Periocular Biometrics in Mobile Environment

Tiago de Freitas Pereira, Sébastien Marcel
IDIAP Research Institute
hhtp://www.idiap.ch
tiago.pereira@idiap.ch, sebastien.marcel@idiap.ch

Abstract

In this work we study periocular biometrics in a challenging scenario: a mobile environment, where person recognition can take place on a mobile device. The proposed technique, that models session variability, is evaluated for the authentication task on the MOBIO database, previously used in face recognition, and on a novel mobile biometric database named the CPqD Biometric Database, as well as compared to prior work. We show that in this particular mobile environment the periocular region is complementary to face recognition, but not superior, unlike shown in a previous study on a more controlled environment. We show also that a combination with face recognition brings a relative improvement of 7.93% in terms of $HTER$. Finally, the results of this paper are reproducible using an open software and a novel Web platform.

1. Introduction

Face recognition in mobile environment introduces many variations such as pose, illumination, expression and occlusion. Pose variation appears, when the user holds the mobile phone in different ways during the acquisition. Illumination variation occurs because the acquisition can take place indoor (with artificial illumination) or outdoor (with natural illumination). Facial expression variation happens when the user is speaking. Occlusion variation also likely happens due to artefacts (scarf, sunglasses) or to parts of the face which are not visible (the user is not properly presenting its face). Since the periocular region is a small portion of the whole face, this part can be suitable for recognition in such environment.

When the periocular region was firstly introduced by Park et al. [13], several studies in the literature demonstrated that the region itself can be used for recognition or can complement both face [6] and iris recognition [4] to reduce the error rates.

The contribution of this paper is two-fold: First we present a novel approach based on Inter-Session Variability Modelling [10] to recognize people based on the periocular region and evaluated in a very challenging condition (mobile devices). To the best of our knowledge, this evaluation was never done before. Next, we provide an entire reproducible experimental procedure. Results of this paper are available for reproducibility using the BEAT (Biometrics Evaluation and Testing) platform[2][1].

The organization of the paper is the following. In Section 2 we present the prior work for periocular recognition. In Section 3 the proposed approach is presented in details. In Section 4 the experimental setup and results are presented. Finally in Section 5 the conclusions are presented.

2. Related work

As aforementioned, [13] presented the first evaluation of the periocular modality. Their work relied on three different features for three different periocular recognition systems: Local Binary Patterns histograms (LBP) [1], Scaled-invariant Features Transform ($SIFT$) [8] and Histograms of oriented Gradients (HoG). As a distance measure, euclidean distance was used to match the systems based on LBP and HoG. To match the $SIFT$ based system, the distance ratio based on [8] was used. The systems were evaluated on a private database, made specifically for this study. They achieved an accuracy of 80% in Rank identification combining those three features.

Miller et al. [11], presented an evaluation of the image quality assessment of the periocular region. They introduced different types of noise in the FRGC database convolving the images with different parametrizations of Gaussian filter (varying the kernel and the variance). They used LBP histograms as features and city block distance as a distance measure. Comparing the error rates of the periocular region with face recognition under the same conditions, they concluded that the periocular region is more robust than face when noise is introduced showing a relative improvement of 185% in the most noisy scenario.

Bharadwaj et al. [4] proposed an approach based on

[1] https://www.beat-eu.org/platform/

shape and texture features to complement the iris recognition. For the shape, they used five perceptual dimensions (image naturalness, openness, roughness, expansion and ruggedness) and for the texture, LBP histograms. As distance measure, the chi-square distance was used for both approaches. The final score was a fusion (score-level) between both systems. The evaluation, conducted with the UBIRIS database, which was designed to be an iris at-a-distance reference, shown an accuracy of $\sim 70\%$ in Rank-1 identification. Also trying to complement an iris recognition system, Tan et al. [15] encoded the periocular with several features (LBP histograms, HoG, $SIFT$, $GIST$ [12] and LMF (Leung-Malik Filters) [7]). Both systems (iris and periocular), were analysed alone and combined with fusion of scores presenting a relative improvement of 132.3% on the UBIRIS database.

