Binary Biometrics: An Analytic Framework to Estimate the Performance Curves under Gaussian Assumption.

E.J.C. Kelkboom, G. Garcia Molina, J. Breebaart, R.N.J. Veldhuis, T.A.M. Kevenaar, and W. Jonker

***Abstract*—In recent years the protection of biometric data has gained increased interest from the scientific community. Methods such as the fuzzy commitment scheme, helper data system, fuzzy extractors, fuzzy vault and cancellable biometrics have been proposed for protecting biometric data. Most of these methods use cryptographic primitives or error-correcting codes (ECC) and use a binary representation of the real-valued biometric data. Hence, the difference between two biometric samples is given by the Hamming distance or bit errors between the binary vectors obtained from the enrollment and verification phases respectively. If the Hamming distance is smaller (larger) than the decision threshold, then the subject is accepted (rejected) as genuine. Because of the use of ECCs, this decision threshold is limited to the maximum error-correcting capacity of the code, consequently limiting the false rejection rate (FRR) and false acceptance rate (FAR) trade-off. A method to improve the FRR consists in using multiple biometric samples in either the enrollment or verification phase. The noise is suppressed, hence reducing the number of bit errors and decreasing the Hamming distance. In practice, the number of samples is empirically chosen without fully considering its fundamental impact. In this work, we present a Gaussian analytical framework for estimating the performance of a binary biometric system given the number of samples being used in the enrollment and the verification phase. The error detection trade-off (DET) curve that combines the false acceptance and false rejection rates is estimated to assess the system performance. The analytic expressions are validated using the FRGC v2 and FVC2000 biometric databases.**

***Index Terms*—Binary biometrics, Binary template matching, Performance estimation, Template protection.**

I. Introduction

WITH the increased popularity of biometrics and its application in society, privacy concerns are being raised by privacy protection watchdogs. This has stimulated research into methods for protecting the biometric data in order to mitigate these privacy concerns. Numerous methods such as the *fuzzy commitment scheme* [1], *helper data system* [2], [3], [4], *fuzzy extractors* [5], [6], *fuzzy vault* [7], [8] and *cancellable biometrics* [9] have been proposed for transforming the biometric data in such a way that the privacy is safeguarded. Several of these privacy or template protection techniques use some cryptographic primitives (e.g. hash functions) or error-correcting codes (ECC). Therefore they use a binary representation of the biometric data, referred to as the *binary vector*. The transition from real-valued to binary representation of the biometric allows the difference between two biometric samples to be quantified by the Hamming distance (HD), i.e. the number of different bits (bit errors) between two binary vectors.

Eventually the biometric system has to verify the claimed identity of a subject. If verified, this identity is considered as genuine. The decision of either rejecting or accepting the subject as genuine depends on whether the Hamming distance is larger than a predetermined decision threshold (T). In template protection systems that use an ECC, T is usually determined by its error-correcting capacity. Hence, the false rejection rate (FRR) depends on the number of genuine matches that produce a Hamming distance larger than the decision threshold.

Attackers may attempt to gain access by impersonating a genuine user. The associated comparisons are referred to as the imposter comparisons and will be accepted if the Hamming distance is smaller or equal to T, thus leading to a false accept. The success rate of impersonation attacks is quantified by the false acceptance rate (FAR).

Therefore, the performance of a biometric system can be expressed by its FAR and FRR, which depends on the genuine (ϕ_{ge}) and imposter (ϕ_{im}) Hamming distance probability mass functions (pmf) and the decision threshold T. A graphical representation is given in Fig. 1.

One of the problems with template protection systems based on ECCs is that the FRR is lower bounded by the error-correcting capacity of the ECC. A large FRR makes the biometric system inconvenient, because many genuine subjects will be wrongly rejected. In some practical cases [2], [3] high FRR values were obtained because it was impossible to further increase the decision boundary, since the used ECC was unable to correct more bits. The method they used to improve the FRR consists in using multiple biometric samples in order to suppress the noise and thus reducing the number of bit errors resulting in a smaller Hamming distance.

The main objective of this study is to analytically estimate, under the Gaussian assumption, the performance of a biometric system based on binary vectors under Hamming distance comparison and considering the use of multiple biometric samples. We present a framework for analytically estimating both the genuine and imposter Hamming distance pmfs from the analytically estimated bit-error probability presented in [10] under the assumption that both the wihin- and between-class of the real-valued features are Gaussian distributed. Firstly, due to the central limit theorem we can assume that the real-valued features will tend to approximate a Gaussian distribution when they result from a linear combinations of many components, e.g. feature extraction techniques based on the principle component analysis (PCA) or linear discriminant

E.J.C. Kelkboom, G. Garcia Molina, J. Breebaart, and W. Jonker are with Philips Research, The Netherlands {Emile.Kelkboom, Gary.Garcia, Jeroen.Breebaart, Willem.Jonker}@philips.com

T.A.M. Kevenaar is with priv-ID, The Netherlands Tom.Kevenaar@priv-id.com

R.N.J. Veldhuis and W. Jonker are with the University of Twente, The Netherlands R.N.J.Veldhuis@utwente.nl and jonker@cs.utwente.nl

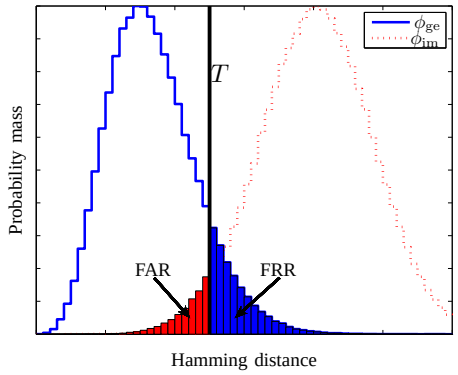


Fig. 1. FRR and FAR from the genuine and imposter Hamming distance pmfs, ϕ_{ge} and ϕ_{im}, respectively.

analysis (LDA). PCA or LDA techniques are often being used to perform dimension reduction in order to prevent overfitting or to simplify the classifier [11], and in the field of template protection PCA is also used to decorrelate the features in order to guarantee uniformly distributed keys extracted from the biometric sample [5]. Secondly, the Gaussian assumption makes it possible to obtain an analytical closed-form expression for the Hamming distance pmf.

This paper is organized as follows. In Section II we present a general description of a biometric system with template protection and model each processing component. We present the Gaussian model assumption describing the probability density function (pdf) of the real-valued biometric features extracted from the biometric sample, the binarization method under consideration, and the interpretation of the template protection block. Then, we present the analytic expression for estimating the genuine and imposter Hamming distance pmfs, and the FRR and FAR curves in Section III. In Section IV we validate these analytic expressions with two different real biometric databases namely, the FRGC v2 3D face images [12] and the FVC2000 fingerprint images [13]. We further extend the framework in Section V and VI in order to relax the assumptions made in Section II. Furthermore, some practical considerations are discussed in Section VII. Section VIII concludes this paper and outlines the future work.

II. Modeling of a Biometric System with Template Protection

A general scheme of a biometric system with template protection based on helper data is shown in Fig. 2. In the enrollment phase a biometric sample, for example a 3D shape image of the face of the subject, is obtained by the acquisition system and presented to the Feature-Extraction module. The biometric sample is preprocessed (enhancement, alignment, etc.) and a real-valued *feature vector* $\mathbf{f}_{\mathrm{R}}^{\mathrm{e}} \in \mathbb{R}^{N_{\mathrm{F}}}$ is extracted, where N_{F} is the number of feature components or dimension of the feature vector. In the Bit-Extraction module, a binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \in \{0,1\}^{N_{\mathrm{B}}}$ is extracted from the real-valued feature vector, where N_{B} is the number of bits and in general does not need to be equal to N_{F}. Quantization schemes range from simple, extracting a single bit out of each feature component [3], [2] to more complex, extracting multiple bits per feature component [14], [15]. Hereafter, the binary vector is protected within the Bit-Protection module. The Bit-Protection module safeguards the privacy of the users of the biometric system by enabling accurate comparisons without the need to store the original biometric data $\mathbf{f}_{\mathrm{R}}^{\mathrm{e}}$ or $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$. We focus on the helper data system that is based on ECCs and cryptographic primitives, for example hash functions. A unique but renewable key is generated for each user and kept secret by using a hash function. Robustness to measurement noise and biometric variability is achieved by effectively using error-correcting codes. The output is a pseudo identity (PI), represented as a binary vector, accompanied by some auxiliary data also known as helper data (AD) [16]. Finally, PI and AD have to be stored for use in the verification phase.

In the verification phase, another live biometric measurement is acquired from which its real-valued feature vector $\mathbf{f}_{\mathrm{R}}^{\mathrm{v}}$ is extracted followed by the quantization process, which produces the binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$. In the Bit-Protection module a candidate pseudo identity PI^* is created using AD and the binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$. There is an exact match between PI and PI^* when the same AD is presented together with a biometric sample with similar characteristics as the one presented in the enrollment phase. In a classical biometric system, the comparator bases its decision on the similarity or distance between the feature vectors $\mathbf{f}_{\mathrm{R}}^{\mathrm{e}}$ and $\mathbf{f}_{\mathrm{R}}^{\mathrm{v}}$. For a binary biometric system, the decision is based on the difference between $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ and $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$, which can be quantified using the Hamming distance. For a template protection system, there is an acceptance only when PI and PI^* are identical.

In summary, the biometric system incorporating template protection can be divided into three blocks; (i) the Acquisition and Feature-Extraction modules where the input is the subject's biometric and the output is a real-valued feature vector $\mathbf{f}_{\mathrm{R}} \in \mathbb{R}^{N_{\mathrm{F}}}$, (ii) the Bit-Extraction module that extracts a binary vector $\mathbf{f}_{\mathrm{B}}$ out of $\mathbf{f}_{\mathrm{R}}$, and (iii) the Bit-Protection and Bit-Matching modules which protects the binary vector and performs the matching and decision making based on PI and PI^*. *To build an analytical framework, we have to model each block.* In this Section we present a simple model for each block. However, the simple model incorporating the Acquisition and Feature-Extraction block is built under strong assumptions and will be relaxed later in the paper.

A. Acquisition and Feature-Extraction Block

The input of the Acquisition and Feature-Extraction block is a captured biometric sample of the subject and the output is a real-valued feature vector $\mathbf{f}_{\mathrm{R}} = [f_{\mathrm{R}}[1], f_{\mathrm{R}}[2], \ldots, f_{\mathrm{R}}[N_{\mathrm{F}}]]'$ of dimension N_{F}, where ' ' is the transpose operator. The feature vector $\mathbf{f}_{\mathrm{R}}$ is likely to be different between two measurements, even if they are acquired immediately after each other. Causes for this difference include sensor noise, environment conditions (e.g. illumination) and biometric variabilities (e.g. pose or expression).

To model these variabilities, we consider Parallel Gaussian Channels (PGC) as portrayed in Fig. 3. We assume

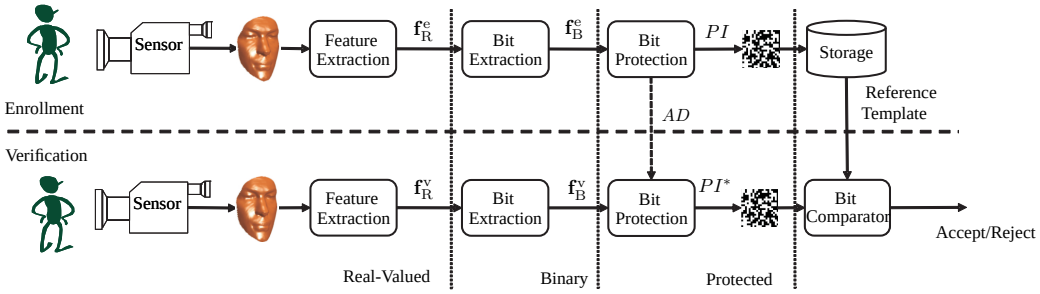


Fig. 2. A general scheme of a biometric system with template protection based on helper data.

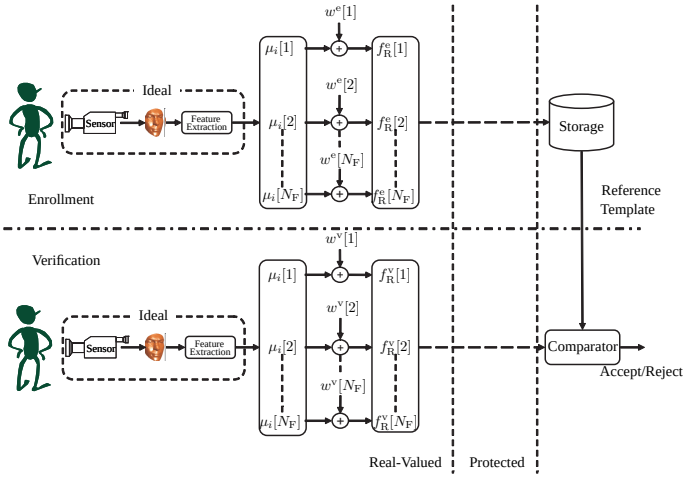


Fig. 3. The Parallel Gaussian Channel for both the enrollment and verification phase.

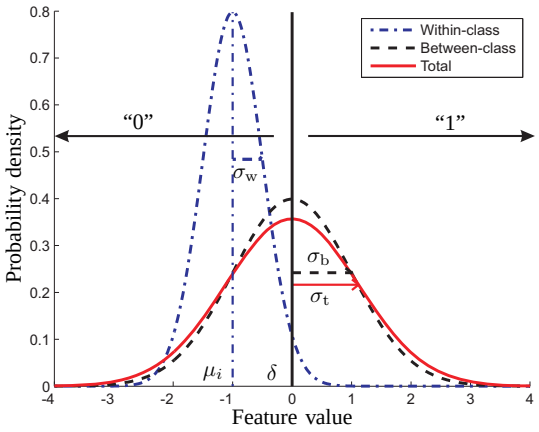


Fig. 4. Modeling of a single feature component of the real-valued biometric.

an ideal Acquisition and Feature-Extraction module which always produces the same feature vector $\boldsymbol{\mu}_i$ for subject i. Such ideal module is thus robust against all aforementioned variabilities. However, the variability of component j is modeled as an additive zero-mean Gaussian noise $w[j]$ with its pdf $p_{\mathrm{w}[j],i} \sim \mathcal{N}(0, \sigma^2_{\mathrm{w},i}[j])$. Adding the noise $w[j]$ with the mean $\mu_i[j]$ results into the noisy feature component $f_{\mathrm{R}}[j]$, in vector notation $\mathbf{f}_{\mathrm{R}} = \boldsymbol{\mu}_i + \mathbf{w}$. The observed variability within one subject is characterized by the variance of the *within-class* pdf and is referred to as within-class variability. We assume that each subject has the same within-class variance, i.e. homogeneous within-class variance $\sigma^2_{\mathrm{w},i}[j] = \sigma^2_{\mathrm{w}}[j], \forall i$. For each component, the within-class variance can be different and we assume the noise to be independent.

