FACE AUTHENTICATION BASED ON MORPHOLOGICAL SHAPE DECOMPOSITION

C. Kotropoulos A. Tefas I. Pitas
Department of Informatics, Aristotle University of Thessaloniki
Box 451, Thessaloniki 540 06, GREECE
{costas,tefas,pitas}@zeus.csd.auth.gr

1 INTRODUCTION

Many theories for shape decomposition and recognition have been developed in the past two decades. Shape analysis is an important step towards shape description that aims at describing a shape (or object) before shape matching in which it is desirable to establish the equivalence of two shapes [1]. The algorithms for shape description are generally classified in two classes, namely the *external* algorithms and the *internal* ones. For example, contour description algorithms belong to the former class whereas region-based algorithms belong to the later. Another taxonomy of shape description algorithms is to *information-preserving* and *non information-preserving* algorithms depending on whether or not the original shape can be reconstructed from the descriptor. In this paper, we deal with an internal and information-preserving shape description algorithm, namely the *morphological shape decomposition* (MSD).

It is well known that mathematical morphology is very reach in providing means for the representation and analysis of binary and grayscale images [2, 3]. 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 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 [4]. A flexible search-based shape representation scheme that typically gives more efficient representations than the morphological skeleton and MSD is developed in [5].

In this paper, we propose the use of morphological shape decomposition to extract an appropriate feature vector that is used in a pattern matching algorithm, namely the *Dynamic Link Architecture* (DLA) [6] for face recognition. 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. Face modeling may be considered as a special case of 3D object modeling, where two coordinates are devoted to the description of the spatial information and a third coordinate is related to grayscale information. 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 [7]) and those using grey-level information (e.g. the eigenface approach [8]). A different approach that uses both grey-level information and shape information has been proposed in [6]. 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 [9].

A novel dynamic link architecture that combines morphological shape decomposition and elastic graph matching is developed and tested for face authentication. That is, we propose the substitution of the responses of a set of Gabor filters by the grey level value 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 [3]. (2) It allows arbitrary amounts of detail to be com-

puted and also allows the abstraction from detail [3]. (3) The representation is unique. Moreover, it is information-preserving in contrast to Morphological Dynamic Link Matching (MDLA) proposed in [10]. (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 [3].

The first experimental results reported in this paper indicate the superiority of the proposed novel variant of DLA, to be called *Morphological Shape Decomposition-Dynamic Link Architecture (MSD-DLA)*, over the (standard) dynamic link matching with Gabor-based feature vectors.

The outline of this 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 morphological shape decomposition 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*) attempting to detect a facial region at a coarse resolution and subsequently to validate the outcome by detecting facial features at the next resolution level [11]. Towards this goal, the method employs a hierarchical knowledge-based pattern recognition system. Recently, a variant of this method has been proposed [12]. The above-mentioned variant treats efficiently scenes where a single person appears and the background is fairly uniform, for the following reasons: (a) It allows for rectangular cells in contrast to the square cells used in [11]. (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 [11], 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. 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.

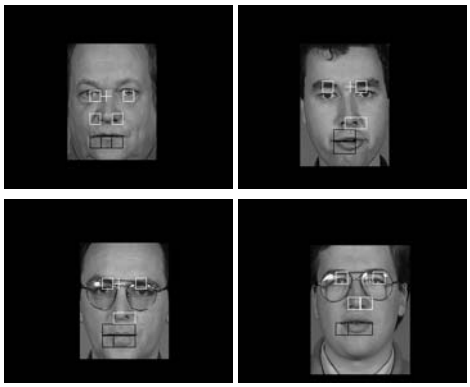


Figure 1: Face detection results.

Figure 1 shows the results of the face detection algorithm on frontal views of eight persons extracted from the European ACTS project M2VTS database [13]. The white overlaid rectangles correspond to eyebrows/eyes and nostrils/nose candidates. The black overlaid rectangles correspond to mouth candidates and the white cross indicates the characteristic bright point between the eyes.

Morphological shape decomposition is applied to the output of the face detection algorithm shown in Figure 1. 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. Therefore, 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 [2, 3]:

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

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:

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

where $l_i(\mathbf{x})$ is the so called *spine* [3] and

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

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

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

Accordingly, the first spine is given by:

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

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

$$\Big[(f - \hat{f}_{i-1}) \ominus (n_i + 1)g\Big]\,(\mathbf{x}) < 0. \tag{8}$$

Step 3. Calculate the i-th component by

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

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

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

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 2 shows the block diagram of the morphological shape decomposition. The module Component Extraction (CE) implements the Steps 2 and 3 of the algorithm outlined above. It can easily be seen that morphological shape decomposition is an information-preserving shape description method.

