Multi-Sample Fusion with Template Protection

E.J.C. Kelkboom and J. Breebaart
Philips Research, The Netherlands
{Emile.Kelkboom, Jeroen.Breebaart}@philips.com

R.N.J. Veldhuis
University of Twente, Fac. EEMCS, The Netherlands
R.N.J.Veldhuis@utwente.nl

X. Zhou and C. Busch
Fraunhofer Institute for Computer Graphics Research IGD, Germany
{Xuebing.Zhou, Christoph.Busch}@igd.fraunhofer.de

Abstract: The widespread use of biometrics and its increased popularity introduces privacy risks. In order to mitigate these risks, solutions such as the helper-data system, fuzzy vault, fuzzy extractors, and cancelable biometrics were introduced, also known as the field of template protection. Besides these developments, fusion of multiple sources of biometric information have shown to improve the verification performance of the biometric system. Our work consists of analyzing feature-level fusion in the context of the template protection framework using the helper-data system. We verify the results using the FRGC v2 database and two feature extraction algorithms.

1 Introduction

More applications are using biometrics ranging from simple home or business applications with a small and limited group of enrolled people (for example access control to buildings or rooms) to large-scale systems used by an entire nation or even the entire world (for example identity cards with biometrics or the electronic passport e-Passport). Unfortunately, its widespread use increases the related privacy risks such as identity theft or activity monitoring by cross-matching between biometric databases of different applications. However, the field of template protection provides the technology that enables the mitigation of these privacy risks by transforming the biometric template with a one-way operation in order to guarantee the irreversibility property and by randomizing the biometric template that guarantees that multiple protected templates from the same biometric sample cannot be linked to each other. In the literature, different types of technologies have been presented, for example the *Helper-Data Systems* (HDS) [KGK+07, KSA+05, TAK+05], *Fuzzy Vaults* [JS02, NJP07], *Fuzzy Extractors* [CR07, DRS04], and *Cancelable Biometrics* [RCCB07].

Besides the template protection developments, fusion of multiple sources of biometric information has shown to improve the verification performance of the biometric system. As

described in [RNJ06], the basic principle of fusion is the reconciliation of evidence presented by multiple sources of biometric information in order to enhance the classification performance. Furthermore, different sources of biometric information can be extracted from the same biometric modality by: (i) capturing a sample of multiple instances (left and right index fingerprint or iris) with the same sensor, (ii) using different types of sensors to acquire a different biometric sample from the same instance, (iii) capturing several samples using the same sensor and instance, and (iv) extracting dissimilar feature representations of the same biometric sample using different algorithms. These cases are referred to as the multi-instance, multi-sensor, multi-sample, and multi-algorithm systems, respectively. Furthermore, the fifth type is the multi-modal system, which is the fusion of sources of biometric information from multiple modalities, for example fingerprint, face, iris, voice, palm or retina. To complete the summary from [RNJ06], the sixth type is referred to as the hybrid system, which consists of a combination of the aforementioned fusion types. The most common implementations of multi-biometric systems address fusion at the feature-level, score-level or decision-level.

In the work of [NJ08], the Fuzzy Vault template protection system is used for applying multi-sample, multi-instance, and multi-modal fusion. In case of multi-sample fusion, they create a single mosaiced template from multiple fingerprint impressions from which they construct the vault. For multi-instance fusion they take the union of the minutiae sets of the left and right index fingers for constructing the vault. For multi-modal fusion, a fingerprint and a iris sample are combined by concatenating the unordered minutiae set with the transformed iriscode extracted from the fingerprint and iris samples, respectively. The vault is constructed using the concatenated unordered set. The verification performance has improved for all three cases as well as the claimed security.

Furthermore, the works of [KGK$^+$07, KSA$^+$05, LT03] based on the HDS template protection system inherently apply multi-sample fusion at feature-level by averaging the multiple enrolment samples. However, no arguments are provided for applying feature-level fusion instead of either score-level or decision-level.

Our work also consists of applying multi-sample fusion using the HDS, but we analyze the performance improvements of fusion at feature-, score-, and decision-level fusion. We use 3D face range images of the FRGC v2 dataset [PFS$^+$05] and verify the performance improvement on two recognition algorithms.

The outline of this paper is as follows. In Section 2 we briefly discuss the HDS system, while in Section 3 we discuss the application of multi-sample fusion at feature-, score-, and decision-level using the HDS system together with the experimental setup and results. We finish with the conclusions in Section 4.

