - a tecnologia existente requer o uso de alguns produtos químicos, bem como competências de um perito e, por isso, não está orientada para o reconhecimento não invasivo em tempo real;
- Problemas de privacidade - a informação acerca da susceptibilidade da pessoa ter uma doença, pode ser obtida através do padrão de ADN e, essa é uma preocupação pois a informação do código genético pode ser usada com o intuito discriminatório, como por exemplo, em práticas de contratação.

2.2.1.2 Face

É um método não intrusivo, sendo o mais utilizado pelos humanos no reconhecimento. Este dado biométrico pode ser usado com base na localização e na forma dos atributos da face como, olhos, sobrancelhas, nariz, lábios e queixo. Assim como através da análise global da imagem que representa a face como uma combinação ponderada de um número de faces canónicas. Quando se utiliza o reconhecimento facial uma grande quantidade de problemas podem ocorrer que prejudicam o reconhecimento, tais como:

- A forma como se adquire a imagem da face;
- Por vezes é requerido um fundo ou iluminação especial;
- Dificuldade de reconhecimento em duas poses muito diferentes, e em condições de iluminação diferentes;
- Expressões faciais, maquilhagem, pelos faciais e estilos de cabelos.

2.2.1.3 Facial, Mão, e Termograma Infravermelho das Veias da Mão

O padrão de calor irradiado pelo corpo humano é uma das características de um indivíduo e pode ser captada por uma câmara infravermelha de uma forma discreta muito parecido com uma foto normal (no espectro visível). Deste modo, esta tecnologia poderia ser usada sem que o alvo de reconhecimento tivesse conhecimento. Um sistema baseado num termograma não requer contacto e é não-invasivo, mas a aquisição das imagens é um desafio em ambientes não controlados, onde

as superfícies que emanam calor (por exemplo, aquecedores e tubos de escape dos veículos) estão presentes na proximidade do corpo. A tecnologia relacionada com infravermelhos de curto alcance é usado para digitalizar a forma de um punho fechado para determinar a estrutura das veias da mão. Os sensores de infravermelhos a nível monetário, são extremamente dispendiosos, o que poderá provocar inibição na utilização deste tipo de reconhecimento de uma forma mais generalizada. Visto que, nesta forma de reconhecimento é necessária uma participação ativa e cooperativa do sujeito durante a identificação, a sua identificação tem maior probabilidade de chegar ao sucesso, visto ser necessário estar num ambiente controlado, o que facilita a identificação.

2.2.1.4 Impressão Digital

Têm sido bastante usadas ao longo dos anos com uma alta taxa de reconhecimento. A impressão digital é única e desenvolve-se ao longo dos primeiros meses de gestação fetal, mesmo em casos de gêmeos verdadeiros. Por isso, e também pelos preços mais baixos dos scanners, a identificação por impressão digital tornou-se um método mais comum de se usar em diversos dispositivos (como por exemplo, os portáteis pessoais). A quando do uso da impressão digital, os problemas de localização, artefatos de sombra e variações de luminosidade não se colocam, visto que, a recolha da impressão digital é feita em ambiente controlado.

2.2.1.5 Andar (Gait)

É uma forma peculiar de reconhecimento, pois é um dado biométrico complexo tal no espaço como no tempo. O andar pode não ser notoriamente diferente, no entanto é diferente o suficiente para que seja discriminante num sistema que necessite uma segurança pouco complexa (menos confiável e possivelmente com mais falhas). Contudo, se a captura do andar for feita através da recolha de um vídeo, para captar diversos movimentos de cada articulação, irá tornar o sistema computacionalmente complexo.

2.2.1.6 Geometria da Mão e do Dedo

Os sistemas baseados na geometria da mão incluem medidas como a forma, o tamanho da palma, o comprimento e largura dos dedos. Estes sistemas têm a vantagem de serem fáceis de usar, baratos e muito simples. Contudo, problemas como a falta de invariância ao longo do tempo (crescimento de uma criança até à idade adulta), o uso de joalharia (como anéis) e a possibilidade de doença (artrites), podem alterar a forma geométrica da mão. Esta forma de reconhecimento por ser cooperativa e em ambiente controlado é facilmente encaminhada para o sucesso.

2.2.1.7 Íris

A textura da íris é formada durante o desenvolvimento fetal e estabiliza durante os dois primeiros anos de vida. A sua textura complexa torna bastante distinta a informação útil para o reconhecimento pessoal. A precisão e a velocidade destes sistemas aponta para uma identificação em grande escala. Cada íris é única tal como a impressão digital, até para íris de gêmeos idênticos as íris são diferentes. A íris caracteriza-se por ser extremamente difícil de adulterar cirurgicamente e é bastante fácil detetar íris artificiais (como lentes de contacto). Embora seja necessário, nestes sistemas, uma participação do utilizador, e sejam dispendiosos, os novos sistemas mais recentes têm-se tornado mais amigáveis para o utilizador e melhores na relação custo-eficiência. Ao contrário dos scanners de retina que requerem uma aproximação do scanner para a sua identificação, os scanners de íris foram comprovados ser eficazes a distância superiores a 46cm. A sua desvantagem é o tamanho da

íris ser muito pequeno quando a distância aumenta, o que faz da recolha de imagens à distância problemática.

2.2.1.8 Keystroke

Existe a possibilidade de que cada pessoa digite num teclado de uma forma característica. Este comportamento biométrico não deverá ser único para cada indivíduo, mas oferece informações discriminatórias suficientes para permitir a verificação de identidade. A dinâmica de digitação é um comportamento biométrico, podendo-se esperar para observar grandes variações em padrões típicos de digitação. Além disso, esta verificação poderia ser facilmente feita, de uma forma discreta quando alguém digita informações via teclado.

2.2.1.9 Odor

Sabe-se que cada objecto emana um odor que é característico na sua composição química e isso poderia ser utilizados para distinguir os vários objectos. Uma lufada de ar em torno de um objecto é soprado através de uma variedade de sensores químicos, cada um sensível a um determinado grupo (aromático) de compostos. Um componente do odor emitido pelo corpo de um ser humano (ou qualquer outro animal) é distinto de um determinado indivíduo. Não está claro se a invariância no odor do corpo poderia ser detectados, apesar da existência de desodorizantes de cheiros, e da variância da composição química do meio ambiente.

2.2.1.10 Palmas da Mão

As palmas da mão humana contêm um padrão de irregularidades (rasgos e vales) que são semelhantes à impressão digital. Porém, a área da palma é muito maior, pelo que é esperado que sejam ainda mais distintas do que a impressão digital. Porém como os scanners da palma necessitam abranger uma área maior, são mais corpulentos e mais caros que os de impressão digital. Ainda assim, como as palmas têm características distintas como as linhas principais e as rugas isso permite que a captura seja feita numa resolução menor que torna esses sensores mais baratos. Caso o objectivo seja construir um sistema biométrico de maior eficácia, um scanner de alta resolução irá permite essa construção, visto que irá detetar todos os detalhes da palma humana como a sua geometria, as suas irregularidades (rasgos e vales), as linhas principais e as rugas que combinadas contribuem para a eficiência deste sistema.

2.2.1.11 Scanner de Retina

A vascularização da retina é rica na sua estrutura e é suposto ser uma característica de cada indivíduo e de cada olho. Alega-se ser o mais seguro sistema biométrico pois não é fácil mudar ou replicar a vascularização da retina. A aquisição das imagens requer que a pessoa olhe para um objeto ocular e se concentre num ponto específico no campo visual para que uma parte predeterminada da vasculatura da retina possa ser capturada. A aquisição das imagens envolve a cooperação do sujeito, o que implica o contacto com a ocular, e requer um esforço consciente por parte do utilizador. Todos esses fatores afetam adversamente a aceitação pública da retina como sinal biométrico. A vascularização da retina pode revelar algumas condições médicas, como hipertensão, que é outro fator de dissuasão para a aceitação pública da verificação da retina como sinal de biometria.

2.2.1.12 Assinatura

A forma como uma pessoa assina o seu nome é conhecido por ser uma característica daquele indivíduo. Embora nas assinaturas seja necessário o contacto com o instrumento de escrita e um esforço por parte do usuário, foram aceites no governo, nos tribunais e no comércio como um método de verificação. As assinaturas são um comportamento biométrico que muda ao longo de um período de tempo e são influenciados pelas condições físicas e emocionais dos signatários. As assinaturas de algumas pessoas variam substancialmente: até mesmo as impressões sucessivas de sua assinatura são significativamente diferentes. Além disso, os falsários profissionais podem ser capaz de reproduzir as assinaturas, e conseqüentemente irão enganar o sistema.

2.2.1.13 Voz

A voz é uma combinação de dados biométricos, fisiológicos e comportamentais. As características da voz de um indivíduo são baseados no tamanho e forma dos apêndices (por exemplo, trato vocal, boca, fossas nasais e lábios) que são usados na síntese do som. Estas características fisiológicas da fala humana são invariantes para um indivíduo, mas a parte comportamental do discurso de uma pessoa muda ao longo do tempo devido à idade, condições médicas (como uma constipação), estado emocional, entre outros. A voz também não é muito distinta, e pode não ser apropriada para a identificação em larga escala. Um sistema de reconhecimento de texto dependente de voz é baseado na emissão de uma frase fixa pré-determinada. Um sistema de reconhecimento de voz independente de texto reconhece a pessoa independente do que ela fala. Um sistema texto-independente é mais difícil conceber que um sistema de texto-dependente, mas oferece maior proteção contra a fraude. Uma desvantagem de reconhecimento de voz baseados é que as características do discurso são sensíveis a uma série de fatores como o ruído de fundo. O reconhecimento de voz é mais adequado em aplicações baseadas no sinal telefónico, mas o sinal de voz por telefone é afetado por uma degradação na qualidade, pelo micro e / ou pelo canal de comunicação.

2.2.2 Modos de Operação

No modo de verificação, o sistema valida uma identidade de uma pessoa através da comparação dos dados biométricos capturados com o modelo biométrico próprio armazenado na base de dados do sistema. Em tal sistema, um indivíduo que deseja ser reconhecido reivindica uma identidade, geralmente através de um número de identificação pessoal (PIN), um nome de utilizador ou um cartão inteligente e o sistema realiza uma comparação um-para-um para determinar se a afirmação é verdadeira ou não (por exemplo, "Será que estes dados biométricos pertencem ao Bob? "). A verificação de identidade é tipicamente utilizada para o reconhecimento positivo, onde o objetivo é prevenir que várias pessoas usem a mesma identidade [Way01].

Por outro lado, no modo de identificação o sistema reconhece um indivíduo pesquisando os modelos de todos os usuários da base de dados. Portanto, o sistema realiza uma comparação um-para-muitos para estabelecer a identidade de um indivíduo (ou não se o sujeito não estiver inscrito na base de dados do sistema) sem o sujeito ter que reivindicar uma identidade. A identificação é uma componente crítica em aplicações de reconhecimento negativo, quando o sistema verifica se a pessoa é quem ela (implicitamente ou explicitamente) nega ser. O objetivo do reconhecimento negativo é o de impedir que uma pessoa use identidades múltiplas [Way01]. A identificação também pode ser usada em reconhecimento positivo por conveniência (o utilizador não necessita reivindicar uma identidade). Embora os métodos tradicionais de reconhecimento pessoal, como senhas, PINs, chaves e tokens possam trabalhar para o reconhecimento positivo, o reconhecimento negativo só

pode ser estabelecido através de biometria.

2.2.3 Medidas de Desempenho

As medidas de desempenho têm o intuito de avaliar a qualidade das medições biométricas, sendo elas um indicador do bom ou do menos bom reconhecimento. Estas medidas servem para avaliar o grau de segurança em que podem ser utilizadas dependendo da área em que vai ser usada (ex: numa zona de acessos restritos é melhor errar numa possível autorização do que dar acesso a pessoas não autorizadas).

	Verdade	Falso
Classificação Verdadeira	TP	FN
Classificação Falsa	FP	TN

Tabela 2.2: Matriz de confusão para o retorno do sistema de classificação [LS07] .

Note-se que o cálculo das medidas True Positive, False Negative, False Positive e True Negative são calculados da seguinte forma:

$$TP = \frac{TP}{TP + FN} \quad (2.1)$$

$$FP = \frac{FP}{FP + TN} \quad (2.2)$$

$$FN = \frac{FN}{FN + TP} \quad (2.3)$$

$$TN = \frac{TN}{FP + TN} \quad (2.4)$$

2.2.3.1 Curva ROC

A curva ROC é usada para tomar decisões em diversas tarefas de decisão. Esta estabelece uma relação entre as deteções true positives (denominadas assim quando a classe é classificada verdadeira e de facto, é verdadeira) e as false positives (que ao contrário das anteriores são classificadas positivas, mas que na verdade são falsas) como varia o limiar de decisão. É uma medida bastante utilizada no reconhecimento da fala, e menos usada noutros tipos de reconhecimento.

2.2.3.2 Curva DET

Esta curva DET mostra a relação entre as taxas de erro em ambos os eixos, dando tratamento uniforme aos tipos de erro. O gráfico que representa a medida da curva DET pode ser visualizada usando eixos logarítmicos para uma melhor distinção do caso a analisar.

2.2.3.3 AUC

Esta medição AUC é usada para comparar duas curvas ou dois modelos, normalmente usada para ambas as curvas ROC e PR.

2.2.3.4 PR Curve

A curva de PR (Precision 2.5-Recall 2.6) mostra o que acontece quando o limiar de decisão é ajustado.

$$Precision = \frac{TP}{TP + FP} \quad (2.5)$$

$$recall = \frac{TP}{TP + FN} \quad (2.6)$$

2.2.3.5 F-Mesure

É uma medida que reflete ambos os valores da precision e recall e o seu deslocamento 2.7.

$$F - measure = \frac{2 \times precision \times recall}{precision + recall} \quad (2.7)$$

2.2.3.6 Decidibilidade

A decidibilidade de um sistema biométrico proposta por Daugman [DW96] é quantificada aproximadamente pela separação da diferença de resultados dados pelas comparações entre as intra e as inter classes, como denota a fórmula 2.8 onde μ_E e σ_E são a média e o desvio padrão das inter-classes e μ_I e σ_I são os valores similares correspondentes das intra-classes [Pro09].

$$d' = \frac{|\mu_E - \mu_I|}{\sqrt{\frac{1}{2}(\sigma_I^2 + \sigma_E^2)}} \quad (2.8)$$

2.2.3.7 Histogramas

Os histogramas podem representar as distribuições estatísticas de uma forma normalizada de uma característica. Os histogramas podem ser vistos como um conjunto de vectores que recolhem informação dos valores da característica a analisar que cada dado contém. Para dados que sejam muito densos ou contínuos é necessário definir um limiar para o número de bins que o histograma irá representar, o que torna o histograma mais simples para a análise e recolha da informação. Os histogramas representam um número barras de cada bin e o comprimento de cada barra representa a densidade desse bin.

2.2.4 Reconhecimento Biométrico da Orelha

Nos últimos anos, o reconhecimento biométrico usando a orelha foi alvo de grande estudo. Tal como a face, a íris tem diversas falhas que fazem com que o seu uso em aplicações para fins comerciais sejam limitadas. Ao contrários da face, a íris, mantém-se consistente ao longo do tempo e são mais fáceis de extrair, contudo a sua recolha numa resolução necessária, à distancia, é difícil. Por estas

razões uma nova abordagem baseada na orelha tem tido bastantes desenvolvimentos devidos à sua robustez, confiança e de fácil extração à distância.

2.2.4.1 Visão Global

As orelhas são difíceis de descrever, porque não temos vocabulário para tal. Existem poucos adjetivos que possam ser aplicados às orelhas, como largas ou achatadas. Mesmo que pareça haver poucos meios para que o reconhecimento possibilite distinguir uma orelha das restantes, a sua estrutura é rica e única, e por isso podemos usá-la como fator biométrico para identificação passiva.

Assim como não podemos afirmar com total certeza de que a nossa impressão digital é única, o mesmo se aplica à orelha. Em vez disso, é provável que o seja, dado que temos estudos que o provam, como os dois estudos de Iannarelli [Ian89], onde o primeiro compara cerca de 10,000 orelhas escolhidas de uma forma aleatória de uma amostra recolhida na Califórnia, e o segundo estudo mostra que ao examinar gémeos idênticos, onde características fisiológicas são conhecidas por serem semelhantes, a estrutura é consideravelmente diferente. Após a divulgação destes estudos, foi comprovado que a estrutura da orelha não era semelhante, especialmente na Concha e na área do lobe (nódulo).

É do conhecimento geral que a estrutura da orelha não muda radicalmente ao longo do tempo. Estudos médicos mostram que o crescimento da orelha nos primeiros quatro meses após o nascimento, é proporcional. Se pensarmos que o crescimento da orelha é proporcional, a gravidade pode provocar na orelha um estiramento na direção vertical. Este estiramento é mais notório no lobe da orelha, onde medições feitas, mostram que essa mudança não é linear. É possível usar a orelha como fonte adicional de prova e como sistema de reconhecimento, por exemplo, como reconhecimento adicional num caixa multibanco, na qual poderia ser reconhecida a pessoa através da face e da orelha. O reconhecimento pode ser usado no acesso a um local seguro, assim como detetar, à posteriori, um acesso não autorizado.

O que distingue a orelha dos outros sinais biométricos é:

- a sua maior consistência quando comparada com a face na medida da variação devido a expressões faciais,
- tem uma estrutura rica e estável que é preservada desde o nascimento até à idade adulta,
- a sua localização na parte lateral da cabeça torna a sua deteção mais fácil,
- um conjunto de dados é conveniente em comparação com as tecnologias mais invasivas como a íris, a retina e a impressão digital,
- o momento de captura não causa ansiedade, como nas medições da retina e da íris. Portanto são comparáveis com a íris, retina e impressão digital, mas são mais fáceis de capturar,
- o conjunto de dados não tem associado problemas higiénicos, como pode acontecer com o contacto direto na medição da impressão digital.

O que não invalida o facto de também ter alguns problemas que dificultem o reconhecimento como oclusões causadas pelos cabelos e brincos, problemas na fase de captura como o excesso / falta de luminosidade e a pose da cabeça. Podemos portanto, usar a orelha como um dado biométrico de uma abordagem viável e promissora para a identificação humana automática. É viável por isso, usar a orelha como dado biométrico para uma identificação passiva.

2.2.4.2 Métodos de Reconhecimento

Na fase de reconhecimento biométrico usando a orelha, existem diversos tipos de métodos que podem e têm sido usados ao longo dos últimos anos. Nesta secção será feita uma breve descrição dos métodos que existem na literatura.

