Access control: adaptation and real-time implantation of a face recognition method

J. Mitéran
J. P. Zimmer
F. Yang
M. Paindavoine
University of Burgundy
Laboratory Le2i
Aile des Sciences de l'Ingénieur
BP 47870
21078 Dijon Cedex, France
E-mail: miteranj@u-bourgogne.fr

Abstract. An improvement of a face recognition method is proposed. The goal is to develop easy-to-use access control software, allowing personal-computer or building access control with minimal constraints for the users. This approach requires a high-speed classification method (about 8 images/s) and a high global recognition rate. We obtain good results using a method derived from principal-component analysis, a geometric transformation of the feature space, and a fast decision algorithm.
[DOI: 10.1117/1.1355256]

Subject terms: face recognition; image processing; access control.

Paper 990393 received Oct. 7, 1999; revised manuscript received Nov. 1, 2000; accepted for publication Nov. 8, 2000.

1 Introduction

To improve security by controlling the access to personal computers (PCs) or buildings (industrial, public, or private), we have adapted to public use and implemented in real time a robust and fast face recognition method. The aim of this paper is to present the adaptations needed for public use of this kind of control, and a performance evaluation of the modified method.

Numerous algorithms have been proposed for face recognition in the literature. Some of the most common methods are based on direct images of faces, known as eigenfaces,[1–3] on profiles,[4] on geometry,[5,6] or on 3D data acquisition of the surface of the face.[7]

For access control of a PC, it is necessary to have a low-cost and fast system, requiring very few adjustments. The system must also be easy to use and place a minimum of constraints on the users. This excludes all active methods, based on the projection of a pattern, for example.[8]

In view of its performance in recognition and computational time, we have chosen in our laboratory to implement the method based on *eigenfaces*, which uses principal-component analysis (PCA) for dimensionality reduction. For the classification and recognition step, we based our implementation on a neural network. A fast implementation of this method is described in detail by Yang et al.[9,16] and by J. P. Zimmer et al.[10] The system is composed of a standard video conferencing camera, connected to a PC via either the parallel port or a PCI acquisition board. We present in this paper an adaptation of this method to systems where the computational time is critical, and where the user can enter his name or a code before proceeding to the recognition test. That means that we can just perform a verification of the identity. In terms of performance, this improvement was necessary for our applications (security access), where a very low rate of confusion is needed, but where the success identification rate is not so critical.

In Sec. 2 we recall the theoretical basis of the eigenface method. In Sec. 3 we present the problems caused by the learning and decision protocol, and the solution we have chosen. In Sec. 4 we present the results of the implementation.

2 Review of the Eigenvector-Based Method for Face Recognition

This general face recognition method, implemented by Yang,[16] is based on the PCA and the perceptron. It is composed of two phases: the learning phase, during which approximately ten images of each user must be stored in the database, and the decision phase, during which a new image must be acquired and classified as unknown or as a match for an existing user.

2.1 *Learning Phase*

During the first step, we create the face image database used to compute the input stimuli of the learning network. We based our choice on the method described by Abdi,[11,12] and we decompose each image (width L and height l) into a singular-value vector.

The advantage of this method is in presenting small-sized vectors to the first layer of the network. For that, initially, we build a human base matrix $\mathbf{H}_{[l,L,n]}$, in which np face images are assembled in a row. $\mathbf{H}$ is constructed with all the images of n human faces to be recognized (p images per person). This matrix can be decomposed in singular values as follows:

$$\mathbf{H}=\mathbf{P}\Lambda\mathbf{Q}^T, \tag{1}$$

where

Λ = matrix of eigenvalues of $\mathbf{HH}^T$ and $\mathbf{H}^T\mathbf{H}$
$\mathbf{P}$ = matrix of eigenvectors of $\mathbf{HH}^T$
$\mathbf{Q}$ = matrix of eigenvectors of $\mathbf{H}^T\mathbf{H}$.

