TURNING A FRONTAL FACE VERIFICATION TECHNIQUE TO A SYSTEM PROTOTYPE

Anastasios Markoulidis *Anastasios Tefas* *Constantine Kotropoulos* *Ioannis Pitas*

Department of Informatics
Aristotle University of Thessaloniki
Box 451, Thessaloniki 54124, GREECE
e-mail: {amarkoul,tefas,costas,pitas}@zeus.csd.auth.gr

ABSTRACT

In this paper a Windows-based platform for face verification is described. The platform implements the Morphological Dynamic Link Architecture (MDLA). The paper deals with the calculation of the thresholds during the training procedure as well as the computational gain that results from using incremental methods for updating the face verification system. Experimental results for two different databases are also presented in this paper; the first one is the extended M2VTS database which contains 295 persons' video data in 8 shots recorded under optimal conditions, while the second database was created in the Department of Informatics in the Aristotle University of Thessaloniki under real conditions. The experiments indicate that MDLA is a quite successful method for face verification.

1. INTRODUCTION

Biometrics is a well-known research area, whose goal is to discriminate one person from another based on certain characteristics. Research has been focused on characteristics that are related either to person's physiology (e.g. face, fingerprints, hand and finger geometry, iris, retinal scan, ear geometry) or to person's behavior (e.g. signature, voice, keystroke rhythm), or even to person's biology (e.g. DNA, thermogram, human scent). Obviously the most interesting part of biometrics is automatic face recognition. Its significant development over the past fifteen years is mainly driven by application demands, such as identification for law enforcement and authentication for remote banking and access-control applications. A survey on face recognition techniques can be found in [1].

During the last 15 years many approaches for face authentication have been presented. These methods can be classified in three categories: a) approaches based on geometrical features of the face, such as distance between eyes, size of the lips etc., b) algorithms based on comparisons between well defined regions- of the face (e.g. mouth, nose, eyes etc.), and c) approaches based on statistical methods of linear projection of the face image. One of these algorithms, which belongs to the third category, is the *DLA* (Dynamic Link Architecture) or *EGM* (Elastic Graph Matching) [2]. In this method a sparse grid is overlaid on the face image during the training procedure, and the output of the 2-D Gabor wavelets is computed on every grid node; the result is the feature vector for the person whose face is displayed on the image. The grid can be placed onto the face region that is either manually defined or extracted using a face detection algorithm. During the test procedure the reference grid is placed on the facial image of the test person and deformed in order to minimize the corresponding distance between the grids.

A variant of DLA, which is called *MDLA* (Morphological Dynamic Link Architecture) [3] has been introduced. MDLA's major difference from DLA is that it uses the multiscale dilation and erosion of the image instead of the Gabor wavelets. The most important advantage of dilation and erosion is that they deal with the local maxima and minima of an image, and the basic features of the face are associated either to local minima (eyebrows/eyes, nostrils, mouth corners) or maxima (nose tip).

In Section 2, the mathematical framework and the concept of the MDLA algorithm is briefly reviewed. Section 3 contains the algorithm for calculating the thresholds during the training procedure, while Section 4 describes the process of incremental training of the MDLA face verification system when the database is modified. In Section 5 the problems occured in real world tests are described and the performance of MDLA is assessed using 2 databases. Finally, Section 6 presents *BioProof*, a system prototype which implements the algorithms described in this paper.

2. MATHEMATICAL FRAMEWORK OF MDLA

Let us suppose that we have a digital image $f(\mathbf{x}) : \mathcal{D} \subseteq \mathcal{Z}^2 \rightarrow \mathcal{R}$ and a function $g(\mathbf{x}) : \mathcal{G} \subseteq \mathcal{Z}^2 \rightarrow \mathcal{R}$, which is called *structuring function*. The *dilation* of $f(\mathbf{x})$ by $g(\mathbf{x})$ [4] is defined by:

$$(f \oplus g)(\mathbf{x}) = \max_{\mathbf{z}\in\mathcal{G},\mathbf{x}-\mathbf{z}\in\mathcal{D}} \{f(\mathbf{x}-\mathbf{z}) + g(\mathbf{z})\}. \quad (1)$$

