Template Protection and its Implementation in 3D Face Recognition Systems

Xuebing Zhou

Fraunhofer IGD, Fraunhoferstr. 5, 64283 Darmstadt, Germany
E-mail: xuebing.zhou@igd.fhg.de

ABSTRACT

As biometric recognition systems are widely applied in various application areas, security and privacy risks have recently attracted the attention of the biometric community. Template protection techniques prevent stored reference data from revealing private biometric information and enhance the security of biometrics systems against attacks such as identity theft and cross matching. This paper concentrates on a template protection algorithm that merges methods from cryptography, error correction coding and biometrics. The key component of the algorithm is to convert biometric templates into binary vectors. It is shown that the binary vectors should be robust, uniformly distributed, statistically independent and collision-free so that authentication performance can be optimized and information leakage can be avoided. Depending on statistical character of the biometric template, different approaches for transforming biometric templates into compact binary vectors are presented. The proposed methods are integrated into a 3D face recognition system and tested on the 3D facial images of the FRGC database. It is shown that the resulting binary vectors provide an authentication performance that is similar to the original 3D face templates. A high security level is achieved with reasonable false acceptance and false rejection rates of the system, based on an efficient statistical analysis. The algorithm estimates the statistical character of biometric templates from a number of biometric samples in the enrollment database. For the FRGC 3D face database, the small distinction of robustness and discriminative power between the classification results under the assumption of uniquely distributed templates and the ones under the assumption of Gaussian distributed templates is shown in our tests.

Keywords: template protection, security and privacy enhancement, 3D face verification system, biometric-based cryptography

1. INTRODUCTION

Biometrics techniques uniquely combine authentication and user friendliness. Biometrics enabled authentication has been widely employed in numerous application area such as electronic access and border control in the recent years. In enrollment of biometrics system, biometric template, which represents features of biometric modality, is extracted and stored in data storage. In verification process, stored template is compared with lively extracted template to authenticate user identity. Nevertheless, storage and transmission of biometric template leads to many security and privacy problems. In common biometrics system, biometric templates can be easily obtained by attackers, malicious system providers and verifier. Once a biometric template is compromised, the corresponding personal identity based on biometrics is compromised. Damage of misusing and abusing biometric information is high, since it is not able to be revoked, reissued or destroyed. Furthermore, when the same biometric templates are deployed in manifold applications, user actions can be linked and malicious data holder can track user actions in other applications. From privacy points of view, employment of biometrics is very critical, since biometrics information includes sensitive information like genetic and disease information. Still, collecting and storing biometric information are critical in numerous countries because of privacy legislation. In this paper, we address template protection techniques, which enhance security and protect privacy in biometrics system.

Template protection techniques convert biometric template into a secure reference with help of random variables. The resulting secure reference reveals very little information of the original biometric template, meanwhile, it is robust to biometric variation. Secure references are compared directly and recovering the original template is not necessary. Alternatively, “on card matching” mechanism can also protect privacy. It enables to store

biometric templates and to conduct matching on a smart card. A biometric template is never released from the corresponding storage, which is held by an individual. Nevertheless, the performance of this mechanism is limited by capacity of card storage and communication channel. Furthermore, encryption of the transmission channel is desired to prevent eavesdropping attack. The complexity of system rises due to the ongoing key management. Each smart card must be authenticated prior to the usage of the contained template. Additionally, the integration of smart cards increases obviously cost of biometrics system and identification is impossible. In contrast, template protection overcomes security and privacy weakness of biometrics system. Private biometric information is efficiently protected. Centralized database is allowed without conflict of privacy law. The randomness of template protection allows to generate many uncorrelated secure templates from the same biometric template. Cross matching can be avoided and new functionalities as renewability, revocation are possible.

