

A Novel Application of Universal Background Models for Periocular Recognition

João C. Monteiro and Jaime S. Cardoso

INESC TEC and Faculdade de Engenharia, Universidade do Porto
Campus da FEUP, Rua Dr. Roberto Frias, 378, 4200-465 Porto, Portugal
dee12007@fe.up.pt, jaime.cardoso@inescporto.pt

Abstract. In recent years the focus of research in the fields of iris and face recognition has turned towards alternative traits to aid in the recognition process under less constrained acquisition scenarios. The present work assesses the potential of the periocular region as an alternative to both iris and face in such conditions. An automatic modeling of SIFT descriptors, using a GMM-based Universal Background Model method, is proposed. This framework is based on the Universal Background Model strategy, first proposed for speaker verification, extrapolated into an image-based application. Such approach allows a tight coupling between individual models and a robust likelihood-ratio decision step. The algorithm was tested on the UBIRIS.v2 and the MobBIO databases and presented state-of-the-art performance for a variety of experimental setups.

Keywords: Biometrics, Iris segmentation, Unconstrained environment, Gradient flow, Shortest closed path

1 Introduction

Over the past few years face and iris have been on the spotlight of many research works in biometrics. The *face* is a easily acquirable trait with a high degree of uniqueness, while the *iris*, the coloured part of the eye, presents unique textural patterns resulting from its random morphogenesis during embryonic development [1]. These marked advantages, however, fall short when low-quality images are presented to the system. It has been noted that the performance of iris and face recognition algorithms is severely compromised when dealing with non-ideal scenarios such as non-uniform illumination, pose variations, occlusions, expression changes and radical appearance changes [1]. Several recent works have tried to explore alternative hypothesis to overcome this problem, either by developing more robust algorithms or by exploring new traits to allow or aid in the recognition process [28].

The *periocular* region is one of such unique traits. Even though a true definition of the periocular region is not standardized, it is common to describe it as the region in the immediate vicinity of the eye [10]. Periocular recognition can be motivated as a representation in between face and iris recognition. It has

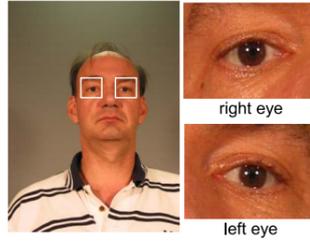


Fig. 1. Example of periocular regions from both eyes, extracted from a face image [28].

been shown to present increased performance when only degraded facial data [7] or low quality iris images [26] are made available, as well as promising results as a soft biometric trait to help improve both face and iris recognition systems in less constrained acquisition environments [2].

Periocular biometrics is a recent area of research, proposed by the first time in a feasibility study by Park et al. [12]. In this pioneer work, the authors suggested the periocular region as a potential alternative to circumvent the significant challenges posed to iris recognition systems working under unconstrained scenarios. In recent years, a number of relevant works have helped further the potential of periocular recognition in the field of biometrics. Padole and Proença [10] explored the effect of scale, pigmentation and occlusion, as well as gender, and propose an initial region-of-interest detection step to improve recognition accuracy. Ross et al. [21], following a recent trend in biometric research, explored information fusion based on several feature extraction techniques, to handle the significant variability of input periocular images. Tan et al. [26] studied the benefits of periocular recognition when highly degraded regions result from the traditional iris segmentation step. A more in depth analysis of the state-of-the-art in periocular recognition can be found in our previous work [8] and in the thorough review work of Santos and Proença [23].

The present work serves as an extension of [8], where we first proposed an innovative approach to periocular recognition under less ideal acquisition conditions. Such proposal is based on the idea of maximum *a posteriori* adaptation of Universal Background Model, as proposed by Reynolds for speaker verification [19]. We evaluate the proposed algorithm on two datasets of color periocular images acquired under visible wavelength (VW) illumination. Multiple noise factors such as varying gazes/poses and heterogeneous lighting conditions are characteristic to such images, thus representing the main challenge of the present work. In addition to the aforementioned objectives, we also aimed to assess the performance of the proposed algorithm with variable distances between the captures individuals and the acquisition apparatus. As no objective definition exists concerning which anatomical structures compose the periocular region, this analysis might provide significant insight for further research in the field of periocular recognition.

2 PROPOSED METHODOLOGY

2.1 Algorithm overview

The proposed algorithm is schematically represented in Figure 2. The two main blocks - *enrollment* and *recognition* - refer to the typical architecture of a biometric system. During enrollment a new individual's biometric data is inserted into a previously existent database of individuals. Such database is probed during the recognition process to assess either the validity of an identity claim - *verification* - or the k most probable identities - *identification* - given an unknown sample of biometric data.

