Advanced Identification Technologies for Human-Computer Interaction in Crisis Rooms

Massimo Tistarelli[1], Rob Van Kranenburg[2], and Enrico Grosso[1]

[1] Computer Vision Laboratory
Università di Sassari, 07100 Sassari, Italy
[2] Virtual Platform and Utrecht University
The Netherlands
{tista,grosso}@uniss.it, kranenbu@xs4all.nl

Abstract. The advances in computer and communication technologies increased both the number of users and the amount of data shared over the Network. Many times the amount of complex and articulated information available makes it difficult to retrieve what is really required for a given task. For these reasons, the efficient, easy and trustworthy transfer of data is now of paramount importance in many everyday scenarios, especially concerning environments and situations where security and data protection are mandatory. On the other hand, data protection often implies the adoption of security means which create virtual (and sometimes even physical) barriers to data retrieval. In this paper, advanced identification technologies, based on the processing of biometric data, are presented. These techniques provide a number of tools to facilitate the seamless human interaction with the data, and the security barriers, by enabling the environment to recognize and learn from the user, shaping the data available on the basis of his/her identity. The presented techniques are based on the extraction of invariant features from face and fingerprint images to process static biometric features, also allowing the enhancement of identification accuracy by data fusion.

Keywords: Personal identification, Visual recognition, Computer Vision, Biometrics, Pattern Recognition.

1 Introduction

The advent of Internet in the mid-80's boosted the requirement for new technological solutions for fast and reliable data communication. In the first 10-15 years the population in "the Net" increased from few hundred thousands to more than 40 million hosts. In the last 6 years, from the beginning of the current century, the number of hosts in Internet increased to more than 400 million: an increase of ten times in only six years. The growth in the number of users carried a gigantic amount of data transmitted over the Network by many different means. For these reasons, the efficient, easy and trustworthy transfer of data is now of paramount importance in many everyday scenarios, especially concerning environments and situations where

C. Stephanidis (Ed.): Universal Access in HCI, Part I, HCII 2007, LNCS 4554, pp. 553–562, 2007.

security and data protection are mandatory. On the other hand, data protection often implies the adoption of security means which create virtual (and sometimes even physical) barriers to data retrieval. In this context, advanced identification technologies (also termed biometric identification), based on image understanding techniques, may allow to provide a number of tools to enable the environment to recognize and learn from the user, shaping the space and the data available on the basis of his/her identity. At the same time, the user may learn or receive advice from the environment on how to parse the data or dialogue with other users [1-5].

The identification process can be based on several measurements performed on the human body. This requires the placement of several sensors in the environment to record visual, tactile and acoustic data. The identification is based on a confidence level which can be different depending on the sensing modality. At the entry level, face images, or image streams, can be used to identify each user in the room [6,7].

Depending on the required reliability of the identification more sensing modalities can be required. The user can then be required to touch a particular spot in the room where a fingerprint sensor is located and/or to utter a phrase or his/her name to perform speech-based user identification. Different sensing modalities can be invoked within a layered architecture, depending on both the security level required (for example visual data only for entry level security, visual and tactile for enhanced security and include speech for highest the security level) and the confidence level associated to the classification performed [8].

A proper interaction of the users with the environment can not be limited to the identification of each user's identity but also requires the understanding of gestures, motion and general attitude of the users which convey messages to be converted into actions by the environment. The cameras placed in the environment, deployed to identify the user's identity, can also be used to allow the recognition of motion and gestures of the users [9-12]. The basic recognition of human motion and gestures can be performed in a similar manner as the face recognition process, in this case the extraction of shape information must be coupled with the analysis of motion, to describe the temporal evolution of dynamic features [3,4].

In this paper two advanced identification technologies, based on the processing of biometric data, are presented. These techniques provide a number of tools to facilitate the seamless human interaction with the data, and the security barriers, by enabling the environment to recognize and learn from the user, shaping the data available on the basis of his/her identity. The presented techniques are based on the extraction of invariant features from face and fingerprint images to process static biometric features, also allowing the enhancement of identification accuracy by data fusion [7-12,15,16].