2 Template Protection Scheme

In the literature, many presented template protection schemes are based on the capability of generating a robust binary vector or key from biometric measurements of the same subject. This also holds for the HDS system we consider and is depicted in Figure 1. For the sake of

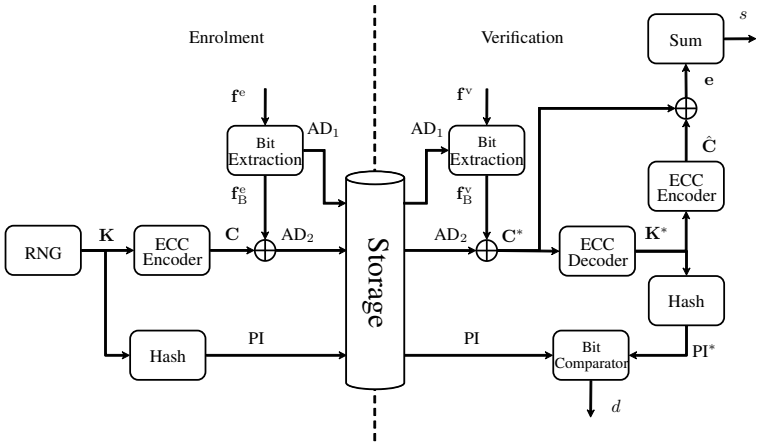


Figure 1: The HDS template protection scheme.

coherence we use the terminology *auxiliary data* (AD) and *pseudo identity* (PI) proposed in [BBGK08], which is in line with standardization activities in ISO [ISO09]. Within the *Bit Extraction* module, a binary vector $\mathbf{f}_{\mathrm{B}} \in \{0,1\}^{N_{\mathrm{B}}}$ is extracted from the real-valued representation of the biometric sample, $\mathbf{f} \in \mathbb{R}^{N_{\mathrm{F}}}$. We use a single bit quantization scheme based on thresholding and the *reliable component selection* (RCS) algorithm. We select the N_{B} most reliable components based on the estimated z-score of each component. With use of the multiple (N_{e}) enrolment samples, the z-score is estimated as the ratio between the distance of the estimated mean with respect to the quantization threshold and the estimated standard deviation, see [KGK+07] for a more detailed description of the z-score estimation and the quantization scheme. The index information of the selected reliable components is stored as auxiliary data AD_1.

The binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ could be used as a key for any encryption purposes, however it is not considered as being practical because of the high probability that it is not exactly the same in both the enrolment and verification phase ($\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \neq \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$), due to measurement noise and biometric variability that lead to *bit errors*. The number of bit errors is also referred to as the Hamming distance $d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}})$. To deal with the bit errors, we use error-correcting codes (ECC). The combination of the ECC with a cryptographic hash function forms the scheme also known as the Fuzzy Commitment scheme [JW99]. In the enrolment phase, a binary secret or message vector $\mathbf{K} \in \{0,1\}^{k_{\mathrm{c}}}$ is randomly generated by the *Random-Number-Generator* (*RNG*) module. A codeword $\mathbf{C}$ of an error-correcting code is obtained by encoding $\mathbf{K}$ in the *ECC-Encoder* module. As the ECC we use the linear block type code "Bose, Ray-Chaudhuri, Hocquenghem" (BCH) [BRC60], which is specified by the codeword length (n_{c}), secret length (k_{c}), and the corresponding number of bits that can be corrected (t_{c}), in short $[n_{\mathrm{c}}, k_{\mathrm{c}}, t_{\mathrm{c}}]$. Some practical BCH settings are provided in Table 1, where the bit error rate (BER) is the ratio $t_{\mathrm{c}}/n_{\mathrm{c}}$. The codeword is XOR-ed with $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ in order to obtain auxiliary data AD_2. Hence, $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ should have the same dimension as $\mathbf{C}$, implying $N_{\mathrm{B}} = n_{\mathrm{c}}$. Furthermore, the hash of $\mathbf{K}$ is taken in order to obtain the pseudo identity PI. Under the assumption that the bits of $\mathbf{f}_{\mathrm{B}}$ are independent, from [TG] we can use the secret

size k_c as a measurement of the difficulty of guessing the enrollment binary vector $\mathbf{f}_B^e$ from the protected template $\{AD_1, AD_2, PI\}$, hence safeguarding the privacy. The larger the secret size the more difficult it is to either guess $\mathbf{f}_B^e$ or $\mathbf{K}$ from PI.

