Face Verification based on Morphological Shape Decomposition

A. Tefas C. Kotropoulos I. Pitas

Department of Informatics, Aristotle University of Thessaloniki
Box 451, Thessaloniki 540 06, GREECE
{tefas,costas,pitas}@zeus.csd.auth.gr

Abstract

Morphological shape decomposition is used to model a facial image region as a sum of components and to extract a feature vector at the nodes of a sparse grid overlaid over the facial area in dynamic link matching. The feature vector is comprised of the greylevel values at this node in the reconstructed images at several decomposition levels. This feature vector is subsequently employed in Dynamic Link Architecture to verify the identity of each person from a training set. The experimental results indicate that the proposed combination of morphological shape decomposition and dynamic link matching practically offers the same verification capability to the standard dynamic link matching with Gabor wavelets.

1. Introduction

It is well known that mathematical morphology is very rich in providing means for the representation and analysis of binary and grayscale images [5, 12]. The morphological representation of images is well suited for the description of the geometrical properties of image objects. The morphological skeleton and the morphological shape decomposition are two popular approaches for morphological shape representation. Morphological Shape Decomposition (MSD) is the decomposition of an image object (in our case of the facial region) into a union of simple components by using morphological operations, i.e., the erosion and the dilation. It has successfully been applied to the decomposition of a binary shape into a union of simple binary shapes, that is, the maximal inscribable disks [11]. A flexible search-based shape representation scheme that typically gives more efficient representations than the morphological skeleton and MSD is developed in [13].

The main aim of MSD is to extract an appropriate feature vector that is used in a pattern matching algorithm, namely the *Dynamic Link Architecture* (DLA) for face verification [9]. A potential application of the proposed method is in face modeling and subsequently in model-based retrieval of a frontal facial image that corresponds to a specific person from a video sequence that contains frontal facial images of several persons. Another possible application of the proposed method is as a recognition technique in teleshopping applications. In the following, the state-of-the-art in face recognition techniques is briefly outlined.

Two main categories for face recognition techniques can be identified in the literature: those employing geometrical features (for example [1]) and those using grey-level information (e.g. the eigenface approach [14]). A different approach that uses both grey-level information and shape information has been proposed in [9]. More specifically, the response of a set of 2D Gabor filters tuned to different orientations and scales is measured at the nodes of a sparse grid overlaid on the face image of a person from a reference set. The responses of Gabor filters form a *feature vector* at each node of the grid. In the recall phase, the grid of each person in the reference set is overlaid on the face image of a test person and is deformed so that a criterion based both on the feature vectors and the grid distortion (i.e., the geometry) is minimized. An implementation of DLA based on Gabor wavelets is described in [4].

A novel dynamic link architecture that combines morphological shape decomposition and elastic graph matching is developed and tested for face verification. That is, we propose the substitution of the responses of a set of Gabor filters by the set of grey level values of the reconstructed images at the several levels of decomposition. There are several reasons supporting this decision, namely: (1) The decomposition of a complex object yields simple components that conform with our intuition. In our case the component is the maximal inscribable cylinder of unit height. In addition, the method is object-independent [12]. (2) It allows arbitrary amounts of detail to be computed and also allows the abstraction from detail [12]. (3) The representation is unique. Moreover, it is information-preserving in contrast to Morphological Dynamic Link Matching (MDLA) pro-

posed in [6]. (4) MSD employs grayscale erosions and dilations with a flat structuring function, namely, a cylinder of unit height having a circular cross-section of radius 2. Grayscale erosions and dilations with a flat structuring function can be computed very fast by using running min/max selection algorithms [12].

The outline of the paper is as follows. Facial region modeling using MSD is outlined in Section 2. The proposed MSD-DLA is described in Section 3. The evaluation of performance of MSD-DLA with respect to its Receiver Operating Characteristic (ROC) is treated in Section 4. Conclusions are drawn and further research directions are indicated in Section 5.