- Baseada em medidas geométrica da orelha - sendo as 12 medidas de Iannarelli [RAC05, Ian89] ou a construção de grafos de vizinhança [ANRR06] ou de diagramas Voronoi das arestas detetadas da orelha [BB99, BB02]. As características detetadas por estas abordagens são robustas e garantem um certo nível de confiança quando extraídas à distância. No entanto o principal problema destas abordagens é a deteção de segmentos de curvas erógenas, que podem ocorrer devido à mudança de iluminação, sobras e/ou oclusões. Esta extração das 12 medidas de Iannarelli pode ser observada na figura 2.1 (a), na figura 2.1 (b) pode-se observar a extração utilizando os diagramas de Voronoi e do detetor de arestas de Canny na figura 2.1 (c).
- PCA para o reconhecimento da face e da orelha [VBS02, CBSV03]. Para as metodologias que usa o método PCA é necessário indicar manualmente as coordenadas da Triangular Fossa e da Antitragus como se pode observar na imagem 2.2. Um conjunto de imagens de treino é processado para produzir uma matriz especial de projeção que permite o cálculo de um conjunto de pesos que, quando usados para dimensionar um conjunto de «imagens-vectores» especiais já se encontram codificadas na matriz de projeção, o que produz uma representação da imagem original. Estes pesos formam vectores de características convenientes que representam cada imagem e são perfeitos para comparação biométrica, pois permitem o cálculo de uma distância entre as imagens, fazendo uma subtração de vectores simples.
- Force field transformation (transformação do campo de força) para determinar as linhas de energia, poços e canais da orelha [HNC05]. Esta metodologia é diferente das anteriores pois o campo de força é gerado da imagem inteira sem qualquer pré-processamento, suavizando toda a imagem com um kernel gigante $1/r$, que é 4 vezes maior que o tamanho original da imagem retirando os pequenos detalhes, mas mantendo toda a informação contida na imagem. Para esta aplicação são esperadas que condições como luminosidade e o «ponto de vista» sejam controlados e que o sujeito esteja numa posição lateral para a câmara.
- Classificadores neurais e macro características extraídas por redes de compressão [MSV99]. Neste caso é treinada uma rede de compressão na imagem original da orelha, para extrair suas propriedades estatísticas mais importantes da imagem ou de macro características. Estas propriedades estatísticas ficam guardadas num vector, que é uma representação intermediária codificada da imagem original, chamado vector de compressão. Constitui-se a entrada de um único percetor que executa a tarefa de identificação própria (cada uma de suas saídas corresponde a cada um dos indivíduos a serem identificados).
- Descritor local de superfície para representação do orelha em 3D [CB07, CB05, YB05]. Nas abordagens que usam representação 3D da orelha, partindo do Local Surface Path para estimar a rotação e a translação inicial do par base de dados / prova e só após este passo é aplicado o algoritmo ICP para compensar a distorção [CB07].
- Análise Acústica - Akkermans, Kevenaer e Schobben et al. propuseram a utilização da transformada de transferência acústica para registar a assinatura acústica da cada orelha e daí estabelecer a correlação entre orelhas para aplicação no reconhecimento biométrico [AKS05].

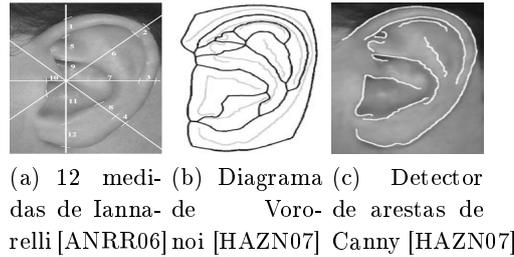


Figura 2.1: Exemplos de abordagens baseadas na geometria da orelha.



Figura 2.2: Localização da Triangular Fossa e da Antitragus [CBSV03].

2.3 Base de Dados UBEAR

Esta secção descreve a base de dados construída no SOCIA Lab num trabalho realizado no âmbito do projeto BIOREC. Esta base de dados tem como principal objetivo a construção de um conjunto de dados biométricos onde a orelha é capturada em condições diferentes e em que as condições do cenário não são controladas. O que significa que os voluntários foram capturados em movimento e em diversos tipos de rotação da cabeça. Também situações em que o excesso ou a falta de luminosidade afetam a imagem capturada foi notória, para o mesmo sujeito em diferentes sessões. Nesta base de dados, os voluntários cooperaram em duas sessões em dias diferentes com o intuito de que as condições fossem notoriamente diferentes. Desta captura resulta um vídeo onde o voluntário está de perfil para a câmara e executa movimentos de rotação da cabeça para cima, para baixo, para o lado da câmara e para o lado oposto, e finalmente dá alguns passos que tentam similar o movimento de um sujeito que passe pela câmara. O vídeo resultante foi separado por frame e selecionados manualmente 17 frames representativos dos movimentos descritos. Na tabela 2.3 estão indicados dados como a câmara e a lente utilizada, os parâmetros utilizados e algumas estatísticas relativas à idade, sexo, número de imagens capturadas e o numero orelhas capturadas dos voluntários. Um exemplo do conteúdo desta base de dados com as respetivas máscaras binárias estão na figura 2.3.

Deste trabalho de construção da base de dados UBEAR surgiram inúmeros problemas no reconhecimento relacionados com a qualidade das imagens capturadas. Isto impulsionou o trabalho desta tese, visto que o trabalho de análise da qualidade de uma imagem foi feita manualmente. Daqui

Framework de Aquisição de Vídeo	
Câmara	Stingray F-504B 2/3"
Focal length	35mm
Representação de cor	escala de cinzentos
Resolução de Vídeo	1280x960 pixels
Frames por segundo	15
Codec de vídeo	Avi sem compressão
Detalhes dos frames selecionados	
Resolução de Imagem	1280x960 pixels
Representação de cor	escala de cinzentos
Codec de Imagem	tiff
Voluntários	
Totais = Sujeitos 126; Orelhas 252; Imagens 4430	Sexo = Masculino: 44.62% Feminino: 55.38%
Idade = [0, 20] 43,85%	
[21, 25] 47.69%	
[26, 30] 3.85%	
[31, 35] 2.31%	
[36, 99] 2.30%	

Tabela 2.3: Framework usado na construção da base de dados UBEAR.

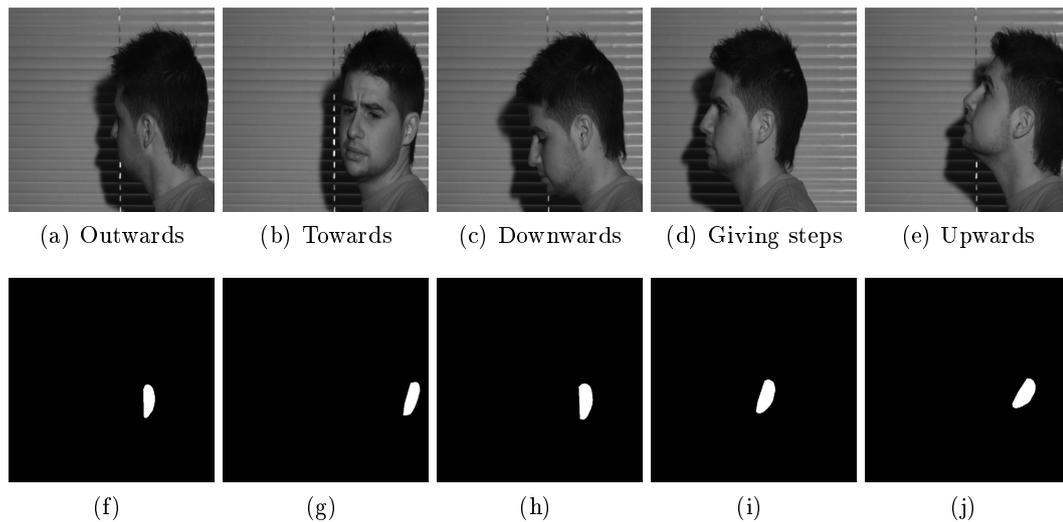


Figura 2.3: Imagens da base de dados UBEAR e a sua máscara binária correspondente [RHPP].

retira-se uma avaliação que não é exata e dependente de diversos fatores que podem influenciar esta avaliação.

2.4 Image Quality Assessment

Os avanços da tecnologia que usam imagem digital, velocidade computacional, capacidade de armazenamento e rede, resultaram na proliferação de imagens e vídeos digitais. Como as imagens digitais são capturados, os dados armazenados, transmitidos, e exibidos em diferentes dispositivos, existe necessidade de manter a qualidade da imagem. Os utilizadores finais dessas imagens, de um número incrivelmente grande de aplicações, são observadores humanos.

Nesta secção, vamos estudar objetivamente os critérios para a avaliação da qualidade de uma imagem percebida por um observador humano médio. Mesmo que se use o termo qualidade de imagem,

estamos interessados principalmente na fidelidade de imagem, isto é, o quão fidedigna uma imagem foi adquirida em comparação com a imagem original ou de referência. A avaliação da qualidade de imagem desempenha um papel fundamental na concepção e avaliação de imagem e sistemas de processamento de imagem. Como exemplo, algoritmos de controle de qualidade podem ser usados para avaliar sistematicamente o desempenho de algoritmos de compressão de imagens diferentes que tentam minimizar o número de bits necessários para armazenar uma imagem, mantendo a qualidade da imagem suficientemente elevada. Da mesma forma, os algoritmos de controle de qualidade podem ser usados para avaliar a aquisição de imagens e sistemas de visualização. As redes de comunicação desenvolveram-se muito nos últimos dez anos, imagens e vídeo são frequentemente transportados sobre fibra ótica, redes de comutação de pacotes como a Internet, sistemas sem fio, eficiência, etc. A largura de banda de aplicações tais como vídeo-conferência e vídeo sob demanda pode ser melhorada utilizando sistemas QA para avaliar os efeitos de erros no canal, nas imagens e vídeos transportados. Além disso, os algoritmos de controle de qualidade podem ser usados no design de vários componentes de um sistema de comunicação de imagem. Finalmente, a avaliação de imagem e a psicofísica da visão humana são áreas relacionadas. A investigação sobre QA de imagem e vídeo podem fornecer profundos conhecimentos sobre o funcionamento de HVS, o que seria de grande valor científico.

2.4.1 Critérios dos Sistemas de Avaliação de Imagens

Existem diferentes critérios de classificação da qualidade de imagem de acordo com tipos de conhecimento disponíveis. Os critérios existentes são: de referência completa (FR), sem referência (NR) e referência reduzida (RR) [WB06].

O primeiro critério para classificar as medidas objetivas de qualidade da imagem é a disponibilidade de uma "imagem original", que é considerado de qualidade e sem distorção ou perfeito, e pode ser utilizado como referência na avaliação de uma imagem distorcida. A maioria das propostas de medidas objetivas de qualidade na literatura assumem que a imagem de referência não falseada existe e esteja totalmente disponível. Apesar de a "qualidade de imagem" ser frequentemente usada por razões históricas, o termo mais preciso para este tipo de métrica seria similaridade de imagem, medida de fidelidade, ou a referência completa (FR) da avaliação da qualidade da imagem.

Em muitas aplicações práticas, uma imagem do sistema de avaliação de qualidade não tem acesso às imagens de referência. Portanto, é desejável desenvolver abordagens de medição que possam avaliar a qualidade da imagem cegamente. A avaliação da qualidade da imagem cega ou de não referência (NR) acaba por ser uma tarefa muito difícil, apesar de observadores humanos geralmente poderem eficaz e fiávelmente avaliar a qualidade de imagens distorcidas, sem usar qualquer referência. Isso porque é provável que o cérebro humano possua um grande conhecimento sobre como as imagens devem ou não ficar.

No terceiro tipo de método de avaliação da qualidade da imagem RR, a imagem de referência não está totalmente disponível. Em vez disso, certas características são extraídas da imagem de referência e utilizadas pelo sistema de avaliação da qualidade como informações para ajudar a avaliar a qualidade da imagem distorcida. Isto é, referido como referência reduzida, a avaliação da qualidade (RR) da imagem. A ideia de RR avaliação da qualidade foi inicialmente proposta como um meio para controlar o grau de degradação da qualidade visual dos dados de vídeo transmitidos através de redes de comunicações complexas, e o seu âmbito de aplicação foi expandido mais tarde. O quadro para a implantação de sistemas RR de avaliação de qualidade da imagem é mostrado na figura 2.4. Ele inclui um processo de extração do recurso no lado do remetente e um de extração

de características / processo de comparação no lado do receptor. As características são extraídas descrevendo a imagem de referência que são transmitidas para o receptor como informação lateral através de um canal auxiliar. O método de extração de recurso no lado do receptor, pode ser ajustado de acordo com as características extraídas no lado do emissor (mostrado como a seta tracejada). Um parâmetro importante de um sistema de RR é a largura de banda disponível para a transmissão de informação. O sistema RR deve selecionar os recursos mais eficazes e eficientes para otimizar a precisão da previsão de qualidade de imagem sob a restrição da largura de banda disponível.

Contudo nesta dissertação focar-se-á no critério de avaliação da qualidade de imagem sem referência ou cega.

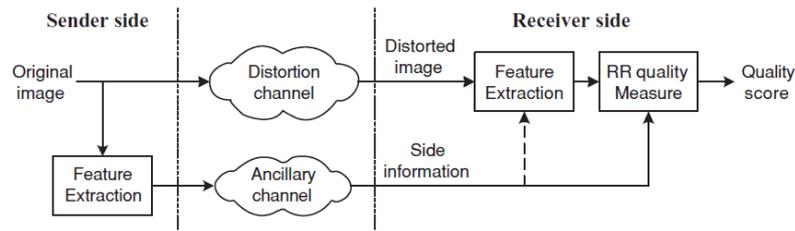


Figura 2.4: Diagrama de um sistema de avaliação da qualidade de imagem de RR [WB06].

Uma métrica comum para os fatores de desfoque ótico e de desfoque de movimento foi proposta por Lee et al. [LSTC10], onde após a localização da região a avaliar, é aplicado o LGP por oito máscaras de padrões, como podemos ver na figura 2.5. Estas máscaras são usadas na fórmula 2.10 onde n e m são respetivamente, a altura e largura da zona localizada para a avaliação da qualidade, $h(i,j)$ é o valor do pixel da zona localizada e S_p é um dos LGP. Ao aplicar as oito máscaras LGP para calcular oito valores de característica o objetivo é o de representar os vários padrões de arestas existentes na zona localizada. A classificação é feita recorrendo à fórmula 2.11 que junta os oito valores das características.

-1	-1	-1	-1	0	1	0	1	1	1	1	0
0	0	0	-1	0	1	-1	0	1	1	0	-1
1	1	1	-1	0	1	-1	-1	0	0	-1	-1
-1	-1	0	-1	0	0	0	0	1	0	1	1
0	0	0	-1	0	1	-1	0	1	0	0	0
0	1	1	0	0	1	-1	0	0	-1	-1	0

Figura 2.5: Máscaras de LGP [LSTC10].

$$L_p(i, j) = \sum_{a=-1}^1 \sum_{b=-1}^1 [h(i+a, j+b)S_p(a, b)]^2, p = 1, 2, \dots, 8 \quad (2.9)$$

$$L_p = \sum_{j=1}^m \sum_{i=1}^n L_p(i, j) \quad (2.10)$$

$$q_1 = L_1 + L_2 + \dots + L_8 \quad (2.11)$$

2.4.2 Pose da Cabeça

Um dos fatores que influencia a qualidade de imagem na fase de reconhecimento é a pose da cabeça, que ao contrário da detecção do rosto ou do reconhecimento de expressões, é fácil imaginar a recuperação do movimento da cabeça, sem saber muito acerca da sua estrutura. Sendo o movimento do corpo rígido, a detecção da cabeça é igual a qualquer outro padrão, varia com rotações e translações, visto serem de origem biológica. No entanto, há uma série de vantagens que advêm, quando consideramos a estrutura da cabeça:

- A cabeça está coberta de características não-rígidas, estimuladas (o rosto) ou apenas passivamente não-rígidas (cabelos longos). É importante considerar estas características, apenas para evitá-las;
- Para recuperar o movimento rígido a partir de informações visuais, algum conhecimento da forma da cabeça terá que ser recuperado. Para fazer isso, ajuda a existência de um modelo de parâmetros para usar como comparação;
- Deve haver uma maneira de relacionar a postura da cabeça para as características da superfície real que sejam importantes para a interface, - por exemplo, os olhos e a boca. Para isso é indispensável uma inicialização manual, ou algum substituto automático, que dependerá necessariamente de um conhecimento especial da estrutura da cabeça.

Uma vez que, o objetivo é criar um sistema NR os passos acima descritos não são possíveis de efetuar, dado que não teremos uma imagem inalterada para comparar. Para um melhor entendimento, os tipos de rotação da cabeça foram renomeados como podemos ver na figura 2.6.

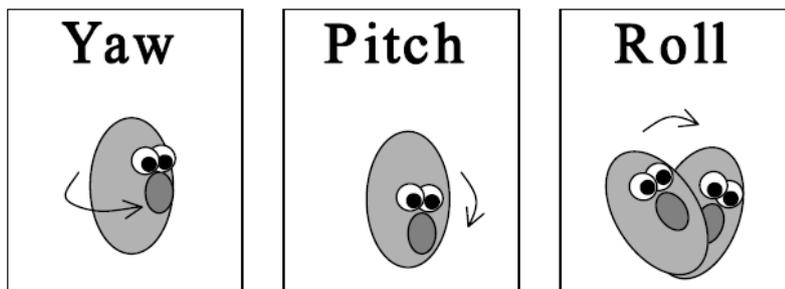


Figura 2.6: Nomenclatura dos tipos de rotação da cabeça [Fit01].

Dos vários métodos de estimativa de pose da cabeça existentes, podemos agrupá-los pela sua forma de operar, vantagens e desvantagens [MCT09]. Segue-se uma breve descrição dos métodos que têm em comum o mesmo modo de operação.

2.4.2.1 Métodos de Modelos de Aparência

Esta metodologia fundamenta-se numa comparação entre uma nova imagem da cabeça e um conjunto de exemplos, de diferentes poses, por forma a encontrar a vista mais semelhante. Este processo é bastante vantajoso para casos em que queiramos expandir o modelo, para adaptá-lo a condições de mudança de uma forma mais simples. Outro dos pontos favoráveis é o facto de não ser necessário exemplos negativos de treino ou pontos faciais característicos, apenas precisam de

imagens cortadas onde esteja apenas a face e ter anotações da pose da cabeça. Por outro lado, se não for usado um método de interpolação, este tipo de modelos apenas é capaz de localizar poses discretas, os erros na localização da cabeça podem influenciar negativamente a eficiência do sistema e quanto maior for a expansão do modelo mais demorado computacionalmente irá ser, dado que implica um maior número de comparações. No entanto para este último inconveniente pode-se aplicar um filtro Gaussiano ou uma wavelet de Gabor.

2.4.2.2 Métodos Detetores de Vector

Os métodos de detecção de vector são treinados uma série de detetores de cabeças cada um deles apontado para uma pose específica e atribui uma pose discreta ao detetor com melhor sustento. O que caracteriza esta abordagem é o facto de treinar vários detetores de face cada um deles para uma pose, para vector de classificadores binários, que detetam corretamente a face através da pose da cabeça, os vectores são semelhantes aos modelos de aparência, no entanto os vectores usam caminhos de imagens em vez de comparar uma imagem com um grande conjunto de modelos individuais. Portanto a imagem é avaliada pelo detetor que foi treinado em muitas imagens com um algoritmo de aprendizagem supervisionada. Como pontos favoráveis temos o facto de que na fase de detecção e localização da cabeça não seja necessário o uso de algoritmos de treino pois são bons para ignorar variações de aparência que não correspondem a variações de pose e os vectores de detetores são mais adequados para imagens de grande e pequena resolução. Contudo é uma metodologia que é penosa de treinar com muitos detetores para cada pose discreta. Quando dois detetores estão a localizar poses muito similares, as imagens que são exemplos positivos de treino para um, deve ser um exemplo negativo para o outro, os requisitos computacionais aumentam linearmente com o número de detetores, o que o torna difícil de implementar num sistema de tempo real com um vector grande. Mas para resolver este último problema é possível usar um classificador para escolher um detetor subsequente para usar na estimativa da pose.

2.4.2.3 Métodos de Regressão Não-lineares

O uso de métodos de regressão não-linear usa ferramentas de regressão não-linear para desenvolver um mapeamento funcional da imagem ou da característica de medição da pose da cabeça. Partindo de um conjunto de dados de treino anotados, podemos construir um modelo que fornece estimativas da pose discreta ou contínua para qualquer nova amostra de dados. O problema desta abordagem é o facto de não ser explícita a forma de como a ferramenta de regressão é capaz de aprender adequadamente o mapeamento sendo usada principalmente com redes neuronais, seja com MLP ou com LLM. Sintetizando, estes métodos são rápidos, apenas requerem as faces recortadas e anotadas para treino, funcionam bem com imagens próximas e afastadas e dá bons resultados práticos na estimativa da pose. Porém são propensos a erros da localização da cabeça, pelo que para solucionar esta limitação foi proposto o uso de uma rede de convolução que aumenta a MLP modelando algumas invariâncias de mudança, escalas e distorções para serem usadas por forma a reduzir a fonte de erro.

