A Comparative Analysis of Two Approaches to Periocular Recognition in Mobile Scenarios

João C. Monteiro[1(✉)], Rui Esteves[2], Gil Santos[3], Paulo Torrão Fiadeiro[4], Joana Lobo[2], and Jaime S. Cardoso[1]

[1] INESC TEC and Faculdade de Engenharia, Universidade do Porto, Campus da FEUP, Rua Dr. Roberto Frias, 378, 4200-465 Porto, Portugal
jcmonteiro89@gmail.com
[2] Associação Fraunhofer Portugal Research, Rua Alfredo Allen 455/461, 4200-135 Porto, Portugal
[3] IT - Instituto de Telecomunicacoes, Lisboa, Portugal
[4] Departamento de Física, Unidade de Detecção Remota Universidade da Beira Interior, Rua Marquês D'Ávila e Bolama, 6201-001 Covilhã, Portugal

Abstract. In recent years, periocular recognition has become a popular alternative to face and iris recognition in less ideal acquisition scenarios. An interesting example of such scenarios is the usage of mobile devices for recognition purposes. With the growing popularity and easy access to such devices, the development of robust biometric recognition algorithms to work under such conditions finds strong motivation. In the present work we assess the performance of extended versions of two state-of-the-art periocular recognition algorithms on the publicly available CSIP database, a recent dataset composed of images acquired under highly unconstrained and multi-sensor mobile scenarios. The achieved results show each algorithm is better fit to tackle different scenarios and applications of the biometric recognition problem.

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, is composed by a set of irregular 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. With the increasing popularity and availability of mobile devices capable of performing the whole biometric recognition framework, from data acquisition to final decision, serves as further motivation for research in the field of unconstrained biometrics [2]. Several recent works have tried to explore alternative hypotheses to overcome this challenge, either by developing more robust algorithms or by exploring new traits to allow or aid in the recognition process.

G. Bebis et al. (Eds.): ISVC 2015, Part II, LNCS 9475, pp. 268–280, 2015.
DOI: 10.1007/978-3-319-27863-6_25

The *periocular* region is one of such unique traits. It is common to describe the periocular region as the region in the immediate vicinity of the eye. Periocular recognition can be motivated as a representation in between face and iris recognition. It has been shown to present increased performance when only degraded facial data or low quality iris images are made available. Even in mobile application scenarios, the periocular region does not require rigid capture or complex imaging systems, thereby making it easy to acquire even by an inexperienced user. Nevertheless, several problems arise when attempting to perform periocular biometrics in mobile environments. The wide variety of camera sensors and lenses used in mobile devices produce discrepancies in working images, as they might be acquired with both color distortions and multiple resolutions. On-the-go acquisition by inexperienced subjects will result in demanding pose, illumination, and expression changes, thereby yielding variable acquisition angles and scales, or rotated images. All these limitations are intrinsic to the nature of mobile devices and must, thus, be handled by the recognition algorithm.

On the present work we aim to compare the performance of two state-of-the-art approaches to periocular recognition - an extension of Monteiro et al. [3] to multiple features and fusion strategies and Santos et al. [2] - when exposed to images acquired on multiple mobile scenarios, using a recently collected multi-sensor periocular database [2]. We evaluate both approaches with regards both to recognition performance as well as the processing time, with real-world applications in mind. Finally, we present some preliminary results on cross-sensor periocular recognition, thourgh the analysis of whether or not multiple sensors from varying manufacturers present meaningful interoperability.

2 Related Work

Periocular biometrics is a recent area of research, proposed by the first time in a feasibility study by Park et al. [4]. 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. The same authors analysed the effect of degradation on the accuracy of periocular recognition [5]. Padole and Proença [6] also explore 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. [7] explored information fusion based on several feature extraction techniques, to handle the significant variability of input periocular images. Information fusion has become one of the trends in biometric research in recent years and periocular recognition is no exception.

Some works have explored the advantages of the periocular region as an aid to more traditional approaches based on iris. Joshi et al. [8] proposed feature level fusion of wavelet coefficients and LBP features, from the iris and periocular regions respectively, with considerable performance improvement over both singular traits. A more recent work by Tan et al. [9] has also explored the benefits of periocular recognition when highly degraded regions result from the traditional

iris segmentation step. The authors have observed discouraging performance when the iris region alone is considered in such scenarios, whereas introducing information from the whole periocular region lead to a significant improvement. A thorough review of the most relevant method in recent years concerning periocular recognition and its main advantages can be found in the work by Santos and Proença [10].