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

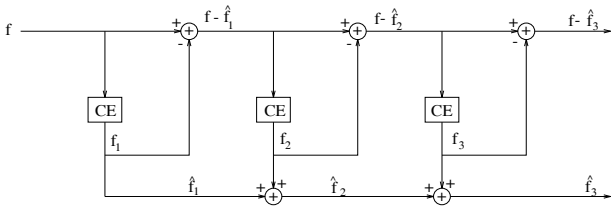


Figure 2: Block diagram of morphological shape decomposition.

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 successive levels of decomposition $i = 1, \ldots, K$ for K=16. That is, the grey level information $\hat{f}_i$ at the node $\mathbf{x}$ of a sparse grid for the levels of decomposition $i = 1, \ldots, 16$ along with the grey level information f is concatenated to form a feature vector $\mathbf{J}(\mathbf{x})$, the so called *jet*, in Dynamic Link Architecture (DLA) [6]: yielding the Morphological Shape Decomposition-Dynamic Link Architecture (MSD-DLA):

$$\mathbf{J}(\mathbf{x}) = \Big(f(\mathbf{x}), \hat{f}_1(\mathbf{x}), \ldots, \hat{f}_K(\mathbf{x})\Big) \tag{11}$$

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:

$$\mathbf{J}'(\mathbf{x}) = \Big(f(\mathbf{x}) - \hat{f}_K(\mathbf{x}), \hat{f}_1(\mathbf{x}), \ldots, \hat{f}_K(\mathbf{x})\Big) \tag{12}$$

Figure 3 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 first sixteen reconstructed images have been employed in the dynamic link architecture.

It is seen that the approximation error is negligible everywhere except the salient minima of the facial region that correspond to eyes, nostrils etc. This observation suggests that a further improvement can be obtained by applying the same approach to the negative of the original image so that image minima are approximated more accurately.

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.:

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

As in DLA [6], the quality of a match is evaluated by taking into account the grid deformation as well.

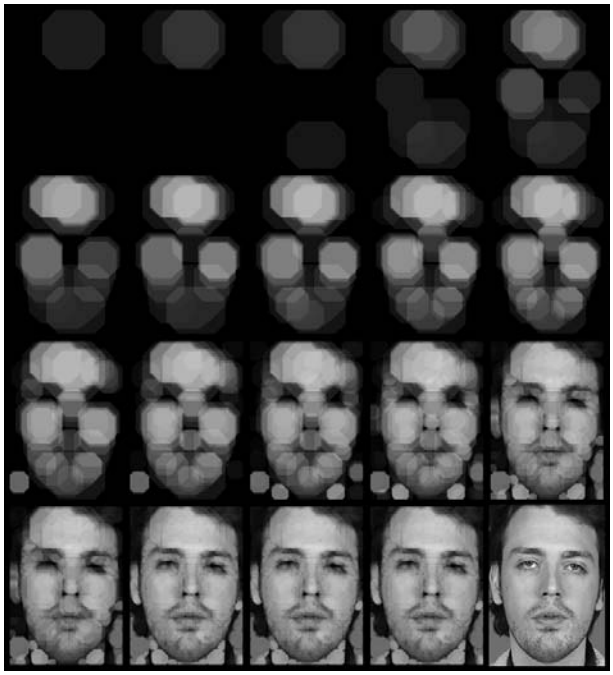


Figure 3: Reconstructed images at the nineteen levels of the decomposition. The image at the bottom right is the original one.

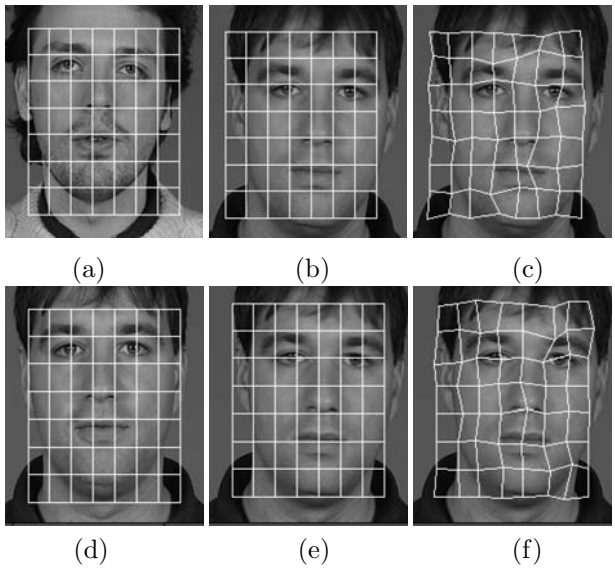


Figure 4: Three stages of the graph matching procedure: model grid, best grid for the test person after coarse matching, best grid for the test person after fine matching. Cases (a)–(c): The test person is different from the reference one. Cases (d)–(f): The test person is identical to the reference one.