2.4.2.4 Métodos de Incorporações Múltiplas

Os processos que usam as incorporações múltiplas procuram modelar a variação contínua da pose da cabeça. Imagens novas podem ser incorporadas nessas múltiplas e então ser usadas para incorporar ou regredir a correspondência. Podem usar um PCA para a estimar pose, projetando a imagem no subespaço da PCA e comparar os resultados com um conjunto de modelo. As técnicas de incorporação múltipla podem ser lineares ou não-lineares. Sendo que as técnicas lineares têm

a vantagem de a incorporação poder ser concebida por multiplicações de matrizes, falta-lhes a representatividade de técnicas não-lineares.

2.4.2.5 Métodos Flexíveis

Este tipo de métodos ajustam um modelo não rígido à estrutura da face de cada indivíduo no plano da imagem. A pose é estimada com base nas comparações feitas com as características ou com a instanciação de parâmetros do modelo. Para esta abordagem um EGM é criado um grafo para cada pose discreta e cada uma delas é comparada com a nova vista da cabeça. A desvantagem é que a estimativa da pose é discreta, o que requer muitos grafos para refinar a estimativa e a comparação dos grafos é computacionalmente pesada quando confrontando com outras técnicas. Um destes métodos flexíveis é o AAM que aprende os modos primários das variações da forma facial e da textura de uma perspectiva 2D. Para construir um AAM é necessário primeiro gerar um ASM do conjunto de treino. As imagens de faces são deformadas até que os pontos das características correspondam à forma média. Estas imagens deformadas devem ser normalizadas e de seguida usadas para construir um modelo de texturas sem formas. No final, a correlação entre a forma e a textura são aprendidas e usadas por modelo de combinação de textura e formas. Um ASM tem a capacidade de representar os modelos primários da variação da forma olhando apenas com os componentes principais.

2.4.2.6 Métodos Geométricos

Partindo da localização de características como os olhos, a boca e a ponta do nariz estes métodos determinam a pose. A configuração das características faciais pode ser explorada de muitas formas para estimar a pose. Usando 5 pontos do rosto (cantos exteriores dos olhos, nariz e boca), a simetria da face pode ser encontrada ligando a linha entre os pontos médios dos olhos da boca e do nariz. Outra forma de localização é usar 5 pontos diferentes (os cantos interiores e exteriores dos olhos e a ponta do nariz). Contudo assumindo que os quatro pontos dos olhos são coplanares, o ângulo de rotação (yaw) pode ser determinado pela diferença de tamanho do olho direito e esquerdo devido à distorção projetiva dos parâmetros conhecidos da câmara. O ângulo de rotação (roll) pode ser encontrado simplesmente partindo do ângulo dessa linha com uma linha horizontal. O ângulo de rotação (pitch) é determinado comparando a distância entre a ponta do nariz e a linha do olhar para um modelo antropométrico. Estes modelos são chamados degenerativos visto que requerem uma precisão muito grande para estimar corretamente a pose da cabeça com estes modelos. Uma abordagem mais recente proposta baseia-se nos cantos interno e externo dos olhos e nos cantos da boca, onde qualquer desvio do paralelo na imagem resulta numa distorção projetiva. O ponto de fuga (onde estas linhas se intersectam) pode ser calculado com recurso aos mínimos quadrados, para minimizar a solução sobre-determinado por três linhas. Este ponto pode ser usado para determinar a orientação 3D das linhas paralelas, se a razão do seu comprimento for conhecido e possa usado para calcular a posição absoluta 3D de cada ponto de característica se o comprimento da linha for conhecido. Este tipo de métodos são rápidos e simples, com poucas características faciais conseguimos obter uma estimativa da pose decente. A dificuldade destes métodos está na correta deteção das características com precisão. Imagens mais distantes podem ser problemáticas para localização precisa das características. E também são mais sensíveis a oclusões do que os modelos baseados na aparência que usam a informação de toda a região da face.

2.4.2.7 Métodos de Localização

Os métodos de localização recuperam a mudança de pose global da cabeça analisando os movimentos observados entre frames de vídeo. Têm a seu favor o facto de possuírem um alto nível de precisão, contudo necessitam de uma inicialização da posição da cabeça. Operam de forma bottom-up seguindo marcas faciais de baixo nível de frame para frame. Não há uma fronteira nítida entre as abordagens bottom-up e top-down. A separação dos algoritmos em bottom-up ou top-down a divisão é uma conveniência conceptual, e os algoritmos podem (e provavelmente devem) conter elementos de ambas as categorias. As categorias não devem ser confundidas com a noção de transformação de baixo nível contra o processamento de alto nível que se encontra frequentemente na literatura de computador e visão humana. O "baixo nível" refere-se aos processos de extração de características básicas (por exemplo, deteção de bordas, filtragem) e "alto nível" implica processos complexos multi-componentes ou cognitivos, tais como reconhecimento de objetos ou de deteção de rosto humano. Em vez disso, a dicotomia bottom-up/top-down usada aqui é baseada no processar da informação visual que é feita no nível dos componentes ou a nível do sistema.

Um trabalho inicial considera 6 pontos de características (localizados usando janela de correlação) e em seguida determina o movimento da cabeça de uma fraca perspectiva geométrica. Mais tarde, uma outra abordagem assume que a face humana é uma superfície plana de um espaço ortogonal. Neste caso dois DOF (graus de liberdade) podem ser recuperados usando os mínimos quadrados ponderados para determinar a melhor transformação entre dois frames. A localização pode alternativamente empregar um modelo baseado na aparência encontrando a transformação do modelo que melhor representa o movimento observado da cabeça. Para a estimativa da pose é normalmente usado um modelo rígido 3D da forma da cabeça. E assim basta verificar qual a melhor rotação e translação do modelo que melhor se ajusta a cada nova imagem. Podemos usar estes métodos para medições contínuas da pose, usando a procura de gradiente descendente e aperfeiçoá-las com fluxo ótico para orientar a otimização.

A principal vantagem destes métodos é a sua capacidade de localizar a cabeça com grande eficácia encontrando pequenas mudanças na pose entre frames. Se usarmos estes métodos baseados em modelos também é possível construir um arquétipo individualizado da cabeça de uma pessoa. Isso permite que estas abordagens consigam evitar os efeitos negativos da variação da aparência. O problema destes métodos está na necessidade de fornecer uma inicialização da pose da cabeça para gerar ou adaptar o modelo. As melhorias deste tipo de métodos pode estar na utilização de um filtro de partículas para incorporar informação à priori sobre a dinâmica da cabeça. Num filtro de partículas clássico, uma observação do estado do objeto é observado em cada passo temporal e assumido como sendo ruidoso, a ótima localização pode ser encontrada através da maximização da probabilidade à posteriori do movimento, dada a observação, utilizando um conjunto de amostras simuladas. Para os métodos baseados na aparência, o filtro de partículas, ao invés de observar uma amostra ruidosa da posição e da orientação absoluta da cabeça, é obtida uma imagem da cabeça a cada intervalo temporal. Um conjunto de amostras de poses são geradas então à priori com um modelo dinâmico que servirá para tornar as vistas do modelo com diferentes transformações. Cada imagem virtual pode ser diretamente comparada com a imagem observada e estas comparações auxiliam a atualizar os pesos do filtro de partículas. Desta forma, permite-nos usá-los em sistema de tempo real, em diferentes ambientes incluindo a captura aproximada de vídeo, o vídeo de baixa resolução com sub-espacos adaptativos como PCA etc.

2.4.2.8 Métodos Híbridos

Como está implícito à sua designação estes métodos combinam um ou mais destes métodos anteriormente referidos para melhorar as limitações que estão inerentes a uma abordagem simples. Uma modalidade comum é o de completar uma abordagem de estimativa de pose com um sistema de rastreamento. O modelo estático será responsável pela inicialização e o sistema de rastreamento pelo reajustar da estimativa da pose ao longo do tempo. Se o rastreamento começar à deriva, o sistema estático reinicializa o sistema de rastreamento. Outras possíveis combinações são:

- um método geométrico automático com um localizador de pontos,
- um modelo de correspondência PCA com um fluxo ótico,
- um modelo de correspondência PCA com uma camada contínua escondida de um modelo de Markov,
- um modelo de correspondência PCA com uma correspondência para o frame chave com localização stereo por modelos de constância em tons de cinzentos e profundidade,
- modelos de aparência de cor e textura com um filtro de partículas baseado em imagens.

Os sistemas híbridos têm na sua principal vantagem a possibilidade de poder usar mais de dois sistemas simples e fundir os seus resultados de cada sistema independente num único resultado.

2.4.3 Desfoque Ótico

O desfoque pode ser causado por diversas causas, mas em geral, o desfoque ocorre quando o ponto de focagem está fora do campo de profundidade do objeto a ser capturado. Quanto mais distante um objeto está do intervalo do campo de profundidade, quanto maior será o grau de desfocagem. No domínio das frequências afeta toda a gama de frequências e principalmente aumenta a altas frequências.

Uma image de $M \times M$ pixels normalizada pelo espectro de força $P(u,v)$ é obtida pela fórmula 2.12, onde F é a transformada de Fourier, (u,v) são as duas coordenadas do domínio das frequências e M^2 é o número total de pixels da imagem [CY10].

$$P(u, v) = \frac{|F(u, v)|^2}{M^2} \quad (2.12)$$

O comportamento do espectro da potência de uma imagem natural pode ser vista através da função 2.13 assim definida por Cohen et al.

$$P(f) \propto \frac{1}{f^\gamma} \quad (2.13)$$

Uma imagem não-isotrópica média do espectro de potência (utilizando coordenadas polar) pode ser apresentada de uma forma mais geral, como função de orientação (ângulo) 2.14, onde $\gamma(\theta)$ é a inclinação espectral em função da orientação (média em relação ao ângulo θ).

$$P(f, \theta) \propto \frac{1}{f^{\gamma(\theta)}} \quad (2.14)$$

Para acentuar a visibilidade do efeito do desfoque, o potencial do espectro suavizado é ponderado, multiplicando o quadrado da frequência espacial f^2 para obter o MIS 2.15, onde $P(f)$ é a média radial normalizada (no que diz respeito ao ângulo) do espectro do potencial da imagem, S é o kernel médio de suavização e $*$ denota o operador de convolução.

$$MIS(f) = [P(f) * S]f^2 \quad (2.15)$$

O cálculo do MIS necessita ser normalizado pela fórmula NDMIS 2.16.

$$NDMIS(i) = \frac{MIS(i+1) - MIS(i)}{[MIS(i+1) + MIS(i)]/2} \quad (2.16)$$

O comportamento da função NDMIS 2.16 pode ser classificado por intervalos da seguinte forma:

$$\left\{ \begin{array}{l} \text{steep drop, } NDMIS(i) < -\delta \\ \text{moderate drop, } -\delta \leq NDMIS(i) \leq 0 \\ \text{rise, } 0 < NDMIS(i) \end{array} \right. \quad (2.17)$$

À mudança de "drop" ("steep" ou "moderate") para "rise" na MIS 2.15 é um mínimo local chamado defocus bending point. Supondo uma potência de desfoque constante ("moderate drop") e um poder de imagem decrescente ("steep drop"), quando aumenta a potência do desfoque, a curvatura do MIS 2.15 (em direção a um ângulo positivo) ocorre a uma frequência mais baixa. Isso significa que o número de componentes espectrais da imagem são significativamente afetados pelo desfoque superior ao defocus bending point, aumentando com o subida da potência do desfoque. Desta forma podemos calcular quais os pontos afetados pelo desfoque através da fórmula 2.18.

$$impacto\ do\ desfoque = \frac{pontos\ afetados\ pelo\ desfoque}{total\ de\ pontos} \quad (2.18)$$

2.4.3.1 Métodos Baseado num Kernel de Convolução

Este fator pode ser avaliado através da medição das altas frequências na globalidade da imagem ou apenas na zona de aplicação segmentada. O que difere este método do anterior, é a utilização de um kernel de convolução de 8×8 2.4 e medindo a potência total na resposta proposto por Daugman et al. O potencial do espectro da image 2D é então passado por uma fórmula não linear 2.19 proposta por Kalka et al. para obter um valor normalizado de 0 a 100. Nesta fórmula podemos dizer que X é o parâmetro que representa o potencial do espectro medido pelo kernel de convolução (8×8) e c é metade do potencial do valor de foque correspondente a 50%. Esta medida do espectro do foque resulta bem quando as imagens estão dominadas pela zona a analisar. Uma imagem está desfocada quando não contém informação de altas frequências. Contudo, em imagens que exista alguma oclusão, este resultado representa uma pontuação de foque global, pelo que é necessário avaliar o foque local da área de interesse. Para que isto não aconteça podemos aplicar a mesma medição espectral do foque localmente ao invés de globalmente numa área que seja conhecido como sendo menos provável a existência de oclusão.

Tabela 2.4: Kernel de convolução de 8x8 de Daugman.

-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	3	3	3	3	-1	-1
-1	-1	3	3	3	3	-1	-1
-1	-1	3	3	3	3	-1	-1
-1	-1	3	3	3	3	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1

$$f(x) = 100 \frac{x^2}{x^2 + c^2} \quad (2.19)$$

2.4.4 Desfoque de Movimento

Este fator de avaliação da qualidade de uma imagem resulta do movimento relativo entre o objeto ou a câmara durante o tempo de exposição, o que irá suprimir principalmente as altas frequências. Este facto pode resultar em desfoque de movimento linear ou não linear, contudo este método baseia-se apenas na estimativa de desfoque de movimento linear [KZSC10].

2.4.4.1 Método Baseado na Direção do Desfoque de Movimento

Calcular o desfoque de movimento linear é equivalente a estimar a direção primária do desfoque de movimento numa imagem, ao longo da força do desfoque de movimento nessa direção. Para calcular o ângulo são aplicados filtros direcionais no espaço de Fourier. Primeiro, à imagem inicial é aplicada a transformada de Fourier (seta A da figura 2.7). Depois é feito o produto interno da magnitude do logaritmo do resultado da transformada com os filtros direcionais (seta B da figura 2.7). O poder total dado por cada uma das direções será uma estimativa do ângulo aproximado (seta C da figura 2.7). A direção do desfoque de movimento é dada pela equação 2.20, sendo I a imagem, F(I) a transformada de Fourier da imagem, são aplicados filtros direcionais do escalar α e $H(\theta : \alpha)$ é a resposta de um filtro direcional aplicado à imagem na direção θ .

$$\hat{\theta} = \arg \max_{\theta \in [0:5:180]} \|F(I)H(\theta : \alpha)\|^2 \quad (2.20)$$

Uma vez que a direção seja determinada, a força é calculada analisando a parte da imagem com magnitude log perpendicular ao ângulo estimado do desfoque de movimento e que passa pelo centro. A largura do nóculo principal é inversamente proporcional à força do desfoque de movimento. Se medirmos o nóculo principal podemos obter uma estimativa da força do desfoque de movimento de movimento linear. Para obtermos este nóculo, os coeficientes são suavizados usando o B-spline com menos otimizações quadradas (seta D da figura 2.7). Finalmente adotamos uma abordagem baseada em gradientes para localizar o ponto central (seta D da figura 2.7).

2.4.4.2 Métodos Baseado no Espectro da Imagem

Tendo por sustentáculo a força do espectro 2.12, a diferença normalizada do espectro modificado da imagem 2.16 podemos usar esses cálculos para obter uma estimativa do desfoque de movimento existente numa dada imagem. A primeira mudança de "steep drop" 2.17 para "moderate drop"

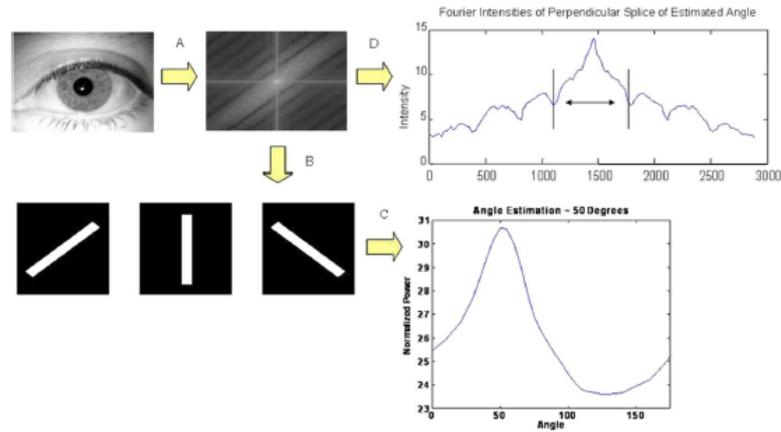


Figura 2.7: Diagrama de blocos do processo de estimativa do Desfoque de Movimento [KZSC10].

na MIS é um ponto de inflexão chamado motion blur bending point. Este ponto reflete o comportamento do desfoque de movimento da MTF. Sendo que o desfoque de movimento reduz as altas frequências do espectro da imagem para zero ou próximo de zero, a localização do motion blur bending point deve ser identificado pelas frequências com intensidade mais baixas do espectro. Portanto, um valor limiar pequeno é aplicado (cerca de 2% acima de zero) para distinguir a região do espectro de baixa intensidade. À medida que o desfoque de movimento cresce, é esperado que os pontos do potencial espectral sejam atenuados, resultando num aumento do número de componentes acima do motion blur bending point. Com isto podemos demarcar quais os pontos considerados com e sem desfoque de movimento, recorrendo à fórmula 2.21.

$$\textit{impacto do desfoque de movimento} = \frac{\textit{pontos afectados pelo Desfoque de Movimento}}{\textit{total de pontos}} \quad (2.21)$$

2.4.5 Fusão de Fatores de Qualidade de Imagem

Com estes resultados acima descritos 2.21 e 2.18, torna-se fácil o cálculo do impacto da distorção na qualidade da imagem causada pelo desfoque ótico e pelo desfoque de movimento 2.22 e a sua inversa que representa a classificação da qualidade de imagem 2.23.

$$\textit{Impacto da distorcao} = \textit{impacto de rudo} + \textit{impacto do desfoque} \quad (2.22)$$

$$\textit{Image quality score} = 1 - \textit{Impacto da Distorcao} \quad (2.23)$$

Uma forma de combinar fatores de qualidade de imagem é usar a teoria DS visto que tem uma baixa complexidade temporal relativamente à teoria Bayesiana, tratamento explícito relativamente a estatísticas dependentes entre dimensões e pela sua boa performance em muitas aplicações, como robótica, software de previsão de falhas e biometria. A teoria DS baseia-se em conceitos de crenças, proposições e provas. As crenças para proposições (semelhantes aos eventos da teoria Bayesiana) começam a zero, com a incerteza de ser igual a 1. As evidência a se que chega atualizam as crenças, o que resulta no decréscimo da incerteza. Na teoria DS, o modelo de crenças é construído numa álgebra Booleana de exclusão mútua de proposições conhecidas como imagem de discernimento

(denotado por θ). A crença na proposição $\text{Bel}(A)$ é a medida de certeza de que a evidencia A é verdadeira. Shaffer dá esta expressão para determinar as crenças.

Se θ é uma imagem de discernimento então

uma função $m : 2^\theta \rightarrow [0, 1]$ é chamada uma atribuição de probabilidade básica então é observado

1. $m(\emptyset) = 0$;
2. $\sum_{A \subset \theta} m(A) = 1$.

Para medir uma crença da proposição A , uma deve ser adicionada a todos os subconjuntos B pertencentes a A . $\text{Bel}(A) = \sum_{B \subset A} m(B)$.

