2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2010. 67 - 73

DOI: http://dx.doi.org/10.1109/CVPRW.2010.5543231

Scenario-Based Score Fusion for Face Recognition at a Distance

Pedro Tome, Julian Fierrez, Fernando Alonso-Fernandez and Javier Ortega-Garcia
Biometric Recognition Group - ATVS, Escuela Politecnica Superior - Universidad Autonoma de Madrid
Avda. Francisco Tomas y Valiente, 11 - Campus de Cantoblanco - 28049 Madrid, Spain
{pedro.tome, julian.fierrez, fernando.alonso, javier.ortega}@uam.es

Abstract

The effect of different acquisition distances on the performance of face verification is studied. In particular, we evaluate two standard approaches using popular features (DCT and PCA) and matchers (GMM and SVM) under variation in the acquisition distance, as well as their score-level combination. The DCT-GMM-based system is found to be more robust to acquisition distance degradation than the PCA-SVM-based system. We exploit this fact by introducing an adaptive score fusion scheme based on a novel automatic scenario estimation which is shown to improve our system in uncontrolled environments.

1. Introduction

The increasing need for reliable automated personal identification in the current networked society and the recent advances in pattern recognition have resulted in the current interest in *biometric systems* [12]. In particular, automatic face recognition [23, 21, 3, 1] has received great attention in the last years because of: i) the commonly accepted distinctiveness of the face pattern, ii) the widespread deployment of electronic acquisition devices for acquisition at a distance, iii) the low intrusivity of this kind of systems, and iv) the wide variety of practical applications ranging from access control to criminal identification.

Our first objective in this work is to investigate the effects of acquisition distance variation on the performance of automatic face recognition systems. This is motivated by the analysis of the results from the recent Multiple Biometric Grand Challenge (MBGC 2009) [17] and the Face Recognition Vendor Test (FRVT 2006) [19], which show that a lot of research is still needed to overcome these problems. As a result, the National Institute of Standards and Technology (NIST) has proposed a new challenge called "the Good, the Bad and the Ugly" [18] which makes use of three partitions of the MBGC Still Face dataset of frontal images [17]. This new challenge is designed by NIST to develop new face algorithms capable to match correctly difficult face pairs. In this sense, we have studied the degradation effects in three different scenarios defined by the acquisition distance between subject and camera, namely "close", "medium" and "far" distance.

Li *et al.* [16] consider the problem of Biometrics at a Distance as having no restrictions over conditions such as scale, pose, lighting, focus, resolution, facial expression, accessories, makeup, occlusions, background, or photographic quality. Many solutions have been proposed in the literature to deal with these factors individually [4, 9, 14, 15, 22, 24] but a suitable solution to the global problem of unconstrained environments has not been developed yet.

As a second objective, we propose a novel scenario estimator that enables system adaptation depending on the predited acquisition conditions. This is expoted in the present paper in a new scenario-based fusion scheme.

As to the experiments, we have studied the effect of training and testing with images acquired at different distances using two classical face recognition approaches (DCT-GMM- and PCA-SVM- based systems). We also investigate experimentally the effects of acquisition distance variation on a multi-algorithm approach [11] based on these matchers. Finally, we also evaluate the proposed scenario-based fusion approach that exploits the scenario estimator presented. In particular, we study an adaptive score-level fusion technique [7] based on the estimated scenario.

The paper is structured as follows. Sect. 2 summarizes related work on the characterization of face acquisition distance and its effects on automatic recognition, and describe the face acquisition distance estimator proposed. Sect. 3 summarizes the individual face verification systems used. The proposed scenario-based score fusion scheme is introduced in Sect. 4 and Sect. 5 summarizes the experimental setup. Results obtained are given in Sect. 6, and finally, conclusions are drawn in Sect. 7.

