MODEST FACE RECOGNITION

Vitomir Štruc, Janez Križaj, Simon Dobrišek

Faculty of Electrical Engineering, University of Ljubljana,
Tržaška cesta 25, SI-1000 Ljubljana, Slovenia
vitomir.struc, janez.krizaj, simon.dobrisek@fe.uni-lj.si

ABSTRACT

The facial imagery usually at the disposal for forensics investigations is commonly of a poor quality due to the unconstrained settings in which it was acquired. The captured faces are typically non-frontal, partially occluded and of a low resolution, which makes the recognition task extremely difficult. In this paper we try to address this problem by presenting a novel framework for face recognition that combines diverse features sets (Gabor features, local binary patterns, local phase quantization features and pixel intensities), probabilistic linear discriminant analysis (PLDA) and data fusion based on linear logistic regression. With the proposed framework a matching score for the given pair of probe and target images is produced by applying PLDA on each of the four feature sets independently - producing a (partial) matching score for each of the PLDA-based feature vectors - and then combining the partial matching results at the score level to generate a single matching score for recognition. We make two main contributions in the paper: *i)* we introduce a novel framework for face recognition that relies on probabilistic MOdels of Diverse fEature SeTs (MODEST) to facilitate the recognition process and *ii)* benchmark it against the existing state-of-the-art. We demonstrate the feasibility of our MODEST framework on the FRGCv2 and PaSC databases and present comparative results with the state-of-the-art recognition techniques, which demonstrate the efficacy of our framework.

Index Terms— Face recognition, probabilistic modeling, diverse feature sets, modest framework

1. INTRODUCTION

Unconstrained face recognition still represents an open problem that has not yet been satisfactory solved by today's (face) recognition technology. In unconstrained settings the variability in the facial-image data caused, for instance, by the ambient lighting conditions, self-occlusions, varying viewing angles and alike, represents a major source of difficulty for the existing technology. Fig. 1, which shows a few sample frames from a couple of real-world videos containing faces, illustrates this problem. Large pose variations, poor resolution and occluded or missing facial data are commonly encountered in settings where controlled conditions cannot (or could not) be assured for the data acquisition procedure. This is typically the case in forensics applications, which also often experience difficulties in utilizing the available (video) evidence due to the poor performance of the existing recognition technology on data captured in uncontrolled conditions.

Organized efforts towards improving the state of the face recognition technology, such as the Face Recognition Grand Challenge [1], the Labeled Faces in the Wild [2], or the Good, the Bad and the Ugly Face Recognition Challenge [3] have spurred the development of more efficient face recognition techniques in recent years and helped to improve the recognition performance on unconstrained data significantly compared to what was possible before. Improvements in face detection and registration [4], [5], facial-feature extraction [6], [7], [8], [9] and modeling-and-classification [10], [11], [12], [13], [14] have all contributed significantly to these developments.

Fig. 1. Sample frames from real-world videos (Point-and-Shoot-Camera (PaSC) database [15]) - typical problems encountered with unconstrained face recognition, such as different lighting conditions and varying viewing angles are shown.

Equally important are recent findings about the complementary information carried by different feature types. Using more than a single feature type for describing a facial image can significantly boost the performance of the given face recognition system and can help with the robustness in unconstrained settings. Tan and Triggs [16], for example, showed that using Gabor features together with local binary patterns (and local ternary patterns) improves upon the case, where either of the feature types is used on its own. A similar result was also shown by Yuan et al. [17] for the case of local binary and local phase quantization patterns.

In this paper, we build on the ideas presented above and use a rich set of diverse features to represent a given face image for recognition. In particular, we exploit Gabor features, local binary patterns (LBPs), local phase quantization features (LPQs) and pixel intensities to describe the facial images and use the features together with a probabilistic form of linear discriminant analysis (PLDA) [10] to produce partial matching scores for each feature type. We then combine the partial matching scores into a final score for recognition through a weighted sum, where the weights of the summation are learned using linear logistic regression. Our novel framework (exploiting probabilistic **MO**dels of **D**iverse f**E**ature **S**e**T**s - MODEST) is in more detail presented in the remainder of the paper.