A regra combinação de Dempster é usada para combinar crenças ao longo da mesma imagem de discernimento que são de diferentes fontes de evidencia. Esta é calculada pela soma ortogonal de todas as funções de crenças m , que resultam numa nova função baseada na combinação das evidencias 2.24 onde f é dada por 2.25.

$$m(C) = \frac{\sum_{A_i \cap B_j = C} f(m_1(A_i)m_2(B_j))}{1 - \sum_{A_i \cap B_j = \emptyset} f(m_1(A_i)m_2(B_j))} \quad (2.24)$$

$$f(m_1(A_i)m_2(B_j)) = [m_1(A_i)m_2(B_j)]^n, n \in [0, 1] \quad (2.25)$$

Esta regra caracteriza n como o método para pesar a evidencia. Se $n < 0.5$ dá maior importância quando combinando com uma nova evidencia, se $n < 0.5$ dá menor peso. Outros defensores da regra de Murphy caracterizam n como o valor de correlação entre as evidencias. $n > 0.5$ será mais independência entre as evidencias e $n < 0.5$ será correlação entre ambas e finalmente $n = 0.5$ é considerado neutro e o peso aplicado será igual para todas as evidencias. Porém antes de fundir os fatores é necessário normalizá-los. Para isso usamos a fórmula 2.26 para normalizar o desfoque ótico e o desfoque de movimento, onde para o desfoque ótico Q_{old} representa os valores da energia sem preparação obtida do bandpass kernel. No que diz respeito ao desfoque de movimento Q_{old} representa os valores de energia sem preparação correspondentes ao cume localizado, que foi extraído do filtro direcional que é perpendicular ao ângulo estimado.

$$Q_{new} = \frac{Q_{old} - Q_{min}}{Q_{max} - Q_{min}} \quad (2.26)$$

No quadro de discernimento estão contidas 2 proposições opostas que representam crenças de que a:

1. A – Qualidade de Imagem má
2. B – Qualidade de Imagem boa

Cada fator de qualidade foi normalizado e atribuído à proposição A . Dado que, as proposições representam crenças, são atribuídas a B as proposições complementares de A . Aplicando a regra de Murphy para combinar as crenças com o parâmetro $n = 0.5$, para k fatores de $m_1 a m_k$ usamos a fórmula 2.27, onde $m_i(B) = 1 - m_i(A)$. O problema é que se tivermos 7 fatores, vamos ter 5040 combinações ($7! = 5040$). É igualmente importante encontrar os valores máximo e mínimo dos fatores, que representariam o melhor e o pior caso do fator de qualidade. Mladenovski et

al. [Mla04] provou que o valor máximo pode ser alcançado pela primeira triagem das crenças em ordem crescente com $n = 0,5$. Da mesma forma, se classificados em ordem decrescente, o valor mínimo pode ser obtido.

$$m_i(\hat{A}) = \frac{(m_{i-1}(A).m_i(A))^n}{(m_{i-1}(A).m_i(A))^n + (m_{i-1}(B).m_i(B))^n}, i = 2, \dots, k \quad (2.27)$$

Capítulo 3

Métodos Propostos

Neste capítulo serão apresentados todos os métodos testados e implementados para discussão dos resultados obtidos no capítulo seguinte.

3.1 Pré Processamento das Imagens

O conjunto de imagens da base de dados utilizada contém uma grande variação de luminosidade, pelo que foi introduzida uma fase de pré processamento onde iremos suavizar essas diferenças. Este tratamento dos dados tal como foram captados, servirá para atenuar os efeitos diferenciadores das luminosidades aquando da captura das imagens do conjunto em análise, como exemplo na figura 3.1. Este pré processamento visa preparar totalidade das imagens de input de modo a que, na fase posterior a esta, seja analisada com a maior equidade possível.

3.1.1 Equalização de Histograma

A equalização de histograma baseia-se na necessidade existente de que os valores do histograma numa imagem ocuparem a maior parte da gama disponível e que é expressa numa imagem que aparenta um maior contraste facilitando a visualização de detalhes.

A função de equalização de histograma (`histeq` do Matlab) realça o contraste das imagens através da transformação dos valores numa imagem intensa, ou os valores no mapa de cores de uma imagem indexada, de modo que o histograma da imagem de saída corresponda a cerca de um histograma especificado.

3.2 Reconhecimento Biométrico

De forma a que fosse possível avaliar o impacto de cada um dos fatores no reconhecimento biométrico houve a necessidade de usar sistemas de reconhecimento. Como o intuito deste trabalho não era o sistema de reconhecimento, foram usadas dois sistemas, um que com o método SIFT [Low04] e outro com wavets diádicas.

3.2.1 SIFT

Um método SIFT tem na sua principal função a extração de características distintas invariantes de uma imagem, que pode ser usada para captar eficazmente correspondências entre diferentes objetos num determinado cenário. Estas características extraídas de uma imagem são invariantes à escala e à rotação, assim como permite uma correspondência robusta num conjunto com distorções, mudanças de ponto de vista em 3D, à existência de ruído, assim como mudanças de luminosidade. As características são altamente diferenciadas, no sentido de que uma única característica pode ser corretamente combinada com alta probabilidade contra um grande conjunto de características de muitas imagens diferentes.

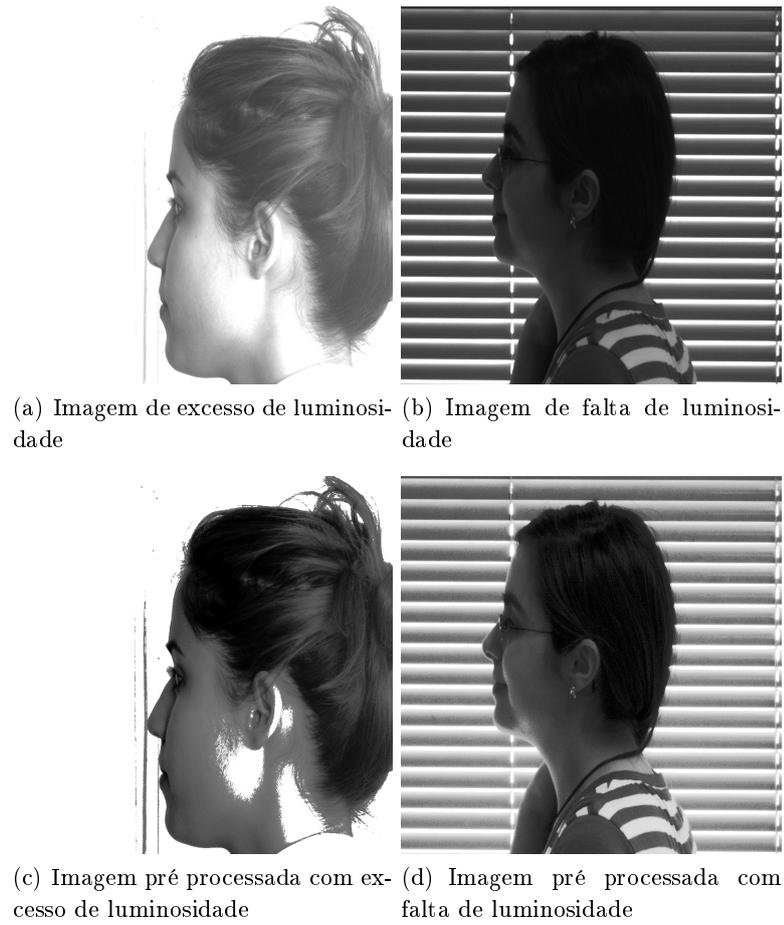


Figura 3.1: Exemplos de níveis de luminosidade muito diferenciada antes e depois da fase de pré processamento.

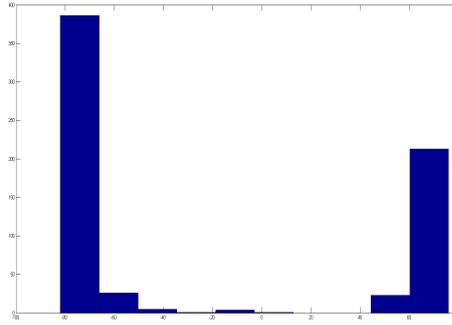


Figura 3.2: Histograma da densidade relativa ao fator Pose dos dados utilizados.

3.2.2 Wavelets Diádicas

Mallat e Zhong [MZ92] propuseram uma abordagem generalizada do detetor de arestas de Canny, e apresentaram uma transformada multiplas escalas de wavelets diádicas para a caracterização de sinais 1D e 2D. Com uma função wavelet $\psi(x) \in L^2(\mathbb{R})$, uma wavelet contínua transformação de $f(x)$ é dada pela fórmula 3.1.

$$W_{a,b}f(x) = \langle f, \psi_{a,b} \rangle = \int_{-\infty}^{+\infty} f(x) \frac{1}{a} \psi\left(\frac{x-b}{a}\right) dx \quad (3.1)$$

3.3 Factores Implementados

3.3.1 Pose

O cálculo automático da pose da cabeça de uma imagem de rosto é um problema de sub-análise de rosto humano com aplicações difusas, como deteção de olhar direção e como a interação homem - computador. A maioria dos métodos atuais estimam pose dentro de uma gama limitada ou tratam a pose como um problema de classificação, atribuindo ao rosto uma das muitas poses discretas. A abordagem do problema da estimativa da pose como um problema de regressão contínua de imagens de mundo real com grandes variações na iluminação de fundo e de expressão. A proposta feita por Aghajanian [AP09] considera uma estrutura probabilística com uma representação geral que não depende de localizar características faciais. Em vez disso, representam um cara com uma grelha de correções não sobrepostas. Esta representação usa um modelo generativo para o cálculo automático da pose numa representação variando de -90° a 90° em imagens capturadas em ambientes não controlados.

3.3.2 Luminosidade

A quantidade de luminosidade numa imagem aquando da sua captura tem uma influência determinante para a qualidade. Esta variação pode impossibilitar a visualização de certos detalhes da imagem real. Tanto o excesso como a falta de luminosidade são prejudiciais à imagem (como visualizamos no exemplo 3.1), uma vez que se houver um grande nível de luminosidade existe um brilho que não possibilita a visualização correta dos seus detalhes. Inversamente, também um nível de luminosidade demasiadamente baixo é penalizador no que diz respeito aos detalhes que poderiam ser visíveis com um nível de luminosidade correto. É possível quantificar o nível de

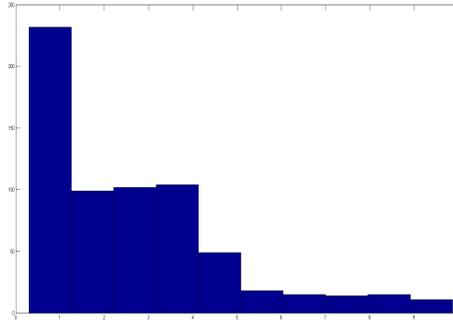


Figura 3.3: Histograma da densidade relativa ao fator Luminosidade dos dados utilizados.

luminosidade de uma determinada imagem através da formula 3.2, proposta por Kalka [KZSC10]. A densidade do fator luminosidade das imagens utilizadas neste trabalho pode ser analisada pelo histograma 3.3.

$$Lighting = \sum (X_i - \mu_R)^2 N, \text{ onde } \begin{cases} \mu_R = \frac{1}{N} \sum X_i \\ N = \# \text{ de regiões} \end{cases} \quad (3.2)$$

3.3.3 Desfoque Ótico

O Desfoque Ótico de uma imagem afeta a nitidez com que podemos observar os objetos. O Desfoque Ótico de uma imagem é caracterizado pelo menor número de componentes localizados na gama das altas frequência quando analisamos o resultado da aplicação da transformada de Fourier [Pro11]. Desta forma podemos determinar o quão focada a imagem está medindo a quantidade de altas frequências, isto é a quantidade de poder da convulsão entre a imagem I e o filtro passa-alto H, como denota a fórmula 3.3. Esta fórmula retorna bons resultados quando a imagem de entrada é dominado pelo ouvido. Desta forma, para este trabalho e tendo por base a máscara binária que nos indica a localização do ouvido, foi dada como entrada uma imagem adequada para obter o melhor resultado.

$$\alpha_f = \int_x \int_y |I * H|^2 dx dy \quad (3.3)$$

3.3.4 Desfoque de Movimento

O fator desfoque de movimento pode alterar a configuração do objetos caso ocorra um movimento quando a captura está a ocorrer. A quantidade de desfoque de movimento de uma imagem pode ser determinada por fórmula 3.4 proposta por Proença [Pro11]. A densidade de desfoque de movimento existente nos dados utilizados neste trabalho pode ser analisada através do histograma da figura 3.5.

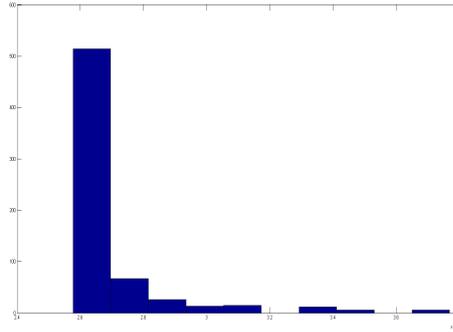


Figura 3.4: Histograma da densidade relativa ao fator Desfoque Ótico dos dados utilizados.

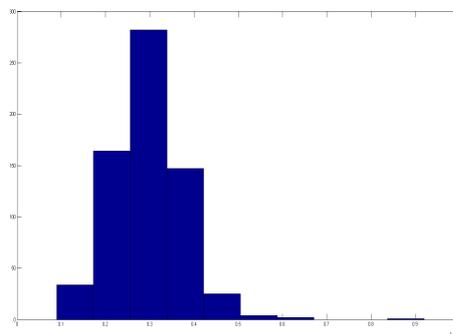


Figura 3.5: Histograma da densidade relativa ao fator desfoque de movimento dos dados utilizados.

$$\hat{\theta} = \underset{\theta \in [0.5:180]}{\operatorname{argmax}} \| F(I)H(\theta : \alpha) \|^2, \text{ sendo } \begin{cases} F(I) \text{ a transformada de Fourier,} \\ H(\theta : \alpha) \text{ o filtro resultante} \\ \text{de uma dada orientação } \theta \end{cases} \quad (3.4)$$

3.3.5 Oclusão

Consideramos oclusão como a quantidade não visível do objeto que estamos a analisar. Desta forma, a métrica implementada para medição da oclusão do ouvido tem por base um descritor angular. A Função Angular (Cumulativa Normalizada) transforma o contorno de uma forma 2D fechada num sinal unidimensional, invariante à translação e à rotação. Os passos seguintes descrevem os procedimentos seguidos para a implementação de uma métrica que possibilite a medição da quantidade de oclusão numa imagem.

1. Partindo da máscara binária (que nos informa a localização exata do ouvido), consideramos apenas o contorno dessa zona de interesse sem qualquer possível buraco que possam conter na imagem.
2. Aplicação da função que calcula o descritor angular (visível na imagem proposto por Nixon e Aguado [NA02], e daqui temos um descritor angular que mede a curvatura do ouvido (como

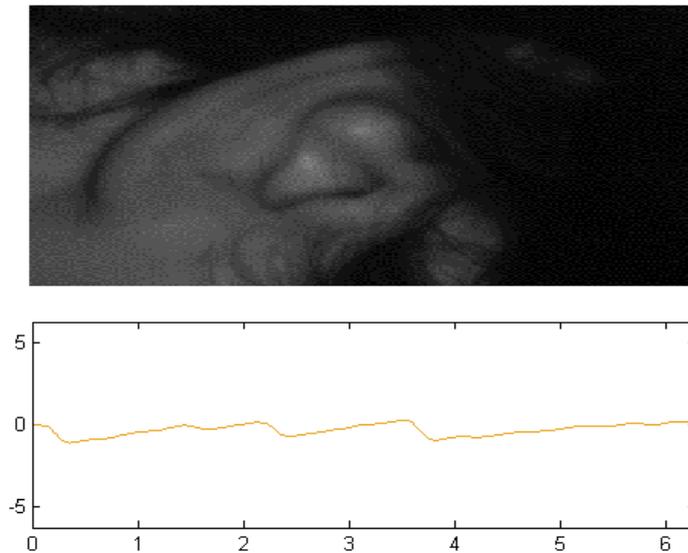


Figura 3.6: Exemplo de um ouvido e o correspondente descritor angular.

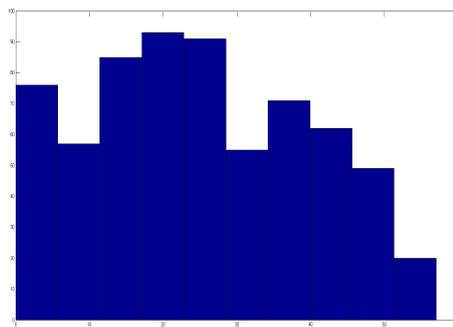


Figura 3.7: Histograma da densidade relativa ao fator Oclusão dos dados utilizados.

podemos observar o exemplo da imagem 3.6) que iremos comparar.

3. Comparar o ouvido objeto da nossa análise com todas as restantes imagens do conjunto de treino (que não têm qualquer tipo de oclusão) e encontrar o descritor angular que mais se aproxima (quando comparada através do cálculo da sua distância). Desta forma, conseguimos encontrar um ouvido que se mais se aproxima do ouvido que estamos analisar, ao nível da curvatura para nos possibilitar o cálculo da percentagem de oclusão.
4. Para calcular da percentagem de oclusão, consideramos a área do ouvido que mais aproximou os resultados do descritor angular e do qual temos garantia que não tem qualquer oclusão, com o ouvido que estamos a analisar.

O domínio de imagens com oclusão pode ser observado no histograma 3.7, para melhor compreensão dos resultados que iremos apresentar no capítulo seguinte.

Capítulo 4

Resultados e Discussão

4.1 Cálculo dos Pesos de Impacto de Cada Fator

O principal objetivo deste trabalho é o de investigar qual o impacto que os fatores acima referidos influenciam a qualidade de uma imagem e posteriormente o seu reconhecimento biométrico. Como tal, nesta seção serão explicitados todos passos que foram percorridos até aos resultados finais e consequentes conclusões.

4.1.1 Função de Distribuição Cumulativa

A probabilidade cumulativa refere-se à probabilidade de que o valor de uma variável aleatória se encontrar dentro de um intervalo especificado. Frequentemente, as probabilidades cumulativas referem-se à probabilidade de uma variável aleatória ser menor ou igual do que a um valor especificado, tal como especificado na fórmula 4.1.

$$P(x) = \int_{-\infty}^x P(t) dt \quad (4.1)$$

No âmbito deste trabalho foi utilizada a função de distribuição cumulativa à distancia resultante das comparações feitas no sistema de reconhecimento biométrico. No caso de a comparação que estamos a analisar ser resultante de uma intra classe, então o valor de retorno é apenas o valor resultante da função de distribuição cumulativa. Se a comparação analisado seja um inter classe, então o retorno devolvido será o valor de um subtraído ao valor de saída da função (1-CDF(d)). Desta forma, é possível diferenciar os casos de comparações intra e inter classes.

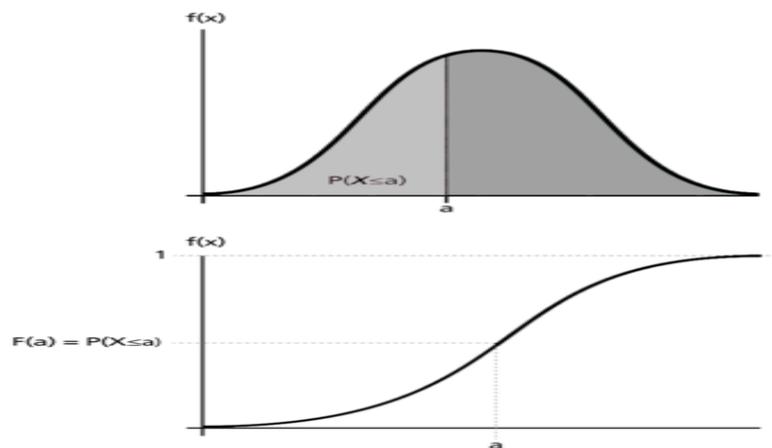


Figura 4.1: Exemplo da função de probabilidade cumulada.

4.1.2 Método de Ajustamento de Curva

Os métodos de ajustamento de curva, também conhecido como análise de regressão, são usados para encontrar o “melhor ajuste”, linha ou curva para uma série de dados. Na maioria das vezes, o ajustamento de curva vai produzir uma equação que pode ser usada para localizar pontos em qualquer ponto ao longo da curva. Em alguns casos, pode não ser necessário encontrar uma equação. Em vez disso, é possível apenas usar um ajustamento de curva para suavizar os dados e melhorar a aparência do seu enredo.