2. Facial region modeling using morphological shape decomposition

The modeling of a grayscale facial image region by employing MSD is described in this section. To begin with let us briefly describe a necessary preprocessing step that aims at detecting facial regions in frontal views. A very attractive approach for face detection is based on multiresolution images (also known as *mosaic images*). It attempts to detect a facial region at a coarse resolution and subsequently to validate the outcome by detecting facial features at the next resolution level [15]. Towards this goal, the method employs a hierarchical knowledge-based pattern recognition system. Recently, a variant of this method has been proposed [7]. It offers the following features: (a) It allows for rectangular cells in contrast to the square cells used in [15]. (b) It is equipped with a preprocessing step that determines an estimate of the cell dimensions and the offsets so that the mosaic model fits the face image of each person. (c) It has very low computational demands compared to the original algorithm [15], because the iterative nature of the algorithm is avoided due to the preprocessing step that has been employed. (d) It employs more general rules that are close to our intuition for a human face. However, the above-described variant treats efficiently scenes where a single person appears and the background is fairly uniform. By using this method, we may define roughly a region where the face is included, and control the placement of a sparse grid over the face in order to store a model for each person in dynamic link matching, as is described later on.

MSD is applied to the output of the face detection algorithm. Let us define by $f(\mathbf{x}) : \mathcal{D} \subseteq \mathbb{Z}^2 \rightarrow \mathbb{Z}$ the image at the output of the preprocessing step employed with $\mathbb{Z}$ denoting the set of integer numbers and $\mathcal{D}$ being the domain of $f(\mathbf{x})$. Without any loss of generality it is assumed that the image pixel values are non-negative, i.e., $f(\mathbf{x}) \geq 0$. Let $g(\mathbf{x}) = 1, \forall \mathbf{x} : \|\mathbf{x}\| \leq R$ denote the *structuring function*. The value $R = 2$ has been used in all experiments. It is seen that by definition, $g(\mathbf{x})$ is symmetric. Accordingly, symmetric operators will not explicitly denoted hereafter. Furthermore, it can easily be seen that our structuring function is a cylinder of unit height with a circular cross-section of radius 2. Given $f(\mathbf{x})$ and $g(\mathbf{x})$, the *grayscale dilation* of the image $f(\mathbf{x})$ by the structuring function $g(\mathbf{x})$ is defined as [5, 12]:

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

The complementary operation, the *grayscale erosion*, is defined as:

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

The objective of shape decomposition is to decompose $f(\mathbf{x})$ into a sum of components, i.e.:

$$f(\mathbf{x}) = \sum_{i=1}^{K} f_i(\mathbf{x}) \quad (3)$$

where $f_i(\mathbf{x})$ denotes the i-th component that should be a simple function. That is, it can be expressed as follows:

$$f_i(\mathbf{x}) = [l_i \oplus n_i\, g]\,(\mathbf{x}) \quad (4)$$

where $l_i(\mathbf{x})$ is the so called *spine* [12] and

$$n_i\, g(\mathbf{x}) = \underbrace{[g \oplus g \oplus \cdots \oplus g]}_{n_i \text{ times}}(\mathbf{x}). \quad (5)$$

An intuitively sound choice for $n_1 g(\mathbf{x})$ is the maximal function in $f(\mathbf{x})$, that is, to choose n_1 such that

$$[f \ominus (n_1+1)g]\,(\mathbf{x}) \leq 0 \quad \forall x \in \mathcal{D}. \quad (6)$$

Accordingly, the first spine is given by:

$$l_1(\mathbf{x}) = [f \ominus n_1\, g]\,(\mathbf{x}). \quad (7)$$

Morphological shape decomposition can then be implemented recursively as follows.

Step 1. Initialization: $\hat{f}_0(\mathbf{x}) = 0$.