It is important to highlight that the source codes of the aforementioned papers were not made explicitly public by the authors. Also given the lack of public database for the work in [13] it is not possible to reproduce the prior work directly.

3. Proposed approach

The next subsection explains in details each step of the proposed approach.

3.1. Image preprocessing and feature extraction

For this work, each image is cropped, rotated and registered to 58×25 pixels with 33 pixels of eyes distance (based on manually eyes annotations). To suppress the illumination effects we apply the Tan & Triggs pre-processing [16]. Each pre-preprocessed periocular image is sampled in blocks of 12×12 pixels moving the sampling window in one pixel (see Figure 1). Each block is mean and variance normalized and the first 45 2D-DCT coefficients are extracted. For each block the first coefficient (DC component) is discarded resulting in a feature vector with 44 elements per block.

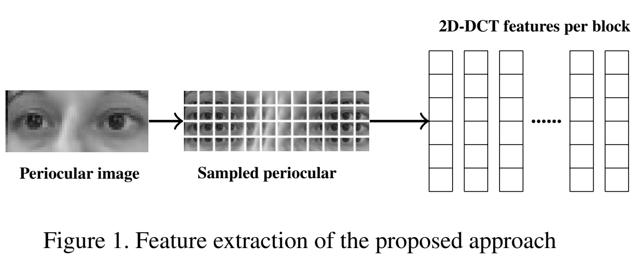

Figure 1. Feature extraction of the proposed approach

3.2. Inter-session variability modelling (ISV)

In this work we propose to model the feature distribution of the periocular region using Gaussian Mixture Models (GMM)[14]. A GMM is a weighted sum of C multivariate gaussian components:

$$p(o|\Theta_{gmm}) = \sum_{c=1}^{C} w_c \mathcal{N}(o; \mu_c, \Sigma_c), \tag{1}$$

where $\Theta_{gmm} = \{w_c, \mu_c, \sigma_c\}_{\{c=1...C\}}$ are the weights, means and the covariances of the model.

Built on top of GMMs, ISV proposes to model the variations between different sessions of the same client[2] and compensate them during the enrolment and testing time. For instance, session variability refers to variations between images of the same person. The issues discussed in Section 2 regarding pose, illumination and expression are considered session variations.

Basically, ISV assumes that the session variability is an additive offset to the mean supervector space combined with the client specific offset. For the given GMM in the equation 1, the mean supervector m is the concatenation of the means of all gaussian components $[\mu_{c=1}^T, \mu_{c=2}^T ... \mu_{c=C}^T]$.

At **training time**[3] (made offline), the 2D-DCT features described in the last section are used to train an Universal Background Model (UBM)[14], which is a GMM trained with the data of the whole training set. Basically this UBM describes how the features of the training set are distributed. The mean supervectors derived from this UBM (a maximum a posteriori adaptation of the whole training set) are used to estimate the within class variation matrix, so called U matrix.

Given the j_{th} periocular sample $\mathcal{O}_{i,j}$ of the client i, the mean supervector $\mu_{i,j}$ of the GMM that best represents the sample is:

$$\mu_{i,j} = m + Ux_{i,j} + Dz_i, \tag{2}$$

where m is th UBM, U is the subspace that contains all possible session effects, $x_{i,j}$ is its associated latent session variable ($x_{i,j} \sim \mathcal{N}(0, I)$), while Dz_i represents the client offset.

At **enrolment time**, the model for the client i is obtained by estimating $x_{i,j}$ and z_i. The effect of the session variability for each periocular image ($Ux_{i,j}$ in (2)) is then excluded from the client model. In the end, the final client model is:

$$s_i^{ISV} = m + Dz_i \tag{3}$$

At **test time**, the scoring is defined as the log-likelihood ratio (LLR) between the client model and the UBM. Given a set of observations $\mathcal{O} = \{o_1....o_T\}$ claimed to be of the client i, the LLR is defined as follows:

[2] In an authentication scenario, clients are the users to recognize while impostors are the ones to reject.

