Face Authentication Using Morphological Dynamic Link Architecture

Constantine Kotropoulos[1], Ioannis Pitas[1], Stefan Fischer[2], Benoît Duc[2]

[1] Department of Informatics, Aristotle University of Thessaloniki
Thessaloniki 54006, Greece. E-mail: {costas, pitas}@zeus.csd.auth.gr
[2] Signal Processing Laboratory, Swiss Federal Institute of Technology,
CH-1015 Lausanne, Switzerland. E-mail: {fischer, duc}@ltssg3.epfl.ch

Abstract. A very attractive approach for face detection is based on multiresolution images (also known as *mosaic images*). Motivated by the simplicity of this approach, a rule-based face detection algorithm in frontal views is developed first. Second, a novel dynamic link architecture based on *multiscale morphological dilation-erosion* is proposed for face authentication. More specifically, a sparse grid is placed over the outcome of face detection stage for each person in a reference set. Subsequently, multiscale morphological operations are employed to yield a feature vector at each node of the grid and dynamic link matching is applied to verify the identity of each person from a test set. The first experimental results reported in this paper verify the superiority of the proposed method over the (standard) dynamic link matching that is based on Gabor wavelets.

1 Introduction

Face recognition has been an active research topic in computer vision for more than two decades. A critical survey of existing literature on human and machine face recognition is given in [1]. One of the key problems in building automated systems that perform face recognition tasks is face detection. Recently, the research on model-assisted coding schemes in addition to the need for multimodal verification techniques in tele-services and teleshopping applications has reinforced the interest on face detection algorithms. Many algorithms have been proposed for face detection in still images that are based either on texture, depth, shape and color information or a combination of them. 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 [2].

Two main categories for face recognition techniques can be identified in the literature: those employing geometrical features (for example [3]) and those using grey-level information (e.g. the eigenface approach [4]). A different approach that uses both grey-level information and shape information has been proposed in [5]. 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., geometry) is minimized. However, norms of the difference between feature vectors based either on the magnitude or on the phase of the Gabor filters response are proved inadequate to discriminate an impostor against the authentic person [6]. Accordingly, a local discriminant criterion is proposed to overcome the sensitivity of the method [7].

Motivated by the simplicity of the face detection approach in [2] and the need for a mechanism that controls the placement of a sparse grid over a face in order to store a model for each person in dynamic link matching, we develop first a variant of this method that has the following features: (1) It allows for rectangular cells in contrast to the square cells used in [2]. (2) 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. (3) It has very low computational demands compared to the original algorithm [2], because the iterative nature of the algorithm is avoided due to the preprocessing step that has been employed. (4) It employs more general rules that are close to our intuition for a human face.

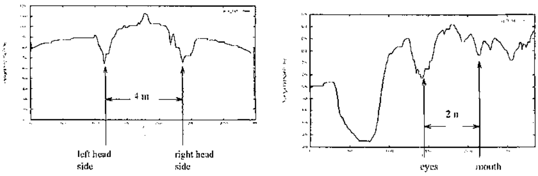
Second, motivated by the limitations of Gabor based feature vectors to capture discriminant characteristics of facial images as is pointed out above, a novel dynamic link architecture based on multiscale morphological dilation-erosion is proposed and tested for face authentication. That is, we propose 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]. The first experimental results reported in this paper indicate the superiority of the proposed method over the (standard) dynamic link matching with Gabor-based feature vectors.

The outline of this paper is as follows. The face detection algorithm developed is briefly presented in Section 2. The Morphological Dynamic Link Architecture (MDLA) is described in Section 3. The evaluation of performance of MDLA with respect to its Receiver Operating Characteristics (ROC) is treated in Section 4. Conclusions are drawn and further research directions are indicated in Section 5.

