



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

Hugo Proença*, Gil Santos

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

ARTICLE INFO

Article history:

Received 13 May 2011

Accepted 24 October 2011

Available online 2 November 2011

Keywords:

Iris recognition

Biometrics

Security and surveillance

Forensics

Visible-light data

ABSTRACT

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

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

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

The iris texture is characterized by the irregular distribution of local features such as furrows, crypts, freckles or spots. Hence, traditional approaches are texture-based and decompose the iris data into a set of numerical coefficients that correspond to specific orientations or frequencies of predominant patterns. The recognition in less controlled environments has been gaining relevance and was the focus of many recent proposals, among which the “Iris-on-the-move” project [1] should be highlighted: it is a major example of engineering an image acquisition system to make the

recognition process less intrusive for subjects. The goal is to acquire near infra-red close-up iris images as a subject walks at normal speed through an access control point. *Honeywell Technologies* applied for a patent [2] on a very similar system, which was also able to recognize irises at a distance. Previously, Fancourt et al. [3] concluded that it is possible to acquire sufficiently high-quality images at a distance of up to 10 m.

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

* Corresponding author.

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

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

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

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

2. Proposed method

A cohesive perspective of the proposed recognition strategy is given in Fig. 1. A color constancy technique is used for regularization purposes, and data are normalized into a Polar coordinate system of constant dimensions, from which global MPEG.7 color descriptors are extracted. Next, a self-organizing data technique

divides the noise-free iris data into spatially and color coherent regions that feed the local color and shape MPEG.7 descriptors. Finally, fusion is performed by means of a weighted sum rule.

2.1. Retinex

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

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

being δ_k given by

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

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

2.2. Noise-free iris segmentation

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

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

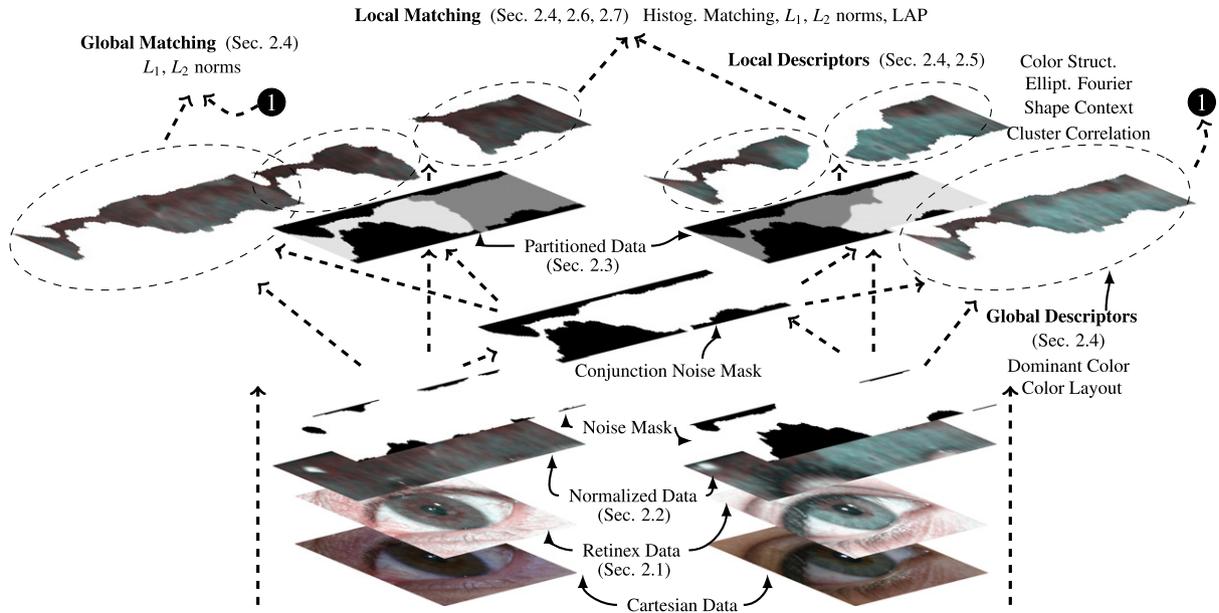


Fig. 1. Cohesive perspective of the proposed method, describing its major phases and the corresponding sections of this paper.



