Face Authentication Based on Morphological Grid Matching

C. Kotropoulos I. Pitas

Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki 540 06, GREECE
E-mail: {costas, pitas}@zeus.csd.auth.gr

Abstract

A novel dynamic link architecture based on multiscale morphological dilation-erosion is proposed for face verification in a cooperative scenario where the candidates claim an identity that is to be checked. The performance of the morphological dynamic link architecture (MDLA) is evaluated in terms of the receiver operating characteristic (ROC) for several threshold selections on the matching error in the M2VTS database. The experimental results indicate that the proposed method outperforms the dynamic link matching with Gabor based feature vectors.

1 Introduction

Face recognition has exhibited a tremendous growth for more than two decades. A critical survey of the literature related to human and machine face recognition can be found in [1]. Two main categories for face recognition techniques are identified: those employing geometrical features and those using grey-level information (e.g. the eigenface approach). An approach that exploits both sources of information, that is, the grey-level information and shape information, is the so-called *Dynamic Link Architecture* (DLA) [2]. This algorithm is split in the training phase and the recall phase. In the training phase, the objective is to build a sparse grid for each person included in the reference set. Towards this goal a sparse grid is overlaid on the facial region of a person's digital image and the response of a set of 2D Gabor filters tuned to different orientations and scales is measured at the grid nodes. The responses of Gabor filters form a *feature vector* at each node. In the recall phase, the reference grid of each person is overlaid on the face image of a test person and is deformed so that a cost function is minimized. The cost function is based on a norm of differences between the feature vectors stored at the nodes of the reference grids and the feature vectors computed at variable pixel coordinates in the test image as well as the grid distortion between the reference grid and the variable graph built on the image of the test person. Therefore, the cost function is a quality measure of the elastic graph matching of the reference (or model) grid to the variable test graph. Several norms of the difference between the feature vectors based either on the magnitude or on the phase of the response of Gabor filters are proved inadequate to discriminate impostors against the authentic persons [3]. An automatic weighting of the nodes according to their significance by employing local discriminants is proposed in [4]. It is demonstrated that such an approach yields a significant performance improvement in DLA.

In this paper, we shall confine ourselves to the standard DLA without weighting the contribution of each node. That is, the experimental results presented do not rely on linear discriminant analysis. DLA employs Gabor-based feature vectors at each node. Their computation relies on floating point arithmetic operations (i.e., FFTs) and it is generally time consuming. Motivated by this fact, a novel dynamic link architecture based on multiscale morphological dilation-erosion, the so-called *Morphological Dynamic Link Architecture* (MDLA), is proposed and tested for face authentication. That is, we propose the substitution of the responses of a set of Gabor filters by the multiscale dilation-erosion of the original image by a scaled structuring function [6]. There are several reasons supporting this decision. Among others, dilations and erosions deal with the local extrema in an image. Therefore, they are well-suited for facial feature representation, because key facial features are associated either to local minima (e.g. eyebrows/eyes, nostrils, endpoints of lips etc.) or to local maxima (e.g. the nose tip).

Several structuring functions have been tested, e.g. the hemisphere, the flat structuring function and the circular paraboloid. Efficient algorithms for computing the multiscale dilation-erosion are described. Another issue studied in detail is the elastic graph matching procedure outlined above.

The performance of the MDLA is evaluated in terms of the false acceptance and false rejection rates as well as the receiver operating characteristics (ROCs) for several threshold selections on the matching error in M2VTS database [9]. A few details of the M2VTS database are included in Section 3. The experimental results verify the superiority of the proposed method over the dynamic link matching with Gabor-based feature vectors.