2 Face detection in frontal views

To begin with, let us describe the framework proposed in [2]. The original algorithm is based on mosaic images of reduced resolution that attempt to capture the macroscopic features of the human face. It is assumed that there is a resolution level where the main part of the face occupies an area of about 4×4 cells. Accordingly, a mosaic image, the so called *quartet* image, can be created for this resolution level. By subdividing each quartet image cell to 2×2 cells of half dimensions the *octet* image results where the main facial features such as the eyebrows/eyes, the nostrils/nose and the mouth can be detected. Therefore, an hierarchical knowledge-based system can be designed that aims at detecting facial candidates by establishing rules applied to the quartet image and subsequently at validating the choice of a facial candidate by establishing rules applied to the octet image that detect the key facial features mentioned above.

The implementation proposed in [2] is computationally intensive. The algorithm is applied iteratively for the entire range of possible cell dimensions in order to determine the best cell dimensions for creating the quartet image of each person. Another limitation is that only square cells are employed. In order to avoid the iterative nature of the original method, we propose to estimate the cell dimensions in the quartet image by processing the horizontal and the vertical profile of the image. It is worth noting that such a preprocessing step can be applied if a single person appears in the scene and the background is fairly uniform. Let us denote by n and m the quartet cell dimensions. The horizontal profile of the image is obtained by averaging all pixel intensities in each image column. By detecting abrupt transitions upwards, the quartet cell dimension in the horizontal direction is estimated, as is depicted in Fig. 1a. Similarly, the vertical profile of the image is obtained by averaging all pixel intensities in each image row. It is fairly easy to locate the image row where the eyebrows/eyes appear in the image by detecting the local minimum after the first abrupt transition in the vertical profile. The row where the upper lip appears in the image can be found by searching for the steepest minimum after a significant maximum that occurs below the eyes. The distance between the eyebrows/eyes and the upper lip is used for estimating the quartet cell dimension in the vertical direction, as is shown in Fig. 1b. Having estimated the quartet cell dimensions



(a) (b)

Fig. 1. (a) Horizontal profile. (b) Vertical profile.

we proceed to the description of facial candidate detection rules. As in [2], first a homogeneous region of 2×2 quartet cells in the middle is detected. Then, we search for homogeneous connected components of significant length in a Π-shaped region around the facial candidate. A significant difference in average grey levels between these two regions should be detected. Beard detection rules have been added in the original model as well. Next eyebrows/eyes, nostrils/nose and mouth detection rules are developed to validate the facial candidates determined by the procedure outlined above. The rules developed enhance the ones proposed in [2] by highlighting the key role of symmetry. A detailed description of the rules implemented is given in [12].

The proposed algorithm has been applied to the European ACTS project M2VTS database [11]. The database includes the videosequences of 37 different persons. The algorithm provides a correct facial candidate in *all* cases. However, the detected facial features that validate the choice of the facial candidate are not always correct. The true success rate of the proposed method in detecting simultaneously eyebrows/eyes, nostrils/nose and mouth is 86.5 % under the most strict evaluation conditions [12]. Fig. 2 shows facial candidates that have successfully been detected and the facial features are located correctly. The octets for eyebrows/eyes and nostrils/nose are shown as white overlaid rectangles. Mouth candidates are shown as black overlaid rectangles. The white cross indicates the characteristic bright octet between the eyes.

Fig. 2. Successful face detection results.

3 Morphological 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 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 $\mathbb{R}$ and $\mathbb{Z}$ denote the set of real and integer numbers respectively. Given an image $f(\mathbf{x}) : \mathcal{D} \subseteq \mathbb{Z}^2 \rightarrow \mathbb{R}$ and a structuring function $g(\mathbf{x}) : \mathcal{G} \subseteq \mathbb{Z}^2 \rightarrow \mathbb{R}$ the dilation of the image $f(\mathbf{x})$ by the structuring function $g(\mathbf{x})$ and its complementary operation, the erosion, are denoted by $(f \oplus g)(\mathbf{x})$ and $(f \ominus g)(\mathbf{x})$, respectively. Their definitions can be found in [9, 10]. 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 that is given by [8]:

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

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

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

The output of multiscale dilation-erosion for $\sigma = -9 \ldots 9$ forms the feature vectors located at the node $\mathbf{x}$ of the grid yielding the so called *Morphological Dynamic Link Architecture* (MDLA):

$$\mathbf{J}(\mathbf{x}) = ((f \star g_9)(\mathbf{x}), \ldots, f(\mathbf{x}), \ldots, (f \star g_{-9})(\mathbf{x})) \ . \quad (3)$$

Fig. 3 depicts the output 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 such as the eyebrows, eyes, nose tip, nostrils, lips, face contour etc. Let the superscripts

Fig. 3. Responses of dilations and erosions for scales 1–9.