We denote by $\boldsymbol{\Delta}=\boldsymbol{\Lambda}^{1/2}$ the diagonal matrix of singular values.

 0091-3286/2001/$15.00

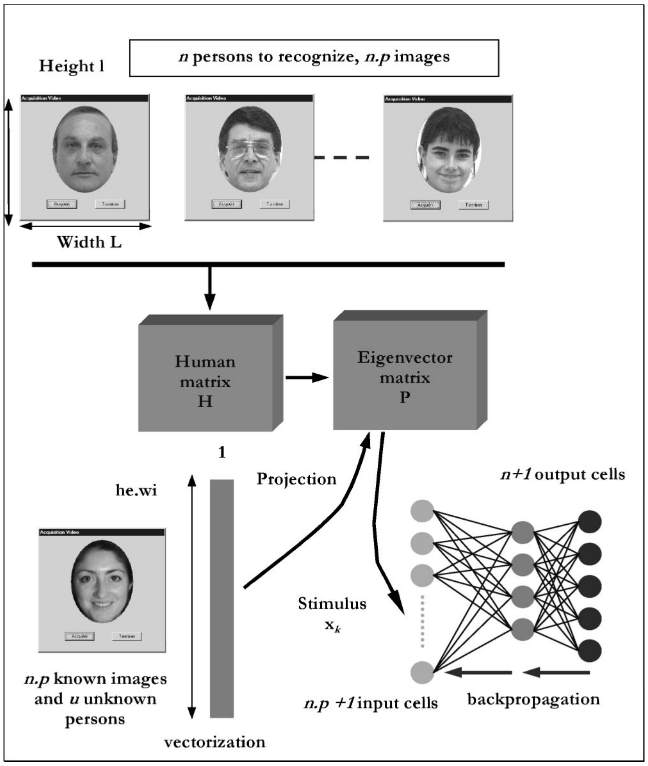

Fig. 1 Learning phase in case of recognition.

Determining $\mathbf{P}$ directly by computing $\mathbf{HH}^T$, of dimension $[lL,lL]$, requires an inordinate amount of memory. Therefore, we first determine the matrix $\mathbf{Q}$. Secondly, we compute $\mathbf{P}$ with

$$\mathbf{P}=\mathbf{HQ}\mathbf{\Delta}^{-1}. \tag{2}$$

Each stimulus $\mathbf{x}_k$ presented to the network (Fig. 1) is the result of the projection of the k'th face image vector onto the matrix $\mathbf{P}$ ($k=1,\ldots,np+u$, where np is the image number of $\mathbf{H}$, and u is the image number of unknown persons).

The second step is the determination of the neural network parameters. The learning structure we use is a network with one hidden layer, as in Solheim et al.[13] and Frasconi et al.[14] Therefore, for the k'th stimulus $\mathbf{x}_k$, we define, $\mathbf{h}_k$ as the output value of the hidden layer, $\mathbf{o}_k$ as the value of the output layer of the network, and $\mathbf{r}_k$ as the theoretical value of the output layer of the network, assuming that the membership of $\mathbf{x}_k$ in a certain class is known.

If $\mathbf{W}$ is the matrix of the weight connections from the input layer to the hidden layer, and $\mathbf{Z}$ is the matrix of weight connections from the hidden layer to the output layer, the cells of the hidden layer compute their activation and convert it into a response by using their transfer function, according to

$$\mathbf{h}_k=f(\mathbf{W}\mathbf{x}_k). \tag{3}$$

Then, the activation of the output layer is computed using $\mathbf{Z}$ and the transfer function f included in each cell according to

$$\mathbf{o}_k=f(\mathbf{Z}\mathbf{h}_k) \tag{4}$$

Here f is a nonlinear differentiable function. We use the common logistic function.[12]

