

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

Emile J. C. Kelkboom, Gary Garcia Molina, Jeroen Breebaart, Raymond N. J. Veldhuis, Tom A. M. Kevenaar, and Willem 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 cancelable biometrics have been proposed for protecting biometric data. Most of these methods use cryptographic primitives or error-correcting codes (ECCs) and use a binary representation of the real-valued biometric data. Hence, the difference between two biometric samples is given by the Hamming distance (HD) or bit errors between the binary vectors obtained from the enrollment and verification phases, respectively. If the HD 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 tradeoff. A method to improve the FRR consists of 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 HD. In practice, the number of samples is empirically chosen without fully considering its fundamental impact. In this paper, 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 tradeoff curve that combines the false acceptance and false rejection rates is estimated to assess the system performance. The analytic expressions are validated using the Face Recognition Grand Challenge v2 and Fingerprint Verification Competition 2000 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

Manuscript received November 30, 2008. First published March 25, 2010; current version published April 14, 2010. This paper was recommended by Guest Editor K. W. Bowyer.

E. J. C. Kelkboom, G. Garcia Molina, and J. Breebaart are with the Philips Research Laboratories, 5656 Eindhoven, The Netherlands (e-mail: Emile.Kelkboom@philips.com; Gary.Garcia@philips.com; Jeroen.Breebaart@philips.com).

R. N. J. Veldhuis is with the University of Twente, 7500 AE Enschede, The Netherlands (e-mail: R.N.J.Veldhuis@utwente.nl).

T. A. M. Kevenaar is with the priv-ID, 5656 Eindhoven, The Netherlands (e-mail: Tom.Kevenaar@priv-id.com).

W. Jonker is with the Philips Research Laboratories, 5656 Eindhoven, The Netherlands, and also with the University of Twente, 7500 AE Enschede, The Netherlands (e-mail: Willem.Jonker@philips.com; jonker@cs.utwente.nl).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSMCA.2010.2041657

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]–[4], *fuzzy extractors* [5], [6], *fuzzy vault* [7], [8], and *cancelable 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 HD 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 an HD that is larger than the decision threshold.

Attackers may attempt to gain access by impersonating a genuine user. The associated comparisons are referred to as the impostor comparisons and will be accepted if the HD 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 impostor (ϕ_{im}) HD probability mass functions (pmfs) and the decision threshold T . A graphical representation is shown in Fig. 1.

One of the problems with template-protection systems based on ECCs is that the FRR is lower (LB) 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

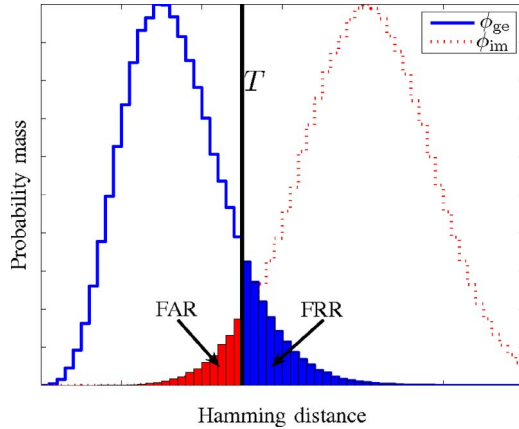


Fig. 1. FRR and FAR from the genuine and impostor HD pmfs, ϕ_{ge} , and (ϕ_{im}), respectively.

suppress the noise and thus reduce the number of bit errors resulting in a smaller HD.

The main objective of this paper is to analytically estimate, under the Gaussian assumption, the performance of a biometric system based on binary vectors under HD comparison and considering the use of multiple biometric samples. We present a framework for analytically estimating both the genuine and impostor HD pmfs from the analytically estimated bit-error probability presented in [10] under the assumption that both the within and between class of the real-valued features are Gaussian distributed. First, 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 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]. Second, the Gaussian assumption makes it possible to obtain an analytical closed-form expression for the HD 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 impostor HD 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 Face Recognition Grand Challenge (FRGC) v2 3-D face images [12] and the Fingerprint Verification Competition (FVC) 2000 fingerprint images [13]. We further extend the framework in Sections 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 3-D 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* $\mathbf{f}_R^e \in \mathbb{R}^{N_F}$ 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 $\mathbf{f}_B^e \in \{0, 1\}^{N_B}$ 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 [2], [3] 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}_R^e or \mathbf{f}_B^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 ECCs. The output is a pseudoidentity (PI), represented as a binary vector, accompanied by some auxiliary data that are 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}_R^v is extracted followed by the quantization process, which produces the binary vector \mathbf{f}_B^v . In the Bit-Protection module, a candidate pseudoidentity PI^* is created using AD and the binary vector \mathbf{f}_B^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}_R^e and \mathbf{f}_R^v . For a binary biometric system, the decision is based on the difference between \mathbf{f}_B^e and \mathbf{f}_B^v , which can be quantified using the HD. 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: 1) the Acquisition and Feature-Extraction modules, where the input is the subject's biometrics and the output is a real-valued feature vector $\mathbf{f}_R \in \mathbb{R}^{N_F}$; 2) the Bit-Extraction module that extracts a binary vector \mathbf{f}_B out of \mathbf{f}_R ; and 3) the Bit-Protection and Bit-Matching modules which protect the binary vector and perform 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 this paper.