Step 2. i-th level of decomposition: Starting with $n_i = 1$ increment n_i until

$$\left[(f - \hat{f}_{i-1}) \ominus (n_i+1)g\right](\mathbf{x}) \leq 0. \quad (8)$$

Step 3. Calculate the i-th component by

$$f_i(\mathbf{x}) = \left\{ \underbrace{\left[(f - \hat{f}_{i-1}) \ominus n_i\, g\right]}_{l_i(\mathbf{x})} \oplus n_i\, g \right\}(\mathbf{x}) \quad (9)$$

Step 4. Calculate the reconstructed image at the i-th level of decomposition:

$$\hat{f}_i(\mathbf{x}) = \hat{f}_{i-1}(\mathbf{x}) + f_i(\mathbf{x}). \qquad (10)$$

Step 5. Let $\mathcal{M}(f-\hat{f}_i)$ be a measure of the approximation of the image $f(\mathbf{x})$ by its reconstruction $\hat{f}_i(\mathbf{x})$ at the i-th level of decomposition. Increment i and go to Step 2 until $i > K$ or $\mathcal{M}(f - \hat{f}_{i-1})$ is sufficiently small.

Figure 1 shows the block diagram of the MSD. The module Component Extraction (CE) implements the Steps 2 and 3 of the algorithm outlined above.

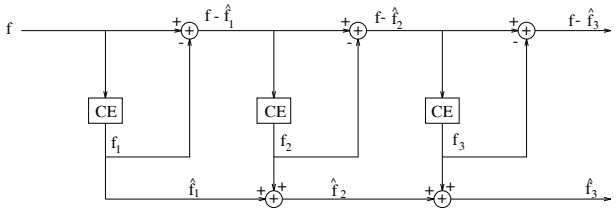

Figure 1. Block diagram of MSD.

3. Combined use of Morphological Shape Decomposition and Dynamic Link Architecture

Traditionally, linear methods like the Fourier transform, the Walsh-Hadamard transform, Gaussian filter banks, wavelets, Gabor elementary functions have dominated thinking on algorithms for generating the information pyramid. An alternative to linear techniques is to use morphological shape decomposition techniques. In this paper, we propose the substitution of Gabor-based feature vectors used in dynamic link matching by feature vectors that are extracted from the reconstructed images $\hat{f}_i(\mathbf{x})$ at the last K successive levels of decomposition $i = L-K, \ldots, L$ where L denotes the maximal number of decomposition levels. We have found that the value $K = 15$ gives good results in practice. That is, the grey level information $\hat{f}_i$ at the node $\mathbf{x}$ of the sparse grid for the levels of decomposition $i = L-15, \ldots, L$ along with the grey level information f is concatenated to form the feature vector $\mathbf{J}(\mathbf{x})$, the so called *jet* [9]:

$$\mathbf{J}(\mathbf{x}) = \left(f(\mathbf{x}), \hat{f}_{L-K}(\mathbf{x}), \ldots, \hat{f}_L(\mathbf{x})\right) \qquad (11)$$

The resulted variant of DLA is the so called Morphological Shape Decomposition-Dynamic Link Architecture (MSD-DLA). Alternatively, one may also use the feature vector:

$$\mathbf{J}'(\mathbf{x}) = \left(f(\mathbf{x}) - \hat{f}_L(\mathbf{x}), \hat{f}_{L-K}(\mathbf{x}), \ldots, \hat{f}_L(\mathbf{x})\right) \qquad (12)$$

Figure 2 depicts a series of reconstructed images at nineteen levels of decompositions for the facial image region of a sample person from the database. The 20th image at the bottom right is the original facial image region that is decomposed. Only the last fifteen reconstructed images have been employed in the MSD-DLA.