2 Dynamic link matching with fast multiscale morphological dilation-erosion

An alternative to linear techniques used for generating an information pyramid is the scale-space morphological techniques. In this paper, we propose the substitution of Gabor-based feature vectors used in dynamic link matching by the *multiscale morphological dilation-erosion* [6]. The multiscale morphological dilation-erosion is based on the two fundamental operations of the grayscale morphology, namely the *dilation* and the *erosion*. Let $\mathbb{R}$ and $\mathbb{Z}$ denote the set of real and integer numbers, respectively. Given an image $f(\mathbf{x}) : \mathcal{D} \subseteq \mathbb{Z}^2 \rightarrow \mathbb{R}$ and a structuring function $g(\mathbf{x}) : \mathcal{G} \subseteq \mathbb{Z}^2 \rightarrow \mathbb{R}$, the dilation of the image $f(\mathbf{x})$ by $g(\mathbf{x})$ and its complementary operation, the erosion, are denoted by $(f \oplus g)(\mathbf{x})$ and $(f \ominus g)(\mathbf{x})$, respectively. Their definitions can be found in [8]. If the structuring function is chosen to be scale-dependent, that is $g_\sigma(\mathbf{z}) = |\sigma| g(|\sigma|^{-1}\, \mathbf{z})\ \forall \mathbf{z} \in \mathcal{G} :, \|\mathbf{z}\| \leq |\sigma|$, then the morphological operations become scale-dependent as well. Suitable structuring functions are described in [6, 8]. The following structuring functions have been tested: (i) the scaled hemisphere [6]:

$$g_\sigma(\mathbf{z}) = |\sigma| \left(\sqrt{1 - (|\sigma|^{-1} \|\mathbf{z}\|)^2} - 1 \right) \qquad (1)$$

(ii) the flat structuring function $g_\sigma(\mathbf{z}) = 0$ that enables fast running max/min selection calculations [8], and (iii) the circular paraboloid structuring function [7]:

$$g_\sigma(\mathbf{z}) = -|\sigma| \frac{\|\mathbf{z}\|^2}{\sigma^2} \qquad (2)$$

that allows for separable morphological operations, because $g_\sigma(z_1, z_2) = g_\sigma^{(1)}(z_1) + g_\sigma^{(1)}(z_2)$ where $g_\sigma^{(1)}(\cdot)$ denotes the one-dimensional structuring function. It can easily be seen that [7]:

$$(f \oplus g_\sigma)(x_1, x_2) = \max_{z_1 \in \mathcal{G}_\sigma^{(1)}} \left(\gamma(x_1 - z_1, x_2) + g_\sigma^{(1)}(z_1) \right) \qquad (3)$$

$$\gamma(x_1, x_2) = \max_{z_2 \in \mathcal{G}_\sigma^{(2)}} \left(f(x_1, x_2 - z_2) + g_\sigma^{(2)}(z_2) \right) \qquad (4)$$

where $\mathcal{G}_\sigma^{(i)}, i = 1, 2$ are the projections of $\mathcal{G}$ on the axes. A similar decomposition can be found for the erosion as well. Moreover, let us suppose that the maximum occurs in (4) for $z_2 = \xi$. It can be proved that:

$$\gamma(x_1, x_2 + 1) = \max_{z_2 = -\sigma}^{\xi+1} \left(f(x_1, x_2 + 1 - z_2) + g_\sigma^{(1)}(z_2) \right). \qquad (5)$$

Eq. (5) enables a recursive separable implementation of the grayscale dilation. Similarly, the recursive implementation for grayscale erosion is based on:

$$\hat{\gamma}(x_1, x_2 + 1) = \min_{z_2 = \hat{\xi} - 1}^{\sigma} \left(f(x_1, x_2 + 1 + z_2) - g_\sigma^{(1)}(z_2) \right) \qquad (6)$$

where $\hat{\gamma}(x_1, x_2)$ is defined by:

$$\hat{\gamma}(x_1, x_2) = \min_{z_2 \in \mathcal{G}_\sigma^{(2)}} \left(f(x_1, x_2 + z_2) - g_\sigma^{(1)}(z_2) \right) \qquad (7)$$

and $z_2 = \hat{\xi}$ is the coordinate where $\hat{\gamma}(\cdot)$ attains the minimum value. The derivation of (5) and (6) follows similar lines as in [7]. However, they should substitute the wrong end results presented in [7, pp. 83-84].