[3] Here training does not mean enrolment, but means the construction of the prior model used for the enrolment (see Section 4.3).

$$h_{ISV}(\mathcal{O}|s_i^{ISV}) = \sum_{t=1}^{T}[ln(\frac{p(o_t|s_i^{ISV} + Ux_{i,j})}{p(o_t|m + Ux_{UBM,j})})] \quad (4)$$

A full derivation on how the U matrix, the latent variable $x_{i,j}$ and the client offset z_i are estimated can be found on [10].

To train the U matrix we use 512 Gaussian Components and we set to 160 the dimension of the session factor. These parametrizations were tuned using the development set of MOBIO database (see Section 4.2). In total the U projection matrix has dimension $[512*44 \times 160] = [22528 \times 160]$. The first dimension of the matrix is the number of gaussians components multiplied by the dimension of the feature vector (see Section 3.1).

4. Experiments

This section describes the experimental procedure that was used to evaluate our proposed approach in the claimed scenario. The section is organized as follows. First we describe the reproducible research procedure implemented in this work. In sub-section 4.2 the databases used in this work are presented. The sub-sections 4.3 and 4.4 present, respectively, the evaluation protocol and the performance measures applied. Sub-section 4.5 presents the baselines used for comparison with our approach. The sub-sections 4.6 and 4.7 present our results in the aforementioned databases. Finally, sub-section 4.8 discusses the results.

4.1. Reproducibility procedure

Experiments in this paper were conducted using the Bob library [3] as well as using the BEAT platform. The BEAT platform is a free, open-source, online web-based platform for the development and attestation of software-based biometric systems. With this platform, it is possible to inspect the source code of the proposed approach and regenerate the obtained results. Furthermore, it is possible to reset the hyper-parameters of the algorithms to execute different experiments.

4.2. Databases

We choose two databases for our particular scenario (faces collected with mobile phones). The first one is the MOBIO database [9]. The second one is the CPqD Biometric Database [17]. Both databases are available on the BEAT platform for reproducible research purposes.

The **MOBIO database** is a bi-modal (face/speaker) video database recorded from 152 people. The database has a female-male ratio of nearly 1:2 (100 males and 52 females) and was collected from August 2008 until July 2010 in 6 different sites from 5 different countries. In total 12 sessions were captured for each individual.

The database was recorded using two types of mobile devices: mobile phones (NOKIA N93i) and laptop computers (standard 2008 MacBook). In this paper we only use the mobile phone data. The MOBIO database is challenging since the data are acquired with uncontrolled illumination (see Figure 2), facial expression, and face pose, and sometimes only parts of the face are visible.

Figure 2. Samples from the MOBIO database

For biometric recognition, one image was extracted from each video by choosing the video frame after 10 seconds. The eye positions were manually labelled and distributed with the database.

More technical details about the MOBIO database can be found in [9] and on its official web page[4], which also contains instructions on how to obtain the data. Based on the gender of the clients, two evaluation protocols are available: **female** and **male**.

The **CPqD biometric database** is also a bi-modal (face/speaker) video database recorded from 222 people (128 males and 98 females). In total 5 sessions were captured and for each one 27 recordings were made with 3 different devices: laptops, smartphones and phones callings (this one only audio).

For the laptops, three different devices were used: a Compaq 510 with built-in camera and microphone, a Dell Latitude using the camera (with built-in microphone) Logitech QuickCam Pro 9000 and a Dell Latitude with built-in camera and microphone. For the smartphones, a Samsung Galaxy S II and a Apple iPhone 4 were used. Phone calls were done with land line phones in the office and with the mobile phones of the participants when the recording session was outside the office. As it can be observed in Figure 3 the database is quite challenging in terms of illumination.

