LINEAR PROJECTION ALGORITHMS AND MORPHOLOGICAL DYNAMIC LINK ARCHITECTURE FOR FRONTAL FACE VERIFICATION

Constantine Kotropoulos *Ioannis Pitas*
Dept. of Informatics, Aristotle University of Thessaloniki,
Box 451, Thessaloniki 540 06, GREECE
Tel: +30 31 998225, 996304 Fax: +30 31 996304
e-mail: {costas, pitas}@zeus.csd.auth.gr

ABSTRACT

The combined use of linear projection algorithms with a variant of dynamic link architecture based on the multiscale morphological dilation-erosion is proposed for face verification. The performance of the combined scheme 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 incorporation of linear projections in the morphological dynamic link architecture improves significantly the verification capability of the method.

1 INTRODUCTION

Face recognition has exhibited a tremendous growth for more than two decades. A critical survey on face recognition can be found in [1]. 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]. The principles of this pattern recognition scheme can be traced back to the origins of self-organisation in neural networks. The 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 minimised.

The research on DLA and its applications has been an active research topic since its invention. A different topology cost for a particular pair of nodes has been proposed in [3]. It is based on the radius of the Apollonius sphere defined by the Euclidean distances between the nodes being matched. Three major extensions to the DLA have been introduced in [4] in order to handle larger galleries and larger variations in pose and to increase the matching accuracy. Recently, a novel variant of DLA that is based on multiscale morphological dilation-erosion, the so-called *Morphological Dynamic Link Architecture* (MDLA), has been proposed for face authentication [5]. More specifically, we have proposed 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].

A problem in elastic graph matching that has received much attention is the weighting of graph nodes according to their discriminatory power. Several methods have been proposed in the literature. For example, a Bayesian approach yields the more reliable nodes for gender identification, beard and glass detection in bunch graphs [7]. An automatic weighting of the nodes according to their significance by employing local discriminants is proposed in [8].

Frequently, linear projection algorithms are used to reduce the dimensionality of the 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 [9]. Two are the most popular linear projection algorithms. The *Karhunen-Loeve or Principal Component Analysis* (PCA) that does not employ category information and the *Linear Discriminant Analysis* (LDA) that exploits the category labels.

In this paper, we propose the combined use of MDLA with linear projection algorithms aiming at enhancing the verification capability of the MDLA. More specifically, we have used the PCA for two reasons: (i) to reduce the dimensionality of the feature vectors at the grid nodes, and (ii) to obtain uncorrelated feature vectors. We have employed the LDA in order to derive a class-dependent weighting vector for the feature vector at each grid node such that the dispersion within each class is minimised and the dispersion between that class and the set of all other classes is maximised. It is demonstrated that the combined use of both linear projection algorithms improves the performance of the MDLA in M2VTS database. The performance of the algorithm is evaluated in terms of its receiver operating characteristic.

2 LINEAR PROJECTIONS IN MORPHOLOGICAL DYNAMIC LINK MATCHING

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 study 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 gray-scale 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 $f(\mathbf{x}) : \mathcal{D} \subseteq \mathcal{Z}^2 \rightarrow \mathcal{R}$ 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 denoted by $(f \oplus g)(\mathbf{x})$. Its complementary the erosion is denoted by $(f \ominus g)(\mathbf{x})$. Their definitions can be found in any book on Digital Image Processing. 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. In this paper the *scaled hemisphere* is employed, i.e. [6]:

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

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

$$(f \star g_\sigma)(\mathbf{x}) = \begin{cases} (f \oplus g_\sigma)(\mathbf{x}) & \text{if } \sigma > 0 \\ f(\mathbf{x}) & \text{if } \sigma = 0 \\ (f \ominus g_\sigma)(\mathbf{x}) & \text{if } \sigma < 0. \end{cases} \tag{2}$$

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

$$\mathbf{j}(\mathbf{x}) = \left((f \star g_9)(\mathbf{x}), \ldots, (f \star g_1)(\mathbf{x}), f(\mathbf{x}), (f \star g_{-1})(\mathbf{x}), \ldots, (f \star g_{-9})(\mathbf{x}) \right). \tag{3}$$

An 8×8 sparse grid has been created by measuring the feature vectors $\mathbf{j}(\mathbf{x})$ at equally spaced nodes over the output of the face detection algorithm described in [10]. It has been demonstrated that such feature vectors captures important information for the key facial features [5].