In the verification phase, a new biometric sample is taken and transformed into its binary representation within the *Bit Extraction* module with help of auxiliary data AD_1. The new word $\mathbf{C}^*$ is computed by XOR-ing $\mathbf{f}_B^v$ with AD_2, and for a genuine case it is expected that $\mathbf{C}^*$ is close to $\mathbf{C}$. The candidate secret $\mathbf{K}^*$ is obtained by decoding $\mathbf{C}^*$ in the *ECC-Decoder* module. Subsequently, the candidate pseudo identity PI^* is computed by hashing $\mathbf{K}^*$. The decision in the *Bit-Comparator* module is based on whether PI and PI^* are bitwise identical.

The *Bit-Comparator* module outputs a match as its decision d only if PI and PI^* are identical, which occurs when the number of bit errors between the binary vectors $\mathbf{f}_B^e$ and $\mathbf{f}_B^v$ is smaller or equal to the error-correcting capability t_c of the ECC. Thus, there is a match when the Hamming distance is smaller than t_c, $d_H(\mathbf{f}_B^e, \mathbf{f}_B^v) = ||\mathbf{f}_B^e \oplus \mathbf{f}_B^v||_1 \leq t_c$. Therefore, the fuzzy commitment scheme can be considered as a Hamming distance classifier with threshold t_c. Note, that the maximum number of bits that the BCH can correct t_c^* is close to 25% of the codeword length. In the remainder of the text, we indicate this limitation as the *ECC-limitation*.

As a distance score s we use the number of bits that had to be corrected by the ECC decoder. The candidate secret $\mathbf{K}^*$ is encoded to its corresponding codeword $\hat{\mathbf{C}}$ and is XOR-ed with $\mathbf{C}^*$ in order to obtain the error pattern $\mathbf{e}$. The error pattern is equal to the bit differences between the enrolment and verification binary feature vectors ($\mathbf{f}_B^e \oplus \mathbf{f}_B^v$) as follows

$$\begin{aligned} \mathbf{e} &= \hat{\mathbf{C}} \oplus \mathbf{C}^* \\ &= \hat{\mathbf{C}} \oplus (\mathbf{f}_B^v \oplus AD_2) \\ &= \hat{\mathbf{C}} \oplus (\mathbf{f}_B^v \oplus (\mathbf{f}_B^e \oplus \mathbf{C})) \\ &= (\hat{\mathbf{C}} \oplus \mathbf{C}) \oplus (\mathbf{f}_B^e \oplus \mathbf{f}_B^v) \\ &= (\mathbf{f}_B^e \oplus \mathbf{f}_B^v) \text{ if } \hat{\mathbf{C}} = \mathbf{C}, \end{aligned} \tag{1}$$

where $\hat{\mathbf{C}}$ is equal to $\mathbf{C}$ when there is a match, i.e. $\mathbf{K}$ and $\mathbf{K}^*$ are equal. The distance score s is thus the sum of the error pattern, hence equal to $d_H(\mathbf{f}_B^e, \mathbf{f}_B^v)$ and only a valid score when there is a match, i.e. $d_H(\mathbf{f}_B^e, \mathbf{f}_B^v) \leq t_c$. If the score is not valid we only know that $d_H(\mathbf{f}_B^e, \mathbf{f}_B^v) > t_c$.

Table 1: Some examples of the BCH code given by the codeword (n_c) and secret (k_c) length, the corresponding number of correctable bits (t_c), and the bit error rate (BER) t_c/n_c.

n_c	k_c	t_c	BER $= t_c/n_c$
127	8	31	24.4%
	15	27	21.3%
255	9	63	24.7%
	21	55	21.6%
511	10	127	24.9%
	31	109	21.3%

3 Experiments

In this section we present the methods for multi-sample fusion at feature-, score-, and decision-level and empirically validate the best performance achieved at each level by means of a biometric database and two feature extraction algorithms.

3.1 Experiment Setup

3.1.1 Biometric Databases

All the results in this work are obtained using the FRGC v2 dataset [PFS+05] containing a total of 4007 3D shape samples from 465 subjects.

However, one of the two 3D shape recognizers we used could not successfully extract a feature vector out of each sample, hence reducing the dataset to 3507 samples from 454 subjects. As the template protection algorithm works best at multiple enrolment samples, the subset of subjects with at least 6 (5 as enrolment samples with at least one for the verification) samples or more is created. This resulted into a subset of 261 subjects with in total 2970 samples.