Let us denote by $\mathcal{V}$ the set of grid nodes. Then, an additional cost function is used:

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

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

$$\begin{aligned} C(\{\mathbf{x}_i^t\}) &= \sum_{i \in \mathcal{V}} \{S_v(\mathbf{J}(\mathbf{x}_i^t), \mathbf{J}(\mathbf{x}_i^r)) + \\ &+ \lambda \sum_{j \in \mathcal{N}(i)} S_e(\mathbf{d}_{ij}^t, \mathbf{d}_{ij}^r)\}. \end{aligned} \tag{15}$$

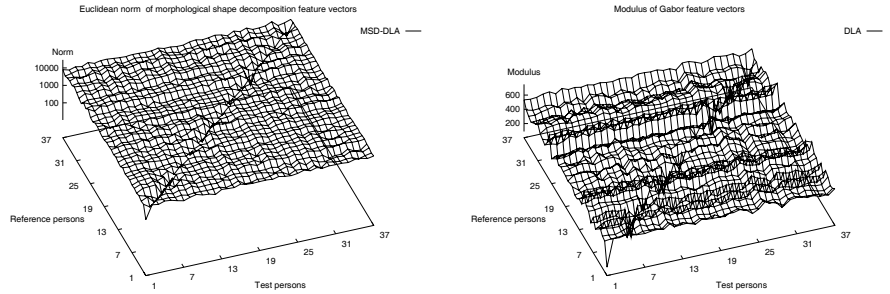
In the minimization of (15), the coarse-fine approach proposed in [6] has been used. The reference grid (i.e., the model grid) has been placed over the output of face detection algorithm described in Section 2. A sparse grid of 8×8 equally spaced nodes has been employed. Figure 4 depicts the grids formed in the procedure of matching the same person with two reference persons.

4 PERFORMANCE EVALUATION OF MSD-DLA

The MSD-DLA has been tested on the M2VTS database of 37 persons [13]. A reference set has been created by choosing a frontal view of each person from the third shot. A test set has been created by choosing another frontal or near frontal view of each person from this shot. In this paper, we have used solely the sum of L_2 norms of feature vector differences at all grid nodes as a distance measure between a test and a reference person. The distance for all pairs of test and reference persons when MSD-DLA is used is plotted in Figure 5a. For comparison purposes, the same experiment has been repeated with DLA. In the latter case, the sum of moduli of Gabor feature vectors at the grid nodes has been used to yield a distance between a reference and a test person. The plot of distances is shown in Figure 5b. It is clearly seen that MSD-DLA has much more minima in the main diagonal than DLA. The DLA has succeeded to identify 17 persons from 37 whereas the MSD-DLA has succeeded to identify 31 persons from 37.

Both DLA and MSD-DLA always yield a minimum value of the distance measure that has been used irrespective of whether or not a corresponding image of the same person is contained in the reference set. Therefore, we need to introduce a threshold T that will permit either to accept or to reject the outcome of the minimal distance rule. Let t_i and r_j denote the i-th test and j-th reference person respectively. Let $\Delta(t_i, r_j)$ denote their distance. Let also N be the number of persons in the test (i.e., N=37). We define the *false rejection rate* at threshold T, $FRR(T)$, as follows:

$$FRR(T) = 1 - \frac{1}{N} \text{ card } \{\forall i : i = \arg\min\{\Delta(t_i, r_j),$$



(a) (b)

Figure 5: Plots of elastic graph distances for all pairs of test and reference persons in MSD-DLA and DLA. (a) Euclidean norm of MSD feature vectors. (b) Modulus of Gabor feature vectors.

$$j = 1, \ldots, N\} \quad \text{and} \quad \Delta(t_i, r_i) \leq T\} \quad (16)$$

where card{ } denotes the cardinality of a set. Accordingly, the *false acceptance rate* at threshold T, $FAR(T)$, is given by:

$$FAR(T) = \frac{1}{N} \operatorname{card}\{\forall i : k = \arg\min\{\Delta(t_i, r_j);\ j = 1, \ldots, N;\ j \neq i\} \quad \text{and} \quad \Delta(t_i, r_k) \leq T\}. \quad (17)$$

