Audio and Video Based Biometric Person Authentication (AVBPA '99)
Washington DC (USA), March 22nd – 24th, 1999

Face Authentication by retinotopic sampling of the Gabor decomposition and Support Vector Machines

F. Smeraldi, N. Capdevielle, J. Bigün
Microprocessor and Interface Laboratory
Swiss Federal Institute of Technology (EPFL)
CH-1015 Lausanne

Abstract

We describe a face authentication algorithm based on retinotopic sampling of the Gabor decomposition. A log–polar sampling grid is used to encode information from relevant regions of the face. Decision is based on a multi expert scheme, in which each expert independently authenticates the client based on a different facial feature. Implementation of the expert as Support Vector Machines is discussed. An alternative implementation using K–Nearest–Neighbours classifiers is presented for comparison.

1 Introduction

As services which require automatic access of eligible persons (clients) to services (privileges) become increasingly common, the importance of face authentication has been growing. Neurological investigations have shown a strong analogy between the responses of the simple cells in the Visual Cortex and a classic and powerful computer vision tool, the Gabor decomposition [13]. Some aspects of the human visual information processing have been captured by the elastic graph matching algorithm described in [8]. On the other hand, perceptive studies have shown that certain parts of the face, the eye region in particular, tend to be more relevant than the others for face characterisation [12]. This observation, combined with the neurophysical and computational studies on attentional mechanisms such as retinotopic sampling and the human saccadic system [18], motivated us to develop an approach based on the retinotopic sampling of the Gabor decomposition of facial images in the eye region. A rigid log–polar sampling grid is used to extract three sets of Gabor features for each client. These are obtained with the retina positioned on the subject's right eye, left eye and on the point midway between them, to which we will refer as "the nose" in the rest of this paper. Three classifiers are then separately trained on the features set, so that a left eye expert, a right eye expert and a nose expert are available which can independently authenticate the person. Such a multiple expert approach has already been shown to be advantageous in [7].

Two choices have been tested for the classifier: a classical K–Nearest–Neighbours solution and Support Vector Machines [17, 4]. The latter have already been applied to the related problem of face detection [15]. Comparison of the results shows that optimal performance is obtained with the use of linear SVM's, which also confirms the discriminating efficiency of the features employed.

2 Log polar sampling and the Gabor decomposition

Our feature extraction step is based on the use of a sparse retinotopic sampling grid. The grid has log–polar geometry, meaning that the density of sampling point decreases exponentially from the centre. This sampling topology automatically implements a "focus of attention" concept, concentrating the computational effort on the current fixation point. During the feature extraction step, the retinotopic grid is placed on the two eyes and the nose of the subject being considered for processing 1. For each of these placements, a feature set is acquired by computing a vector of Gabor filter responses at each point of the grid. The three feature sets so obtained constitute all the information we retain about a person, and are treated separately by the left eye, right eye and nose experts.

The log–polar mapping has also been applied to the design of the Gabor filters in the frequency domain. Standard complex valued Gabor functions in the frequency domain are scaled, translated and shifted versions of the following function:

$$\hat{\mathcal{G}}(\vec{\omega}|\sigma_x, \sigma_y, \omega_0) = \exp\left(-\frac{(\omega_x - \omega_0)^2}{2\sigma_x^2}\right) \cdot \exp\left(-\frac{\omega_y^2}{2\sigma_y^2}\right)$$

Figure 1: The retinotopic sampling grid placed on the right eye of a subject for feature extraction.

The parameters σ_x, σ_y, ω_0 and the rotation parameter are chosen to cover the frequency plane as completely at possible.

However, when only a small number of logarithmically spaced frequency channels is used, problems arise in obtaining a uniform coverage of the frequency plane. Given that the spacing between the centres of the filters increases exponentially, the symmetric Gaussian shape doesn't appear to be optimal, since it extends the same distance towards the (well sampled) central region of the frequency space as well as towards the loosely sampled periphery. For these reasons, we choose to substitute for the commonly used Gabor function a modified filter

$$\hat{\mathcal{G}}'(\vec{\omega}|\sigma_\rho, \sigma_\phi, \rho_0) = \exp\left(-\frac{(\rho-\rho_0)^2}{2\sigma_\rho}\right) \cdot \exp\left(-\frac{\omega_\phi}{2\sigma_\phi^2}\right)$$

where $(\rho, \phi) = (\ln(|\vec{\omega}|), \tan^{-1}(\omega_y/\omega_x))$ is the conformal mapping of the frequency plane to log polar coordinates [1]. These filters have been previously used in texture analysis problems, showing high discrimination power [2]. We therefore construct a uniform grid of Gaussian filters in the log–polar frequency domain, which in turn yields the desired consistent and exponential coverage of the Fourier plane (figure 2).

