COMPENSATING FOR VARIABLE RECORDING CONDITIONS IN FRONTAL FACE AUTHENTICATION ALGORITHMS

A. Tefas† Y. Menguy‡ C. Kotropoulos† G. Richard‡ I. Pitas† P. Lockwood‡

†Department of Informatics
Aristotle University of Thessaloniki
Box 451, Thessaloniki 540 06, GREECE
e-mail: costas@zeus.csd.auth.gr

‡MATRA NORTEL COMMUNICATIONS
Rue JP Timbaud-BP 26
78392 Bois d' Arcy Cedex, FRANCE
e-mail: yann.menguy@matranortel.com

ABSTRACT

This paper addresses the problem of compensating for variable recording conditions such as changes in illumination, scale differences, varying face position. It is well known that the performance of any face authentication/ recognition algorithm deteriorates significantly in the presence of the aforementioned conditions as well as the expression variations. The use of simple and powerful pre-processing techniques aiming at compensating for variable recording conditions prior to the application of any authentication algorithm is proposed. It is shown that such an approach overcomes indeed the image variations and guarantees an almost stable performance for the Morphological Dynamic Link Architecture developed within the European research project M2VTS.

1. INTRODUCTION

The interest and the research activities in automatic face recognition have increased significantly over the past few years. This growth is mainly driven by application demands, such as identification for law enforcement and authentication for remote banking and access-control applications. A recent survey on face recognition can be found in [1].

A comparative study has been performed for three well known face recognition techniques, namely, the eigenfaces, the auto-association and classification neural networks, and the elastic graph matching in [2]. It has been found that the eigenfaces work well when the face images have relatively small lighting and moderate expression variations. Their performance deteriorates significantly as lighting variation increases. On the contrary, elastic graph matching (EGM) is found relatively insensitive to variations in lighting, face position and expressions. Zhang et al. attributed the robustness of EGM to the use of Gabor filters and to the rigid and deformable matching stages of the recognition algorithm. Furthermore, they reported that the performance of the neural networks they considered is upper bounded by that of the eigenface approach. In the closely related work [3], Adini et al. presented an empirical study that evaluates the sensitivity of several image representations (e.g. edge maps, directional and non-directional derivatives of Gaussian filters, images convolved by 2-D Gabor-like filters) in changes of the illumination conditions. It has been found that all the aforementioned image representations are insufficient to overcome variations due to changes in illumination direction, view-point and expressions.

Within the framework of EU-ACTS project M2VTS several frontal-face authentication algorithms have been developed. Among others the morphological dynamic link architecture (MDLA) and the morphological signal decomposition - dynamic link architecture [6, 7], the elastic graph matching (residual matching and local discriminants) [8], the optimized robust correlation [9] are mentioned. Their performance has been assessed with respect to their receiver operating characteristics (ROCs) on two databases, i.e., the M2VTS database (37 persons in 4 shots) [4] and the extended M2VTS (295 persons in 8 shots) [5]. A scalar figure of merit that can be determined from a ROC is the Equal Error Rate (EER). For example MDLA has achieved an EER of 3.7 % on M2VTS database[1] when discriminatory power coefficients were employed [7] and 6.3%–6.57% on the extended M2VTS database[2], respectively. However, both databases have been collected in a controlled environment, i.e., face size and position are constant, uniform lighting is obtained by spotlight, facial expression is neutral etc. Motivated by [2, 3], it is interesting to test the performance of the MDLA on a database recorded under conditions that simulate a real access-control to a building system or a tele-banking application. Towards this goal, a small database is collected under "real conditions" by MATRA NORTEL Communications (shortly MATRA-NORTEL database). Several sources of degradation are included in the database, such as, varying face size (i.e., scale differences), varying face position, changes in lighting and expression variations. Instead of searching for a robust image representation or a robust verification algorithm, the use of simple and powerful pre-processing techniques aiming at compensating for the aforementioned variations between the reference and the test images is proposed. Any authentication algorithm can be applied afterwards. In this paper, it is shown that such an approach overcomes indeed the image variations and guarantees an almost stable per-

This work was carried out within the framework of European ACTS project "Multi-modal Verification Techniques for Tele-services and Security Applications" (M2VTS).