2 Face Recognition from Invariant Features

The recognition of human faces relies on a two step process, which is common to most biometric identification technologies:

- Characteristic features of the biometric trait (the face appearance in this case) are learnt by the system by means of an ***enrollment*** phase. This process is

accomplished by acquiring several instances of the user's biometric data to synthesize a compact and hopefully unique representation of the user's biometrics called *template*. The user's template is stored in either a personal storage device (such as a *smart card*) or in a centralized database. This procedure is generally performed off-line, within a controlled environment.

- During on-line ***identification*** or ***verification*** of the user's identity, biometric data is acquired from the user's and compared to the stored template (verification) or to a number of templates in a database (identification). Toward this end, the raw data acquired is processed to obtain the same representation stored as template. This is a fast, on-line process, which requires only a fraction of a second.

The face recognition system presented in this section is based on a complete graph representation derived from face images. The graph representation is drawn on feature points extracted using the SIFT operator [13,14]. The on-line verification or identification is performed by matching the graph representation (the template) according to some constraints [16].

2.1 Invariant and Robust SIFT Features

The scale invariant feature transform, called SIFT descriptor, has been proposed by Lowe [13] and proved to be invariant to image rotation, scaling, translation, partly illumination changes, and projective transform. The basic idea of the SIFT descriptor is detecting feature points efficiently through a staged filtering approach that identifies stable points in the scale-space. This is achieved by the following steps:

- select candidates for feature points by searching peaks in the scale-space from a difference of Gaussians (DoG) function,
- localize the feature points by using the measurement of their stability,
- assign orientations based on local image properties,
- calculate the feature descriptors which represent local shape distortions and illumination changes.

After candidate locations have been found, a detailed fitting is performed to the nearby data for the location, edge response, and peak magnitude. To achieve invariance to image rotation, a consistent orientation is assigned to each feature point based on local image properties. The histogram of orientations is formed from the gradient orientation at all sample points within a circular window of a feature point. Peaks in this histogram correspond to the dominant directions of each feature point.

In order to achieve illumination invariance, eight orientation planes are defined. The gradient magnitude and the orientation are smoothed by applying a Gaussian filter and then sampled over a 4 x 4 grid with 8 orientation planes.

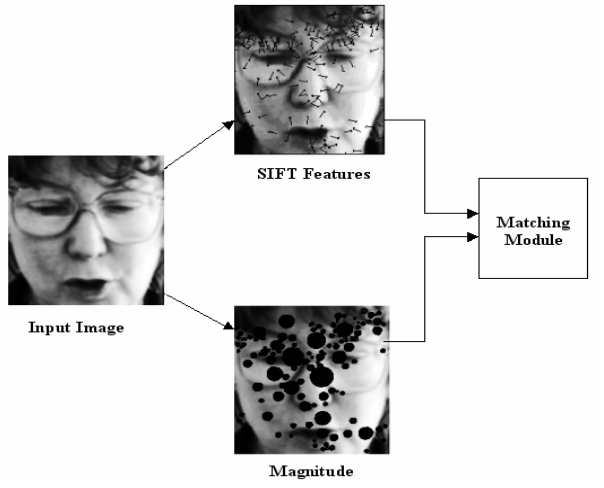

Fig. 1. Example image used for face recognition showing the SIFT Features locations and orientation (top) and the magnitude (bottom) coded by circles of different radius

2.2 Graph-Based Matching

From each face image a set of invariant and stable features (SIFT) are extracted. All features extracted from a single, static face image, are used to build an invariant representation of the subject's face. This representation is obtaining building a complete graph, whose nodes are the feature points extracted using the SIFT operator [13,14]. In order to perform the identification or verification process, several matching constraints can be applied to reduce the probability of mismatching when pairing the template and user's graphs [15,16]:

- *Gallery image based match constraint*. It is assumed that matching points will be found around similar positions i.e., fiducial points on the face image. In order to eliminate false matches a minimum Euclidean distance measure is computed by mean of the Hausdroff metric.
- *Reduced point based match constraint.* In addition to the previous constraint, all false matches due to one way assignments are eliminated by removing the links which do not have any corresponding assignment from the other side.
- *Regular grid based match constraint.* To further improve the robustness of the graph matching, the image is divided into a set of overlapping sub images. The matching is determined by computing the distances between all pairs of corresponding sub-image graphs, and finally averaging them with dissimilarity scores for a pair of sub-images.