O método de Least Squares é um método de ajustamento de curva, amplamente conhecido ao longo tempo. O método de ajustamento de curva minimiza o erro quadrado entre os dados originais e os valores previstos pela equação que desejamos ajustar. Embora método possa não ser o mais robusto no ajuste de uma função para um conjunto de dados, tem a vantagem de ser relativamente simples (em termos de poder de computação necessário) e de ser bem compreendidos.

4.1.2.1 Função Lsqcurvefit

De forma a que nos fosse possível analisar o impacto dos fatores de qualidade numa imagem, retirámos partido da utilização da função de ajuste (o `lsqcurvefit` do Matlab). Esta função tenta ajustar os coeficientes que melhor ajustam a função de input aos dados resultantes do reconhecimento. Esta função tenta resolver problemas não-lineares de ajuste de uma curva (dados de entrada), no sentido de mínimos quadrados. Isto é, dando os dados de entrada $xdata$, e sendo $ydata$ o resultado observado, encontram os coeficientes de x que “melhor ajustam” a equação. Um dos parâmetros de entrada do ajustamento de curva é a função que queremos ajustar aos dados, pelo que a função usada para os fatores em separado foi a função 4.2, sendo $F(x)$ o valor da função CDF 4.1.1 e X o valor do fator em análise, m o declive da reta e o $(0, b)$ é o ponto de interseção da reta com o eixo dos yy .

$$F(x) = mX + b; \quad (4.2)$$

4.1.2.2 Análise Individualizada dos Fatores

Numa fase preliminar a análise incidiu em cada um dos fatores separadamente. A seguir, serão listados os resultados obtidos para o ajuste dos dados, de cada fator. Cada um deles podemos examinar o dados e a reta que melhor ajusta os pesos associados ao fator graficamente nos gráficos 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 4.10, 4.11, 4.12 e 4.13. Sendo que o reconhecimento biométrico se traduz na comparação de dois sujeitos (no caso dois ouvidos), temos dois ouvidos cada um com a respetiva quantificação da qualidade. A forma adotada para harmonizar os valores de reconhecimento com o fatores de qualidade foi considerar o sempre o pior caso, mais concretamente considerou-se sempre o ouvido com pior qualidade.

Como podemos observar na tabela 4.2 resultantes do métodos de ajuste dos fatores aos dados de output do reconhecimento, apenas o fator de desfoque ótico tem um declive positivo o que significa que é o único que podemos afirmar que influencia diretamente o reconhecimento. Uma vez que o seu valor absoluto é alto concluímos que a sua importância é maior do que para os restantes fatores. Isto pode justificar-se pelo fato de que quando existe uma grande desfocagem ótica a imagem fica praticamente impercetível a sua análise.

De forma inversa, temos os restantes fatores (Luminosidade, Desfoque de Movimento, Oclusão e Pose) influenciam inversamente o reconhecimento biométrico. Entre estes, podemos ainda destacar

		Foco	Luminosidade	Motion Blur	Oclusion	Pose
Dyadic	m	1885000	-1.125	-0.00006081	-0.4348	-17.22
	b	25930000	3.291	0.0003313	25.34	-13.3
SIFT	m	373700	-0.2552	-0.000005057	-0.2295	-21.43
	b	26820000	2.775	0.0002993	25.2	-13.15

Figura 4.2: Tabela resumo dos coeficientes encontrados para cada fator individualmente.

		Foco	Luminosidade	Motion Blur	Oclusion	Pose
SSE	Dyadic	2.21E+15	2897	4.13E-06	1.43E+05	3.01E+06
	SIFT	2.31E+15	2930	4.23E-06	1.43E+05	3.00E+06
R-Square	Dyadic	0.04242	0.01191	0.02413	3.65E-05	0.002713
	SIFT	0.002166	0.0007963	0.0002167	1.32E-05	0.005454
Adjusted R-Square	Dyadic	0.04096	0.0104	0.02264	-0.001486	0.001195
	SIFT	0.0006472	-0.0007246	-0.001305	-0.001509	0.003941
RMSE	Dyadic	1.84E+06	2.1	7.93E-05	14.75	67.69
	SIFT	1.87E+06	2.112	8.02E-05	14.75	67.6

Figura 4.3: Tabela resumo de erro dos coeficientes encontrados para cada fator individualmente.

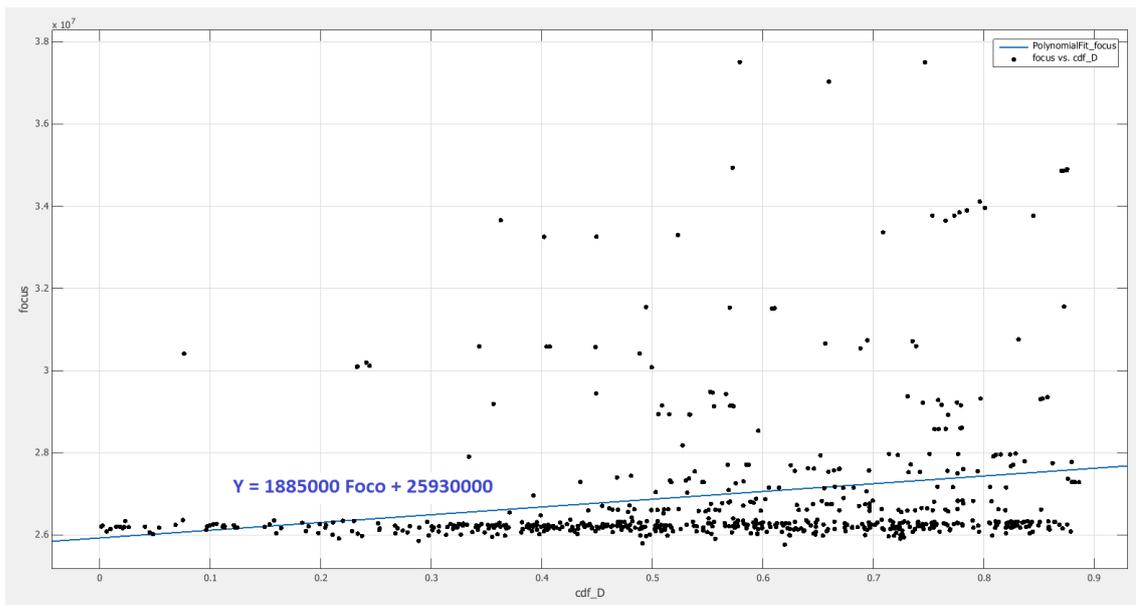


Figura 4.4: Fitting resultante do fator de Desfoque Ótico usando o método de wavelets diádicas.

o fator pose visto que afeta em maior grau o reconhecimento do que os remanescentes, visto o seu valor absoluto ser superior. Análise semelhante é aplicada aos restantes fatores, pelo que desta forma ficamos com o conhecimento de quais e de que forma os fatores analisados influenciam o reconhecimento. O fato de a pose se destacar dos demais fatores pode se explicar porque na presença de oclusão poderemos estar a ocultar detalhes importantes no reconhecimento biométrico. Em relação aos restantes fatores, a justificação da sua relevâncias prende-se com os valores apresentados na tabela 4.2 consoante seja maiores (em valores absoluto e com o seu sinal que nos informa se a sua influência é direta (com declive positivo) ou indireta (com uma reta de declive negativo).

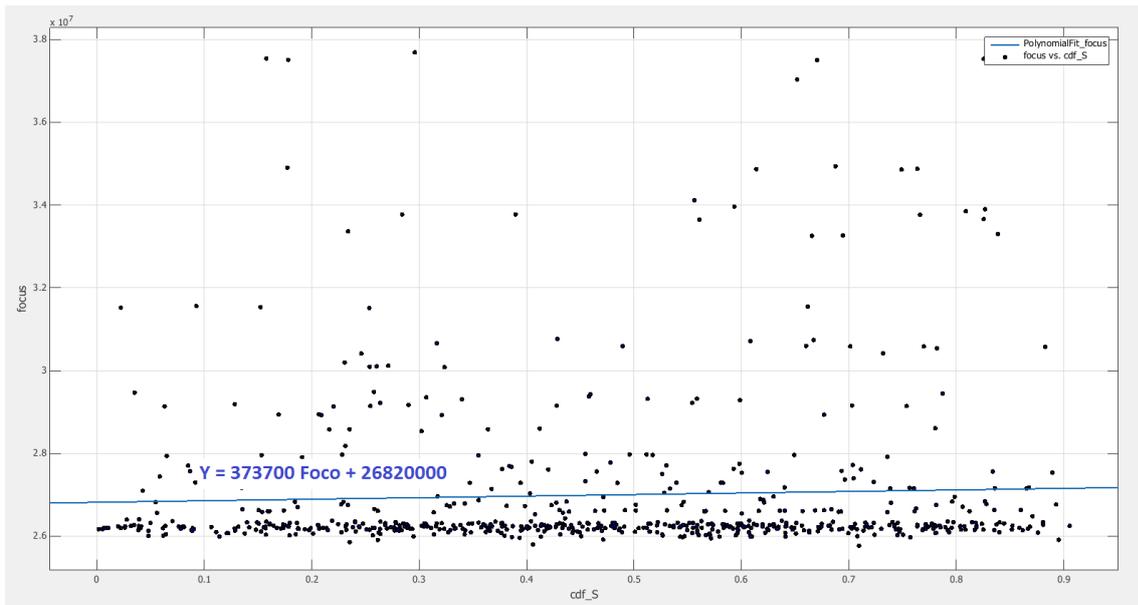


Figura 4.5: Fitting resultante do fator de Desfoque Ótico usando o método de SIFT

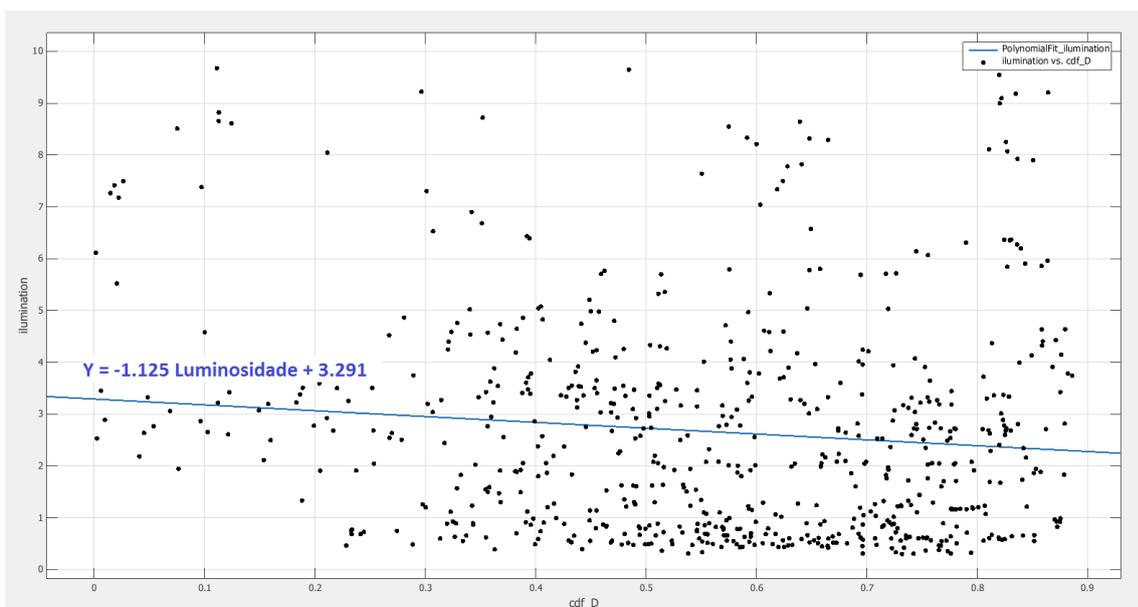


Figura 4.6: Fitting resultante do fator de Luminosidade usando o método de wavelets diádicas.

4.1.2.3 Correlação de Pearson

A correlação entre duas variáveis reflete o grau em que as variáveis estão relacionadas. A medida mais comum para este cálculo é a correlação de Pearson ou correlação do momento do produto. Quando medida numa população, a correlação de Pearson é designada pela letra grega rho (ρ). Quando calculada numa amostra, é designada pela letra "r" e é algumas vezes chamado de "r de Pearson." A correlação de Pearson reflete o grau de relação linear entre duas variáveis. Ela varia de 1 a -1. A correlação de um significa que existe uma relação positiva linear perfeita entre as variáveis.

Como forma certificar os resultados apresentados em 4.2 foi usada a correlação de Pearson entre a CDF do resultado do algoritmo de reconhecimento e cada um dos fatores separadamente, como

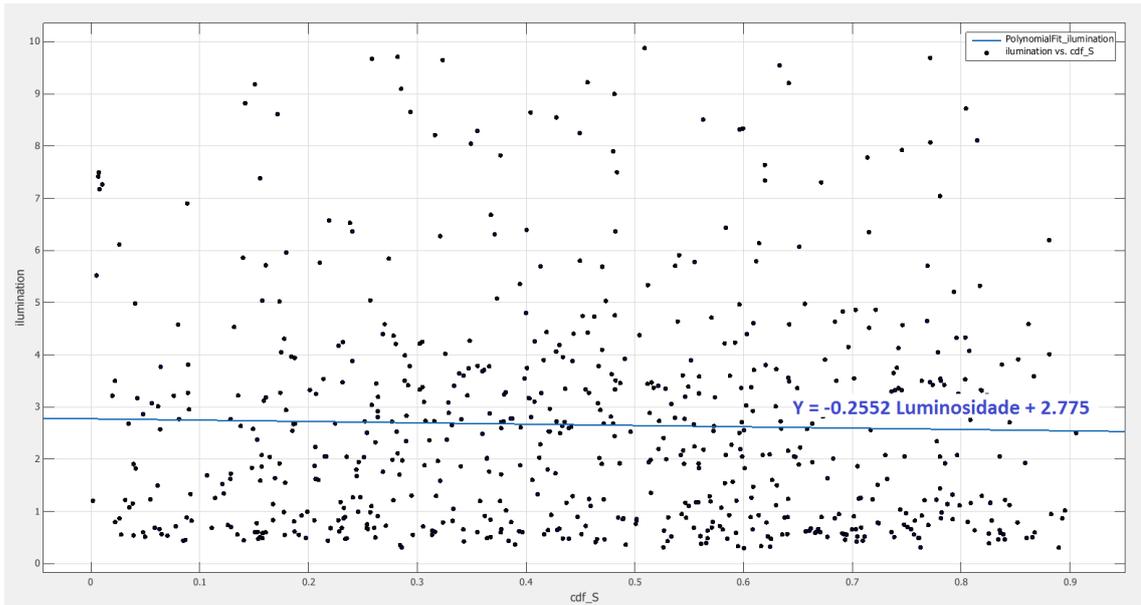


Figura 4.7: Fitting resultante do fator de Luminosidade usando o método de SIFT

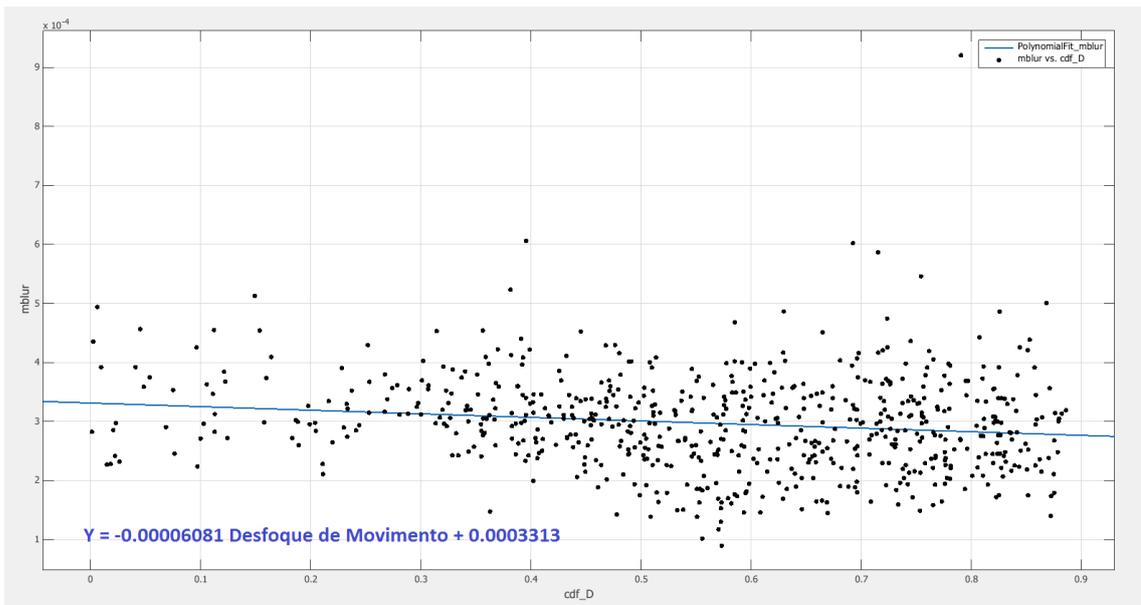


Figura 4.8: Fitting resultante do fator de Desfoque de Movimento usando o método de wavelets diádicas.

pode ser observado na tabela 4.1 os valores da correlação confirmam os resultados apresentados anteriormente.

Algoritmos	Fatores				
	Desfoque Ótico	Luminosidade	Desfoque de Movimento	Oclusão	Pose
Wavelets diádicas	0.2060	-0.1091	-0.1553	-0.0060	-0.0521
SIFT	0.0465	-0.0282	-0.0147	-0.0036	-0.0739

Tabela 4.1: Resultados obtidos pelo cálculo da correlação de Pearson

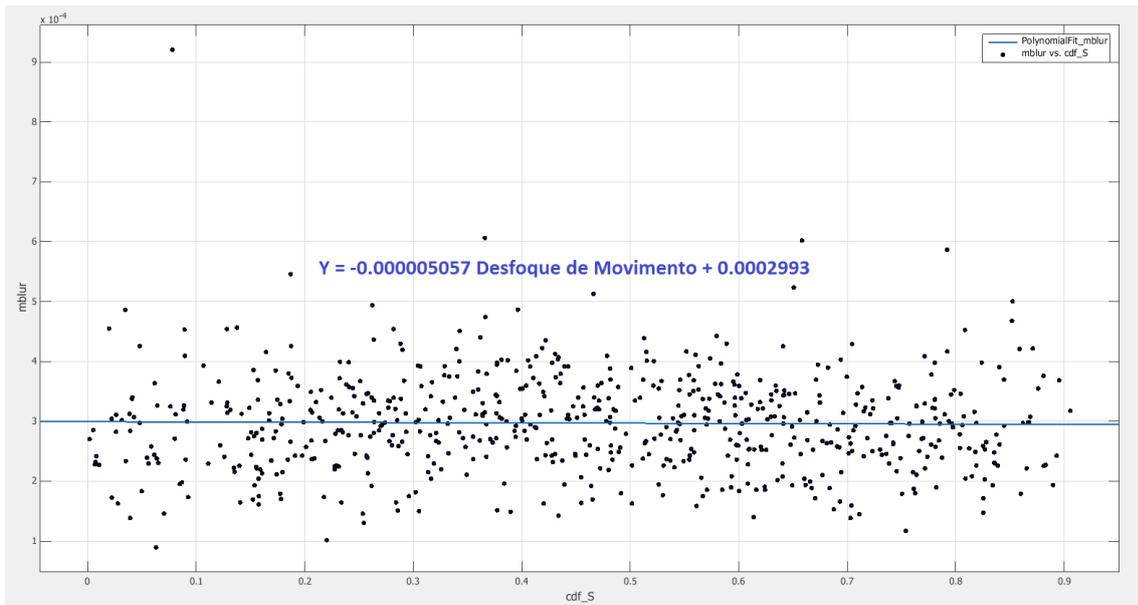


Figura 4.9: Fitting resultante do fator de Desfoque de Movimento usando o método de SIFT

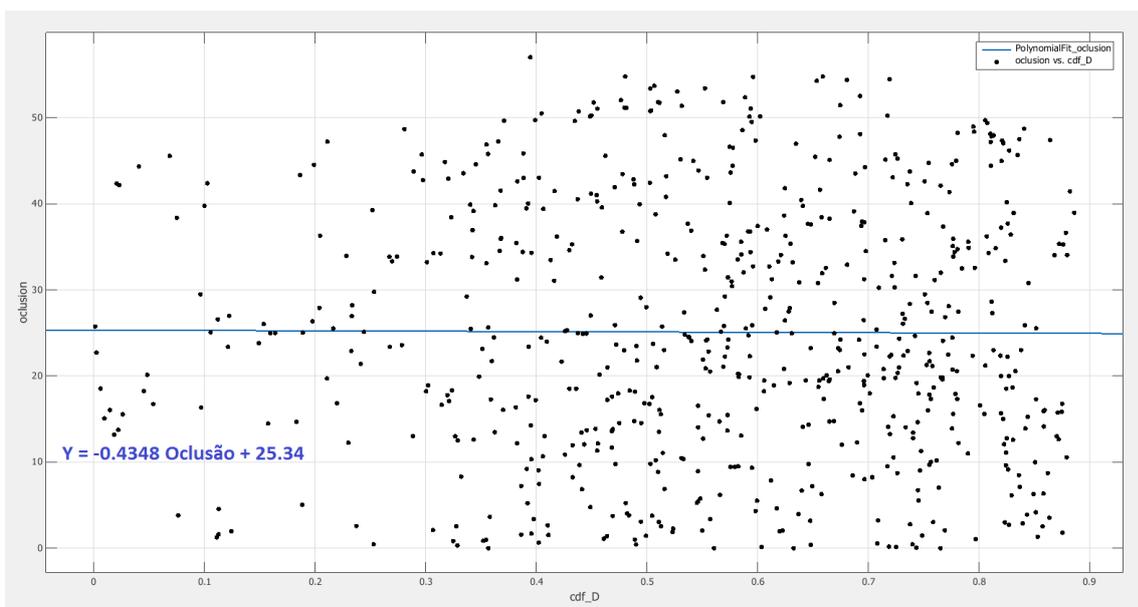


Figura 4.10: Fitting resultante do fator de Oclusão usando o método de wavelets diádicas.