The recordings were done in three different environments. **Office**: in that environment the illumination is artificial and w.r.t audio recordings, the background noise is moderate. **Garden**: in that environment the illumination is natural and has a lot of variations because of the weather and w.r.t audio recordings, the background noise is moder-

[4]http://www.idiap.ch/dataset/mobio

ate as well. **Restaurant**: in that environment the illumination is artificial and there are lots of people passing by the recording. W.r.t the audio recordings the environment is very noisy since the sessions were taken when the restaurant was crowded.

As in the MOBIO database, for this experiment one image was extracted from each video by choosing 40th frame. The eye positions were manually labelled and distributed with the database. Furthermore the CPqD Biometric Database has two evaluation protocols based on the gender of the clients: **female** and **male**.

Figure 3. Samples from the CPqD Biometric Database; respectively in the **office**, **garden** and **restaurant**

4.3. Evaluation protocol

In order to have an unbiased evaluation, the clients of both databases are split up into 3 different sets:

Training set The files from training set can, e. g., be used to compute projection matrices for PCA or to compute the UBM and the U matrix (this is the case for our approach). For some systems, these images also serve as the negative examples in a classification system. Furthermore, these images can be used as cohort images for score normalization.

Development set (dev) The development set can be used to choose the meta-parameters of the algorithms, e. g., the dimension of a PCA projection matrix and the distance function. For each client in this set, the images are subdivided into the enrolment and the probe group. For the MOBIO database **five files** per client are provided for enrolment and for the CPqD Biometric Database, only **one** is provided for that purpose.

Evaluation set (eval) The evaluation set is used to assess the final unbiased system performance. The subdivision into enrollment and probe groups is similar to the one in the development set and also the model enrollment uses similar data. However, individual subjects are not the same between each set.

4.4. Performance measures

The performance on the test set of each database is reported with the Half Total Error Rate ($HTER$):

$$HTER = \frac{FAR(\tau) + FRR(\tau)}{2}, \tag{5}$$

where $FAR(\tau)$ and $FRR(\tau)$ are, respectively, the False Acceptance Rate and the False Rejection Rate[5]. Both rates are computed as a function of τ which is the threshold corresponding to the Equal Error Rate (EER) on the development set of each database. We present also Detection Error Trade-off (DET) curves.

4.5. Baselines for comparison

As mentioned in Section 2, it is not possible to reproduce directly prior work. To provide a baseline system for comparison with our proposed approach, we re-implemented one of the systems described in [13] and [11]. The baseline system is based on LBP histograms. The LBP bins are computed and histogrammed in the periocular area of 58×25 (see Section 3). As a similarity measure, the chi-Square distance is computed.

Additionally to the baseline provided, we also compare the performance of our approach with a face recognition system. The face recognition system used is the one published in [10], which inspired our approach for periocular recognition using session variability modelling and which is one of the best performing system on the MOBIO database.

As mentioned in Section 1, we hypothesize that the periocular region can be used alone or combined with face recognition. Hence, we also perform a score level fusion between periocular and the above mentioned face recognition system. Such fusion is simply achieved with linear logistic regression between the scores of both modalities.

4.6. Results on the MOBIO Database

It is possible to observe in Table 1 and in Figure 4 that the result of the proposed periocular recognition approach is far better than the baseline, based on LBP histograms. For instance, the $HTER$, in the development set, decreases from 24.24% to 9.37% in the protocol male and from 26.19% to 9.37% in the protocol female.

Also, comparing the results with our approach, on periocular, with the face recognition baseline it is possible to observe a drop in terms of $HTER$ and in the DET plot. The $HTER$, in the development set, increases from 3.69% to 9.37% on the protocol male. The same behaviour can be observed for the protocol female.

The fusion between face and periocular decreases slightly the error rates in comparison with face recognition. For instance, the $HTER$ decreases from 3.69% to 3.53% on the protocol male and from 6.56% to 5.93% on the protocol female. The results obtained in the evaluation set follow the same trends observed in the development set.