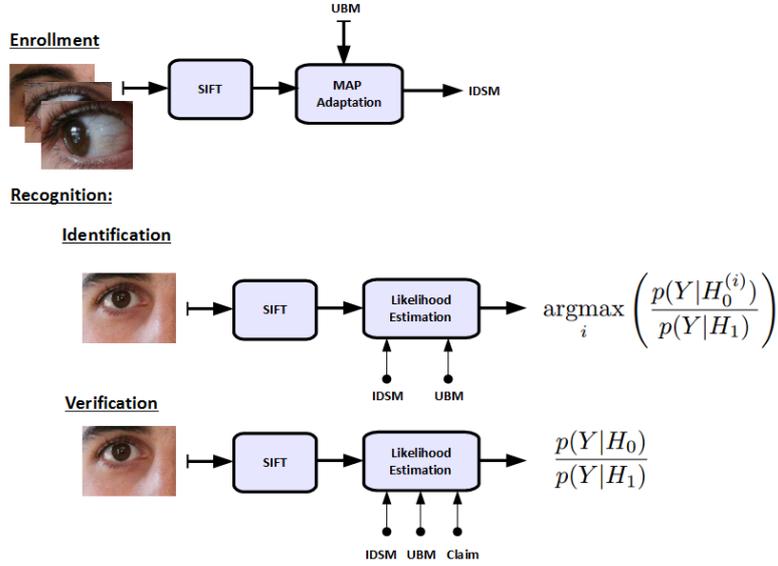


Fig. 2. Graphical representation of the main steps in both the enrollment and recognition (verification and identification) phases of the proposed periocular recognition algorithm.

During the enrollment, a set of N models describing the unique statistical distribution of biometric features for each individual $n \in \{1, \dots, N\}$ is trained by maximum *a posteriori* (MAP) adaptation of an Universal Background Model (UBM). The UBM is a representation of the variability that the chosen biometric trait presents in the universe of all individuals. MAP adaptation works as a specialization of the UBM based on each individual's biometric data. The idea of MAP adaptation of the UBM was first proposed by Reynolds [19], for speaker verification, and will be further motivated in the following sections.

The recognition phase is carried out through the projection of the features extracted from an unknown sample onto both the UBM and the individual specific

models (IDSM) of interest. A likelihood-ratio between both projections outputs the final recognition score. Depending on the functioning mode of the system - verification or identification - decision is carried out by thresholding or maximum likelihood-ratio, respectively.

2.2 Universal Background Model

Universal background modeling is a common strategy in the field of voice biometrics [13]. It can be easily understood if the problem of biometric verification is interpreted as a basic hypothesis test. Given a biometric sample Y and a claimed ID, S , we define:

$$\begin{aligned} H_0: & Y \text{ belongs to } S \\ H_1: & Y \text{ does not belong to } S \end{aligned}$$

as the null and alternative hypothesis, respectively. The optimal decision is taken by a *likelihood-ratio test*:

$$\frac{p(Y|H_0)}{p(Y|H_1)} \begin{cases} \geq \theta & \text{accept } H_0 \\ \leq \theta & \text{accept } H_1 \end{cases} \quad (1)$$

where θ is the decision threshold for accepting or rejecting H_0 , and $p(Y|H_i), i \in \{0, 1\}$ is the likelihood of observing sample Y when we consider hypothesis i to be true.

Biometric recognition can, thus, be reduced to the problem of computing the likelihood values $p(Y|H_0)$ and $p(Y|H_1)$. It is intuitive to note that H_0 should be represented by a model λ_{hyp} that characterizes the hypothesized individual, while, alternatively, the representation of H_1 , $\overline{\lambda_{hyp}}$, should be able to model *all the alternatives to the hypothesized individual*.

From such formulation arises the need for a model that successfully covers the space of alternatives to the hypothesized identity. The most common designation in literature for such a model is *universal background model* or *UBM* [20]. Such model must be trained on a large set of data, so as to faithfully cover a representative user space and a significant amount of sources of variability. The following section details the chosen strategy to model $\overline{\lambda_{hyp}}$ and how individual models, λ_{hyp} , can be adapted from the UBM in a fast and robust way.

2.3 Hypothesis Modeling

On the present work we chose Gaussian Mixture Models (GMM) to model both the UBM, i.e. $\overline{\lambda_{hyp}}$, and the individual specific models (IDSM), i.e. λ_{hyp} . Such models are capable of capturing the empirical probability density function (PDF) of a given set of feature vectors, so as to faithfully model their intrinsic statistical properties [19]. The choice of GMM to model feature distributions in biometric data is extensively motivated in many works of related areas. From the most common interpretations, GMMs are seen as capable of representing broad “hidden” classes, reflective of the unique structural arrangements observed in the

analysed biometric traits [19]. Besides this assumption, Gaussian mixtures display both the robustness of parametric unimodal Gaussian density estimates, as well as the ability of non-parametric models to fit non-Gaussian data [18]. This duality, alongside the fact that GMM have the noteworthy strength of generating smooth parametric densities, confers such models a strong advantage as generative model of choice. For computational efficiency, GMM models are often trained using diagonal covariance matrices. This approximation is often found in biometrics literature, with no significant accuracy loss associated [29].

All models are trained on sets of Scale Invariant Feature Transform (SIFT) keypoint descriptors [6]. This choice for periocular image description is thoroughly motivated in literature [21, 11], mainly due to the observation that local descriptors work better than their global counterparts when the available data presents non-uniform conditions. Furthermore, the invariance of SIFT features to a set of common undesirable factors (image scaling, translation, rotation and also partially to illumination and affine or 3D projection), confer them a strong appeal in the area of unconstrained biometrics.