Let us also define the *receiver operating characteristic* (ROC) of an identification technique as the plot of FRR versus FAR with the threshold T being a varying parameter. Ideally, we want to find a threshold T_o such that $FRR(T_o) \to 0$ and $FAR(T_o) \to 0$. What usually happens is a trade-off between FRR and FAR. Therefore, between identification techniques the smaller the area under the ROC the better the technique is. The ROC of the MSD-DLA and the DLA for the distance measures used is plotted in Figure 6a. It is seen that MSD-DLA outperforms DLA. For the same FAR=20%, the FRR for MSD-DLA is 30% whereas the FRR for DLA is 56%. The plots of FAR/FRR versus the threshold T are given in Figures 6b and 6c respectively.

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 first experimental results collected are very encouraging and indicate that the proposed method outperforms the (standard) dynamic link matching that is based on Gabor wavelets. However, further experiments should be conducted to validate the first results reported in this paper. Furthermore, the derivation of a theoretically sound distance measure between a pair of a test and reference person that combines signal as well as geometrical distortions is another subject of future research.

References

[1] Levine, M.D., *Vision in Man and Machine*. New York:McGraw-Hill, 1985.

[2] Haralick, R.M., Sternberg, S.R., and Zhuang, X., "Image analysis using mathematical morphology," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 4, pp. 532–550, July 1987.

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

[4] Pitas, I., and Venetsanopoulos, A.N., "Morphological shape decomposition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 38–45, January 1990.

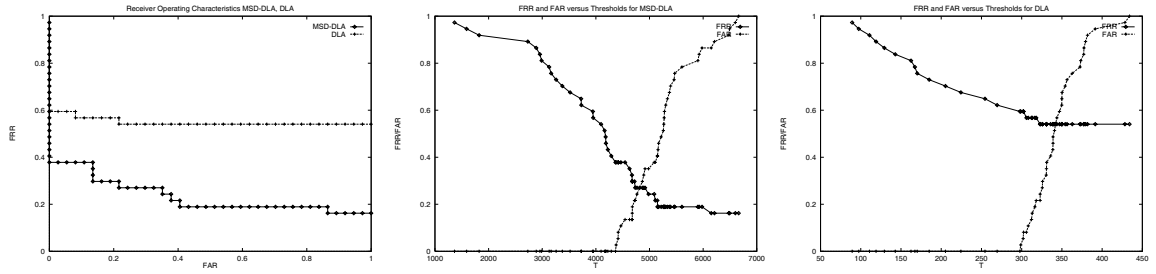
[5] Reinhardt, J.M., and Higgins, W.E., "Efficient morphological shape representation," *IEEE Trans. on Image Processing*, vol. 5, no. 1, pp. 89–101, January 1996.

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

[7] Brunelli, R., and Poggio, T., "Face recognition: Features versus templates," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042–1052, 1993.

[8] Turk, M., and Pentland, A., "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[9] Fischer, S., Duc, B. and Bigün, J., "Face recognition with Gabor Phase and Dynamic Link



(a) (b) (c)

Figure 6: (a) Receiver Operating Characteristics for the MSD-DLA and the DLA. (b) Plot of False Rejection/Acceptance Rate versus a threshold in MSD-DLA. (c) Plot of False Rejection/Acceptance Rate versus a threshold in DLA.

Matching for Multi-Modal Identification," Technical Report LTS 96.04, Signal Processing Laboratory, Swiss Federal Institute of Technology, 1996.

[10] Kotropoulos, C., Pitas, I., Fischer, S., and Duc, B., "Face authentication using morphological Dynamic Link Architecture," in *Lecture Notes in Computer Science: Audio- and Video- based Biometric Person Authentication (J. Bigün, G. Chollet, and G. Borgefors, Eds.)*, vol. 1206, pp. 169–176, Springer-Verlag, 1997.

[11] Yang, G., and Huang, T.S., "Human face detection in a complex background," *Pattern Recognition*, vol. 27, no. 1, pp. 53–63, 1994.

[12] Kotropoulos, C., and Pitas, I., "Rule-based face detection in frontal views," in *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP-97)*, Munich, Germany, April 20-24, 1997, to appear.

[13] Pigeon, S., and Vandendorpe, L., "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, Springer-Verlag, 1997.