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

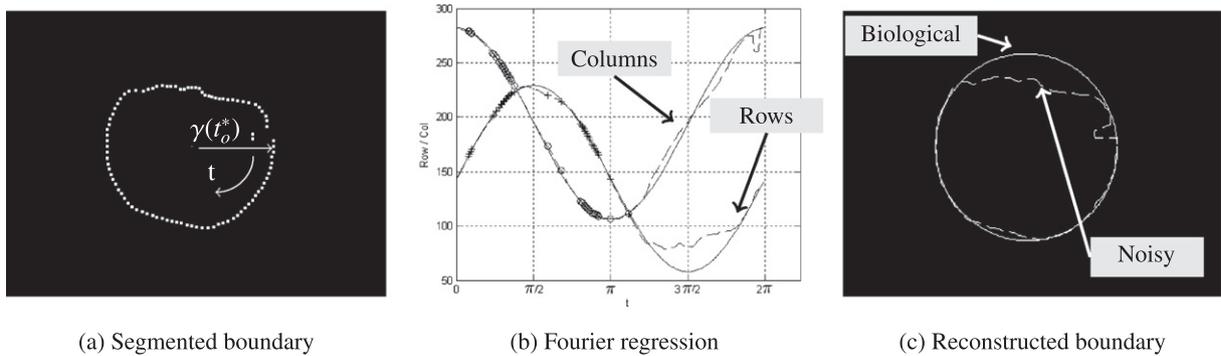


Fig. 3. Parameterization of the biological iris boundaries. According to the values of (3), smoother regions with low energy of the cumulative angular descriptor (a) were deemed to belong to the biological boundaries (cross and circular data points of b) and used to reconstruct the deemed biological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

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

$$\gamma(t) = \int_0^{t^*} k(r)dr - k(0) + t \quad (2)$$

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

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

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

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

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

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

2.3. Partitioning the iris into regions

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

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

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

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

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

Fig. 4 illustrates how the typical appearance of the generated clusters would vary with respect to different ω_i values. Black pixels denote regions that were classified as *noisy* by the segmentation method and, as such, were not considered in the clustering process. The remaining intensities represent the clusters assigned to each pixel of the normalized iris data when privileging the (ω_1, ω_2) weights (associated with spatial features) and the $(\omega_3, \omega_4, \omega_5)$ (associated with color features) (bottom left image). The image at the bottom illustrates clusters generated for the *optimal* weight values $\hat{\omega}_i$, which constitutes a trade-off between space and color, as follows:

$$\hat{\omega}_i = \arg \min_{\omega_i} \sum_k \sum_j \alpha \phi_k(\omega_i, \{C_{k,j}\}) + (L_{k,j}^2 - 4\pi A_{k,j}), \quad (7)$$

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

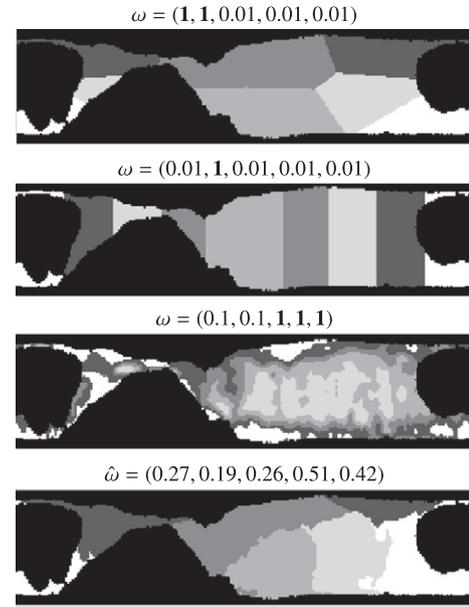


Fig. 4. Comparison between the regions resulting from different clustering processes with respect to weights given to each feature. The image at the top resulted from large weights (denoted by bold font) for spatial features, whereas in the case of the second image at the top, a large weight was given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

2.4. Color descriptors

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

2.4.1. Dominant color descriptor

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

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

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

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

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

¹ <http://www.csse.uwa.edu.au/du/Software/graphics/xyz2lab.m>

2.4.2. Color layout descriptor

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

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

Finally, the $X_{\{i\}}$ coefficients were zigzag ordered [23], resulting in a vector \vec{v} of $(r \times c)/n^2$ real components. Matching between vectors \vec{v}_1 and \vec{v}_2 was performed according to the L_2 norm of the vector $\vec{v}_1 - \vec{v}_2 = \sqrt{\sum_{i=1}^{(r \times c)/n^2} (v_1(i) - v_2(i))^2}$.