2. How to Estimate the Acquisition Distance

The concept of estimating the acquisition distance in order to define different scenarios has not been traditionally used in face recognition. Automatic scenario estimation gives us knowledge about the variability level that affects

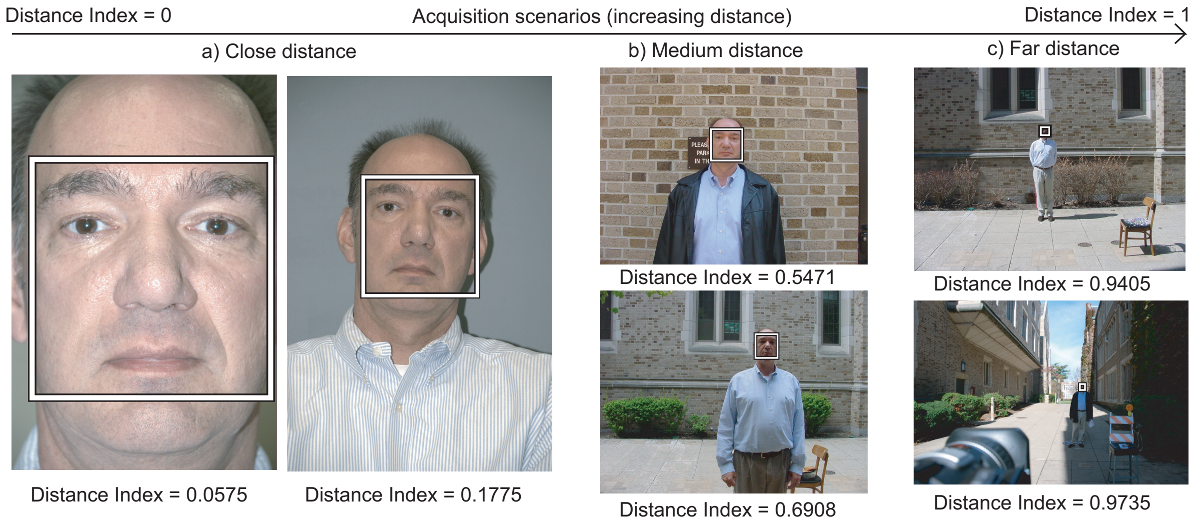

Figure 1. Acquisition scenarios defined: a) close distance, b) medium distance and c) far distance, together with the distance index (normalized using Eq. 2 to be in the $[0, 1]$ range).

the system (i.e., different scenarios usually present different variability factors) and therefore is a valuable tool for system adaptation.

2.1. Face Acquisition Distance Index

The proposed estimator of the acquisition distance, is based on the difference between areas of segmented face and full image. We define the "Distance Index" D as:

$$D = -\log_{10}\left(\frac{A_s}{A_f}\right) \tag{1}$$

where A_s and A_f are respectively the segmented face area and the full image area. Therefore D is a function of the distance to the acquisition device. The minimum possible value is 0, when the segmented face occupies the whole image. As the person goes away of the camera, the "Distance Index" increases, until it reaches a maximum value of Inf (i.e. $A_s = 0$ or no face). In Fig. 1, we plot examples of images acquired at different distances, as well as their computed distance measures. Fig. 2 shows the distribution of the proposed index D in the database used in this paper, where it is possible to appreciate the differences between the three scenarios defined by the acquisition distance.

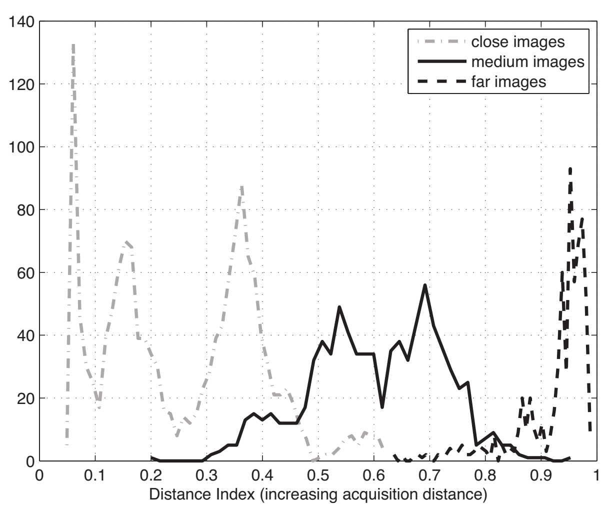

Figure 2. Histogram of the Distance Index on the NIST MBGC v2.0 Face Stills dataset (normalized using Eq. 2 to be in the $[0, 1]$ range).

3. Face Verification Systems

The architecture of the face recognition system used is shown in Fig. 1. The preprocessing stage is divided into: i) automatic localization of the face, ii) segmentation, iii) size normalization to a constant size (64×80 in our experiments), and iv) pose and illumination compensation.

The preprocessing stage was executed using VeriLook SDK v2.0 (commercial system) and the few produced errors were manually corrected as described in previous works [5].