Its complementary operation, the *erosion* is defined by:

$$(f \ominus g)(\mathbf{x}) = \min_{\mathbf{z}\in\mathcal{G},\mathbf{x}+\mathbf{z}\in\mathcal{D}} \{f(\mathbf{x}+\mathbf{z}) - g(\mathbf{z})\}. \quad (2)$$

The multiscale dilation-erosion is defined as:

$$(f \star g_\sigma)(\mathbf{x}) = \begin{cases} (f \oplus g_\sigma)(\mathbf{x}) & \text{if } \sigma > 0 \\ f(\mathbf{x}) & \text{if } \sigma = 0 \\ (f \ominus g_{|\sigma|})(\mathbf{x}) & \text{if } \sigma < 0 \end{cases} \quad (3)$$

By applying the multiscale dilation-erosion on the facial image, the feature vector for each grid node $\mathbf{x}$ is received, i.e. $\mathbf{j}(\mathbf{x}) = ((f \star g_{\sigma_m})(\mathbf{x}), ..., (f \star g_1)(\mathbf{x}), f(\mathbf{x}), (f \star g_{-1})(\mathbf{x}), ..., (f \star g_{\sigma_{-m}})(\mathbf{x}))$. In the experiments described in Section 5 the function $g(\mathbf{x}) = 0$ was used as a structuring function, because it allowed the use of recursive calculation according to the scale [3]. Also, in these experiments, σ_m was equal to 9 and an 8×8 sparse grid was created

Research presented in this paper was reported in the framerowk of COST275.

by measuring the feature vectors $\mathbf{j}(\mathbf{x})$ at equally spaced nodes over the output of the face detection algorithm described in [5].

During the verification procedure, the reference grid (i.e. the grid which belongs to the person whose identity is claimed by the test person) is placed on the test image and is deformed. The L_2 norm of the difference of the feature vectors at the node i is defined as a measure of comparison between the reference and the deformed grid at the node i:

$$S_\upsilon(\mathbf{j}(\mathbf{x}_i^{\mathbf{t}}), \mathbf{j}(\mathbf{x}_i^{\mathbf{r}})) = \| \mathbf{j}(\mathbf{x}_i^{\mathbf{t}}) - \mathbf{j}(\mathbf{x}_i^{\mathbf{r}}) \| . \tag{4}$$

Assuming that $\mathcal{V}$ is the set of the grid nodes and that $\mathcal{N}(i)$ is an area around the node i (in our experiments a set of 4 nodes has been used), the target set during the deformation procedure is to minimize the following cost function:

$$C(\{\mathbf{x}_i^t\}) = \sum_{i \in \mathcal{V}} \left\{ S_\upsilon(\mathbf{j}(\mathbf{x}_\mathbf{i}^{\mathbf{t}}), \mathbf{j}(\mathbf{x}_\mathbf{i}^{\mathbf{r}})) + \lambda \sum_{j \in \mathcal{N}(i)} S_e(i,j) \right\} \tag{5}$$

where $S_e(i,j) = S_e(\mathbf{d}_{ij}^t, \mathbf{d}_{ij}^r) = \| \mathbf{d}_{ij}^t - \mathbf{d}_{ij}^r \|, \quad \mathbf{d}_{ij} = (\mathbf{x}_i - \mathbf{x}_j), \quad \forall i \in \mathcal{V}, j \in \mathcal{N}(i)$, is an additional cost function which can be used to penalize grid deformations. In order to minimize the function (5), a propabilistic hill climbing algorithm, like simulated annealing, can be used. The whole way grid matching procedure works can be seen in Figure 1.