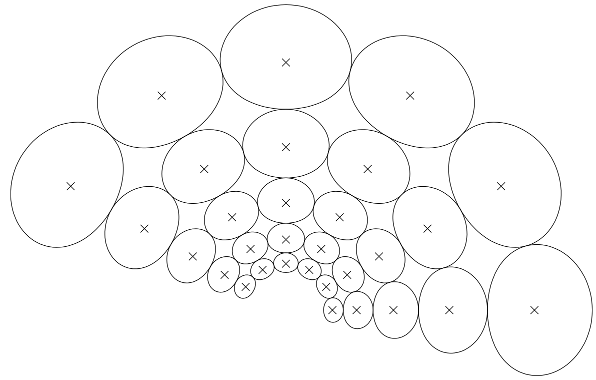

Figure 2: Iso–curves of the Gabor filters created by uniform sampling of the frequency plane in log–polar coordinates. The crosses represent maxima, whose positions are slightly biased towards origin.

3 Expert implementation

Acceptance or rejection of the identity claims is delegated to three independent experts, separately trained on the features extracted from the left eye region, the right eye region and the nose region of each single subject. Each expert is implemented as a classifier. The sampling retina used for feature extraction consists of five concentric rings for a total of 50 points. Gabor responses along six directions and five frequency channels are extracted at each point, so that each expert is fed with 1500-dimensional data.

Experts are trained to discriminate between data from the client and data from 35 training impostors. The database we employed for testing allows using an average training set of 10 images from the client and 350 images of training impostors, for a total of around 360 vectors. Since this number is small when compared with the dimensionality of the feature space, we chose to implement the experts as Support Vector Machine (SVM) classifiers. SVM's are known to assure a good generalisation ability even in the case of small training sets. The SVM engine we used was developed following the ideas in [11] and [14].

3.1 SVM Experts

The theory behind SVM's is rather complex, and we therefore refer the reader to specific literature. Nevertheless, the general working principle of a Support Vector Machine can be intuitively described by a schematic diagram like the one in Figure 3. After an (optional) nonlinear mapping to a higher dimensional space, data are separated by means of the Optimal Separating Hyperplane, which is the unique one having a maximum distance from the training examples of the two classes. The decision function in this higher dimensional space is the sign of

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + \mathbf{b}, \tag{1}$$

where $\mathbf{x}$ is the vector to be classified and b is the constant. The normal vector to the hyperplane, $\mathbf{w}$, is a

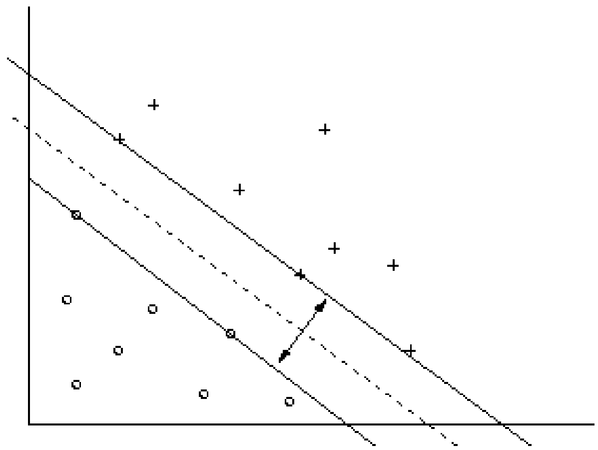

Figure 3: The Optimal Separating Hyperplane (schematic). The distance between the hyperplane and the closest examples from the two classes is maximised.

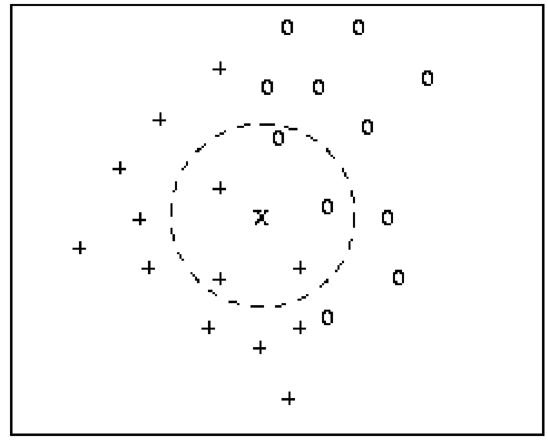

Figure 4: The K–Nearest–Neighbours classifier (schematic, $K = 5$). X is assigned to class $\{+\}$.

linear combination of some of the training examples (or their images after the nonlinear mapping), which are known as Support Vectors. If the norm of $\mathbf{w}$ is chosen so that $f(\mathbf{x}) = \pm\mathbf{1}$ for the support vectors, $|f(\mathbf{x})|$ can be considered a measure of the certainty of the classifier. Given that the three experts are expected to have comparable performance, it therefore makes sense to combine their output by averaging the value of the three classification functions.