2.4 H_1 : UBM Parameter Estimation

To train the Universal Background Model a large amount of “impostor” data, i.e. a set composed of data from all the enrolled individuals, is used, so as to cover a wide range of possibilities in the individual search space [25]. The training process of the UBM is simply performed by fitting a k -mixture GMM to the set of PCA-reduced feature vectors extracted from all the “impostors”.

If we interpret the UBM as an “impostor” model, its “genuine” counterpart can be obtained by adaptation of the UBM’s parameters, λ_{hyp} , using individual specific data. For each enrolled individual, ID , an *individual specific model* (IDSM), defined by parameters λ_{hyp} , is therefore obtained. The adaptation process will be outlined in the following section.

2.5 H_0 : MAP Adaptation of the UBM

IDSMs are generated by the *tuning of the UBM parameters* in a maximum *a posteriori* (MAP) sense, using individual specific biometric data. This approach provides a tight coupling between the IDSM and the UBM, resulting in better performance and faster scoring than uncoupled methods [29], as well as a robust and precise parameter estimation, even when only a small amount of data is available [25]. This is indeed one of the main advantages of using UBMs. The determination of appropriate initial values (i.e. seeding) of the parameters of a GMM remains a challenging issue. A poor initialization may result in a weak model, especially when the data volume is small. Since the IDSM are learnt only from each individual data, they are more prone to a poor convergence than the GMM for the UBM, learnt from a big pool of individuals. In essence, UBM constitutes a good initialization for the IDSM.

The adaptation process, as proposed by Reynolds [19], resembles the Expectation-Maximization (EM) algorithm, with two main estimation steps. The first is similar to the expectation step of the EM algorithm, where, for each mixture of the UBM, a set of sufficient statistics are computed from a set of M individual specific feature vectors, $X = \{\mathbf{x}_1 \dots \mathbf{x}_M\}$:

$$n_i = \sum_{m=1}^M p(i|\mathbf{x}_m) \quad (2)$$

$$E_i(\mathbf{x}) = \frac{1}{n_i} \sum_{m=1}^M p(i|\mathbf{x}_m) \mathbf{x}_m \quad (3)$$

$$E_i(\mathbf{x}\mathbf{x}^t) = \frac{1}{n_i} \sum_{m=1}^M p(i|\mathbf{x}_m) \mathbf{x}_m \mathbf{x}_m^t \quad (4)$$

where $p(i|\mathbf{x}_m)$ represents the probabilistic alignment of \mathbf{x}_m into each UBM mixture. Each UBM mixture is then adapted using the newly computed sufficient statistics, and considering diagonal covariance matrices. The update process can be formally expressed as:

$$\hat{w}_i = [\alpha_i n_i / M + (1 - \alpha_i) w_i] \xi \quad (5)$$

$$\hat{\boldsymbol{\mu}}_i = \alpha_i E_i(\mathbf{x}) + (1 - \alpha_i) \boldsymbol{\mu}_i \quad (6)$$

$$\hat{\Sigma}_i = \alpha_i E_i(\mathbf{x}\mathbf{x}^t) + (1 - \alpha_i) (\boldsymbol{\sigma}_i \boldsymbol{\sigma}_i^t + \boldsymbol{\mu}_i \boldsymbol{\mu}_i^t) - \hat{\boldsymbol{\mu}}_i \hat{\boldsymbol{\mu}}_i^t \quad (7)$$

$$\boldsymbol{\sigma}_i = \text{diag}(\Sigma_i) \quad (8)$$

where $\{w_i, \boldsymbol{\mu}_i, \boldsymbol{\sigma}_i\}$ are the original UBM parameters and $\{\hat{w}_i, \hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\sigma}}_i\}$ represent their adaptation to a specific speaker. To assure that $\sum_i w_i = 1$ a weighting parameter ξ is introduced. The α parameter is a data-dependent adaptation coefficient. Formally it can be defined as:

$$\alpha_i = \frac{n_i}{r + n_i} \quad (9)$$

where r is generally known as the *relevance factor*. The individual dependent adaptation parameter serves the purpose of weighting the relative importance of the original values and the new sufficient statistics in the adaptation process. For the UBM adaptation we set $r = 16$, as this is the most commonly observed value in literature [19]. Most works propose the sole adaptation of the mean values, i.e. $\alpha_i = 0$ when computing \hat{w}_i and $\hat{\boldsymbol{\sigma}}_i$. This simplification seems to bring no nefarious effects over the performance of the recognition process, while allowing faster training of the individual specific models [3].

2.6 Recognition and Decision

After the training step of both the UBM and each IDSM, the recognition phase with new data from an unknown source is somewhat trivial. As referred in previous sections, the identity check is performed through the projection of the new test data, $X_{test} = \{\mathbf{x}_{t,1}, \dots, \mathbf{x}_{t,N}\}$, where $\mathbf{x}_{t,i}$ is the i -th PCA-reduced SIFT vector extracted from the periocular region of test subject t , onto both the UBM and either the claimed IDSM (in verification mode) or all such models (in identification mode). The recognition score is obtained as the average likelihood-ratio of all keypoint descriptors $\mathbf{x}_{t,i}, \forall i \in \{1..N\}$. The decision is then carried out by checking the condition presented in Equation (1), in the case of verification, or by detecting the maximum likelihood-ratio value for all enrolled IDs, in the case of identification.