[1] 5,328 client claims and 5,328 impostor claims.

[2] 400 client claims and 112,000 impostor claims.

formance for the MDLA. Similar results are expected for the other authentication algorithms developed within the M2VTS project as well.

The paper reports results on face verification and therefore it complements the previously reported works [2, 3]. Its outline is as follows. Section 2 describes the problems that occur in field tests performed under real conditions and the MATRANORTEL database. Section 3 presents the proposed face normalization technique. Experimental results that quantify the success of the proposed technique are included in Section 3 as well. Conclusions are drawn in Section 4.

2. PROBLEMS OCCURRED IN FIELD TESTS PERFORMED UNDER REAL CONDITIONS

The evaluation of any authentication algorithm on a database collected under optimal conditions is inadequate when the algorithm is to be integrated in a platform for commercial exploitation. In such a case, it is of outmost importance to evaluate the verification performance of the algorithm on field tests. This is the case of MATRANORTEL database. Although the experiments reported in this paper have been conducted on 21 persons, the size of the database progressively increases (51 persons are currently included in the database). Several sources of degradation are modeled in the database:
(a) Face size and position. In practice it is very difficult to control the position of the subject with respect to the camera.
(b) Changes in illumination. If a spotlight is not used, lighting variations occur. For example close to a window, the lighting depends strongly on the day-time and the weather.
(c) Facial expressions. In practice it is almost impossible to control the mood of the subject. The smile causes probably the largest variation of facial expressions.
In addition to images belonging to the just-mentioned cases, the database contains one set of training images (1 image per person) and one set of test images (2 images per person) recorded under "optimal" conditions. That is, a uniform white background exists in the images, uniform lighting conditions are used, the face is of neutral expression and is located at the center of the image. All the images in the database are recorded in 256 grey levels and they are of dimensions 144 $\times$ 192. Samples images from MATRANORTEL database before normalization are shown in Figure 1(a).

A set of field tests has been implemented so that the performance of three authentication algorithms developed within the M2VTS research project is correlated to each degradation source. More specifically, 441 impostor claims and another 21 client claims have been tested under each degradation source. In addition to these claims, 781 impostor claims and 38 client claims have been tested under the just-mentioned "optimal" conditions. This is tantamount to 2,445 impostor claims and 117 client claims in total. The first column in Table 1 summarizes the performance of MDLA for each source of degradation separately and in the general case that encompasses all varying conditions. From the inspection of the first column in Table 1, it becomes evident that the performance of the algorithm depends on the variable recording conditions. To alleviate such a dependency, a compensation for the different conditions is needed. To do so, the use of simple and powerful face normalization technique prior to the application of any authentication algorithm is proposed in Section 3.

3. FACE NORMALIZATION TECHNIQUE

The proposed technique is based on the detection of the facial region in the image and its splitting in two segments, the left segment and the right one. We assume that: (1) the background of the images is uniform, and (2) one only person appears in the scene. The technique is comprised of the following steps:

Step 1. Discard the uniform background by using a region growing grass-fire algorithm.

Step 2. Fit an ellipse to the facial region. Determine the near vertical semimajor axis length and the near horizontal semiminor axis length. Compensate for a rotation of the ellipse by finding its orientation.

Step 3. Split the elliptical facial region in two segments (left and right) by taking into account the symmetry about the vertical major axis of the ellipse.

Step 4. Apply a K-means clustering in each segment aiming at relating one specific cluster to the skin-like area.

Step 5. Fit an ellipse to the union of left and right clusters that are related to the skin-like areas.