Different approaches of template protection exist. One of the ideas is to combine cryptography with error correction coding so that cryptographic hashing can be applied to noisy data like biometric data[1,2,].[3] Realization of this idea strongly depends on the characteristics of biometric features. For ordered features, where number of components are stable, the fuzzy commitment scheme is proposed as shown in[1,].[3] In[4,5,],[6] the helper data scheme is introduced to construct the fuzzy commitment. The security of this schemes is proved in.[7] The algorithm was integrated in 2D face recognition system using texture information,[8] fingerprints recognition system[9] and ear identification.[10] Their results are successful. For non-order features like minutiae of fingerprints, whose components vary and can not be described as a vector, the fuzzy vault scheme can be adopted.[11] Another possible approach is cancelable biometrics, which utilizes "non-invertible" function as scrambling, morphing[12,].[13] "Non-invertibility" of employing function must be strictly satisfied and authentication performance of biometrics should not be reduced.

In this paper, we describe a template protection using helper data scheme. Properties of key components in the scheme are described mathematically. The scheme is implemented in a face recognition system using 3D depth information. It is shown that the performance of this helper data scheme depends on the statistical analysis of the given biometric templates. Estimation of statistical characteristics should be adapted to input data. In section 2, the detailed description of the helper data scheme is given. The functionality of the kernel components such as binarization and selecting reliable component are presented. The scheme is integrated in 3D face recognition system and the results is evaluated in section 3. A conclusion is given in section 4.

2. TEMPLATE PROTECTION USING HELPER DATA SCHEME

In this section, the template protection using helper data scheme is presented. The helper data scheme can extract secure templates from biometric data. This secure template is stable to biometric variation and it is impossible to retrieve original biometric information from it. The mathematical formulation of these properties is summarized as delta-contracting and epsilon-revealing by J. P. Linnartz et al.[6] The block diagram of the helper data scheme is depicted in figure 1.

It is assumed that M is a biometric template extracted from a biometric measurement. In the enrollment process, the binarization converts the biometric template M into a binary vector Q. The binarization should make the resulting binary string uniformly distributed for different users and invariant for identical user. The detailed description of binarization is given in section 2.1. The random number generator creates randomly a secret code S, which is hashed and stored. Thus, it enables randomness in the system so that distinct references can be created from the same biometric characteristics for different applications. The error correction encoder adds redundancy in the secret S. As a consequence, the resulting codeword C is longer than S. Depending on property of bit errors, different error correction code can be adopted. When bit errors are uniformly distributed in codeword, a BCH- code, which has a codeword length of $2^L - 1$ (L is a nature number), can be employed. If the length of the bit string Q extends the length of the codeword C, then the most reliable bits in Q are selected so that the resulting binary string X is as long as the codeword C and robustness is improved. R indicates the position of reliable bits. W, the bitwise XOR of X and C, is so called helper data. With help of W, the same secret S can be extracted in verification process. Only the position vector R, the helper data W, the hashed secret code $h(S)$ and user identity information are stored in data storage. It can be proved that both W and