This is a second big advantage of using UBM. The ratio between the IDSM and the UBM probabilities of the observed data is a more robust decision criterion than relying solely on the IDSM probability. This results from the fact that some subjects are more prone to generate high likelihood values than others, i.e. some people have a more “generic” look than others. The use of a likelihood ratio with an universal reference works as a normalization step, mapping the likelihood values in accord to their global projection. Without such step, finding a global optimal value for the decision threshold, θ , presented in Equation 1 would be a far more complex process.

3 Experimental results

In this section we start by presenting the datasets and the experimental setups under which performance was assessed. Further sections present a detailed analysis regarding the effect of model complexity and fusion of color channels in the global performance of the proposed algorithm.

3.1 Tested datasets

The proposed algorithm was tested on two noisy color iris image databases: UBIRIS.v2 and MobBIO. Even though both databases were designed in an attempt to promote the development of robust iris recognition algorithms for images acquired under VW illumination, their intrinsic properties make them attractive to study the feasibility of periocular recognition under similar conditions. The following sections detail their main features as well as the reasoning behind their choice.

UBIRIS.v2 database Images in UBIRIS.v2 [17] database were captured under non-constrained conditions (at-a-distance, on-the-move and on the visible wavelength), with corresponding realistic noise factors. Figure 3 depicts some examples of these noise factors (reflections, occlusions, pigmentation, etc.). Two acquisition sessions were performed with 261 individuals involved and a total of

11100 300×400 color images acquired. Each individual's images were acquired at variable distances with 15 images per eye and per season.

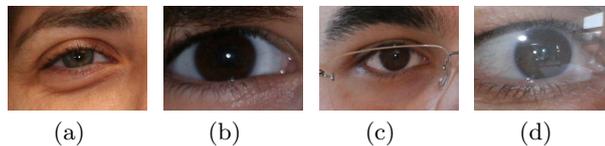


Fig. 3. Examples of noisy image from the UBIRIS.v2 database.

MobBIO database The MobBIO multimodal database [24] was created in the scope of the 1st Biometric Recognition with Portable Devices Competition 2013, integrated in the ICIAR 2013 conference. The main goal of the competition was to compare various methodologies for biometric recognition using data acquired with portable devices. We tested our algorithm on the iris modality present on this database. Regarding this modality the images were captured under two alternative lighting conditions, with variable eye orientations and occlusion levels, so as to comprise a larger variability of unconstrained scenarios. Distance to the camera was, however, kept constant for each individual. For each of the 105 volunteers 16 images (8 of each eye) were acquired. These images were obtained by cropping a single image comprising both eyes. Each cropped image was set to a 300×200 resolution. Figure 4 depicts some examples of such images.

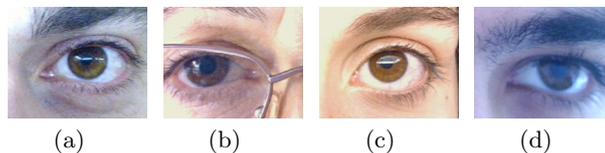


Fig. 4. Examples of iris images in the MobBIO database.

The MobBIO database presents a face modality which has also been explored for comparative purposes in the present work. Images were acquired in similar conditions to those described above for iris images, with 16 images per subject. Examples of such images can be observed in Figure 5.

3.2 Evaluation metrics

Performance was evaluated for both verification and identification modes. Regarding the former we analyzed the *equal error rate* (EER) and the *decidability*



Fig. 5. Examples of face images in the MobBIO database.

index (DI). The EER is observed at the decision threshold, θ , where the errors of falsely accepting and falsely rejecting H_0 occur with equal frequency. The global behavior of both types of errors is often analyzed through receiver operating characteristic (ROC) curves. On the other hand, the DI quantifies the separation of the “genuine” and “impostor” likelihood score distributions, as follows:

$$DI = \frac{|\mu_g - \mu_i|}{\sqrt{0.5(\sigma_g^2 + \sigma_i^2)}} \quad (10)$$

where (μ_g, σ_g) and (μ_i, σ_i) are the mean and standard deviation of the genuine and impostor score distributions, respectively.

For identification we analyze cumulative match curves (CMC). These curves represent the rate of correctly identified individuals, by checking if the true identity is present in the N highest ranked identities. The N parameter is generally referred to as rank. That allows us to define the *rank-1 recognition rate* as the value of the CMC at $N = 1$.

3.3 Experimental setups

Our experiments were conducted in three distinct experimental setups, two of them regarding the UBIRIS.v2 database and the remaining one the MobBIO database:

1. In the first setup, for the UBIRIS.v2 images, six samples from 80 different subjects were used, captured from different distances (4 to 8 meters), with varying gazes/poses and notable changes in lighting conditions. One image per individual was randomly chosen as probe, whereas the remaining five samples were used for the UBM training and MAP adaptation. The results were cross-validated by changing the probe image, per subject, for each of the six chosen images.
2. Many works on periocular biometrics evaluate their results using a well-known subset of the UBIRIS.v2 database, used in the context of the NICE II competition [14]. This dataset is divided in train and test subsets, with a total of 1000 images from 171 individuals. In the present work we choose to use test

subset, composed by 904 images from 152 individuals. Only individuals with more than 4 available images were considered, as 4 images were randomly chosen for training and the rest for testing. Results were cross-validated 10-fold. The train dataset composed by the remaining 96 images from 19 individuals was employed in the parameter optimization step described in further sections.