2.4.3. Color structure descriptor

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

2.4.4. Scalable color descriptor

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

2.5. Order statistics of dominant colors

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

$$d^*(\vec{c}_i, \vec{c}_j) = |d(\vec{c}_i, \vec{c}) - d(\vec{c}_j, \vec{c})|.$$

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

$$\vec{c}_i \succ \vec{c}_j \iff d(\vec{c}_i, \vec{c}) > d(\vec{c}_j, \vec{c}) \quad (10)$$

. The rationale behind \succ is to consider as *greater* colors those that are more distant from the dominant color of the iris. From this definition, it is straightforward to infer that \succ is irreflexive, asymmetric and transitive, which is particularly useful for our purposes. Let $X = \{\vec{x}_1, \dots, \vec{x}_n\}$ be a random variable that represents the distance between the dominant colors inside each cluster and the whole iris.

According to the elementary theory of rank tests and using (10), the k th order statistic $\vec{x}_{(k)}$ of a statistical sample $\{\vec{x}_1, \dots, \vec{x}_n\}$ is equal to its k th smallest value. Let $\vec{x}_{(1)}, \dots, \vec{x}_{(n)}$ be the order statistics of a set of independent observations, that is, $\vec{x}_{(1)} < \vec{x}_{(2)} < \dots < \vec{x}_{(n)}$. Assuming that \vec{x}_i is mutually independent, the distribution function of $\vec{x}_{(k)}$ is equal to [24]:

$$F_{(k)}(y) = P(\vec{x}_{(k)} \leq y) = \sum_{i=k}^n \binom{n}{i} [F(y)]^i [1 - F(y)]^{n-i} \quad (11)$$

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

2.6. Linear assignment problem

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

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

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

2.7. Histogram matching

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

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

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

2.8. Shape context descriptor

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

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

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

2.9. Robustness to data variation factors

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

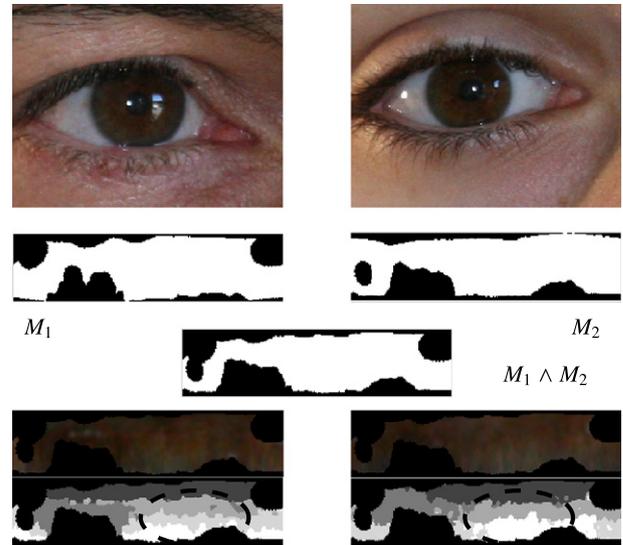


Fig. 6. Clusters generated for two different heavy pigmented irises, where local contrast inside the iris ring is hardly perceived by a human observer. Even so, the appearance of the resulting group is evidently different. Images are “C_1_S1_I12.tiff” and “C_101_S1_I10.tiff” of the UBIRIS.v2 data set.