We have chosen the known backpropagation algorithm in order to compute the learning parameters. This method is used to modify, in a iterative way, the weights of the connections $\mathbf{W}$ and $\mathbf{Z}$ in order to reduce the final error value of classification. The error $\mathbf{e}_k$ obtained at the output layer is estimated by comparing the results given by the $\mathbf{o}_k$ cells with the theoretical response $\mathbf{r}_k$ of this layer:

$$\mathbf{e}_k=\mathbf{r}_k-\mathbf{o}_k. \tag{5}$$

The output error signal $\delta_{s,k}$, taking into account the error due to the cell and its activation state, is given by

$$\delta_{s,k}=f'(\mathbf{Z}\mathbf{h}_k)*\mathbf{e}_k, \tag{6}$$

where $*$ is the Hadamard product. The connection matrix $\mathbf{Z}$ is corrected at each iteration and becomes for the t'th iteration

$$\mathbf{Z}_{t+1}=\mathbf{Z}_t+\eta\delta_{s,k}\mathbf{h}_k^T. \tag{7}$$

The error then propagates to the hidden layer, and we estimate the error signal $\delta_{h,k}$ for the cells of this layer by

$$\delta_{h,k}=f'(\mathbf{W}\mathbf{x}_k)*(\mathbf{Z}_t^T\delta_{s,k}). \tag{8}$$

The connection matrix $\mathbf{W}$ is corrected for iteration $t+1$ according to

$$\mathbf{W}_{t+1}=\mathbf{W}_t+\eta\delta_{h,k}\mathbf{X}_k^T. \tag{9}$$

Then the algorithm iterates until we get the desired minimum error (a threshold is fixed by the user).

2.2 *Decision Phase*

During this step, an image of the face is acquired. Then, the image is vectored and projected onto the eigenvector matrix $\mathbf{P}$. The projection result is used as a stimulus to the network, which gives a response (set of $n+1$ values of the output cells.) Finally, a decision threshold is used in order to eliminate the weakest values.

2.3 *Performance of the Method*

The performance of the eigenface method is widely described in the literature (Ref. 15, for example). In order to estimate our own implementation, including the acquisition system and the learning protocol, we studied the performance using a database of 250 images.[17] This database was divided into subsets D1, D2, and D3:

- D1 contains the images of 10 persons to recognize (10 images per person) and was used to compute the eigenvectors (Fig. 2).
- D2 contains the same images as D1, and 5 unknown persons (2 images per person). These images allow us to compute the neural network parameters (Fig. 3).

Fig. 2 Extract of D1.

- D3 contains 140 images, with 10 images per known person (10 persons), 4 images per unknown person, and 4 images per new person (5 persons) never learned by the system. (Fig. 4).

The best global error rate we obtained concerning confusion (known or unknown persons classified as other person) is 2%, but if we take account of the proportion of unknown persons in the D3 base, 10% of unknown persons are classified as known persons. This performance was not acceptable for access control. We must then adapt the method.

3 Adaptation to Face Verification

3.1 *Data Organization*

In the case of access control of a PC, the user must enter his name and password before access is granted. Since the system knows the name of the individual, we just need to verify his identity, not to distinguish him from all of the other users. Therefore, we can separate all of the persons to be identified, using one set of data (**H**,**P**,**W**,**Z**) per person (Fig. 5).

This allows the decision phase to be very fast. In the case of access control of buildings, the person must push a button corresponding to his name. This choice allows an unlimited number of persons to be recognized. The practical limit is the disk space used to store the learning images and matrices **H**,**P**,**W**,**Z** (about 1 Mbyte per person).

3.2 *Definition of the Acquisition Protocol*

In the first version of the face recognition software, the decision protocol was defined as follows. The user, during the recognition phase, must hit a key of the keyboard to grab his face after positioning. The person was accepted or not, using only this one image. This protocol required a high success rate during the recognition phase. Moreover, it was not practical to use.