- Concerning the MobBIO database, 8 images were randomly chosen from each of the 105 individuals for the training of the models, whereas the remaining 8 were chosen for testing. The process was cross-validated 10-fold. For comparative purposes a similar experiment was carried out on face images from the same 105 individuals, using the same 8 + 8 image distribution.

As both databases are composed by color images, each of the RGB channels was considered individually for the entire enrollment and identification process. For the parameter optimization described in the next section images were previously converted to grayscale.

3.4 Recognition Results

The results obtained for both databases and experimental setups are represented through ROC and CMC curves on Figures 6(a) to 6(f). A comparison with some state-of-the-art algorithms in the UBIRIS.v2 database is also presented in Table 1. In this table results are grouped according to the experimental setup of each reported work and also the studied trait: P - Periocular, I - Iris or $P + I$ - Fusion of both traits.

<i>Work</i>	<i>Setup</i>	<i>Traits</i>	R_1	EER	D_i
Proposed	1	P	97.73%	0.0452	4.9795
Proposed	2	P	88.93%	0.0716	3.6141
Moreno et al. [9]	1	P	97.63%	0.1417	-
Tan et al. [26]	2	$P + I$	39.4%	-	-
Tan et al. [27]	2	$P + I$	-	-	2.5748
Kumar et al. [5]	2	I	48.01%	-	-
Proença et al. [16]	2	I	-	≈ 0.11	2.848

Table 1. Comparison between the average obtained results with both experimental setups for the UBIRIS.v2 database and some state-of-the-art algorithms.

Besides testing each of the RGB channels individually, a simple sum-rule score-level fusion strategy [4] was also considered. It is easily discernible, from the observation of Figure 6, that the fusion of information from multiple color channels brings about a significant improvement in performance for all the tested datasets. When comparing the results obtained with this approach with some state-of-the-art algorithms a few points deserve further discussion. First, the proposed algorithm is capable of achieving and even surpassing state-of-the-art

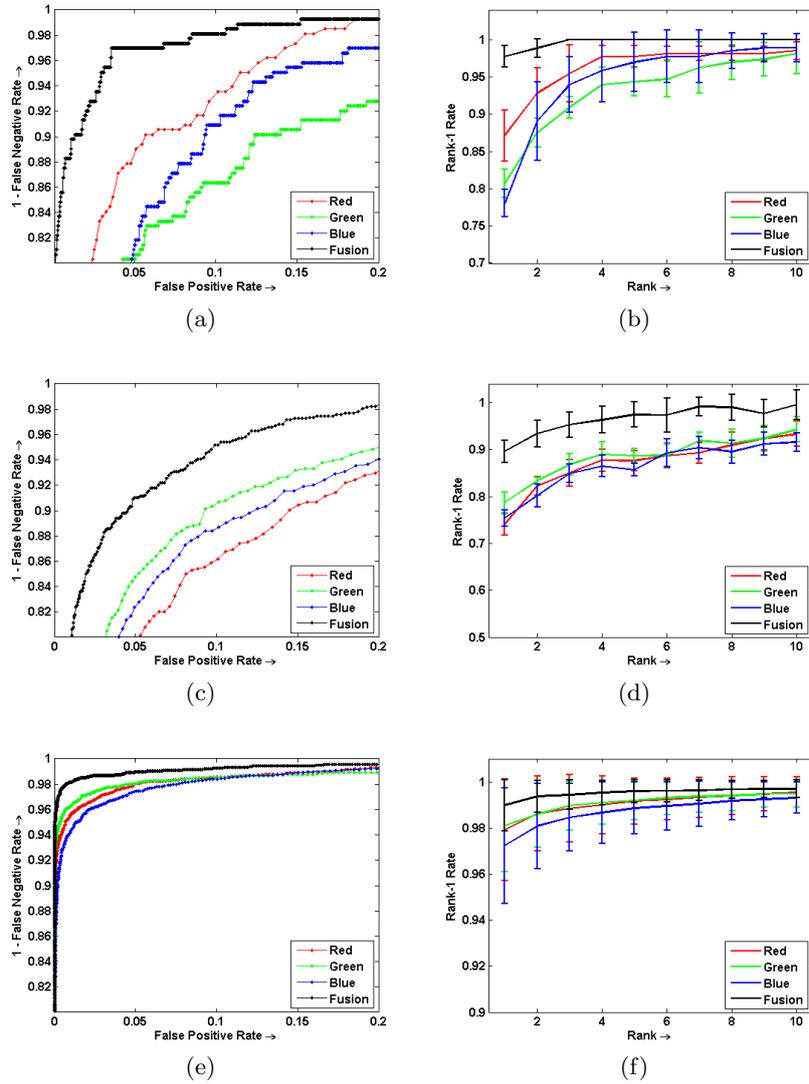


Fig. 6. ROC and CMC curves for: (a-b) UBIRIS.v2 database, setup (1); (c-d) UBIRIS.v2 database, setup (2) and (e-f) MobBIO database, setup (3);. ROC curves present the average results of cross-validation, whereas CMCs present the average value and error-bars for the first 10 ranked IDs in each setup.