On the present work we chose to perform a comparative analysis of two recent approaches to the issue of periocular recognition, when working with images acquired in unconstrained mobile scenarios. These works, by Monteiro et al. [3] and Santos et al. [2], will be analyzed in further detail in the following section.

3 Recognition Algorithms

This section will detail both previously referred methodologies for periocular recognition. Both approaches will be analysed in a comparative scenario so as to ascertain their main advantages and disadvantages in the mobile acquisition environments that serve as motivation for the present work.

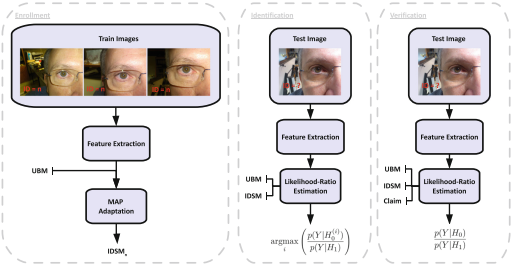

Fig. 1. Flow diagram of the main composing blocks of the methodology proposed in [3].

3.1 Method 1: GMM-UBM

The GMM-UBM algorithm for periocular recognition, first proposed by Monteiro et al. [3], is schematically represented in Fig. 1. During the enrollment, a set of N models describing the unique statistical distribution of biometric features for each individual $n \in \{1, \ldots, N\}$ 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 [11], for speaker verification. The tuning of the UBM parameters in a maximum *a posteriori* sense, using individual specific biometric data, provides a tight coupling between the individual models and the UBM, resulting in better performance and faster scoring than uncoupled methods, as well as a robust and precise parameter estimation, even when only a small amount of data is available.

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. The use of a likelihood-ratio score 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 would be a far more complex process.

Gaussian Mixture Models (GMM) were chosen to model both the UBM and the individual specific models (IDSM). Regarding feature extraction while the original algorithm used SIFT keypoint descriptors alone, on the present work we present an extension to four distinct descriptors, as detailed later.

3.2 Method 2: Santos et al.

This algorithm, proposed in [2], may be divided into four main blocks: normalization using a device-specific color correction and region-of-interest (ROI) definition; feature encoding using information from both the iris and the periocular region; feature matching and score-level fusion. The flow of information through the aforementioned blocks is schematically depicted in Fig. 2.

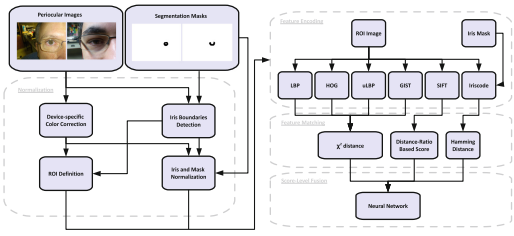

Fig. 2. Flow diagram of the main composing blocks of the methodology proposed in [2].

During the normalization block a device-specific color-correction was applied so as to compensate for possible chromatic distortions observed in real-life scenarios. Another variability source that is commonly observed in data acquired with mobile devices is variable scale. In order to overcome such problem, and making use of a state-of-the-art iris segmentation algorithm [12], the authors propose a segmentation of the iris boundary, to serve as a reference for the periocular region. Using the previously calculated radius of the iris, r_i, the periocular ROI was defined as 35 square patches that formed a 7×5 grid, where each patch had an area equivalent to $1.4r_i^2$.

Periocular data was encoded using a similar feature extraction scenario as the one described for the previous methodology. SIFT, HOG, uLBP (as well as the original LBP) and GIST, were also tested independently and used in a conjugated manner. For the iris region, in addition to these descriptors, a fifth

approach was also explored, using the original iriscode algorithm proposed by Daugman [13]. For a single image a total of 11 feature descriptors is, therefore, extracted: 5 for the periocular region and the same 5 plus the iriscode for the iris region.

Matching was carried out by comparing the 11 pairs of feature descriptors extracted from a pair of images, using the matching algorithm specified for each of them (χ^2 for histogram-based algorithms, distance-based score for pairs of SIFT keypoints and Hamming Distance for iriscode), resulting, thus, in 11 individual scores. Performance can then be evaluated either for each descriptor individually or by exploring more complex fusion strategies. In the original algorithm, a multi-layer perceptron artificial neural network was used to achieve such fusion.