As mentioned previously, two approaches are used for face verification. These two matchers receive a normalized face from the preprocessing stage:

- **PCA-SVM system**. This verification system uses Principal Component Analysis (PCA). The evaluated system uses normalized and cropped face images of size 64×80 pixels (width $\times$ height) to train a PCA vector space where 96% of the variance is retained. This leads to a system where the original image space of 5120 dimensions is reduced to 249 dimensions. Similarity scores are computed in this PCA vector space using a SVM classifier with linear kernel.

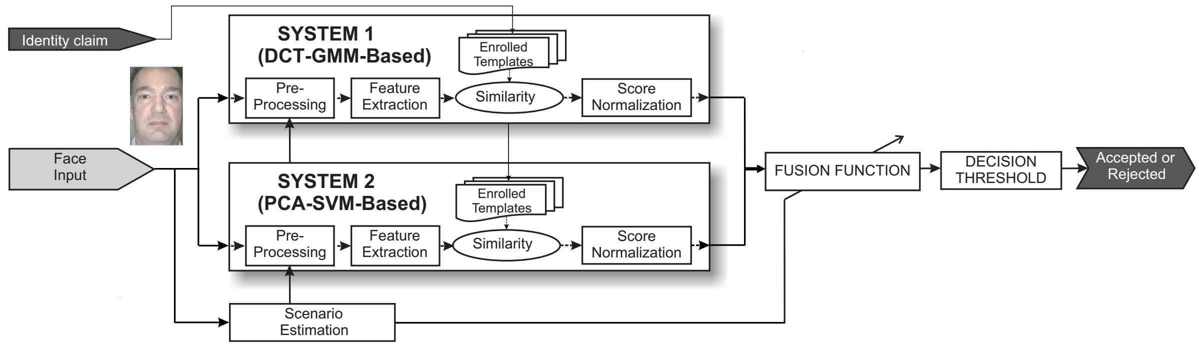

Figure 3. Scenario-based multi-algorithm approach for face verification at a distance.

- **DCT-GMM system**. This verification system divides the 64×80 face image into 8×8 blocks with horizontal and vertical overlap of 4 pixels. This process results in 285 blocks per segmented face. From each block a feature vector is obtained by applying the Discrete Cosine Transform (DCT) from which only the first 15 coefficients ($N = 15$) are retained. The blocks are used to derive a world GMM Ω_w and a client GMM Ω_c [8]. From previous experiments we obtained that using $M = 1024$ mixture components per GMM gave the best results. The DCT feature vector from each block is matched to both Ω_w and Ω_c to produce a log-likelihood score.

To carry out the fusion stage, scores of the two systems are first normalized to the $[0, 1]$ range using the tanh-estimators described in [10]:

$$s_k = \frac{1}{2}\left\{\tanh\left(C \cdot \frac{s'_k - \mu_{SD}}{\sigma_{SD}}\right) + 1\right\}, \qquad (2)$$

where s'_k is the kth score, s_k denotes the normalized score, C is a constant ($C = 0.4$ in our experiments), and μ_{SD} and σ_{SD} are respectively the estimated mean and standard derivation of the score distribution.

4. Scenario-Based Score Fusion

The fusion method used is based on the combination of the two sytems at the score-level following the sum rule approach [6]. It consists of averaging the matching scores provided by the different matchers. Under some mild statistical assumptions [2, 13] and with the proper matching score normalization [10] described previously, this simple method is demonstrated to give good results for the biometric authentication problem. This fact is corroborated in a number of studies [13, 20] in other modalities. Let the similarity scores s_{DG} (DCT-GMM-based system) and s_{PS} (PCA-SVM-based system) provided by the two matchers after the normalization stage. The fused result using the sum rule is $s = (s_{DG} + s_{PS})/2$.

Our basic assumption for the adaptive scenario-based fusion approach implemented is that verification performance of one of the algorithms drops significantly compared to the other one under image distance degradation. This fact is exploited with the following adaptive distance-based fusion strategy:

$$s_D = \left(\tfrac{D}{2}\right) s_{DG} + \left(1 - \tfrac{D}{2}\right) s_{PS}, \qquad (3)$$

where D is the "Distance Index" normalized to the range $[0, 1]$ using the Eq. 2. In other words, as the face adquisition distance decreases, D approaches 0 and more importance is given to the matching score of the more robust system s_{PS} (PCA-SVM-based system). When the distance increases ($D \rightarrow 1$), more importance is gradually given to the DCT-GMM system.

