FACE AUTHENTICATION WITH SPARSE GRID GABOR INFORMATION

Benoît Duc, Stefan Fischer and Josef Bigün

Signal Processing Laboratory, Swiss Federal Institute of Technology,
CH-1015 Lausanne, Switzerland

ABSTRACT

This paper investigates the application of statistical pattern recognition methods in the framework of the *Dynamic Link Matching* approach. This method describes objects by means of local frequency information on nodes of a sparse grid. Matching of an input image with a reference is achieved by displacement and deformation of the grid. This method is applied here to the authentication of human faces in a cooperative scenario where candidates claim an identity that is to be checked. The matching error is not powerful enough to provide satisfying results in this case. We introduce an automatic weighting of the nodes according to their significance. Results show that for regular grids, this weighting leads to a significant improvement of the performance.

1. INTRODUCTION

Recognition done by the human being can be seen at the same time as a holistic and a feature analysis approach[4]. Computer-based face recognition often privileges one of these aspects. Template-based methods that attempt to match well-defined portions of the face (eye, mouth) belong to the analysis category [3, 13]. The eigenface approach [11] describes images in terms of linear combinations of basis images, and thus represents a global approach. Methods that constrain local features by adding geometrical constraints can be considered as a mixture of both aspects. One can cite the Dynamic Link Architecture (DLA) [8] and related graph-based feature matching approaches [9], as well as methods based on Neural Networks, and feature-based approaches where features are geometrical measures [3].

This paper investigates the applicability of statistical pattern recognition methods in the framework of the *Dynamic Link Matching* approach [8, 12]. The method is illustrated by the verification of identity from human faces. Here, a special emphasis is put on the development of discriminant measures for face images, which leads to significantly improved performance.

In the identity verification scenario considered here, the person willing to access a building or service is cooperative, i.e. it is assumed that one can ask the person to face the camera and to claim his identity, which may be either true, or false in case of an impostor.

An authentication procedure is decomposed into two steps. First, the input face is matched with the prototype of the claimed person. Then the matching is evaluated by using a discriminant function, leading to the acceptance or rejection of the candidate. This kind of hypothesis testing function is designed on training data, which consist of several frontal views of each person in the database.

The paper is organised as follows: the matching of a grid is described in Section 2. The determination of the discriminant measure in Section 3. Some results are shown in Section 4. Finally, conclusions are drawn and future developments are discussed.

2. GRID MATCHING FOR FACE RECOGNITION

2.1. Feature Vectors

Each face is represented by a set of feature vectors positioned on nodes of a coarse, rectangular grid placed on the image. Comparing two face images is accomplished by matching and adapting a grid taken from one image to the features of the other image [8]. We use the modulus of complex Gabor responses as feature vectors from filters with 6 orientations and 3 resolutions. These are sets of features that describe local properties of points in the image, similar to those in [1].

2.2. Grid Matching

The grid matching serves two purposes: firstly, it aims at normalising the input, in order to make the subsequent comparison invariant with respect to translation and a reasonable amount of deformation. Secondly, the residual error accounts for the difference between the normalised input and the reference pattern. Intuitively, the higher the error, the higher the probability to have an impostor.

The method for grid matching we employ consists of two steps. The first minimises an error function depending on the difference of the reference and the test feature vectors, by translating the undeformed grid that has been stored in a database and attempting to find the best correspondence between the grid and the image to be verified. The second step translates every feature vector separately to find individual minima, thus resulting in a deformed grid. At this point, a penalising term for grid deformation is added to the error function, so that the coarse geometry of the grid is conserved.

3. FEATURE EXTRACTION

The first step of the authentication consists in matching the image with the prototype grid of the claimed class (in the following, each person in the database is considered as

0-8186-7919-0/97 $10.00

a *class* of the classification problem). This prototype is taken as the mean of the feature vectors provided by all views of the considered person in the training set. It is expected that if the claimed identity is correct, the feature vector will be close to the prototype of the class; in case of an impostor, the matching will perform poorly. Unfortunately, early experiments showed that a simple Euclidean distance measure is not sufficient to discriminate between an impostor and the authentic person, see Section 4.2. This is partly due to the presence of noise in the measurement. Indeed, the feature space considered here is very large: for an 8 by 8 grid comprising 18 Gabor responses at each node, a total of 1152 features is obtained.

Reducing the dimensionality is an efficient way to reduce the influence of noise [6]. From a training set consisting of several frontal views of each person, one establishes subspaces which maximise the dispersion of all classes while minimising the dispersion within the classes.

This is achieved by designing a local discriminant measure for each resolution, at each node of the grid, so that a subsequent selection of most significant nodes can be accomplished. Furthermore, the dimension of the local feature space becomes small compared to the number of training samples, so that an "over-training" of the discriminant measure with respect to training samples is reduced.