4.1.2.4 Análise Global dos Fatores

Finalmente, para que possamos obter uma visão genérica do impacto combinados de todos os fatores no reconhecimento, pode-se realizar uma nova execução da função `Lsqcurvefit` 4.1.2.1 acima descrita onde se pretende ajustar a equação 4.3. Deste ajuste foram obtidos os coeficientes da tabela 4.2 com os valores residuais da tabela 4.3, dos quais podemos vislumbrar algumas diferenças em relação às conclusões obtidas quando a análise foi feita a cada um dos fatores separadamente em 4.1.2.2. Perante os valores obtidos podemos afirmar que no conjunto dos fatores analisados o fator desfoque de movimento é o que demonstrámos ser mais importante, visto que tem o valor absoluto mais elevado. De destacar também o fator pose que nesta condições de análise se demonstrou ser revelante e contribuir com um efeito direto para o reconhecimento.

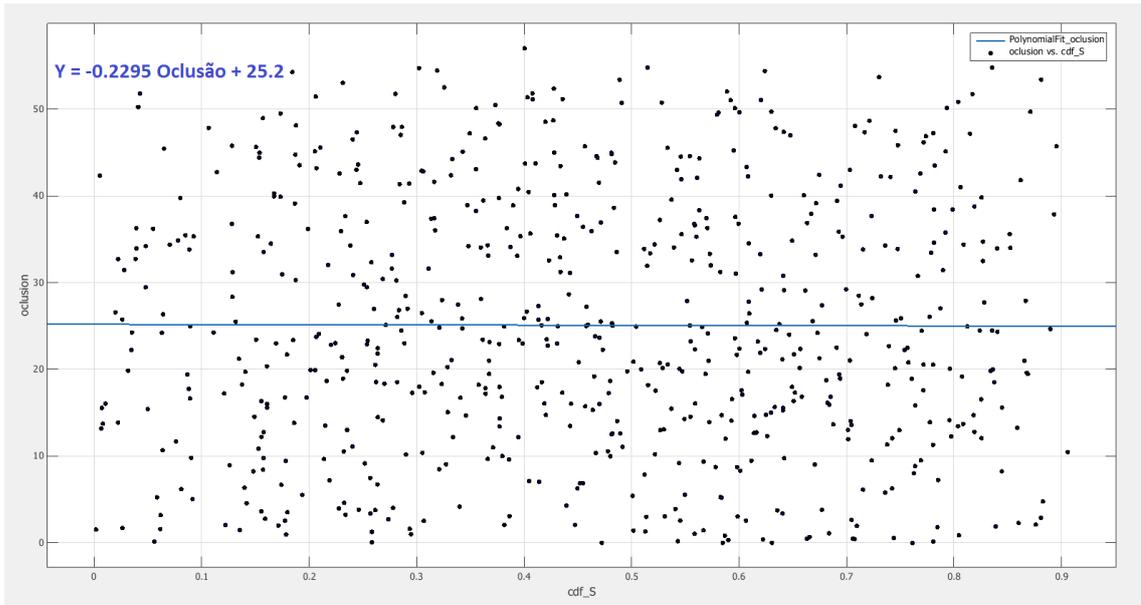


Figura 4.11: Fitting resultante do fator de Oclusão usando o método de SIFT

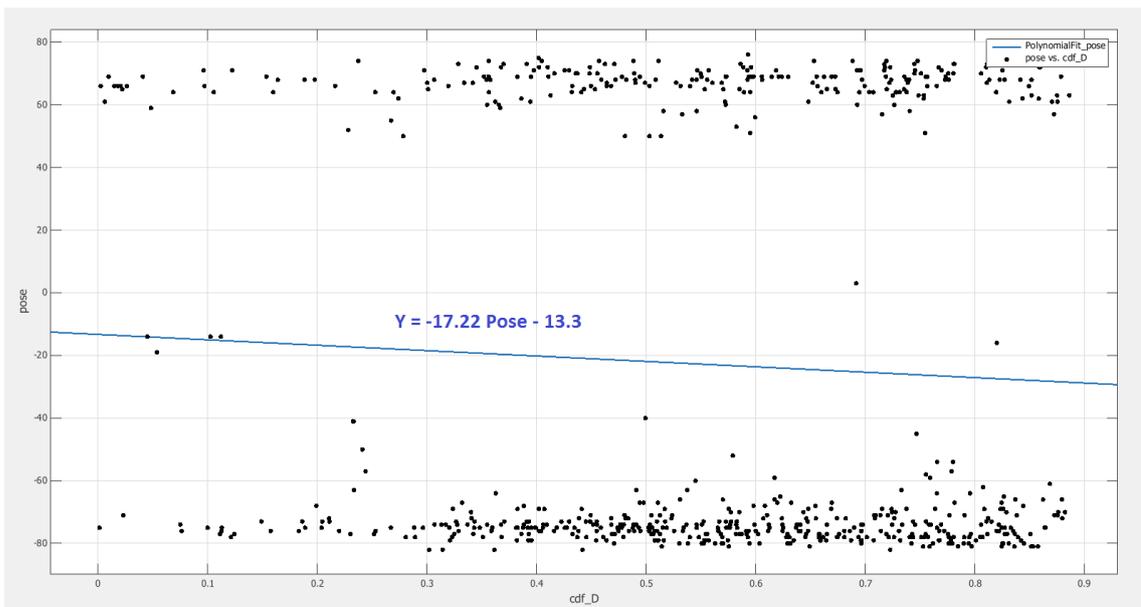


Figura 4.12: Fitting resultante do fator de Pose usando o método de wavelets diádicas.

$$F(x) = m_{Desfoquetico} \times Desfoquetico + m_{Luminosidade} \times Luminosidade + \quad (4.3)$$

$$m_{DesfoquedeMovimento} \times DesfoquedeMovimento + m_{Ocluso} \times Ocluso + \quad (4.4)$$

$$m_{Pose} \times Pose + b \quad (4.5)$$

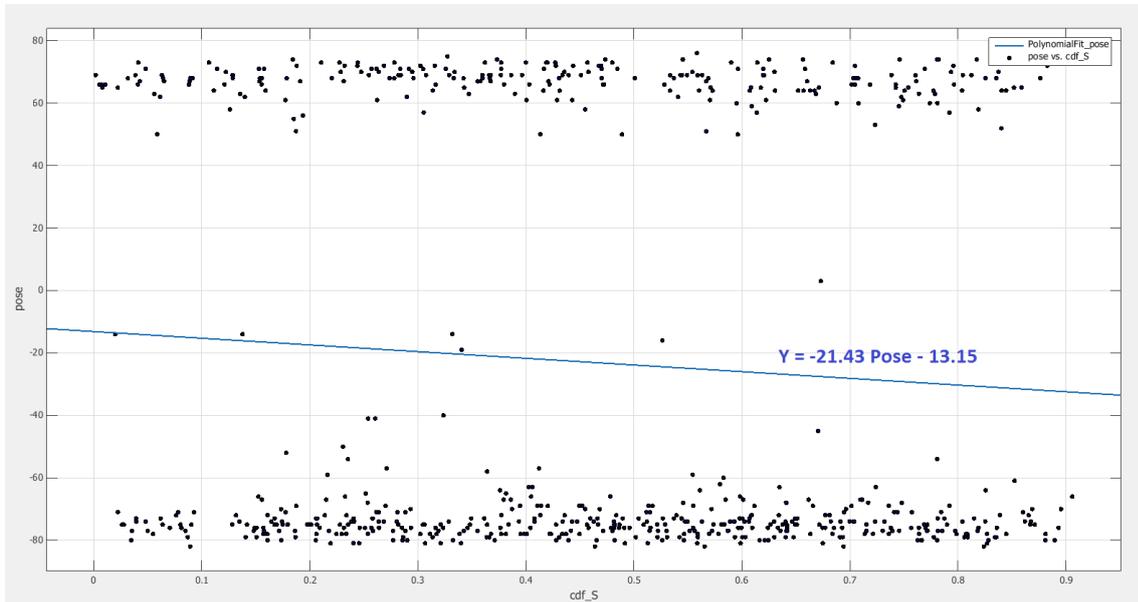


Figura 4.13: Fitting resultante do fator de Pose usando o método de SIFT

Algoritmo	$m_{Desfoquetico}$	$m_{Luminosidade}$	$m_{DesfoquedeMovimento}$	m_{Ocluso}	m_{Pose}	b
Wavelets Diádicas	0.0000	-0.0052	-225.4084	0.0002	-0.0002	0.1637
SIFT	0.0000	-0.0012	11.7770	-0.0001	-0.0003	0.2730

Tabela 4.2: Tabela resultante do calculo dos coeficientes com todos fatores.

Algoritmo	Resnorm
Wavelets Diádicas	26.0703
SIFT	35.5733

Tabela 4.3: Tabela dos valores residuais no ajuste realizado pela função lsqcurvefit.

Capítulo 5

Conclusões

5.1 Resumo do Trabalho Realizado

Existem, neste momento, diversas formas de aplicação de reconhecimento biométrico sendo que, as características mais comuns na aplicação desta técnica baseiam-se nas impressões digitais, voz, geometria das mãos ou do rosto e íris.

A biometria como método tem ganho, cada vez mais, notoriedade em outras áreas, começando a substituir pelos métodos mais tradicionais, como a utilizações de códigos (pin).

Este êxito deve-se à obrigatoriedade da presença física do indivíduo junto ao ponto de identificação e à eliminação da necessidade de memorizar senhas ou transportar documentos identificativos que poderiam ser facilmente esquecidos ou roubados.

Como referido, existem outras formas de biometria sendo que, este trabalho tem como principal objetivo realizar um estudo sobre o impacto que a qualidade da imagem tem no reconhecimento biométrico da orelha. Como tal, para a concretização deste propósito, foram percorridas diversas fases na tentativa desse reconhecimento, sendo elas:

- Em primeiro lugar, foram tiradas diferentes imagens num ambiente não controlo que à posteriori foi disponibilizado à comunidade científica [RHPP], de inúmeros indivíduos, em diversas condições;
- Posteriormente, a essas imagens foi aplicado um pré-processamento de entrada, de forma a esbater as discrepâncias de luminosidade, que se encontram sempre presentes em ambientes não controlados (como o caso);
- A fase seguinte consistiu em quantificar cada um dos fatores que considerámos mais relevantes, para cada imagem individualizada;
- Tendo a qualidade da imagem quantificada, cada uma das imagens foi devidamente processada nos sistemas de reconhecimento referenciados, para alcançarmos os valores relacionados com o reconhecimento biométrico da orelha;
- Após a obtenção desses valores, foram realizadas as devidas adaptações para moldar devidamente os dados de forma a prover o método de ajuste dos dados corretos, para que se encontre desta forma o melhor ajuste nos dados para cada fator individualmente;
- Posto isto, a função de ajuste, desta feita, foi aplicada a todos os fatores com um todo, em concordância com a função por nós definida, prevê os melhores valores que ajustam os dados de entrada com os fornecidos do sistema de reconhecimento biométrico.
- Finalmente foram retiradas as devidas conclusões discutidas no capítulo antecedente demonstrando o valor real de cada fator individualmente e analisados como um todo.

Várias investigações, principalmente conduzidos por Mark Nixon, têm sido realizadas no âmbito do reconhecimento biométrico apresentando um elevado interesse por parte da ciência e da segurança, por ser um método inovador.

Como apresentado ao longo do trabalho, para além das técnicas biométricas tradicionais, através da orelha também se obtém uma definição biométrica do indivíduo. E, como em cada uma dessas técnicas, esta também é uma técnica individual, ou seja, cada indivíduo possui características diferentes da orelha.

Como em todas as áreas, o reconhecimento biométrico também possui vantagens e desvantagens. O professor Nixon apontou, como vantagem desta biometria em relação à mais tradicional como o reconhecimento do rosto, que os ouvidos possuem uma estrutura rica e estável desde o nascimento até à velhice? ao contrário do que acontece com a face. No entanto, como desvantagem existe o facto de as orelhas estarem (ou não) escondidas pelo cabelo ou de haver oclusões por brincos, problemas na fase de captura como o excesso/falta de luminosidade e a pose da cabeça.

O estudo incidiu em diversos fatores e diferentes sistemas de reconhecimento biométrico para que fosse possível uma análise mais aprofundada nas suas conclusões. Esse foi um objetivo cumprido com sucesso, embora permaneça a ideia de que se pode fazer mais e melhor. Tendo por base os resultados obtidos ao longo deste trabalho, podemos concluir que os fatores que no geral mais influência tiveram no reconhecimento são o desfoque de movimento e a pose.

Com a realização deste trabalho, esperamos que este estudo tenha sido um contributo para futuros melhorias tanto na fase de captura de novas bases de dados em ambientes não controlados, como também na avaliação de sistemas de reconhecimento biométrico (sejam eles, novas propostas, como para existentes com novas abordagens).

5.2 Trabalho(s) Futuro(s)

Vivemos num mundo da globalização e, portanto, a cada dia surgem novas técnicas e novas áreas de investigação. Nunca nenhuma área na sua totalidade se encontra totalmente investigada e, assim sendo, como trabalho futuro poderão ser explorados novos fatores, como por exemplo:

- a resolução; visto que, imagens de baixa resolução podem não ter resolução suficiente para que seja possível concretizar o reconhecimento;
- a compressão; uma vez que cada vez mais vivemos num mundo global, uma análise válida será este balancear o tamanho físico das imagens com a qualidade que poderão guardar;
- a presença de ruído, por inúmeras razões as imagens podem conter ruído que pode ser um obstáculo na análise da imagem real.

Tão importante como investigar novos fatores que influenciam a qualidade de imagem, importa também estudar formas distintas de calcular os fatores atualmente usados, para nos possibilitar a comparação e estudo das melhores.

Um dos fatores que pode ser susceptível de melhoria é o fator oclusão. Pode ser vantajoso um método de deteção automática de oclusão, assim como a análise de formas distinguir os tipos de

oclusão (por brincos ou provocado pelo cabelo) e pelo estudo das zonas do ouvido mais relevantes para o reconhecimento.

A exploração de outras formas de avaliação da qualidade de imagem são igualmente importantes, tais como Foveated Wavelet Image Quality Index [WBL03] e Structural SIMilarity Indexs [Mor92]. Adicionalmente seria importante estudar novas formas de ajustar os pesos da fórmula que contempla todos os fatores.

Bibliografia

- [AKS05] Anton H. M. Akkermans, Tom A. M. Kevenaer, and Daniel W. E. Schobben. Acoustic ear recognition for person identification. In *AutoID*, pages 219–223. IEEE Computer Society, 2005. Available from: <http://doi.ieeecomputersociety.org/10.1109/AUTOID.2005.11>. 13
- [ANRR06] A. F. Abate, M. Nappi, D. Riccio, and S. Ricciardi. Ear recognition by means of a rotation invariant descriptor. In *ICPR*, pages IV: 437–440, 2006. Available from: <http://doi.ieeecomputersociety.org/10.1109/ICPR.2006.465>. 13, 14
- [AP09] J. Aghajanian and S. J. D. Prince. Face pose estimation in uncontrolled environments. In *BMVC*, pages xx–yy, 2009. Available from: <http://www.bmva.org/bmvc/2009/Papers/Paper256/Paper256.pdf>. 31
- [BB99] M. Burge and W. Burger. Ear biometrics. In *BIDNS99*, pages 273–286, 1999. 13
- [BB02] M. Burge and W. Burger. Ear biometrics in computer vision. pages 826–830, 2002. 13
- [CB05] H. Chen and B. Bhanu. Contour matching for 3D ear recognition. In *WACV*, pages I: 123–128, 2005. Available from: <http://doi.ieeecomputersociety.org/10.1109/ACVMOT.2005.38>. 13
- [CB07] H. Chen and B. Bhanu. Human ear recognition in 3D. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(4):718–737, apr 2007. Available from: <http://dx.doi.org/10.1109/TPAMI.2007.1005>. 13
- [CBSV03] Kyong Chang, K.W. Bowyer, S. Sarkar, and B. Victor. Comparison and combination of ear and face images in appearance-based biometrics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1160 – 1165, sept. 2003. xi, 13, 14
- [cib] CIBIM 2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management. <http://www.ieee-ssci.org/2011/cibim-2011>. 51
- [CY10] Erez Cohen and Yitzhak Yitzhaky. No-reference assessment of blur and noise impacts on image quality. *Signal, Image and Video Processing*, 4(3):289–302, 2010. Available from: <http://dx.doi.org/10.1007/s11760-009-0117-4>. 22
- [DW96] J. Daugman and G. Williams. A proposed standard for biometric decidability. pages 223–234. *Proceedings of the CardTech/SecureTech Conference*, 1996. 11
- [Fit01] Paul Fitzpatrick. Head pose estimation without manual initialization, January 21 2001. Available from: <http://citeseer.ist.psu.edu/609822.html>; <http://www.ai.mit.edu/projects/lbr/hrg/2001/paulfitz-headpose.pdf>. xi, 18
- [Fu86] K. S. Fu. A step towards unification of syntactic and statistical pattern recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(3):398 – 404, May 1986. 5
- [HAZN07] DJ Hurley, B Arbab-Zavar, and MS Nixon. The ear as a biometric. In *EUSIPCO 2007*, pages 25–29, 2007. Available from: <http://eprints.ecs.soton.ac.uk/14771/>. 14