5. Experimental Framework

5.1. Database and Scenarios

The database used for the experimental work presented in this paper is a subcorpus called "Face Stills dataset" of the NIST Multiple Evaluation Grand Challenge (MBGC) v2.0 [17]. The database is comprised of 3842 face images from 147 subjects acquired at diferent distances. We further classify all the face images into three acquisition distance groups as follows. We consider three different scenarios: 1) “close” distance, in which the shoulders may be present; 2) “medium” distance, including the upper body; and 3) “far” distance, including the full body. Using these three general definitions we manually labeled all the 3482 face images from the dataset. Some example images are depicted in Fig. 1.

Only subjects with at least 4 images were kept in each scenario considered. A portion of the dataset was discarded

	Development	Evaluation	Discarded	Total
# users	56	78	13	147
Condition: at least 4 images per scenario: 2 train and 2 test.				
# images	484	2595	403	3482

Table 1. Number of users and images of NIST MBGC v2.0 Face Stills dataset used.

Development Set

Scenario	*close*	*Medium*	*Far*	*Mix*
# images	222	132	130	484

Evaluation Set

Scenario	*close*	*Medium*	*Far*	*Mix*
# images	1290	727	578	2595
Train	661	386	304	1351
Test	629	341	274	1244

Table 2. Configuration of the datasets (close, medium, far and mix combination of all of them) of each adquisition scenario.

(360 images from 89 subjects), because the face was occluded or the illumination completely degraded the face. A reduced number of subjects (13) were completely discarded (less than 4 image per scenario) discarding a total 403 images of the whole dataset. The data selection process is sumarized in Table 1, where we can see that the two considered subcorpora result in 134 subjects, using 484 images of 56 subjects for the development of the systems and 2595 images of 78 subjects for the evaluation.

5.2. Protocol

For the experiments in this paper we have divided the data in 56 subjects as development for tuning the systems and the remaining 78 subjects as evaluation (see Table 1).

The dataset was then divided according to the three acquisition distance scenarios defined in Section 5.1. The resulting subsets are shown in Table 2. The development set is used to train a PCA subspace and GMM world model per scenario (close, medium, far and mix). Here it is important to note that we have tuned the systems with an equal number of images (130 images, given by the smaller scenario, i.e. the "far" one).

On the other hand, the evaluation set was equally divided into a train and a test set, the first one for training the models of SVM and GMM per user and the other to test the system performance. Table 2 shows the different divisions of data in the three scenarios defined. It is possible to appreciate that the number of images is not perfectly distributed between these two sets (train and test) due to an imbalance in the number of samples per user.

Four main experiments are defined for verification performance assessment across scenarios:

- *close2x*. This is designed to obtain the performance of the systems in situations where only high quality controlled images are used to train the system. This will be considered as the Baseline system. In this case, only the 661 images of the close train set are used to train the GMM and SVM classifiers.

- *medium2x*, This protocol uses 386 images as training set from the medium distance dataset.

- *far2x* protocol. This protocol uses 304 images as training set from the far distance dataset.

- *mix2x*. This is designed to study the effects of combining several kinds of information (training with different acquisition distances). The train dataset is comprised of the sum of the three acquisition distance datasets (1351 images).

6. Results

Verification performance results are given in Fig. 4 for the individual matchers: a) DCT-GMM- and b) PCA-SVM-based. This figure shows all the possible combinations between training and test sets. The four curves represent the Equal Error Rate (EER) of each **train set** defined (*close2x*, *medium2x*, *far2x*, and *mix2x*) matched with each **test set** (Close, Medium, Far, and Mix).

As the person is going away the acquisition device the face information available decreases. This effect is appreciated on the system performance in Fig. 4 where both systems degrade their performance when the acquisition distance and the variability increases.

By analyzing these curves, it can be seen that the DCT-GMM-based matcher is more robust against an increasing acquisition distance [5], especially when training with the medium distance dataset. Conversely, althought being much better in ideal conditions (*close2close*), the PCA-SVM-based matcher degrades quick when increasing the acquisition distance .

When the system is trained with the highest quantity of information possible (*mix2x* protocol), we obtain better performance in general but we must be careful in the comparison because in this case we are training with a higher number of images compared to other scenarios.