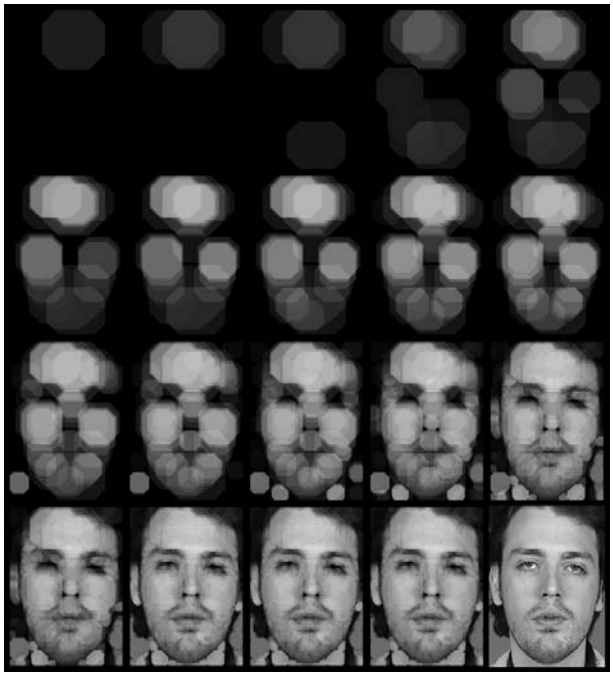

Figure 2. Reconstructed images at the nineteen levels of the decomposition. The image at the bottom right is the original one.

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

$$S_v(\mathbf{J}(\mathbf{x}_i^t), \mathbf{J}(\mathbf{x}_i^r)) = \|\mathbf{J}(\mathbf{x}_i^t) - \mathbf{J}(\mathbf{x}_i^r)\|. \qquad (13)$$

As in DLA [9], 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. Then, an additional cost function is used:

$$S_e(i,j) = S_e(\mathbf{d}_{ij}^t, \mathbf{d}_{ij}^r) = \|\mathbf{d}_{ij}^t - \mathbf{d}_{ij}^r\| \quad \forall i \in \mathcal{V}; j \in \mathcal{N}(i) \qquad (14)$$

where $\mathcal{N}(i)$ denotes the neighborhood of a vertex i (e.g. a four-connected neighborhood in our case) and $\mathbf{d}_{ij} = \mathbf{x}_i - \mathbf{x}_j$. It can easily be seen that (14) 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

$$C(\{\mathbf{x}_i^t\}) \quad = \quad \sum_{i \in \mathcal{V}} \{S_v(\mathbf{J}(\mathbf{x}_i^t), \mathbf{J}(\mathbf{x}_i^r)) +$$

$$+ \quad \lambda \sum_{j \in \mathcal{N}(i)} S_e(\mathbf{d}_{ij}^t, \mathbf{d}_{ij}^r)\}. \qquad (15)$$

One may interpret (15) as a simulated annealing with an additional penalty (i.e., a constraint on the objective function). Since the cost function (14) does not penalize translations of the whole graph. The random configuration $\mathbf{x}_i^t$ can be of the form of a random translation $\mathbf{d}$ of the (undeformed) reference grid node and a bounded local perturbation $\underline{\delta}_i$, i.e.:

$$\mathbf{x}_i^t = \mathbf{x}_i^r + \mathbf{d} + \underline{\delta}_i \quad ; \quad \|\underline{\delta}_i\| \leq \delta_{\max} \qquad (16)$$

where the choice of $\delta_{\max}$ controls the rigidity/plasticity of the graph. Figure 3 depicts the grids formed in the procedure of matching.