3.1. Local Discriminants

Suppose that the dimensionality of the considered feature space is small compared to the number of training elements in each of the c considered classes. One would like to establish a decision criterion for the acceptance or rejection of the candidate. This criterion should be "small" if the candidate is the right person, and "large" in case of an impostor. Obviously, this decision has to be made on the difference between the prototype of the claimed class and the measured feature vector. The components of this difference do not bear the same significance, as some may be more relevant than others for the given class. Therefore, we propose the following discriminant criterion:

$$d_k(\mathbf{r}) = \left(\sum_{i=1}^{N_G} v_{k_i} (r_i - \hat{\mu}_{k_i}) \right)^2 \qquad (1)$$

for class $k, k = 1 \ldots c$, where r_i are the components of the measurement vector $\mathbf{r}$, N_G is the dimension of the feature space. The unknown coefficients v_{k_i}'s are determined on the training set by minimising the ratio:

$$D_k = \frac{\sum_{\mathbf{r} \in S_k} d_k(\mathbf{r})}{\sum_{\mathbf{r} \in (S - S_k)} d_k(\mathbf{r})}, \qquad (2)$$

where S_k is the set of training vectors belonging to class k, S is the whole training set, so that $(S - S_k)$ is the set of all impostors for class k. This formulation leads to an eigenvalue problem, and $\mathbf{v}_k$ is given by the eigenvector corresponding to the smallest eigenvalue.

All local responses have to be combined in order to provide a unique, global discriminant value for the considered face. This is a problem related to sensor or decision fusion [5]. Here, we build the global response by simply adding the contributions from the local discriminants.

4. EXPERIMENTS

4.1. Face Database and Experimental Setup

This work is part of a project aiming at authentication methods based on several modalities, such as speech, frontal and profile views of the face. The use of several modalities required the acquisition of a multi-modal database which contains both sound and image information [10]. It includes 4 shots of 37 individuals, which were taken at one week intervals. For each shot, people were asked to rotate the head from 0 to -90 degrees, again to 0, then to +90 and back to 0 degrees.

For image-based authentication, the rotation sequences were considered, by using the luminance information only, at the QCIF format (144×176). Frontal images are selected automatically using a symmetry measure. We select images that have a symmetric gray-level distribution in a rectangular region. This region has the shape of a horizontal bar that covers the face region and the background on the two sides of the face. We compute the horizontal centre of gravity and the symmetry in respect to this centre and select images with an extremal symmetry measure. We have also evaluated edge-based methods to compute a symmetry measure. These methods fail mainly because the strong contrast between hair and background and between hair and skin dominate the image. Asymmetric hair cuts can result in a failure of the frontal face detection scheme.

The experiments were conducted following a combination of the left-one-out and the rotation estimates [6]. Alternatively, each person is labelled as an *impostor*, while the 36 others are considered as *clients*. For each combination, three shots of the 36 clients build the training set while the fourth shot is used as evaluation set in the following way: each client tries to access under its own identity, and the impostor tries to access under the identity of the 36 clients. This sums up to 36 authentic tests and 36 impostor tests. This procedure is repeated four times, by considering each shot for evaluation successively. In total, the client and impostor verifications amount each to $37 \times 4 \times (37-1) = 5328$.

4.2. Results

An example of grid matching is shown in Figure 1. In order to motivate the process of feature extraction, we first want to show that the Euclidean distance measure between features is not sufficient for a reliable decision. Figure 2 shows as an example distances of training and test samples with person 15 used as reference. It turns out that the distance to the reference view is clearly not sufficient to detect impostors.

A representation of discriminant values for the same person is shown in Figure 3. Now the discrimination of impostors is much more powerful. One can notice that there seems to be some over-training, as the discrimination measure is almost zero for all members of the considered class in the training set, and significantly larger for images of the same class in the test set, while remaining smaller than the threshold. This is due to the small number of training samples for each person in the database.

At that point, the discriminant values taking values in the $[0, \infty[$ interval are normalised to the $[0, 1]$ interval,

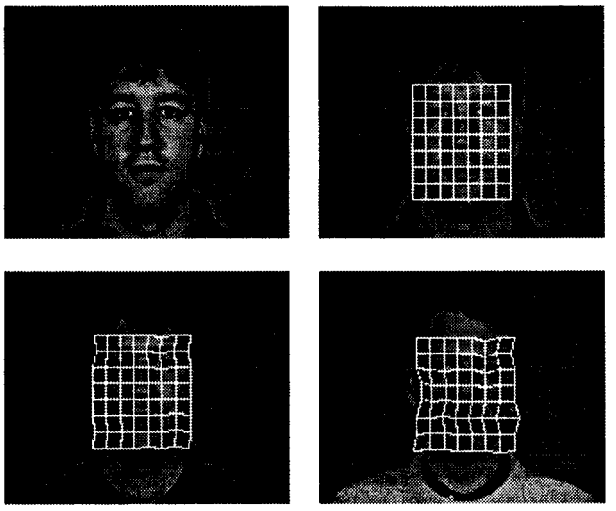

Figure 1: Example of a grid matching. Top left: reference image. Top right: reference grid. Bottom left: matched grid on another image of the same person (distance to reference grid: 740). Bottom right: matched grid on another person (distance to reference grid: 1406).