All these techniques are aimed to find the corresponding sub-graph in the probe face image given the complete graph in the gallery image.

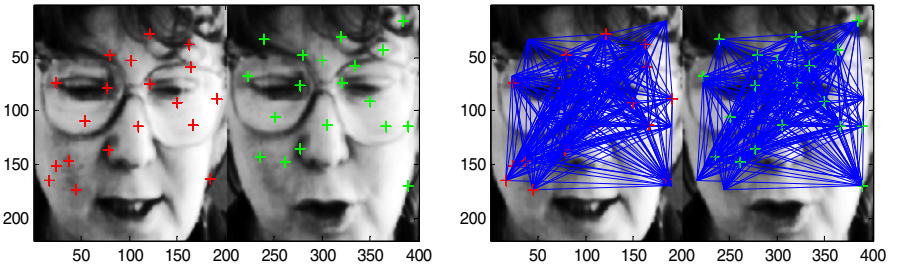

Fig. 2. An example of reduced point based match constraint. On the left, all matches computed from the left to the right image. On the right, the resulting complete graphs with a few numbers of false matches.

Table 1. Prior EER on G1 and G2 for the two methods: 'GIBMC' stands for gallery image based match constraint, 'RPBMC' stands for reduced point based match constraint and 'RGBMC' stands for regular grid based match constraint

	GIBMC	RPBMC	RGBMC
Prior EER on G1	10.13%	6.66%	4.65%
Prior EER on G2	6.92%	1.92%	2.56%
Average	8.52%	4.29%	3.6%

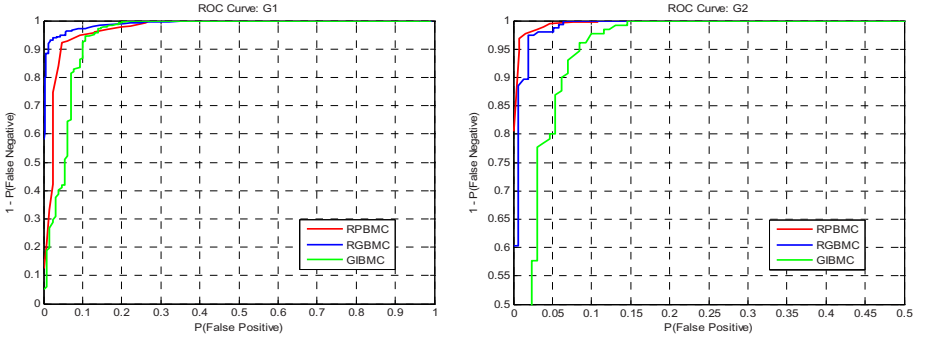

Fig. 3. ROC curves for G1 (left) and G2 (right): 'GIBMC' stands for gallery image based match constraint, 'RPBMC' stands for reduced point based match constraint and 'RGBMC' stands for regular grid based match constraint

2.3 Experimental Testing

The proposed graph matching technique was tested on the BANCA database [17]. The database used is composed of four image sets captured from 52 subjects in different acquisition sessions. For this experiment, the Matched Controlled (MC) protocol is followed, where the images from the first session are used for training,

whereas second, third, and fourth sessions are used for testing and generating client and impostor scores. The testing images are divided into two groups, G1 and G2, of 26 subjects each. The error rate was computed using the standard procedure described in [17]. The reported errors are detailed in figure 3 and table 1. The obtained results show the capability of the system to cope for illumination changes and occlusions. This is a desirable feature for real applications in environments where both the lighting conditions and the subject's face position can only be partially controlled. From the ROC curves in figure 3, it is worth noting that the level of security can be increased by reducing the number of wrong acceptance (i.e. the risk of a false positive). This is obtained by tuning the matching threshold, which has the effect of moving the error on the abscissa toward the left, at the cost of an increased number of false alarms (enrolled users which are not recognized).