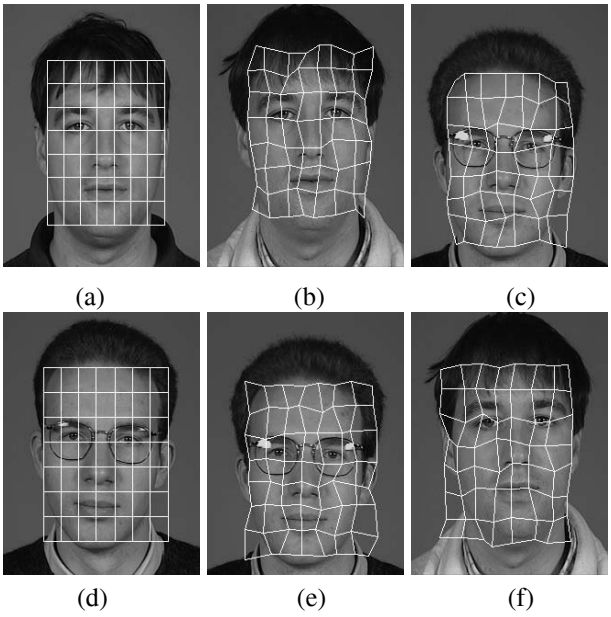

Figure 3. The graph matching procedure in MSD-DLA: model grid, best grid for the test person after translation and deformation of the grid. Figures (a),(d): Reference person. Figures (b),(e): The test person is identical to the reference one (Distance b-a=2776, Distance e-d=1751). Figures (c),(f): The test person is different from the reference one (Distance c-a=6273, Distance f-d=5003).

4. Performance evaluation of MSD-DLA

The MSD-DLA has been tested on the M2VTS database [10]. 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, ..., 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 $\times$ 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. This amounts to 3 $\times$ 36 = 108 images. 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 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 Shape Decomposition - 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:

$$T_{BS}(01, BP) = D_{(j+Q)}, \quad Q = 0, 1, 2, \ldots. \qquad (17)$$

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. By repeating the procedure four times, $4 \times 37 \times 36 = 5328$ authentic and imposture tests have been realized in total.

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 MSD-DLA. 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.,

$$\begin{aligned} D(BP_{01}, \{BS\}) \quad = \quad & \min\{D(BP_{01}, BS_{02}), \\ & D(BP_{01}, BS_{03}), \\ & D(BP_{01}, BS_{04})\} \end{aligned} \tag{18}$$

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 for each training set is plotted separately in Figure 4a. The corresponding curve for the entire experiment is shown in Figure 4b. The Equal Error Rate (EER) of MSD-DLA (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 MSD-DLA is found to be 11.89 %. Table 1 summarizes the FRR achieved for FAR $\approx 10\%$ for each shot left out. It is seen that due to the variations in the appearance of the persons included in the database and the recording conditions (e.g. illumination changes) that occur in the four shots the performance of the method is not constant. However, Table 1 suggests that a compensation of illumination conditions as well as the use of linear discriminant analysis may improve further the verification efficiency of the method. Another 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 [2].

Table 1. False rejection rates achieved for a false acceptance rate $\approx$ 10 % when each shot in turn is left out.

Shot left out	FAR (%)	FRR (%)
1	11.33	7.50
2	10.73	18.92
3	10.06	8.11
4	10.58	18.09

Table 2. Comparison of the Verification techniques Equal Error Rates.

Verification Technique	EER (%)
MSD-DLA	11.89
GDLA	10.8-14.4

Table 2 compares the EER achieved by MSD-DLA to the same figure of Gabor-based DLA [3]. It can be seen that the proposed combination of morphological shape decomposition and dynamic link matching practically offers the same verification capability to the standard dynamic link matching with Gabor wavelets.

5. Conclusions

A novel morphological dynamic link architecture that employs morphological shape decomposition as feature extraction mechanism has been developed and has been tested. The experimental results collected are very encouraging. A compensation of illumination conditions as well as the use of linear discriminant analysis may improve further the verification efficiency of the method. Recently, by weighting the signal similarity measure (13) at each grid node with an appropriately derived coefficient that quantifies the discriminatory power of the the grid node we achieved an EER of 6.58% following the same experimental setup [8].

Acknowledgement

This work has been carried out within the framework of the European ACTS-M2VTS project.