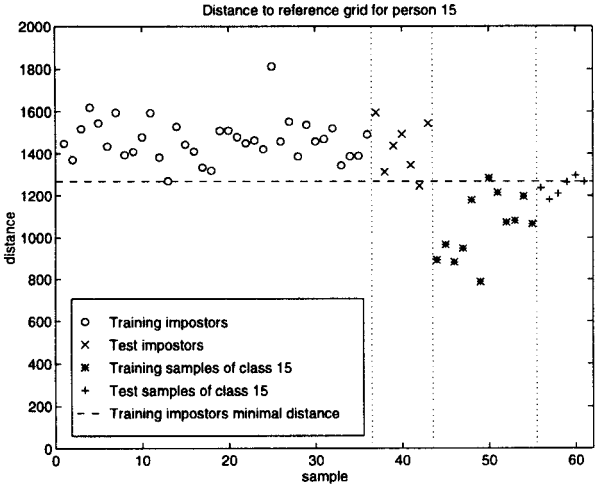

Figure 2: Plot of distances for person (or class) 15. The distance of the grids of different kind of images, namely impostors in the training and the test set, members of the class in the training and test set, are shown. If one uses the minimal distance on the training impostors as a threshold for the decision, some members of the class in the training and test set are misclassified.

so that they can be combined with or compared to other verification modalities like speech. Nevertheless, using a hard threshold is useful for comparing performance of different alternatives. As an illustration of the usefulness of the discriminant measure over all classes, we show the Receiver Operating Characteristics (ROC) of the Euclidean distance and the global discriminant value in Figure 4. Such curves reflect the performance of a given solution averaged on *all*

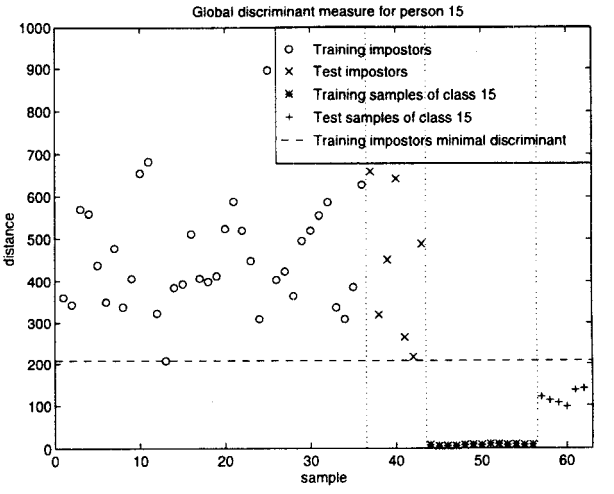

Figure 3: Plot of discriminant values for person (or class) 15. The image indices correspond to the ones of Figure 2. If one uses the minimal discrimination measure over the training impostors as a threshold for the decision, then all points are correctly classified. Members of the class and impostors are better separated than in Figure 2.

classes. The points on the ROC were obtained by scaling the minimum threshold displayed in Figures 2 and 3 with a varying factor. Clearly, the discriminant measure outperforms the simple distance everywhere.

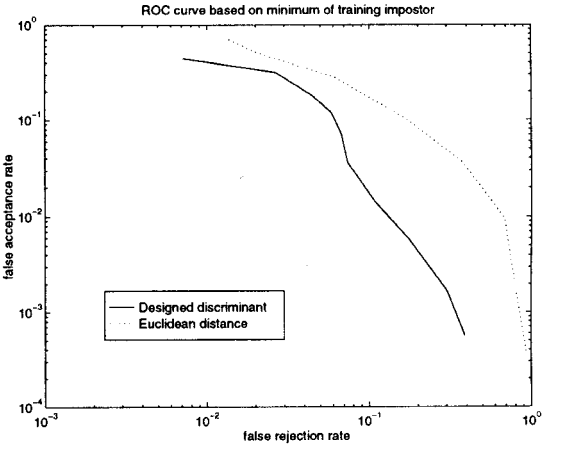

Figure 4: Experimental ROC curve for the distance and the discriminant measure in a log-log scale.

