The VIRSBS Project:

Visual Intelligent Recognition for Secure Banking Services

M. Tistarelli[1], E. Grosso[1]
J. Bigun[2], C. Sacerdoti[3], J. Santos-Victor[4], D. Vernon[5]

[1] DIST, University of Genoa, Via Opera Pia 13 - 16145 Genoa, Italy
[2] École Polytechnique Fédérale de Lausanne - Switzerland
[3] Logitron S.r.l. - Italy
[4] Instituto Superior Técnico de Lisboa - Portugal
[5] Maynooth College - Ireland

Abstract. Secure access control is a key issue in banking services. Magnetic cards and personal identification numbers, currently adopted for accessing automatic tellers, do not provide a sufficient degree of security and are likely a source for unauthorized operations. As far as the access to restricted areas is concerned, it usually requires direct surveillance by guards or indirect surveillance by a human operator through a monitoring system. It is often difficult, due to fatigue or other distracting factors, to guarantee continuous and high performance in this task.

The VIRSBS Reactive LTR project faces the problem of the automatic detection of person's identity by using advanced computer vision techniques. The goal of the VIRSBS project is to realize a prototype autonomous station for personal identification. This station will include all the features required to be integrated into a new generation of automated security check-point along corridors, passageways or access doors, and in the next-generation of automatic teller machines. This prototype will be used to perform a significant set of statistical tests on personal identification.

1 Introduction

Banking services have been characterized in the recent years by a considerable diffusion of automatic tellers, points of sale and other widespread home or virtual facilities. Considering the technological advances in telecommunications, this trend will continue in the next years.

Security is a key problem in automatically handled and remote transaction services. Available statistics demonstrate that the use of cards and personal identification numbers cannot guarantee a sufficient degree of security. Moreover they are likely a source for unauthorized operations, including swindling and illicit withdrawals.

The VIRSBS project is focused on banking services, in particular for secure and safe control of access to key areas in the bank building and for cross-checking the identity of persons requesting banking transactions. A system based on visual recognition will have a major impact on man machine interaction, providing a more natural way for the customer to interact with the banking security system.

2 Innovation and Technological Developments

Many different techniques have been proposed for personal identification. They span from handwriting interpretation to fingerprint recognition and biological motion analysis. While most of these techniques are based on vision or image processing there has been no or little effort in intelligent visual recognition of human persons. Nevertheless, visual recognition has indoubted advantages for real applications:

- it is extremely safe, due to the passive nature of video cameras;
- it is simple from a logistical point of view; compact cameras can be installed anywhere and protected from misuse and treacherous actions;
- it does not require cooperation from the subject; this feature is particularly important for handicapped and elderly people;
- it is flexible for implementation in different application scenarios.

The interest in this topic is evident from specialized literature. Recently, face recognition is also attracting the attention of private research centers and government institutions (NIST, FBI, Interpol, National Police Departments).

So far the approaches proposed for visual recognition are based on pattern matching or extraction and comparison of characteristic features (eyes, nose, ears and mouth). The major problem of this approach is the low resolution of acquired visual information, mainly related to the impossibility to actively direct the camera toward the examined person. The steady fall in the cost of video equipment and the availability of low cost real-time systems can lead in a short time to a breakthrough in current technology. Active vision system will provide much more accurate information and will allow active, real-time exploration of persons and faces.

3 Approach

In order to reach the objectives of the project, newly-developed advanced techniques, based on computer vision and robotics, will be exploited. Iconic and feature-based techniques will be utilized in the first instance and in case of ambiguities, performances will be improved by using stereo analysis and the general theory of projective invariants. A breakthrough with respect to current technology will be given by the use of space-variant image representation and by the realization of an active robotic system, able to fixate and track the examined subject.

Two main problems are addressed in the research workpakages. The former concerns the reliable location of head and facial features for any arbitrary (frontal) perspective, which is a prerequisite for any visual face recognition. Previous experience of one of the partners has demonstrated that a variant of the generalized Hough transform can provide a very robust person-independent mechanism for locating the head.

The latter concerns the visual recognition problem, that is the identification of

individuals given a database of known faces. Considering the rich representation required to characterize a human face, iconic data seems naturally best suited to convey all the needed information. On the contrary, image features of a face, like the edges of the eyes, eyebrows or lips are not always stable slightly varying the illumination or the face expression.
Some key aspects of the proposed approach are detailed in the following.

