A multi-biometric verification system for the privacy protection of iris templates

S. Cimato, M. Gamassi, V. Piuri, R. Sassi and F. Scotti

Dipartimento di Tecnologie dell'Informazione, Università di Milano,
Via Bramante, 65 – 26013 Crema (CR), Italy
{cimato,gamassi,piuri,sassi,fscotti}@dti.unimi.it

Abstract. Biometric systems have been recently developed and used for authentication or identification in several scenarios, ranging from institutional purposes (border control) to commercial applications (point of sale). Two main issues are raised when such systems are applied: reliability and privacy for users. Multi-biometric systems, i.e. systems involving more than a biometric trait, increase the security of the system, but threaten users' privacy, which are compelled to release an increased amount of sensible information. In this paper, we propose a multi-biometric system, which allows the extraction of secure identifiers and ensures that the stored information does not compromise the privacy of users' biometrics. Furthermore, we show the practicality of our approach, by describing an effective construction, based on the combination of two iris templates and we present the resulting experimental data.

1 Introduction

Nowadays, biometric systems are deployed in several commercial, institutional, and forensic applications as a tool for identification and authentication [1, 2]. The advantages of such systems over traditional authentication techniques, like the ones based on the possession (of a password or a token), come from the fact that identity is established on the basis of physical or behavioral characteristics of the subject taken into consideration and not on something he/she carries. In fact, biometrics cannot be lost or stolen, they are difficult to copy or reproduce, and in general they require the presence of the user when the biometric authentication procedure takes place.
However, side to side with the widespread diffusion of biometrics an opposition grows towards the acceptance of the technology itself. Two main reasons might motivate such resistance: the *reliability* of a biometric system and the possible threatens to users' *privacy*. In fact, a fault in a biometric system, due to a poor implementation or to an overestimation of its accuracy could lead to a security breach. Moreover since biometric traits are permanently associated to a person, releasing the biometric information acquired during the enrollment can be dangerous, since an impostor could reuse that information to break the biometric authentication process. For this reason, privacy agencies of many countries have ruled in favor of a legislation which limits the biometric information that can be centrally stored or carried on a personal ID. For example, templates, *e.g.* mathematical information derived from a fingerprint, are re-

tained instead of the picture of the fingerprint itself. Also un-encrypted biometrics are discouraged.

A possible key to enhance the reliability of biometric systems might be that of simultaneously using different biometric traits. Such systems are termed in literature *multi-biometric* [3] and they usually rely on a combination of one of several of the followings: (*i*) multiple sensors, (*ii*) multiple acquisitions (*e.g.*, different frames/poses of the face), (*iii*) multiple traits (*e.g.*, an eye and a fingerprint), (*iv*) multiple instances of the same kind of trait (*e.g.*, left eye, and right eye). As a rule of thumb, the performances of two of more biometric systems which each operate on a single trait might be enhanced when the same systems are organized in a single multimodal one. This is easy to understand if we refer to the risk of admitting an impostor: two or more different subsequent verifications are obviously more difficult to tamper with than a single one (AND configuration). But other less obvious advantages might occur. Population coverage might be increased, for example, in an OR configuration since some individuals could not have one biometric traits (illnesses, injuries, etc.). Or the global fault tolerance of the system might be enhanced in the same configuration, since, if one biometric subsystem is not working properly (*e.g.*, a sensor problem occurred), the multimodal system can still keep working using the remaining biometric submodules. On the other hand, the usage of multimodal biometric systems has also some important drawbacks related to the higher cost of the systems, and user perception of larger invasiveness for his/her privacy.

In the following, we will derive a multi-biometric authentication system which limits the threats posed to the privacy of users while still benefiting from the increase reliability of multiple biometrics. It was introduced in [4] and it is based on the secure sketch, a cryptographic primitive introduced by Dodis *et al*. in [5]. In fact, a main problem in using biometrics as cryptographic keys is their inherent variability in subsequent acquisitions. The secure sketch absorbs such variability to retrieve a fixed binary string from a set of similar biometric readings.

In literature biometric authentication schemes based on secure sketches have been presented and applied to face and iris biometrics [6, 7]. Our proposal is generally applicable to a wider range of biometric traits and, compared to previous works, exploits multimodality in innovative way. In the following we describe the proposed construction and show its application to the case where two biometrics are used, the right and left iris. Iris templates are extracted from the iris images and used in the enrolment phase to generate a secure identifier, where the biometric information is protected and any malicious attempt to break the users' privacy is prevented.