We chose the following protocol:

1. The user enters his name.
2. The set of data corresponding to the name is loaded in memory, and the test number T is set to 0.
3. One image is acquired, and the verification test is performed. T is incremented.
4. If the test succeeds, the user is authorized and the protocol stop. If the test fails, the process loops to the step 3, while $T<100$.
5. If the user is still unrecognized after $T=100$ tests, the protocol stops and the user is rejected.

This is possible because approximately eight images per

Fig. 3 Extract of D2.

Fig. 4 Extract of D3.

second, in terms of recognition, can be handled by the system (these results were obtained using a Pentium II 233-MHz processor, with 64 MB of memory, and a standard PCI acquisition board based on the BT848 chip). This allows the system to discard misaligned images but also increases the probability of encountering a classification errors: a *confusion*, which occurs when an unknown user is identified as a known user.

We modified the method by adding these new features. We thereby reduced the confusion rate, but also increased the per-image nondetection rate. However, the large number of images processed rendered this increase in the nondetection rate insignificant.

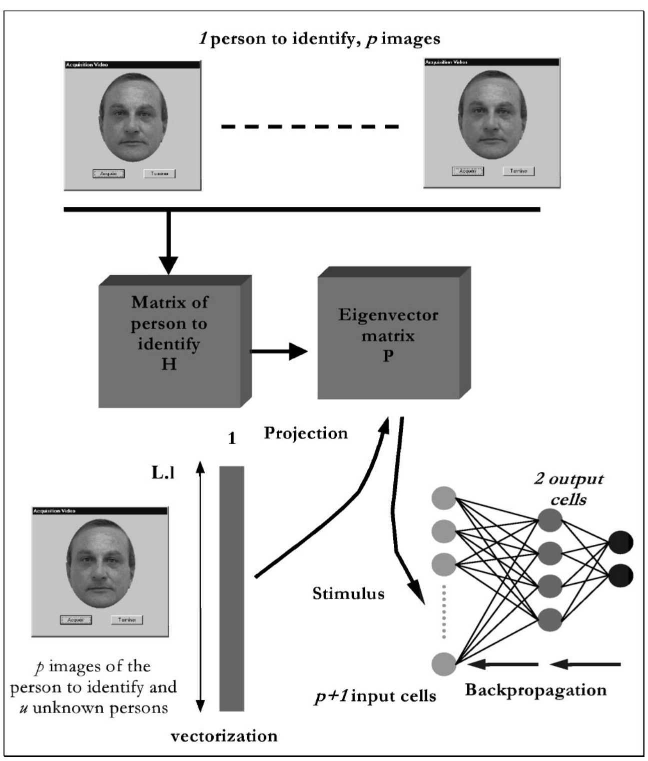

Fig. 5 Learning phase for one person in verification.

3.2.1 *Diminution of confusion rate*

For a better understanding of the problems that the system must overcome, it is necessary to recall that the first eigenvector represents the low frequencies in the image, and the last eigenvectors the high frequencies. These high frequencies contain the most significant visual features of an individual (Fig. 6).

Here, the projections are computed using the luminance of the images, causing two problems. Firstly, the image acquisition quality depends on external lighting conditions, which is not crucial because the camera has good automatic gain control. Secondly, if someone in the database has a particular feature (a beard, for example), where the luminance takes a dominant place, the neural network will converge regarding mainly this first eigenvector, and thus regarding low frequencies only. This implies that his features are very different from those of the unknown persons in the learning data set, and if the system attempts to recognize another bearded person, there will be some confusion (Fig. 7).

To solve this problem, we studied two methods. The first method, inspired by Belhumeur et al.[15] consisted in the elimination the first three eigenvectors to perform the classification. The second method, inspired by Yang,[16] was to use edge detection (a low-pass filter followed by a Sobel filter) instead of using direct luminance (Fig. 8).