The rest of the paper is structured as follows. In Section 2 we introduce our MODEST framework, evaluate it Section 3 and conclude the paper in Section 4.

978-1-4799-8105-2/15/$31.00 © 2015 IEEE

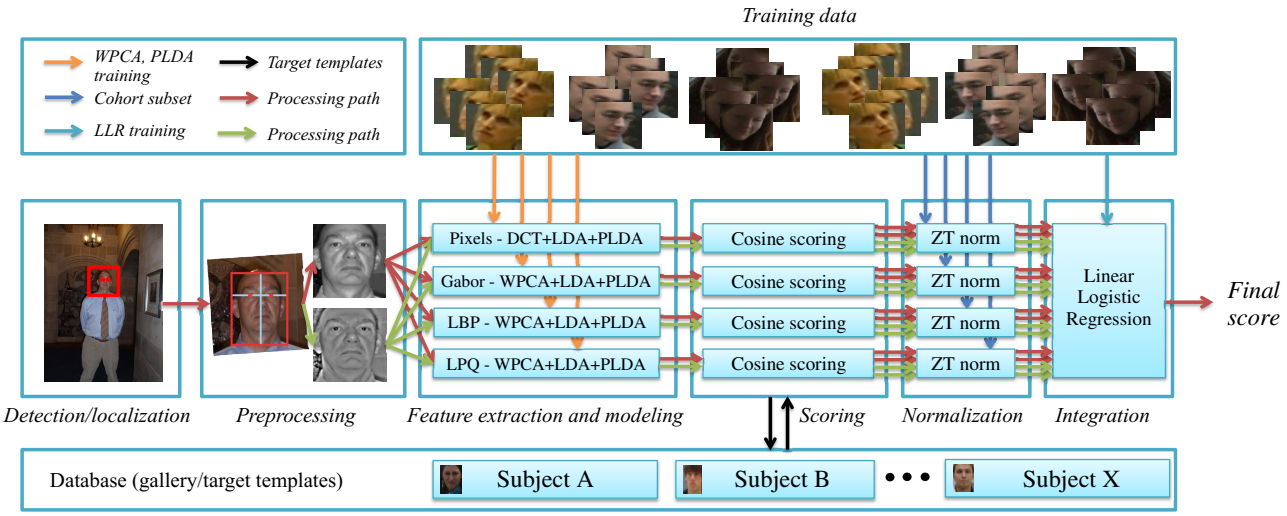

Fig. 2. Schematic representation of our MODEST framework.

Fig. 3. Illustration of the preprocessing procedure.

2. THE MODEST FRAMEWORK

A short overview of our MODEST framework is presented in Figure 2. In the first step, the framework detects the facial area and eye-center locations in the given image (or video frame) and based on the detection results geometrically normalizes the facial area. The normalized facial area is then converted to gray-scale and photometrically normalized using a retina-based modeling technique [18]. In the feature extraction step diverse sets of features, i.e., Gabor features, LBPs, LPQs and pixel intensities, are first extracted from the grey-scale and photometrically normalized images and then subjected to a whitening procedure (i.e., whitened principal component analysis - WPCA) followed by PLDA for dimensionality reduction. In the last step, the low-dimensional PLDA feature vectors extracted from the given probe image (or frame) are matched against the corresponding PLDA feature vectors of the given target image (or frame). Note that for a given image 8 distinct PLDA feature vectors (4 features types $\times$ 2 facial images (i.e., Y and P)) are computed, resulting in 8 matching scores. These scores are normalized using a symmetric form of ZT normalization and ultimately combined using a fusion procedure relying on linear logistic regression. It should be noted that in case our framework is applied to video data (instead of still images) a single PLDA feature vectors is computed for each feature type by averaging the corresponding PLDA feature vectors of all frames of the given video sequence.

2.1. Detection and preprocessing

The detection and preprocessing steps of our MODEST framework aim at detecting the presence of a face in the given image or video frame and geometrically and photometrically normalizing the detected facial region to a form that is suitable for feature extraction. Since we use pre-annotated face- and eye-location data as well as automatic techniques for the detection step in our experiments, we defer the description of the detection procedure to the experimental section and assume at this point that the eye-coordinates for the given face image are known.