3 Multibiometric Recognition

From a system point of view, redundancy can always be exploited to improve accuracy and robustness which is achieved in many living systems as well. Human beings, for example, use several perception cues for the recognition of other living creatures. They include visual, acoustic and tactile perception. Multibiometric systems [6,7] remove some of the drawbacks of the uni-biometric systems by grouping the multiple sources of information. These systems utilize more than one physiological or behavioral characteristic for enrollment and on-line matching [8]. It has been demonstrated that a biometric system that integrates information at an earlier stage of processing provides more accurate results than the integration of information at a later stage, because of the availability of richer information [7]. This section illustrates a novel approach to fuse face and fingerprint biometrics at the feature extraction level. The improvement obtained applying the feature level fusion is presented over the score level fusion technique.

3.1 Face Representation Based on SIFT Features

The face recognition system introduced in the previous section, is based on the SIFT [13,14] features extracted from images of the query and database face. The face representation module is based on the spatial, orientation and keypoint descriptor information of each extracted SIFT point. Thus the input to the face module is represented by the face image itself, while the output is the set of extracted SIFT features $s=(s_1, s_2,\ldots\ldots s_m)$, where each feature point $s_i=(x, y, \theta, k)$ consist of the (x, y) spatial location, the local orientation θ and k is the keydescriptor of size $1x128$.

3.2 Fingerprint Representation Based on Minutiae

The fingerprint recognition module has been developed using the minutiae-based technique discussed in [18]. In order to achieve the same rotation invariance as the facial SIFT features, the fingerprint image is processed to detect the left, top and right edges of the foreground to calculate the overall slope, and by fitting a straight line to each edge by linear regression. A rectangle is fitted to the segmented region and rotated with the same angle to nullify the effect of rotation. The input to the

fingerprint module is the fingerprint image and the output is the extracted minutiae $m=(m_1, m_2,\ldots\ldots m_m)$ where each feature point $m_i=(x, y, \theta)$ consist of the (x, y) spatial location and the local orientation θ [19].

3.3 Feature Level Fusion Scheme

The feature level fusion is realized by simply concatenating the feature points obtained from different sources of information. The concatenated feature pointset has better discrimination power than individual feature vectors. The procedure followed for concatenation is as follows:

- *Feature set compatibility and normalization*. The minutiae feature pointset is made compatible with the SIFT feature pointset by making it rotation and translation invariant and introducing the keypoint descriptor along with the minutiae position. The keypoint descriptors of each face and fingerprint points are then normalized using *min-max normalization* technique (s_{norm} and m_{norm}) to ensure that all the *128* values of a keypoint descriptor are within *0 to 1* range.
- *Feature Reduction and Concatenation.* The feature level fusion is performed by concatenating the two feature pointsets; This results in a fused feature pointset, which we call $concat=(s_{1norm}, s_{2norm},\ldots s_{mnorm},\ldots m_{1norm}, m_{2norm}, m_{mnorm})$. Feature reduction is applied to eliminate irrelevant feature points both before or after feature concatenation.
- *Matching*. The concatenated features pointset *(concat and concat')* of the database and the query images are processed by the matcher to compute the proximity between the two pointsets. The point pattern matching technique is applied, where the similarity between two concatenated features is based on the number of matched pairs found in the two sets, for both monomodal traits and for the fused feature pointset.

3.4 Experimental Testing

The multibiometric system has been tested on two different databases: the first consists of 50 chimeric individuals composed of 5 face and fingerprint images for each individual. The face images are taken from the controlled sessions of BANCA Database [17], while the fingerprint images were collected by the authors. The fingerprint images were acquired using an optical sensor at 500 dpi.

The following procedure has been established for testing mono-modals and multimodal system:

– *Training*: one image per person is used for enrollment in the face and fingerprint verification system; for each individual, one pair face-fingerprint is used for training the fusion classifier.
– *Testing*: this represents a thorough test made on the first chimeric dataset. Four samples per person are used for testing and generating client scores. Impostor scores are generated by testing the client against the first sample of the rest of the individuals, in the case of monomodal systems. In case of multimodal testing the client is tested against the first face and fingerprint samples of the rest of the chimeric users thus in total *50x4=200* client scores and *50x49=2450* imposters scores for each of the uni-modal and the multimodal systems are generated.