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

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

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

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

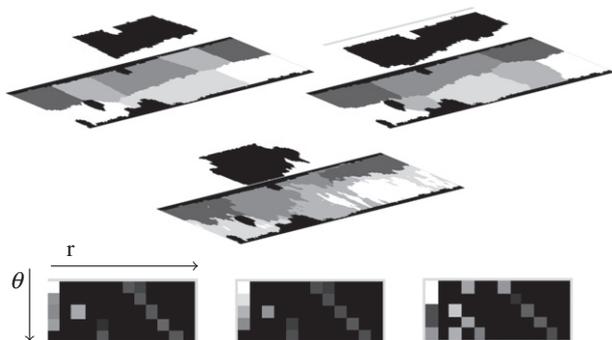


Fig. 5. Illustration of the shape descriptor used to characterize each iris region. The upper row shows two similar shapes, from which shape context descriptors were extracted. The image at the center has a significantly different shape. Images in the bottom row illustrate the corresponding shape descriptors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

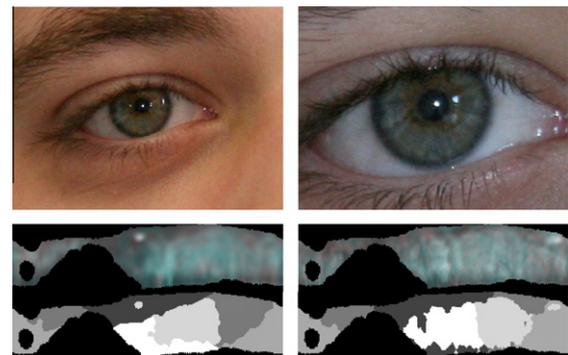


Fig. 7. Robustness to changes in scale. Images are “C_111_S1_I4.tiff” and “C_111_S1_I13.tiff” of the UBIRIS.v2 data set.

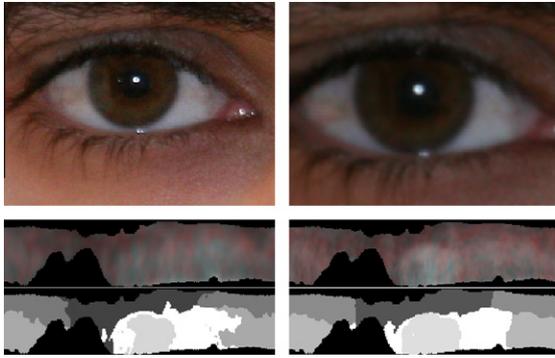


Fig. 8. Robustness to defocused data. Images are “C_183_S2_I10.tiff” and “C_183_S2_I13.tiff” (defocused by a Gaussian kernel of $\sigma = 1.4$) of the UBIRIS.v2 data set.

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

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

Off-angle images are of special interest because gaze is known to be a primary source of error in traditional recognition strategies, particularly when circular iris parameterization techniques introduce differences in the phase of the normalized data and the bias phase-based in encoding/matching methods. The translation into the polar coordinate system implies that the data are sampled at different rates with respect to the length of the iris ring at each angle ($a \gg b$ in the right image of Fig. 10, but $a \approx b$ in the left image), which does not significantly affect the color perception of the resultant data. This relationship was observed even in cases where exaggerated deviations occlude portions of the iris. Furthermore, this figure gives a typical failure situation motivated by iris seg-

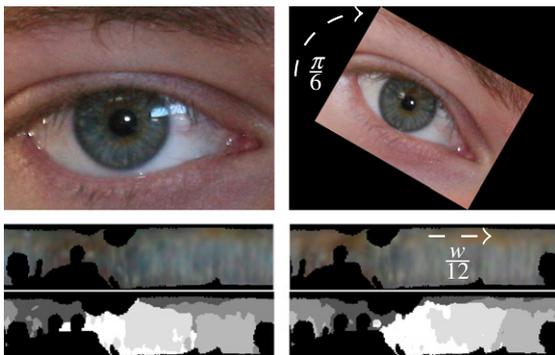


Fig. 9. Robustness to changes in rotation. Images are “C_171_S1_I10.tiff” and “C_171_S2_I10.tiff” (rotated by $\frac{\pi}{6}$) of the UBIRIS.v2 data set.

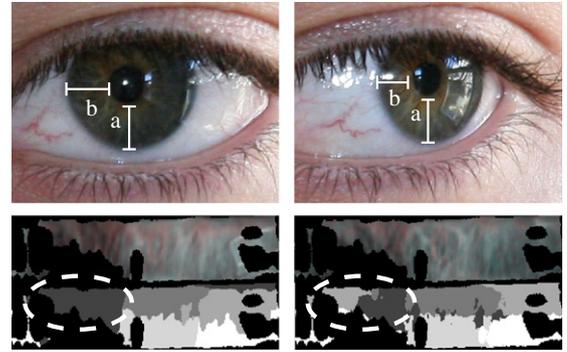


Fig. 10. Robustness to off-angle image acquisition. Images are “C_24_S1_I13.tiff” and “C_24_S1_I15.tiff” of the UBIRIS.v2 data set.

mentation inaccuracies: the region delimited by the dashed ellipse in the right figure should have been classified as noise (corresponds to the upper part of the iris, partially occluded by eyelashes) but was erroneously considered for the clustering process and induced substantial differences in the resultant clusters near that region.