6.1. Fusion Results

The combination of these systems through the sum fusion rule, and the proposed scenario-based weighted sum for different face acquisition distance groups is presented in Fig. 5. As can be seen, the fixed fusion strategy based on the sum rule only leads to improved performance over the best individual system in *medium2x* and *mix2x* scenarios, shown in Fig. 5b) and Fig. 5d). The proposed adaptive fusion approach results in improved performance for all the acquisition distance groups, outperforming the standard sum rule approach, especially in *medium2close* testing conditions in Fig. 5b), where the performance of the individual matchers are very different.

As shown in Fig. 5, the best results against increased acquisition distance are obtained when the system is trained with medium distance images and the mix of acquisition distance groups (*medium2x* and *mix2x* protocols). The baseline scenario (close distance training images) show less ro-

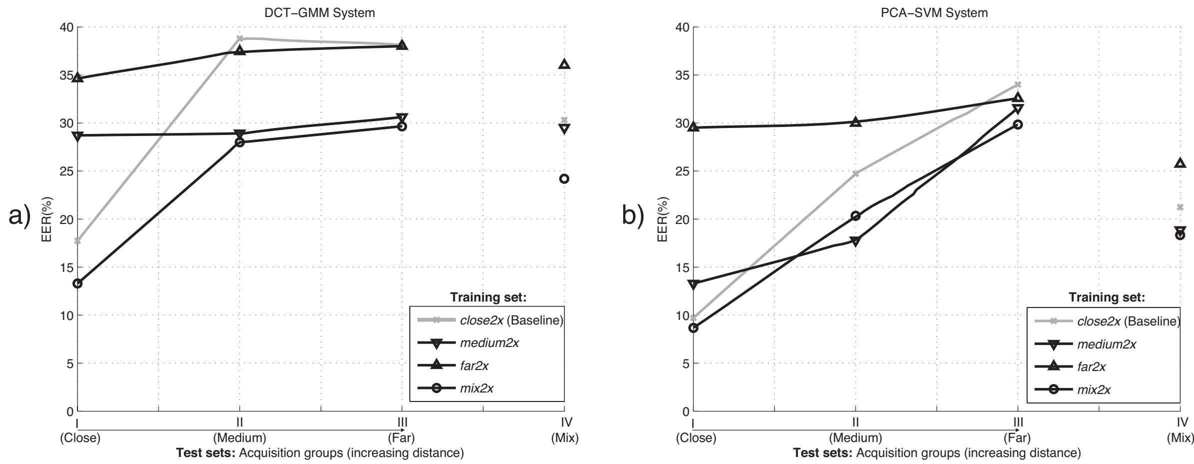

Figure 4. Verification performance of the individual matchers (DCT-GMM- and PCA-SVM- based) and their work in different conditions of training and test sets with different acquisition scenarios.

bustness, with great degradation as the acquisition distance increases.

Training with medium distance images is a good way to control the performance degradation due to varying distance. The DCT-GMM system generates low performance but stable results and the PCA-SVM system provides better performance but deteriorates quickly with the distance. Here the proposed fusion provides better results for far distance where both systems have a similar performance, and the new adaptive fusion is capable to equal the best system in closed distance testing.

The best performance is obtained with the *mix2* protocol where we are using the whole information of different acquisition distances in the training stage. As can be observed the fusion has an important role, increasing the system performance of the best individual system in all the cases. Also worth noting, in far distance conditions the fusion schemas improve the performance remarkably.

7. Discussion and Conclusions

A simple technique for scenario estimation based on the distance between the subject and the acquisition device has been proposed. The effects of face acquisition distance on the performance of two common approaches for face verification have been studied using this new scenario estimator. It has been found that the approach based on PCA subspace information and SVM classifier outperforms the DCT-GMM-based approach in close acquisition distance conditions but the approach based on DCT and GMM classifier is more robust to increasing acquisition distance.

It must be emphasized that this evidence is based on particular implementations of well known algorithms, and should not be taken as a general statement. Other implementations may lead to improved performance of any approach over the other in acquisition distance variation conditions (different scenario). On the other hand, the robustness observed of the DCT-GMM-based approach as compared to the PCA-SVM-based system has been observed in other studies [5], where far acquisition distance images degrade the performance in a PCA-SVM system, but the DCT-GMM system remains robust.

As has been demostrated, the variability present in at a distance scenarios can be used in the training stage in order to stabilize the system performance degradation ocurring in unpredicted aquisition conditions.