The two methods gave similar results regarding confusions, in that they eliminate all confusions in our test. For practical reasons, we chose to implement the second method, which avoids loss of information that may be useful for discrimination. This choice allows the method to be

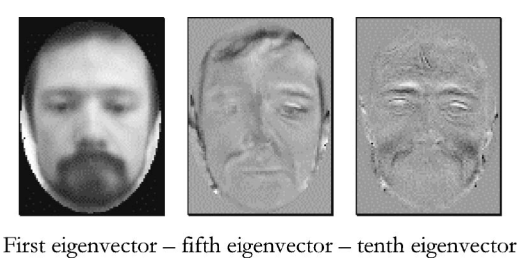

Fig. 6 Eigenvector representation.

Fig. 7 Example of confusion.

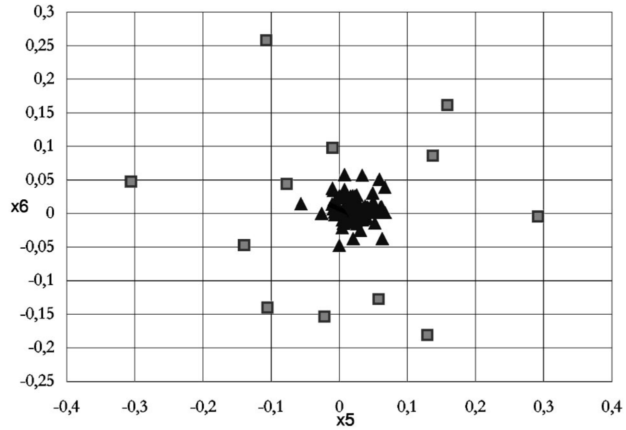

Fig. 9 Projection of fifth and sixth eigenvectors.

globally more robust (the rate of false alarms is reduced), even if the recognition rate per individual image is less than when using direct luminance. We do not give here any rate of success, because the system must be evaluated globally, including the dynamic aspect of the acquisition and classification (see the results given in Sec. 4).

3.3 *Problems of Convergence*

Since we eliminate low frequencies, the method is now more discriminatory, using the high frequencies of the images, which represent the identity of the person. However, the main problem of neural networks, well known in the literature, is that sometimes, depending on the person to be recognized, the backpropagation algorithm does not converge. The main reason is that the high-level eigenvector distributions are not easily separable in the feature space. To solve this problem, it is known that we can improve the neural network, choosing more hidden layers, for example, but this solution was not chosen, mainly for computing-time reasons. Moreover, since the distribution has a particular symmetry in the feature space, it is simplest to compute a transformation using this symmetry.

We have represented in Fig. 9 an example of projections of the learning images on the fifth and sixth eigenvectors (high-level eigenvectors). The known user is represented by squares, and unknown persons by triangles. One can see that the samples corresponding to the known user are randomly distributed around the samples corresponding to unknown persons.

After investigating the geometrical shape of the distributions, we chose to keep the first three projections without any modification, and the fourth feature of the classification method is computed as

$$x_3' = \left(\sum_{i=1}^{np} x_i^2 \right)^{1/2},$$

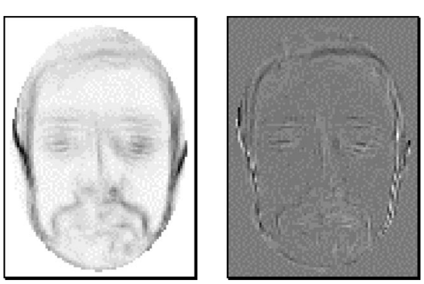

Fig. 8 First and tenth eigenvectors after edge detection.

where x_k is the projection of the image on the k'th eigenvector. This method allows us to obtain the same results as for a circular threshold around the distribution of unknown persons, but is better suited to the classification method.

We present in Figs. 10 and 11 an example of the distribution obtained with one person to be recognized. The figures show that the two classes are now very easily separable. Moreover, the decision phase is now faster, since the number of cells in the first layer of the neural network no longer depends on the number of learning images of one person.