Table 2. FAR, FRR and accuracy values obtained from the multimodal fusion

Algorithm	*FRR(%)*	*FAR(%)*	*Accuracy*
Fingerprint	5.384	10.97	91.82
(Face+Finger) score level	5.66	4.78	94.77
(Face+Finger) Feature Level	1.98	3.18	97.41

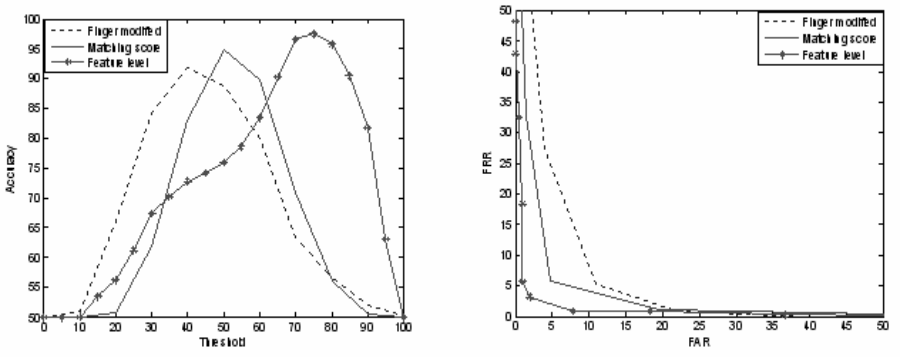

Fig. 4. Comparison of recognition performance. The accuracy (left) and ROC curve (right) for modified fingerprint, fusion at matching score and at feature level.

As multiple information sources can be exploited to gain a better recognition performance, this section outlined the possibility to augment the verification accuracy by integrating multiple biometric traits. The reported results show that a remarkable improvement in the accuracies are obtained when properly fusing feature sets which do not constitute an end in itself, but rather suggests to attempt a multimodal data fusion as early as possible in the processing pipeline. The actual feasibility of this approach may heavily depend on the physical nature of the acquired signal and on the overall set-up of the data acquisition scenario. For this reason the architecture of the data acquisition system must be purposively designed to allow an optimal functionality of the system to identify individuals within the context of crisis rooms.

4 Conclusion

The dynamic man-machine interaction within the context of crisis rooms requires the implementation of advanced identity recognition strategies to provide a personalized access to secure data. Two advanced systems were presented based on unimodal and multimodal biometric systems, the latter based on the integration of face and a fingerprint traits at the feature extraction level.

Considering the general problem of identity recognition, redundancy can be always exploited to improve accuracy and robustness. This can be accomplished in several ways: processing repeated measurements of the same biometric trait, applying different algorithms to the same data, or processing data from different biometric

traits. Human beings, for example, use several perception cues for the recognition of other living creatures. They include visual, acoustic and tactile perception. Starting from these considerations, this paper outlined the possibility to augment the verification accuracy by integrating multiple biometric traits. There are several advantages in multimodal biometric systems. Not only they provide an augmented performance, but also increase the easy of use (by minimizing the probability of false alarms and false rejections of registered users), robustness to noise, and the possibility to use low-cost, off-the-shelf hardware for data acquisition.

In most of the examples presented in the literature, fusion is performed either at the score level or at the decision level, always improving the performance of each single modality. In this paper a novel approach has been presented were both fingerprint and face images are processed with compatible feature extraction algorithms to obtain comparable features from the raw data. Reported results show that a remarkable improvement in the accuracies is obtained when properly fusing feature sets.

The presented approach does not constitute an end in itself, but rather suggests to attempt a multimodal data fusion as early as possible in the processing pipeline. Possibly this may even imply a normalization of the data at the sensor acquisition level to allow an early fusion of the raw data. The actual feasibility of this approach may heavily depend on the physical nature of the acquired signal. Several experiments, performed with different algorithmic solutions, have been presented on a chimeric multimodal database. Further experiments will allow to better validate the overall system performances.