On the other hand, each subject should have a unique mean in order to be distinguishable. Across the population we assume $\mu_i[j]$ to be another Gaussian random variable with density $p_{\mathrm{b}[j]} \sim \mathcal{N}(\mu_{\mathrm{b}}[j], \sigma^2_{\mathrm{b}}[j])$. The variability of $\mu_i[j]$ across the population is referred to as the *between-class* variability. Fig. 4 shows an example of the within-class and between-class pdfs for a specific component and a given subject. The *total* pdf describes the observed real-valued feature value $f_{\mathrm{R}}[j]$ across the whole population and is also Gaussian with $p_{\mathrm{t}[j]} \sim \mathcal{N}(\mu_{\mathrm{t}}[j], \sigma^2_{\mathrm{t}}[j])$, where $\mu_{\mathrm{t}}[j] = \mu_{\mathrm{b}}[j]$ and $\sigma^2_{\mathrm{t}}[j] = \sigma^2_{\mathrm{w}}[j] + \sigma^2_{\mathrm{b}}[j]$. For simplicity but without loss of generality we consider $\mu_{\mathrm{t}}[j] = \mu_{\mathrm{b}}[j] = 0$.

As depicted in Fig. 3, in both the enrollment and verification phase the PGC adds random noise $\mathbf{w}^{\mathrm{e}}$ and $\mathbf{w}^{\mathrm{v}}$ with the same probability density to $\boldsymbol{\mu}_i$, resulting in $\mathbf{f}^{\mathrm{e}}_{\mathrm{R}}$ and $\mathbf{f}^{\mathrm{v}}_{\mathrm{R}}$, respectively. Thus $\boldsymbol{\mu}_i$ is sent twice over the same Gaussian channel.

B. Bit-Extraction Block

The function of the Bit-Extraction block is to extract a binary representation from the real-valued representation of the biometric sample. As the bit extraction method, we use the thresholding version used in [2], [3], where a single bit is extracted from each feature component. Hence, the obtained binary vector $\mathbf{f}_{\mathrm{B}} \in \{0,1\}^{N_{\mathrm{F}}}$ has the same dimension as $\mathbf{f}_{\mathrm{R}}$. Furthermore, the binarization threshold for each component $\delta[j]$ is set equal to the mean of the between-class pdf $\mu_{\mathrm{b}}[j]$; if the value of $f_{\mathrm{R}}[j]$ is smaller than $\delta[j]$ then it is set to "0" otherwise it is set to "1", see Fig. 4. More complex binarization schemes could be used [14], [15], but the simple binarization is used more frequently. Therefore, we only focus on the single bit binarization method. Note that the binarization method is similar in both the enrollment and verification phase. In the case where multiple biometric samples are used in either the enrollment (N_{e}) or verification (N_{v}) phase, the average of all the corresponding $\mathbf{f}_{\mathrm{R}}$ is taken prior to the binarization process.

C. Bit-Protection and Bit-Comparator Block

Many bit protection or template protection schemes are based on the capability of generating a robust binary vector

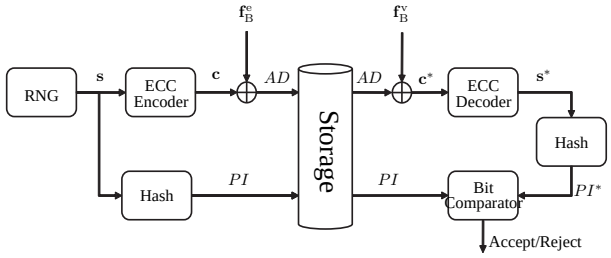


Fig. 5. Fuzzy commitment scheme.

or key out of different biometric measurements of the same subject. However, the binary input vector $\mathbf{f}_{\mathrm{B}}$ itself cannot be used as the key because it is most likely not exactly the same in both the enrollment 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 $d_H(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}})$. Therefore, error-correcting codes are used to deal with these bit errors. A possible way of integrating an ECC is shown in Fig. 5, which is also known as the fuzzy commitment scheme [1].

In the enrollment phase, a binary secret or message vector $\mathbf{s}$ is randomly generated by the *Random-Number-Generator* (*RNG*) module. The security level of the system is higher at larger secret lengths. A codeword $\mathbf{c}$ of an error-correcting code is obtained by encoding $\mathbf{s}$ in the *ECC-Encoder* module. The codeword is XOR-ed with $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ in order to obtain the auxiliary data AD. Furthermore, the hash of $\mathbf{s}$ is taken in order to obtain the pseudo identity PI. For the sake of coherence we use the terminology proposed in [17], [16].

In the verification phase, the possibly corrupted codeword $\mathbf{c}^*$ is created by XOR-ing $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ with AD. The candidate secret $\mathbf{s}^*$ is obtained by decoding $\mathbf{c}^*$ in the *ECC-Decoder* module. We compute the candidate pseudo identity PI^* by hashing $\mathbf{s}^*$. The decision in the Bit-Comparator block is based on whether PI and PI^* are bitwise identical.

We focus on the linear block type ECC "Bose, Ray-Chaudhuri, Hocquenghem" (BCH), which is specified by the codeword length (n_{c}), message length (k_{c}), and the corresponding number of bits that can be corrected (t_{c}), in short $[n_{\mathrm{c}}, k_{\mathrm{c}}, t_{\mathrm{c}}]$. Because the BCH ECC can correct random bit errors, the Bit-Protection module yields equivalent PI and PI^* when the number of bit errors between the binary vectors $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ and $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ is smaller or equal to the error-correcting capability t_{c}. Thus, there is a match when the Hamming distance is smaller than t_{c}, $d_H(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) = ||\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \oplus \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}||_1 \leq t_{\mathrm{c}}$, and the Bit-Protection module can be modeled as a Hamming distance classifier with threshold t_{c}. The basic BCH error correction code is used here. While more sophisticated ECCs can be used, the BCH accommodates our framework due to its Hamming distance classifier property. For example if we consider the binary symbol based Reed-Solomon code, the number of bits it can correct depends on the error pattern. Hence, their probabilistic decoding behavior also needs to be modelled which is out of the scope of this work. Some $[n_{\mathrm{c}}, k_{\mathrm{c}}, t_{\mathrm{c}}]$ settings of the BCH code are given in Table I. Note, that the maximum number of bits that can be corrected lies between 20-25% of the binary vector.

D. Modeling Summary

Here follows a summary of the modeling choices and assumptions that we have made:

- **Acquisition and Feature-Extraction Block $\mathbf{f}_{\mathrm{R}}$**
  - Modeled as a Parallel Gaussian Channel, where each feature component is defined by:
    * Within-class pdf $\sim \mathcal{N}(0, \sigma_{\mathrm{w}}^2[j])$
      - Describes the genuine biometric variability and measurement noise
      - Homogeneous variance across subjects $\sigma_{\mathrm{w},i}^2[j] = \sigma_{\mathrm{w}}^2[j], \forall i$
      - Noise is independent across channels, measurements, and subjects
    * Between-class pdf $\sim \mathcal{N}(0, \sigma_{\mathrm{b}}^2[j])$
      - Characterizes the $\mu_i[j]$ variability across the population
      - Feature components are independent
    * Total pdf $\sim \mathcal{N}(0, \sigma_{\mathrm{t}}^2[j])$
      - Defines $f_{\mathrm{R}}[j]$ across the population
- **Bit-Extraction Block $\mathbf{f}_{\mathrm{B}}$**
  - Single bit extraction method, with binarization threshold $\delta[j] = \mu_{\mathrm{b}}[j]$
- **Bit-Protection and Bit-Comparator Block**
  - Hamming distance classifier with the ECC settings defining its decision boundary.

III. Analytical Estimation of Bit-Error Probabilities, FRR and FAR.

The goal of this study is to analytically estimate the performance of the presented general template protection system. In Section II, we have presented a comprehensive description of such a system including the modeling approach or properties of each block that forms the basis of our analytic framework. In case of a Hamming distance classifier, the goal is to analytically estimate the expected genuine and imposter Hamming distance pmfs ϕ_{ge} and ϕ_{im}, respectively (see Fig. 1). With these pmfs we can compute the false rejection rate β (FRR) and the false acceptance rate α (FAR), where β is the probability that a genuine subject is incorrectly rejected and α

TABLE I
Some examples of the BCH code given by the codeword (n_{c} and message (k_{c}) length, the corresponding number of correctable bits (t_{c}), and the bit error rate $t_{\mathrm{c}}/n_{\mathrm{c}}$.

n_{c}	k_{c}	t_{c}	BER $= t_{\mathrm{c}}/n_{\mathrm{c}}$
15	5	3	20.0%
	11	1	6.7%
31	6	7	22.6%
	16	3	9.7%
63	7	15	23.8%
	16	11	17.5%

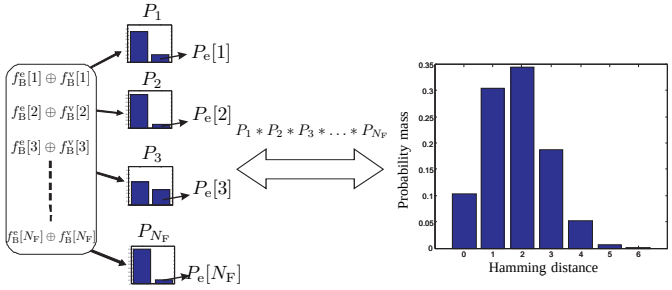


Fig. 6. A toy example of the convolution method given by (2).

is the probability that an imposter is incorrectly accepted by the biometric system.

The Hamming distance between two binary vectors is the number of bit errors between them. Knowing the bit-error probability for each bit $P_e[j]$, the expected Hamming distance $\bar{d}_H$ between $\mathbf{f}_B^e$ and $\mathbf{f}_B^v$ is

$$\bar{d}_H(\mathbf{f}_B^e, \mathbf{f}_B^v) = \sum_{j=1}^{N_F} P_e[j]. \tag{1}$$

Further, we define the pmf of the number of bit errors of component j as $P_j = [1 - P_e[j], P_e[j]]$, where $P_j(0)$ is the probability of no bit error ($d_H = 0$) and $P_j(1)$ is the probability of a single bit error ($d_H = 1$). Under the assumption that the bit-error probabilities are independent, the pmf of $d_H(\mathbf{f}_B^e, \mathbf{f}_B^v)$ is defined as

$$\begin{aligned} \phi(k) &\stackrel{\text{def}}{=} \mathcal{P}\{d_H(\mathbf{f}_B^e, \mathbf{f}_B^v) = k\} \\ &= (P_1 * P_2 * \ldots * P_{N_F})(k), \end{aligned} \tag{2}$$

where the convolution is taken of the pmf of the number of bit errors per component. A toy example is shown in Fig. 6. For the two extreme cases of (2) we have

$$\phi(0) = \prod_{j=1}^{N_F} P_j(0) = \prod_{j=1}^{N_F} (1 - P_e[j]), \tag{3}$$

$$\phi(N_F) = \prod_{j=1}^{N_F} P_j(1) = \prod_{j=1}^{N_F} P_e[j], \tag{4}$$

which are the probabilities of having zero or N_F errors, respectively. The FRR corresponding to a Hamming distance threshold T, $\beta(T)$, is the probability that the Hamming distance for a genuine comparison is greater than T, therefore

$$\begin{aligned} \beta(T) &= \mathcal{P}\{d_H(\mathbf{f}_{B,i}^e, \mathbf{f}_{B,i}^v) > T\} \\ &= \sum_{k=T+1}^{N_F} \phi_{ge}(k). \end{aligned} \tag{5}$$

Furthermore, $\alpha(T)$ is the probability that the Hamming distance for an imposter comparison is smaller or equal to the threshold T, hence we have

$$\begin{aligned} \alpha(T) &= \mathcal{P}\{d_H(\mathbf{f}_{B,i}^e, \mathbf{f}_{B,j}^v) \leq T, \forall i \neq j\} \\ &= \sum_{k=0}^{T} \phi_{im}(k). \end{aligned} \tag{6}$$

In other words, if we want to estimate $\beta(T)$ and $\alpha(T)$ analytically we have to obtain an analytic closed-form expression of the average bit-error probability $P_e[j]$ across the population for both the genuine and imposter case, $P_e^{ge}[j]$ and $P_e^{im}[j]$ respectively. Because of the PGC modeling approach, $P_e^{ge}[j]$ will depend on the within-class and between-class variances $\sigma_w^2[j]$ and $\sigma_b^2[j]$, respectively. Furthermore, we also want to find the relationship between $P_e^{ge}[j]$ and the number of enrollment N_e and verification N_v samples. As mentioned in Section II-B, in case of multiple samples the average of the extracted $\mathbf{f}_R$ of each samples is taken prior to the binarization process.

A. P_e Estimation for the Imposter Case: P_e^{im}

For the imposter case, we are considering the comparison between binary vectors of two different subjects, $d_H(\mathbf{f}_{B,i}^e, \mathbf{f}_{B,j}^v), \forall i \neq j$. As mentioned in Section II-B, we focus on the binarization method based on thresholding with $\delta = \mu_b = \mu_t$ (see Fig. 4). Because the total pdf is assumed to be Gaussian with mean μ_t, we have equiprobable bit values. This implies that the bit-error probability of randomly guessing a bit is 1/2, $P_e^{im}[j] = 1/2, \forall j$. Thus, under the assumption that the feature components are independent, imposter comparisons are similar to matching $\mathbf{f}_B^e$ with a random binary vector.