performance in multiple experimental setups. Concerning the most common of such setups (2), it is interesting to note that a few works attempted to explore the UBIRIS.v2 dataset for iris recognition. The obtained performance has been considered “discouraging” in the work by Kumar et al. [5]. Comparing the rank-1

recognition rate obtained with our algorithm (88.93%) with the 48.1% reported in the former work, we conclude that the periocular region may represent a viable alternative to iris in images acquired under visible wavelength (VW) illumination. Such acquisition conditions are known to increase light reflections from the cornea, resulting in a sub-optimal signal-to-noise ratio (SNR) in the sensor, lowering the contrast of iris images and the robustness of the system [15]. More recent works have explored multimodal approaches, using combined information from both the iris and the periocular region. Analysis of Table 1 shows that none of such works reaches the performance reported in the present work for the same experimental setup. Such observation might indicate that most discriminative biometric information from the UBIRIS.v2 images might be present in the periocular region, and that considering data from the heavily noisy iris regions might only result in a degradation of the performance obtained by the periocular region alone.

Concerning the MobBIO database, an alternative comparison was carried out to analyze the potential of the periocular region as an alternative to face recognition. The observed performance for periocular images was considerably close to that using full-face information, with rank-1 recognition rates of 98.98% and 99.77% respectively. These results are an indication that, under more ideal acquisition conditions, there is enough discriminative potential in the periocular region alone to rival with the full face in terms of recognition performance. In scenarios where some parts of the face are purposely disguised (scarves covering the mouth for example) this observation might indicate that a non-corrupted periocular region can, indeed, overperform recognition with the occluded full-face images. Such conditions were not tested in the present work but might be the basis for an interesting follow-up. Even though the observed results are promising, it must be noted that the noise factors present in the MobBIO database are still far from a highly unconstrained scenario.

The robustness of the likelihood-ratio decision step was also assessed. We compared the performance observed for the scores obtained with Equation 1 and the scores obtained using only its numerator, i.e. only the likelihood of each test image without the UBM normalization. For the experimental setup (2) we obtained an average rank-1 recognition rate of 43.6%, whereas the MobBIO experimental setup (3) resulted in 90.5% for the same metric. It is easily noted that performance is less compromised in the MobBIO database. Considering only the numerator of Equation 1 is the same as considering a constant denominator value for every tested image. As the denominator represents the projection of the tested images on the UBM, this alternative decision strategy might be interpreted as assuming a constant background for every tested image. From the observed results we might conclude that such assumption fits better the images from the MobBIO database. We also note that for more challenging scenarios, where the constant background assumption fails, the use of background normalization produces a significant improvement in performance.

A few last considerations regarding the discriminative potential of the proposed algorithm may be taken from the observation of Figure 7. On each row

we analyze the 4 highest ranked models for the images presented in the first column. The first two rows depict correct identifications. It is interesting to note how each of the 4 highest ranked identities in the second row correspond to individuals wearing glasses. Such observation seems to indicate that the proposed modeling process is capable of describing high-level global features, such as glasses. Furthermore, the fact that the correct ID was guessed also demonstrates its capacity of distinguishing between finer details separating individual models. The third and fourth rows present some test images whose ID was not correctly assessed by the algorithm. In the third row we present a case where even though the correct ID and the most likely model were not correctly paired, the correct guess still appears in the top ranked models. We note that even a human user analyzing the four highest ranked models would find it very difficult to detect significant differences. The fourth row presents the extreme case where none of the top ranked models correspond to the true ID. It is worth noting how the test images presented in the third and fourth rows are very similar to a large number of images present in other individual's models. This observation leads to the hypothesis that *some users are easier to identify than others inside a given population*, an effect known as the *Doddington zoo effect* [22]. It also shows that the proposed algorithm is capable of narrowing the range of possible identities to those subjects who “look more alike”.

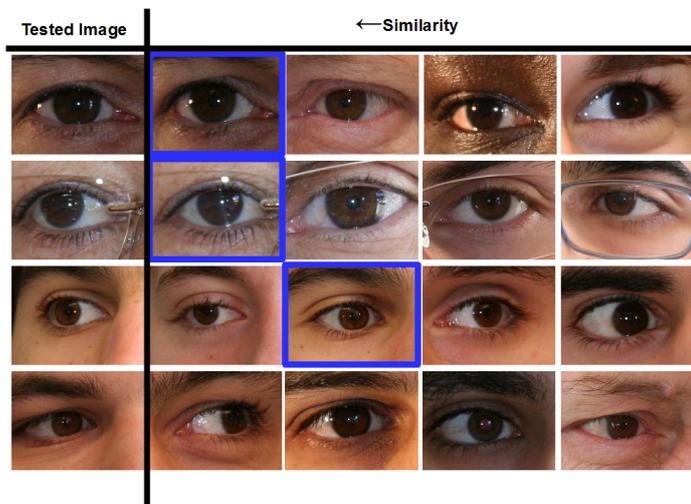


Fig. 7. Identification results for rank-4 in the UBIRIS.v2 database. The first column depicts the tested images while the remaining 4 images exemplify representative images from the 4 most probable models, after the recognition is performed. The blue squares mark the true identity.