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

3. Experiments

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

3.1. Feature selection

In a training set of 1000 images used by the participants of the NICE:II contest (available at²), the discriminating ability of a large set of features was assessed, testing different values for the number of clusters (between two and seven) and for the most relevant parameters of the described encoding strategies, yielding a total of 112 features. Fig. 12 gives the probability density functions and the corresponding cumulative density functions of the ten most discriminating features, selected based on mutual information and the

² <http://nice2.di.ubi.pt>

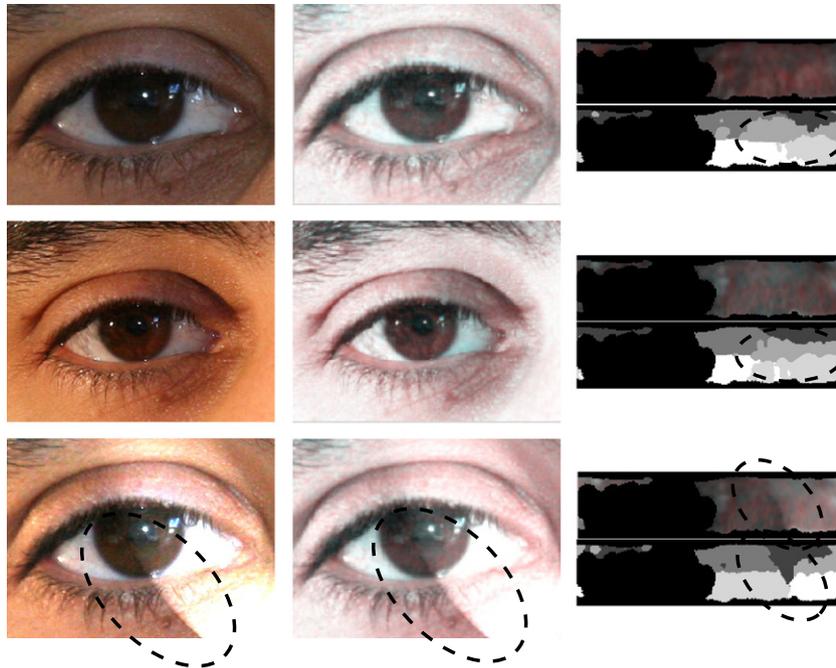


Fig. 11. Robustness to global and nonuniform lighting changes. Images are “C_137_S1_110.tiff” (top) and “C_137_S1_17.tiff” (middle) and “C_137_S1_110.tiff” (bottom, with a directional artificial light effect) of the UBIRIS.v2 data set.