Since $P_e^{im}[j] = 1/2, \forall j$, we can simplify $\phi_{im}(k)$ as the binomial pmf

$$\phi_{im}(k) = (P_1 * P_2 * \ldots * P_{N_F})(k) \tag{7}$$

$$= \binom{N_F}{k} (P_e^{im}[j])^k (1 - P_e^{im}[j])^{N_F - k} \tag{8}$$

$$= \binom{N_F}{k} 2^{-N_F}, \tag{9}$$

where the simplification step from (7) to (8) holds because of $P_e^{im}[i] = P_e^{im}[j],\ \forall\ i \neq j$. Furthermore, $\alpha(T)$ turns into

$$\alpha(T) = \sum_{k=0}^{T} \phi_{im}(k) = 2^{-N_F} \sum_{k=0}^{T} \binom{N_F}{k}, \tag{10}$$

which corresponds to what is used in [18].

B. P_e Estimation for the Genuine Case: P_e^{ge}

We focus on estimating the bit-error probability for each component $P_e^{ge}[j]$, and for convenience purposes we omit the component index j. Using the Gaussian model approach as defined in Section II and depicted in Fig. 7, the expected bit-error probability P_e^{ge} over the whole population is defined by

$$\begin{aligned} P_e^{ge} &= E\left[P_e^{ge}(\mu)\right] \\ &= \int_{-\infty}^{\infty} p_b(\mu) P_e^{ge}(\mu)\, d\mu, \end{aligned} \tag{11}$$

where $P_e^{ge}(\mu)$ is the bit-error probability given μ and p_b is the between-class pdf. With the binarization threshold $\delta = \mu_b = 0$, this problem becomes symmetric with respect to δ.

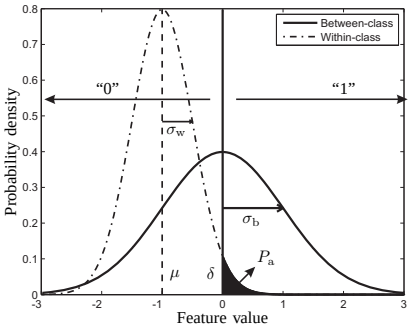


Fig. 7. Measurement error $P_{\rm a}$.

Consequently, (11) becomes

$$\begin{aligned} P_{\rm e}^{\rm ge} &= 2\int_{-\infty}^{0} p_{\rm b}(\mu)P_{\rm e}^{\rm ge}(\mu)\,{\rm d}\mu \\ &= 2\int_{-\infty}^{0} \tfrac{1}{\sqrt{2\pi}\sigma_{\rm b}}{\rm e}^{-\left(\frac{\mu}{\sqrt{2}\sigma_{\rm b}}\right)^2} P_{\rm e}^{\rm ge}(\mu)\,{\rm d}\mu \\ &= \tfrac{2\lambda}{\sqrt{\pi}}\int_{-\infty}^{0} {\rm e}^{-(\lambda\mu)^2} P_{\rm e}^{\rm ge}(\mu)\,{\rm d}\mu, \end{aligned} \tag{12}$$

where $\lambda = \frac{1}{\sqrt{2}\sigma_{\rm b}}$.

We define the measurement or acquisition error probability $P_{\rm a}$, depicted by the shaded area in Fig. 7, as the probability that the measured bit is different than the bit defined by the mean μ of the feature value. $P_{\rm a}$ becomes smaller at either a larger distance between μ and the binarization threshold δ or a smaller within-class variance. Since multiple enrollment ($N_{\rm e}$) and verification ($N_{\rm v}$) samples are considered, $P_{\rm a}$ also depends on the number of samples N, given as

$$P_{\rm a}(\mu;N) = \int_{0}^{\infty} \tfrac{\sqrt{N}}{\sqrt{2\pi}\sigma_{\rm w}} e^{-\left(\frac{\sqrt{N}(x-\mu)}{\sqrt{2}\sigma_{\rm w}}\right)^2}\,{\rm d}x, \tag{13}$$

where we used the fact that when averaging N samples the within-class variance decreases as

$$\sigma_{{\rm w},N}^2 = \frac{\sigma_{\rm w}^2}{N} \Rightarrow \sigma_{{\rm w},N} = \frac{\sigma_{\rm w}}{\sqrt{N}}. \tag{14}$$

With use of the error function

$$\texttt{erf}(z) \;\; = \tfrac{2}{\sqrt{\pi}}\int_{0}^{z} {\rm e}^{-t^2}\,{\rm d}t. \tag{15}$$

and by defining $\eta = \frac{\sqrt{N}}{\sqrt{2}\sigma_{\rm w}}$, $P_{\rm a}(\mu;N)$ can be rewritten as

$$\begin{aligned} P_{\rm a}(\mu;N) &= \tfrac{\eta}{\sqrt{\pi}}\int_{0}^{\infty} {\rm e}^{-(\eta(x-\mu))^2}{\rm d}x \\ &= \tfrac{1}{\sqrt{\pi}}\int_{-\eta\mu}^{\infty} {\rm e}^{-z^2}{\rm d}z, \text{ with } z=\eta(x-\mu) \\ &= \tfrac{1}{\sqrt{\pi}}\left[\int_{0}^{\infty} {\rm e}^{-z^2}{\rm d}z - \int_{0}^{-\eta\mu} {\rm e}^{-z^2}{\rm d}z\right], \text{for } \mu\le 0 \\ &= \tfrac{1}{\sqrt{\pi}}\left[\tfrac{\sqrt{\pi}}{2} - \tfrac{\sqrt{\pi}}{2}\texttt{erf}(-\eta\mu)\right] \\ &= \tfrac{1}{2}\left[1-\texttt{erf}(-\eta\mu)\right], \end{aligned} \tag{16}$$

where we used the well known result $\int_0^\infty \lambda{\rm e}^{-(\lambda\mu)^2}\,{\rm d}\mu = \frac{\sqrt{\pi}}{2}$.

There is a bit-error probability only when there is a measurement error at either the enrollment or the verification phase. If there is a measurement error in both phases then the measured bits still have the same bit value, thus no bit error. Hence, $P_{\rm e}(\mu)$ of (12) becomes

$$\begin{aligned} P_{\rm e}^{\rm ge}(\mu;N_{\rm e},N_{\rm v}) &= (1-P_{\rm a}(\mu;N_{\rm e}))P_{\rm a}(\mu;N_{\rm v}) \\ &\quad + P_{\rm a}(\mu;N_{\rm e})(1-P_{\rm a}(\mu;N_{\rm v})) \\ &= \tfrac{1}{4}\left[(1+\texttt{erf}(-\eta_{\rm e}\mu))(1-\texttt{erf}(-\eta_{\rm v}\mu))\right. \\ &\quad \left. + (1-\texttt{erf}(-\eta_{\rm e}\mu))(1+\texttt{erf}(-\eta_{\rm v}\mu))\right] \\ &= \tfrac{1}{2}\left[1-\texttt{erf}(-\eta_{\rm e}\mu)\texttt{erf}(-\eta_{\rm v}\mu)\right], \end{aligned} \tag{17}$$

where $\eta_{\rm e} = \frac{\sqrt{N_{\rm e}}}{\sqrt{2}\sigma_{\rm w}}$ and $\eta_{\rm v} = \frac{\sqrt{N_{\rm v}}}{\sqrt{2}\sigma_{\rm w}}$. By substituting (17) into (12) we obtain

$$\begin{aligned} P_{\rm e}^{\rm ge}(N_{\rm e},N_{\rm v}) &= \frac{\lambda}{\sqrt{\pi}}\int_{-\infty}^{0} {\rm e}^{-(\lambda\mu)^2}\left[1-\texttt{erf}(-\eta_{\rm e}\mu)\texttt{erf}(-\eta_{\rm v}\mu)\right]{\rm d}\mu \\ &= \frac{\lambda}{\sqrt{\pi}}\int_{0}^{\infty} {\rm e}^{-(\lambda\mu)^2}\left[1-\texttt{erf}(\eta_{\rm e}\mu)\texttt{erf}(\eta_{\rm v}\mu)\right]{\rm d}\mu \\ &= \frac{1}{2} - \frac{\lambda}{\sqrt{\pi}}\int_{0}^{\infty} {\rm e}^{-\lambda^2\mu^2}\texttt{erf}(\eta_{\rm e}\mu)\texttt{erf}(\eta_{\rm v}\mu)\,{\rm d}\mu. \end{aligned} \tag{18}$$

The integral of the erf function can be solved using the general solution of erf integrals [19] given as

$$\int_{0}^{\infty} {\rm e}^{-\gamma x^2}\texttt{erf}(ax)\texttt{erf}(bx)\,{\rm d}x = \frac{\arctan\left(\frac{ab}{\sqrt{\gamma(a^2+b^2+\gamma)}}\right)}{\sqrt{\gamma\pi}}. \tag{19}$$

Thus, (18) can be solved by using (19) with $\gamma=\lambda^2$, $a=\eta_{\rm e}$, and $b=\eta_{\rm v}$ as

$$\begin{aligned} P_{\rm e}^{\rm ge}(N_{\rm e},N_{\rm v},\sigma_{\rm w},\sigma_{\rm b}) &= \tfrac{1}{2} - \tfrac{\lambda}{\sqrt{\pi}}\frac{\arctan\left(\frac{\eta_{\rm e}\eta_{\rm v}}{\sqrt{\lambda^2\left(\eta_{\rm e}^2+\eta_{\rm v}^2+\lambda^2\right)}}\right)}{\lambda\sqrt{\pi}} \\ &= \tfrac{1}{2} - \tfrac{1}{\pi}\arctan\left(\frac{\eta\sqrt{N_{\rm e}N_{\rm v}}}{\lambda\sqrt{N_{\rm e}+N_{\rm v}+\left(\frac{\lambda}{\eta}\right)^2}}\right) \\ &= \tfrac{1}{2} - \tfrac{1}{\pi}\arctan\left(\frac{\sigma_{\rm b}\sqrt{N_{\rm e}N_{\rm v}}}{\sigma_{\rm w}\sqrt{N_{\rm e}+N_{\rm v}+\left(\frac{\sigma_{\rm b}}{\sigma_{\rm w}}\right)^{-2}}}\right), \end{aligned} \tag{20}$$

where we also included $\sigma_{\rm w}$ and $\sigma_{\rm b}$ as an argument of the estimation function. As can be observed, $P_{\rm e}^{\rm ge}$ is dependent on the $\sigma_{\rm b}/\sigma_{\rm w}$ ratio, $N_{\rm e}$, and $N_{\rm v}$.

C. Summary

We have presented the analytic expressions of the genuine ($\phi_{\rm ge}$) and imposter ($\phi_{\rm im}$) Hamming distance pmfs and the corresponding FRR ($\beta(T)$) and FAR ($\alpha(T)$) curves. Because of the choice of the binarization scheme the imposter bit-error probability $P_{\rm e}^{\rm im}[j]$ does not need to be estimated and can be assumed to be equal to 1/2 for each feature component. However, the genuine bit-error probability $P_{\rm e}^{\rm ge}[j]$ has to be estimated using the analytic expression in (20). Therefore, in the remainder of this study we only need to estimate

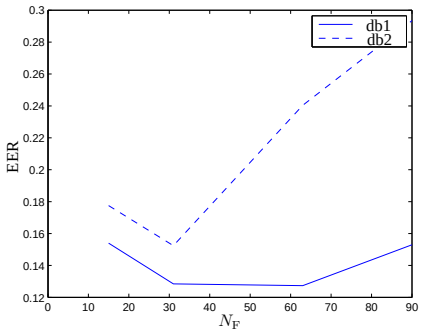


Fig. 8. EER of the training set after applying PCA for different reduced number of features N_F.

$P_e^{ge}[j]$ and for convenience reason we frequently omit the ge superscript.

IV. Experimental Evaluation with Biometric Databases

In this section, the analytic expressions and the effect of the Gaussian assumption are validated using two real biometric databases, which are discussed in Section IV-A. To estimate $P_e[j]$ using (20), we need to estimate the within- and between-class variances $\sigma_w^2[j]$ and $\sigma_b^2[j]$, respectively. In Section IV-B we show that the within-class variance influences the between-class variance estimation and we present a corrected estimator. Due to the limited size of the databases, estimation errors do occur when estimating $P_e[j]$ even in the case when the underlying model is correct. We account for these errors by estimating the 95 percentile boundaries in Section IV-C. We then present the results of estimating $P_e[j]$ in Section IV-D, and the effect of using PCA as a mean to generate uncorrelated features in Section IV-E. We conclude by portraying the experimental $\phi_{ge}(k)$, $\phi_{im}(k)$, $\beta(T)$, $\alpha(T)$, and DET curves in Section IV-F.

A. Biometric Databases and Feature Extraction

The first database (db1) consists of 3D face images from the FRGC v2 dataset [12], where we used the shape-based 3D face recognizer of [20] to extract feature vectors of dimension $N_{orig} = 696$. Subjects with at least 8 samples were selected resulting in $N_s = 230$ subjects with a total of $N_t = 3147$ samples. The number of samples per subject varies between 8 and 22 with an approximate average of $\bar{N}_i = 14$ samples per subject. The second database (db2) consists of fingerprint images from Database 2 of FVC2000 [13], and uses a feature extraction algorithm based on Gabor filters and directional fields [21] resulting in 1536 features ($N_{orig} = 1536$). There are $N_s = 110$ subjects with $N_i = 8$ samples each. An overview is given in Table II.

The components of the original feature vectors are dependent. Therefore, we applied the principle component analysis (PCA) technique to decorrelate the features and reduce the dimension of the feature space if necessary. Furthermore, we partitioned both databases into a training and testing set containing 25% and 75% of the number of subjects, respectively.

TABLE II
Overview of the biometric databases

Database	N_{orig}	N_s	N_t	$\bar{N}_i = N_t/N_s$
FRGC v2 (db1)	696	230	3147	≈ 14
FVC2000 (db2)	1536	110	880	8

TABLE III
Variance estimation table as defined in [23].

