Face Recognition at a Distance: Scenario Analysis and Applications

R. Vera-Rodriguez, J. Fierrez, P. Tome, and J. Ortega-Garcia

Abstract. Face recognition is the most popular biometric used in applications at a distance, which range from high security scenarios such as border control to others such as video games. This is a very challenging task since there are many varying factors (illumination, pose, expression, etc.) This paper reports an experimental analysis of three acquisition scenarios for face recognition at a distance, namely: close, medium, and far distance between camera and query face, the three of them considering templates enrolled in controlled conditions. These three representative scenarios are studied using data from the NIST Multiple Biometric Grand Challenge, as the first step in order to understand the main variability factors that affect face recognition at a distance based on realistic yet workable and widely available data. The scenario analysis is conducted quantitatively in two ways. First, an analysis of the information content in segmented faces in the different scenarios. Second, an analysis of the performance across scenarios of three matchers, one commercial, and two other standard approaches using popular features (PCA and DCT) and matchers (SVM and GMM). The results show to what extent the acquisition setup impacts on the verification performance of face recognition at a distance.

1 Introduction

The growth of biometrics has been very significant in the last few years. A new research line growing in popularity is focused on using biometrics in less constrained scenarios in a non-intrusive way, including acquisition "On the Move" and "At a Distance" [7], which are user-friendly, and often do not need user cooperation.

The most common biometric modes used for recognition at a distance are face, iris and gait, being face the most popular of them. Face recognition is a challenging problem in the field of computer vision which has been the subject of active

R. Vera-Rodriguez, J. Fierrez, P. Tome, and J. Ortega-Garcia
ATVS, Escuela Politecnica Superior - Universidad Autonoma de Madrid,
Avda. Francisco Tomas y Valiente, 11 - 28049 Madrid, Spain
e-mail: {ruben.vera,julian.fierrez,pedro.tome}@uam.es,
javier.ortega@uam.es

A.P. de Leon F. de Carvalho et al. (Eds.): Distrib. Computing & Artif. Intell., AISC 79, pp. 341–348.
springerlink.com

research for the past decades because of its many applications in domains such as surveillance, covert security and context-aware environments. Face recognition is very appealing as a biometric as it offers several advantages in terms of being non-intrusive, non-invasive, cost-effective, easily accessible (i.e., face data can be conveniently acquired with a few inexpensive cameras) and relatively acceptable to the general public. However, employing the face for recognition also presents some difficulties since the appearance of the face can be altered by intrinsic factors such as age, expression, facial hair, glasses, make up, etc., as well as extrinsic ones such as scale, lighting, focus, resolution, or pose amongst others [13].

This paper is focused on the study of 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 NIST Multiple Biometric Grand Challenge (MBGC 2009) [8] and the Face Recognition Vendor Test (FRVT 2006) [9], which show that a lot of research is still needed to overcome these problems. In this sense, three different scenarios have been defined from the NIST MBGC depending on the acquisition distance between the subject and the camera, namely "close", "medium" and "far" distance. We use a subset of this benchmark dataset consisting of images of a total of 112 subjects acquired at different distances and varying conditions regarding illumination, pose/angle of head, and facial expression. This analysis is conducted quantitatively at two levels for the considered scenarios: 1) main data statistics such as information content, and 2) performance of recognition systems: one commercial, and two other based on popular features (PCA and DCT) and matchers (SVM and GMM).

Depending on the distance to the camera, face recognition could be applied in two different applications [1, 7]:

- Requiring cooperative users (near distance), such as in border control (e-passport) or security access (for example access to stadium in 2008 Olympic Games). In these cases a verification (one to one) of the identity is carried out.
- Not requiring cooperative users (medium and far distances), such as face surveillance (for example subway watch-list) or in large database search (such as national registration data or black-list data). In these cases an identification (one to many) is normally carried out.

Other applications could be on social network webs for automatic face tagging and finding people[1]. Apart from the person recognition applications, there are other applications in which face recognition technology can be useful such as activity detection (for smart homes [14], ambient assisting living [3] or video games [6, 12]), or in pedestrian detection to avoid accidents. In this last case a possible fusion between face and gait would be of interest [5]. Figure 1 shows some examples of applications of face recognition.

The paper is structured as follows. Sect. 2 describes the dataset and scenarios under study. Sect. 3 analyzes the main data statistics of the scenarios. Sect. 4 studies the performance of the three considered recognition systems on the different scenarios. Sect. 5 finally discusses the experimental findings and outlines future research.