2 A Multimodal Sketch based (MSB) Verification Scheme

The MSB verification scheme we propose is composed of two basic modules: the first one (enroll module) creates an identifier (ID) for each user starting from the biometric samples. The ID can then be stored and must be provided during the verification phase. The second one, the (verification module) performs the verification process starting from the novel biometric readings and the information contained into the ID.

Verification is successful if the biometric matching succeeds when comparing the novel reading with the stored biometrics, concealed into the ID.

2.1 Enrollment module

The general structure of the enroll module is depicted in Figure 1 in its basic configuration where the multimodality is restricted at two biometrics. The scheme can be generalized and we refer the reader to [5] for further details. First, two independent biometrics are acquired and processed with two feature extraction algorithms F_1 and F_2 to extract sets of biometric features. Each set of features is then collected into a template, a binary string. We refer to each template as I_1 and I_2. The feature extraction algorithms can be freely selected; they represent the single biometric systems which compose the multimodal one. Let us denote with r_i the binary tolerable error rate of each biometric subsystem, i.e., the rate of bits in the templates which could be modified without affecting the biometric verification of the subject.

The second biometric feature I_2 is given as input to a pseudo random permutation block, which returns a bit string of the same length, having almost uniform distribution.

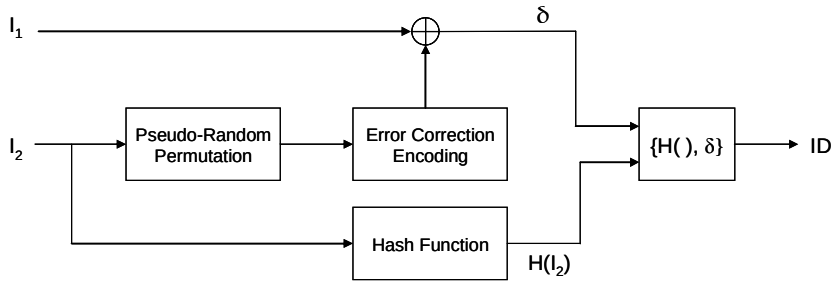

Fig. 1. The MSB Enroll Module

The string is then encoded by using an error correcting code and the resulting codeword c is xored with the other biometric feature I_1 to obtain δ. Given N_1, the bit-length of I_1, the code must be selected so that it corrects *at most* r_1N_1 single bit errors on codewords which are N_1 bits long. Finally, I_2 is given as input to a hash function and the digest $H(I_2)$, together with δ, and other additional information possibly needed (to invert the pseudo random permutation) are collected and published as the identifier of the enrolled person.

2.2 Verification module

Figure 2 shows the structure of the verification module. Let us denote with I'_1 and I'_2 the biometric features freshly collected. The *ID* provided by the subject is split into δ, the hash $H(I_2)$ and the key needed to invert the pseudo random permutation. A corrupted version of the codeword c, concealed at enrollment, is retrieved by xoring the fresh reading I'_1 with δ. Under the hypothesis that both readings I_1 and I'_1 belong to

the same subject, the corrupted codeword *c'* and *c* should differ for at most r_1 bits. Thus the subsequent application of the error correcting decoding and of the inverse pseudo random permutation, should allow the exact reconstruction of the original reading I_2.