Accordingly, the multiscale dilation-erosion of the image $f(\mathbf{x})$ by $g_\sigma(\mathbf{x})$ is defined by [6]:

$$(f \star g_\sigma)(\mathbf{x}) = \begin{cases} (f \oplus g_\sigma)(\mathbf{x}) & \text{if } \sigma > 0 \\ f(\mathbf{x}) & \text{if } \sigma = 0 \\ (f \ominus g_\sigma)(\mathbf{x}) & \text{if } \sigma < 0. \end{cases} \qquad (8)$$

The outputs of multiscale dilation-erosion for $\sigma = -9, \ldots, 9$ form the feature vectors located at the grid node $\mathbf{x}$:

$$\mathbf{J}(\mathbf{x}) = ((f \star g_9)(\mathbf{x}), \ldots, f(\mathbf{x}), \ldots, (f \star g_{-9})(\mathbf{x}))\,. \qquad (9)$$

Table 1 summarizes the time needed to compute the multiscale dilation-erosion for $\sigma = -9, \ldots, 9$ for a typical facial image of dimensions 286×350 from M2VTS database in a SUN Sparc Ultra 1 Enterprise 3000 workstation with 64 MB RAM. It is seen that the recursive separable implementation with the paraboloid structuring function is the fastest method. Moreover, it has been found that the choice of the structuring function does not affect the verification efficiency of the proposed MDLA.

Table 1: Computational time for multiscale dilation-erosion

Structuring function	Algorithm	Time (sec)
hemisphere	straightforward	65.18
flat	straightforward	25
	running	6.98
paraboloid	separable	4.63
	recursive-separable	4.61

Let the superscripts t and r denote a test and a reference person (or grid), respectively. The L_2 norm of the difference between the feature vectors at the i-th grid node has been used as a (signal) similarity measure, i.e.:

$$S_v(\mathbf{J}(\mathbf{x}_i^t), \mathbf{J}(\mathbf{x}_i^r)) = \|\mathbf{J}(\mathbf{x}_i^t) - \mathbf{J}(\mathbf{x}_i^r)\|. \qquad (10)$$

As in DLA [2], the quality of a match is evaluated by taking into account the grid deformation as well. Let us denote by $\mathcal{V}$ the set of grid nodes. The grid nodes are simply the vertices of a graph. Let also $\mathcal{N}(i)$ denote the neighborhood of vertex i. A four-connected neighborhood has been used in our case. An additional cost function:

$$S_e(i,j) = S_e(\mathbf{d}_{ij}^t, \mathbf{d}_{ij}^r) = \|\mathbf{d}_{ij}^t - \mathbf{d}_{ij}^r\| \quad \forall i \in \mathcal{V}; j \in \mathcal{N}(i) \qquad (11)$$

can be used to penalize grid deformations. In (11), $\mathbf{d}_{ij} = (\mathbf{x}_i - \mathbf{x}_j)$. It can easily be seen that (11) does not penalize translations of the whole graph. The objective is to find the test grid node coordinates

$\{\mathbf{x}_i^t,\ i \in \mathcal{V}\}$ that minimize

$$C(\{\mathbf{x}_i^t\}) = \sum_{i \in \mathcal{V}} \left\{ S_v(\mathbf{J}(\mathbf{x}_i^t), \mathbf{J}(\mathbf{x}_i^r)) + \lambda \sum_{j \in \mathcal{N}(i)} S_e(i,j) \right\}. \tag{12}$$

The reference grid (i.e., the model grid) has been placed over the output of face detection algorithm described in [10]. An 8×8 sparse grid of equally spaced nodes has been employed. The outputs of multiscale dilation-erosion for scales $\sigma = -9, \ldots, 9$ have been concatenated to form the feature vector at each grid node. The cost function (12) is actually a matching error that defines a distance measure between two persons.

In [2] the authors argue that a two stage coarse-to-fine optimization procedure suffices for the minimization of (12). In our experiments, the above mentioned approach is proved inadequate. Accordingly, we propose: (i) to exploit the face detection results that are provided by the hierarchical rule-based system described in [10] for initializing the minimization of the cost function, and (ii) to replace the two stage optimization procedure by a probabilistic hill climbing algorithm (i.e., a simulated annealing algorithm) that is reminiscent of the Algorithm 1.4 [5, p. 12] that does not make distinction between coarse and fine matching. That is, we propose a random translation of the (undeformed) reference grid and subsequent local perturbations of all grid nodes to find an overall grid deformation that minimizes the cost function (12). Figure 1 depicts the grids formed in the procedure of matching a test person with himself and another person for a pair of test persons extracted from the M2VTS database.