In particular, we have shown how the proposed distance estimator can be used in an adaptive score-level fusion approach to control this degradation. The proposed scheme leads to enhanced performance over the best matcher and the standard sum fusion rule over a wide range of face acquisition distances.

Acknowledgements

Pedro Tome is supported by a FPU Fellowship from Univ. Autonoma de Madrid, Julian Fierrez is supported by a Marie Curie Fellowship from the European Commission and Fernando Alonso-Fernandez is supported by a Juan de la Cierva Fellowship from the Spanish MICINN. This work has been partially supported by projects Bio-Challenge (TEC2009-11186), Contexts (S2009/TIC-1485), TeraSense (CSD2008-00068) and "Cátedra UAM-Telefónica".

References

[1] A. Abate, M. Nappi, D. Riccio, and G. Sabatino. 2D and 3D face recognition: A survey. *Pattern Recognition Letters*, 28(14):1885–1906, October 2007.

[2] E. Bigun, J. Bigun, B. Duc, and S. Fischer. Expert conciliation for multi modal person authentication systems by

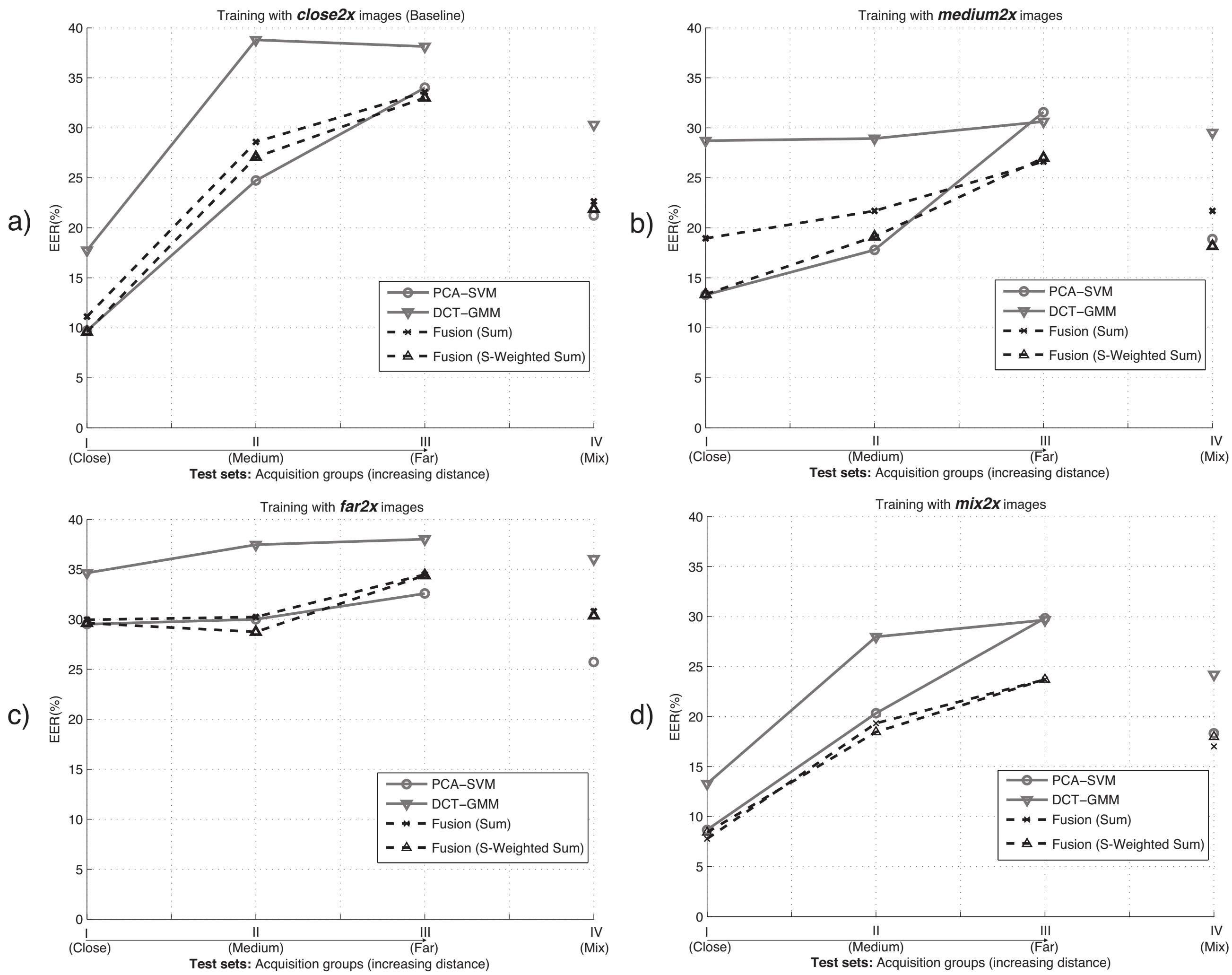

Figure 5. Verification performance of the individual matchers (DCT-GMM- and PCA-SVM- based), their combination through the sum fusion rule, and the proposed distance/scenario-based weighted sum for increasing the system performance at a distance. The results are displayed in the different acquisition scenarios under study.

bayesian statistics. *Proc. of AVBPA, Springer LNCS-1206*, pages 291–300, 1997.

[3] R. Chellappa, C. Wilson, and S. Sirohey. Human and machine recognition of faces: a survey. *Proceedings of the IEEE*, 83(5):705 –741, may 1995.

[4] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang. Total variation models for variable lighting face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:1519–1524, 2006.

[5] J. Fierrez, P. Tome, and J. Ortega-Garcia. Acquisition scenario analysis for face recognition at a distance. In *IEEE Proc. Intl. Conf. on Image Processing, ICIP (submitted)*, 2010.

[6] J. Fierrez-Aguilar, Y. Chen, J. Ortega-Garcia, and A. K. Jain. Incorporating image quality in multi-algorithm fingerprint verification. In *Proc. IAPR Intl. Conf. on Biometrics, ICB*, volume 3832 of *LNCS*, pages 213–220. Springer, January 2006.

[7] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun. Discriminative multimodal biometric authentication based on quality measures. *Pattern Recognition*, 38(5):777–779, May 2005.