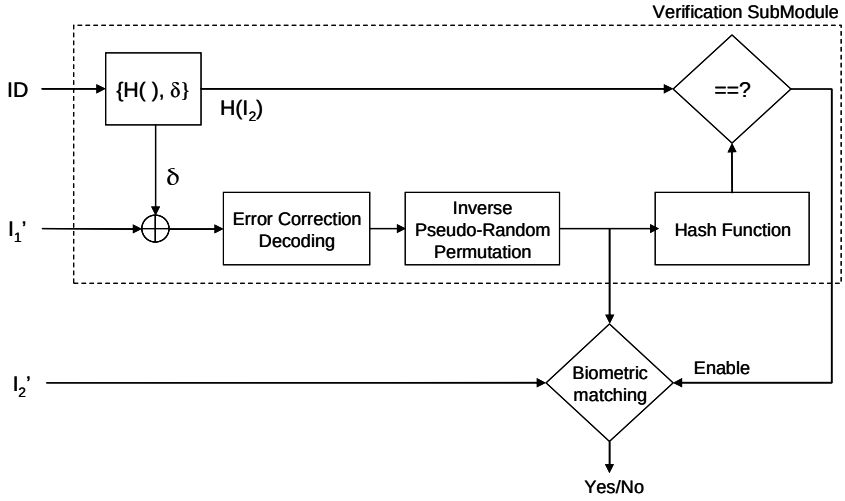

Fig. 2. The MSB Verification Module

The identity of the user is verified in two steps. First a check is performed to compare the hash of the retrieved value for I_2 with the value $H(I_2)$ stored into the identifier. If the check succeeds it means that the readings of the first biometric trait did not differ more than what permitted by the biometric employed. Then a second biometric matching is performed using as input the retrieved value of I_2 and the fresh biometric reading I'_2. The authentication is successful when also this second match is positive.

3. Experimental data and results

3.1 Dataset creation

The proposed scheme has been tested by using the public CASIA dataset [8]. (version 1.0) which contains seven images of the same eye obtained from 108 subjects. The images were collected by the Chinese Academy of Science waiting at least one month between two capturing stages using near infrared light for illumination (3 images during the first session and 4 for the second one). We used the first 3 images in the enroll operations, and the last 4 images in the verification phase. At the best of our knowledge, there is no public dataset containing the left and right eyes sample of each individual with the sufficient iris resolution to be effectively used in identification tests. For this reason we synthetically created a new dataset by composing two irises of different individuals taken from the CASIA dataset. Table 1 shows the details of the composition method used to create the synthetic dataset from the CASIA samples.

Table 1. Creation of the synthetic dataset.

CASIA Individual Identifier	CASIA File Name	Enroll/ Validation	Synthetic DB Individual Identifier	Notes
001	001_1_1.bmp	Enroll	01	Right eye, Enroll, Sample 1
	001_1_2.bmp	Enroll		Right eye, Enroll, Sample 2
	001_1_3.bmp	Enroll		Right eye, Enroll, Sample 3
	001_2_1.bmp	Validation		Right eye, Validation, Sample 1
	…	…		…
	001_2_4.bmp	Validation		Right eye, Validation, Sample 4
002	002_1_1.bmp	Enroll		Left eye, Enroll, Sample 1
	002_1_2.bmp	Enroll		Left eye, Enroll, Sample 2
	002_1_3.bmp	Enroll		Left eye, Enroll, Sample 3
	002_2_1.bmp	Validation		Left eye, Validation, Sample 1
	…	…		…
	002_2_4.bmp	Validation		Left eye, Validation, Sample 4

The method we used to create the dataset can be considered as a pessimistic estimation of real conditions, since the statistical independence of the features extracted from the iris samples coming from the left and right eye of the same individual is likely to be equal or lower than the one related to the eyes coming from different individuals. In the literature it has been showed that the similarities of the iris templates coming from the left and right eyes of the same individuals are negligible when Iriscodes templates are used [9]

3.2 Template creation

The iris templates of the left and right eyes were computed using the code presented in [10] (a completely open implementation which builds over the original ideas of Daugman [9]). The code has been used to obtain the iris codes of the right and left eye of each individual present in the synthetic database.

The primary biometric template I_1 has been associated to the right eye of the individual by using a 9600 bits wide iris template. As suggested in [10], the 9600 bits have been obtained by processing the iris image with a radial resolution (the number of points selected along a radial line) of 20. The author suggested for the CASIA database a matching criterion with a separation point of $r_1 = 0.4$ (Hamming distance between two different iris templates). Using such a threshold, we independently verified that the algorithm was capable of a false match rate (FMR, the probability of an individual not enrolled being identified) and false non-match rate (FNMR, the probability of an enrolled individual not being identified by the system) of 0.028% and 9.039%, respectively using the CASIA version 1.0 database. Such rates rise to 0.204% and 16.799% respectively if the masking bits are not used. The masking bits mark bits in the iris code which should not be considered when evaluating the Hamming distance between different patterns due to reflections, eyelids and eyelashes coverage, etc.

Due to security issues, we preferred to not include the masking bits of the iris code in the final templates since the distribution of zero valued bits in the masks is far from being uniform. The higher FNMR compared with the work of [10] can be explained

by considering that using the adopted code failed segmentations of the pupil were reported to happen in the CASIA database in 17.4% of the cases.