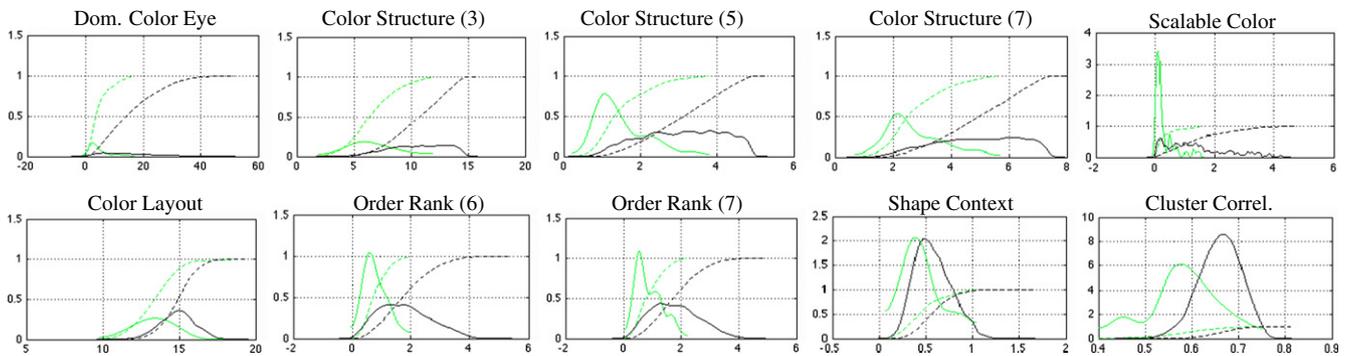


Fig. 12. Probability density (continuous lines) and cumulative density functions (dashed lines) of the features selected for the biometric recognition process. The *non-match* comparisons are represented by darker lines and the *match* comparisons by lighter lines.

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

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

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

3.2. Verification mode

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

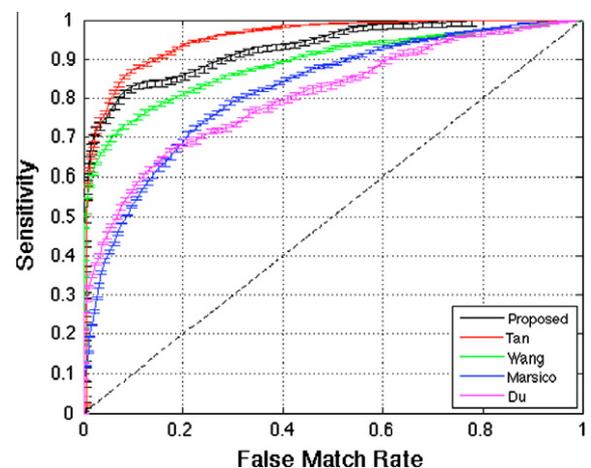


Fig. 13. Comparison between the receiver operating characteristic curves of the proposed methods and others used for contextualization purposes.