3.1.2 Feature Extraction Algorithms

The first algorithm is the shape-based 3D face recognizer from [GIA06] and is referred to as Algo1. It has two main steps: 1) the alignment of faces, and 2) the extraction of surface features from 3D facial data. In the alignment step, each face is registered to a generic face model (GFM) and the central facial region is cropped. The GFM is computed by averaging correctly aligned images from a training set. After the alignment step, we can assume that all faces are transformed in such a way that they best fit the GFM, and have the same position in the common coordinate system.

After alignment, the facial surface is divided into 174 local regions. For each region, the maximum and minimum principal curvature direction are computed. Each of the two directions is presented by the azimuthal and the polar angle in the spherical coordinate system. Combining all the regions leads to a feature vector dimension $N_{\mathrm{F}} = 174 \times 2 \times 2 = 696$.

The second algorithm, Algo2, is a histogram-based feature extraction method. After the pre-registration of the face data, a frontal view of the face model is obtained, where the tip of the nose is at the origin in the Cartesian coordinate system. The distribution of depth values of the normalized face model describes the characteristics of an individual facial surface. In order to obtain more detailed information about the local geometry, the 3D model is divided into several sub areas which are orthogonal to the symmetry plane of the face. The features are extracted from the depth value distribution in each sub-area. The feature vector dimension is $N_{\mathrm{F}} = 476$. A full description of this algorithm is provided in [ZSBF08].

For both feature extraction algorithms, the raw feature vectors they produce are used as input of the template protection system as described in Section 2. Hence, no further signal processing is performed.

3.1.3 Testing Protocols

The performance testing protocol consists of randomly selecting 50% (130) subjects as the training set and the other subjects as the test set, this is referred to as the training-test-set split. The template protection system parameters such as the quantization thresholds, used within the *Bit Extraction* module, are estimated on this training set. Hereafter, the test set is randomly split into an equally sized fusion-training and evaluation set containing around 65 subjects each. All the training needed for fusion is thus performed on the fusion-training set and the reported performance is obtained from the evaluation set. From the evaluation set, 5 samples of each subject are randomly selected as the enrolment samples while the remaining samples are considered as the verification samples. This split is referred to as the enrolment-verification split. The protected template is generated using all the enrolment samples and compared with each verification sample.

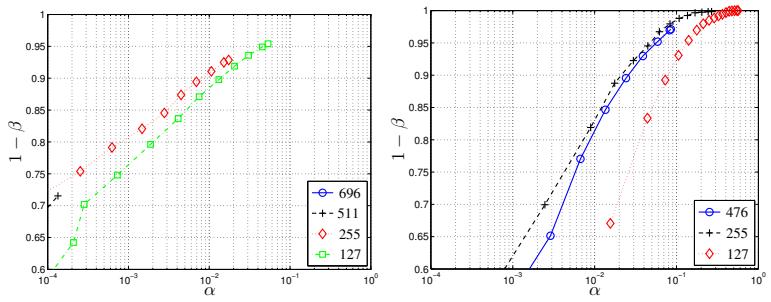
The training-test-set split is performed five times, while for each split the enrolment-verification split is performed five times. From each enrollment-verification split we measure the β_{tar} (the false non-match rate (FNMR, β) at the targeted false match rate (FMR, α) of $\alpha_{tar} = 0.25\%$) and the equal-error rate (EER), which is the error rate achieved at the operating point where both FNMR and FMR are equal. With use of the 25 measurements we estimate the 95% confidence interval (ci) defined as $ci = 1.96\sigma_{EER}/\sqrt{(25)}$ for the EER case while using $\sigma_{\beta_{tar}}$ for the β_{tar} case, respectively. Note, that the splits are performed randomly, however the seed at the start of the protocol is always the same, hence all the splits are equal for the performance tests at feature-, score-, and decision-level fusion. Hence, the splitting process does not contribute to any performance differences.

3.2 Experiment Results

3.2.1 Feature-level Fusion

Similar to the works [KGK$^+$07, KSA$^+$05, LT03], we average the $N_{\mathrm{e}} = 5$ enrolment samples before entering the template protection scheme. By averaging the samples the measurement noise and the biometric variability are suppressed. Hence there will be less bit-errors and the binary representation will be more robust.