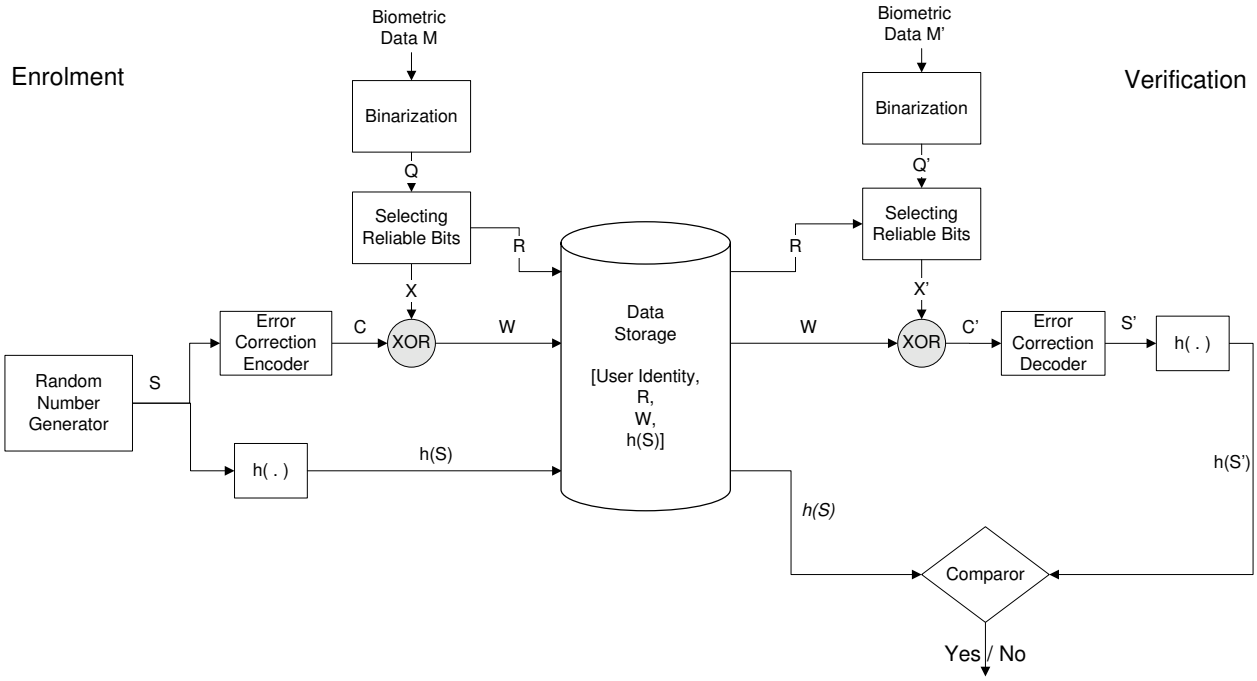

Figure 1. Block diagram of helper data scheme

$h(S)$ reveal little information about S, X as well as the biometric template M.[7]

During the verification process, with claimed identity, R, W and $h(S)$ are released from data storage. The binary string Q' is extracted from biometric template M', which in fact is M distorted by noise. The binary string X' is estimated with M' and R. A corrupted codeword C' can be acquired from W and X'. The following error correction decoder removes errors in C' and retrieves secret code S'. Comparing $h(S)$ with $h(S')$, a positive or negative response for a verification query can be given. In contrast to common biometrics system, only "hard decision" (rejected or accepted) is given and no similarity scores is available in comparer of template protection system due to hash function. Hill climbing attack, which iteratively reconstructs biometrics using matching scores[14,,15] can be prevented.

The security of the system is among others dependent on length of secret code. If the length of the codeword is fixed, the length of the secret code is restricted by the error correction ability. The maximum length of the codeword relies on the entropy for the considered biometric characteristics. Obviously, binarization and selecting reliable component are decisive to the performance of template protection. In the following sections we introduce their functionalities and construction.

2.1. Binarization

Binarization is the core component of helper data scheme. The requirements of its output binary vector can be summarized as follows. There, P denotes probability, B is the binarization function, M_i, M_j are biometric templates of user i and j, M_i' is the distorted biometric template of user i, Q is binary vector with length L and $\{0/1\}^L$ is L-dimensional binary vector space, δ is a small positive number:

1. For biometric template M of any user, a binary vector Q with length L exists.
2. Binary vectors should be uniformly distributed:

$$P\{B(M_i) = Q\} \approx \frac{1}{2^L}, \forall Q \in \{0/1\}^L \tag{1}$$

3. Statistical independence for binary vectors of different users:

$$P\{B(M_i)|B(M_j)\} \approx P\{B(M_i)\}, \quad \text{for} \quad i \neq j \tag{2}$$

4. Collision free for biometric data of different users:

$$P\{B(M_i) = B(M_j)\} = 0, \quad \text{for} \quad i \neq j \tag{3}$$

5. Binary vector of a specific user should be robust to biometric data variation:

$$P\{B(M_i) = B(M_i')\} = 1, \quad \text{if} \quad |M_i - M_i'| < \delta \tag{4}$$

Requirement 1 indicates versatility of binarization. Requirement 2 guarantees that no blind estimation of binary vector is possible and discriminability of binary vector is optimized. Requirement 2 and 3 ensure that no information of a user can be retrieved using binary vectors of other users. The last one indicates the desired noise resilience.