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.: $S_v(\mathbf{J}(\mathbf{x}_i^t), \mathbf{J}(\mathbf{x}_i^r)) = ||\mathbf{J}(\mathbf{x}_i^t) - \mathbf{J}(\mathbf{x}_i^r)||$. As in DLA [5], 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 that form the vertices of a graph. Then, an additional cost function is used:

$$S_e(i,j) = S_e(\mathbf{d}_{ij}^t, \mathbf{d}_{ij}^r) = ||\mathbf{d}_{ij}^t - \mathbf{d}_{ij}^r|| \quad \forall i \in \mathcal{V}; j \in \mathcal{N}(i) \quad (4)$$

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 (4) 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

$$C(\{\mathbf{x}_i^t\}) = \sum_{i \in \mathcal{V}} \left\{ S_v(\mathbf{J}(\mathbf{x}_i^t), \mathbf{J}(\mathbf{x}_i^r)) + \lambda \sum_{j \in \mathcal{N}(i)} S_e(\mathbf{d}_{ij}^t, \mathbf{d}_{ij}^r) \right\} . \quad (5)$$

In the minimization of (5) the coarse-to-fine approach proposed in [5] 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. The outputs of multiscale dilation-erosion for scales $\sigma = -9, \ldots, 9$ has been concatenated to form the feature vectors at the grid nodes.

4 Performance evaluation of Morphological Dynamic Link Architecture

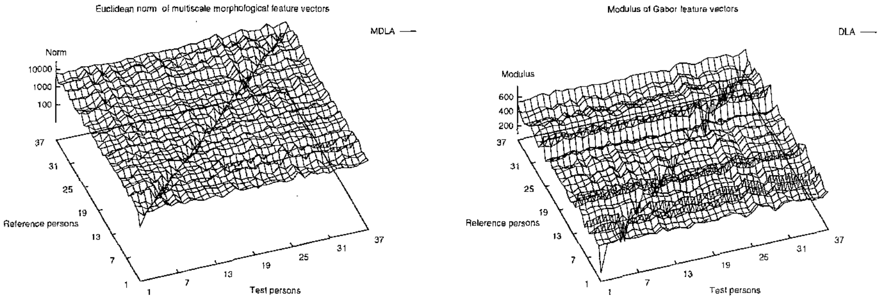
The MDLA has been tested on the M2VTS database of 37 persons [11]. A reference set has been created by choosing a frontal view of each person from the third shot. A test set has been created by choosing another frontal or near frontal view of each person from this shot. Although equation (5) provides a nice instrumental ad-hoc cost function to perform grid matching, it does not combine signal differences and geometrical deformations in a theoretically sound manner yielding thus a distance measure between a test and a reference person. In this paper we shall resort only to a measure of signal similarity. That is, we have used solely the sum of L_2 norms of multiscale morphological feature vectors 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 is plotted in Fig. 4, when MDLA is used. For comparison purposes the same experiment has been repeated with DLA. In the latter case, the sum of moduli of Gabor feature vectors at all grid nodes has been used as a distance between a reference and a test person. The plot of distances is shown in Fig. 4b. It is clearly seen that MDLA has much more minima in the main diagonal than DLA. The DLA has succeeded to identify 17 persons from 37 whereas the MDLA has succeeded to identify 31 persons from 37.