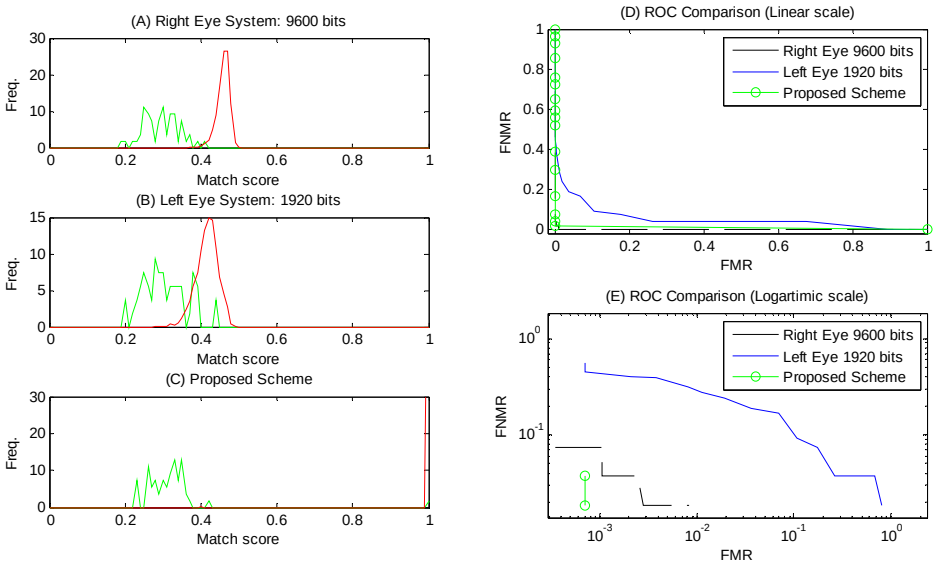

Fig. 3. Impostor and genuine frequency distributions of the iris templates composed by 9600 bits (A) and 1920 bits (B) using the synthetic dataset and for the proposed scheme (C and D respectively). The corresponding FNMR versus FMR are plotted in linear (D) and logarithmic scale (E).

3.3 Enroll and verification procedures

The enroll procedure for the right eye has been executed according to the following steps. The three iris codes available in the enroll phase (Table 1) of each individual were evaluated for quality, in term of number of masking bits. The iris code with the highest "quality" was retained for further processing. The *best of three* approach was devised to avoid that segmentation errors might further jeopardize the verification stage. Then, the remaining enroll phases were performed according to the description previously made. A Reed-Solomon $[9600,1920,7681]_{m=14}$ correction code has been adopted with $n_1 = 9600$ and $r_1 = 0.4$. In such set up, the scheme allows for up to $k = 1920$ bits for storing the second biometric template. If list decoding is taken into consideration the parameters should be adapted to take into account the enhanced error correcting rate of the list decoding algorithm. The former has been chosen by selecting the available left iris template with highest quality (*best of three* method) in the same fashion adopted for the right eye. Using this approach, a single identifier ID has been created for every individual present in the synthetic dataset. In particular, the shorter iris code was first subjected to a pseudo random permutation (we used AES in CTR mode) and then it was encoded with the RS code and then xored with the first one to obtain δ. Note that the RS codewords are 14 bits long. The unusual usage of

the RS code (here we didn't pack the bits in the iris code to form symbols, as in typical industrial application) is due to the fact that here we want to correct "at most" a certain number of error (and not "at least"). Each bit of the iris code was then inserted in a separate symbol adding random bits to complete the symbols. Finally an hash value of the second biometric template was computed to get the final ID with δ. In the implementation we used the hash function SHA-1 (Java JDK 6).

In the verification procedure, the left eye related portion was processed only if one of the iris codes was able to unlock the first part of the scheme. Otherwise the matching was considered as failed, and a maximum Hamming distance of 1 was associated to the failed matching value. If the first part of the scheme was successful, the recovered left eye template was matched by using a classical biometric system with the left eye template selected for the validation. The Hamming distance between the two strings is used to measure the distance between the considered templates. The *best of four* strategy is applied using the four left eye images available in the validation partition of the synthetic dataset.