[1] For example http://picasaweb.google.com

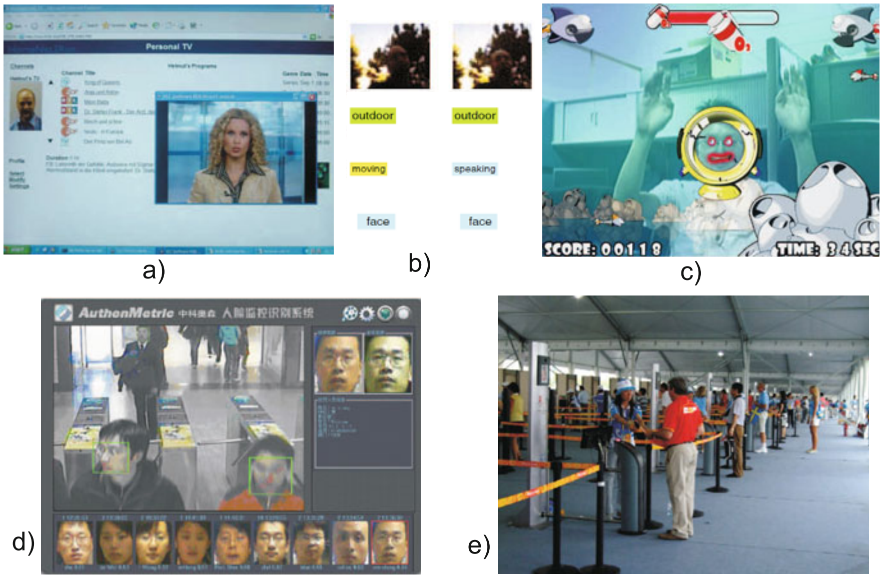

Fig. 1 Example images of different applications of face recognition: a) Web interface for smart TV program selection by face recognition [14]. b) Classification results of activity detection [2]. c) Example of video game using face and activity detection [12]. d) Example of a watch-list surveillance and identification system [7]. e) Face verification system used in Beijing 2008 Olympic Games [1].

2 Scenario Definition

The three scenarios considered are: 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, the 3482 face images from the 147 subjects present in the dataset NIST MBGC v2.0 Face Stills [8] were manually tagged. Some sample images are depicted in Fig. 2. A portion of the dataset was discarded (360 images from 89 subjects), because the face was occluded or the illumination completely degraded the face. Furthermore, although this information is not used in the present paper, all the images were marked as indoor or outdoor.

Finally, in order to enable verification experiments considering enrollment at close distance and testing at close, medium, and far distance scenarios, only the subjects with at least 2 images in close and at least 1 image in both of the two other scenarios were kept. The data selection process is summarized in Table 1, which shows that the three considered scenarios result in 112 subjects and 2964 face images.

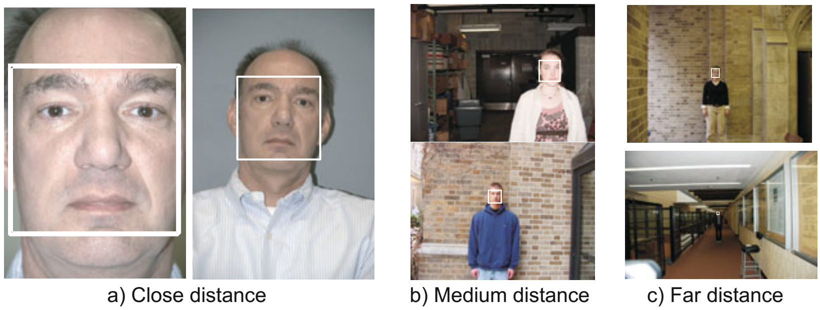

Fig. 2 Example images of the three scenarios defined: a) close distance, b) medium distance, and c) far distance. Images are collected indoors and outdoors and with different illuminations.

Table 1 Number of images of each scenario constructed from NIST MBGC v2.0 Face Visible Stills.

Num. users	Close distance	Medium distance	Far distance	Discarded images	Total
147	1539	870	713	360	**3482**
	*At least **2** images per user*	*At least **1** images per user*			
112	1468	836	660		**2964**

Table 2 Segmentation results based on errors produced by face Extractor of VeriLook SDK.

	Close distance	Medium distance	Far distance	Discarded	Total
Num. Images	1468	836	660	360	**3324**
Errors	21	151	545		**848**
Errors(%)	1.43%	18.06%	**82.57%**		

3 Scenario Analysis: Data Statistics

First of all, faces were localized and segmented (square areas) in the three acquisition scenarios using the VeriLook SDK discussed in Sect. 4.1. Segmentation results are shown in Table 2, which shows that segmentation errors increase significantly across scenarios, from only 1.43% in close distance to 82.57% in far distance. Segmentation errors here mean that the VeriLook software could not find a face in the image. For all the faces detected by VeriLook we conducted a visual check, where we observed 3 and 10 segmentation errors for medium and far distance respectively. All the segmentation errors were then manually corrected by manually marking the eyes. The face area was estimated based on the marked distance between eyes.