Source of variation	*Sum of squares*	*d.f.*	Auxiliary
Within	$\sum_{i=1}^{N_s}\sum_{j=1}^{N_i}(f_{i,j}-\hat{\mu}_i)^2$	$N_t - N_s$	$\hat{\mu}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} f_{i,j}$
Between	$\sum_{i=1}^{N_s} N_i(\hat{\mu}_i-\hat{\mu})^2$	$N_s - 1$	$\hat{\mu} = \frac{1}{N_t}\sum_{i=1}^{N_s}\sum_{j=1}^{N_i} f_{i,j}$
Total	$\sum_{i=1}^{N_s}\sum_{j=1}^{N_i}(f_{i,j}-\hat{\mu})^2$	$N_t - 1$	

The size of the test set is a very important factor in this analytic framework, thus we traded off the size of the training set and limited it to 25 % of the number of subjects. We applied PCA on the training set and reduced the dimensionality (N_F) of the feature vectors to the codeword lengths presented in Table I and computed the equal error rate (EER) (see Fig. 8), which is defined as the point where FAR equals FRR. The optimal performance is computed using the bit-extraction method in Section II-B and a Hamming distance classifier. The optimal number of features for both db1 and db2 are in the range of 15, 31, and 63. Note that the best EER of 12.7% for db1 and 15.2% for db2 is higher than the reported performance of template protection systems based on these databases in the literature ($\approx 8\%$ for db1 in [2] and $\approx 5\%$ for db2 in [22])[1]. However, our proposed analytic framework is not focused on optimizing the performance but on analytically estimating the performance. The effect of the PCA transformation on the feature value distribution and the error probability estimation is discussed in Section IV-E. Unless stated otherwise, the remainder of this analysis is based on the PCA transformed test set using the PCA matrix obtained from the training set. For convenience, the remainder of this work is mainly focussed on the optimal setting of $N_F = 31$.

B. Variance Estimation of σ_w^2 and σ_b^2

The analytic expression $P_e^{ge}(N_e, N_v, \sigma_w, \sigma_b)$ in (20) requires the standard deviations σ_w and σ_b. The estimated values $\hat{\sigma}_w$ and $\hat{\sigma}_b$ are obtained from the test set of the database under consideration. The variances $\hat{\sigma}_w^2$ and $\hat{\sigma}_b^2$ are estimated according to the *variance estimation* table given in Table III from [23], where $f_{i,j}$ is the jth real-valued feature vector of subject i, N_s is the number of subjects, N_i is the number of samples or feature vectors of subject i and N_t is the total number of samples $N_t = \sum_{i=1}^{N_s} N_i$. This table is also used in ANOVA (analysis of variance) models and describes the

[1]In [2] the most reliable feature components were selected and in [22] six enrollment samples were used.

method for computing the *sum of squares* of the source of the within-class (SSW), between-class (SSB), and the total (SST) variation. Two important facts deriving from this table are that (i) the total sum of squares is equal to sum of the within-class and between-class sum of squares SST $=$ SSW $+$ SSB, and (ii) the total number of *degrees of freedom* (d.f.) is equal to the sum of the between-class and the within-class degrees of freedom. The details are in [23]. With the use of the table, the variance estimation is given as the sum of squares divided by the d.f., thus

$$\begin{aligned}\hat{\sigma}^2_{\mathrm{w}} &= \tfrac{1}{N_{\mathrm{t}}-N_{\mathrm{s}}}\sum_{i=1}^{N_{\mathrm{s}}}\sum_{j=1}^{N_i}\left(f_{i,j}-\hat{\mu}_i\right)^2,\\ \hat{\sigma}^2_{\mathrm{b}} &= \tfrac{1}{\bar{N}_{\mathrm{i}}(N_{\mathrm{s}}-1)}\sum_{i=1}^{N_{\mathrm{s}}}N_i\left(\hat{\mu}_i-\hat{\mu}\right)^2 \text{ with } \bar{N}_{\mathrm{i}}=\tfrac{N_{\mathrm{t}}}{N_{\mathrm{s}}}\\ \hat{\sigma}^2_{\mathrm{t}} &= \tfrac{1}{N_{\mathrm{t}}-1}\sum_{i=1}^{N_{\mathrm{s}}}\sum_{j=1}^{N_i}\left(f_{i,j}-\hat{\mu}\right)^2,\end{aligned} \tag{21}$$

with the exception of $\hat{\sigma}^2_{\mathrm{b}}$, which is also divided by the average number of samples per subject $\bar{N}_{\mathrm{i}}$. Notice that $\hat{\sigma}^2_{\mathrm{w}}$ is calculated as the variance of the aggregated zero-mean samples of subjects, while taking into account that N_{s} degrees of freedom are lost because of the need to estimate the mean of each subject $\hat{\mu}_i$. Furthermore, $\hat{\sigma}^2_{\mathrm{w}}$ is also equal to the weighted average of the variance of each subject, because (21) can also be written as

$$\begin{aligned}\hat{\sigma}^2_{\mathrm{w}} &= \tfrac{1}{N_{\mathrm{t}}-N_{\mathrm{s}}}\sum_{i=1}^{N_{\mathrm{s}}}(N_i-1)\hat{\sigma}^2_{\mathrm{w},i}\\ &= \tfrac{1}{N_{\mathrm{s}}\left(\bar{N}_{\mathrm{i}}-1\right)}\sum_{i=1}^{N_{\mathrm{s}}}(N_i-1)\hat{\sigma}^2_{\mathrm{w},i}\\ &= \tfrac{1}{\frac{1}{N_{\mathrm{s}}}\sum_{i=1}^{N_{\mathrm{s}}}(N_i-1)}\tfrac{1}{N_{\mathrm{s}}}\sum_{i=1}^{N_{\mathrm{s}}}(N_i-1)\hat{\sigma}^2_{\mathrm{w},i}, \text{ with}\\ \hat{\sigma}^2_{\mathrm{w},i} &= \tfrac{1}{N_i-1}\sum_{j=1}^{N_i}\left(f_{i,j}-\hat{\mu}_i\right)^2,\end{aligned} \tag{22}$$

which turns into $\hat{\sigma}^2_{\mathrm{w}} = \frac{1}{N_{\mathrm{s}}}\sum_{i=1}^{N_{\mathrm{s}}}\hat{\sigma}^2_{\mathrm{w},i}$ when N_i is equal for each subject.

The variance estimators are validated using a synthetically generated database of $N_{\mathrm{s}} = 1000$ subjects with $N_i = 4$ samples each. The parameters $\{\sigma^2_{\mathrm{w}}, \sigma^2_{\mathrm{b}}\}$ are used during the synthesis and we estimated $\{\hat{\sigma}^2_{\mathrm{w}}, \hat{\sigma}^2_{\mathrm{b}}, \hat{\sigma}^2_{\mathrm{t}}\}$ using Eqs. 21, 21, and 21, respectively. The synthesis and estimation processes are performed ten times (10-fold) and the average of the result is taken. Fig. 9 shows the estimation results of $\hat{\sigma}^2_{\mathrm{w}}$ for different values of σ^2_{w} with $\sigma^2_{\mathrm{b}} = 2$, and both $\hat{\sigma}^2_{\mathrm{b}}$ and $\hat{\sigma}^2_{\mathrm{t}}$ for different values of σ^2_{b} with $\sigma^2_{\mathrm{w}} = 2$. We can conclude that the $\hat{\sigma}^2_{\mathrm{w}}$ and $\hat{\sigma}^2_{\mathrm{t}}$ estimators give values that closely resemble the underlying model parameters σ^2_{w} and σ^2_{t}, but we observe a constant estimation error for the $\hat{\sigma}^2_{\mathrm{b}}$ estimator . This estimation error is examined for different values of σ^2_{w} and N_i, as shown in Fig. 10(a) and (b), respectively. The figures show that the estimation error increases when σ_{w} increases or when N_i decreases.

The constant estimation error of $\hat{\sigma}^2_{\mathrm{b}}$ is caused by the estimation error of the sample mean of each subject $\hat{\mu}_i$. From [23], we know that the variance of the sampling distribution

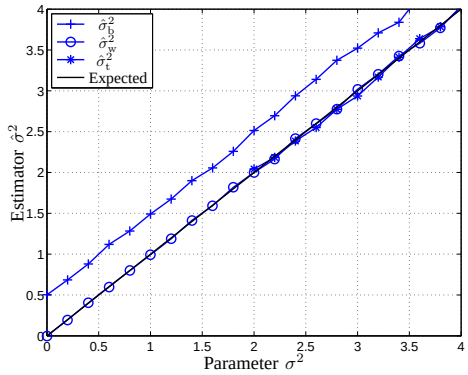


Fig. 9. The within-class, between-class, and total variance estimation for different settings of $\{\sigma^2_{\mathrm{w}}, \sigma^2_{\mathrm{b}}\}$.

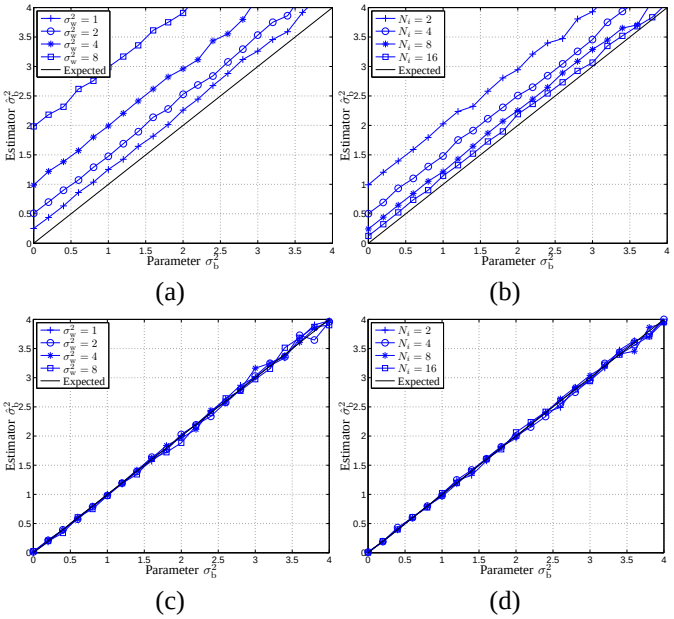


Fig. 10. The between-class estimation of (21) at (a) different values of σ^2_{w} with $N_i = 2$ and (b) different values of N_i with $\sigma^2_{\mathrm{w}} = 2$, with its corrected version (25) in (c) and (d), respectively.

of the sample mean $\hat{\mu}_i$ is given by

$$\sigma^2_{\hat{\mu}_i} = \tfrac{\sigma^2_{\mathrm{w},i}}{N_i}. \tag{23}$$

If more samples are taken to estimate the sample mean, the estimation variance decreases. This implies that the estimation $\hat{\sigma}^2_{\mathrm{b}}$ of (21) is in fact

$$\hat{\sigma}^2_{\mathrm{b}} = EST(\sigma^2_{\mathrm{b}} + \sigma^2_{\hat{\mu}}) = EST(\sigma^2_{\mathrm{b}} + \tfrac{\sigma^2_{\mathrm{w}}}{\bar{N}_{\mathrm{i}}}), \tag{24}$$

where $EST(\tau) \triangleq \hat{\tau}$ is the estimation of parameter τ. The corrected version of the between-class estimation $\check{\sigma}^2_{\mathrm{b}}$ thus becomes

$$\check{\sigma}^2_{\mathrm{b}} = \hat{\sigma}^2_{\mathrm{b}} - \tfrac{\hat{\sigma}^2_{\mathrm{w}}}{\bar{N}_{\mathrm{i}}}. \tag{25}$$

Fig. 10(c)(d) shows the results of applying this correction on the results of Fig. 10(a)(b) and the estimation has clearly improved.

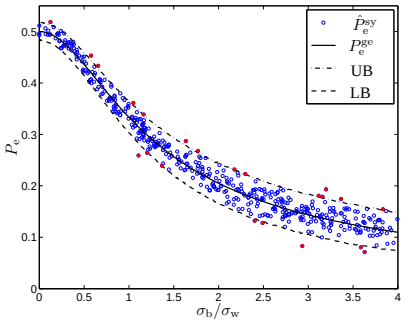


Fig. 11. Random estimation errors due to the random nature and the upper (UB) and lower (LB) boundaries.

C. Boundaries of Tolerated Estimation Errors

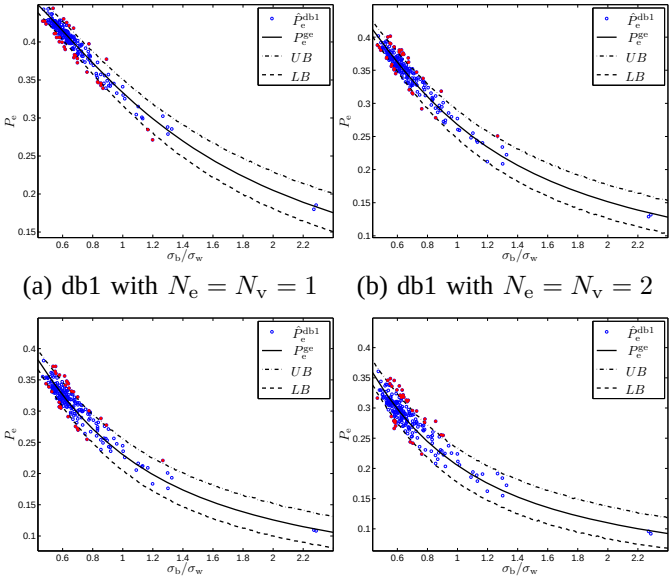
When estimating $P_{\rm e}[j]$ of a given biometric database, there are always estimation errors because of its random nature. Even if we randomly generate a synthetic database that fully complies with the Gaussian modeling assumption, there are still estimation errors. These estimation errors are caused by the random nature of the problem and should be tolerated. Hence, we compute the upper (UB) and lower (LB) tolerance bounds for the estimation errors. Such an example is depicted in Fig. 11 for a synthetic dataset of similar size as db2 ($N_{\rm s} = 110$ and $N_i = 8$) but with $N_{\rm F} = 500$ and $\sigma_{\rm w}^2[j] = 1$ with $\sigma_{\rm b}^2[j]$ randomly drawn from the uniform distribution $U(0, 16)$ with minimum and maximum value of 0 and 16, respectively. Fig. 11 compares the estimated bit-error probability of the synthetic dataset $\hat{P}_{\rm e}^{\rm sy}[j]$ with the corresponding analytically obtained $P_{\rm e}^{\rm ge}[j]$, which stands for $P_{\rm e}^{\rm ge}(N_{\rm e}, N_{\rm v}, \hat{\sigma}_{\rm w}[j], \check{\sigma}_{\rm b}[j])$ of (20), where $\hat{\sigma}_{\rm w}[j]$ and $\check{\sigma}_{\rm b}[j]$ are estimated using (21) and (25), respectively. $\hat{P}_{\rm e}^{\rm sy}[j]$ is reported by a circle ('o') at its estimated $\hat{\sigma}_{\rm b}[j]/\hat{\sigma}_{\rm w}[j]$ ratio and its analytic estimation is the value of the solid line at the same $\hat{\sigma}_{\rm b}[j]/\hat{\sigma}_{\rm w}[j]$ ratio. A greater vertical distance implies a greater analytical estimation error.