3.2 KNN Experts

In order to allow estimating the role played by SVM's in the classification result as compared with the discriminating power of the feature extraction step we also implemented the experts by means of simple K–Nearest–Neighbours (KNN) classifiers [6]. In a KNN classifier, all of the positive and negative example vectors are retained to constitute a description of the two classes. When presented with a vector $\mathbf{x}$ to be classified, the classifier finds the K training examples which happen to be closest to $\mathbf{x}$ according to some metric (which in our case is the Euclidean distance). The vector is then assigned to the class which is mostly represented in the set of K neighbours, see Figure 4. We emphasise that KNN based authentication is different from minimum Euclidean distance based recognition in that the training impostors are considered as a whole to represent the "reject" class. That means that in case the set of K neighbours happens to contain two example vectors belonging to the client and $K-2$ examples each from a different training impostor, the claim is rejected even if the client is more represented than each of the impostors separately taken. Since KNN classifiers give an intrinsically binary output, these outputs are combined by a majority voting scheme.

4 The test procedure

The database used for testing was extracted from the first multi–modal verification oriented database collected by the M2VTS consortium. It consists of four series of images of 37 people taken at different periods of time. Each series contains two to four frontal images from each subject. Appearance of subjects varies widely across the different series.

Tests are performed according with a "leave–one–out" rotation scheme known as "Brussels Protocol" [3, 5, 10]. At each step, one person is removed from the database to act as impostor. Also, one entire image series is set aside for client tests. The 36 persons remaining act as registered clients, and their three image series are available for training. Client tests are performed taking, for each client, an image from the test series; a single image of the impostor is used for impostor tests. This amounts to 36 client tests and 36 impostor tests. By choosing the impostor and the test series in all the possible ways, $36 \times 37 \times 4 = 5328$ essentially different client tests and the same number of impostor tests can be generated.

In the training, each expert (independently of its implementation) uses the three training series of images from the client which it should learn as positive examples, and all of the images from the remaining 35 clients known to the system as negative examples.

SVM Experts	FA	FR
Linear	0.0%	9.2%
Poly - deg 2	0.0%	9.4%
Poly - deg 3	0.0%	9.5%
Poly - deg 4	0.0%	11.6%
RBF	0.0%	14.1%

Figure 5: False Acceptance and False Rejection with SVM experts — linear, polynomial and Radial Basis Function kernels.

KNN Experts	FA	FR
K=1	1.4%	2.0%
K=3	1.0%	5.4%
K=5	0.5%	10.7%

Figure 6: False Acceptance and False Rejection with KNN experts.

5 Experimental results

We performed our tests by manually supplying the eye coordinates to the system for all of the images. The results we obtained indicate that a very low error rate can be achieved by means of both SVM based and KNN based experts. Results are summarised in Figures 5 and 6. As can be observed, the best KNN results are obtained in the Nearest Neighbour case ($K = 1$). This can be ascribed to the relatively low density of client examples in the feature space. Due to the wider variability present in the training impostors, as compared with the intra–class client variation, SVM experts display a consistent tendency towards false rejection. This can be compensated for by modifying the SVM expert fusion decision function

$$\sum_{c=1}^{3} f_c(\mathbf{x}) > \mathbf{0} \tag{2}$$

with the introduction of a variable threshold τ:

$$\sum_{c=1}^{3} f_c(\mathbf{x}) > \tau \tag{3}$$

In this way, False Acceptance (FA) and False Rejection (FR) curves can be obtained (Figure 8), and system performance can be assessed in terms of Equal Error Rate (EER — Figure 7). The performance obtained with SVM experts is, as expected, slightly superior to the results afforded by the KNN technique. Although the system profits from the optimal margin separation assured by SVM's, the dimensionality boost obtained by resorting to nonlinear kernels does not appear to improve classification results any further. This fact, together with the comparable performance of KNN experts, certifies the good discriminating power of the features employed.

SVM Experts	EER
Linear	1.4%
Poly - deg 2	1.4%
Poly - deg 3	1.4%
Poly - deg 4	1.4%
RBF	1.4%

Figure 7: Equal Error Rates with SVM experts and different types of kernel.

As a final remark, since the decision function 1 applies directly in the feature space in the case of linear Support Vector Machines, SVM experts have the advantage that training data are condensed into a single vector $\mathbf{w}$. This is an advantage (over SVM's with non-linear mappings) in case a large client database must be maintained or biometric data must be stored on a smartcard.