As a result of the defined scenarios, we observe that the sizes of the segmented faces decrease with the acquisition distance. In particular, the average face size in pixels for each scenario is: 988×988 for close, 261×261 for medium, and

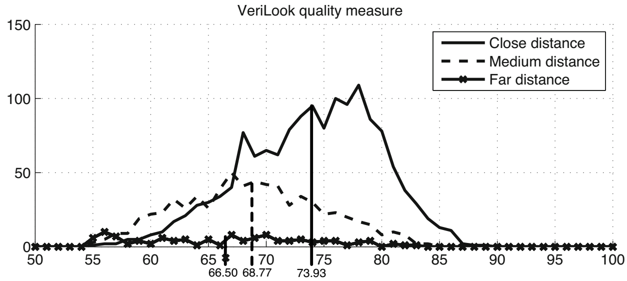

Fig. 3 Histogram of face quality measures produced by VeriLook SDK.

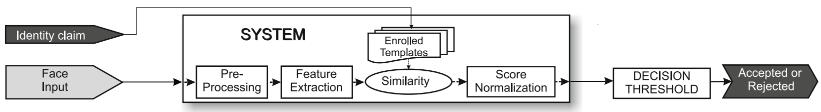

Fig. 4 Diagram of face recognition system used for VeriLook SDK, DCT-GMM and PCA-SVM.

78×78 for far distance. For the experimental work, the face size is normalized to 64×80 pixels.

Another data statistic that was computed for the three scenarios was the average face quality index provided by VeriLook (0 = lowest, 100 = highest): 73.93 for close, 68.77 for medium, and 66.50 for far distance (see Fig. 3, computed only for the faces correctly segmented by VeriLook). As stated by VeriLook providers, this quality index considers factors such as lightning, pose, and expression.

4 Scenario Analysis: Verification Performance Evaluation

4.1 Face Verification Systems

The architecture of the face recognition system used is shown in Fig. 4. In a similar way as in previous work [10], three approaches are used for face verification:

- **VeriLook SDK.** Commercial face recognition system developed by Neurotechnology[2].
- **PCA-SVM system.** This verification system uses Principal Component Analysis (PCA). The evaluated system uses normalized and cropped face images of size 64×80 (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.

[2] http://www.neurotechnology.com/

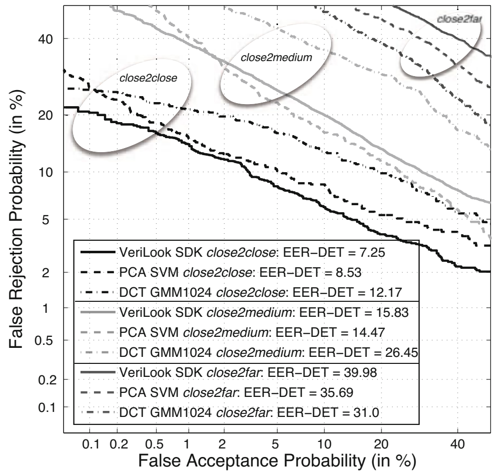

Fig. 5 Verification performance results for the three scenarios and three systems considered.

- **DCT-GMM system.** This verification system also uses face images of size 64 $\times$ 80 divided into 8 $\times$ 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 [4]. 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 [4].

4.2 Experimental Protocol

Three main experiments are defined for the verification performance assessment across scenarios:

- *Close2close.* This experiment gives an idea about the performance of the systems in ideal conditions (both enrollment and testing using close distance images). About half of the close distance subcorpus (754 images) is used for development (training the PCA subspace, SVM, etc.), and the rest (714 images) is used for testing the performance.
- *Close2medium*, and *close2far* protocol. These two other experiments use as training set the whole close distance dataset (1468 face images). For testing the performance of the systems the two other datasets are used: 836 medium distance images for *close2medium*, and 660 far distance images for *close2far*.

4.3 Results

Fig. 5 shows the verification performance for the three considered scenarios: *close2close*, *close2medium*, and *close2far*. As can be seen, VeriLook is the best of the three systems in *close2close* with an EER of around 7%. At the same time, this commercial system is the most degraded in uncontrolled conditions, with an EER close to 40% in *close2far*, much worse than the other two much simpler systems. This result corroborates the importance of analyzing and properly dealing with variability factors arising in biometrics at a distance.