During geometric normalization the facial area is rotated based on the eye-center coordinates in such a way that the line connecting eyes is in a horizontal position (Fig. 3 left). The facial area is then cropped with respect to the inter-occular distance dx and scaled to a fixed size. The coefficients k_1, k_2, k_3, k_4 (Fig. 3 left) are chosen empirically based on our experience with the face recognition task, i.e., $k_1 = 0.9$, $k_2 = 2.2$, $k_3 = k_4 = 1.3$.

Once the facial area is geometrically normalized, we convert the normalized crop to gray-scale and photometrically normalize it using the normalization technique from [18] that exploits the reflectance-luminance model of image formation to remove illumination artifacts from the facial images. An example of the gray-scale and photometrically normalized image is shown in Fig. 3 (right).

2.2. The feature sets

To extract as much discriminative information from the gray-scale (Y) and photometrically normalized (P) facial images as possible, we extract four types of feature vectors from the two images (Y and P), i.e., Gabor magnitude features [9], local binary patterns (LBPs) [7], local phase quantization features (LPQs) [19] and feature vectors comprised of pixel intensities. Hence, for a given input image, our MODEST framework computes eight distinct feature vectors that are then subjected to a dimensionality reduction procedure before being fed to the PLDA modeling technique. Note that the common processing chain prior to PLDA is to use a whitened version of PCA (WPCA) together with linear discriminant analysis (LDA) to whiten the feature vectors and reduce their dimensionality [11], [13]. However, based on our preliminary experiments we concluded that for raw pixel intensities we can use an additional discrete-cosine-transform (DCT) step for dimensionality reduction prior to whitening without reducing performance, but with an significant reduction in the mem-

Table 1. Dimensionality reduction and whitening

	DCT	$WPCA$	LDA
Intensities (from *Y* and *P*)	•	•	•
Gabor, LBP, LPQ (from *Y* and *P*)	-	•	•

ory footprint of our MODEST framework. A short summary of the dimensionality reduction steps for each feature type is given in Table 1.

2.3. Probabilistic modeling and matching

In the last step of our MODEST framework, we use probabilistic linear discriminant analysis (PLDA) [10] to further reduce the dimensionality of the feature vectors and enhance their discrimination information. While there exists several variants of PLDA, the variant used in this paper can formally be described as follows: let $\{\eta_r : r = 1, .., R\}$ denote a collection of feature vectors extracted from a set of face images (or frames) of a distinct individual. Then PLDA decomposes each feature vector into the following form:

$$\eta_r = m + \Phi\beta + \Gamma\alpha_r + \varepsilon_r, \tag{1}$$

where m denotes a global offset, representing the average feature vector, Φ provides the basis for the identity-specific subspace, β represents a latent identity vector with a standard normal distribution, Γ provides the basis for the channel subspace, α_r denotes a latent vector distributed according to a standard normal distribution and ε_r denotes a sample-dependant residual term, which is assumed to be normally distributed with a mean of zero and a diagonal covariance matrix Σ. It has to be noted at this point that the parameters of the PLDA model $\{m, \Phi, \Gamma, \Sigma\}$ are not determined analytically as with LDA. Instead, they are learned from some development data via an EM algorithm, e.g., [13]. Once the PLDA model parameters are known, inferences about the identity of a given feature vector η_r can be made based on the hidden identity variable β.

Note that within the MODEST framework one hidden identity variable β is computed for each of the eight feature vectors. A cosine-similarity-based scoring procedure is then used with these identity variables to produce eight matching scores for each matching attempt (i.e., for each probe-to-target comparison). These matching scores are then normalized using a symmetric variant of ZT score normalization and ultimately combined into a final similarity score based on a weighted summation, where the weights of the sum are learned on some annotated development data using linear logistic regression.

3. EXPERIMENTS

3.1. Databases and protocols

We assess our MODEST framework on two publicly available databases, i.e, the second version of the Face Recognition Grand Challenge (FRGCv2) and the Point and Shoot Face Recognition Challenge (PaSC) databases.