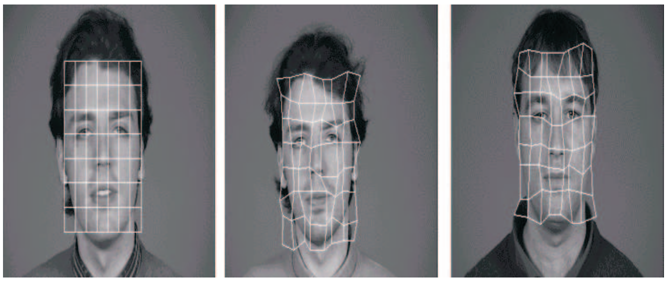

Figure 1: Grid matching procedure in MDLA. (a) Reference grid for person BP. (b) Best grid for test person BP after elastic graph matching with the reference grid. (c) Best grid for test person BS after elastic graph matching with the reference grid for person BP. (The images are from the M2VTS database)

3. TRAINING PROCEDURE

The previously mentioned algorithm is dealing with the calculation of the "distance" between a reference grid and a deformed one computed over a test face image. But if a person wants to be identified successfully the distance between deformed grids computed over his/her daily frontal face images and his/her reference grids stored in a database must exceed a threshold. The threshold for each person should be calculated during the training procedure and tested for its efficiency. In the present work the following algorithm for the calculation of the thresholds and their testing is proposed:

1. The distances between the reference grids of the users recorded into the system database are calculated.
2. For each pair of users m and n, the minimum distance between the reference grids of the user m and the deformed grids of the user n, $D_{min}(m,n)$ is found. With these elements, the $N \times N$ ($N =$ amount of users) table $MinDist$ is created. The table has the following form:

$$MinDist = \begin{bmatrix} D_{min}(1,1) & \dots & D_{min}(1,N) \\ D_{min}(2,1) & \dots & D_{min}(2,N) \\ D_{min}(3,1) & \dots & D_{min}(3,N) \\ \vdots & \vdots & \vdots \\ D_{min}(N,1) & \dots & D_{min}(N,N) \end{bmatrix} \tag{6}$$

   During the search for the $D_{min}(m,n)$ a reference grid from each user is excluded and, of course, the distances which "involve" this grid. The set of reference grids is called $ShotOut$ and will be used during the test procedure.
3. The following procedure is iterated N times:
   - A user, who is called *Left* is temporarily excluded from the training set. On each iteration a different user is excluded and the training procedure is repeated (leave-one-out procedure).
   - A threshold for each user i in the training set is defined by:

     $$T_i(\mathit{Left}; Q) = OS_{(1+Q)}\{D_{min}(j,i)\} \tag{7}$$

     where $j \neq i$ and $j, i \neq \mathit{Left}$, $OS_{(1+Q)}\{\mathcal{A}\}$ is the $(1+Q)$-th order statistic of a set $\mathcal{A}$ and Q is a measure of "strictness" of the threshold selection.
4. The final threshold $T(i)$ for a user i is defined by the average of the thresholds for the user i previously calculated on step 3.

After the training procedure, the test procedure must be executed in order to examine the performance of the system using the thresholds obtained during the training procedure. Let us denote the set of the recorded users as $\mathcal{S}$. The test procedure includes the following steps:

1. Let us denote the set of deformed grids of the user *Left* as $\mathcal{L}$. The number of *False Acceptances* (*FA*) is obtained as follows:
   - For each $X \in \mathcal{S}$ let us denote its set of reference grids as $\mathcal{Y}$. For each $i \in \mathcal{L}$ and each $j \in \mathcal{Y}$, compute the distance D between the j reference grid of the user X and the i deformed grid of the user $Left$, according to equation (5).
   - Count a false acceptance if $D \leq T(X)$.
2. For each $X \in \mathcal{S}$ (apart from *Left*) the $ShotOut$ images $X_{ShotOut}$ are taken into account. The number of *False Rejections* (*FR*) is calculated as follows:
   - For each $j \in \mathcal{Y}$ compute the distance D between the j reference grid of user X and the $ShotOut$ deformed grid of the same user according to equation (5).
   - Count a false rejection if $D > T(X)$.

This procedure is being iterated until every user and every training grid is excluded. The ratio between the sum of these results and the total number of tests gives us the total *False Acceptance Rate* (*FAR*) and *False Rejection Rate* (*FRR*) of the system for the currently chosen threshold. By calculating FAR and FRR for all possible Q the *ROC* (Receiver Operating Characteristics) curve of the system is produced.