3.5 Effect of acquisition distance on performance

An additional test was carried out to assess how the distance between the acquisition apparatus and the captured individual affects the performance of the proposed algorithm. With this in mind an alternative evaluation setup was designed. We chose the UBIRIS.v2 database for this analysis, as it is comprised by images acquired at variable distances (4 – 8 meters). Images were divided into training and testing as defined in Section 3.3 for the training setup 2. During the test process, the test images were divided into 5 different groups, according to the acquisition distance at which they were captured. Figure 8 depicts the evolution of the decidability index values with respect to the acquisition distance parameter, d_a . It is interesting to note that images acquired at farther distances present higher recognition rates, with exception of the images acquired at the closest distance. It was expected that farther images presented consistently better performance, as more anatomical landmarks (eyebrows, for example) become accessible to the recognition block, thus increasing the amount of discriminative information present in each tested image. The fact that the closest images present an increased performance might be connected either to a peculiarity of the tested data or to the fact that, in closer images, the iris contributes with a non-negligible role to the recognition process. The possibility that some of the closer images present enough detail in the iris region to improve the performance in such conditions seems legitimate. It is, however, also discernible that even in this possible semi-iris recognition scenario, performance is still significantly lower than the one observed when the periocular region is set to incorporate a larger portion of the vicinity of the eye. It may be argued that this scenario is rapidly approaching the ideal conditions of full-face recognition. However, as no objective description of the periocular region is standardized in the research community, these observations might help in the process of establishing an uniform definition of such region.

4 CONCLUSIONS AND FUTURE WORK

In the present work we propose an automatic modeling of SIFT descriptors, using a GMM-based UBM method, to achieve a canonical representation of individual's biometric data, regardless of the number of detected SIFT keypoints. We tested the proposed algorithm on periocular images from two databases and achieved state-of-the-art performance for all experimental setups. Periocular recognition has been the focus of many recent works that explore it as a viable alternative to both iris and face recognition under less ideal acquisition scenarios.

Even though we propose the algorithm for periocular recognition, the framework can be easily extrapolated for other image-based traits. To the extent of our knowledge, GMM-based UBM methodologies were solely explored for speaker recognition so far. The proposed work may, thus, represent the first of a series of experiments that explore its main advantages in the scope of multiple trending biometric topics. For example, the fact that any number of keypoints triggers a recognition score may be relevant when only partial or occluded data is available

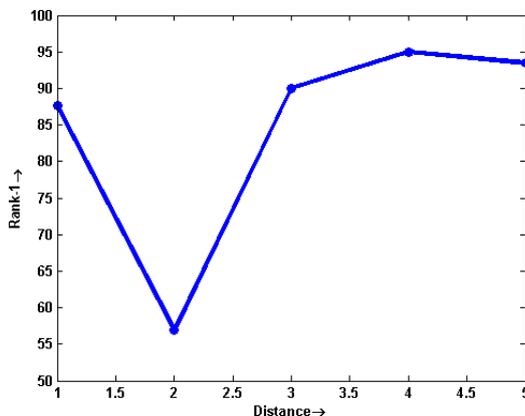


Fig. 8. Evolution of the rank-1 recognition rate values observed for increasing distance between the captured individual and the acquisition apparatus in the UBIRIS.v2 database.

for recognition. Scenarios, like the one described in the last section, where faces are purposely occluded may be an interesting area to explore.

Besides from the conceptual advantages of the proposed algorithm, a few technical details may be improved in further works. Exploring further color channels besides the RGB space could bring benefits to the proposed algorithm. Regarding fusion, exploring individual specific parameters instead of a global parametrization, would enable the algorithm to be trained to counter the Dodington zoo effect. As not all people are as easy to identify, fitting the properties of the designed classification block to adapt to different classes of individuals seems like an interesting idea.

Finally, and regarding the training setup, some questions might be worthy of a more thorough research. In the case of voice recognition it is common to train two separate UBMs for male and female speakers. Extrapolating this idea to image-based traits, multiple UBMs trained on homogeneous sets of equally or similarly zoomed images might improve the results when more realistic and dynamic conditions are presented to the acquisition system. In a related topic it is also not consensual whether the left and right eyes, due to the intrinsic symmetry of the face, should be considered in a single model or as separate entities. All the aforementioned questions demonstrate how much the present results can be improved, leaving some promising prospects for future works.

ACKNOWLEDGEMENTS

The first author would like to thank Fundação para a Ciência e Tecnologia (FCT) - Portugal the financial support for the PhD grant SFRH/BD/87392/2012.