3 Performance evaluation of Morphological Dynamic Link Architecture

The MDLA has been tested on the M2VTS database [9]. The database contains 37 persons' video data, which include speech consisting of uttering digits and image sequences or rotated heads. Four recordings (i.e., shots) of the 37 persons have been collected. Let $BP, BS, CC, \ldots, XM$ be the identity codes of the persons included in the database. In our experiments, the sequences of rotated heads have been considered by using only the luminance information at a resolution of 286×350 pixels. From each image sequence, one frontal image has been chosen based on symmetry considerations. Four experimental sessions have been implemented by employing the "leave one out" principle. Each experimental session consists of a training and a test procedure that are applied to their training set and test set, respectively.

First let us describe the training procedure. The training set is built of 3 (4 are available) shots of 36 (37 are available) persons. By using these images (i.e., the samples for each trained class) one may compute: (i) 6 distance measures for all pairwise combinations between the different samples in the same class, and, (ii) another 6 distance measures for each pairwise combination between the samples of any two different classes. In all pairwise combinations samples that

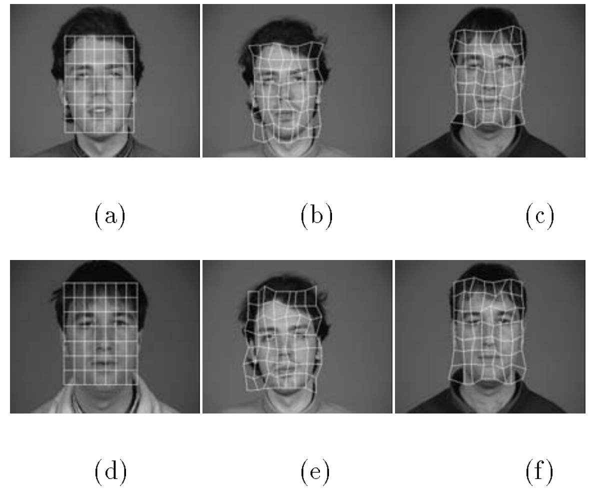

Figure 1: Grid matching procedure in MDLA: (a) Model grid for person BP. (b) Best grid for test person BP after elastic graph matching with the model grid. (c) Best grid for test person BS after elastic graph matching with the model grid for person BP. (d) Model grid for person BS. (e) Best grid for test person BP after elastic graph matching with the model grid for BS. (f) Best grid for test person BS after elastic graph matching with the model grid.

originate from different shots are taken into consideration. In other words, 6 intra-class distance measures and 210 inter-class distance measures are computed for each of the 36 trained classes. Morphological Dynamic Link Architecture has been used to yield all the distance measures required.

Having computed all the 216 distance measures for each trained class, the objective in the training procedure is to determine a threshold on the distance measures that should ideally enable the distinction between the test samples that belong to the trained class under study, and the test samples that belong to any other class. For example, by leaving out shot 01 and person BP, the following 35 thresholds are determined: $T_{BS}(01, BP)$, $T_{CC}(01, BP)$, ..., $T_{XM}(01, BP)$. The threshold $T_{BS}(01, BP)$ is used to discriminate samples of person BS that originate from shots 02, 03, and 04 against all the samples of the remaining 35 classes which originate from any of the above-mentioned shots, when the samples of person BP from these shots are not considered at all. The thresholds have been computed as follows. The minimum intra-class distance and the minimum inter-class distance (i.e., impostor distance) have been found. The vector of 36 minimum distances is ordered in ascending order according to their magnitude. Let $D_{(j)}$ denote the minimum impostor distance for BS when shot 01 is left out and person BP is excluded. The threshold is chosen as follows:

$$T_{BS}(01, BP) = D_{(j+Q)}, \quad Q = 0, 1, 2, \ldots. \tag{13}$$

In the test procedure, three shots create the training set while the fourth one has been used as a test set. Each person of the test set has been considered in turn as an impostor while the 36 others have been used as clients. Each client tries to access under its own identity while the impostor tries to access under the identity of each of the 36 clients in turn. This is tantamount to 36 authentic tests and 36 imposture tests.