Moreover, binarization tries to extract a long binary vector from biometric template without any degradation of authentication performance. The construction of binarization depends on the statistic analysis of the input biometric templates. Assuming that a training template set contains N users and each user has K samples and $M_{n,k} = [m_{n,k,1}, m_{n,k,2}, \cdots, m_{n,k,T}]$ is the template with T components extracted from the k-th samples of the user n with $k \in \{1, \cdots, K\}$ and $n \in \{1, \cdots, N\}$. If each component is statistic independent and at least one bit can be extract from each component, the binarization function can be defined as:

$$q_{n,t} = B\{m_{n,k,t}|k \in [1, \cdots, K]\} = \begin{cases} 1 & \text{if} \quad \mu_{n,t} \geq \mu_t \\ 0 & \text{if} \quad \mu_{n,t} < \mu_t \end{cases} \tag{5}$$

where $\mu_{n,t}$ is an estimation of the real template for user n and μ_t is the threshold of binarization. In order to achieve uniform distribution of binary vector, μ_t should be the median of $\mu_{n,t}$ of all the users. Instead of median, mean can also be adopted. If the training data set is large enough, there is no significant difference between median and mean. In practice, we suggest to use median, which is resistant to extreme values caused by measure errors.

2.2. Selecting reliable bits

Selecting reliable bits contributes to the robustness of the system. It is based on the estimation of the error probability for each bit. Only the bits with the lowest error probability are selected. Error probability depends on the distance between $\mu_{n,t}$ and μ_t as shown in equation 5. $\mu_{n,t}$ of a relative stable bit should derive from μ_t. On the other hand, intra class variation is also decisive for error probability. The smaller the intra class variation is, the more reliable the corresponding bit is.

Statistical analysis of intra class characteristics for each user has a major effect on the performance of selecting reliable bits. If biometric templates are Gaussian distributed, then:

$$\mu_{n,t} = E\{m_{n,k,t}|k \in [1, \cdots, K]\} \tag{6}$$

$$p_{n,t} \propto \frac{|\mu_{n,t} - \mu_t|}{\sigma_{n,t}} \tag{7}$$

where E is the function calculating expected value, $p_{n,t}$ is the error probability of the t-th component of user n, $\sigma_{n,t}$ the standard deviation of $m_{n,k,t}$ for $k \in [1, \cdots, K]$ (see also[8]).

If it is not the case, assuming that distribution of templates is uniquely distributed and it is impossible to estimate intra class variation, then:

$$\mu_{n,t} = MEDIAN_{k=1}^{K}\{m_{n,k,t}\} \tag{8}$$

$$p_{n,t} \propto |\mu_{n,t} - \mu_t| \tag{9}$$

Actually, efficient estimation of error probabilities can only be achieved with sufficient number of samples. In the next section we show how the template protection using the helper data scheme is integrated in 3D face recognition system.

3. IMPLEMENTATION IN THE 3D FACE RECOGNITION SYSTEMS

We have implemented the template protection algorithm in 3D face recognition system. The 3D facial images of face recognition grant challenge (FRGC)[16] database version 1 are used as testing data. During the test, 99 users from all 289 users are chosen, who has at least 4 samples. Three samples per user are chosen as enrollment data and one sample as verification data. A different sample for the verification is chosen for each test and the tests are repeated 4 times.

In the feature extraction process, the 3D facial data is normalized to compensate the pose variation in the acquisition. The normalized 3D facial data is projected into regular grids. Then a fixed face region is selected for each resulting range image. The selected face region is divided into 68 sub-areas. A histogram-based extraction algorithm is applied in each sub-area. A feature vector containing $68 \times 6 = 408$ real values is obtained. The false acceptance rate (FAR) and false rejection rate (FRR) using the correlation classifier is plotted in figure 2. The equal error rate (EER) is equal to 3.38%.