4 Results of the Improved Method

Finally, we have tested our software in the case of face verification. The first test was made with fixed lighting. The second test was performed to illustrate robustness against lighting variations.

4.1 *Standard Test*

In order to evaluate the performance of the system, we defined an acquisition protocol, which reflects the real use of the software. All the measurements were made in real

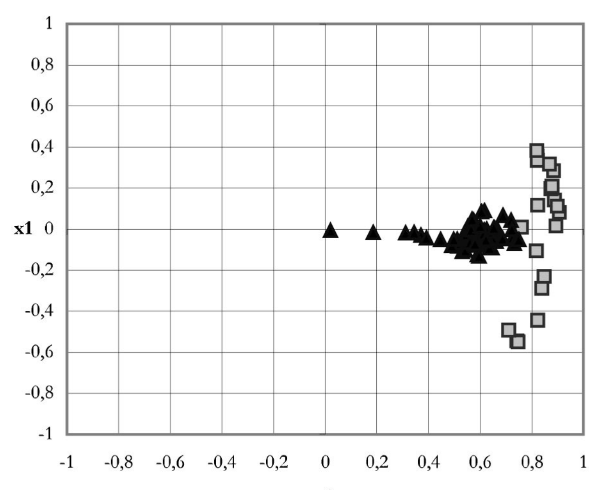

Fig. 10 Projection of first and second eigenvectors.

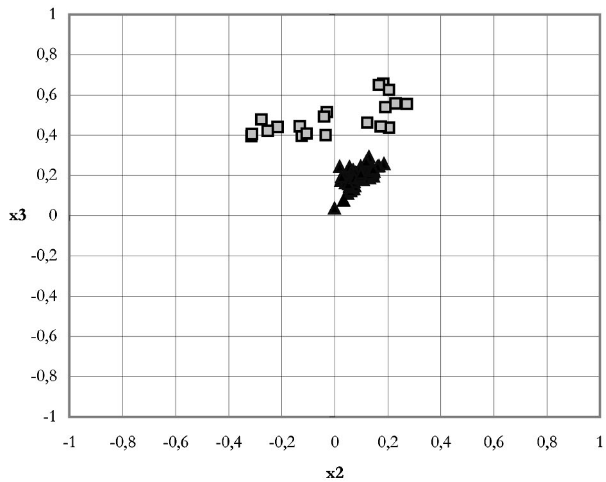

Fig. 11 Projection of third eigenvector and x_3'.

conditions: the user himself acquires the images of the learning phase, and performs the test by presenting his face in front of the camera.

For practical reasons, we divided a set of 30 persons to be recognized in six subsets of five persons each. The steps of the learning and recognition protocol are as follows:

1. Record five images of each person in the subset.
2. Record five new images of each person. The separation between the first and the second step allows to take into account positional variations.
3. Execute the learning phase (compute **H**,**P**,**W**,**Z** for each person).
4. Perform three tests for each person (verified on 100 images), so that he knows how to present himself in front of the camera in order to be well recognized. Add a few images if necessary. The maximum number of learning images is fixed to 15.
5. Perform for each person alternatively five identification tests entering his real name, and five tests of intrusion using a false name.

The results of the test are reported in Table 1. One can see that we obtain a 90% success rate in terms of verification. It is important to note that 100% of users were well identified at least one time, and that the success rate is better for experienced users (96%). The most important result is that we obtained only two confusions. After analysis, the origin of these confusions was seen to be in the bad positioning of the person during the learning phase. One can note that the quality of the learning phase, as in all classification problems, determines the performance of the system. If someone moves in front of the camera during the acquisition phase, the registered image will contain only a part of his face, and the risk of intrusion for this individual increases. The user must verify each learning image before accepting or rejecting it. After removing the bad images, the two confusions of the test disappeared. This fact is usually not taken into account in the literature, where authors often test their face recognition using a fixed database of pictures. The originality of this protocol is to characterize the whole system, including face positioning, during image acquisition and final decision.