The test protocol for calculating $\hat{P}_{\rm e}^{\rm sy}[j]$ is as follows: for each feature component, $\hat{P}_{\rm e}^{\rm sy}[j]$ is calculated as the average across the bit-error probability of each subject $\hat{P}_{{\rm e},i}^{\rm sy}[j]$. The subject bit-error probability $\hat{P}_{{\rm e},i}^{\rm sy}[j]$ results from performing 200 matches and determining the relative number of errors. For each match, $N_{\rm e}$ distinct feature vectors are randomly selected, averaged and binarized (enrollment phase). The obtained bit is compared to the bit obtained from averaging and binarizing $N_{\rm v}$ different randomly selected feature vectors of the same subject (verification phase).

We empirically estimate the upper (UB) and lower (LB) boundaries by clustering the points into equidistant intervals on the x-axis and compute the 95 percentile range of the $\hat{P}_{\rm e}^{\rm sy}[j]$ values in each interval. The circles (discs) correspond to cases where $\hat{P}_{\rm e}^{\rm sy}[j]$ is within (outside) the 95 percentile boundaries.

D. Validation of the Analytic Expression $P_{\rm e}^{\rm ge}$

In this section we experimentally validate the analytic expression of the bit-error probability $P_{\rm e}^{\rm ge}$. In the previous section, we have discussed the use of PCA for decorrelating the feature components and for reducing the dimension to



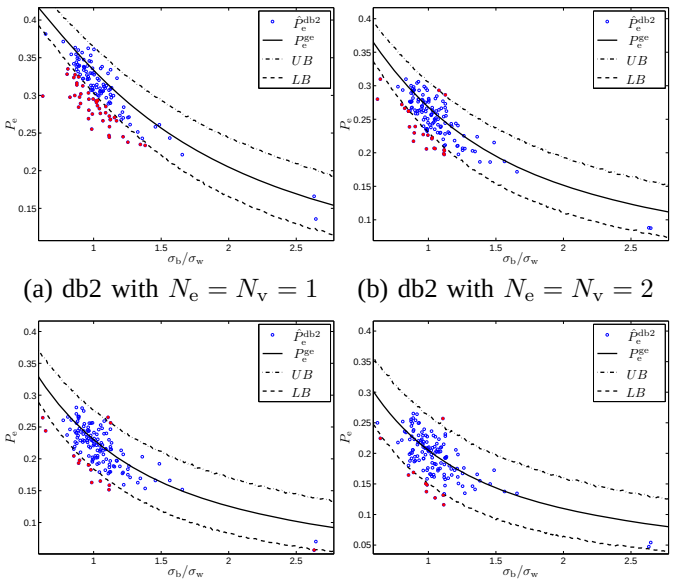
(a) db1 with $N_{\rm e} = N_{\rm v} = 1$ (b) db1 with $N_{\rm e} = N_{\rm v} = 2$

(c) db1 with $N_{\rm e} = N_{\rm v} = 3$ (d) db1 with $N_{\rm e} = N_{\rm v} = 4$

Fig. 12. Comparison between $P_{\rm e}^{\rm ge}[j]$ and $\hat{P}_{\rm e}^{\rm db1}[j]$ for different settings (a) $N_{\rm e} = N_{\rm v} = 1$, (b) $N_{\rm e} = N_{\rm v} = 2$, (c) $N_{\rm e} = N_{\rm v} = 3$, and (d) $N_{\rm e} = N_{\rm v} = 4$. The circles (discs) correspond to cases where $\hat{P}_{\rm e}^{\rm db1}[j]$ falls within (outside) the boundaries.

$N_{\rm F} = 31$. In order to have more components for the validation we apply PCA but without reducing the number of features. Hence, we consider the original number of features (696) for database db1. However, for database db2 we only consider 223 components since 25% of the total number of subjects (i.e. 28 subjects) with a total of 224 feature vectors were used to derive the PCA projection. Thus, to avoid singularities we have reduced the number of features to 223.

To assess the model assumptions, we compared the estimated bit-error probability of the biometric database $\hat{P}_{\rm e}^{\rm db}[j]$ with the corresponding analytically obtained $P_{\rm e}^{\rm ge}[j]$. The same test protocol is used as discussed in Section IV-C. The experimental results for db1 and db2 for different values of $N_{\rm e}$ and $N_{\rm v}$ are shown in Fig. 12 and Fig. 13, respectively. The circles (discs) correspond to cases where $\hat{P}_{\rm e}^{\rm db}[j]$ is within (outside) the 95 percentile boundaries. We refer to the number of discs as the estimation error $\epsilon_{P_{\rm e}}$. If all the assumptions hold then we expect the relative $\epsilon_{P_{\rm e}}$ to be around 5%. Table IV reports the absolute and relative $\epsilon_{P_{\rm e}}$. Because $\epsilon_{P_{\rm e}}$ is noisy due to the random selection of $N_{\rm e}$ and $N_{\rm v}$ samples within the test protocol, we repeat the estimation 20 times and report its mean. For db1, $\epsilon_{P_{\rm e}}$ is 16.7% for $N_{\rm e} = N_{\rm v} = 1$ and decreases to 13% for $N_{\rm e} = N_{\rm v} = 4$. In the case of db2, $\epsilon_{P_{\rm e}}$ is very large; 27.3% for $N_{\rm e} = N_{\rm v} = 1$ but decreases significantly when both $N_{\rm e}$ and $N_{\rm v}$ are increased, reaching 6.3% when $N_{\rm e} = N_{\rm v} = 4$. Thus, for both databases there is a clear improvement when increasing the number of samples. We conjecture that the improved bit-error probability estimation performance is due to the fact that the feature value distribution becomes more Gaussian when averaging multiple samples as stated by the central limit theorem [24]. Also note that many $\hat{P}_{\rm e}^{\rm db1}[j]$ estimations of db1 are very close to the



(a) db2 with $N_e = N_v = 1$ (b) db2 with $N_e = N_v = 2$

(c) db2 with $N_e = N_v = 3$ (d) db2 with $N_e = N_v = 4$

Fig. 13. Comparison between $P_e^{ge}[j]$ and $\hat{P}_e^{db2}[j]$ for different settings (a) $N_e = N_v = 1$, (b) $N_e = N_v = 2$, (c) $N_e = N_v = 3$, and (d) $N_e = N_v = 4$. The circles (discs) correspond to cases where $\hat{P}_e^{db2}[j]$ falls within (outside) the boundaries.

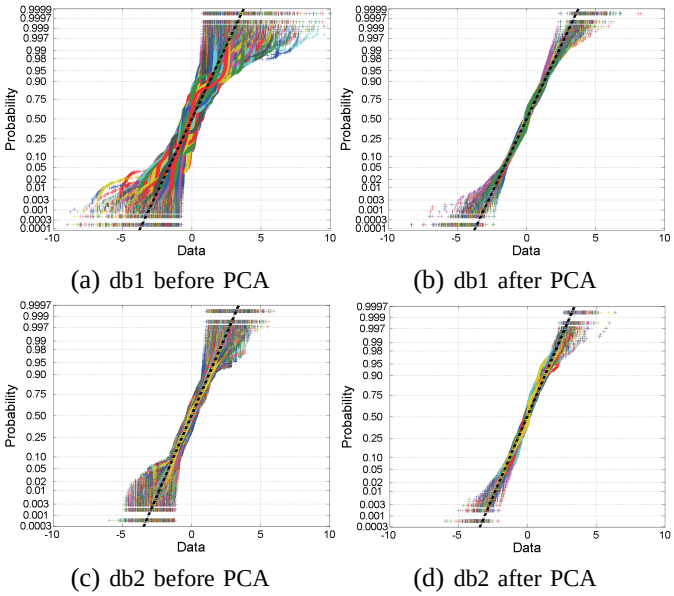
TABLE IV
THE NUMBER OF CASES ϵ_{P_e} WHERE $\hat{P}_e^{db}[j]$ IS OUTSIDE THE 95% PERCENTILE BOUNDARIES PER DATABASE AND $\{N_e, N_v\}$ SETTING.

	db1		db2	
Setting	Abs. ϵ_{P_e}	Rel. ϵ_{P_e}.	Abs. ϵ_{P_e}	Rel. ϵ_{P_e}
$N_e = N_v = 1$	116	16.7 %	61	27.3%
$N_e = N_v = 2$	103	14.8 %	33	14.8%
$N_e = N_v = 3$	91	13.1 %	18	8.1%
$N_e = N_v = 4$	92	13.2 %	14	6.3%

95 percentile boundaries, hence small estimation errors can lead to large variation in ϵ_{P_e} that could explain the bit-error probability estimation performance differences between db1 and db2 observed in the table.

E. The effect of PCA on the Gaussian Assumption

As described in Section II, the analytic framework is based on the Gaussian model assumption. Figs. 14(a)(c) show the normal probability plot for each component of the feature vectors of db1 and db2 respectively, before applying the PCA transformation. The normal probability plot is a graphical technique for assessing the degree to which a dataset approximates a Gaussian distribution. If the curve of the data closely follows the dashed-thick line then the data can be assumed to be approximately Gaussian distributed. Prior to comparing, we normalized each feature so that it has zero-mean and unit-variance. For both databases it is evident that the distributions before applying PCA are not Gaussian, because they significantly deviate from the dashed-thick line that represents a perfect Gaussian distribution. Figs. 14 (b)(d) depict the normal probability plot for each of the 696 components of db1 and



(a) db1 before PCA (b) db1 after PCA

(c) db2 before PCA (d) db2 after PCA

Fig. 14. Normal probability plot of each feature vector component of db1 and db2 before and after applying PCA.

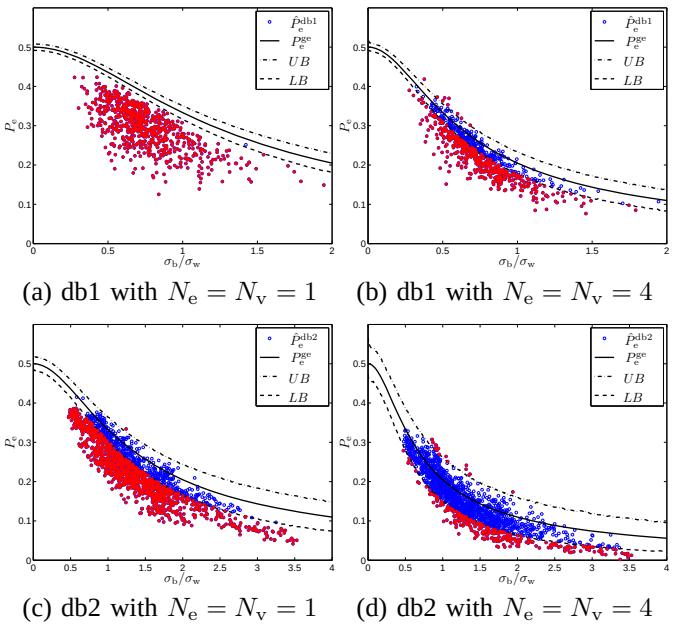
the 223 components of db2 respectively, after applying PCA. For both databases the figures show that after applying PCA the features tend to behave more like Gaussians. Yet, the tails deviate the most from being Gaussian where for the most cases the empirical distribution is wider.

Fig. 15 shows the P_e estimations before applying PCA for both databases in two cases: $N_e = N_v = 1$ and $N_e = N_v = 4$. Note that before PCA db1 and db2 have 696 and 1536 components, respectively. For db1 ϵ_{P_e} is equal to 99.8% for the $N_e = N_v = 1$ and 61.2% for the $N_e = N_v = 4$ case, while for db2 ϵ_{P_e} is 71% and 18%, respectively. Comparing these results with the ϵ_{P_e} values when applying PCA, see Table IV, we can also conclude that applying PCA makes the features significantly more Gaussian.

F. Validation of the Analytic Expression of FRR and FAR

For both db1 and db2, we analytically estimate the genuine $\phi_{ge}(k)$ and imposter $\phi_{im}(k)$ Hamming distance pmfs, and the $\beta(T)$ and $\alpha(T)$ curves. The results are presented in Fig. 16 and Fig. 17 for db1 and db2, respectively. The experimentally calculated pmfs are indicated by 'Exp' while the ones obtained using the analytical model are indicated by 'Mod'. The experimental results are obtained using the same protocol as the one discussed in Section IV-C, but storing the Hamming distance pmfs of each subject instead. We focus on the cases corresponding to $N_F = 31$ with $N_e = N_v = 1$ and $N_e = N_v = 4$.

