FACE VERIFICATION BASED ON MORPHOLOGICAL DYNAMIC LINK ARCHITECTURE

Constantine Kotropoulos and Ioannis Pitas

Department of Informatics, Aristotle University of Thessaloniki
Box 451, Thessaloniki 54006, 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. More specifically, a sparse grid is placed over the facial region of the image of each person in the reference set. Multiscale morphological operations are employed then to yield a feature vector at each grid node. Subsequently, dynamic link matching is applied to establish an isomorphism between the reference grid of the claimed person and the variable graph built over the face image of the candidate. 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 are found in [1]. Two main categories for face recognition techniques are identified: those employing geometrical features (for example [2]) and those using grey-level information (e.g. the eigenface approach [3]). 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) [4]. This algorithm is split in two phases, i.e., the training 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 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 an impostor against the authentic person [5]. An automatic weighting of the nodes according to their significance by employing local discriminants is proposed in [6]. 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 [8]. There are several reasons supporting this decision, namely: (1) Scale-space morphological techniques are able to find the true size of the object in an image smoothed to a particular size. (2) The scale parameter has a straightforward interpretation since it is associated with an integer number of pixels. (3) Dilations and erosions can be computed very fast since they employ min/max selection operations [9]. (4) 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).

Another issue studied in detail is the elastic graph matching procedure outlined above. In [4] the authors argue that a two-stage coarse-to-fine optimization procedure suffices for the minimization of the cost function that measures the quality of the elastic graph matching. The experiments conducted indicate that the replacement of the two stage optimization procedure by a probabilistic hill climbing al-

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

gorithm (i.e., a simulated annealing algorithm) that is reminiscent of the Algorithm 1.4 [7, p. 12] yields better results in terms of the verification efficiency.

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 [11]. 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.

Frequently, linear projection algorithms are used to reduce the dimensionality of the input feature vectors. The type of linear projection used in practice is influenced by the availability of category information about the feature vectors in the form of labels on the feature vectors. In this paper we test the Karhunen-Loeve method or principal component analysis (PCA) to reduce the dimensionality of feature vectors at the grid nodes. It is shown that by keeping six only principal components at each grid node we attain an approximation error less than 5 % without any deterioration in the verification efficiency of the method.

The outline of the paper is as follows. The proposed variant of dynamic link matching that is based on multiscale dilation-erosion is described in Section 2. Its performance evaluation is assessed in Section 3. The application of PCA to the feature vectors measured at each grid node is presented in Section 4. Conclusions are drawn and further research directions are indicated in Section 5.

2. DYNAMIC LINK MATCHING WITH MULTISCALE MORPHOLOGICAL DILATION-EROSION

Traditionally, linear methods like the Fourier transform, the Walsh-Hadamard transform, the Gaussian filter banks, the wavelets, the Gabor elementary functions have dominated thinking on algorithms for generating an information pyramid. An alternative to linear techniques 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* [8].

The multiscale morphological dilation-erosion is based on the two fundamental operations of the grayscale morphology, namely the *dilation* and the *erosion*. Let $\mathcal{R}$ and $\mathcal{Z}$ denote the set of real and integer numbers, respectively. Given an image $f(\mathbf{x}) : \mathcal{D} \subseteq \mathcal{Z} \rightarrow \mathcal{R}$ and a structuring function $g(\mathbf{x}) : \mathcal{G} \subseteq \mathcal{Z}^2 \rightarrow \mathcal{R}$, the dilation of the image $f(\mathbf{x})$ by $g(\mathbf{x})$ is defined by [9, 10]:

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

Its complementary operation, the erosion is given by:

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

If the structuring function is chosen to be scale-dependent, that is $g_\sigma(\mathbf{z}) = |\sigma| g(|\sigma|^{-1}\mathbf{z})$, then the morphological operations become scale-dependent as well. Suitable structuring functions are described in [8, 9]. In this paper the *scaled hemisphere* is employed, i.e. [8]:

$$g_\sigma(\mathbf{z}) = |\sigma| \left(\sqrt{1 - \left(|\sigma|^{-1}\|\mathbf{z}\|\right)^2} - 1 \right) \quad \forall \mathbf{z} \in \mathcal{G} : \|\mathbf{z}\| \leq \sigma. \quad (3)$$

Accordingly, the multiscale dilation-erosion of the image $f(\mathbf{x})$ by $g_\sigma(\mathbf{x})$ is defined by [8]:

$$(f \star g_\sigma)(\mathbf{x}) = \begin{cases} (f \oplus g_\sigma)(\mathbf{x}) & \text{if } \sigma > 0 \\ f(\mathbf{x}) & \text{if } \sigma = 0 \\ (f \ominus g_\sigma)(\mathbf{x}) & \text{if } \sigma < 0. \end{cases} \quad (4)$$