The achieved performances for different n_{c} settings are portrayed by the ROC curves in Figure 2(a) and (b) for algorithms Algo1 and Algo2, respectively. Furthermore, the EER and β_{tar} details are given in Table 2. The table provides the ci for both EER and β_{tar} and their operating point provided as the relative Hamming distance (RHD). The right column of the table provides the effective secret size $|\mathbf{K}_f|$ of the template protection system at the specific fusion level. Because a single protected template is created at feature-level



(a) Algo1 (b) Algo2

Figure 2: ROC curves at feature-level fusion for different n_c settings for the (a) Algo1 and (b) Algo2 algorithm.

fusion, $|\mathbf{K}_f|$ is equal to k_c of the ECC. On the other hand, k_c is determined by the t_c setting that leads to a α close to the target α_{tar}, but smaller. This is exactly the ECC setting with a BER just larger than the operating point in RHD corresponding to β_{tar}. Entries in the table indicated with quotes cannot be reached in practice because of the ECC-limitation, however we are able to estimate them because of the Hamming distance classifier assumption as discussed in Section 2. Entries with "x" can neither be reached nor estimated.

Note that the ROC curves are limited because of the ECC-limitation. In order to reach larger α and smaller β values it is required to tolerate and thus correct more bit errors. However, the error correcting capability of an ECC is limited. From the results we can conclude that both algorithms perform optimally at a codeword size of $n_c = 255$. These settings are used in the score- and decision-level fusion analysis. Compared to the Algo2 algorithm, Algo1 has a better performance but a smaller secret size (see Table 2, right column).

Table 2: The EER and β_{tar}, and their ci and operating point for the individual algorithms Algo1 and Algo2 at different settings of n_c. The last column is the effective secret size $|\mathbf{K}_f|$ which is equal to the secret size k_c of the ECC at the operating point t_c for achieving α_{tar}.

n_c	*EER* [%]	*RHD* [%]	β_{tar} [%]	*RHD* [%]	$\lvert\mathbf{K}_f\rvert$ [bits]
Algo1					
696	"3.76 ± 0.25"	"38.8"	"16.13 ± 1.93"	"33.62"	x
511	"3.69 ± 0.30"	"35.2"	"15.19 ± 1.79"	"28.77"	x
255	"4.02 ± 0.41"	"27.5"	15.84 ± 2.10	19.61	21
127	4.88 ± 0.47	23.6	18.95 ± 2.01	14.96	29
Algo2					
476	5.44 ± 0.35	22.1	37.69 ± 3.14	11.76	x
255	5.06 ± 0.30	10.2	30.25 ± 2.88	1.96	215
127	8.92 ± 0.33	3.9	89.57 ± 1.20	0.00	120

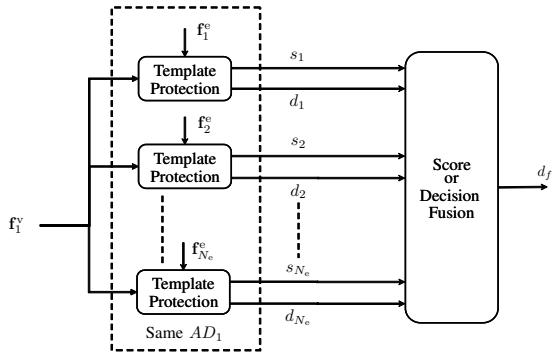
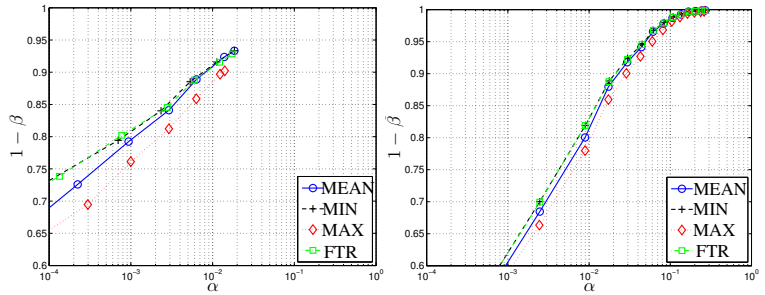


Figure 3: The general implementation of multi-sample fusion at score- or decision-level.