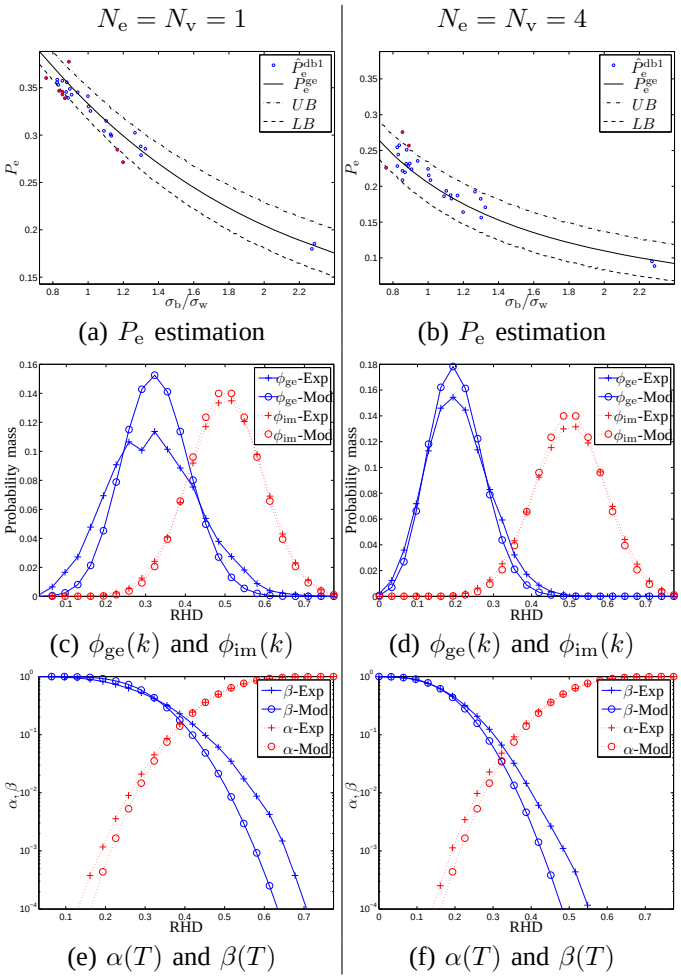
Both Fig. 16 and Fig. 17 indicate that there is a good agreement between $\phi_{im}(k)$-Exp and $\phi_{im}(k)$-Mod. Large differences are observed between $\phi_{ge}(k)$-Exp and $\phi_{ge}(k)$-Mod. However, the differences decrease when both N_e and N_v are increased. Averaging multiple independent samples leads to a higher Gaussianity degree in accordance with the central limit theorem. This effect was also observed for the P_e estimation results in previous section. It is interesting to note



(a) db1 with $N_e = N_v = 1$ (b) db1 with $N_e = N_v = 4$

(c) db2 with $N_e = N_v = 1$ (d) db2 with $N_e = N_v = 4$

Fig. 15. $\hat{P}_e^{dbx}[j]$ at different settings of N_e and N_v for both db1 and db2 before applying the PCA transform.



(a) P_e estimation (b) P_e estimation

(c) $\phi_{ge}(k)$ and $\phi_{im}(k)$ (d) $\phi_{ge}(k)$ and $\phi_{im}(k)$

(e) $\alpha(T)$ and $\beta(T)$ (f) $\alpha(T)$ and $\beta(T)$

Fig. 16. Results for db1 with $N_F = 31$, (a)(b) $\hat{P}_e^{db1}$ and the analytical estimation of P_e^{ge}, (c)(d) $\phi_{ge}(k)$ and $\phi_{im}(k)$ pmfs, and (e)(f) the $\alpha(T)$ and $\beta(T)$ curves. The graphs on the left (right) correspond to $N_e = N_v = 1$ ($N_e = N_v = 4$).

the differences between the estimation errors of $\phi_{ge}(k)$ of db1 and db2. For db1 the center of gravity of $\phi_{ge}(k)$-Exp and $\phi_{ge}(k)$-Mod practically coincide. The only difference is the width of the pmfs, since the experimentally obtained pmf is wider than the theoretical one. In case of db2, we see that there is both an alignment and a width error, $\phi_{ge}(k)$-Exp is skewed to the left.

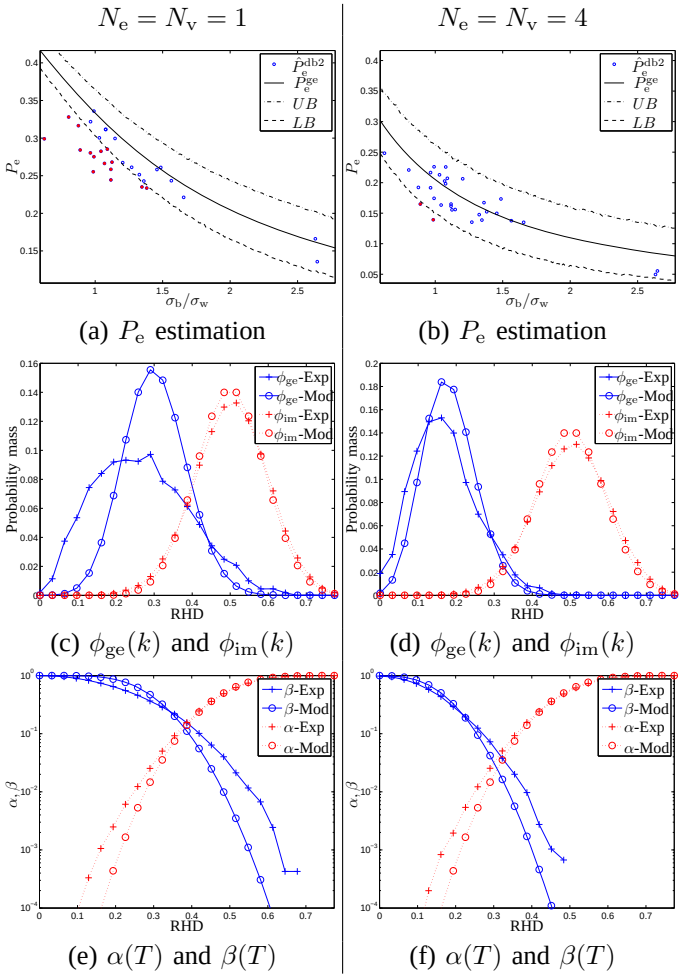
Eventually, we are interested in estimating the DET curves. Because the DET curves combine both β and α, they are thus prone to estimation errors associated with β or α. The DET curves for db1 and db2 for $N_F = 31$ with different values of N_e and N_v are shown in Fig. 18. From these figures we can conclude that increasing N_e and N_v leads to greater estimation errors of the DET curve, which contradicts the previous finding that increasing N_e and N_v leads to better estimations of P_e and $\phi_{ge}(k)$. This can be explained by the fact that in the $N_e = N_v = 4$ case, the area of interest with $\beta(T) \in [0.01, 0.1]$ occurs for smaller values of $\alpha(T)$, because the number of bit errors decreases when N_e and N_v increase, i.e. the performance improves. As shown by the $\alpha(T)$ curves in Fig. 16 and Fig. 17, there is a greater estimation error at smaller values of $\alpha(T)$ thus amplifying the estimation error of the DET curve.

A summary of the probable causes for the observed differences, starting from the most probable, are (i) the non-homogeneous within-class variance (ii) the dependency between features, and (iii) the dependency between bit errors. Database db2 seems to be clearly not adhering to the homogeneous within-class variance assumption, resulting into a skewed $\phi_{ge}(k)$ with a large tail. Such a tail is caused by subjects that have on average a worse performance than the other subjects. These subjects have many feature components with a larger within-class variance leading to larger $P_e[j]$ values and thus greater Hamming distances. In the literature these subjects are referred to as goats [25], [26]. If the features are dependent, then the Hamming distance pmf becomes wider while keeping its original mean. This effect is visible for both $\phi_{ge}(k)$ and $\phi_{im}(k)$ for both databases. On the other hand, certain disturbances such as occluded biometric images or strong biometric variabilities can cause multiple errors to occur simultaneously. Thus, the bit errors are dependent causing the tails on the right side of the genuine Hamming distance pmf. A right tail is slightly visible for db1, but is clearly present for db2 as illustrated in Figs. 16(c)(d) and Figs. 17(c)(d), respectively.

In Section V we propose a modified model that incorporates the non-homogeneous within-class variance property, while in Section VI we further extend the model to include dependencies.

V. Relaxing the Homogenous Within-Class Variance Assumption

In this section we propose a modified model that takes the non-homogeneous property into account, while still assuming

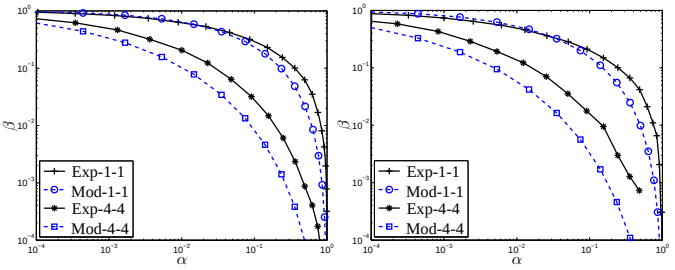


(a) P_e estimation (b) P_e estimation

(c) $\phi_{ge}(k)$ and $\phi_{im}(k)$ (d) $\phi_{ge}(k)$ and $\phi_{im}(k)$

(e) $\alpha(T)$ and $\beta(T)$ (f) $\alpha(T)$ and $\beta(T)$

Fig. 17. Results for db2 with $N_F = 31$, (a)(b) $\hat{P}_e^{db2}$ and the analytical estimation of P_e^{ge}, (c)(d) $\phi_{ge}(k)$ and $\phi_{im}(k)$ pmfs, and (e)(f) the $\alpha(T)$ and $\beta(T)$ curves. The graphs on the left (right) correspond to $N_e = N_v = 1$ ($N_e = N_v = 4$).



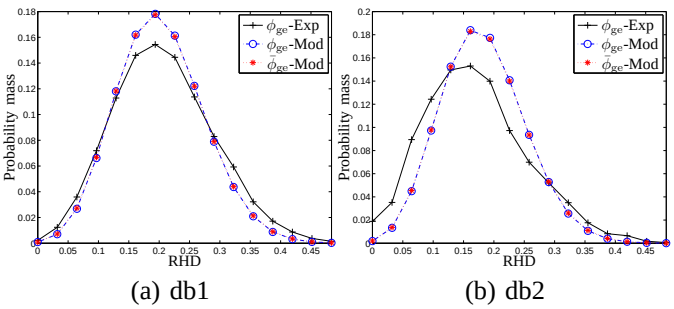
(a) db1 with $N_F = 31$ (b) db2 with $N_F = 31$

Fig. 18. DET curves for both db1 and db2 for $N_F = 31$ with different values of N_e, and N_v. The values N_e and N_v are indicated in the legend in the subsequent order. The experimentally obtained curves are denoted by 'Exp' while the analytical by 'Mod'.

independent feature components. The proposed method makes use of the approximation of the convolution of (2) with the binomial pmf. For the genuine case, this would be

$$\bar{\phi}_{ge}(k) = \binom{N_F}{k} (\bar{P}_e^{ge})^k (1 - \bar{P}_e^{ge})^{N_F - k}, \quad (26)$$

where $\bar{P}_e^{ge}$ is the average bit-error probability across the feature components $\bar{P}_e^{ge} = 1/N_F \sum_{j=1}^{N_F} P_e^{ge}[j]$. The approximate



(a) db1 (b) db2

Fig. 19. The approximation of the genuine Hamming distance pmf as binomial with $\bar{P}_e$ ((26)) for the $N_e = N_v = 4$ case with $N_F = 31$.

pmfs $\bar{\phi}_{ge}(k)$ are depicted in Fig. 19(a) for db1 and Fig. 19(b) for db2 for the $N_e = N_v = 4$ case with $N_F = 31$. For both databases, the approximation is reasonably accurate.

Thus we can model the non-homogeneous effect by assuming that $\bar{P}_{e,i}^{ge}$ is not equal for each subject and is distributed according to a probability density $p_{\bar{P}_e^{ge}}$. The following step consists in determining the pdf $p_{\bar{P}_e^{ge}}$ across the population and computing the average genuine Hamming distance pmf defined as

$$\bar{\Phi}_{ge}(k) = \int_0^{1/2} p_{\bar{P}_e^{ge}}(\tau) \bar{\phi}_{ge}(k|\tau) d\tau, \quad (27)$$

where the integral limits are due to the fact that $P_e \in [0, 1/2]$ and $\bar{\phi}_{ge}(k|\tau)$ is the generic case of (26) as

$$\bar{\phi}_{ge}(k|\tau) = \binom{N_F}{k} (\tau)^k (1 - \tau)^{N_F - k}. \quad (28)$$

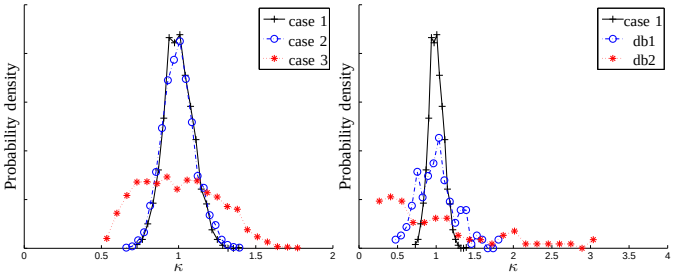
We propose a method for estimating $p_{\bar{P}_e^{ge}}$ using only the estimated within-class variance of each subject $\hat{\sigma}_{w,i}^2[j]$. Because of the limited number of samples N_i, we know from [23] that the estimation ratio $((N_i - 1)\hat{\sigma}_{w,i}^2[j])/\sigma_w^2[j]$ follows the χ^2 distribution with $N_i - 1$ degrees of freedom, where $\sigma_w^2[j]$ is the underlying within-class variance that has to be estimated and is assumed to be homogeneous. However, in practice $\sigma_w^2[j]$ is unknown, therefore we have to replace it by its estimate $\hat{\sigma}_w^2[j]$. It is well known that the mean associated with a χ^2 distribution is equal to its number of degrees of freedom, thus by omitting the $(N_i - 1)$ multiplications it becomes unit mean.

The next step is to take the average ratio over all feature components as

$$\kappa_i = \frac{1}{N_F} \sum_{j=1}^{N_F} \hat{\sigma}_{w,i}^2[j] / \hat{\sigma}_w^2[j]. \quad (29)$$

We can model the non-homogeneous property by assuming that for all components of subject i the within-class variance is $\sigma_{w,i}^2[j] = \kappa_i \sigma_w^2[j]$. If the homogeneous assumption holds and the number of features is large, then the pdf of κ_i across the whole population becomes Gaussian with unit mean and a variance that decreases when N_F increases. The variance decreases at larger values of N_F because this would be similar to having N_F times more samples and therefore a better estimation of its mean. When there are "goat-like" subjects, the homogeneous assumption does not hold, then the variance of the pdf of κ_i increases.