The first, the FRGCv2 database [1], represents a large database of facial images featuring more than 40000 still images of 466 distinct subjects. For the experiments on the FRGCv2 database we select the most challenging experimental configuration defined for the database, commonly referred to as Experiment 4. This experiment defines a target (or gallery) set of 16028 images, a probe (or query) set of 8014 images, and a training set of 12776 images that need to be used during experimentation. Note that the images of these image sets were captured in adverse conditions and, therefore, represent quite a challenge to the existing face-recognition technology. The result of the experiments on the FRGCv2 database is a similarity matrix (8014×16028) based on which various performance metrics and performance curves can be computed. We report our results in the form of ROC curves and the verification rate at the false accept rate of 0.1% - VER@01FAR. To facilitate comparisons against other results reported in the literature, we follow the experimental protocol and present so-called ROC-III curves (and corresponding operating points), which are computed from a subset of the scores in the 8014×16028 similarity matrix that correspond to more challenging verification attempts.

The second database used in our experiments - the PaSC database [15] - represents a very recent database that features still- as well as video-data of more than 250 subjects. In our experiments we focus on the video part of the database, which features video recordings of 265 subjects. Note that unlike other databases designed for face-recognition experiments, the video data in the PaSC database does not feature subjects facing the camera directly, instead the data represents real-world videos, where the subjects perform various tasks and do not pay special attention to the fact that they are being recorded. Due to this setup, the video frames and images from the PaSC database exhibit variability in terms of viewing angles, self-occlusion, varying lighting conditions, motion blur, poor focus and alike. The PaSC database contains 4688 still images and 1401 video recordings for experimentation and another 2872 still images and 280 videos that can be utilized during training. Using this data we conduct two types of experiments on the PaSC database:

- *still-vs-video* recognition experiments, where each of the 4688 still images is compared against each of the 1401 video recordings resulting in a similarity matrix of size 4688×1401, and
- *video-vs-video* recognition experiments, where 1401 video recordings are compared against each other, resulting in a similarity matrix of size 1401×1401.

Similar as with the FRGCv2 database we report the results for the PaSC database in terms of ROC curves (computed from the similarity matrices) and a selected operating point, i.e., the verification rate at the false accept rate of 1% - VER@1FAR. A few example images from both databases (after scaling to a fixed size) are shown in Fig. 4. Here, the group of images on the left shows example images from the FRGCv2 database, and the group of images on the right shows the images from the PaSC database. Each column represents the same identity.

3.2. Results and discussion

3.2.1. Experiments on the FRGCv2 database

Our first series of verification experiments uses only the FRGCv2 database and aims at demonstrating some characteristics of our MODEST framework. For this series of experiments we skip the face detection/localization step and use the ground-truth eye-center coordinates provided with the database to geometrically normalize the facial images and scale them to 100×100 pixels.

The first issue worth investigating is the use of the original as well as photometrically normalized face images for extraction of the feature sets (i.e., Gabor features, LBP features, LPQ features and pixel intensities) used for our MODEST framework. Commonly, only the photometrically normalized images are used

Fig. 4. Sample images from the two experimental databases: (left) images from FRGCv2, (right) images from the PaSC challenge. Note that each column represents images of the same subject. The last column in Fig. 4 (left) and Fig. 4 (right) represents the same person, which is present in both the FRGCv2 as well as the PaSC database. Note how the unconstrained settings (in Fig. 4 (right)) result in images that are much more challenging for the recognition system.

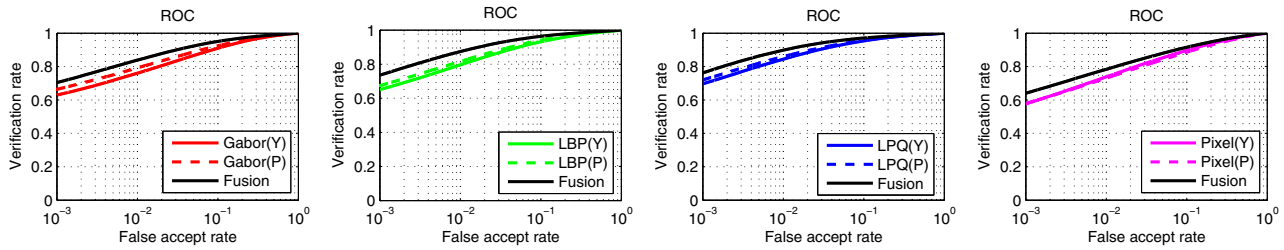

Fig. 5. Performance ensured by the individual feature sets when extracted from the original and the photometrically normalized images as well as the combined result (Exp. 4 - ROC III). From left to right: for Gabor features, for LBP features, for LPQ features and for pixel intensities.