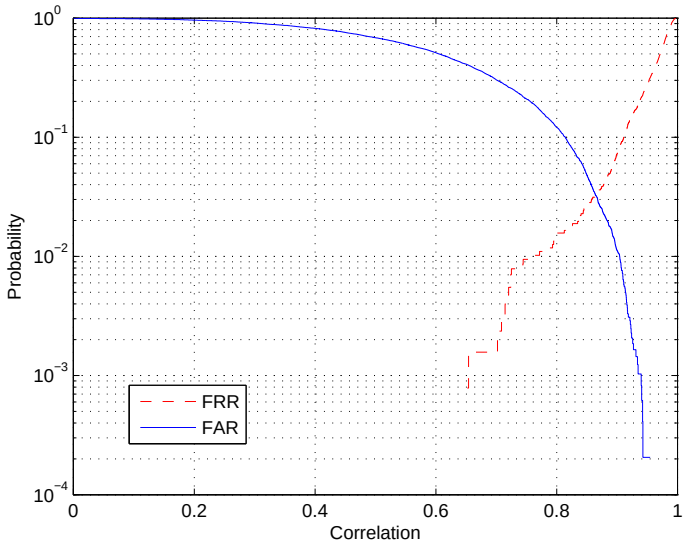

Figure 2. Classification results of the histogram- based face recognition algorithm

Then, we use the above mentioned binarization function to convert the extracted feature vectors into binary strings. To compare the authentication performance before and after binarization, we show the receiver operation characteristic (ROC) curves in figure 3. The solid line of the binary feature vectors is obviously above the dashed line of the real-valued feature vector. That is to say, binarization function improves slightly the authentication performance. Generally, a good binarization can be applied with acceptable changing on the authentication performance.

In the above binarization process, the median was adopted to calculate the binarization threshold. If we compare the FAR and FRR curves of the binarization using median (figure 4) and mean (figure 5), there is no significant difference regarding authentication performance. Both EER are around 3%, however, the FRR-curve of mean-based binary vectors deviates from the probability-axis in comparison with the one of median-based binary vectors. The median-based binarization has higher robustness to noise. This makes it better than mean-based binarization, since the performance of template protection is restricted by errors occuring in the binary feature

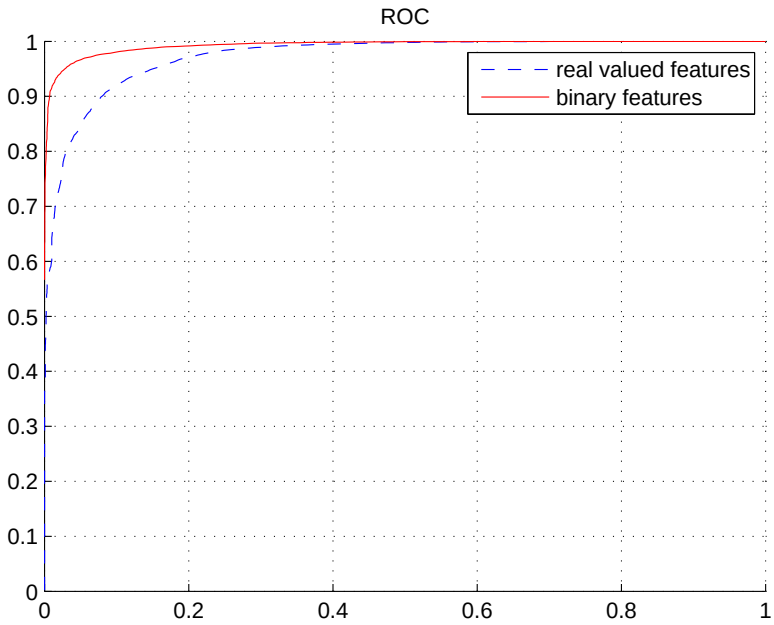

Figure 3. ROC curves of real-valued feature vectors and binary feature vectors

vectors.

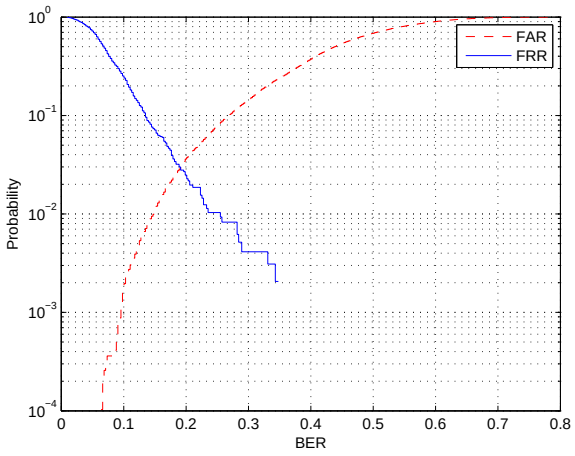

Figure 4. The classification results for the binary vectors using the median-based binarization