3.3 Algorithm Extensions

Besides the comparative analysis between the two methodologies described in the previous sections, we also present some extensions to their original formulations. Such modifications will be present in the following sections.

Pre-processing. An no pre-processing strategy was included in the original formulation of the GMM-UBM algorithm, we aimed to assess if its presence could bring about a significant improvement in performance. We chose the Discrete Cosine Transform (DCT), as proposed by Chen et al. [14] for illumination normalization in face images, as it yielded the best overall performance. This technique is based on the removal of low-frequency coefficients of the DCT, in order to compensate for the variations in lighting conditions, that are known to lie, mainly, on such frequency band. We also tested the device-specific colorcorrection technique proposed in Method 2.

Periocular Segmentation. As the GMM-UBM approach presented no preliminary periocular segmentation in its original form, and in order to achieve an uniform set of conditions for performance comparison, we chose to perform segmentation with the same methodology used for Method 2, as described in Sect. 3.2.

Feature Descriptors. Similar to segmentation, we chose to explore the performance of multiple features using Method 1, as we believed that multiple sources of information might offer the algorithm an increased robustness when dealing with more complex and realistic datasets. Similar to Method 2 we chose to test the GMM-UBM with the LBP, HOG and GIST descriptors, besides the original SIFT formulation.

Fusion Strategies. Fusion scenarios can contribute, in some complex situations, to an overall improvement of system performance. On the present work,

two fusion strategies at score level were evaluated: performance-weighted score-level fusion and neural network score-level fusion. Both are novel to Method 1, which did not include any score-level fusion in its original formulation, while the simpler performance-weighted strategy is tested as an alternative to the original version of Method 2.

- **Performance-Weighted Score-Level Fusion:** the fusion score, s_f, is obtained by a weighted sum of the individual scores obtained for each feature individually:

$$s_f = w_{feat_1} \times s_{feat_1} + w_{feat_2} \times s_{feat_2} + ... + w_{feat_N} \times s_{feat_N} \quad (1)$$

  where s_{feat_n} is the individual score obtained for feature $n \in 1...N$ and w_{feat_n} is its corresponding weight in the final score. The weight of each feature is computed in relation to its individual performance:

$$w_{feat_n} = \frac{p_{feat_n}}{\sum_{i=1}^{N} p_{feat_i}} \quad (2)$$

  where p_{feat_n} is the individual performance obtained with a specific metric for feature n and the denominator term is introduced so that $\sum_i w_{feat_i} = 1$.
- **Neural Network Score-Level Fusion:** the final recognition score, s_f, is obtained by a multilayer perceptron artificial neural network (MLP-NN), trained on a small data partition which is not included in the test phase. NN-based methods have been applied widely to classification problems because of their high learning capacities and good generalization [2]. In the present study, two hidden layers NN were trained using back-propagation. The architecture of the NN was as follows: the first hidden layer had the same number of neurons as the number of individual scores derived from the matching stage, i.e. 4 for Method 1 and 11 for Method 2; the second hidden layer had 2 and 6 neurons for Methods 1 and 2 respectively, while the final (output) layer presented a single layer outputting the final s_f value.

Besides these strategies, we also analyzed the effect of fusing information from different channels of the *RGB* colorspace. Integration of information, in this case, was performed by treating each color channel individually and computing three independent recognition scores for each one: r_R, r_G and r_B. The final score was obtained by simple averaging these three values.

4 Results and Discussion

The present section will serve as a detailed analysis of the comparison carried out between the two algorithms presented in the last section. We start by offering some insight regarding the specific details of the multi-sensor periocular database on which both algorithms were assessed, as well as the experimental setups and performance metrics used for such assessment. We then present and discuss the main results regarding both recognition performance as well as processing time and possible limitations and advantages of each algorithm in real-world scenarios.

4.1 CSIP Database

The CSIP database, created for the assessment of the original version of Method 2 [2], is a recent and publicly available dataset, designed with the main goal of gathering periocular images from a representative group of participants, acquired using a variety of mobile sensors under a set of variable acquisition conditions. Given the heterogeneity of the camera sensors and lens setups of consumer mobile devices, 10 different setups were used during the dataset acquisition stage: four different devices, some of which had both frontal and rear cameras, and LED flash. This variety of sensors confers a strong appeal to the CSIP database regarding its potential use for the assessment of algorithms under a highly heterogeneous set of conditions. A visual example of an image for each subset of the same individual is depicted in Fig. 3. Each participant was imaged using all of the test setups.