3.1 Foveal image representation

The topology of the human retina allows to coarsely divide visual processing into foveal and peripheral processing. This characterization is very useful to tailor the scale space analysis of image data to that which is best suited to the given visual process. Several space-variant image representations have been proposed like the log-polar transform, the Radial Wedge Transform and the Gaussian pyramid. The log-polar and the Gaussian pyramid space-variant representations are applied to improve existing algorithms for face detection and verification.

3.2 Multi-resolution multi-orientation analysis

The multiresolution nature of the foveal image representation allows for easy implementations of multi orientation analysis techniques. The documented biological vision mechanisms, as well as computer vision research learns that a very precious set of features for many visual tasks are those which can exhibit sensitivity to the presence of lines or edges with particular orientations. Multi-resolution multi-orientation decompositions, as well as features based on these, are expected to help in the tasks of recognition and tracking.

3.3 Active systems

The detection and recognition of individuals requires the capability to analyze the scene and collect images at variable resolution. This process is best performed with a motorized camera which allows to direct the gaze toward the detected face and acquire space-variant images from selected positions within the face.

Issues related to the processing time required to detect and recognize human faces will be considered. At present, the most time consuming process is the detection of a face, while the identity verification is relatively fast, depending on the size of the face database. In general, by adopting a space-variant image representation coupled with an active camera it is possible to limit the amount of data to be processed by using the image resolution strictly required.

3.4 Stereo and invariant techniques

Both feature based and iconic based methods fall in the class of 2D techniques and, as such, are particularly sensitive to head orientation. The use of a stereo setup could be very important in order to discriminate 3D features of the subject and to improve the accuracy of visual recognition. Projective invariants have been widely used in the literature for 2D and 3D recognition. These methods appear to be very important in a large number of situations, particularly in presence of clutter and partial occlusions.

4 Preliminary results

The VIRSBS project started on November 1996 and reached in May the first milestone. The main research tasks started on December 1996 and some interesting results have been obtained concerning the segmentation of the human face, the detection of facial features and the recognition from space-variant fixations. These results are detailed in the following sections. Most of the work has been done on Khoros 2.1, the development platform adopted by the consortium.

4.1 Face segmentation

Figure 1 shows 8 images, from a sequence of 64 acquired using a fixed b/w camera. In order to segment the face of the moving subject these images were analyzed using an elliptical Hough transform which takes as input a 3x3 neighborhood approximation to the gradient of the image. The Hough transform uses the gradient direction to improve performance by reducing the number of solutions to the ellipse equation. The Hough space is actually a 5-D space, with variables in ellipse center (x, y), aspect ratio $(e = a/b)$, size (a), and orientation θ. Variation in θ is restricted to ± 20 degrees approximately. The face attitude denoted by the little ellipse is based on an algorithm which processes the density of facial features in the ellipse.

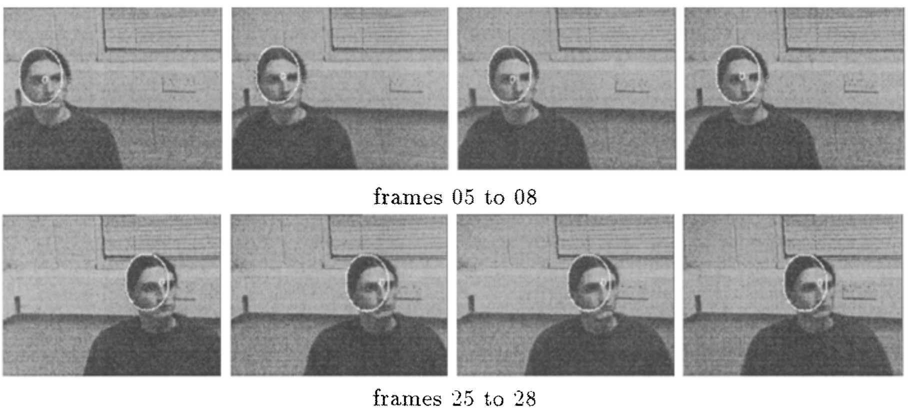


Fig. 1. Few frames, from a sequence of 64, showing the face position (large ellipse) and attitude (small ellipse) computed applying an elliptical Hough transform.

4.2 Facial features detection

The consortium has been working on the application of generalized Hough transform and morphological operators for the localization of facial features. Generalized Hough transform is used to determine a starting point for the localization of the eyes. Due to the concentricity of the contours on the eyes region this method

Fig. 2. Result of the morphological processing: (from left to right) original image, contours, valleys, peaks.

turned out to be very robust, even in case of low resolution images. Concerning the morphological operators, three basic filters have been applied to detect contours, valleys and peaks in the grey levels (see figure 2).