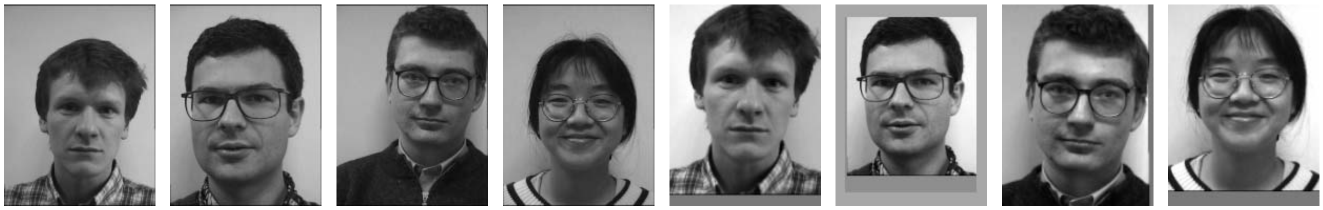
In addition to the utility of the just-described technique in detecting the facial area in the image, the technique provides: (i) an estimate of the face center that can be used in compensating for face translations, and (ii) estimates of the face height and width which are related to the length of the major and minor axes, respectively.

3.1. Face detection by ellipse fitting

The oval shape of a face can be approximated by an ellipse. Therefore, the detection of the facial area in an image can be performed by detecting an object of elliptical shape. To do so, first we have to discard the image background by using a region growing grass-fire algorithm applied after edge detection. Accordingly, the image is segmented into two regions one of which contains the facial region and the other contains the background.

The next step is to model the face-like region by an ellipse using moment-based features [10]. Let us denote the face-like area by C and the best-fit ellipse by E. An ellipse is defined by its center (x_0, y_0), its orientation θ and the length a and b of its semimajor and semiminor axes. The center of the ellipse is estimated by the center of mass of the region C. The orientation of the ellipse is computed by determining the angle between the axis of the least moment of inertia and the horizontal axis of the coordinating system, i.e.:

$$\theta = \frac{1}{2} \arctan\left(\frac{2\rho_{1,1}}{\rho_{2,0} - \rho_{0,2}}\right) \tag{1}$$



(a) (b)

Figure 1: Sample images from MATRANORTEL database (a) before (b) after normalization.

where $\rho_{i,j}$ denotes the (i,j)-central moment of the region C. The length of the semimajor axis a and the length of the semiminor axis b can be computed by evaluating the least and the greatest moments of inertia. The least and the greatest moments of inertia of an ellipse with orientation θ, I_{min} and I_{max}, are given by:

$$I_{min} = \sum_{(x,y)\in C} [(y-y_0)\cos\theta - (x-x_0)\sin\theta]^2 \quad (2)$$

$$I_{max} = \sum_{(x,y)\in C} [(y-y_0)\sin\theta + (x-x_0)\cos\theta]^2 \quad (3)$$

where x, y denote the horizontal and vertical coordinate of a pixel. Accordingly, a and b are given by:

$$a = \left(\frac{4}{\pi}\right)^{\frac{1}{4}} \left[\frac{(I_{max})^3}{I_{min}}\right]^{\frac{1}{8}}, \qquad b = \left(\frac{4}{\pi}\right)^{\frac{1}{4}} \left[\frac{(I_{min})^3}{I_{max}}\right]^{\frac{1}{8}} \quad (4)$$

respectively. The objective is to find the ellipse which models better the given region. The objective is satisfied by iteratively maximizing the measure:

$$M = \sum_{(x,y)\in E\cap C} 1 - \sum_{(x,y)\in E\cap C^c} 1 \quad (5)$$

where C^c denotes the complement of the region C (background). The maximization of the above measure corresponds to the maximization of the number of correctly modeled pixels (i.e., $(x,y) \in E\cap C$) and the minimization of the number of incorrectly modeled pixels (i.e., $(x,y) \in E \cap C^c$). In each step of the algorithm the region C is modified to be the intersection between the ellipse E and the region C of the previous step. The parameters of the ellipse are recomputed and the measure M is re-evaluated. The procedure is repeated until a maximum for the measure M is reached. By using this iterative algorithm the ellipse fitting becomes more robust to noise, i.e., to pixels that correspond to clothes, hair, etc.