6 Conclusions

We have described a face authentication system based on log–polar sampling of the Gabor decomposition in the eye and nose regions. Experimental results show that very low error rates (1.4% EER) can be achieved by splitting the decision among multiple experts, each of which is trained on a different facial feature. Comparison of results from different classifier types as well as previous results ($\sim$ 6.5% EER) using the "Brussels Protocol" on the M2VTS database [9] evidence the discriminating power of the features themselves. The choice of a retinotopic sampling grid is also motivated by the possibility of performing attention–driven face and eye detection, as described in [16]. In the future, we plan to integrate this facial feature detection step into the identity verification system.

Acknowledgement

This work has been supported by the VIRSBS project within the European IT-LTR programme.

References

[1] J. Bigun. Speed, frequency, and orientation tuned 3-d gabor filter banks and their design. In *Proceedings of International Conference on Pattern Recognition, ICPR, Jerusalem*, pages C–184–187. IEEE Computer Society, 1994.

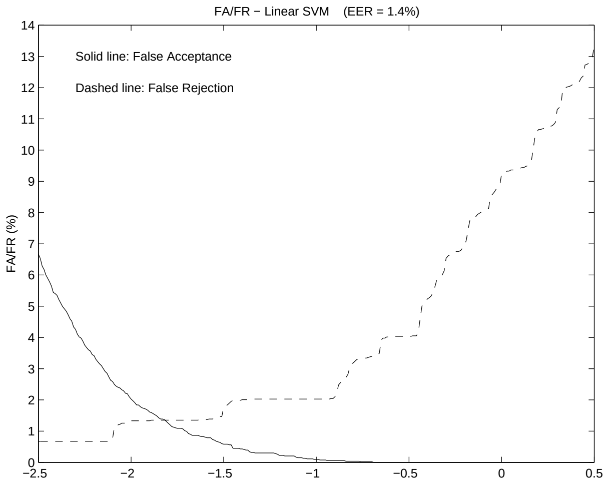

Figure 8: False Acceptance and False Rejection curves for the SVM experts, linear kernel.

[2] J. Bigun and J. M. H. du Buf. N-folded symmetries by complex moments in Gabor space. *IEEE-PAMI*, 16(1):80–87, 1994.

[3] J. Bigun, B. Duc, S. Fischer, A. Makarov, and F. Smeraldi. Multi modal person authentication. In H. W. et. al., editor, *Nato-Asi advanced study on face recogniton*, volume LNCS. Springer, Scheduled Feb. 98 1997.

[4] C. Cortes and V. Vapnik. Support–Vector Networks. *Machine Learning*, 20:273–297, 1995.

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

[6] L. Devroye, L. Györfi, and G. Lugosi. *A probabilistic theory of pattern recognition.* Number 31 in Applications of mathematics. Springer–Verlag, 1996.

[7] B. Duc, E. S. Bigun, J. Bigun, G. Maitre, and S. Fischer. Fusion of audio and video information for multi modal person authentication. *Pattern Recognition Letters*, 18:835–843, 1997.

[8] B. Duc, S. Fischer, and J. Bigün. Face authentication with gabor information on deformable graphs. *Accepted for publication in* IEEE Transactions on Image Processing, 1997.

[9] B. Duc, S. Fischer, and J. Bigun. Face authentication with sparse grid gabor information. In *IEEE Proc. of ICASSP, Munich*, volume 4, pages 3053–3056, 1997.

[10] B. Efron and R. J. Tibshirani. *An introduction to the Bootstrap.* Chapman & Hall, New York, 1993.

[11] T. Joachims. *Making Large-scale SVM Learning Practical*, chapter 11 of Advances in Kernel Methods - Support Vector Learning. MIT Press, 1998. Eds. B. Schölkopf, C. J. C. Burges, A. J. Smola.

[12] C. F. Keating and E. G. Keating. Monkeys and mug shots: cues used by Rhesus monkeys (Macaca mulatta) to recognize a human face. *Journal of Comparative Psychology*, 107(2):131–139, 1993.

[13] G. A. Orban. *Neuronal operations in the visual cortex.* studies of brain functions. Springer, 1984.

[14] E. Osuna, R. Freund, and F. Girosi. Improved training algorithm for Support Vector Machines. In *Proceedings of IEEE NNSP'97*, pages 276–285, September 1997.

[15] E. Osuna, R. Freund, and F. Girosi. Training Support Vector Machines: an application to face detection. In *Proceedings of CVPR '97*, June 17–19 1997.

[16] F. Smeraldi and J. Bigun. Facial features detection by saccadic exploration of the Gabor decomposition. In *International Conference on Image Processing, ICIP-98, Chicago, October 4-7*, 1998.

[17] V. N. Vapnik. *The nature of statistical learning theory.* Springer–Verlag, 1995.

[18] A. L. Yarbus. *Eye movements.* Plenum, New York, 1967.