These informations are gathered together to make hypotheses for the presence of specific facial features. For example, the visible part of the sclera of the eye corresponds to a peak in the grey levels. On the other hand, the nostrils correspond to valleys. In order to verify the hypotheses a deformable model is finally used to check the location and shape/size of the facial features. In this phase a cost function is used to adapt the model to the feature in the image. Preliminary results obtained applying the procedure described above are shown in figure 3.

Fig. 3. Preliminary results concerning the facial features extraction for a subset of 4 images among 20 acquired. The symbol $\times$ indicates the starting point for the eyes localization procedure; ( ) denotes the final localization for the eyes while + denotes the final localization of the nostrils.

4.3 Face recognition

Face recognition algorithms, based on space-variant fixations of the subjects, are under development. A promising technique, currently investigated, uses four log-polar images per subject and exploits the principal component analysis (from the singular value decomposition) to make hypotheses of possible matches with a given subject. The gray level histogram is also matched to verify the consistency of the hypotheses. The algorithm has been tested both for identity verification and recognition.

Preliminary results of experiments on the acquisition of space-variant images and recognition from a limited database of 128 views from 16 subjects are shown. Figure 4 shows the eight views acquired for one subject. They correspond to different orientations of the head, various facial expressions and partial occlusions. Figure 5 shows the four (30x64 pixels) log-polar images extracted from each view; these images are used for principal component analysis. Figure 6 shows an

Fig. 4. All views acquired for one subject.

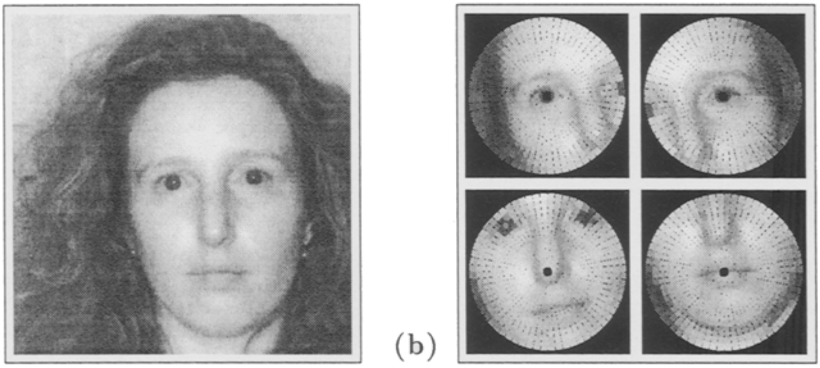
(a)  (b)

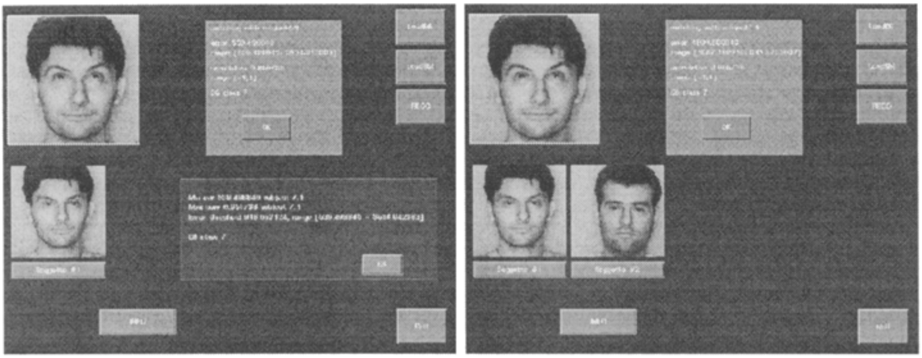
Fig. 5. (a) First view of one subject. (b) Log-polar images (eyes, nose and lips) extracted from the subject in (a), represented in the Cartesian plane. This information is coded in the face-space.

example of recognition. In this case the reference database contains the first view of each subject and the recognition is performed against all the views. The figure also shows how a limited number of fixations (two instead of four) generates ambiguities in the matching. However, the correct subject still corresponds to the highest histogram correlation score. The importance of histogram correlation is furthermore clarified by figure 7.