These results are attested for reproducibility in the BEAT

[5] In this work the False Acceptance Rate (FAR) is synonym of False Match Rate (FMR) and False Rejection Rate (FRR) is synonym of False Non Match Rate ($FNMR$).

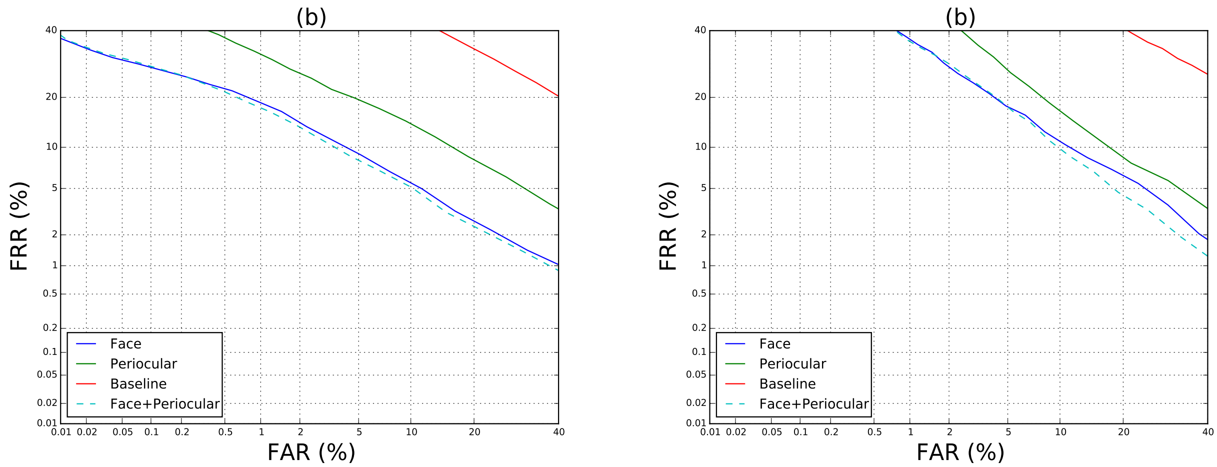

Figure 4. DET Curves in the test set of MOBIO database: (a) protocol male (b) protocol female

Table 1. $HTER(\%)$ of periocular recognition on MOBIO and CPqD Biometric Database

Modality	**MOBIO**				**CPqD Biometric database**			
	protocol male		protocol female		protocol male		protocol female	
	dev	eval	dev	eval	dev	eval	dev	eval
Face	3.69%	7.03%	6.56%	11.36%	3.96%	4.37%	7.45%	6.32%
Periocular	**9.37%**	**12.27%**	**9.37%**	**13.21%**	**7.77%**	**7.27%**	**13.77%**	**13.67%**
Fusion	3.53%	6.58%	5.93%	10.73%	3.24%	3.87%	6.96%	6.41%
LBP as [13] and [11]	24.24%	27.48%	26.19%	30.57%	22.55%	26.39%	26.78%	32.89%

platform in this link[6], for the experiments using the protocol male, and in this link[7], for the experiments using the protocol female.

4.7. Results on CPqD Database

For the CPqD Biometric Database, it is possible to make (based on Table 1 and Figure 5) the same observations made in the last section. The proposed approach for periocular recognition is far better than the baseline in both defined protocols (male and female). For instance, the $HTER$ decreases from 22.55% to 7.77% in the protocol male and from 26.78% to 13.77% in the protocol female (comparisons done in the development set).

Comparing the results with our approach, on periocular, with the face recognition baseline it is possible to observe a drop in terms of $HTER$ and in the DET plot. For instance, the $HTER$, in the development set, increases from 3.96% to 7.77% on the protocol male. The same behaviour can be observed for the protocol female.

The fusion between face and periocular decreases slightly the error rates, in the development set, in comparison with face recognition baseline (from 3.96% to 3.24%) on the protocol male. As before, the same behaviour can be observed on the protocol female. The results obtained in the evaluation set follow the same trends observed in the development set.