3.2.2 Score-Level Fusion

A general implementation of the template protection system at score- or decision-level fusion is depicted in Figure 3. A protected template is created for each of the N_{e} enrolment samples. Note that the RCS quantization scheme as discussed in Section 2 uses multiple enrolment samples in order to estimate the necessary statistics, hence we use all the N_{e} enrolment samples to determine the N_{B} most reliable components and is used as such in each N_{e} template protection systems portrayed in Figure 3. Within the *Score- or Decision-level Fusion* module the scores $\{s_1, s_2, \ldots, s_{N_{\mathrm{e}}}\}$ are combined into a single fused score s_f from which the decision d_f is taken based on a score threshold. Note that a score is valid only when there is a match from the corresponding template protection system and occurs when $s_i \leq t_{\mathrm{c}}$. Therefore we set the error-correcting capacity t_{c} to its maximum (t_{c}^*) in order to obtain a valid score for the largest range possible. Consequently, the secret size used for each of the N_{e} protected templates is equal to nine bits and does not depend on the score threshold. Hence, at score-level fusion the score threshold determines the operating point of the ROC curve and not the ECC setting. Combination methods such as the minimum (MIN), the maximum (MAX), and the mean (MEAN) of the scores are used in order to obtain s_f. For the MEAN method we take the mean of the valid scores only, while the MIN and MAX methods are based on all the scores. We take the maximum based on all the scores because if there is a single invalid score it should lead to a non-match. Furthermore, for each method, if all the scores are not valid it will automatically lead to a non-match.

The ROC curves at the optimal setting of $n_{\mathrm{c}} = 255$ are depicted in Figure 4 with the details in Table 3. As a comparison, we included the ROC curve obtained at feature-level fusion indicated as "FTR". Because it suffices to guess a single $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ from one of the N_{e} protected templates to breach your privacy, the effective secret size $|\mathbf{K}_f|$ of the template protection system at score-level fusion for each method is also nine bits. Consequently we have omitted them from the table. The results indicate that taking the MIN method leads to the best performance, however the difference is not significant when considering the ci. Furthermore, the MIN method ROC curve is very close to the ROC from feature-level fusion (FTR). Note that for the Algo1 algorithm it is not possible to estimate the EER for



(a) Algo1 (b) Algo2

Figure 4: ROC curves at score-level fusion compared to the feature-level (FTR) curves for the (a) Algo1 and (b) Algo2 algorithm.

Table 3: The EER and β_{tar}, and their ci and operating point for the score-level fusion experiments with $n_{\mathrm{c}} = 255$.

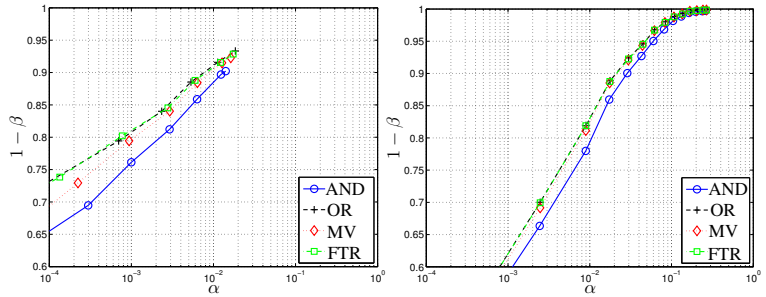
Method	*EER* [%]	*RHD* [%]	β_{tar} [%]	*RHD* [%]
Algo1, $n_{\mathrm{c}} = 255$				
MEAN	x	x	16.45 ± 2.08	20.00
MIN	x	x	15.74 ± 2.09	19.61
MAX	x	x	19.48 ± 2.08	20.39
Algo2, $n_{\mathrm{c}} = 255$				
MEAN	4.96 ± 0.28	10.6	31.46 ± 3.23	2.35
MIN	4.87 ± 0.30	10.2	29.90 ± 3.29	2.35
MAX	5.49 ± 0.29	11.4	33.49 ± 3.08	2.35

all the methods, because the EER is at an operating point greater than t_{c}^{*}, hence there are no valid scores.

We also observed that the ROC curves, especially for Algo2, are very similar. At further analysis we discovered that the ROC curves converge to a single one when decreasing n_{c}. This can be explained as follows. When selecting the most reliable components many enrolment samples from the same subject have an identical binary representation $\mathbf{f}_{\mathrm{B}}$. For example, for the $n_{\mathrm{c}} = 255$ case 75% of the enrolled subjects have no differences between the binary representation $\mathbf{f}_{\mathrm{B}}$ of its N_{e} enrolled samples for the Algo1 algorithm and 92% for the Algo2 algorithm, respectively. For the $n_{\mathrm{c}} = 127$ case, the likelihood increases to 99% and 100%, respectively.