in the processing pipelines of the existing face recognition systems, while the original images are discarded as being affected too much by the external lighting conditions. To explore this issue, we conduct verification experiments for each feature type independently. We extract each type of feature from the original as well as the photometrically normalized images, match them against the corresponding feature vectors of the gallery/target templates and combine the results at the matching score level using a weighted sum. We learn the weights of the weighted-sum on part of the development data of the FRGCv2 database with linear logistic regression. The results of this series of experiments are presented in Fig. 5. As we can see, the performance ensured by the feature sets extracted from the photometrically normalized face images is slightly better than the performance ensured by the features extracted from the original images for all four feature types. Similarly, when original as well as photometrically normalized images are combined a substantial performance increase is again visible for all four feature types. This results shows that despite the generally acknowledged believe that the low-frequency information should be removed from the images to ensure robustness to external lighting conditions, the low-frequency part of the image still contains useful information that can be exploited for recognition.

Another important aspect of our MODEST framework is its overall recognition performance. To further explore this issue, we again apply our MODEST framework on well aligned facial data from the FRGCv2 database. Thus, for this series of experiments we skip the face detection/localization step and use the ground-truth eye-center coordinates provided with the database to geometrically normalize the facial images and to scale them to a fixed

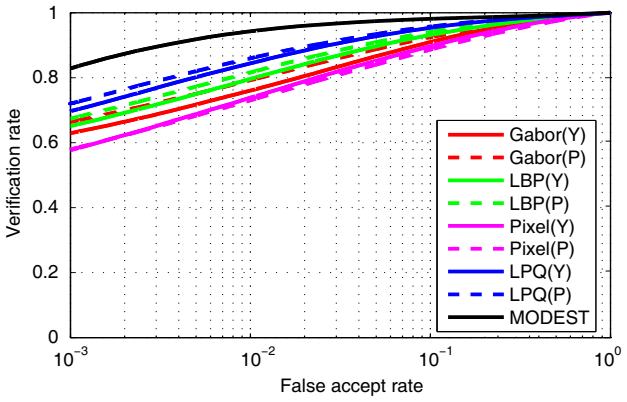

Fig. 6. Comparison of the performance ensured by our MODEST framework and the individual feature sets on the FRGCv2 database (Exp. 4 - ROC III).

size of 100×100 pixels. We then run the processing pipeline of our MODEST framework on the FRGCv2 database and generate corresponding ROC curves. As shown in Fig. 6, where the results of this experiment are presented, our MODEST framework manages to achieve a verification rate of 83.2% at the false acceptance rate of 0.1% improving significantly on the individual feature sets. This shows that our MODEST features contain complementary information that contributes to the overall performance of the MODEST framework. To put this performance into perspective, the reader is referred to [8] for a recent comparison of

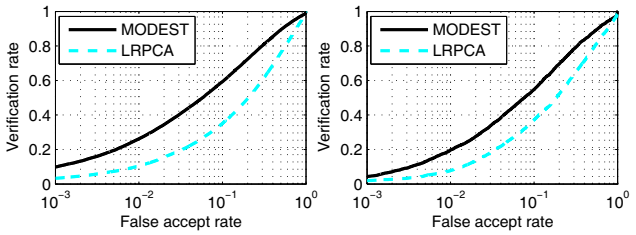

Fig. 7. ROC curves for the still-vs-video (left) and video-vs-video (right) experiments on the PaSC database.

Table 2. Performance comparison with the state-of-the-art on the PaSC data. The table shows VER@1FAR.