[8] J. Galbally, C. McCool, J. Fierrez, S. Marcel, and J. Ortega-Garcia. On the vulnerability of face verification systems to hill-climbing attacks. *Pattern Recognition*, 43(3):1027–1038, March 2010.

[9] R. Gross, I. Matthews, and S. Baker. Appearance-based face recognition and light-fields. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(4):449–465, 2004.

[10] A. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285, December 2005.

[11] A. K. Jain and A. Ross. Multibiometric systems. *Communications of the ACM*, 47(1):34–40, 2004.

[12] A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Trans. on Circuits and Systems*

for Video Technology, 14:4–20, 2004.

[13] J. Kittler, M. Hatef, R. P. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:226–239, 1998.

[14] K.-C. Lee, J. Ho, and D. J. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(5):684–698, 2005.

[15] S. Z. Li, R. Chu, S. Liao, and L. Zhang. Illumination invariant face recognition using near-infrared images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(4):627–639, 2007.

[16] S. Z. Li, B. Schouten, and M. Tistarelli. *Handbook of Remote Biometrics for Surveillance and Security*, chapter Biometrics at a Distance: Issues, Challenges, and Prospects, pages 3–21. Springer, 2009.

[17] MBGC. Multiple Biometric Grand Challenge. *NIST - National Institute of Standard and Technology*, http://face.nist.gov/mbgc/.

[18] NIST - N.I. of Standard and Technology. Future Challenges based on the Multiple Biometric Grand Challenge (MBGC). http://face.nist.gov/mbgc/MBGCFuture/FutureChallengesMBGC.html.

[19] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe. Frvt 2006 and ice 2006 large-scale experimental results. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99, 2009.

[20] A. Ross and A. Jain. Information fusion in biometrics. *Pattern Recognition Letters*, 24:2115–2125, 2003.

[21] X. Tan, S. Chen, and Z. hua Zhou Fuyan Zhang. Face recognition from a single image per person: A survey. *Pattern Recognition*, 39:1725–1745, 2006.

[22] H. Wang, S. Z. Li, Y. Wang, and W. Zhang. Illumination modeling and normalization for face recognition. In *AMFG '03: Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, page 104, Washington, DC, USA, 2003. IEEE Computer Society.

[23] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, 2003.

[24] S. K. Zhou, G. Aggarwal, R. Chellappa, and D. W. Jacobs. Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination-invariant face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(2):230–245, 2007.