Finally, figure 8 summarizes the results obtained running the recognition algorithm under various configurations of the reference database. The output is categorized according to 4 possibilities:

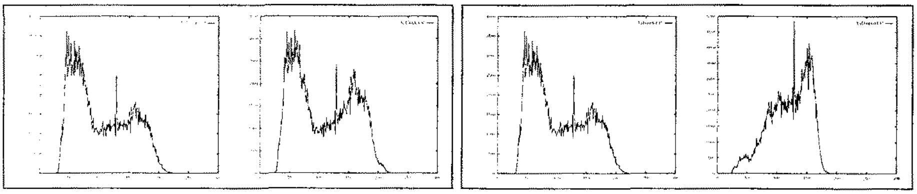
A) the subject is correctly recognized as best match, corresponding to the lowest euclidean distance;

B) the correct subject is selected among the 4 best matches, corresponding to the highest histogram correlation;



(a) (b)

Fig. 6. Example of recognition output for the same subject: (a) all 4 fixations are used, and the correct subject is recognized; (b) only two fixations, centered on the eyes, are used, and two candidates cannot be immediately discriminated by the recognition process. The correct subject corresponds to the highest histogram correlation score.



(a) (b)

Fig. 7. Example gray level histograms obtained: (a) from the same subject under different poses; (b) different subjects under the same pose.

C) the correct subject is not shown among the 4 best matches;
D) a wrong subject is recognized as best match.

5 Conclusion

Some recent outcomes of the VIRSBS project have been presented in this paper. The research effort started on December 1996 and will last till milestone 4, on January 1998. In this framework, many different techniques for face localization/recognition are under development and testing.

Starting from February 1988 the results of the research workpackages will be integrated into a prototype automatic teller, prepared by the industrial partner, and tests on the field will be performed in order to evaluate the performance of the system over a large database of subjects (1,000 - 10,000 different persons).

Acknowledgments

This work has been funded by the Commission of the European Community, under contract LTR Esprit Project 21894 "VIRSBS".

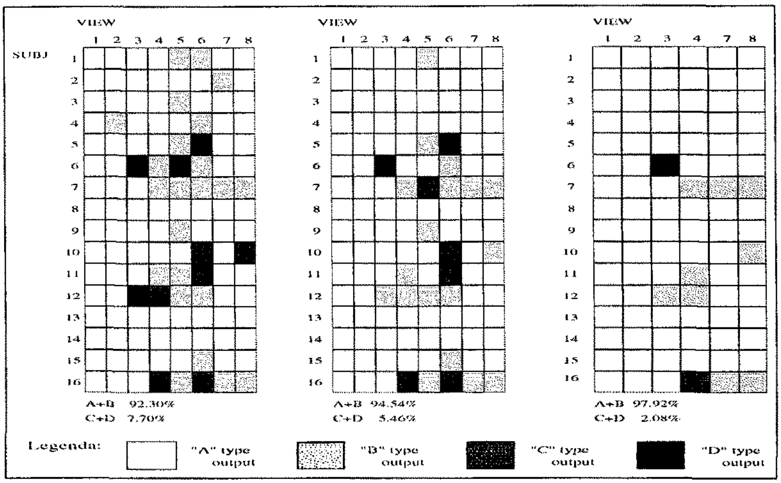


Fig. 8. Diagrams obtained from the results of some tests performed using a database of 16 subjects. The gray level codes the output type (A, B, C or D) for each recognition test. The results were obtained: (left) with a database with one view per subject (view 1). It is worth noting that the recognition score is 100% for view number 2 (smiling expression) and only one test reports output "B"; (middle) with a database with two views per subject (view 1 and 2); (right) same of above, but the side views (5 and 6) have not been used in the recognition test.

References

1. I. Craw, D. Tock, and A. Bennett. Finding face features. In *Proc. of second European Conference on Computer Vision*, pages 92–96, S. Margherita Ligure (Italy), 1992. Springer Verlag.
2. R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE Trans. on PAMI*, PAMI-15(10):1042–1052, 1993.
3. M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–79, March 1991.
4. G. Sandini and M. Tistarelli. Vision and space-variant sensing. In H. Wechsler, editor, *Neural Networks for Perception: Human and Machine Perception*. Academic Press, 1991.
5. F. Tong and Z.N. Li. The reciprocal-wedge transform for space-variant sensing. In *4th IEEE Intl. Conference on Computer Vision*, pages 330–334, Berlin, 1993.
6. E. L. Schwartz. Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception. *Biological Cybernetics*, 25:181–194, 1977.
7. C. F. R. Weiman and G. Chaikin. Logarithmic spiral grids for image processing and display. *Comp. Graphic and Image Process.*, 11:197–226, 1979.
8. M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–591, Lahaina, Maui, Hawaii, 1991. IEEE Computer Society Press.
9. T. Nguyen and T.S. Huang. Segmentation, grouping and feature detection for face image analysis. In *Proc. of the First IEEE International Symposium on Computer Vision*, pages 593–598, Miami (FL), 1995.