Method	*still-vs-video*	*video-vs-video*
LR-PCA [3]	0.10	0.08
ISV-GMM [14]	0.11	0.05
WPCA-SILD [21]	0.23	0.09
Eigen-PEP [22]	0.24	**0.26**
MODEST (ours)	**0.26**	0.19

state-of-the-art methods on this database.

3.2.2. Experiments on the PaSC database

In our third series of verification experiments we aim at evaluating our MODEST approach on the PaSC database, which contains real-world facial imagery captured in unconstrained environments. To fully automate our framework we use a commercial face and eye detector[1] to find faces and eye-centers in the video and still-image data. Note that due to this automatic procedure some faces are missed and, therefore, not all frames and images from the PaSC database can be exploited for the experiments. If in a given video sequence not a single face is found, we use some random coordinate on a randomly selected frame to generate some data to produce the match scores. Different from the FRGCv2 database, the images from the PaSC database are scaled to a size of 50x50 pixels prior to feature extraction, which helps to make the recognition procedure a little more robust. The PaSC database and its associated experimental protocol also define a baseline technique, i.e., local-region principal component analysis (LR-PCA) [3], which was found by the authors of the database to out-perform many of the existing face recognition techniques and is, therefore, also included in our experiments.

A comparison between the performance of our MODEST framework and the baseline LR-PCA technique is presented in Fig. 7 for the still-vs-video and the video-vs-video experiments. Note that for both experiments our framework manages to increase the verification rate at the false accept rate of 1% by more than 2.5 times with respect to the baseline. For the still-vs-video experiment this is (to the best of our knowledge) also the best reported performance on this database by any non-commercial face recognition system.

Next to the comparison with the baseline LR-PCA technique, it is also of interest how our MODEST framework compares to other state-of-the-art methods from the literature. To explore this issue, we provide in Table 2 the results of the recent face recognition competition held in conjunction with IJCB 2014 [20]. The results of the participants (which next to the PaSC baseline are to the best of our knowledge also the only published results on this database by the time of writting) are shown in the first four rows of the table, while the performance of our MODEST framework are presented in the last row. Note that our MODEST framework achieved the best result on the still-vs-video experiment and ranked in second on the video-vs-video experiment. The remaining methods from the table represent: *i)* a technique based on inter-session variability modeling using Gaussian mixture models - ISV-GMM [14], *ii)* a technique build around a probabilistic elastic part model - Eigen-PEP [22],[23], and a technique based on WPCA and side-information-LDA [21]. The reader is referred to the provided references for more information on the techniques included in the comparison.

[1]To be precise, we use eye coordinates generated by the PittPatt eye detector that were provided to the participants of the recent IJCB face recognition competition by the organizers [20]

4. CONCLUSION

We have presented a MODEST framework for face recognition that relies on probabilistic modeling of diverse feature sets to facilitate face recognition from real world-data. We have shown that the proposed framework ensures a recognition performance that is competitive with the existing state-of-the-art. As part of our future work, we plan to include an additional processing path to our framework that provides information on soft biometric cues and quality measures to the recognition system and improve the facie registration step, which seems to be crucial for the recognition performance.

5. ACKNOWLEDGEMENTS

The work presented in this paper was supported in parts by the national research program P2-0250(C) Metrology and Biometric Systems and the European Union's Seventh Framework Programme (FP7-SEC-2011.20.6) under grant agreement number 285582 (RESPECT). The support of COST Actions IC1106 and IC1206 is also appreciated.

6. REFERENCES

[1] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, Jin Chang, K. Hoffman, J. Marques, Jaesik Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," in *Proc. of CVPR'05*, 2005, pp. 947–954.

[2] G.B. Huang, M. Ramesh, T. Berg, and Learned-Miller E., "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," in *Technical Report 07-49, University of Massachusetts*, 2007.

[3] P.J. Phillips, J.R. Beveridge, B.A. Draper, G. Givens, A.J. O'Toole, D.S. Bolme, Dunlop J., Y. Man, Sahibzada H., and S. Weimer, "An Introduction to the Good, the Bad & the Ugly Face Recognition Challenge Problem," in *Proc. of FG'11*, 2011, pp. 346–353.

[4] G.B. Huang, M. Mattar, H. Lee, and Learned-Miller E., "Learning to Align from Scratch," in *Proc. of NIPS'12*, 2012.