3.2.3 Multi-Sample Fusion at Decision Level

Similar to the score-level fusion case a protected template is created for each N_{e} samples and compared with the single verification sample. However, the *Score- or Decision-level Fusion* module combines the decision $\{d_1, d_2, \ldots, d_{N_{\mathrm{e}}}\}$ into a single fused decision d_f. Methods such as the OR-rule, AND-rule, and majority voting (MV) are used in order to obtain d_f. For the AND-rule method, all the decisions have to be a match in order for the



(a) Algo1 (b) Algo2

Figure 5: ROC curves at decision-level fusion compared to the feature-level (FTR) curves for the (a) Algo1 and (b) Algo2 algorithm.

Table 4: The EER and β_{tar}, and their ci and operating point, and the effective secret size $|\mathbf{K}_f|$ for the decision-level fusion experiments with $n_c = 255$.

Method	*EER* [%]	*RHD* [%]	β_{tar} [%]	*RHD* [%]	$\lvert\mathbf{K}_f\rvert$ [bits]
Algo1, $n_c = 255$					
AND	"4.76 ± 0.40"	"29.0"	19.48 ± 2.08	20.39	21
OR	"3.95 ± 0.39"	"27.1"	15.74 ± 2.09	19.61	21
MV	"4.11 ± 0.44"	"27.8"	16.62 ± 2.05	20.00	21
Algo2, $n_c = 255$					
AND	5.49 ± 0.29	11.4	33.49 ± 3.08	2.35	207
OR	4.87 ± 0.30	10.2	29.90 ± 3.29	2.35	207
MV	4.89 ± 0.28	10.2	30.78 ± 3.27	2.35	207

final one to be a match too, while for the OR-rule case only a single match leads to a final match. For the MV method more than half of the decisions should be a match in order to have a final match.

Again, it suffices to break a single protected template for the adversary to know $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$, hence the effective secret size $|\mathbf{K}_f|$ is equal to the secret k_c corresponding to the ECC setting.

The experimental results are portrayed in Figure 5 with the performance details in Table 4. As a comparison, we included the ROC curve obtained at feature-level fusion indicated as "FTR". From these results we can conclude that the OR-rule fusion method consistently leads to a better performance, followed by the MV method, and the worst performance is with the AND-rue method. However, the difference is not significant. Compared to feature-level fusion results, the OR-rule methods leads to a similar ROC curve. The ROC curves, especially for the Algo2 algorithm, are very similar due to the same reason as as discussed in the previous section where it was noticed that the reliable binary representation $\mathbf{f}_{\mathrm{B}}$ is very similar for every N_e samples.

3.2.4 Summary and Discussions

We have compared performances of multi-sample fusion at feature-, score-, and decision-level. At the optimal setting of $n_c = 255$ we do not observe a significant performance differences between either feature-, score-, and decision-level fusion method. The effective secret size $|\mathbf{K}_f|$ is the same at feature- and decision-level fusion, and at its smallest at score-level fusion. Taking into account that at score and decision level fusion a protected template has to be made and stored for each N_e enrolment sample but only a single one at feature level, we can conclude that the best multi-sample fusion method is at feature level. For security and privacy reasons it is also not desired to store multiple protected templates, which could facilitate the attacker with hacking the protected template and either obtain the secret or the biometric data itself. Furthermore, a single protected template has a smaller storage capacity requirement.

When carefully analyzing the score- and decision-level fusion results, we can also conclude that the MIN-score and OR-decision methods have precisely the same performance, similarly for the MAX-score and AND-decision methods. The explanation for the MAX-score and AND-decision case is that if the maximum score is a match it would imply that all the other $N_e - 1$ scores are also a match, which is also the requirement for the AND-decision fusion method. The MIN-score and OR-decision performance similarity can be explained by the fact that both methods require at least a single individual comparison to be a match in order for the final decision to be a match.

4 Conclusions

With this work we have shown that it is possible to apply multi-sample fusion with the HDS system at feature-, score-, and decision-level. Because the HDS system inherently has only a decision as the output, we adapted the system accordingly in order to have a score as output for the score-level fusion. As a distance score we took the number of bits the ECC had to correct. Furthermore, applying fusion with template protection at feature- or decision-level is straightforward and conventional. However, fusion at score-level is different due to the use of an ECC, which has a limited error-correcting capability. Consequently, for each template protection system there is only a valid score when there is a match.

Given the biometric database and feature extraction algorithms, our experimental results showed that at the optimal setting of $n_c = 255$ there are no significant differences between the best performance (ROC curves) obtained at feature-, score-, and decision-level. Because at feature-level fusion only a single protected template is created, which is better in terms privacy and security protection and storage, we can conclude that the optimal multi-sample fusion is at feature-level.