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

$$\mathbf{J}(\mathbf{x}) = ((f \star g_9)(\mathbf{x}), \ldots, (f \star g_1)(\mathbf{x}), f(\mathbf{x}), (f \star g_{-1})(\mathbf{x}), \ldots, (f \star g_{-9})(\mathbf{x})). \quad (5)$$

Figure 1 depicts the outputs of multiscale dilation-erosion for the scales that have been used. The first nine pictures starting from the upper left picture are dilated images and the remaining nine are eroded images. It is seen that multiscale dilation-erosion captures important information for the key facial features, e.g. the eyebrows, the eyes, the nose tip, the nostrils, the lips, the face contour etc. 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.:

Figure 1: Responses of dilations and erosions for scales 1–9.

$$S_v(\mathbf{J}(\mathbf{x}_i^t), \mathbf{J}(\mathbf{x}_i^r)) = \|\mathbf{J}(\mathbf{x}_i^t) - \mathbf{J}(\mathbf{x}_i^r)\|. \quad (6)$$

As in DLA [4], 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:

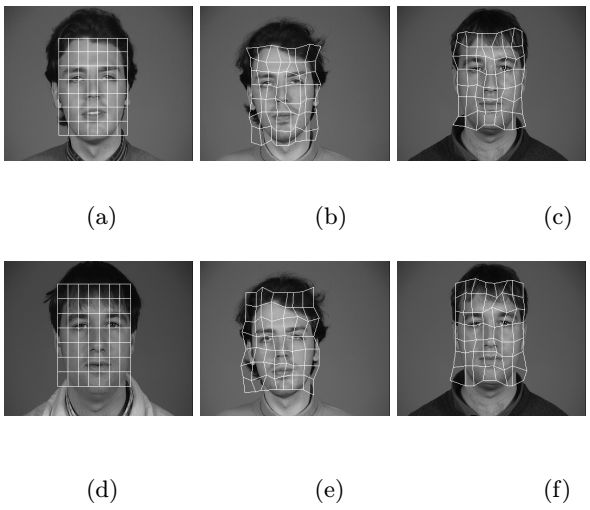
$$S_e(i,j) = S_e(\mathbf{d}_{ij}^t, \mathbf{d}_{ij}^r) = \|\mathbf{d}_{ij}^t - \mathbf{d}_{ij}^r\| \quad \forall i \in \mathcal{V}; j \in \mathcal{N}(i) \quad (7)$$

can be used to penalize grid deformations. In (7), $\mathbf{d}_{ij} = (\mathbf{x}_i - \mathbf{x}_j)$. It can easily be seen that (7) 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

$$C(\{\mathbf{x}_i^t\}) = \sum_{i \in \mathcal{V}} \left\{ S_v(\mathbf{J}(\mathbf{x}_i^t), \mathbf{J}(\mathbf{x}_i^r)) + \lambda \sum_{j \in \mathcal{N}(i)} S_e(\mathbf{d}_{ij}^t, \mathbf{d}_{ij}^r) \right\}. \quad (8)$$

The reference grid (i.e., the model grid) has been placed over the output of face detection algorithm described in [12]. 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 (8) is actually a matching error that defines a distance measure between two persons.

In [4] the authors argue that a two stage coarse-to-fine optimization procedure suffices for the minimization of (8). 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 [12] 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 [7, 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 (8). Figure 2 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.



(a) (b) (c)

(d) (e) (f)

Figure 2: 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 (Distance=2617). (c) Best grid for test person BS after elastic graph matching with the model grid for person BP (Distance=3885). (d) Model grid for person BS. (e) Best grid for test person BP after elastic graph matching with the model grid for BS (Distance=3950). (f) Best grid for test person BS after elastic graph matching with the model grid (Distance=3174).

3. PERFORMANCE EVALUATION OF MORPHOLOGICAL DYNAMIC LINK ARCHITECTURE

The MDLA has been tested on the M2VTS database [11]. 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 286×350 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 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:

$$T_{BS}(01, BP) = D_{(j+Q)}, \quad Q = 0, 1, 2, \ldots. \tag{9}$$

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 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.,

$$D(BP_{01}, \{BS\}) = \min\{D(BP_{01}, BS_{02}), D(BP_{01}, BS_{03}), D(BP_{01}, BS_{04})\} \quad (10)$$

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 3a. The corresponding curve for the entire experiment is shown in Figure 3b. The ROC of DLA (without weighting the contribution of each node) from [6] is included in Figure 3b 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 %. 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 [6].