Both DLA and MDLA 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:

$$FRR(T) = 1 - \frac{1}{N} \text{card}\,\{\forall i : i = \arg\min\{\Delta(t_i, r_j)\} \text{ and } \Delta(t_i, r_i) \leq T\} \tag{6}$$

where card$\{\ \}$ denotes the cardinality of a set and $j = 1, \ldots, N$. Accordingly, the *false acceptance rate* at threshold T, $FAR(T)$, is given by:

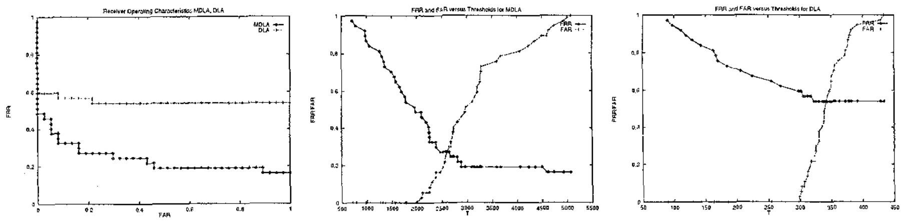
$$FAR(T) = \frac{1}{N} \text{ card}\,\{\forall i : k = \arg\min\{\Delta(t_i, r_j);\ j \neq i\} \text{ and } \Delta(t_i, r_k) \leq T\}\ . \tag{7}$$



(a) (b)

Fig. 4. Plots of elastic graph distances for all pairs of test and reference persons in MDLA and DLA. (a) Euclidean norm of multiscale morphological feature vectors. (b) Modulus of Gabor feature vectors.

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 MDLA and the DLA for the distance measures used is plotted in Fig. 5a. It is seen that MDLA outperforms DLA. For the same FAR=20%, the FRR for MDLA is 27% whereas the FRR for DLA is 56%. The plots of FAR/FRR versus the threshold T are given in Fig. 5b and 5c respectively.



(a) (b) (c)

Fig. 5. (a) Receiver Operating Characteristics for the MDLA and the DLA. (b) Plot of False Rejection/Acceptance Rate versus threshold for MDLA. (c) Plot of False Rejection/Acceptance Rate versus threshold for DLA.

5 Conclusions

A rule-based detection algorithm in frontal views has been outlined in this paper. The output of the algorithm has been used to control the placement of the grid in dynamic link matching. A novel multiscale morphological dynamic link architecture has been proposed and tested. The first experimental results that have been 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 experimental results reported in this paper.

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

References

1. Chellapa, R., Wilson C.L., and Sirohey, S.: Human and machine recognition of faces: A survey. Proceedings of the IEEE, **83** (1995) 705-740
2. Yang, G., and Huang, T.S.: Human face detection in a complex background. Pattern Recognition, **27** (1994) 53–63
3. Brunelli, R., and Poggio, T.: Face recognition: Features versus Templates. IEEE Trans. on Pattern Analysis and Machine Intelligence **15** (1993) 1042–1052
4. Turk, M., and Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience **3** (1991) 71–86
5. Lades, M., Vorbrüggen, J.C., Buhmann, J., Lange J., v.d. Malsburg, C., Würtz, R.P., and Konen, W.: Distortion invarinat object recognition in the Dynamic Link Architecture. IEEE Trans. on Computers **42** (1993) 300-311
6. Fischer, S., Duc, B., and Bigün, J.: Face recognition with Gabor Phase and Dynamic Link Matching for Multi-Modal Identification. Technical Report LTS 96.04 (1996), Signal Processing Laboratory, Swiss Federal Institute of Technology
7. Duc, B., Fischer, S., and Bigün, J.: Face authentication with sparse grid Gabor information. In Proc. of 1997 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (to appear)
8. Jackway, P.T. and Deriche, M.: Scale-space properties of the multiscale morphological dilation-erosion. IEEE Trans. on Pattern Analysis and Machine Intelligence **18** (1996) 38–51
9. Pitas, I., and Venetsanopoulos, A.N.: Nonlinear Digital Filters: Principles and Applications. Norwell, MA: Kluwer Academic Publ. (1990)
10. Haralick, R.M., Sternberg, S.R., and Zhuang, X.: Image analysis using mathematical morphology. IEEE Trans. on Pattern Analysis and Machine Intelligence **9** (1987) 532–550
11. Pigeon, S., and Vandendorpe, L.: The M2VTS multimodal face database. In Proc. of the First International Conference on Audio- and Video-based Biometric Person Authentication (this volume)
12. Kotropoulos, C., and Pitas, I.: Rule-based face detection in frontal views. In Proc. of the 1997 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (to appear)