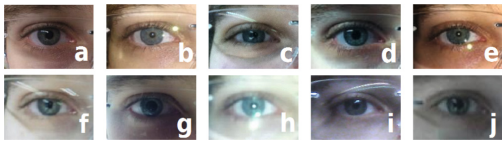

Fig. 3. Examples of images from each subset of the CSIP database. From (a-j) respectively: $AR0$, $AR1$, $BF0$, $BR0$, $BR1$, $CF0$, $CR0$, $CR1$, $DF0$ and $DR0$.

To simulate the variable noise associated with on-the-go recognition, participants were not imaged at a single location, but instead they were enrolled at multiple sites with artificial, natural, and mixed illumination conditions. In total, 50 participants were enrolled, all Caucasian and mostly males (82 %), with ages ranging between 21 and 62 years (mean = 31.18 ± 9.93 years). For each periocular image acquired by the mobile devices, a binary iris segmentation mask was also produced. The masks were obtained automatically using the state-of-the-art iris segmentation approach proposed by Tan et al. [12], which is particularly suitable for uncontrolled acquisition conditions, as demonstrated by its first place ranking at Noisy Iris Challenge Evaluation - Part 1 (NICE.I) (NICE.I) [15].

4.2 Experimental Setup

In order to achieve a fair comparison between both tested algorithms, a uniform experimental setup was defined and adapted to fit the specifics of each method. With that in mind, the set of all images of the CSIP dataset was divided as follows: 50 % of the images per individual and per subset were kept to either train the models in Method 1 or to serve as reference for each identity in Method 2; the remaining 50 %, apart from a small independent set used to train the fusion neural networks, were used to assess the performance of both methodologies. Performance assessment was adapted to fit the nature of the originals algorithms:

- **Method 1:** given an input image I of an unknown source and an associated identity claim S, the score, r_s, associated with this image/claim pair is computed by the likelihood-ratio, $r_s = \frac{proj(desc(I), IDSM_S)}{proj(desc(I), IDSM_{UBM})}$, where $desc(I)$ represents the feature descriptor extracted from image I and $proj(X, GMM)$ represents the projection of feature vector X onto a specific GMM (either the claimed ID's IDSM or the UBM). This process is repeated for every possible ID, so that for image I the assessment block outputs a total of N scores, with N being the total number of individuals enrolled in the database.
- **Method 2:** given an input image I of an unknown source and an associated identity claim S, the score, r_s, associated with this image/claim pair is computed by the averaging of image/image pair similarities, $r_s = \frac{\sum_{k=1}^{n_{ID}} score(I, I_{ID,k})}{n_{ID}}$, where $score(I, I_{ID,k})$ is the comparison score obtained using Method 2 between the unknown image I and the k-th reference image from a specific known ID. The averaging is made in relation to the total number of reference images for the given ID, n_{ID}. This process is repeated for every possible ID, so that for image I the assessment block outputs a total of N scores, with N being the total number of individuals enrolled in the database.

After the N scores are extracted for each image, using both methodologies, performance is assessed for either identification or verification modes. For identification we chose to use the *rank-1 recognition rate* metric, which represents the rate of images for which the highest of the N recognition scores corresponded to the true ID. On the other hand, for verification, we computed the *equal error rate*. This value corresponds to the error rate observed when a specific acceptance threshold is applied to the recognition scores and the resulting false positive and false negative rates are equivalent.

4.3 Performance Comparison

The main results obtained for the setups outlined in the previous sections are summarized in Table 1. All results concern the average performance observed with 10-fold cross-validation for a specific methodology (GMM-UBM or Santos et al. [2]), pre-processing strategy (device specific color correction - DS-CC - or discrete cosine transform - DCT), and single feature (LBP, SIFT, GIST or HOG) or fusion of multiple features (performace-weighted sum-rule - PW-SR - or multilayer perceptron artificial neural network - NN).