Subsequently, a dimensionality reduction of feature vectors is pursued by employing PCA. PCA methods have shown good performance in image reconstruction/compression tasks. Accordingly, the feature vectors produced are called *most expressive features* (MEFs) [11]. In addition to dimensionality reduction PCA decorrelates the feature vectors and facilitates the LDA that is applied next in eigenvalue/eigenvector computations as well as in matrix inversion. Although MDLA has originally been applied to 4 sets of 37 frontal images, one for each person in M2VTS database [13], we now need much more frontal images. We have either extracted other frontal images for each person or we have augmented the original frontal images with others produced by slightly adding Gaussian noise as is proposed in [12]. Let

$$\mathbf{j}'(\mathbf{x}_l) = \mathbf{j}(\mathbf{x}_l) - \mathbf{m}(\mathbf{x}_l) \tag{4}$$

be the normalised feature vector at node $\mathbf{x}_l$ where $\mathbf{j}(\mathbf{x}_l) = (j_1(\mathbf{x}_l), \ldots, j_{19}(\mathbf{x}_l))^T$ and $\mathbf{m}(\mathbf{x}_l)$ is the mean feature vector at $\mathbf{x}_l$. Let N denote the total number of frontal images extracted for all persons. The covariance matrix of the feature vectors at node $\mathbf{x}_l$ is:

$$\mathbf{R}_l = \frac{1}{N} \sum_{i=1}^{N} \mathbf{j}_i{}'(\mathbf{x}_l) \mathbf{j}_i^{'\,T}(\mathbf{x}_l). \tag{5}$$

In PCA we compute the eigenvectors that correspond to the p largest eigenvalues of $\mathbf{R}_l$, say $\mathbf{e}_{1,\,l}, \ldots, \mathbf{e}_{p,\,l}$. The PCA projected feature vector is given by:

$$\hat{\mathbf{j}}(\mathbf{x}_l) = \begin{bmatrix} \mathbf{e}_{1,\,l}^T \\ \vdots \\ \mathbf{e}_{p,\,l}^T \end{bmatrix} \mathbf{j}^{'}(\mathbf{x}_l) = \mathbf{P}_l \mathbf{j}^{'}(\mathbf{x}_l). \tag{6}$$

Its dimensions are $p \times 1$, $p \leq 19$.

Next LDA is applied to feature vectors produced by PCA. It is well known that there is no guarantee that the MEFs are necessarily good for discriminating among classes defined by a set of samples [11, 12]. It is well known that optimality in discrimination among all possible linear combinations of features can be achieved by employing LDA. The feature vectors produced after the LDA projection are called *most discriminating features* (MDFs) [11]. We are interested in applying LDA at each grid node locally. It is evident that if we work at each grid node separately we can deal only with a two-class problem, i.e., $K = 2$. That is, we would like at each grid node to find the projection that will enable discriminating among feature vectors extracted from frontal facial images of the same person (i.e., the clients) and the feature vectors extracted from frontal facial images of the remaining persons in the database (i.e., the impostors). In the following the explicit dependence on $\mathbf{x}$ is omitted for notation simplicity. Let $\mathcal{S}$ be the entire set of feature vectors at a grid node and $\mathcal{S}_k$ be the corresponding set of features vectors at the same node extracted from the frontal facial images of the k-th person in the database. Our local LDA scheme determines a weighting matrix $(d \times p)$ $\mathbf{V}_k$ for the k-th person such that the ratio:

$$\begin{aligned} \mathcal{M}_k &= \frac{\operatorname{tr}\left[\mathbf{V}_k \left\{ \sum_{\hat{\mathbf{j}} \in \mathcal{S}_k} (\hat{\mathbf{j}} - \hat{\mathbf{m}}_k)(\hat{\mathbf{j}} - \hat{\mathbf{m}}_k)^T \right\} \mathbf{V}_k^T \right]}{\operatorname{tr}\left[\mathbf{V}_k \left\{ \sum_{\hat{\mathbf{j}} \in (\mathcal{S} - \mathcal{S}_k)} (\hat{\mathbf{j}} - \hat{\mathbf{m}}_k)(\hat{\mathbf{j}} - \hat{\mathbf{m}}_k)^T \right\} \mathbf{V}_k^T \right]} \\ &= \frac{\operatorname{tr}\left[\mathbf{V}_k \mathbf{W}_k \mathbf{V}_k^T \right]}{\operatorname{tr}\left[\mathbf{V}_k \mathbf{B}_k \mathbf{V}_k^T \right]} \end{aligned} \tag{7}$$

is minimised. In (7) $\hat{\mathbf{m}}_k$ is the class-dependent mean vector of the feature vectors which result after PCA. This is a generalised eigenvalue problem. Its solution is given by the eigenvector that corresponds to the minimal eigenvalue of $\mathbf{B}_k^{-1}\mathbf{W}_k$ or equivalently by the eigenvector that corresponds to the maximal eigenvalue of $\mathbf{W}_k^{-1}\mathbf{B}_k$ provided that both $\mathbf{W}_k$ and $\mathbf{B}_k$ are invertible. Because the matrix $\mathbf{W}_k^{-1}\mathbf{B}_k$ is not symmetric in general, the eigenvalue problem could be computationally unstable. A very elegant method that diagonalises the two symmetric matrices $\mathbf{W}_k$ and $\mathbf{B}_k$ and yields a stable computation procedure for the solution of the generalised eigenvalue problem has been proposed in [14]. In the following, a brief outline of the method is presented. $\mathbf{W}_k$ is by definition a symmetric matrix therefore it is diagonalisable, that is:

$$\mathbf{W}_k = \mathbf{H}_k \mathbf{\Lambda}_k \mathbf{H}_k^T \quad ; \quad \mathbf{H}_k \mathbf{H}_k^T = \mathbf{I} \tag{8}$$

where $\mathbf{\Lambda}_k$ is the diagonal matrix of the eigenvalues of $\mathbf{W}_k$ and $\mathbf{I}$ is the identity matrix. Moreover, the matrix

$$\mathbf{Q}_k = \left(\mathbf{H}_k \mathbf{\Lambda}_k^{-1/2}\right)^T \mathbf{B}_k \left(\mathbf{H}_k \mathbf{\Lambda}_k^{-1/2}\right) \tag{9}$$

is a symmetric one which implies the existence of an orthogonal matrix $\mathbf{U}_k$ and a diagonal matrix $\mathbf{\Sigma}_k$ such that

$$\mathbf{Q}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{U}_k^T. \tag{10}$$

Let $\mathbf{\Delta}_k = \mathbf{H}_k \mathbf{\Lambda}_k^{-1/2} \mathbf{U}_k$. Then:

$$\mathbf{W}_k^{-1}\mathbf{B}_k = \mathbf{\Delta}_k \mathbf{\Sigma}_k \mathbf{\Delta}_k^{-1} \tag{11}$$

which implies that $\mathbf{\Delta}_k$ consists of the eigenvectors of the matrix product $\mathbf{W}_k^{-1}\mathbf{B}_k$ and $\mathbf{\Sigma}_k$ contains its eigenvalues.