4. UPDATING THE SYSTEM THRESHOLDS

The procedure described in the previous section is applied in the case when the system database is fixed. So when the database is updated, i.e. clients are added, removed or have their images updated, the thresholds for the users recorded in the training set must be recalculated. In order to do that, it is not necessary to retrain the system from the beginning. Instead, the system can be updated with the new data. To do so, the following procedure is used:

- When a new user is added the only distances which have to be calculated are those between the deformed grids on the images of the new user and the reference grids of the old users (and vice versa). The corresponding minimum distances D_{min} are found, and the $MinDist$ table is reformed:

$$MinDist' = \left[\begin{array}{c|c} MinDist & D_{min}(i,N) \\ \hline D_{min}(N,i) & D_{min}(N,N) \end{array} \right] \qquad (8)$$

- When a user is deleted from the database the corresponding minimum distances should be also deleted. For example, if the user 2 is deleted, the second column and the second row of the $MinDist$ table will be deleted.
- Finally, during the updating of the images of a particular user i, the distances that should be changed in the $MinDist$ table (this is not compulsory since the new distances may be greater than the already known minimum distances) are those involving the following pairs of users:

  1. $j - i,\ j \neq i$ (or vice versa), and
  2. $i - i$ (distances between reference grids and deformed grids of the same user)

  The dimensions of the $MinDist$ table will remain unchanged.

The computational gain of using incremental training instead of using batch training of the system is quite important, as can be seen in Table 1. Let us suppose that there are N users in the system database, and that every user has been recorded with M images. Then the complexity of the procedure of updating the thresholds (which is the amount of calculated distances) drops from $O(M^2N^2)$ (in batch training) to $O(M^2N)$ when a new user is added; to 0, when there is a user deletion; and to $O(M^2N)$ in the worst case, when a user has his/her images updated.

5. EXPERIMENTAL RESULTS AND PROBLEMS OCCURRED UNDER REAL CONDITIONS

The MDLA algorithm has been tested on two databases. The first one was the extended M2VTS database (or XM2VTS) [7]. This database contains 295 persons' video data which include speech and image sequences of rotated heads. Eight recordings of the 295 persons have been collected under well-controlled recording conditions. The design of a person verification system based on MDLA was split in three procedures: the training, the evaluation and the test procedure [8].

Action	Batch Training	Incremental Training
Add	$M^2N^2 + 2M^2N + M^2$ $-MN - M$	$2M^2N + M^2 - M$
Delete	$M^2N^2 - 2M^2N + M^2$ $-MN + M$	0
Edit	$M^2N^2 - MN$	$a(2MN - a - 1)$ (a=amount of edited images, $1 \leq a \leq M$)

Table 1: Computational savings between batch and incremental training. Each entity denotes the amount of distances between two training grids that are calculated during every action of updating.

1. During the training procedure a subset of 200 persons (i.e., training clients) was used. This resulted in 200 client claims and 39800 impostor claims. An EER≈3.0% was achieved on this training set.
2. A second subset of 25 persons was selected to evaluate the performance of the verification algorithm. These persons acted as evaluation impostors. To implement the evaluation procedure, another three images were used for each client and eight images were employed for each evaluation impostor. By using the just-mentioned set of images 600 client claims and 40000 impostor ones were produced. An EER≈8.0% was obtained from the evaluation set.
3. A third subset of 70 persons was selected to constitute the set of test impostors. two new frontal images were used to implement client accesses, and eight frontal images for each test were used for each test impostor. By using this set of images 400 client claims and 112000 impostor claims were produced. An EER≈6.57% was achieved on the test set.

The evaluation of any authentication algorithm on a database collected under well-controlled conditions is not adequate when the algorithm is to be integrated into a platform for commercial use. In such case, it is very important to evaluate the performance of the authentication algorithm on field tests. However, there are several problems that must be considered during the evaluation of every face recognition-verification algorithm. The most important are:

(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.

(d) *Background*. When the background is uniform, face detection is a relatively easy task. However, most of the face detection algorithms have serious problems when the background is complex.