- [HNC05] David J. Hurley, Mark S. Nixon, and John N. Carter. Force field feature extraction for ear biometrics. *Computer Vision and Image Understanding: CVIU*, 98(3):491–512, jun 2005. 13
- [Ian89] Alfred V. Iannarelli. *Ear Identification (Forensic Identification Series)*. Paramount Publishing Company, Fremont, California, first edition, 1989. 12, 13
- [JDM00] A.K. Jain, R.P.W. Duin, and Jianchang Mao. Statistical pattern recognition: a review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1):4–37, January 2000. xiii, 3, 4
- [JRP04] A.K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1):4–20, jan 2004. 5
- [KZSC10] N.D. Kalka, Jinyu Zuo, N.A. Schmid, and B. Cukic. Estimating and fusing quality factors for iris biometric images. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 40(3):509–524, May 2010. xi, 24, 25, 32
- [Low04] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 29
- [LS07] Yang Liu and E. Shriberg. Comparing evaluation metrics for sentence boundary detection. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 4, pages IV–185–IV–188, april 2007. xiii, 10
- [LSTC10] J-C Lee, Y Su, T-M Tu, and C-P Chang. A novel approach to image quality assessment in iris recognition systems. *Imaging Science Journal, The*, Volume 58, Number 3, June 2010, pp. 136-145(10), 58(3):136–145, June 2010. xi, 17
- [MCT09] E. Murphy-Chutorian and M.M. Trivedi. Head pose estimation in computer vision: A survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(4):607–626, april 2009. 18
- [Mla04] Mladenovski. Information fusion schemes for real time risk assessment in adaptive control systems. Master’s thesis, West Virginia Univ., Morgantown, 2004. 27
- [Mor92] Thomas Mormonn. Structural accessibility and similarity of possible worlds. *Journal of Philosophical Logic*, 21(2):149–172, 1992. 45
- [MSV99] B. Moreno, A. Sanchez, and J.F. Velez. On the use of outer ear images for personal identification in security applications. In *Security Technology, 1999. Proceedings. IEEE 33rd Annual 1999 International Carnahan Conference on*, pages 469–476, 1999. 13
- [MZ92] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 14:710–732, 1992. 31
- [NA02] Mark S. Nixon and Alberto S. Aguado. *Feature extraction and image processing*. Newnes, Oxford, 2002. 33
- [Pro09] Hugo Proença. Biometric recognition: When is evidence fusion advantageous? In George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Yoshinori Kuno, Junxian Wang, Renato Pajarola, Peter Lindstrom, André Hinkenjann, MiguelL. Encarnação, CláudioT. Silva, and Daniel Coming, editors, *Advances in Visual Computing*, volume

- 5876 of Lecture Notes in Computer Science, pages 698–708. Springer Berlin Heidelberg, 2009. Available from: http://dx.doi.org/10.1007/978-3-642-10520-3_66. 11
- [Pro11] Hugo Proença. Quality assessment of degraded iris images acquired in the visible wavelength. *IEEE Transactions on Information Forensics and Security*, 6(1):82–95, 2011. Available from: <http://dx.doi.org/10.1109/TIFS.2010.2086446>. 32
- [RAC05] G. Ruttly, A. Abbas, and D. Crossling. Could earprint identification be computerised? an illustrated proof of concept paper. *International Journal of Legal Medicine*, 119:335–343, 2005. 10.1007/s00414-005-0527-y. Available from: <http://dx.doi.org/10.1007/s00414-005-0527-y>. 13
- [RHPP] Rui Raposo, Edmundo Hoyle, Adolfo Peixinho, and Hugo Proença. Ubear: A dataset of ear images captured on-the-move in uncontrolled conditions. In *CIBIM 2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management*, pages 84 – 90. vii, xi, 1, 15, 43
- [VBS02] B. Victor, K. Bowyer, and S. Sarkar. An evaluation of face and ear biometrics. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 1, pages 429 – 432 vol.1, 2002. 13
- [VJ01] P. Viola and M. J. Jones. Robust real time object detection. In *Workshop on Statistical and Computational Theories of Vision*, 2001. 3
- [Wat85] Satoshi Watanabe. *Pattern Recognition: Human and Mechanical*, pages 75–83. Wiley – Interscience (John Wiley & Sons), 1985. 3
- [Way01] J. L. Wayman. Fundamentals of biometric authentication technologies. *International Journal of Image and Graphics (IJIG)*, 1, January 2001. 9
- [WB06] Zhou Wang and Alan C. Bovik. *Modern Image Quality Assessment. Synthesis Lectures on Image, Video, and Multimedia Processing*. Morgan & Claypool Publishers, 2006. Available from: <http://dx.doi.org/10.2200/S00010ED1V01Y200508IVM003>; <http://dx.doi.org/10.2200/S00010ED1V01Y200508IVM003>. xi, 16, 17
- [WBL03] Zhou Wang, Alan C. Bovik, and Ligang Lu. Foveated wavelet image quality index, September 07 2003. Available from: <http://citeseer.ist.psu.edu/658016.html>; <http://www.cns.nyu.edu/~zwang/files/papers/spie01.pdf>. 45
- [YB05] P. Yan and K. Bowyer. Ear biometrics using 2D and 3D images. In *Advanced 3D Imaging for Safety and Security*, pages III: 121–121, 2005. 13

Apêndice A

Anexos

A.1 UBEAR: A Dataset of Ear Images Captured On-the-move in Uncontrolled Conditions

Neste anexo inclui-se o artigo submetido, aceite e publicado na conferência CIBIM [cib].

UBEAR: A Dataset of Ear Images Captured On-the-move in Uncontrolled Conditions

Rui Raposo

University of Beira Interior
Covilhã, Portugal
m3642@ubi.pt

Edmundo Hoyle

University of Beira Interior
Covilhã, Portugal
edhoyle@ubi.pt

Adolfo Peixinho

University of Beira Interior
Covilhã, Portugal
m4067@ubi.pt

Hugo Proença

University of Beira Interior
IT-Instituto de Telecomunicações
Covilhã, Portugal
hugomcp@di.ubi.pt

Abstract—In order to broad the applicability of biometric systems, the data acquisition constraints required for reliable recognition are receiving increasing attention. For some of the traits (e.g., face and iris) significant research efforts were already made toward the development of systems able to operate in completely unconstrained conditions. For other traits (e.g., the ear) no similar efforts are known. The main purpose of this paper is to announce the availability of a new data set of ear images, which main distinguishing feature is that its images were acquired from on-the-move subjects, under varying lighting conditions and without demanding to subjects any particular care regarding ear occlusions and poses. The data set is freely available to the research community and should constitute a valuable tool in assessing the possibility of performing reliable ear biometric recognition in such d challenging conditions.

I. INTRODUCTION

Due to increasing concerns about safety and security in the modern societies, the use of biometric systems has been encouraged by both governmental and private entities to replace or improve effectiveness of the traditional human recognition systems. Several traits were already acknowledged as possessing the key features of a biometric trait: universality (ability to be collected in as much subjects as possible), collectibility (easiness in performing data acquisition), distinctiveness (high variability between different subjects) and stability (low variability over a single subject in human lifetime). For these traits (e.g., fingerprint, iris, face, retina, palm vein...) several recognition systems were already deployed and operate with remarkable success.

Among other traits that are still in embryonal development stages, the human ear is presently accepted as a promising biometric trait: two studies conducted by Iannarelli [1] provide substantial evidence of the distinctiveness of the ear biometric trait. The first study compared over 10 000 ears drawn from a randomly selected sample in California and the second examined fraternal and identical twins, in which the appearance of most physiological features appears to be similar. These studies support the hypothesis that each ear contains unique physiological features: all examined ears were found to be unique though identical twins were found to have similar ear structures. When compared to other biometric traits, the ear has several major advantages:

- its structure does not change over lifetime, from the birth into mature age;

- its surface is relatively small, allowing systems to cope with reduced spatial resolution images;
- it has a uniform color distribution;
- its appearance does not change according to different facial expressions.

Ears have played a significant role in forensic sciences for many years. In 1949, Iannarelli created his anthropometric identification technique based upon ear biometrics, based in twelve measurements illustrated in figure 1. The identification process relies in these twelve measures plus gender and ethnicity information. In order to support the development of automated ear recognition methods, several data sets were constructed and made publicly available, all of these containing images of relatively good quality acquired in high constrained conditions and environments.

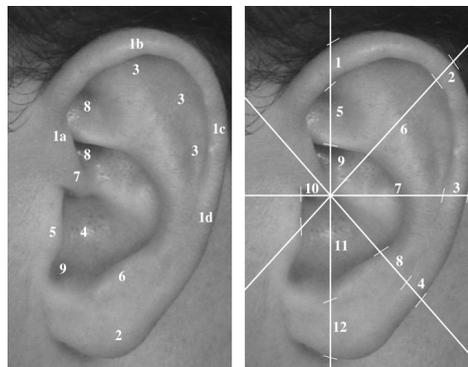


Fig. 1. Iannarelli system [1].

The main purpose of this paper is to announce the availability and to describe a new free available data set (UBEAR) of ear images, which major discriminating features were that data was collected from on-the-move subjects, under dynamic lighting conditions and without requiring to subjects and special care regarding the ears occlusions or poses. This facts turn the UBEAR into a preferable tool in evaluating the robustness of the currently developed ear recognition methods and in the research for new methods able to operate in uncontrolled conditions, toward the applicability of ear recognition systems in real-world scenarios.

The remaining of this paper is organized as follows: section II discusses the related works, section III overviews the existent data sets of ear images. A detailed description of the UBEAR data set, its imaging framework and statistical significance can be found in section IV. Section V describes our experiments and, finally, section VI concludes the paper.

II. RELATED WORK

It's possible to rearrange the proposed methods used in ear recognition, in 3 main categories.

A. 2D Images

- Burge and Burger [2] modeled each subject's ear as an adjacency graph built from the Voronoi diagram of its curve segments. They introduced a novel graph matching based algorithm which takes into account the erroneous curve segments likely to occur, and proposed the use of thermogram images to overcome the problem of hair occlusion.
- Moreno *et al.* [3] investigated the performance of various neural classifiers and combination techniques in ear recognition. The ear image is analyzed by 3 neural classifiers using outer ear feature points, ear morphology and macro features extracted by a compression network.
- Z. Mu *et al.* [4] used two feature vectors to ear recognition. The first vector is composed by features of the outer ear region, the second contains structural features of the inner region. A back propagation neural network is used as classifier.
- B. Arbab-Zavar and M. S. Nixon [5] used a log-Gabor filter to create a template of the ear, which was previously represented in polar coordinates, followed by an occlusion test. Previously, B. Arbab-Zavar *et al.* [6] used the Scale Invariant Feature Transform (SIFT), to detect ear features. They compared the performance of SIFT versus the Principal Components Analysis (PCA) method in the occlusion test.
- A. F. Abate *et al.* [7] proposed the use of a rotation invariant descriptor, the Generic Fourier Descriptor (GFD) [8], to extract meaningful data from ear images. The GFD is applied in the polar representation of the ears images.
- D. Hurley *et al.* [9] treated the ear image as an array of mutually attracting particles that act as source of a Gaussian force field. The original image is described by a set of potential channels and positions of potential wells.
- L. Yaun and Z. Mu [10] used an automatic ear normalization method based on improved Active Shape Model (ASM). The ear images are rotated so that all ears have the same rotation angle. Full-space Linear Discriminant Analysis (FLDA) is applied for ear recognition. Also, they found the acceptable head rotation range between 10 and 20 degrees, to right and left rotations respectively.

B. 3D Images

- Yan and Bowyer [11] used a Minolta VIVID 910¹ range scanner to capture both depth and color information. They developed a fully automatic ear biometric system using Iterative Closest Point (ICP) based in 3D shape matching for recognition, and used both 2D appearance and 3D depth data for automatic segmentation of the iris, also separating it from hair and earrings. They reported a decrease in matching performance where both ear images to be matched have more than 15 degrees of difference. Previously, they tested different approaches [12] and concluded that ICP based matching achieves the best performance.
- Chen and Bhanu [13] proposed to fuse range and color images to detect ears, and both global and local features for the extraction of the meaningful information. Also, they used the ear helix/anti-helix and a LSP (Local Surface Patch) representation for estimating the initial rotation and translation between a gallery/probe pair and then the modified ICP algorithm to compensate the distortion.

C. Acoustic

- Akkermans *et al.* [14] exploited the acoustic properties of the ear for recognition purposes. It turns out that the ear by virtue of its special shape behaves like a filter so that a sound signal played into the ear is returned in a modified form. This acoustic transfer function forms the basis of the acoustic ear signature.

III. RELATED DATASETS

Three different datasets are most widely used in the evaluation of ear recognition proposals: the UND² (University Notre Dame), the XM2VTS³ (the extended M2VTS Database) and the USTB⁴ (University of Science and Technology Beijing). Other datasets were developed or referred in academic research, but are not publicly available and not constitute the scope of this section. In the following we briefly describe the main features of its dataset, together with a set of example images shown in figure 3.

A. UND

The University of Notre Dame supplies four collections of ear datasets:

- CollectionE: 464 visible-light profile ear images from 114 human subjects;
- CollectionF: 942 3D ,plus corresponding 2D profile ear images from 302 human subjects;
- CollectionG: 738 3D, plus corresponding 2D profile ear images from 235 human subjects;
- CollectionJ2: 1800 3D, plus corresponding 2D profile ear images from 415 human subjects.

¹<http://www.konicaminolta.com/instruments/products/3d/non-contact/vivid910/index.html>

²[http://www.nd.edu/\\$\sim\\$scvrl/CVRL/Data/_Sets.html](http://www.nd.edu/\simscvrl/CVRL/Data/_Sets.html)

³<http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/>

⁴<http://www.ustb.edu.cn/resb/>

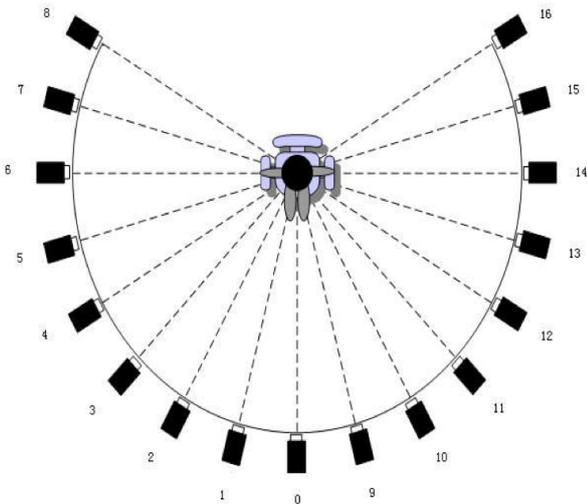


Fig. 2. The platform of the camera system from USTB⁴ dataset.

B. XM2VTS

The University of Surrey supplies several collections of image, audio and video datasets. For ear recognition purposes, the focus should be putted in the head image dataset, described as : 1 left and 1 right profile image (720x576) per person and session, for a total of 2 360 images.

C. USTB

The University of Science and Technology of Beijing (USTB) supplies four datasets of ears, with multi-pose and angles data faces:

- Dataset I - 60 subjects, 3 images each from the right ear, with some of the ears experiencing some shearing and rotation;
- Dataset II - 77 subjects, 4 images per subject. The distance between subject and camera is about 2 meters with variations in terms of illumination and angles. 2 images for different lighting setups and the remaining for pose variations, with rotations of -30 degrees and +30 degrees. Each image is 24-bit true color with 300x400 pixels.
- Dataset III - this dataset is divided into 2 sub sets: the first contains 79 subjects with right and left rotation. The second includes 24 subjects, and each one has 6 images with different ranges of occlusion. The images are 768x576 pixels, 24-bit true color.
- Dataset IV - The capture process consists of 17 CCD cameras which are distributed in a circle with radius being 1 meter and the subject is placed in the center, as illustrated in figure 2. The volunteers (500 in total) were required to look eye level, look upwards, look downwards, look right and look left for the photograph.

IV. UBEAR DATASET

As above stated, our fundamental purpose was to unconstraint the image acquisition scenario, so that images ap-

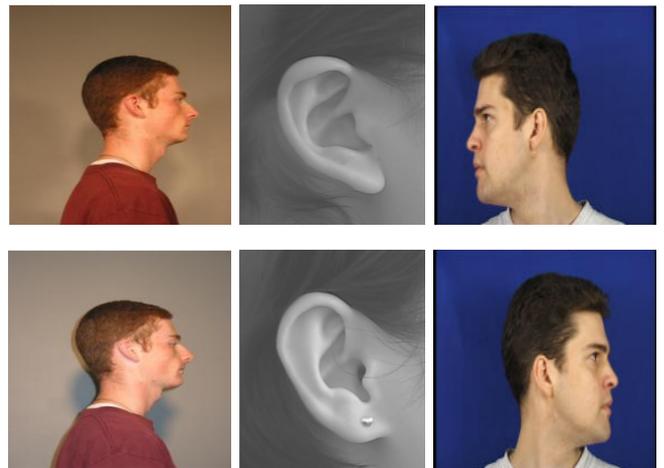


Fig. 3. UND dataset in first column, the second column show USTB dataset and the last column XM2VTS dataset.

pear to be captured in *real-world* conditions, i.e., with subjects on-the-move and without requiring them any particular care about occlusions of the ears and poses. Also, the lighting conditions were highly varying between different sessions so that the typical imagers from each session have notorious differences. This peculiar setup were devised to simulate the covert acquisition of biometric data.

A. Imaging Framework

The setup of the video and imaging framework is given in table I. Samples of the collected images and its corresponding binary ear masks (manually made) are illustrated in figure 4. The video capture sequence starts with all subjects facing front and 3 meters apart from the camera sideways, it's also required of the individual to move his head upwards, downwards, outwards, towards. After these, subjects should step ahead and backwards from the initial position. For each subject both ears were captured in two different sessions, giving a total of four videos per subject. Each video sequence was manually analyzed and 17 frames were selected according to the following criteria:

- 5 frames when the subject was stepping ahead and backwards,
- 3 frames of the subject's head moving upwards,
- 3 frames of the subject's head moving downwards,
- 3 frames of the subject's head moving outwards,
- 3 frames of the subject's head moving towards.

B. Data Variability and Statistics

As illustrated in figure 5, there are three major varying factors in the UBEAR images: 1) lighting variations, either by natural or artificial light; 2) multiple head poses either with yaw and pitch rotations and 3) occlusions due to hair and earrings. Apart from these, images of the UBEAR data set have significantly heterogenous levels of image quality, which can be useful to evaluate the algorithms robustness to changes

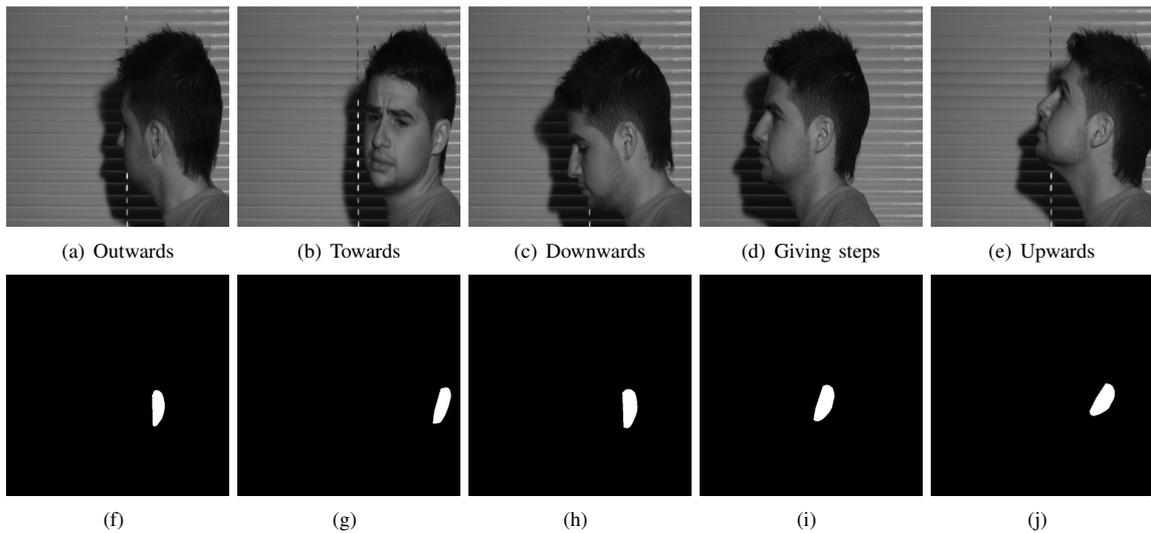


Fig. 4. Images from the UBEAR dataset and this corresponding binary ear mask.