Let the superscripts t and r denote a test and a reference person (or grid), respectively. Having found the weighting matrix $\mathbf{V}_k(\mathbf{x}_l)$, for the l-th node of the k-th person in the database, we project the reference feature vector after PCA at this node onto subspace defined by the row vectors of $\mathbf{V}_k$ as follows:

$$\breve{\mathbf{j}}(\mathbf{x}_l^r) = \mathbf{V}_k \left[\mathbf{P}_l \left(\mathbf{j}(\mathbf{x}_l^r) - \mathbf{m}_l\right) - \hat{\mathbf{m}}_{kl}\right]. \tag{12}$$

Let us suppose that a test person claims the identity of the k-th person. Then the test scalar feature value at the l-th node is given by:

$$\breve{\mathbf{j}}(\mathbf{x}_l^t) = \mathbf{V}_k \left[\mathbf{P}_l \left(\mathbf{j}(\mathbf{x}_l^t) - \mathbf{m}_l\right) - \hat{\mathbf{m}}_{kl}\right]. \tag{13}$$

The L_2 norm of the difference between the MDF vectors at the l-th node has been used as a (signal) similarity measure, i.e.:

$$C_v(\breve{\mathbf{j}}(\mathbf{x}_l^t), \breve{\mathbf{j}}(\mathbf{x}_l^r)) = \|\breve{\mathbf{j}}(\mathbf{x}_l^t) - \breve{\mathbf{j}}(\mathbf{x}_l^r)\| \tag{14}$$

Let us denote by $\mathcal{V}$ the set of grid nodes. The grid nodes are simply the vertices of a graph. Let also $\mathcal{N}(l)$ denote the four-connected neighbourhood of vertex l. The objective is to find the set of test grid node coordinates $\{\mathbf{x}_l^t, \ \ l \in \mathcal{V}\}$ that yields the best matching. As in DLA [2], the quality of the match is evaluated by taking into account the grid deformations as well. Grid deformations can be penalised using the additional cost function:

$$C_e(i,j) = C_e(\mathbf{d}_{l\,\xi}^t, \mathbf{d}_{l\,\xi}^r) = \|\mathbf{d}_{l\,\xi}^t - \mathbf{d}_{l\,\xi}^r\| \quad \xi \in \mathcal{N}(l) \tag{15}$$

with $\mathbf{d}_{l\,\xi} = (\mathbf{x}_l - \mathbf{x}_\xi)$. The penalty (15) can be incorporated to a cost function:

$$C(\{\mathbf{x}_l^t\}) = \sum_{l \in \mathcal{V}} \left\{ C_v(\breve{\mathbf{j}}(\mathbf{x}_l^t), \breve{\mathbf{j}}(\mathbf{x}_l^r)) + \lambda \sum_{\xi \in \mathcal{N}(l)} C_e(\mathbf{d}_{l\xi}^t, \mathbf{d}_{l\xi}^r) \right\}. \tag{16}$$

One may interpret (16) as a simulated annealing with an additional penalty (i.e., a constraint on the objective function). Since the cost function (15) does not penalise translations of the whole graph the random configuration $\mathbf{x}_l^t$ can be of the form of a random translation $\mathbf{d}$ of the (undeformed) reference grid and a bounded local perturbation $\underline{\delta}_l$, i.e.:

$$\mathbf{x}_l^t = \mathbf{x}_l^r + \mathbf{d} + \underline{\delta}_l \quad ; \quad \|\underline{\delta}_l\| \leq \delta_{\max} \tag{17}$$

where the choice of $\delta_{\max}$ controls the rigidity/plasticity of the graph. It is evident that the proposed approach differs from the two stage coarse-to-fine optimisation procedure proposed in [2]. In our approach we replace the two stage optimisation procedure with a probabilistic hill climbing algorithm which attempts to find the best configuration $\{\mathbf{d}, \{\underline{\delta}_l\}\}$ at each step.