The first ellipse that is determined is a coarse approximation of the facial area, because the hair and maybe parts of the clothes are included in it. To overcome this problem the ellipse is subdivided into its left and right segments with respect to the vertical axis. The next step is to apply a clustering algorithm to each segment of the ellipse, separately. By choosing K=2, a K-means algorithm hopefully succeeds to relate the skin-like areas with a single cluster in each segment. The union of clusters in the left and right segment that correspond to skin-like areas is modeled by an ellipse using the algorithm described above. The latter ellipse is the best-fit ellipse we searched for.

3.2. Lighting normalization

Let us compute the mean intensity of the pixel values that belong to each segment. Changes in illumination usually result in a face the left segment of which is darker than the right one or vice versa. Moreover, we assume that the illumination conditions inside the left and right segment are uniform. To compensate for the aforementioned effect, the mean intensities of both face segments should be equalized. Let m_L, m_R be the mean intensity values in the left and the right segment, respectively. The initial image $I(x,y)$ is transformed so that the left and right segments of the normalized image which is obtained, $I_N(x,y)$, have the same (desired) mean intensity I_d:

$$I_N(x,y) = I_d \left(\frac{\frac{1}{m_R} - \frac{1}{m_L}}{1+\exp(\frac{x_0 - x}{\lambda})} + \frac{1}{m_L} \right) I(x,y) \quad (6)$$

where λ simply controls the slope of the sigmoidal function that appears in the denominator of the first fractional term inside parentheses. Let $\mathcal{L}$ be the image region in which $\exp(\frac{x_0-x}{\lambda}) \to \infty$ and $\mathcal{R}$ be the image region in which $\exp(\frac{x_0-x}{\lambda}) \approx 0$. These regions correspond to the left and the right segments of the best-fit ellipse modeling the face, respectively. It can be easily proven that:

$$\mathrm{E}[I_N(x,y),\ (x,y)\in\mathcal{L}] = \mathrm{E}[I_N(x,y),\ (x,y)\in\mathcal{R}] = I_d \quad (7)$$

which satisfies our objective.

3.3. Face position and size

Varying face position and size are easily compensated for if the face is accurately approximated by an ellipse. The problem of varying face position can be solved by translating the initial image so that the center of the ellipse coincides with the image plane center. The width of the face can be approximated by the length of the minor axis $2b$. Similarly, the height of the face can be approximated by the major axis length $2a$. Size (i.e., scale) normalization can be achieved by resizing the image with a horizontal scale factor $\frac{W_d}{2b}$ and a vertical scale factor $\frac{H_d}{2a}$, where W_d and H_d are the desired width and height of the normalized face, respectively. Image resizing has been achieved by linear interpolation. Examples of normalized face images from MATRANORTEL database are shown in Figure 1(b).

3.4. Impact on the verification performance of MDLA

In this paper we report the success of the face normalization technique when it is used prior to MDLA on MATRA-