Acknowledgment

The authors would like to acknowledge the support of the partners within the 3DFACE project, a European Integrated Project funded under the European Commission IST FP6 program.

References

[BBGK08] Jeroen Breebaart, Christoph Busch, Justine Grave, and Els Kindt. A Reference Architecture for Biometric Template Protection based on Pseudo Identities. In *BIOSIG*, Darmstadt, Germany, September 2008.

[BRC60] R. C. Bose and D. K. Ray-Chaudhuri. On a class of error correcting binary group codes. *Information and Control*, 3(1):68–79, March 1960.

[CR07] Ee-Chien Chang and Sujoy Roy. Robust Extraction of Secret Bits from Minutiae. In *Int. Conf. on Biometrics*, pages 750–759, Seoul, South Korea, August 2007.

[DRS04] Y. Dodis, L. Reyzin, and A. Smith. Fuzzy Extractors: How to generate strong secret keys from biometrics and other noisy data. In *Advances in Cryptology - Eurocrypt 2004, LNCS 3027*, pages 532–540, 2004.

[GIA06] Berk Gökberk, M. Okan Irfanoglu, and Lale Akarun. 3D Shape-based Face Representation and Feature Extraction for Face Recognition. *Image and Vision Computing*, 24(8):857–869, August 2006.

[ISO09] ISO/IEC JTC1 SC27. CD 24745 - Information technology - Security techniques - Biometric template protection, 2009.

[JS02] Ari Juels and Madhu Sudan. A Fuzzy Vault Scheme. In *Proc. of the 2002 International Symposium on Information Theory (ISIT 2002)*, Lausanne, 2002.

[JW99] Ari Juels and Martin Wattenberg. A Fuzzy Commitment Scheme. In *6th ACM Conference on Computer and Communications Security*, pages 28–36, November 1999.

[KGK+07] Emile J. C. Kelkboom, Berk Gökberk, T. A. M. Kevenaar, A. H. M. Akkermans, and M. van der Veen. "3D Face": Biometric Template protection for 3D Face Recognition. In *Int. Conf. on Biometrics*, pages 566–573, Seoul, Korea, August 2007.

[KSA+05] Tom A. M. Kevenaar, Geert-Jan Schrijen, Antonius H. M. Akkermans, Michiel van der Veen, and Fei Zuo. Face Recognition with Renewable and Privacy Preserving Binary Templates. In *4th IEEE workshop on AutoID*, pages 21–26, Buffalo, New York, USA, October 2005.

[LT03] Jean-Paul Linnartz and Pim Tuyls. New Shielding Functions to Enhance Privacy and Prevent Misuse of Biometric Templates. In *4th Int. Conf. on AVBPA*, 2003.

[NJ08] Karthik Nandakumar and Anil K. Jain. Multibiometric Template Security Using Fuzzy Vault. In *International Conference on Biometrics: Theory, Applications and Systems*, pages 1–6, 2008.

[NJP07] K. Nandakumar, A. K. Jain, and S. Pankanti. Fingerprint-based Fuzzy Vault: Implementation and Performance. In *IEEE Transactions on Information Forensics and Security*, pages 744–757, Decmber 2007.

[PFS+05] P. Jonathon Phillips, Patrick J. Flynn, Todd Scruggs, Kevin W. Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek. Overview of the Face Recognition Grand Challenge. In *IEEE CVPR*, volume 2, pages 454–461, June 2005.

[RCCB07] Nalini K. Ratha, Sharat Chikkerur, Jonathan H. Connell, and Ruud M. Bolle. Generating Cancelable Fingerprint Templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):561–572, April 2007.

[RNJ06] Arun A. Ross, Karthik Nandakumar, and Anil K. Jain. *Handbook of Multibiometrics*. Springer, 2006.

[TAK+05] Pim Tuyls, Antonius H. M. Akkermans, Tom A. M. Kevenaar, Geert-Jan Schrijnen, A. M. Bazen, and Raymond N. J. Veldhuis. Pratical Biometric Authentication with Template Protection. In *5th International Conference, AVBPA*, Rye Brook, New York, July 2005.

[TG] Pim Tuyls and Jasper Goseling. Capacity and Examples of Template-Protecting Biometric Authentication Systems. In *Biometric Authentication Workshop ECCV2004*.

[ZSBF08] Xuebing Zhou, H. Seibert, C. Busch, and W. Funk. A 3D face recognition algorithm using histogram-based features. In *Eurographics 2008 Workshop on 3D Object Retrieval*, pages 65–71, Crete, Greece, April 2008.