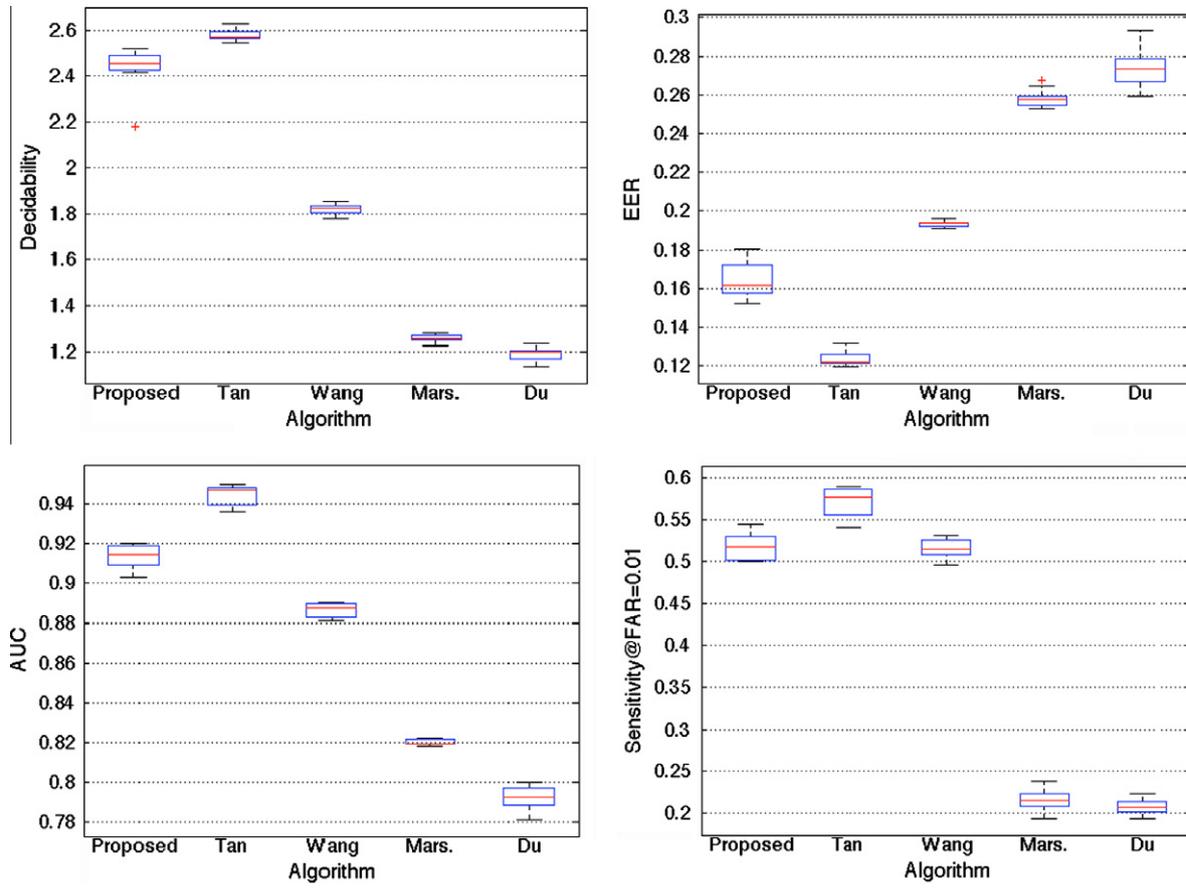


Fig. 14. Comparison between the performance measures obtained by the tested recognition strategies. Each column represents one recognition approach, where the corresponding median value is represented by the horizontal line through the middle of each box. The top and bottom of the boxes denote the first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

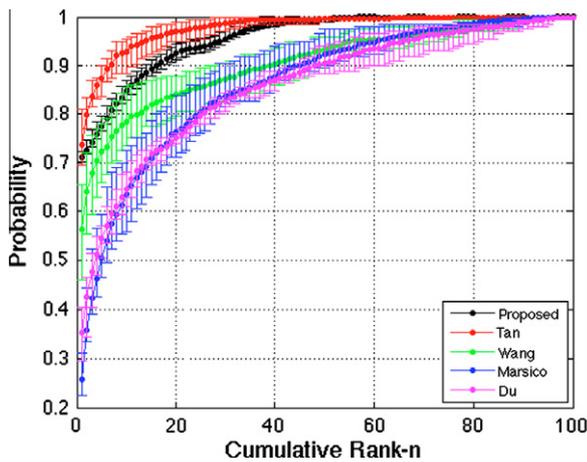


Fig. 15. Average cumulative rank n curves obtained by the proposed method and others used as comparison terms. The bottom and top horizontal lines around the data series denote the worst and best values obtained.

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

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

Table 1

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

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

decidability index (d') given by $d' = \frac{|\mu_E - \mu_I|}{\sqrt{\frac{\sigma_I^2 + \sigma_E^2}{2}}}$, where $\mu_I = \sum_k d_k^I$ and

$\mu_E = \sum_m d_m^E$ are the means of the match/non-match distributions, and $\sigma_I = \sqrt{\frac{\sum_i (d_i^I - \mu_I)^2}{k-1}}$ and $\sigma_E = \sqrt{\frac{\sum_i (d_i^E - \mu_E)^2}{m-1}}$ are their standard deviations.

The approximated equal error rate (EER), the area under the curve (AUC) and the average sensitivity when operating at a FAR of approximately 0.01 were also obtained. The results are expressed in terms of boxplots, showing the median of the observed performance range (horizontal solid line) and the first and third quartile values of the observations (top and bottom of the box marks). The upper and lower whiskers are denoted by the horizontal lines outside each box, and the outliers are denoted by dot points.

3.3. Identification mode

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

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

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

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

3.4. Correlation and fusion

The statistical correlation between the outputs given by our method and others used as comparison terms was analyzed to

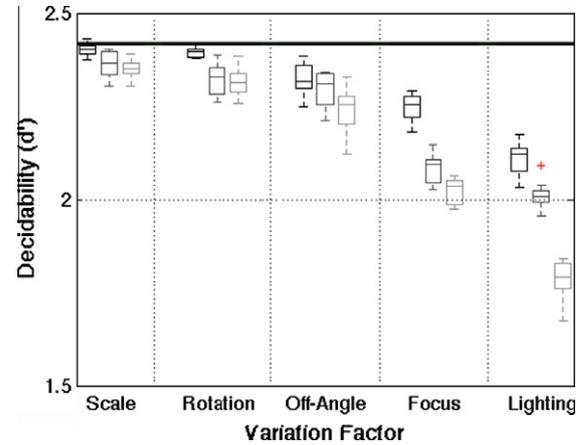


Fig. 17. Degradation in recognition performance, expressed in terms of the decidability values, with respect to variations in different factors. The results are expressed in terms of boxplots, showing the median of the observed performance range (horizontal solid line) and the first and third quartile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