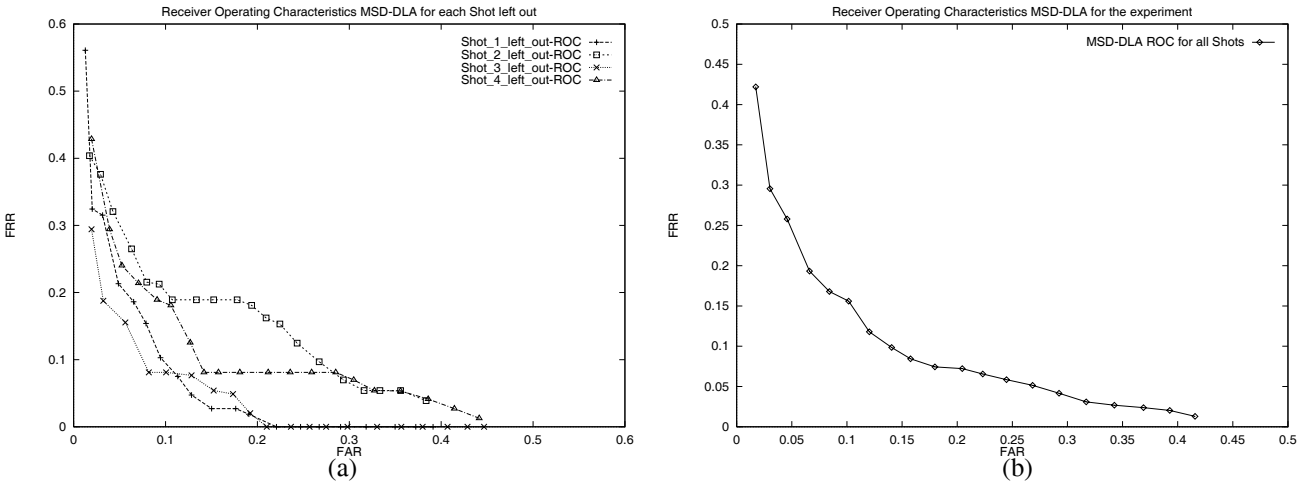

Figure 4. Morphological Shape Decomposition - Dynamic Link Architecture Receiver Operating Characteristics for (a) each training set separately, and (b) the entire experiment.

References

[1] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, 1993.

[2] B. Duc, S. Fischer, and J. Bigün. Face authentication with sparse grid gabor information. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP-97)*, volume IV, pages 3053–3056, Munich, Germany, April 1997.

[3] B. Duc, S. Fischer, and J. Bigün. Face authentication with gabor information on deformable graphs. *IEEE Transactions on Image Processing*, submitted 1997.

[4] 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.

[5] R. Haralick. Image analysis using mathematical morphology. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 9(4):532–550, July 1987.

[6] C. Kotropoulos and I. Pitas. Face authentication based on morphological grid matching. In *Proc. of the IEEE Int. Conf. on Image Processing (ICIP-97)*, volume I, pages 105–108, California, October 1997.

[7] C. Kotropoulos and I. Pitas. Rule-based face detection in frontal views. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP-97)*, volume IV, pages 2537–2540, Munich, Germany, April 1997.

[8] C. Kotropoulos, A. Tefas, and I. Pitas. Frontal face authentication using variants of dynamic link matching based on mathematical morphology. In *1998 IEEE Int. Conf. on Image Processing*, Chicago, October 1998, submitted.

[9] M. Lades, J. Vorbrüggen, J. Buhmann, J. Lange, C. v.d. Malsburg, R. Würtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. on Computers*, 42(3):300–311, March 1993.

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

[11] I. Pitas and A. Venetsanopoulos. Morphological shape decomposition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(1):38–45, January 1990.

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

[13] J. Reinhardt and W. Higgins. Efficient morphological shape representation. *IEEE Trans. on Image Processing*, 5(1):89–101, January 1996.

[14] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[15] G. Yang and T. Huang. Human face detection in a complex background. *Pattern Recognition*, 27(1):53–63, 1994.