The second database was created in the Artificial Intelligence and Information Analysis Laboratory (AIIA) of the Department of Informatics, Aristotle University of Thessaloniki, Greece. This small database comprised by 85 color images from 17 users (5 images per user) at a resolution of 139×109 pixels. In these images the face position was controlled for every user, the illumination level was, in general, stable and the background was complex. Each image contains a face with different face expression. In Figure 2 images from this small database are displayed. The training procedure resulted in 1360 client claims and 1360 impostor claims. The receiver operating characteristics curve of the training set is plotted in Figure 3, and it is seen that an EER≈4.63% is achieved.

Figure 2: Images from the database created in the Department of Informatics of the Aristotle University of Thessaloniki.

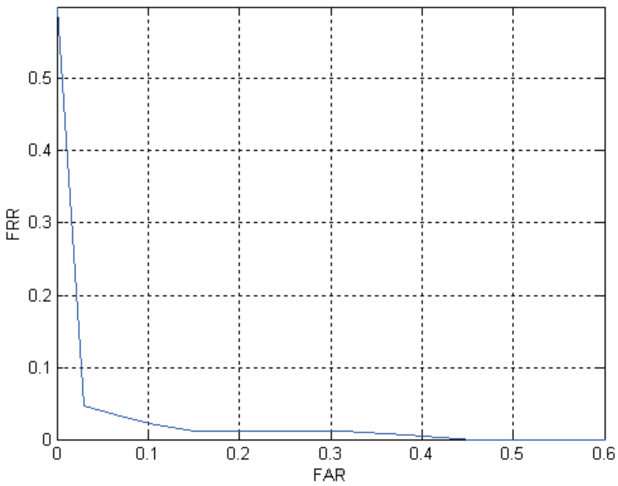

Figure 3: ROC curve for the database created in the Department of Informatics of the Aristotle University of Thessaloniki.

6. IMPLEMENTATION

The MDLA algorithm was implemented in a Windows environment called *BioProof*. Some of the features of the program are the following:

- recording from 1 to 10 images for each user and option for recording sound sample (in the future, voice verification algorithms can be integrated in the system prototype)
- full set of capabilities for manipulating the users' catalog and the catalog of successful and unsuccessful attempts for gaining access (addition, deletion, editing, printing), and
- threshold updating based on the number of successful and unsuccessful attempts for gaining access.

The last feature is based on the fact that if a person already recorded in the system database decides to change some of his/her facial characteristics (e.g. haircut, beard etc.), it will be difficult for the system to authenticate him. In order to solve this problem, the images (not necessarily all of them) of successful and unsuccessful claims are added in the set of the images already recorded in the system database to rebuild the training set. After that, the training procedure is implemented as in Section 3.

7. CONCLUSIONS

In this paper a Windows-based platform for face verification was described. The platform implemented the MDLA algorithm, a well established method for frontal face verification, which is based on multiscale dilation-erosion of the facial image. The experiments indicated that the proposed method is a very successful one in frontal face verification.

8. 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] M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. v.d. Malsburg, R. P. Wurtz and W. Konen, "Distortion invariant object recognition in the Dynamic Link Architecture", *IEEE Transactions on Computers*, vol. 42, no. 3, pp. 300-311, March 1993.

[3] C. Kotropoulos, A. Tefas and I. Pitas, "Frontal face authentication using morphological elastic graph matching", *IEEE Transactions on Image Processing*, vol. 9, pp. 555-560, April 2000.

[4] I. Pitas and A. N. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications*, Kluwer Academic Publications, Norwell, MA, 1990.

[5] C. Kotropoulos and I. Pitas, "Rule-based face detection in frontal views", *Proceedings of the IEEE International Conference on Acoustics, Speece and Signal Processing (ICASSP 97)*, pp. 2537-5240, Munich, Germany 1997.

[6] C. Kotropoulos, A. Tefas, I. Pitas, C. Fernández and F. Fernández, "Performance assessment of Morphological Dynamic Link Architecture under optimal and real operating conditions", *Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image processing (NSIP'99)*, vol. 1, pp. 341-345, Antalya, Turkey, 20-23 June 1999.

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

[8] C. Kotropoulos, A. Tefas and I. Pitas, "Morphological Elastic Graph Matching applied to frontal face authentication under well-controlled and real conditions", *Pattern Recognition*, vol. 33, no. 12, pp. 31-43, October 2000.