Another interesting point is that it is not necessary to take into account all local discriminants in order to obtain good performance. By retaining only a fraction of discriminants that reach the largest values of the criterion in Eq. (2), results are almost as good as with all discriminants taken into account. Figure 5 shows that with the minimum threshold, the false acceptance rate is almost not affected by the number of most significant discriminants taken into account. This behaviour is due to the fact that we conside-

red as a threshold the *minimum* discriminant value of the training impostors. On the contrary, the false rejection rate is sensitive to the number of discriminants: it decreases rapidly and then reaches a plateau at an approximate value of false rejection rate of 0.1, see Figure 5.

If we choose the threshold as the maximum of the training samples of the right class, an inverse behaviour can be expected, namely a false rejection rate insensitive to the number of retained discriminants, and a sensitive false acceptance rate. We chose the "minimum" threshold as it seems to provide lower false acceptance rates, which is a desired property for secured accesses.

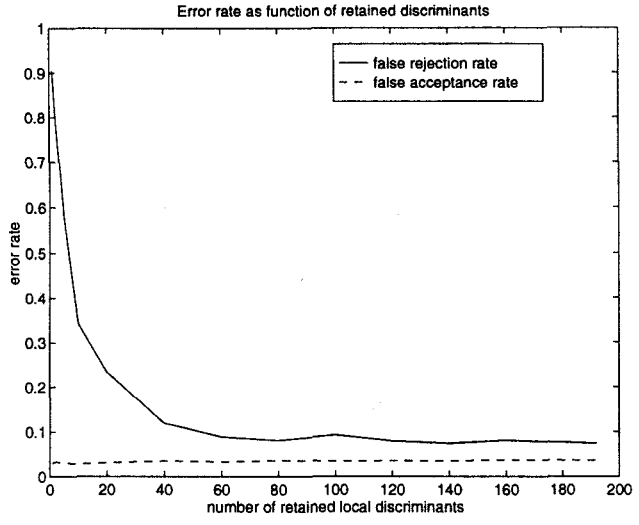

Figure 5: Error rates as a function of the number of most significant discriminant retained, with a threshold given by the minimum value of training impostors.

This provides a natural way of selecting relevant nodes for verification of a given class. From that point, it becomes possible to re-design a grid, adapted to the claimed identity.

5. DISCUSSION AND CONCLUSION

In this contribution, we have shown that matching sparse local frequency information arranged on a regular grid may be used for face verification. A linear discrimination approach has been presented, that weights the contributions of each feature according to its significance for the considered class. It improves the performance significantly.

In future, we will compare different frequency information provided by the Gabor decomposition, namely the complex response, its phase and its modulus. In particular, phase information shows interesting properties, like robustness with respect to illumination.

Finally, this verification application will be embedded into a multi-modal approach for person authentication, which promises lower error rates than mono-modal methods [2, 7].

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

6. REFERENCES

[1] J. Bigün and J. M. H. du Buf. "N-folded symmetries by complex moments in gabor space". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 1, pp. 80–87, January 1994.

[2] R. Brunelli and D. Falavigna. "Person identification using multiple cues". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 10, pp. 955–966, October 1995.

[3] R. Brunelli and T. Poggio. "Face recognitionl; features versus templates". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 10, pp. 1042–1043, October 1993.

[4] R. Chellappa, 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.

[5] B. V. Dasarathy. *Decision Fusion.* IEEE Computer Society Press, 1994.

[6] P.A. Devijver and J. Kittler. *Pattern Recognition: a Statistical Approach.* Prentice-Hall International, London, 1982.

[7] B. Duc, G. Maître, S. Fischer, and J. Bigün. "Person authentication by fusing face and speech information". In *First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'97)*, Crans-Montana, Switzerland, March 12-14 1997.

[8] M. Lades, J. Buhmann J. C. Vorbrüggen, J. Lange, C. v.d. Malsburg, R. P. Würtz, and W. Konen. "Distortion invariant object recognition in the dynamic link architecture.". *IEEE Transactions on Computers*, Vol. 42, No. 3, pp. 300–311, March 1993.

[9] B. S. Manjunath, R. Chellappa, and C. v. d. Malsburg. "A feature based approach to face recognition". In *IEEE Computer Society on Computer Vision and Pattern Recognition*, pp. 373–378. IEEE Computer Society Press, June 1992.

[10] S. Pigeon and L. Vandendorpe. "The m2vts multimodal face database". In *First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'97)*, Crans-Montana, Switzerland, March 12-14 1997. (http://www.tele.ucl.ac.be/M2VTS).

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

[12] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. "Face recognition and gender determination". In M. Bichsel, editor, *International Workshop on Automatic Face- and Gesture Recognition*, pp. 92–97, Zurich, June 1995. MultiMedia Laboratory, Department of Computer Science, University of Zurich.

[13] A. Yuille, D. Cohen, and P. Hallinan. "Feature extraction from faces using deformable templates". In *IEEE Computer Soc. Conf. on Computer Vision and Patt. Recog.*, pp. 104–109. IEEE Computer Society Press, 1989.