Careful observation of the values presented in Table 1 allows for some interesting conclusions to be achieved. Regarding both tested metrics, there is no significantly better algorithm for all the tested subsets. While the UBI algorithm achieves the best average identification performance (Table 1) amongst the whole set of tested subsets, the UBM algorithm still manages to achieve significantly better performances for two of such subsets - $CF0$ and $DR0$ - while managing to achieve values in a very similar range for five other subsets - $AR0$, $AR1$, $BR0$, $BR1$ and $CF0$. In fact, only for the $BF0$, $CR0$ and $DF0$ subsets does the difference in performance between the two algorithms become significant. This non-uniformity in the relative behaviour of performance between the two methodologies might indicate that even though the sources of information

Table 1. Rank-1 recognition rates obtained for each subset of the CSIP database for some variations of both tested methodologies.

				CSIP Subset									
	Pre-Proc.	Feat(s)	Trait(s)	AR0	AR1	BF0	BR0	BR1	CF0	CR0	CR1	DF0	DR0
M.1	DS-CC	PW-SR	P	88.4	97.8	73.5	86.5	93.1	85.0	72.1	92.1	45.5	81.7
	DCT	PW-SR	P	94.9	97.7	75.0	86.5	91.7	91.3	83.7	93.2	55.5	**91.7**
r_1(%)	DCT	RGB PW-SR	P	**97.4**	**100**	83.8	93.2	95.8	**92.5**	82.6	94.3	50.0	95.0
	DCT	NN	P	94.3	99.4	83.8	89.3	96.8	84.2	79.8	94.2	54.8	82.9
M.2	DS-CC	LBP	P	89.8	**100**	75.6	81.7	91.8	75.3	79.6	96.8	59.4	72.0
	DS-CC	SIFT	P	81.8	**100**	73.2	84.2	**100**	61.8	61.3	96.8	82.8	57.3
	DS-CC	GIST	P	96.6	**100**	87.8	93.0	98.8	84.3	**92.5**	94.6	75.0	84.0
r_1(%)	DS-CC	HOG	P	68.2	92.6	56.1	64.6	72.9	42.7	55.9	78.5	40.6	54.7
	DS-CC	NN	P+I	95.5	**100**	**92.7**	**95.3**	95.3	85.4	90.3	97.9	**76.6**	80.0
	DS-CC	PW-SR	P+I	93.2	**100**	90.2	91.4	98.8	87.6	81.7	**98.9**	73.4	82.7
M.1	DS-CC	PW-SR	P	3.3	3.0	8.7	**4.0**	1.4	6.2	8.0	4.5	17.4	7.5
	DCT	PW-SR	P	2.5	2.9	7.9	4.3	1.7	4.6	**5.1**	4.2	**14.8**	5.0
r_1(%)	DCT	RGB PW-SR	P	**1.7**	**0.5**	7.5	4.4	1.5	**4.5**	5.3	4.0	19.3	**4.1**
	DCT	NN	P	2.1	0.7	**7.3**	5.3	1.3	5.9	9.6	4.0	24.0	11.3
M.2	DS-CC	LBP	P	9.6	4.3	14.7	17.4	10.1	13.5	11.0	5.7	23.9	16.0
	DS-CC	SIFT	P	9.9	0.6	16.5	14.5	**0.1**	16.0	19.2	**2.1**	28.3	18.1
	DS-CC	GIST	P	5.2	3.3	11.1	12.0	6.3	10.3	8.9	5.3	20.2	12.0
r_1(%)	DS-CC	HOG	P	17.1	7.4	24.4	23.7	13.6	21.8	20.8	11.8	28.9	22.5
	DS-CC	NN	P+I	6.3	0.6	9.1	11.7	7.1	10.1	9.7	4.0	19.1	13.1
	DS-CC	PW-SR	P+I	7.5	0.8	10.4	12.2	5.9	9.3	10.5	3.5	20.5	12.6

and the feature descriptors used for its encoding are very similar, the modeling strategies used in both works adapt better for some acquisition scenarios. By visual observation of the images it is readily understandable that the $BF0$ and $DF0$ images are the ones that present lower resolution and overall image quality (Fig. 3). The UBM modeling strategy might, therefore, not be able to aptly train GMMs capable of correctly and in a robust way describe such low quality data. As the UBI data uses direct matching algorithms that are optimized for each feature descriptor, the results in low quality data might be improved. On the other hand, when the quality of input images is a bit higher, as far as images acquired with mobile devices can go, the UBM algorithm either reaches the same ranges of the UBI results, or even exceeds it for moderately more complicated scenarios such as $CF0$ and $DR0$. A different set of observations can, however, be carried out by the analysis of the EER values used to assess performance in verification scenarios (Table 1). Here the UBM algorithm consistently achieves better performance regardless of the tested subset. This variable behavior might indicate that while the UBI algorithm present a higher discriminative power between individuals, the UBM algorithm, probably due to the score normalization effect inherent to the method, is more fit to distinguish between classes (genuine and impostor users) in an identity check application.