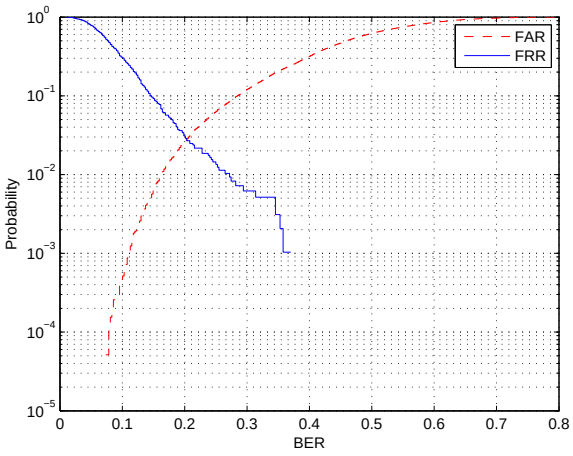

Figure 5. The classification results for the binary vectors using the mean-based binarization

In the implemented scheme, a BCH- code is chosen as error correction code. The maximum length of a codeword under 408 is 255. The 255 most reliable bits is chosen from the 408-bits long binary vector. The classification results under the assumption of uniquely distributed templates and Gaussian distributed templates are shown in figure 6 and in figure 7. Both classification results are similar. Under the assumption of uniquely distribution, the robustness is better than under the assumption of Gaussian distribution, however, the discriminative power is slightly worse.

With codeword of 255 bits, only discrete set of the secret code length s and the correctable errors length e is possible. Several examples and their corresponding bit error rate(BER), FRR and FAR is given in table 1. The FRR under the assumption of uniquely distribution is significantly better than under the assumption of Gaussian distribution, while its FAR decreases slightly.

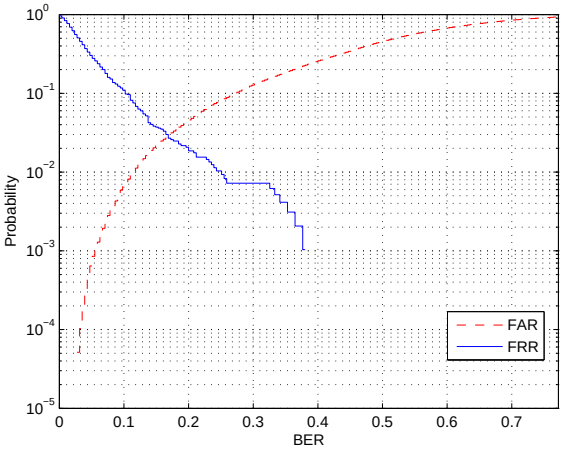

Figure 6. The classification results for the selected binary vectors under the assumption of uniquely distributed templates

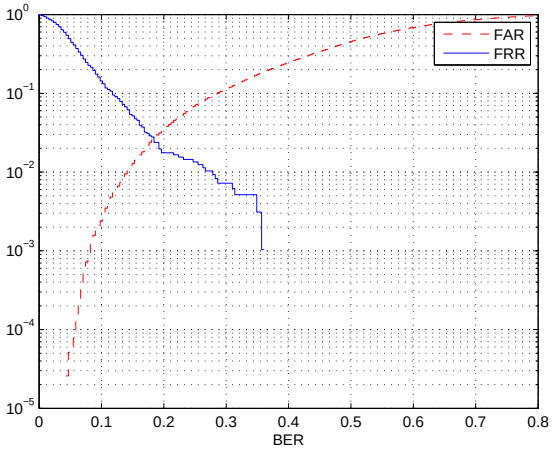

Figure 7. The classification results for the selected binary vectors under the assumption of Gaussian distributed templates

BCH $(c/s/e)$	Correctable BER	Results for uniquely distribution	Results for Gaussian distribution
255/107/22	8.6%	FRR=12%; FAR=0.4%	FRR=21%; FAR$\approx$ 0
255/91/25	9.8%	FRR=11%; FAR=0.6%	FRR=16%; FAR=0.2%
255/79/27	10.5%	FRR=10%; FAR=0.7%	FRR=13%; FAR=0.3%

Table 1. Examples of possible BCH codes and the corresponding FRR and FAR

4. CONCLUSION

This paper presented template protection technique using helper data scheme. The functionalities of the core components of helper data scheme like binarization and selecting reliable component are described in detail. Different realizations of the proposed algorithm are introduced and were tested with FRGC database using a histogram-based 3D face recognition algorithm.