3 PERFORMANCE EVALUATION OF THE COMBINED SCHEME

The combined scheme of MDLA with linear projections has been tested on the M2VTS database [13]. The database contains both sound and image information. 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. Four experimental sessions have been implemented by employing the "leave one out" principle. Details on the experimental protocol used in the performance evaluation can be found in [5].

Let $T_{BS}(01, BP)$ denote the threshold 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 $D_{(1)}(BS;\, 01, BP)$ denote the minimum impostor distance for BS when shot 01 is left out and person BP is excluded. The threshold is chosen as follows:

$$T_{BS}(01, BP) = D_{(1+q)}(BS;\, 01, BP), \quad q = 0, 1, 2, \ldots . \tag{18}$$

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 a plot 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. This plot is the *Receiver Operating Characteristic* (ROC) of the verification technique. The ROCs of the MDLA with one and two MDFs are plotted in Figure 1. In the same plot the ROC of MDLA without linear projections [5] is also depicted. The Equal Error Rate (EER) of a technique (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 with one MDF is 6.8% and with two MDF is 5.4% whereas the EER of MDLA without any linear projections is 9.35 % [5]. It is seen that the incorporation of linear projections improves the EER by 2.55% - 4%. The comparison of EER achieved by the proposed scheme is identical to or better than the ones reported in [8] (i.e., an EER between 5.4% and 9.2 %).

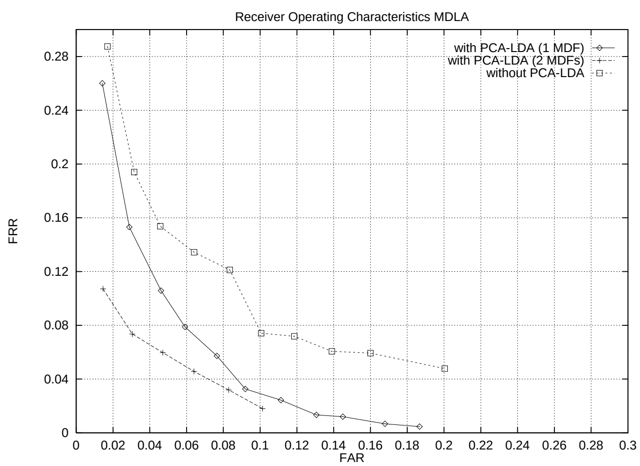

Figure 1: Morphological Dynamic Link Architecture Receiver Operating Characteristics with and without linear projections.

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 invariant object recognition in the Dynamic Link Architecture," *IEEE Trans. on Computers*, vol. 42, no. 3, pp. 300–311, March 1993.

[3] B.S. Manjunath, R. Chellapa, and C. v.d. Malsburg, "A feature based approach to face recognition,' in *Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 373–378, 1992.

[4] L. Wiskott, J.-M. Fellous, N. Krüger, and C. v.d. Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, July 1997.

[5] C. Kotropoulos, and I. Pitas, "Face authentication based on morphological grid matching," in *Proc. of the IEEE Int. Conf. on Image Processing (ICIP 97)*, pp. I-105–I-108, Santa Barbara, California, U.S.A., 1997.

[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] L. Wiskott, "Phantom faces for face analysis," in *Lecture Notes in Computer Science:Computer Analysis of Images and Patterns* (G.Sommer, K. Daniilidis, J. Pauli, Eds.), vol. 1296, pp. 480–487, 1997.

[8] B. Duc, S. Fischer, and J. Bigün, "Face authentication with Gabor information on deformable graphs," *IEEE Trans. on Image Processing*, submitted 1997.

[9] A.K. Jain, and R.C. Dubes, *Algorithms for Clustering Data.* Englewood Cliffs, NJ: Prentice Hall, 1988.

[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.

[11] 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.

[12] 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, G. Chollet and G. Borgefors, Eds.), vol. 1206, pp. 127–142, 1997.

[13] 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.

[14] D.L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. on Pattern Analysis and Machine Intelligence* vol. 18, no. 8, pp. 831–837, August 1996.