Table 1 Results of the test.

Number of verifications	150
Failed	15
Succeeded	135
Number of confusion tests	150
Succeeded (person not authorized)	148
Failed (person authorized)	2

Fig. 12 Learning with homogeneous lighting.

4.2 *Lighting-Variation Test*

To illustrate the robustness of the method, we performed the test using two different learning sets. The first learning set is built using only one standard homogeneous lighting (Fig. 12). The second learning set is built using four different lightings: homogeneous, left illumination, right illumination, and global darkening (Fig. 13). We have chosen to test only realistic variations such as one can find in an office, for example.

For both learning sets, we tested the identification with homogeneous lighting and with lighting variations. To illustrate the improvement of performance using edge detection instead of direct luminance, we also tested the identification using both algorithms. The results (performed with five experimental users and five tests of each lighting for each user) are reported in Figs. 14 and 15.

One can see that the method using edge detection is more robust than using direct luminance, but it is necessary to take into account important lighting variations during learning phase to obtain the best results.

We performed also some qualitative tests showing that the result is not influenced by glasses, and little influence by hair style if the image registration is made carefully[9] (the hair is not taken into account, thanks to the elliptic mask). To avoid too great sensitivity to the slow variations of a beard, we added the possibility of automatically intro-

Fig. 13 Learning with lighting variations.

Result using direct luminance		Example of images
Learning with homogenous lighting	Success rate **23%**	
	Error rate **77%**	
Learning with lighting variations	Success rate **53%**	
	Error rate **47%**	

Fig. 14 Results using luminance.

ducing the image that allowed the last recognition into the learning base. The latter thus is automatically updated. However, the method remains sensitive to abrupt variations of beard.

5 Conclusions

We have implemented a face verification program, used for PC or building access control, improving the feature computation and separability of the learning sets in the feature space.

With regard to performance, we showed that each person can be easily recognized and that the software can be used in a real environment.

We have also shown that the measurement of the recognition rate and the adjustment of the classification method must be adapted to the acquisition protocol to be significant.

Our goal is to eliminate the positional dependence of the user in front of the camera, implementing automatic detection and adjustment of the face position. We are currently developing a mixed method, based on motion detection and the Hough[17,18] transform.

Result using edge detection		Example of images
Learning with homogenous lighting	Success rate **56%**	
	Error rate **44%**	
Learning with lighting variations	Success rate **96%**	
	Error rate **4%**	

Fig. 15 Results using edge detection.

Finally, the improved method can be used in several areas, protecting PC or building access, combining statistical and neural approaches.

Acknowledgments

Thanks to H. Abdi for his advice on neural networks, and to Alan Mangan and Lew Lew Yan Voon for their help with the English translations.