As a conclusion, a large number of samples in enrollment is necessary to estimation the statistical character of biometric templates. In the histogram-based 3D face verification system, binarization does not impact the authentication performance. A high security using template protection can be achieved, when long secret code is enabled in the system.

ACKNOWLEDGMENTS

The work presented in this paper was supported in part by the European Commission under Contract 3D FACE, a European Integrated Project funded under the European Commission IST FP6 program. The author would like to thank Michiel von der Veen, Tom Kevenaar and Christoph Busch for their advices on this topic.

REFERENCES

1. A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *6th ACM Conference on Computer and Communications Security*, pp. pp. 28–36., (http://www.rsasecurity.com/rsalabs/node.asp?id=2048), 1999.
2. A. Juels and M.Sudan, "A fuzzy vault scheme," in *IEEE International Symposium on Information Theory*, (http://biometrics.cse.msu.edu/uludag-jain-fuzzy-fp.pdf), 2002.
3. Y. Dodi, L. Reyzin, and A. Smith, "Fuzzy extrators: How to generate strong secret keys from biometrics and other noisy data," in *In Advances in cryptology - Eurocrypt'04*, L. 3027, ed., pp. 523–540, 2004.
4. P. Tulys, "Privacy protection of biometric templates: cryptography on noisy data," in *Revue HF (Rev. HF) ISSN 0035-3248*, *no3*, pp. 55–64, 2004.
5. P. Tuyls, A. M. Akkermans, T. Kevenaar, G.J.Schrijen, A.M.Bazen, and R.N.J.Veldhuis, "Practical biometric authentication with template protection," in *Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pp. 436–446, 2005.
6. J. P. Linnartz and P. Tuyls, "New shiedling functions to enhance privacy and prevent misuse of biometric templates," in *4th international conference on audio- and video-based biometric person authentication*, 2003.
7. P. Tuyls and J. Goseling, "Capacity and examples of template protecting biometric authentication systems," in *Biometric authentication workshop (BioAW 2004)*, LNCS, ed., (3087), pp. 158–170, Prague, 2004.
8. M. van der Veen, T. Kevenaar, G.-J. Schrijen, T. H. Akkermans, and Z. Fei, "Face biometrics with renewable templates," in *Security, Steganography, and Watermarking of Multimedia Contents VIII*, 1519 January 2006, San Jose, California, USA.
9. E. Verbitskiy, P. Tuyls, D. Denteneer, and J.-P. Linnartz, "Reliable biometric authentication with privacy protection," *24th Benelux Symp. on Info. Theory* , 2003.
10. P. Tuyls, E. Verbitskiy, T. Ignatenko, D. Schobben, and T. H. Akkermans, "Privacy protected biometric templates: ear identification," in *Proceeding of SPIE*, **5404**, pp. 176–182, April 2004.
11. U. Uludag and A. Jain, "Fuzzy fingerprint vault," in *Workshop: Biometrics: Challenges Arising from Theory to Practice*, (citeseer.ist.psu.edu/uludag04fuzzy.html), August 2004.
12. N. Ratha, J. Connell, and R. Bolle, "Enhancing security and privacy of biometric-based authentication systems," *IBM Systems Journal* **40**, p. No. 3, 2002.
13. R. Bolle, J. H. Connell, and N. Ratha, "System and method for distorting a biometric for transactions with enhanced security and privacy." US 6836554 B1, Dec 2004.
14. A. Adler, "Reconstruction of source images from quantized biometric match score data," in *In Biometrics Conference, Washington, DC*, September 2004.
15. C. Soutar, "Biometric system security," in *Information Technology Security Symposium*, *Biometric technology updates* , 2002.
16. "Face recognition grant challenge (frgc)," in *National Institute of Standards and Technology (NIST)*, (http://face.nist.gov/frgc/).