3.4 Experimental results for the proposed scheme

The performances of the proposed method are strictly related to the performance of the code that constructs the iris templates. As such, a fair comparison should be done by considering as reference the performances of the original iris code system working on the same dataset. If we adopt the original iris templates of 9600 and 1920 bits by using the same enroll and verification procedure in a traditional fashion (*best of three* in verification, *best of four* in verification, no masking bits), we obtain the system behaviors described in Figure 3. The right eye system (9600 bits) has good separation between the genuine and impostor distributions and it achieves an equal error rate (ERR, the value of the threshold used for matching at which FMR equals FNMR) that can be estimated to about 0.5% on the synthetic dataset. The left eye system is working only with 1920 bits and achieves a worst separation between the two populations. The corresponding EER has been estimated to be equal to 9.9%.

On the other hand, our multimodal scheme achieves an EER that can be estimated to be equal to 0.96%, and shows then an intermediate behavior between the ROC curves of each single biometric system based on the right or on the left eye (Figure 3). For a wide portion of the ROC curve, the proposed scheme achieves a better performance with respect to the right eye biometric system. That behavior is common for traditional multimodal systems where, very often, the multimodal system can work better than the best single biometric sub-system. The proposed scheme seem to show this interesting property and the slightly worse EER with respect to the best single biometric system (right eye, 9600 bits) is balanced by the protection of the biometric template. We may suppose that the small worsening for the EER is related to the specific code we used to compute the iris code templates and that it might be ameliorated by selecting a different code. Further experiments with enlarged datasets, different coding algorithms and error correction codes will be useful to validate the generality of the discussed results.

4. Conclusions

In this work we proposed a method based on the secure sketch cryptographic primitive to provide an effective and easily deployable multimodal biometric verification system. Privacy of user templates is guaranteed by the randomization transformations which avoid any attempt to reconstruct the biometric features from the public identifier, preventing thus any abuse of biometric information. We also showed the feasibility of our approach, by constructing a biometric authentication system that combines two iris biometrics. The experiments confirm that only the owner of the biometric ID can "unlock" her/his biometric templates, once fixed proper thresholds. More complex systems, involving several biometric traits as well as traits of different kinds will be object of further investigations.

Acknowledgments
The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 216483.

References

1. Jain, A. K., Ross, A., Pankanti, S.: Biometrics: A tool for information security. IEEE Trans. on information forensics and security 1(2), 125–143 (2006)
2. Uludag, U, Pankanti, S., Prabhakar, S., Jain, A.: Biometric cryptosystems: Issues and challenges. Proceedings of the IEEE, Special Issue on Enabling Security Technologies for Digital Rights Management, 92, 948–960 (2004)
3. Snelick, R., Uludag, U., Mink, A., Indovina, M., Jain, A.K: Large scale evaluation of multimodal biometric authentication using state of the art systems. IEEE Trans. Pattern Analysis and Machine Intelligence 27(3), 450–455, (2005)
4. Cimato, S, Gamassi, M., Piuri, V., Sassi, R., Scotti, F.: A biometric verification system addressing privacy concerns. In: IEEE International Conference on Computational Intelligence and Security (CIS 2007), 594-598, (2007).
5. Dodis, Y., Ostrovsky, R., Reyzin, L., Smith, A.: Fuzzy extractors: How to generate strong keys from biometrics and other noisy data, Cryptology Eprint Archive, Tech. Rep. 2006/235, (2006).
6. Bringer, J., Chabanne, H., Cohen, G., Kindari, B., Zemor, G.: An application of the goldwasser-micali cryptosystem to biometric authentication. In: Proceedings of Information Security and Privacy, 12th Australasian Conference, LNCS 4586, 96–106, Springer-Verlag (2007).
7. Sutcu, Y., Li, Q., Memon, N.: Protecting biometric templates with sketch: Theory and practice. IEEE Trans. on Information Forensics and Security, 2(3), (2007).
8. Chinese Academy of Sciences: Database of 756 greyscale eye images; Version 1.0, http://www.sinobiometrics.com/IrisDatabase.htm, (2003).
9. Daugman, J. G.: High confidence visual recognition of persons by a test of statistical independence. IEEE Trans. on Pattern Analysis and Machine Intelligence, 15, 1148–1161 (1993)
10. Masek, L.: Recognition of human iris patterns for biometric identification. Bachelor's Thesis, School of Computer Science and Software Engineering, University of Western Australia (2003)