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

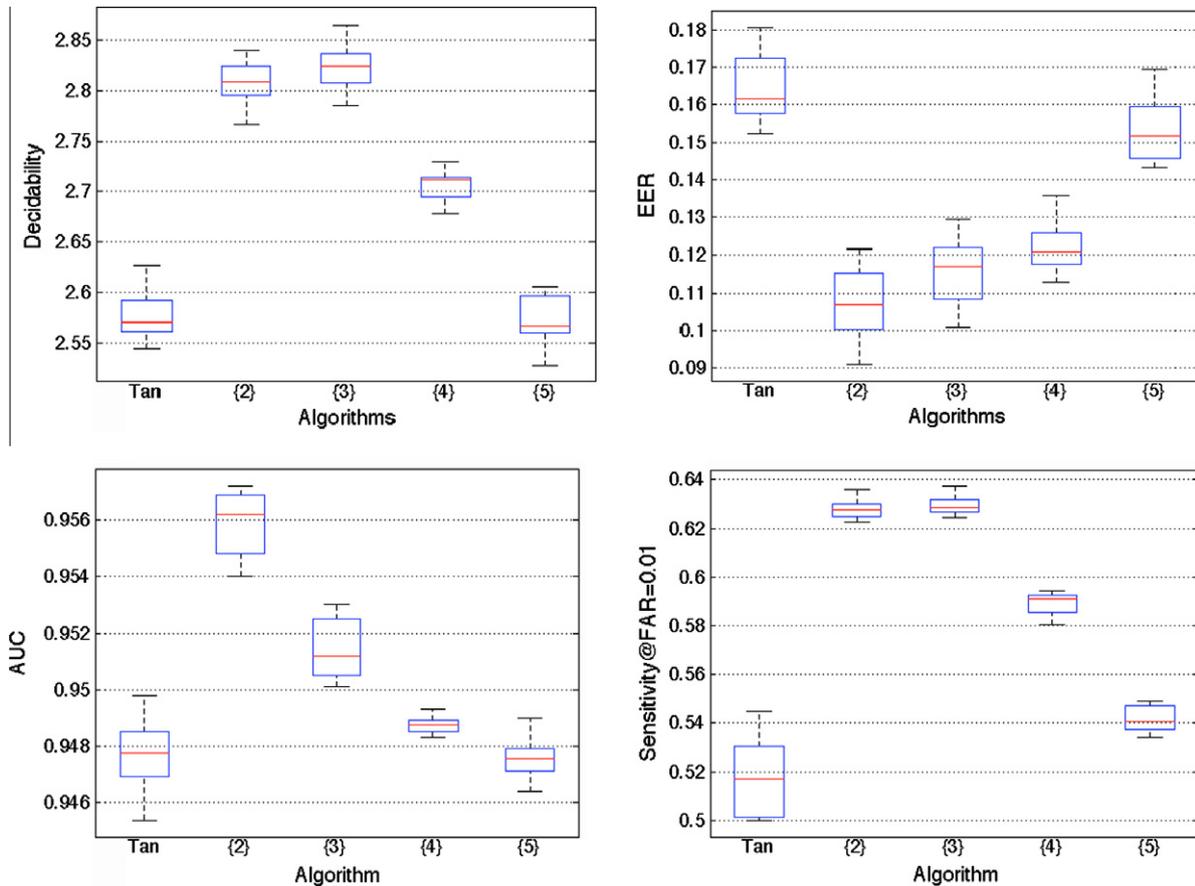


Fig. 16. Comparison between the performance obtained by the best classification ensembles composed of 2–5 recognition methods.

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

$$P(w_j|\bar{x}_i) = \frac{P(\bar{x}_i|w_j)}{\sum_s P(\bar{x}_i|w_s)}. \quad (16)$$

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

3.5. Degradations in performance

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

4. Conclusions

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

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

amount of light and angles of incidence). We are currently working on both these issues.

Acknowledgments

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

References

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

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