These results are attested for reproducibility in the BEAT platform in this link[8], for the experiments using the protocol male, and in this link[9], for the experiments using the protocol female.

4.8. Discussion

In [11], the authors concluded that the periocular region is more robust than the face region when noise (artificial) is introduced (see Section 2). However, this is not the case with the databases considered in this work, where the acquisition conditions are very different than the FRGC database used in [11]. The only assumption that we can support is that the periocular region can complement the face recognition. Comparing the face recognition baseline with the fusion (face recognition and periocular recognition), we obtain a relative average improvement, in $HTER$ terms, of 9.67% and 6.2% respectively for the development and evaluation sets (over the two databases) . In total, averaging the

[6] https://www.beat-eu.org/platform/reports/751803513/
[7] https://www.beat-eu.org/platform/reports/429641009/
[8] https://www.beat-eu.org/platform/reports/643236741/
[9] https://www.beat-eu.org/platform/reports/1952076162/

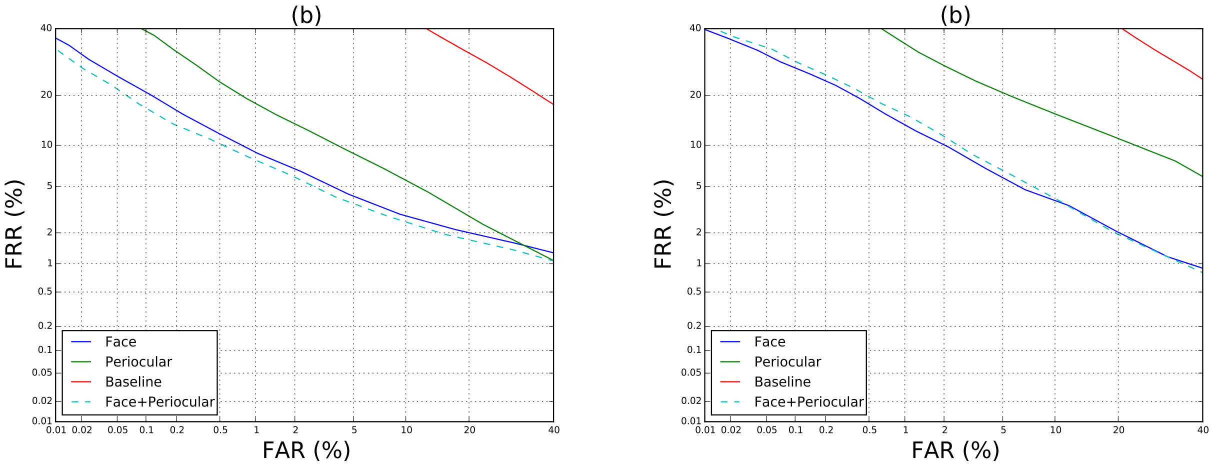

Figure 5. DET curves in the test set of CPqD Biometric database: (a) protocol male (b) protocol female

two sets, this improvement is 7.93%. This is particularly interesting since the periocular region is a subset of the face region and once the feature extraction is already done, for the face recognition, the features for the periocular recognition are ready for usage. Hence, there is no much more computational effort to use the periocular region as a complementary step of the authentication procedure.

It is possible to observe (using the fusion as a reference) that the results for the CPqD Biometric Database are slightly better than the results for the MOBIO Database in both protocols (male and female). We obtain an $HTER$ of 6.58% for the MOBIO Database and 3.87% for the CPqD Biometric Database on the protocol male (evaluation set). For the protocol female we obtain an $HTER$ of 10.73% for the MOBIO Database and 6.41% for the CPqD Biometric Database (evaluation set). Although both databases are quite similar in terms of scenario (both captured in mobile device use case) and in terms of protocols (female and male), the range of performance is very different. We can hypothesize this is related to the camera technology. The MOBIO database was captured with a Nokia N93i (released in 2007) and the CPqD Biometric Database was captured with the Samsung Galaxy S2 (released 2011) and Apple iPhone 4 (released 2010). For each release of new mobile phones, manufacturers increase the quality of the sensor technologies (resolution, automatic focus, etc), hence we can reasonably assume this is playing an important role for the visual recognition.