In each authentic or imposture test, the reference grids derived for each class during the training procedure are matched and adapted to the feature vectors computed at every pixel of the image of a test person that can be either a client or an impostor using MDLA. Then, the distance measure resulted is compared against the threshold having been computed during the training. Again, we have used the minimum intra-class/inter-class distance in the comparisons, i.e., $D(BP_{01}, \{BS\}) = \min_{j=2}^{4}\{D(BP_{01}, BS_{0j})\}$ where the first ordinate in distance computations denotes an image of the test person and the second ordinate denotes a reference grid for a trained class.

For a particular choice of parameter Q, a collection of thresholds is determined that defines an *operating state* of the test procedure. For such an operating state, a false acceptance rate (FAR) and a false rejection rate (FRR) can be computed. By varying the parameter Q several operating states result. Accordingly, we may create plots of FRR versus FAR with a varying operating state as an implicit set of parameters or equivalently by using the scalar Q as a varying parameter. These plots are the *Receiver Operating Characteristics* (ROCs) of the verification technique. The ROC curve for the entire experiment is shown in Figure 2. The ROC of DLA (without weighting

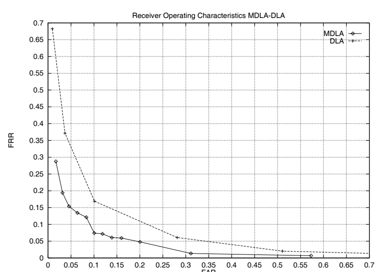

Figure 2: Morphological Dynamic Link Architecture Receiver Operating Characteristic.

the contribution of each node) from [4] is included in the same figure for comparison purposes. It is seen that MDLA clearly outperforms DLA for any FAR. For FAR $\approx$ 10%, the gain is found to be $\approx$ 6.5%. The Equal Error Rate (EER) of MDLA (i.e., the operating state of the method when FAR equals FRR) is another common figure of merit used in the comparison of verification techniques. The EER of MDLA is found to be 9.35 %.

4 Conclusions

A novel multiscale morphological dynamic link architecture has been proposed and tested. The experimental results that have been collected indicate that the proposed method outperforms the (standard) dynamic link matching that is based on Gabor wavelets. Moreover, fast algorithms for the implementation of multiscale dilation-erosion have been developed. It has been found that the choice of the structuring function does not affect the verification efficiency of the method. A further performance improvement is expected by using linear discriminant analysis. An argument that supports such an expectation is that by incorporating local discriminants in the standard DLA an EER of $\approx$ 7.4% has been reported in [4]. The application of linear discriminant analysis in the MDLA is the subject of our future research.

Acknowledgments

The research reported in this paper has been carried out within the framework of the European ACTS-M2VTS project.

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 Würtz, and W. Konen, "Distortion invarinat object recognition in the Dynamic Link Architecture," *IEEE Trans. on Computers*, vol. 42, no. 3, pp. 300–311, March 1993.

[3] S. Fischer, B. Duc, and J. Bigün, "Face recognition with Gabor Phase and Dynamic Link Matching for Multi-Modal Identification," Technical Report LTS 96.04, Signal Processing Laboratory, Swiss Federal Institute of Technology, 1996.

[4] B. Duc, S. Fischer, and J. Bigün, "Face authentication with sparse grid Gabor information," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP-97)*, vol. IV, pp. 3053–3056, Munich, Germany, April 21-24, 1997.

[5] R.H.J.M. Otten, and L.P.P.P. van Ginneken, *The Annealing Algorithm.* Norwell, MA: Kluwer Academic Publ., 1989.

[6] P.T. Jackway, and M. Deriche, "Scale-space properties of the multiscale morphological dilation-erosion," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 1, pp. 38–51, January 1996.

[7] P.T. Jackway, "Morphological scale-space with application to three-dimensional object recognition," Ph.D. Dissertation, Queensland University of Technology, Australia, 1994.

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

[9] 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.

[10] C. Kotropoulos, and I. Pitas, "Rule-based face detection in frontal views," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 97)*, vol. IV, pp. 2537–2540, Munich, Germany, April 21-24, 1997.