References

1. L. Sirovitch and M. Kirby, ''Low-dimensional procedure for the characterization of human faces,'' *J. Opt. Soc. Am.* **2**, 519–526 (1987).
2. M. Turk and A. Pentland, ''Eigenfaces for recognition,'' *J. Cogn Neurosci.* **3**(1), 71–86 (1991).
3. A. Pentland, B. Moghaddam, O. SDtarner, T. Oliyide, and M. Turk, ''View based and modular eigenspaces for face recognition,'' *Comput. Vision Patter. Recogn.*, 84–91 (1994).
4. L. Najman, R. Vaillant, and E. Pernot, ''Face sideviews to identification,'' G. Vemazza, Ed. in *Image Processing: Theory and Applications,* San Remo, Italy (1993).
5. R. Brunelli and T. Poggio, ''Face recognition: features versus template,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, 1042–1052 (1992).
6. A. L. Yuille, D. S. Cohen, and P. W. Hallinan, ''Feature extraction from face using deformable templates,'' *Int. J. Comput. Vis.* **8**(2), 99–111 (1992).
7. J. Y. Cartoux, ''Formes dans les images de profondeur. Application à la reconnaissance et à l'authentification de visages,'' Ph.D. Thesis, Univ. Blaise Pascal de Clermont-Ferrand (1989).
8. R. Vaillant and I. Surin, ''Face reconstruction through active sterovision,'' *Traitement Signal* **12**(2), 201–210 (1995).
9. F. Yang, E. Drege, M. Paindavoine, and H. Abdi, ''Parallel implementation of a face location algorithm based on the Hough transformation,'' in *EUSIPCO-Greece,* pp. 495–498 (1998).
10. J. P. Zimmer, J. Mitéran, F. Yang, and M. Paindavoine, ''Security software using neural networks,'' in *IECON 98,* pp. 72–74, Aachen, Germany (1998).
11. D. Valentin, H. Abdi, and A. J. O'Toole, ''Categorization and identification of human face images by neural networks: a review of the linear autoassociative and principal component approaches,'' *J. Biol. Systems* **2**, 413–429 (1994).
12. H. Abdi, *Les Réseaux de Neurones,* Presses Universitaires de Grenoble, France (1994).
13. I. Solheim, T. L. Payne, and R. Castain, ''The potential in using backpropagation neural networks for facial verification,'' *Simulation* **58**(5), 306–310 (1992).
14. P. Frasconi, M. Gori, ''Face recognition using multi-layered perceptron which also reject never seen faces,'' in *Proc. 4th Italian Workshop on Neural Networks and Parallel Architectures,* pp. 58–66, Vietry, Italy (1992).
15. P. Belhumeur, J. Hespanha, and D. Kriegman, ''Eigenfaces vs. fisherfaces: recognition using class specific linear projection,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 711–720 (1997).
16. F. Yang, ''Traitement automatique d'images de visages. Algorithme et architecture,'' Thèse, Univ. de Bourgogne (1998).
17. P. V. C. Hough, ''Methods and means for recognizing complex patterns'' U.S. Patent No. 3,069,654 (1962).
18. Y. Lei and K. C. Wong, ''Ellipse detection based on symmetry,'' *Patt. Recogn. Lett.* **20**, 41–47 (1999).

J. Mitéran received the PhD degree in image processing from the University of Burgundy, Dijon, France, in 1995. Since 1996, he has been a ''Maitre de conferences'' at the University of Burgundy, in the Laboratory Le2i (Laboratoire Electronique, Informatique et Image). He is now engaged in research on performance evaluations of the classification algorithms, face recognition and access control problems, and real time implantation of these algorithms on software and hardware architectures.

J. P. Zimmer received MSc degree in computing and instrumentation for images from Université de Bourgogne, Dijon, France in 1997. Since 1997, he prepared a thesis in the Le2i Laboratory where he worked as electronic technician. He currently works as computing engineering in the ''Rectorat de Dijon.'' His current research interests are image processing, stereovision and face recognition.

F. Yang received a BS degree in electrical engineering from the University of Lanzhou (China) in 1982. She was a scientific assistant at the Department of Electronics in the University of Lanzhou. She received an MS (DEA) degree in computer science and a PhD degree in image processing from the University of Burgundy, France, in 1994 and in 1998, respectively. Her research interests are in the areas of patterns recognition, neural network, motion estimation based on spatio-temporal Gabor filters, parallelism and real-time implementation, and more specifically, automatic face image processing: algorithms and architectures.

M. Paindavoine received the PhD in electronics and signal processing from the Montpellier University, France, in 1982. He was with Fairchild CCD Company for two years as an engineer specializing in CCD sensors. He joined the Burgundy University in 1985 as ''Maitre de Conferences'' and is currently a full professor and member of LE2I, the Laboratory of Electronic, Computing and Imaging Sciences, Burgundy University, France. His main research interests are image acquisition and real time image processing. He is also a member of ISIS (a research group in signal and image processing of the French National Scientific Research Committee).