A behavior that is easily observable, regardless of the methodology and subset that we choose to focus on, is that images acquired with flash illumination present considerably better results than their non-flash counterparts. This observation was somewhat expected, as flash illumination might serve as a solution to overcome the variable lighting conditions that were referred as a natural limitation of mobile device acquisition in Sect. 1. To the extent of our knowledge no ocular health problems are commonly associated to overexposure to flash illumination in mobile devices, and with the growing technological advances in the manufacturing of such devices, flash illumination might play a crucial role in the implementation of image-based biometrics in mobile environments in the near future. Concerning pre-processing, Table 1 shows that DCT normalization far exceeds the performance obtained with the device-specific color-correction proposed in the original work with the CSIP database. Using a fixed transformation rather than a device-specific approach, that relies on the definition of new transformation matrices for each new device, presents a more robust and reliable alternative as far as the integration of periocular recognition in real-life applications is concerned. Furthermore, it can be seen that the results obtained using information from the three available color channels also results in a non-negligible increase in performance for a variety of subsets. Even though the recognition performance is increased, it must be noted that processing time is increased three times, as the same algorithm must be run in three separate instances. Even though these instances could be ran in parallel, the technological burden for such approach in mobile devices might exceed the current limits.

One more topic to take into consideration regards the alternative fusion strategies that were tested. From Santos et al.s results, it can be observed that the use of neural networks over a simpler performance-weighted approach, results mostly in non-significant variations in the performance. A similar set of conclusions can be drawn for the GMM-UBM results.When comparing this results with the ones obtained for individual features, however, the positive effect of fusion, regardless of its details, is readily discernible. The choice of the fusion strategy should, therefore, be constrained by the specific scenario of application and on how each strategy performs.

On a final note, regarding the processing time of each tested methodology, some considerations can be taken. Given an unknown image I and a single identity check, either by likelihood-ratio or average image similarity, the single-image processing time was computed and averaged for all test images. It was observed that Method 1 using DCT normalization and performance-weighted sum-rule fusion spent an average of 0.018 seconds on this process, whereas Method 2 spent an average of 0.130 seconds for an analogous computation, using device-specific color correction and neural network fusion. This discrepancy is easily explained by the larger amount of features used by Method 2 (11 vs. 4). Real-life applications based on periocular recognition are expected to work as fast as possible, so as to accurately replicate real-time operation. With that in mind, the UBM approach, with its uniform and fast matching algorithm based on GMM projections, seems to present an interesting alternative for further research. Even though the performance obtained for more unconstrained scenarios (for examples $DF0$) is

still far from acceptable, future work on more robust representations or future improvements to the intrinsic architecture of the algorithm might help overcome such limitations. The same thought is applicable to the UBI approach, where a more efficient matching strategy might bring about considerable decreases in processing time, with no significant nefarious effect over performance.

4.4 Cross-Sensor Recognition

Some preliminary experiments were also carried out in cross-sensor scenarios. In this alternative to the biometric recognition problem, enrollment and recognition are carried out using data acquired using different sensors. Some of the most interesting results obtained were that, for both methods, only some specific setups showed considerable interoperability. For example, using the GMM-UBM approach, and using the AR1 set for training and the BR1 set for testing, a recognition rate of 86.1 % was achieved, whereas the single setup scenarios yielded 100 % and 95.8 % respectively. This relatively small loss in performance, when compared to other cross-sensor scenarios, might relate to similarities in the hardware of the rear cameras of devices A and B, as well as to the more uniform conditions in lighting, as a result of flash illumination. A similar behavior was observed for a few other pairs of setups - $BR1/CR1$, $AR0/BR0$ and $BR0/DR0$ for example - but, in general, a signifi- cant drop in performance is observed for cross-sensor scenarios, regardless of the tested methodology. Further research is, therefore, needed to achieve stability in performance when enrollment and testing are carried out in highly variable acquisition conditions.