Fig. 20(a) shows the empirically estimated pdf of κ_i for



(a) Synthetic cases (b) db1 & db2

Fig. 20. Empirical estimated probability density p_{κ_i} using synthetic databases (a) of 2000 subjects with $N_{\rm F} = 31, N_i = 8, \sigma_{\rm b}^2[j] = 1$, where for 'case 1' every subject has the same $\sigma_{{\rm w},i}^2[j] = 1$, in 'case 2' $\sigma_{{\rm w},i}^2[j] = 1 + \nu_i[j]$, and for 'case 3' $\sigma_{{\rm w},i}^2[j] = 1 + \nu_i$ where ν_i is drawn from U(-0.4,0.4) and is redrawn for each feature component separately in 'case 2'. In (b) the comparison between 'case 1', db1, and db2 is shown.

a synthetically generated databases containing 2000 subjects with $N_{\rm F} = 31$, $N_i = 8$, and $\sigma_{\rm b}^2[j] = 1$, where for 'case 1' every subject has the same $\sigma_{{\rm w},i}^2[j] = 1$, in 'case 2' $\sigma_{{\rm w},i}^2[j] = 1 + \nu_i[j]$, and for 'case 3' $\sigma_{{\rm w},i}^2[j] = 1 + \nu_i$ where ν_i is drawn from U(-0.4,0.4) and is redrawn for each feature component separately in 'case 2'. The results imply that the variance of the κ_i pdf increases when $\sigma_{{\rm w},i}^2[j]$ is different for each subject ('case 2') and increases significantly when there is a positive correlation with the variance offset, for example when subjects have all their $\sigma_{{\rm w},i}^2[j]$ larger or smaller than the average value ('case 3'). Hence, in 'case 3' there is a clear existence of goats or doves, where the latter are the subjects that have a small number of bit errors when matched against themselves [27].

Fig. 20(b) compares the κ_i pdf of 'case 1', db1, and db2. The results show that both db1 and db2 do not adhere to the homogeneous property. The κ_i pdf found for db1 looks similar to 'case 3'. However, the pdf found for db2 significantly deviates from the synthetic cases, which confirms the existence of goats and doves. This may also explain the significant discrepancy found when estimating the genuine Hamming distance pmfs of db2 as shown in Fig. 17.

Now we can empirically estimate the probability density $p_{\bar{P}_{\rm e}^{\rm ge}}$ using p_{κ_i}. The relationship between κ_i and $\bar{P}_{{\rm e},i}^{\rm ge}$ is given by

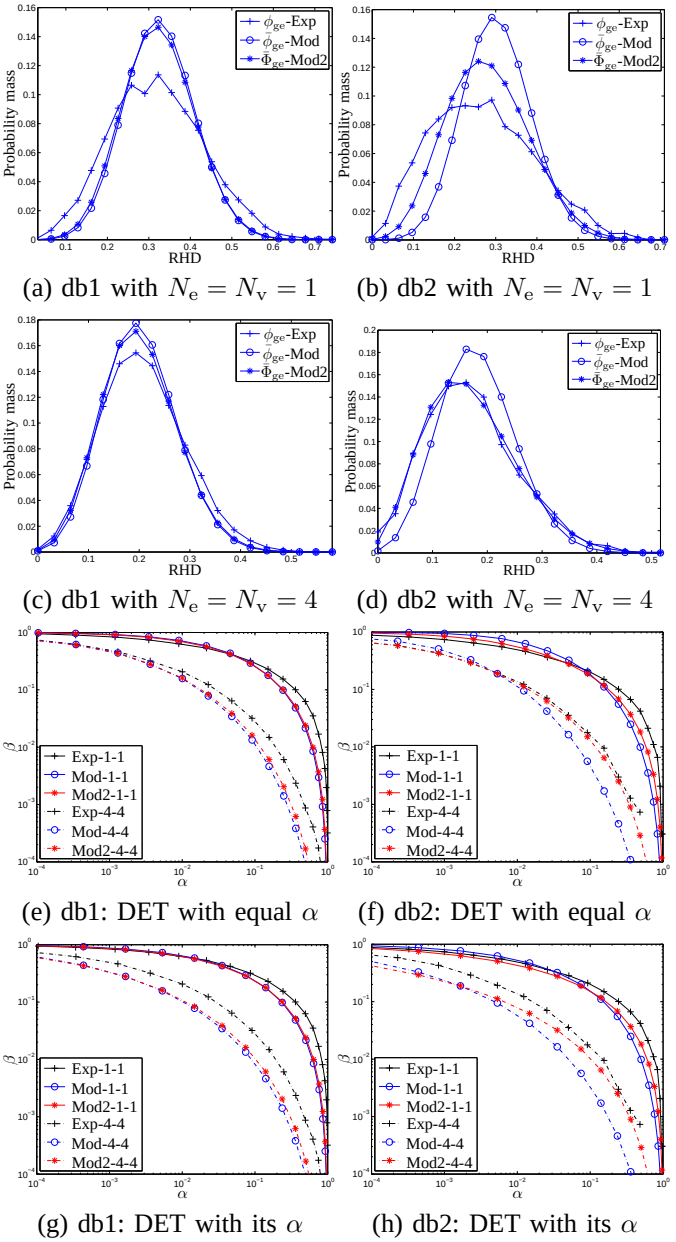
$$\bar{P}_{{\rm e},i}^{\rm ge} = \frac{1}{N_{\rm F}} \sum_{j=1}^{N_{\rm F}} P_{\rm e}^{\rm ge}(N_{\rm e}, N_{\rm v}, \sqrt{\kappa_i \hat{\sigma}_{\rm w}^2[j]}, \hat{\sigma}_{\rm b}[j]), \qquad (30)$$

where we take the average of $P_{\rm e}^{\rm ge}[j]$ across all features, while using $\hat{\sigma}_{\rm b}[j]$ and the modified within-class variance estimation $\sqrt{\kappa_i \hat{\sigma}_{\rm w}^2[j]}$. Because of the nonlinear relationship between $P_{\rm e}^{\rm ge}[j]$ and $\hat{\sigma}_{\rm w}[j]$ we take the average over $P_{\rm e}^{\rm ge}[j]$ instead of estimating $P_{\rm e}^{\rm ge}$ using the average of $\hat{\sigma}_{\rm w}[j]$.

In practice, we can rewrite (27) as:

$$\bar{\Phi}_{\rm ge}(k) = \frac{1}{N_{\rm s}} \sum_{i=1}^{N_{\rm s}} \bar{\phi}_{\rm ge}(k|\bar{P}_{{\rm e},i}^{\rm ge}). \qquad (31)$$

We applied this new method for estimating $\phi_{\rm ge}(k)$ of db1 and db2 and the results are shown in Figs. 21(a-d) for the $N_{\rm e} = N_{\rm v} = 1$ and $N_{\rm e} = N_{\rm v} = 4$ cases with $N_{\rm F} = 31$, where $\phi_{\rm ge}(k)$-Exp is the experimentally obtained pmf, $\phi_{\rm ge}(k)$-Mod is obtained using (2), and $\bar{\Phi}_{\rm ge}(k)$-Mod2 with (31). The results show that $\phi_{\rm ge}$-Exp is better approximated when using the new method $\bar{\Phi}_{\rm ge}(k)$-Mod2. In case of db1 there is a small improvement, but for db2 there is a significant improvement and even a better estimation is obtained when $N_{\rm e} = N_{\rm v} = 4$. Furthermore, Figs. 21(e-h) show the DET curve results. In Figs. 21(e)(f) the same α is used for each DET curve in order to isolate the estimation errors of $\phi_{\rm ge}(k)$, while in Figs. 21(g)(h) α-Exp is used for the 'Exp' curves and α-Mod is used for both the 'Mod' and 'Mod2' curves. With



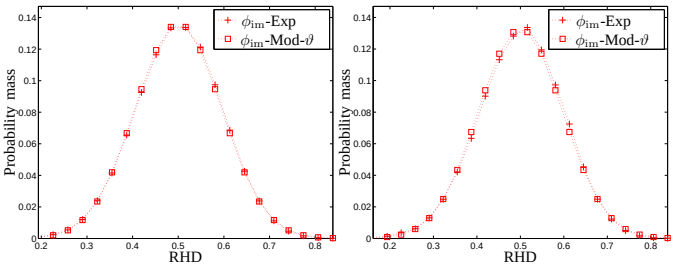
(a) db1 with $N_{\rm e} = N_{\rm v} = 1$ (b) db2 with $N_{\rm e} = N_{\rm v} = 1$

(c) db1 with $N_{\rm e} = N_{\rm v} = 4$ (d) db2 with $N_{\rm e} = N_{\rm v} = 4$

(e) db1: DET with equal α (f) db2: DET with equal α

(g) db1: DET with its α (h) db2: DET with its α

Fig. 21. Results of the proposed method incorporating the non-homogeneous property of db1 and db2 for the cases $N_{\rm e} = N_{\rm v} = 1$ and $N_{\rm e} = N_{\rm v} = 4$ with $N_{\rm F} = 31$. Figures (a-d) show the Hamming distance pmf estimations while figures (e-h) show the DET curves estimation, where 'Mod' and 'Mod2' indicate the modeling method without and with the non-homogeneous property, respectively. In (e) and (f) all the DET curves are plotted using the experimentally obtained α-Exp, while in (g) and (h) we use the α-Exp for the 'Exp' curves and α-Mod for both the 'Mod' and 'Mod2' curves.



(a) db1 $\vartheta_{opt} = 1.11$ (b) db2 $\vartheta_{opt} = 1.17$

Fig. 22. Results of estimating ϑ_{opt} from ϕ_{im}-Exp using (33) for the $N_{\text{e}} = N_{\text{v}} = 1$ case for both databases. The variance corrected Gaussian approximated curve as described by (32) is depicted as ϕ_{im}-Mod-ϑ.

the new method the DET curve estimation has improved, most significantly for db2. However, the differences between Figs. 21(e)(f) and Figs. 21(g)(h) clearly indicate that the remaining estimation errors are caused by the estimation of α. As shown in Figs. 16(c)(d) and Figs. 17(c)(d) there is an estimation error of ϕ_{im}, which we consider to be caused by the fact that the feature components are dependent.

VI. Incorporating Feature Component Dependencies

In previous section we observed that a significant part of the remaining DET estimation errors is related to the estimation errors of the ϕ_{im}-Exp pmf. In this section we propose a further extension of the analytical framework in order to incorporate dependencies between feature components. We propose to estimate the dependency from the ϕ_{im} pmf and apply it to the ϕ_{ge} pmf estimation. Hence, we assume that both pmfs are influenced by the dependency to the same extent.

We estimate the dependency from ϕ_{im}-Exp by fitting it with a Gaussian approximation of the binomial pmf of (9) with the variance as the fitting parameter. For large values of N_{F}, the binomial pmf with probability P_{e} and dimension N_{F} can be approximated by the Gaussian density $\mathcal{N}(N_{\text{F}}P_{\text{e}}, N_{\text{F}}P_{\text{e}}(1-P_{\text{e}}))$, with mean $N_{\text{F}}P_{\text{e}}$ and variance $N_{\text{F}}P_{\text{e}}(1-P_{\text{e}})$. For the imposter case we know that $P_{\text{e}} = 1/2$, from which its mean and variance become $N_{\text{F}}/2$ and $N_{\text{F}}/4$, respectively. Hence, the Gaussian approximation of the ϕ_{im}-Exp pmf with the variance parameter ϑ used for fitting becomes

$$\begin{aligned}\phi_{\text{im}}(k)\text{-Mod-}\vartheta &= \frac{1}{\sqrt{2\pi\vartheta\sigma^2}}\text{e}^{-\frac{(k-\mu)^2}{2\vartheta\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi\vartheta N_{\text{F}}P_{\text{e}}(1-P_{\text{e}})}}\text{e}^{-\frac{(k-N_{\text{F}}P_{\text{e}})^2}{2\vartheta N_{\text{F}}P_{\text{e}}(1-P_{\text{e}})}} \\ &= \frac{2}{\sqrt{2\pi\vartheta N_{\text{F}}}}\text{e}^{-\frac{(2k-N_{\text{F}})^2}{2\vartheta N_{\text{F}}}},\end{aligned} \tag{32}$$

where the optimal ϑ is computed by minimizing the mean-square error (MMSE) as

$$\vartheta_{opt} = \arg\min_{\vartheta} \sum_{k=0}^{N_{\text{F}}} \Big(\phi_{\text{im}}(k)\text{-Exp} - \phi_{\text{im}}(k)\text{-Mod-}\vartheta\Big)^2. \tag{33}$$

The estimation results of ϑ_{opt} for the $N_{\text{e}} = N_{\text{v}} = 1$ case are shown in Fig. 22 for both databases. The optimal value of ϑ_{opt} is 1.11 for db1 and 1.17 for db2. For both databases ϑ_{opt} is very similar, which may indicate that the amount of dependencies between the feature components is relative similar for both databases. Furthermore, the ϕ_{im}-Exp pmf is better estimated when compared to its first estimation disregarding the feature component dependencies as depicted in Fig. 16(c) and Fig. 17(c) for db1 and db2, respectively.