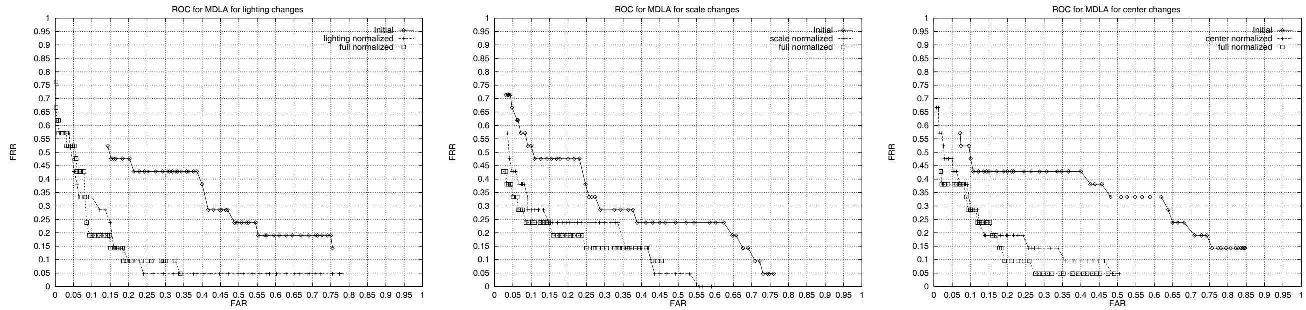


Figure 2: ROC curves of MDLA with and without employing face normalization on MATRANORTEL database.

NORTEL database. The ROCs of MDLA with and without the face normalization technique for three degradation sources, namely, changes in lighting, varying face size and position, are plotted in Figure 2. In each subplot, first the ROC without applying the proposed face normalization technique is given. The ROC, when we compensate for the specific source of degradation, is plotted next. Finally, the ROC when we compensate for the three sources of degradation simultaneously is also shown. The gain in EER when the proposed face normalization technique is employed is shown in Table 1 for all degradation sources as well as in the general case. A significant drop of 7.3% in EER (which amounts to a 33 % relative drop) is achieved in the general case which includes images recorded under any conditions.

Table 1: EERs achieved on MATRANORTEL database for MDLA

Conditions	Initial EER (%)	Normalized EER (%)
Optimal	12	9.5
lighting	33	15
face size	28	19
face position	23	17
expressions	17	13
general	22	14.7

4. CONCLUSIONS

A face normalization method that succeeds to compensate for lighting changes, varying face size and position has been proposed. It guarantees an almost stable verification performance of MDLA in field tests under any recording conditions. As a by-product, facial expressions are compensated for as well. The utility of the method when it is used prior to other authentication algorithms developed within M2VTS is subject of ongoing research.

5. REFERENCES

[1] R. Chellapa, C.L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," Proceedings of the IEEE, vol. 83, no. 5, pp. 705-740, May 1995.

[2] J. Zhang, Y. Yan, and M. Lades, "Face recognition: Eigenface, elastic matching and neural nets," Proceedings of the IEEE, vol. 85, no. 9, pp. 1423-1435, September 1997.

[3] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 721–732, July 1997.

[4] S. Pigeon, and L. Vandendorpe, "The M2VTS multimodal face database," in Lecture Notes in Computer Science: Audio- and Video- based Biometric Person Authentication (J. Bigün, G. Chollet and G. Borgefors, Eds.), vol. 1206, pp. 403–409, 1997.

[5] K. Messer, J. Matas, and J. Kittler. "Acquisition of a large database for biometric identity verification," in Proc. of BioSig 98, June 1998.

[6] C. Kotropoulos, and I. Pitas, "Face authentication based on morphological grid matching," in Proc. of the IEEE Int. Conf. on Image Processing, vol. I, pp. 105–108, Santa Barbara, U.S.A., October 1997.

[7] C. Kotropoulos, A. Tefas, and I. Pitas, "Frontal face authentication using variants of Dynamic Link Matching based on mathematical morphology," in Proc. of the IEEE Int. Conf. on Image Processing, Chicago, U.S.A., October 1998, to appear.

[8] B. Duc, S. Fischer, and J. Bigün, "Face authentication with Gabor information on deformable graphs," IEEE Trans. on Image Processing, accepted for publication 1998.

[9] J. Matas, K. Jonsson, and J. Kittler, "Fast face localisation and verification," in Proc. of British Machine Vision Conference, 1997.

[10] A.K. Jain, Fundamentals of Digital Image Processing. Prentice-Hall, Englewood Cliffs, NJ, 1989.