Acknowledgments. This work has been partially supported by grants from the Italian Ministry of Research, the Ministry of Foreign Affairs and the Biosecure European Network of Excellence. The collaboration of Manuele Bicego, Dakashina Kisku, Andrea Lagorio and Ajita Rattani has been fundamental for the development of this work.

References

1. McNeill.: Hand and Mind: What Gestures Reveal about Thought. University of Chicago Press, Chicago (1992)
2. Essa, I., Pentland, A.: Coding, Analysis, Interpretation and Recognition of Facial Expressions. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19(7), IEEE Computer Society Press, Washington (1997)
3. Weng, J., Cui, Y.: Recognition of hand signs from complex backgrounds. In: Cipolla, Pentland (eds.) Computer Vision for Human-Machine Interaction, Cambridge University Press, Cambridge (1998)
4. Watanabe, H., Hongo, H., Yasumoto, M., Niwa, Y., Yamamoto, K.: Control of Home Appliances Using Face and Hand Sign Recognition. In: Proceedings of International Conference on Computer Vision 2001 - ICCV 2001 (II: 746) (2001)
5. Jain, A.K., Bolle, R., Pankanti, S.: Biometrics: Personal Identification in a Networked Society. Kluwer Academic Publishers, Boston (1999)
6. Jain, A.K., Ross, A.: Multibiometric systems. Communications of the ACM 47(1), 34–40 (2004)

7. Ross, Jain, A.K.: Information Fusion in Biometrics. Pattern Recognition Letters 24, 2115–2125 (2003)
8. Hong, L., Jain, A.K.: Integrating Faces and Fingerprints for Personal Identification. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(12), 1295–1307 (1998)
9. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face Recognition: a Literature Survey. ACM Computing Surveys 35, 399–458 (2003)
10. Tistarelli, M., Grosso, E.: Active vision-based face authentication Image and Vision Computing: Special issue on Facial Image Analysis, M. Image and Vision Computing: Special issue on Facial Image Analysis, M. Tistarelli ed 18(4), 299–314 (2000)
11. Bicego, M., Grosso, E., Tistarelli, M.: Face Authentication Using One-Class Support Vector Machines. In: IEEE Intern.l Workshop on Audio Biometric Recognition Systems 2005, Beijing, China, 22–23 October, Springer, Heidelberg (2005)
12. Tistarelli, M., Lagorio, A., Grosso, E.: What Can I Tell From Your Face? In: Zhang, D., Jain, A.K. (eds.) ICBA 2004. LNCS, vol. 3072, pp. 72–78. Springer, Heidelberg (2004)
13. Lowe, David, G.: Object recognition from local scale invariant features, International Conference on Computer Vision, Corfu, Greece, pp. 1150–1157 (September 1999)
14. Schmid, Mohr, R.: Local grayvalue invariants for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(5), 530–534 (1997)
15. Bicego, M., Lagorio, A., Grosso, E., Tistarelli, M.: On the use of SIFT features for face authentication. In: Proceedings of CVPR Workshop, New York (2006)
16. Kisku, D., Rattani, A., Grosso, E., Tistarelli, M.: On the use of SIFT features for face authentication. In: Proceedings of IEEE Automatic Identification Technologies Workshop 2007, Alghero, Italy (2007)
17. Bailly-Baillire, E., Bengio, S., Bimbot, F., Hamouz, M., Kittler, J., Marithoz, J., Matas, J., Messer, K., Popovici, V., Pore, F., Ruiz, B., Thiran, J.P.: The BANCA database and evalution protocol. In: Proc. Int. Conf. on Audio – and Video-Based Biometric Person Authentication. pp. 625–638, Springer, Heidelberg
18. Ratha, N.K., Karu, K., Chen, S., Jain, A.K.: A Real-time Matching System for Large Fingerprint Databases. IEEE Transactions on PAMI 18(8), 799–813 (1996)
19. Jain, A.K., Prabhakar, S., Hong, L., Pankanti, S.: Filterbank-based fingerprint matching. IEEE Transactions on Image Processing 9(5), 846–859 (2000)