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	10.13	2.70
2	9.53	10.81
3	10.36	10.81
4	10.21	5.33

4. PRINCIPAL COMPONENT ANALYSIS IN MORPHOLOGICAL DYNAMIC LINK ARCHITECTURE

In this section our motivation for applying PCA in conjunction with MDLA and the results we have obtained are described. Representations based on PCA have been studied and extensively used for various applications. Among others PCA has been applied to face recognition e.g. [3]. For a detailed list of applications the interested reader is referred to [13, 14]. PCA aims at reducing the dimensionality of the original feature vectors so that the new vectors after this projection approximate in the sense of minimum mean squared error the original ones. PCA methods have shown good performance in image reconstruction/compression tasks. Accordingly, the feature vectors produced are called *most expressive features* (MEFs) [13].

However, there is no guarantee that the MEFs are necessarily good for discriminating among classes defined by a set of samples [13, 14]. It is well known that optimality in discrimination among all possible linear combinations of features can be achieved by employing Linear Discriminant Analysis (LDA). The feature vectors produced after the LDA projection are called *most discriminating features* (MDFs) [13]. Let us suppose that we would like at each grid node to find the projection that will enable the discriminating among the clients and the impostors in the face verification problem treated in this paper. This is a two class problem and the vector that is needed for such a projection is known as Fisher's linear discriminant that maximizes the ratio of between-class scatter to within-class scatter. The LDA breaks down when there are not enough samples so that the following inequalities are satisfied:

$$n \geq d + K \qquad d \geq K \quad (11)$$

where n is the number of feature vectors available at each grid node, d is the dimensionality of the feature vectors and K is the number of classes to be discriminated. In our case K=2, and n varies between $3 \leq n_C \leq 13$ and $160 \leq n_I \leq 240$ with n_C and n_I denoting the number of feature vectors available for the clients and the impostors, respectively. The small n_C is attributed to the limited number of frontal views for certain persons in the M2VTS database. Therefore, at each grid node although there is no problem in the estimation of the sample covariance matrix of the feature vectors for the impostors, there is indeed such a problem for the estimation of the sample covariance matrix of the feature vectors for the clients, if $d = 19$ as is in MDLA feature vectors. This problem can be resolved by applying the discriminant analysis after performing PCA [13].

We have found that by keeping at each grid node only the first six principal components the mean squared approximation error is less than 5 % at any node. Moreover, the EER achieved is 9.13 + 0.5 %. That is, we have achieved feature dimensionality reduction without sacrificing the verification efficiency of the MDLA.

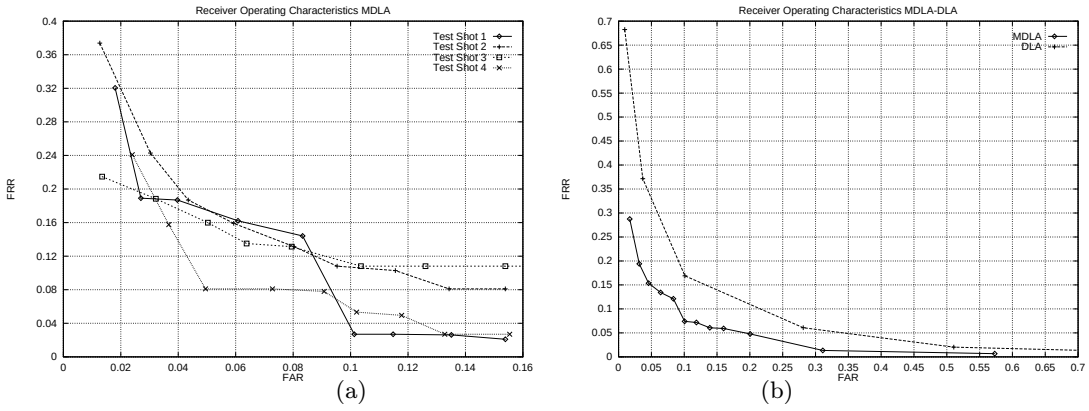


Figure 3: Morphological Dynamic Link Architecture Receiver Operating Characteristics for (a) each training set separately, and (b) the entire experiment.

5. 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. Feature vector dimensionality reduction has been achieved by employing principal component analysis without any deterioration in the verification efficiency of the method. Due to the lack of adequate number of frontal views for several persons in the M2VTS database used in the experiments conducted, principal component analysis serves as a necessary preprocessing step before applying linear discriminant analysis that is expected to improve further the performance of the method proposed. The application of linear discriminant analysis in the MDLA is the subject of our future research.

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

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

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

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

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

[11] 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 et al., Eds.), vol. 1206, pp. 403–409, 1997.

[12] 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.

[13] D.L. Swets, and J. Weng, "Discriminant analysis and eigenspace partition tree for face and object recognition from views," in *Proc. of the IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp.192–197, Killington, Vermont, U.S.A., October 14-16, 1996.

[14] K. Etemad, and R. Chellappa, "Discriminant Analysis for Recognition of Human Face Images," in *Lecture Notes in Computer Science: Audio- and Video-based Biometric Person Authentication* (J. Bigün et al., Eds.), vol. 1206, pp. 127–142, 1997.