[5] X. Xioang and F. De la Torre, "Supervised Descent Method and its Application to Face Alignment," in *Proc. of CVPR'13*, 2013.

[6] D. Cheng., X. Cao, F. Wen, and J. Sun, "Blessing of Dimensionality: High-dimensional Feature and Its Efficient Compression for Face Verification," in *Proc. of CVPR'13*, 2013, pp. 3025–3032.

[7] M. Pietikainen, A. Hadid, G. Zhao, and T. Ahonen, *Computer Visioin using Local Binary Patterns*, Springer, 2011.

[8] Yan Li, Shiguang Shan, Haihong Zhang, Shihong Lao, and Xilin Chen, "Fusing Magnitude and Phase Features for Robust Face Recognition," in *Proc. of ACCV'12*, 2012, pp. 601–612.

[9] D. Cheng., X. Cao, F. Wen, and J. Sun, "Computer Face Recognition Using Early Biologically Inspired Features," in *Proc. of BTAS'13*, 2013, pp. 1–6.

[10] P. Li, Y. Fu, U. Mohammed, J.H. Elder, and S.J.D. Prince, "Probablistic models for inference about identity," *IEEE TPAMI*, vol. 34, no. 1, pp. 144–157, 2012.

[11] B. Vesnicer, J. Žganec Gros, Dobrišek S., and V. Štruc, "Incorporating Duration Information into I-Vector-Based Speaker-Recognition Systems," in *Proc. of Odyssey'14*, 2014.

[12] B. Vesnicer and F. Mihelič, "The likelihood ratio decision criterion for nuisance attribute projection in gmm speaker verification," *EURASIP JASP*, vol. 2008, 2008.

[13] El Shafey, C. L., McCool, R. Wallace, and S. Marcel, "A scalable formulation of probabilistic linear discriminant analysis: Applied to face recognition," *IEEE TPAMI*, vol. 35, no. 7, pp. 1788–1794, 2013.

[14] C. McCool, R. Wallace, M. McLaren, L. El Shafey, and S. Marcel, "Session variablity modelling for face authentication," *IET Biometrics*, vol. 2, no. 3, pp. 117–129, 2013.

[15] Beveridge J.R., Phillips P.J., Bolme D.S., Draper B.A., Givens G.H., Yui Man Lui, Teli M.N., Hao Zhang, Scruggs W.T., Bowyer K.W., Flynn P.J., and Su Cheng, "The challenge of face recognition from digital point-and-shoot cameras," in *Proc. of BTAS'13*, 2013, pp. 1–8.

[16] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transations on Image Processing*, vol. 19, no. 6, 2010.

[17] B. Yuan, H. Cao, and J. Chu, "Combining Local Binary Pattern and Local Phase Quantization for Face Recognition," in *Proc. of ISBAST'12*, 2012.

[18] N. Vu and A. Caplier, "Illumination-robust face recognition using retina modeling," in *Proc. of ICIP'09*, 2009, pp. 2335–2338.

[19] C.H. Chan, Tahir M.A., J. Kittler, and M. Pietikainen, "Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors," *IEEE TPAMI*, vol. 35, no. 7, pp. 1164–1177, 2013.

[20] J.R. Beveridge, H. Zhang, P.J. Flynn, Y. Lee, V.E. Liong, J. Lu, M. de Assis Angeloni, T. de Freitas Pereira, H. Li, G. Hua, V. Struc, Krizaj J., and P.J. Phillips, "The IJCB 2014 PaSC Video Face and Person Recognition Competition," in *Proc. of IJCB'14*, 2014.

[21] M. Kan, S. Shan, D. Xu, and X. Chen, "Side information based linear discriminant analysis for face rcognition," in *Proc. of BMVC'11*, 2011.

[22] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang, "Probabilistic elastic matching for pose variant face verification," in *Proc. of CVPR'13*, 2013, pp. 3499–3506.

[23] H. Li, G. Hua, X. Shen, Z. Lin, and J. Brandt, "Eigen-PEP for Video Face Recognition," in *Proc. of ACCV'14*, 2014.