References

1. Bakshi, S., Kumari, S., Raman, R., Sa, P.K.: Evaluation of periocular over face biometric: A case study. *Procedia Engineering* 38, 1628–1633 (2012)
2. Joshi, A., Gangwar, A.K., Saquib, Z.: Person recognition based on fusion of iris and periocular biometrics. In: *Hybrid Intelligent Systems (HIS), 2012 12th International Conference on*. pp. 57–62. IEEE (2012)
3. Kinnunen, T., Saastamoinen, J., Hautamaki, V., Vinni, M., Franti, P.: Comparing maximum a posteriori vector quantization and gaussian mixture models in speaker verification. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*. pp. 4229–4232. IEEE (2009)
4. Kittler, J., Hatef, M., Duin, R.P., Matas, J.: On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 20(3), 226–239 (1998)
5. Kumar, A., Chan, T.S.: Iris recognition using quaternionic sparse orientation code (qsoc). In: *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. pp. 59–64 (2012)
6. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60(2), 91–110 (2004)
7. Miller, P.E., Rawls, A.W., Pundlik, S.J., Woodard, D.L.: Personal identification using periocular skin texture. In: *2010 ACM Symposium on Applied Computing*. pp. 1496–1500. ACM (2010)
8. Monteiro, J.C., Cardoso, J.S.: Periocular recognition under unconstrained settings with universal background models. In: *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing (BIOSIGNALS) (2015)*, <http://www.inescporto.pt/~jsc/publications/conferences/2015JCMonteiroBIOSIGNALS.pdf>
9. Moreno, J.C., Prasath, V., Proença, H.: Robust periocular recognition by fusing local to holistic sparse representations. In: *Proceedings of the 6th International Conference on Security of Information and Networks*. pp. 160–164. ACM (2013)
10. Padole, C.N., Proenca, H.: Periocular recognition: Analysis of performance degradation factors. In: *5th IAPR International Conference on Biometrics*. pp. 439–445 (2012)
11. Park, U., Jillela, R.R., Ross, A., Jain, A.K.: Periocular biometrics in the visible spectrum. *IEEE Transactions on Information Forensics and Security* 6(1), 96–106 (2011)
12. Park, U., Ross, A., Jain, A.K.: Periocular biometrics in the visible spectrum: A feasibility study. In: *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*. pp. 1–6 (2009)
13. Povey, D., Chu, S.M., Varadarajan, B.: Universal background model based speech recognition. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*. pp. 4561–4564 (2008)
14. Proença, H.: NICE:II: Noisy iris challenge evaluation - part ii (2009), <http://http://nice2.di.ubi.pt/>
15. Proença, H.: Non-cooperative iris recognition: Issues and trends. In: *19th European Signal Processing Conference*. pp. 1–5 (2011)
16. Proença, H., Santos, G.: Fusing color and shape descriptors in the recognition of degraded iris images acquired at visible wavelengths. *Computer Vision and Image Understanding* 116(2), 167–178 (2012)
17. Proença, H., Filipe, S., Santos, R., Oliveira, J., Alexandre, L.A.: The ubiris.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance.

- IEEE Transactions on Pattern Analysis and Machine Intelligence 32(8), 1529–1535 (2010)
18. Reynolds, D.: Gaussian mixture models. *Encyclopedia of Biometric Recognition* pp. 12–17 (2008)
 19. Reynolds, D., Quatieri, T., Dunn, R.: Speaker verification using adapted gaussian mixture models. *Digital signal processing* 10(1), 19–41 (2000)
 20. Reynolds, D.A.: An overview of automatic speaker recognition technology. In: *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*. vol. 4, pp. IV–4072. IEEE (2002)
 21. Ross, A., Jillela, R., Smereka, J.M., Boddeti, V.N., Kumar, B.V., Barnard, R., Hu, X., Pauca, P., Plemmons, R.: Matching highly non-ideal ocular images: An information fusion approach. In: *Biometrics (ICB), 2012 5th IAPR International Conference on*. pp. 446–453. IEEE (2012)
 22. Ross, A., Rattani, A., Tistarelli, M.: Exploiting the doddington zoo effect in biometric fusion. In: *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*. pp. 1–7. IEEE (2009)
 23. Santos, G., Proença, H.: Periocular biometrics: An emerging technology for unconstrained scenarios. In: *Computational Intelligence in Biometrics and Identity Management (CIBIM), 2013 IEEE Workshop on*. pp. 14–21. IEEE (2013)
 24. Sequeira, A.F., Monteiro, J.C., Rebelo, A., Oliveira, H.P.: MobBIO: a multimodal database captured with a portable handheld device. In: *Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP)* (2014)
 25. Shinoda, K., Inoue, N.: Reusing speech techniques for video semantic indexing [applications corner]. *Signal Processing Magazine, IEEE* 30(2), 118–122 (2013)
 26. Tan, C.W., Kumar, A.: Towards online iris and periocular recognition under relaxed imaging constraints. *Image Processing, IEEE Transactions on* 22(10), 3751–3765 (2013)
 27. Tan, T., Zhang, X., Sun, Z., Zhang, H.: Noisy iris image matching by using multiple cues. *Pattern Recognition Letters* (2011)
 28. Woodard, D.L., Pundlik, S.J., Lyle, J.R., Miller, P.E.: Periocular region appearance cues for biometric identification. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. pp. 162–169 (2010)
 29. Xiong, Z., Zheng, T., Song, Z., Soong, F., Wu, W.: A tree-based kernel selection approach to efficient gaussian mixture model–universal background model based speaker identification. *Speech communication* 48(10), 1273–1282 (2006)