TABLE I
VIDEO AND IMAGE FRAMEWORK

Video Acquisition Framework and Set-Up	
Camera	Stingray F-504B 2/3"
Focal length	35mm
Color Representation	gray scale
Video Resolution	1280x960 pixels
Frames per second	15
Videos Codec	Avi uncompressed
Details of the selected frames	
Image Resolution	1280x960 pixels
Color Representation	gray scale
Image Codec	tiff
Volunteers	
Totals = Subjects 126;	Gender = Male: 44.62%
Ears 252; Images 4430	Female: 55.38%
Age = [0, 20] 43.85%	
[21, 25] 47.69%	
[26, 30] 3.85%	
[31, 35] 2.31%	
[36, 99] 2.30%	

in data quality and to assess the actual deterioration in the corresponding error rates.

Figure 6 gives five histograms that describe some of the major features of the ears regions contained in our dataset. The area histogram (figure 6a) gives the values that correspond to the sum of all pixels inside the ear region, in order to give an idea of the ears size. The eccentricity histogram (figure 6b) describes the proportion between the major axis of the ears. For almost round ears, as in figure 7e, it will return a value near to 0, while for more suchlike elliptical ears (figure 7d), the value will approach 1. The length histogram (figure 6c), exhibits the length of the major axis of the visible part of the ears, as illustrated in figure 7. The orientation histogram (figure 6d) gives the typical orientation of the major axis of the ears, that can be high varying, as illustrated in figures 7(b) and (c). Finally, the perimeter histogram (figure 6e), corresponds to the sum of all pixels in the ear boundaries, which can be

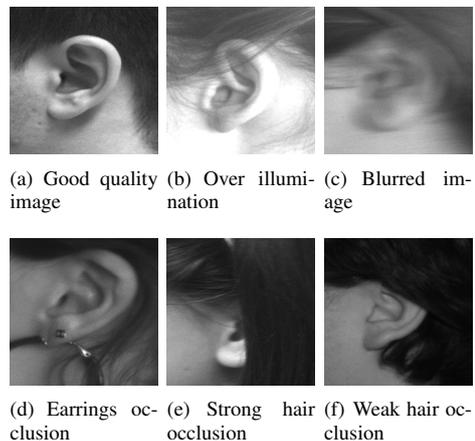


Fig. 5. Comparison between a good quality image and several types of non-ideal images of the UBEAR dataset.

useful for recognition purposes based in boundary descriptors methods.

C. Statistical Significance of the UBEAR dataset

In this section we address the problem of whether the experiments performed on the UBEAR dataset produce statistically significant results. Let α be the confidence interval. Let P be the error rate of a classifier and \hat{P} be the estimated error rate over a finite number of test patterns. At an α -confidence level, we want that the true error rate not exceeds the estimated error rate by an amount larger than $\varepsilon(N, \alpha)$. Guyon et al. [15] fixed $\varepsilon(N, \alpha) = \beta P$ to be a given fraction of P . Assuming that recognition errors are Bernoulli trials, authors concluded that the number of required trials N to achieve $(1 - \alpha)$ confidence in the error rate estimate is $N = -\ln(\alpha)/(\beta^2 P)$. A typical value for α is 0.05 and a typical value for β is 0.2. Based on these values, Guyon et al. [15] recommended the simpler form

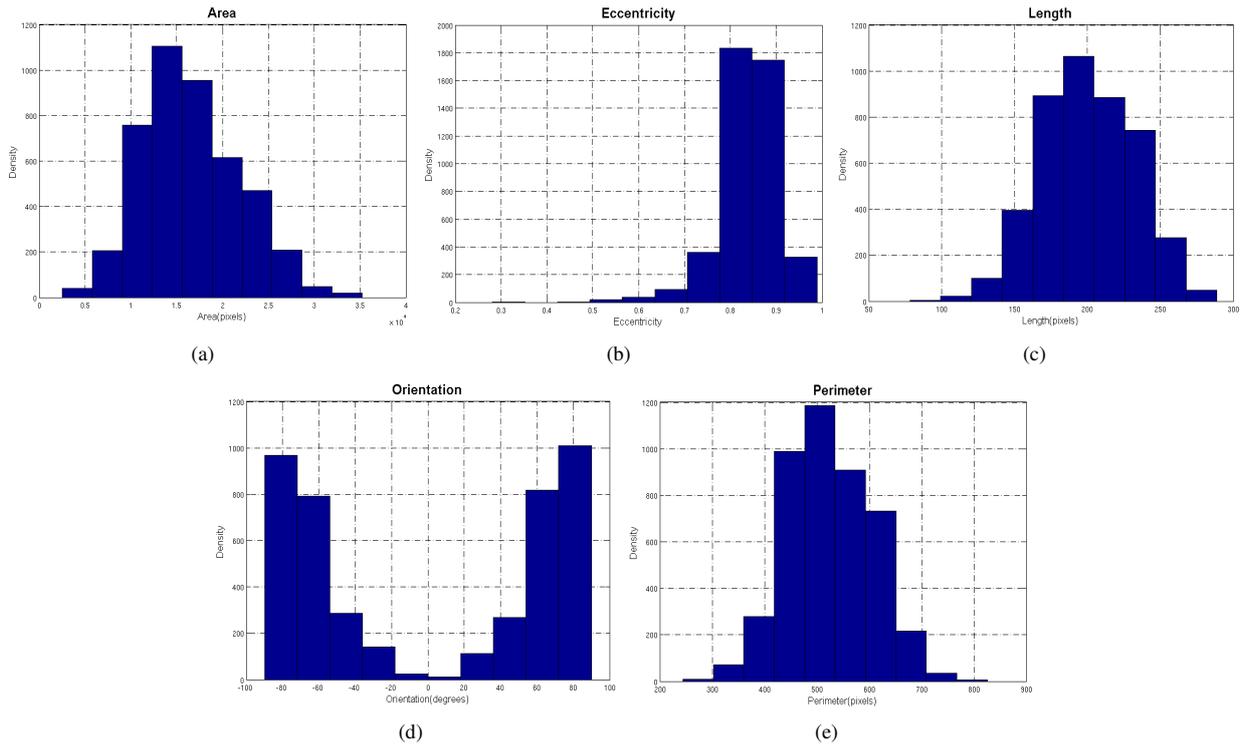


Fig. 6. Statistics of the UBEAR dataset images.

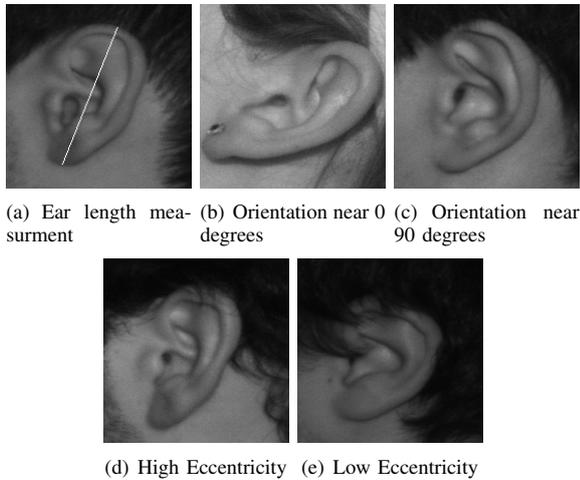


Fig. 7. Some cases that illustrate the histogram extreme cases.

$N \approx \frac{100}{P}$. We had a varying number of subjects that offered as volunteers for the first, second and for both imaging sessions. However, assuming that each iris image can be used to generate a biometric template, that the remaining images from the same eye can be used to analyze intra-class variability and the remaining images from different eyes can be used to analyze inter-class variability, it is possible to obtain a bound for the error that is possible to be tested with statistical significance. The 4 429 images of the UBEAR dataset enable respectively

41 278 and 9 764 528 intra-class and inter-class comparisons. This guarantees statistical significance in experiments with an empirical error rate \hat{P} down to $1,02 \times 10^{-5}\%$. However, we stress that this is a lower bound value that would be increased if we do not assume the independence between images and error correlations are taken into account.

V. EXPERIMENTS

The main purpose of this section is to describe the deteriorations in performance of typical ear recognition methods due to factors that degrade the quality of the acquired data. Based in our earliest experiments, we observed that current methods are particularly sensitive to rotations of the ears, which became the focus of our further experiments. To avoid that other factors bias our results, we selected a sub set of good quality images: those that are sharp, without significant occlusions. Then, data was divided into 5 subsets: 624 images without significant ear rotations (subset 1), 358 images with ears rotated upwards (subset 2), 332 images with ears rotated downwards (subset 3), 309 images with ears rotated outwards (subset 4) and 343 images with ears rotated towards (subset 5). Figure 4 illustrates the four types of rotations that were the main criterium for the division of the data sets.

- Aligned: we compare all images of subset 1,
- Aligned-Upwards: Each image of subset 1 is compared with all images from subset 2,
- Aligned-Downwards: Each image of subset 1 is compared with all images from subset 3,

- Aligned-Outwards : Each image of subset 1 is compared with all images from subset 4,
- Aligned-Towards: Each image of subset 1 is compared with all images from subset 5.

A. SIFT

The Scale Invariant Feature Transform (SIFT) is one of the most popular descriptors for image point matching [16]. The SIFT is known to be invariant to image point scale and rotation and robust to affine distortion, changes in 3D viewpoint, addition of noise and changes in illumination. It's application domain has been extended to human identification and the results are quite promising [17]. Here, keypoints are represented by vectors indicating scale, orientation and location. The keypoints location is refined by fitting it to nearby data and one or more orientations can be assigned using local image gradient directions for each keypoint [17]. The feature descriptor is computed by accumulating the orientation histograms on the 4x4 subregions. Each histogram has 8 bins, thus the SIFT feature descriptor has 128 elements. Finally, the feature vector is normalized to reduce the effects of illumination change [17]. The ratio between the distance of the closest neighbor and the second-closest neighbor, is used in the search for corresponding matching points [17] for recognition strategies. In this paper, we used the D. Lowe's implementation of the SIFT operator⁵.

B. Results

In order to avoid that segmentation errors carry some bias to the obtained results, we manually segmented all the images used in this experiment, producing a binary segmentation mask that distinguishes between the noise-free regions of the ear and all the remaining types of data in the image, as it is illustrated in figure 4. Thus, the SIFT method was applied exclusively to the regions that comprise the un-occluded ears. As performance measures, we elected the well known receiver operating characteristic curves (ROC), the area under curve (AUC) and the equal error rate (EER) and the decidability [18] index, given by $\frac{|\mu_I - \mu_E|}{\sqrt{0.5 \sigma_I^2 + 0.5 \sigma_E^2}}$, where μ_I and μ_E denote the means of the intra-class and inter-class observations and σ_I and σ_E the corresponding 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 positives examples are ranked higher than the negatives ones and the area equal to 1. Any deviation from this ranking decreases the AUC. The EER of a verification system, when the operating threshold for the accept/reject decision is adjusted so that the probability of false acceptance and false rejection becomes equal.

The obtained ROC curves are illustrated in figure 9, and the decidability, EER and AUC results are given in table II. The test 1 obtained the best performance, as shown by the ROC curve, with higher decidability, a lower EER and

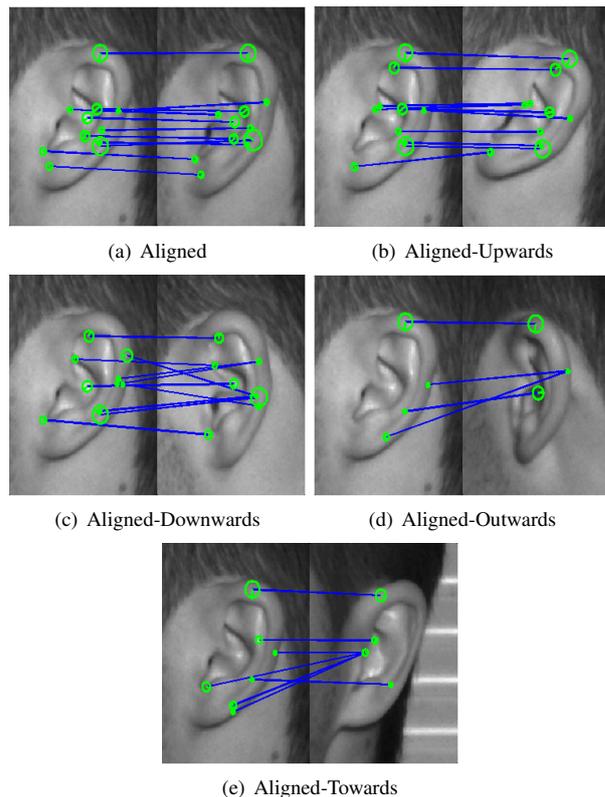


Fig. 8. These images show the matching keypoints according to the used SIFT encoding matching strategies.

higher AUC. Obviously, this was a priori expected, as all the matched images do not have significant differences in pose and SIFT maximally correlates them, as can be seen in figure 8(a). Oppositely, the worst results was obtained in tests 4 and 5, which from our viewpoint can be explained by the fact that significant toward and outward rotations alter the perception of the ear shape, essentially because the ear is far from planar and such rotations lead in some cases to occlusions of portions of the ears. The other types of rotations (upwards and downwards) didn't significantly deteriorate the results, which is in agreement to our perception of the iris structure according to these rotations.

Figure 8 illustrates the insights of the obtained results, giving the comparisons between the key points that were typically matched in each test. It can be confirmed that maximal correlation was obtained when both ears were aligned with the camera (figure a). Upward and backward rotations didn't significantly changed the number of matched key points (figures b and c), in opposition with towards and outwards rotations (figures d and e). From these experiments, we concluded that significant research efforts are required to compensate for different poses of the subjects, which will be a crucial step for the development of ear recognition methods able to operate under less constrained conditions. Hopefully, the UBEAR dataset will contribute for such achievement.

⁵<http://www.cs.ubc.ca/lowe/keypoints/>

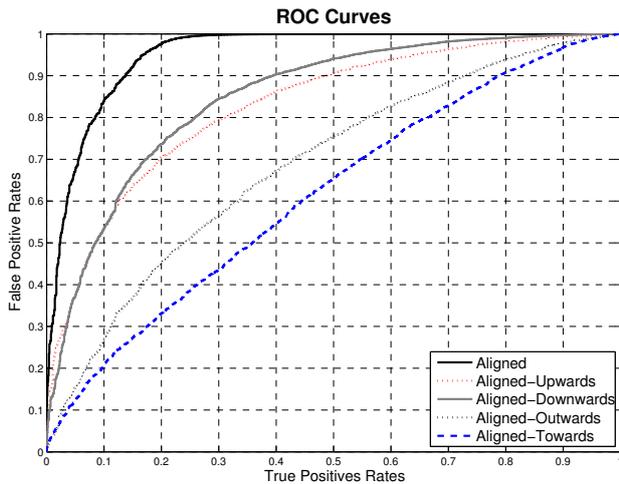


Fig. 9. ROC curves of evaluated tests.

TABLE II
RECOGNITION RATES OF EACH TEST.

	DEC	EER(%)	AUC
Aligned	2.64	12.20	0.95
Aligned-Upwards	1.32	24.99	0.83
Aligned-Downwards	1.43	22.86	0.85
Aligned-Outwards	0.68	36.24	0.69
Aligned-Towards	0.44	42.36	0.62

VI. CONCLUSIONS AND DATA SET AVAILABILITY

In this paper we presented a new dataset of ear images for biometric purposes, which major discriminating point is that it simulates the acquisition of data in real-world scenarios, under varying lighting conditions on moving subjects and without requiring them any particular care regarding ear occlusions and poses. Our experiments show that — as it will be expected — the performance of the most well known ear recognition methods significantly decreases according to the quality of the acquired data. Hence, we hope that the UBEAR dataset constitute as a valuable tool for the research of ear recognition systems more robust to degraded data, namely due to the dynamics lighting recognition, subject movements and perspective. Finally, the value given to the UBEAR dataset should have be directly correspondent to the number of persons that use it in their experiments. Thus, we decided to make UBEAR public and freely available through the UBEAR datasets web site: <http://www.ubear.di.ubi.pt>.

ACKNOWLEDGMENTS

We acknowledge all volunteers that agreed to participate in the imaging sessions of the UBEAR data set. Also, the financial support given by "FCT-Fundação para a Ciência e Tecnologia" and "FEDER" in the scope of the PTDC/EIA/103945/2008 "NECOVID: Negative Covert Biometric Recognition" and

PTDC/EIA/69106/2006 "BIOREC: Non-Cooperative Biometric Recognition" research projects is acknowledged too.

REFERENCES

- [1] A. V. Iannarelli, *Ear Identification (Forensic Identification Series)*, 1st ed. Fremont, California: Paramount Publishing Company, 1989.
- [2] M. Burge and W. Burger, "Ear biometrics in computer vision," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 2, 2000, pp. 822–826 vol.2.
- [3] B. Moreno, A. Sanchez, and J. Velez, "On the use of outer ear images for personal identification in security applications," in *Security Technology, 1999. Proceedings. IEEE 33rd Annual 1999 International Carnahan Conference on*, 1999, pp. 469–476.
- [4] Z. Mu, L. Yuan, Z. Xu, D. Xi, and S. Qi, "Shape and structural feature based ear recognition," in *Advances in Biometric Person Authentication, 5th Chinese Conference on Biometric Recognition, SINOBOMETRICS 2004, Guangzhou, China, December 13-14, 2004, Proceedings*, vol. 3338. Springer, 2004, pp. 663–670.
- [5] B. Arbab-Zavar and M. Nixon, "Robust log-gabor filter for ear biometrics," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, 8-11 2008, pp. 1–4.
- [6] B. Arbab-Zavar, M. Nixon, and D. Hurley, "On model-based analysis of ear biometrics," in *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, 27-29 2007, pp. 1–5.
- [7] A. Abate, M. Nappi, D. Riccio, and S. Ricciardi, "Ear recognition by means of a rotation invariant descriptor," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 4, 2006, pp. 437–440.
- [8] D. Zhang and G. Lu, "Shape-based image retrieval using generic fourier descriptor," *Signal Processing: Image Communication*, vol. 17, no. 10, pp. 825–848, 2002.
- [9] D. J. Hurley, M. S. Nixon, and J. N. Carter, "Force field feature extraction for ear biometrics," *Computer Vision and Image Understanding: CVIU*, vol. 98, no. 3, pp. 491–512, Jun. 2005.
- [10] L. Yuan and Z. C. Mu, "Ear recognition based on 2D images," in *Biometrics: Theory, Applications, and Systems, 2007*, pp. 1–5.
- [11] P. Yan and K. W. Bowyer, "Biometric recognition using 3D ear shape," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1297–1308, Aug. 2007.
- [12] P. Yan and K. Bowyer, "Ear biometrics using 2D and 3D images," in *Advanced 3D Imaging for Safety and Security, 2005*, pp. III: 121–121.
- [13] H. Chen and B. Bhanu, "Human ear recognition in 3d," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 4, pp. 718–737, Apr. 2007.
- [14] A. H. M. Akkermans, T. A. M. Kevenaar, and D. W. E. Schobben, "Acoustic ear recognition," in *Biometric Authentication, 2006*, pp. 697–705.
- [15] R. S. I. Guyon, J. Makhoul, and V. Vapnik, "What size test set gives good error rate estimates?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 52–64, Feb. 1998.
- [16] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [17] K. Mikołajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
- [18] J. Daugman and G. Williams, "A proposed standard for biometric decidability," in *Proceedings of the CardTech/SecureTech Conference*, pp. 223–234, 1996.