5. Conclusion

In this paper we presented a novel approach to recognize people using periocular biometrics in mobile environment. Our approach, based on Inter-Session Variability Modelling, proposes to model the variations (pose, illumination, expression) between different sessions of the same client and compensate them during the enrolment and testing. With experiments conducted using the MOBIO Database and the CPqD Biometric Database, we showed that the periocular region is complementary to face recognition, but not superior, unlike in a previous study. A combination with a face recognition system showed a relative improvement of 7.93% in terms of $HTER$.

It is important to emphasize, that this paper is entirely reproducible using the BEAT platform. With that platform, it is possible regenerate the obtained results and also it is possible to reset the hyper-parameters of the algorithms to execute different experiments.

6. Acknowledgment

The development of this work has received funding from the Swiss National Science Foundation (SNSF) under the HFACE project and from the European Community's Seventh Framework Programme (FP7) under grant agreement 284989 (BEAT) and the Swiss Center for Biometrics Research and Testing.

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *Computer vision-eccv 2004*, pages 469–481. Springer, 2004.

[2] A. Anjos and S. Marcel. Reproducible biometrics evaluation and testing with the beat platform. In *International Biometrics Performance Conference*. NIST, 2014.

[3] A. Anjos, L. E. Shafey, R. Wallace, M. Günther, C. McCool, and S. Marcel. Bob: a free signal processing and machine learning toolbox for researchers. In *20th ACM Conference on Multimedia Systems (ACMMM), Nara, Japan*. ACM Press, Oct. 2012.

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

[5] F. Cardinaux, C. Sanderson, and S. Bengio. Face verification using adapted generative models. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 825–830. IEEE, 2004.

[6] F. Juefei-Xu, K. Luu, M. Savvides, T. D. Bui, and C. Y. Suen. Investigating age invariant face recognition based on periocular biometrics. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–7. IEEE, 2011.

[7] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44, 2001.

[8] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[9] C. McCool, S. Marcel, A. Hadid, M. Pietikainen, P. Matejka, J. Cernocky, N. Poh, J. Kittler, A. Larcher, C. Levy, et al. Bi-modal person recognition on a mobile phone: using mobile phone data. In *Multimedia and Expo Workshops (ICMEW), 2012 IEEE International Conference on*, pages 635–640. IEEE, 2012.

[10] C. McCool, R. Wallace, M. McLaren, L. El Shafey, and S. Marcel. Session variability modelling for face authentication. *IET biometrics*, 2(3):117–129, 2013.

[11] P. E. Miller, J. R. Lyle, S. J. Pundlik, and D. L. Woodard. Performance evaluation of local appearance based periocular recognition. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1–6. IEEE, 2010.

[12] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3):145–175, 2001.

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

[14] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital signal processing*, 10(1):19–41, 2000.

[15] C.-W. Tan and A. Kumar. Towards online iris and periocular recognition under relaxed imaging constraints. *Image Processing, IEEE Transactions on*, 22(10):3751–3765, 2013.

[16] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Image Processing, IEEE Transactions on*, 19(6):1635–1650, 2010.

[17] R. P. V. Violato, M. U. Neto, F. O. Simões, T. de Freitas, and M. d. A. A. Pereira. Biocpqd: uma base de dados biométricos com amostras de face e voz de indivíduos brasileiros. *Cad. CPqD Tecnologia*, 9(2):7–18, 2013.

[18] R. Vogt and S. Sridharan. Explicit modelling of session variability for speaker verification. *Computer Speech & Language*, 22(1):17–38, 2008.

[19] R. Wallace, M. McLaren, C. McCool, and S. Marcel. Inter-session variability modelling and joint factor analysis for face authentication. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–8. IEEE, 2011.