5 Conclusions and Future Work

In the present work we assess the performance of two extended versions of state-of-the-art algorithms for periocular recognition in mobile devices. We extend Monteiro et al. to multiple feature representations and score-level fusion, and adapt Santos et al. to an alternative fusion strategy. The comparative analysis of both approaches shows that, depending on the specific real-world application for which a system is developed, each algorithm presents its advantages and disadvantages. Santos et al. is more fit to identification problems with less restrictions concerning processing time, presenting high performance for a wide variety of noise factors. On the other hand, the GMM-UBM approach presents a faster matching time, with better performance in verification scenarios. Regarding mobile device applications, it would be interesting to explore a joint methodology that managed to keep the fast matching step from Method 1, while achieving good performance in both identification and verification. Further research on both methodologies could lead to interesting improvements. It would also be of relevance to assess how the presented methodologies behave, in their current state, when implemented in existing mobile devices, so as to better understand their current limitations in more realistic scenarios. Another interesting focus

of future work would concern on the improvement of performance in the cross-sensor scenario presented in Sect. 4.4, so that enrollment in a single device could serve for recognition purposes in multiple environments.

Acknowledgments. 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. The second and fifth authors would like to acknowledge the financial support obtained from North Portugal Regional Operational Programme (ON.2 - O Novo Norte), Portuguese National Strategic Reference Framework (NSRF) and the European Regional Development Fund (EDRF) from European Union through project ICT4DCC (NORTE-07-0124-FEDER-000042.). The third and fourth authors would like to acknowledge the financial support provided by FCT through the research grant SFRH/BD/80182/2011, and the RSU - Remote Sensing Unit through PEst-OE-FIS/UI0524/2014.

References

1. Bakshi, S., Kumari, S., Raman, R., Sa, P.K.: Evaluation of periocular over face biometric: A case study. Procedia Eng. **38**, 1628–1633 (2012)
2. Santos, G., Grancho, E., Bernardo, M.V., Fiadeiro, P.T.: Fusing iris and periocular information for cross-sensor recognition. Pattern Recogn. Lett. **57**, 52–59 (2015)
3. 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)
4. 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)
5. Park, U., Jillela, R.R., Ross, A., Jain, A.K.: Periocular biometrics in the visible spectrum. IEEE Trans. Inf. Forensics Secur. **6**, 96–106 (2011)
6. Padole, C.N., Proenca, H.: Periocular recognition: Analysis of performance degradation factors. In: 5th IAPR International Conference on Biometrics, pp. 439–445 (2012)
7. 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: 2012 5th IAPR International Conference on Biometrics (ICB), pp. 446–453. IEEE (2012)
8. Joshi, A., Gangwar, A.K., Saquib, Z.: Person recognition based on fusion of iris and periocular biometrics. In: 2012 12th International Conference on Hybrid Intelligent Systems (HIS), pp. 57–62. IEEE (2012)
9. Tan, C.W., Kumar, A.: Towards online iris and periocular recognition under relaxed imaging constraints. Image Process. IEEE Trans. **22**, 3751–3765 (2013)
10. Santos, G., Proença, H.: Periocular biometrics: An emerging technology for unconstrained scenarios. In: 2013 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM), pp. 14–21. IEEE (2013)
11. Reynolds, D., Quatieri, T., Dunn, R.: Speaker verification using adapted gaussian mixture models. Digital Signal Process. **10**, 19–41 (2000)
12. Tan, T., Zhang, X., Sun, Z., Zhang, H.: Noisy iris image matching by using multiple cues. Pattern Recogn. Lett. **33**, 970–977 (2011)

13. Daugman, J.: High confidence visual recognition of persons by a test of statistical independence. IEEE Trans. Pattern Anal. Mach. Intell. **15**, 1148–1161 (1993)
14. Chen, W., Er, M.J., Wu, S.: Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. IEEE Trans. Syst. Man Cybern. Part B: Cybern. **36**, 458–466 (2006)
15. Proença, H., Alexandre, L.: The nice.i: noisy iris challenge evaluation - part i. In: First IEEE International Conference on Biometrics: Theory, Applications, and Systems, pp. 1–4. IEEE (2007)