Fig. 5 also shows that the GMM-based system works better in far distance conditions than the other systems, although being the less accurate in *close2close* and *close2medium*. This result demonstrates the greater generalization power of this simple recognition approach, and its robustness against uncontrolled acquisition conditions.

5 Discussion and Future Work

An experimental approach towards understanding the variability factors in face recognition at a distance has been reported. In particular, a data-driven analysis of three realistic acquisition scenarios at different distances (close, medium, and far) has been carried out as a first step towards devising adequate recognition methods capable of working in less constrained scenarios.

This analysis has been focused on: 1) data statistics (segmented face sizes and quality measures), and 2) verification performance of three systems. The results showed that the considered systems degrade significantly in the far distance scenario, being more robust to uncontrolled conditions the simplest approach.

Noteworthy, the scenarios considered in the present paper differ not only in the distance factor, but also in illumination and pose (being the illumination variability much higher in far distance than in close distance). Based on the data statistics obtained and the performance evaluation results, a study of the effects of such individual factors is source for future research.

Also, depending on the application, fusion with other biometrics would be of interest, such as in the case of pedestrian detection in order to avoid car crashings it would be very useful a fusion with gait, or also with footsteps [11] in scenarios like walking through an identification bow. This also could be used in ambient intelligence applications such as monitoring the behavior of elderly people [3].

Acknowledgements. This work has been supported by project Contexts (S2009/TIC-1485). P. Tome is supported by a FPU Fellowship from Univ. Autonoma de Madrid.

References

1. Ao, M., Yi, D., Lei, Z., Li, S.Z.: Face Recognition at a Distance: System Issues. In: Handbook of Remote Biometrics for Remote Biometrics, pp. 155–167. Springer, Heidelberg (2009)

2. Casale, P., Pujol, O., Radeva, P.: Face-to-Face Social Activity Detection Using Data Collected with a Wearable Device. In: Pattern Recognition and Image Analysis, pp. 56–63 (2009)
3. CAVIAR: Context Aware Vision using Image-based Active Recognition, `http://homepages.inf.ed.ac.uk/rbf/CAVIAR/`
4. Galbally, J., McCool, C., Fierrez, J., Marcel, S., Ortega-Garcia, J.: On the vulnerability of face verification systems to hill-climbing attacks. Pattern Recognition 43(3), 1027–1038 (2010)
5. Jafri, R., Arabnia, H.R.: Fusion of face and gait for automatic human recognition. In: Third International Conference on Information Technology: New Generations, pp. 167–173 (2008)
6. Lee, Y.J., Lee, D.H.: Research on detecting face and hands for motion-based game using web camera. In: Proc. International Conference on Security Technology (SECTECH 2008), pp. 7–12 (2008)
7. Li, S.Z., Schouten, B., Tistarelli, M.: Biometrics at a Distance: Issues, Challenges, and Prospects. In: Handbook of Remote Biometrics for Surveillance and Security, pp. 3–21. Springer, Heidelberg (2009)
8. MBGC: Multiple biometric grand challenge. NIST - National Institute of Standard and Technology, `http://face.nist.gov/mbgc/`
9. Phillips, P.J., Scruggs, W.T., O'Toole, A.J., Flynn, P.J., Bowyer, K.W., Schott, C.L., Sharpe, M.: Frvt 2006 and ice 2006 large-scale experimental results. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 831–846 (2010)
10. Tome, P., Fierrez, J., Alonso-Fernandez, F., Ortega-Garcia, J.: Scenario-based score fusion for face recognition at a distance. In: Proc. IEEE Confererence on Computer Vision and Pattern Recognition (2010)
11. Vera-Rodriguez, R., Mason, J., Evans, N.: Assessment of a Footstep Biometric Verification System. In: Advances in Pattern Recognition. Handbook of Remote Biometrics. Springer, Heidelberg (2009)
12. Wang, S., Xiong, X., Xu, Y., Wang, C., Zhang, W., Dai, X., Zhang, D.: Face-tracking as an augmented input in video games: enhancing presence, role-playing and control. In: CHI 2006: Proceedings of the SIGCHI conference on Human Factors in computing systems, pp. 1097–1106. ACM, New York (2006)
13. Zhou, S.K., Chellappa, R., Zhao, W.: Unconstrained Face Recognition. Springer, Heidelberg (2006)
14. Zuo, F., de With, P.: Real-time embedded face recognition for smart home. IEEE Transactions on Consumer Electronics 51(1), 183–190 (2005)