With the Gaussian approximation including the variance correction with ϑ_{opt} we have a better estimation of the ϕ_{ge} pmf by rewriting (31) as

$$\bar{\Phi}_{\text{ge}}(k) = \frac{1}{N_{\text{s}}}\sum_{i=1}^{N_{\text{s}}} \frac{1}{\sqrt{2\pi\sigma_{cor}^2}}\text{e}^{\frac{-\left(k-\bar{P}_{\text{e},i}^{\text{ge}}N_{\text{F}}\right)^2}{2\sigma_{cor}^2}}, \tag{34}$$

with $\sigma_{cor}^2 = \vartheta_{opt}N_{\text{F}}\bar{P}_{\text{e},i}^{\text{ge}}(1-\bar{P}_{\text{e},i}^{\text{ge}})$. Because of the Gaussian approximation errors it does not hold that the sum of the probability mass equals to one, therefore we normalize it according to

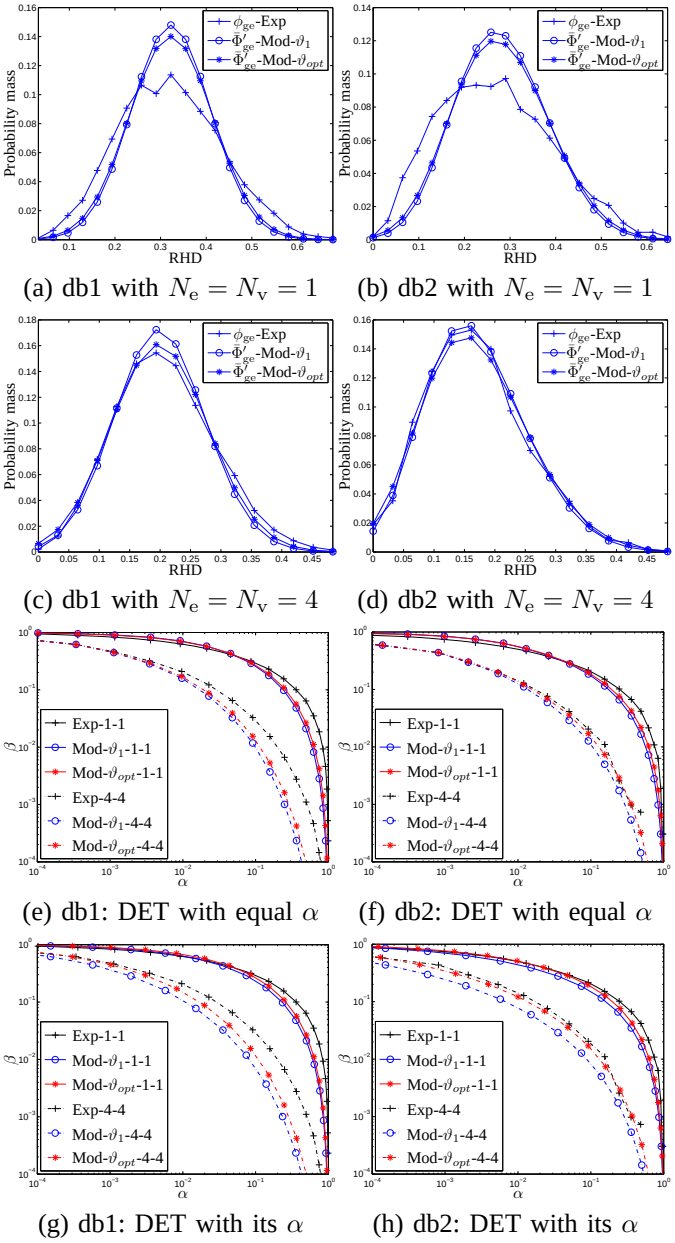
$$\bar{\Phi}'_{\text{ge}}(k) = \frac{1}{\sum_{k=0}^{N_{\text{F}}}\bar{\Phi}_{\text{ge}}(k)}\bar{\Phi}_{\text{ge}}(k). \tag{35}$$

The estimation results using (35) for the cases of $\vartheta = 1$ and $\vartheta = \vartheta_{opt}$ are depicted in Fig. 23. For the $\vartheta = 1$ case the Gaussian approximation is used without the variance correction. Figs. 23(a-d) show that the $\phi_{\text{ge}}(k)$ pmf estimation has slightly improved. The $\bar{\Phi}'_{\text{ge}}$-Mod-ϑ_{opt} curve is closer to $\phi_{\text{ge}}(k)$-Exp than $\bar{\Phi}'_{\text{ge}}$-Mod-ϑ_1. This holds across the whole curve for the $N_{\text{e}} = Nv = 1$ case and mainly for the right tail for the $N_{\text{e}} = Nv = 4$ case. The same conclusions are also portrayed by the DET curves of Figs. 23(e-f), where each DET curve uses the same α curve, namely the experimentally obtained α-Exp in order to isolate the ϕ_{ge}(k) pmf estimation errors. The DET curves in Figs. 23(g-h) use the actual α curves, thus α-Mod-ϑ_1 for the DET-Mod-ϑ_1 curves and α-Mod-ϑ_{opt} for the DET-Mod-ϑ_{opt} curves, respectively. The curves show that the DET-Mod-ϑ_{opt} curve is clearly closer to DET-Exp curve, because α-Mod-ϑ_{opt} is a better approximation of α-Exp as we have shown earlier.

VII. Practical Considerations

In previous sections we have presented several analytical models for estimating the DET performance curve. However, as stated previously, because of the use of an ECC the FRR is lower bounded because of the limited number of bits the ECC can correct. For the setting of $N_{\text{F}} = 31$, which equals the codeword length n_{c}, the BCH ECC can correct up to 7 bits as shown in Table I. The experimentally achieved performance and its analytical estimates at this operating point are given in Table V. The results indicate that at this operating point there is not a significant difference between the estimations using the 'Mod' and 'Mod2' models, while the 'Mod-ϑ_{opt}' estimator leads to the best estimation where its significant improvement is of the α.

Although we have presented an analytical framework for analysis, it could also be used in practical cases. For example, consider the scenario where a database has been collected with a maximum of five samples per subject. Hence, the performance could only be calculated for cases where $N_{\text{e}} + N_{\text{v}} \leq 5$. However, this restriction does not hold for our proposed analytical framework. By estimating σ_{w}^2, σ_{b}^2, κ_i, and ϑ_{opt} from



(a) db1 with $N_e = N_v = 1$ (b) db2 with $N_e = N_v = 1$

(c) db1 with $N_e = N_v = 4$ (d) db2 with $N_e = N_v = 4$

(e) db1: DET with equal α (f) db2: DET with equal α

(g) db1: DET with its α (h) db2: DET with its α

Fig. 23. Results of the proposed method incorporating both the dependency and non-homogeneous property of db1 and db2 for the cases $N_e = N_v = 1$ and $N_e = N_v = 4$ with $N_F = 31$. Figures (a-d) show the ϕ_{ge} estimations, while (e-h) show the DET curves estimation. The label 'Mod-ϑ_1' indicates the new modeling method but with $\vartheta = 1$, hence using only the Gaussian approximation of the binomial pmf including the non-homogeneous property. The label 'Mod-ϑ_{opt}' indicates the cases where $\vartheta = \vartheta_{opt}$. In (e) and (f) all the DET curves are plotted using the experimentally obtained α-Exp, while in (g) and (h) we use the α-Exp for the 'Exp' curves, α-Mod-ϑ_1 for the 'Mod-ϑ_1' curves and α-Mod-ϑ_{opt} for the 'Mod-ϑ_{opt}' curves.

the given database, the performance could be estimated for the cases where $N_e + N_v \geq 5$. Either the performance could be estimated for a specific N_e and N_v setting or the lower bounds of the N_e and N_v setting could be estimated in order to obtain a certain performance or better. Given the same scenario as for Table V where the performance is estimated at the maximum error capability of the ECC for both databases, db1 is expected to reach $\beta \leq 0.1$ when $N_e = N_v \geq 8$, while $N_e = N_v \geq 7$ for db2.

VIII. Conclusions

We have proposed an analytical framework for estimating the DET performance curve of a biometric system, based on binary feature vectors, for different settings of N_e and N_v.

The first proposed estimation method used a simple Parallel Gaussian Channel framework for modeling the pdf of the real-valued features. Each component has its own channel with the corresponding additive Gaussian noise representing the biometric variability and measurement noise, called the within-class variability. The results showed significant estimation errors and were far from optimal, mainly because of the homogeneous within-class variance assumption. Consequently we proposed a modified framework to incorporate the non-homogeneous property, which in fact assumes that the within-class variance is different for each subject. The estimation improved significantly and the remaining estimation error is thought to be caused by the estimation errors of the false acceptance curve due to dependency between feature components and corresponding bits. The final proposed framework also incorporated feature component dependency, whose value was derived from the calculated imposter Hamming distance pmf of the database. This method resulted in the most optimum estimation of the DET performance curves.

References

[1] A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *6th ACM Conference on Computer and Communications Security*, November 1999, pp. 28–36.

[2] E. J. C. Kelkboom, B. 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*, Seoul, Korea, August 2007, pp. 566–573.

[3] T. A. M. Kevenaar, G.-J. Schrijen, A. H. M. Akkermans, M. van der Veen, and F. Zuo, "Face recognition with renewable and privacy preserving binary templates," in *4th IEEE workshop on AutoID*, Buffalo, New York, USA, October 2005, pp. 21–26.

[4] J.-P. Linnartz and P. Tuyls, "New shielding functions to enhance privacy and prevent misuse of biometric templates," in *4th Int. Conf. on AVBPA*, 2003.

TABLE V
THE EXPERIMENTALLY ('EXP') ACHIEVED α AND β AND ITS ANALYTICAL ESTIMATES USING THE SIMPLISTIC MODEL ('MOD'), THE MODEL RELAXING THE HOMOGENOUS PROPERTY ('MOD2'), AND THE MODEL ALSO INCORPORATING THE FEATURE COMPONENT DEPENDENCIES ('MOD-ϑ_{opt}').

	db1			
	$N_e = N_v = 1$		$N_e = N_v = 4$	
	α	β	α	β
Exp	$3.59 \cdot 10^{-3}$	$7.33 \cdot 10^{-1}$	$3.73 \cdot 10^{-3}$	$3.17 \cdot 10^{-1}$
Mod	$1.66 \cdot 10^{-3}$	$8.43 \cdot 10^{-1}$	$1.66 \cdot 10^{-3}$	$2.79 \cdot 10^{-1}$
Mod2	$1.66 \cdot 10^{-3}$	$8.25 \cdot 10^{-1}$	$1.66 \cdot 10^{-3}$	$2.77 \cdot 10^{-1}$
Mod-ϑ_{opt}	$3.06 \cdot 10^{-3}$	$8.15 \cdot 10^{-1}$	$3.06 \cdot 10^{-3}$	$2.94 \cdot 10^{-1}$
	db2			
	$N_e = N_v = 1$		$N_e = N_v = 4$	
	α	β	α	β
Exp	$6.35 \cdot 10^{-3}$	$5.58 \cdot 10^{-1}$	$5.28 \cdot 10^{-3}$	$1.96 \cdot 10^{-1}$
Mod	$1.66 \cdot 10^{-3}$	$7.66 \cdot 10^{-1}$	$1.66 \cdot 10^{-3}$	$1.88 \cdot 10^{-1}$
Mod2	$1.66 \cdot 10^{-3}$	$6.31 \cdot 10^{-1}$	$1.66 \cdot 10^{-3}$	$1.88 \cdot 10^{-1}$
Mod-ϑ_{opt}	$3.80 \cdot 10^{-3}$	$6.31 \cdot 10^{-1}$	$3.94 \cdot 10^{-3}$	$2.01 \cdot 10^{-1}$

[5] E.-C. Chang and S. Roy, "Robust extraction of secret bits from minutiae," in *Int. Conf. on Biometrics*, Seoul, South Korea, August 2007, pp. 750–759.

[6] 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*, 2004, pp. 532–540.

[7] A. Juels and M. Sudan, "A fuzzy vault scheme," *Designs, Codes and Cryptography*, vol. 38, no. 2, pp. 237–257, February 2006.

[8] K. Nandakumar, A. K. Jain, and S. Pankanti, "Fingerprint-based fuzzy vault: Implementation and performance," in *IEEE Transactions on Information Forensics and Security*, Decmber 2007, pp. 744–757.

[9] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Systems Journal*, vol. 40, no. 3, pp. 614–634, 2001.

[10] E. J. C. Kelkboom, G. G. Molina, T. A. M. Kevenaar, R. N. J. Veldhuis, and W. Jonker, "Binary biometrics: An analytic framework to estimate the bit error probability under gaussian assumption," in *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*, 2008, pp. 1–6.

[11] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, January 2000.

[12] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *IEEE CVPR*, vol. 2, June 2005, pp. 454–461.

[13] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "Fvc2000: fingerprint verification competition," *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, vol. 24, no. 3, pp. 402–412, 2002.

[14] Y.-J. Chang, W. Zhang, and T. Chen;, "Biometrics-based cryptographic key generation," in *IEEE Int. Conf. on Multim. and Expo*, vol. 3, June 2004, pp. 2203 – 2206.

[15] C. Chen, R. Veldhuis, T. Kevenaar, and A. Akkermans, "Multi-bits biometric string generation based on the likelihood ratio," in *IEEE Conf. on Biometrics: Theory, Applications and Systems*, Washington DC, September 2007.

[16] J. Breebaart, C. Busch, J. Grave, and E. Kindt, "A reference architecture for biometric template protection based on pseudo identities." in *BIOSIG*, Darmstadt, Germany, September 2008.

[17] N. Delvaux, P. Lindeberg, J. Midgren, J. Breebaart, T. Akkermans, M. van der Veen, R. Veldhuis, E. Kindt, K. Simoens, C. Busch, P. Bours, D. Gafurov, B. Yang, J. Stern, C. Rust, B. Cucinelli, and D. Skepastianos, "Pseudo idenitities based on fingerprint characteristics," in *IEEE 4th international conference on intelligent information hiding and multimedia signal processing (IIH-MSP)*, 2008.

[18] J. Daugman, "The importance of being random: statistical principles of iris recognition," *Pattern Recognition*, vol. 36, no. 2, pp. 279–291, 2003.

[19] A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev, *Integrals and Series, Vol. 2: Special Functions*. New York: Gordon and Breach, 1990.

[20] B. Gökberk, M. O. Irfanoglu, and L. Akarun, "3D shape-based face representation and feature extraction for face recognition," *Image and Vision Computing*, vol. 24, no. 8, pp. 857–869, August 2006.

[21] A. M. Bazen and R. N. J. Veldhuis, "Likelihood-ratio-based biometric verification," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, no. 1, pp. 86–94, 2004.

[22] P. Tuyls, A. H. M. Akkermans, T. A. M. Kevenaar, G.-J. Schrijnen, A. M. Bazen, and R. N. J. Veldhuis, "Pratical biometric authentication with template protection," in *5th International Conference, AVBPA*, Rye Brook, New York, July 2005.

[23] C. Chatfield, *Statistics for Technology: A course in Applied Statistics*, third edition ed. Chapman & Hall, 1983.

[24] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, 2nd ed. Addison Wiley, 1994.

[25] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, "A statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation," in *Proceedings of ICSLP-98*, 1998.

[26] J. Breebaart, A. H. M. Akkermans, and E. J. C. Kelkboom, "Inter-subject differences in false non-match rates for a fingerprint-based authentication system," in *EURASIP J. Advances in Signal Processing (Submitted)*, 2008.

[27] N. Yager and T. Dunstone, "Worms, chameleons, phantoms and doves: New additions to the biometric menagerie," in *AutoID 2007*, 2007.