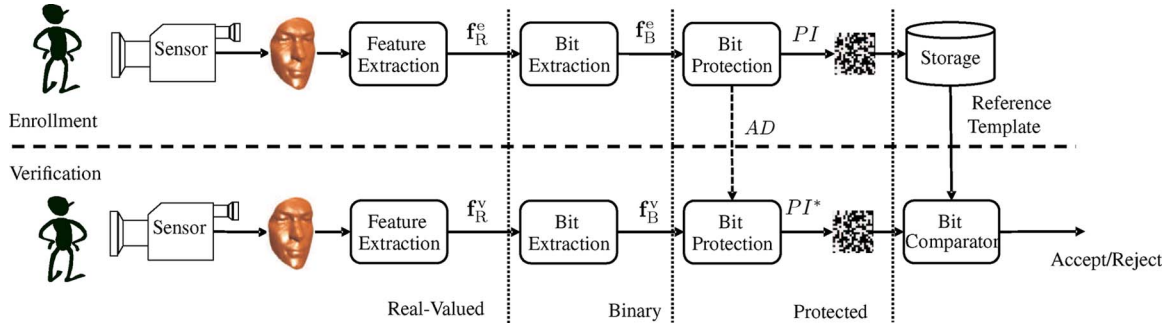


Fig. 2. General scheme of a biometric system with template protection based on helper data.

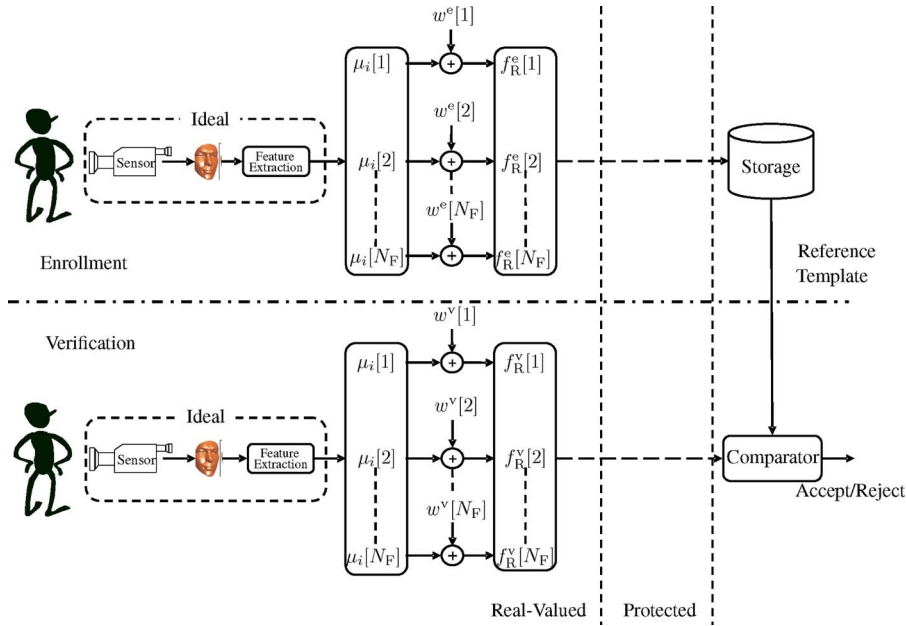


Fig. 3. PGC for both the enrollment and verification phase.

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 $\mathbf{f}_R = [f_R[1], f_R[2], \dots, f_R[N_F]]'$ of dimension N_F , where “ $'$ ” is the transpose operator. The feature vector \mathbf{f}_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 (PGCs) as shown in Fig. 3. We assume an ideal Acquisition and Feature-Extraction module which always produces the same feature vector μ_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 $p_{w[j],i} \sim \mathcal{N}(0, \sigma_{w,i}^2[j])$. Adding the noise $w[j]$ with the mean $\mu_i[j]$ results into the noisy feature component $f_R[j]$; in vector notation, $\mathbf{f}_R = \mu_i + \mathbf{w}$. 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 $\sigma_{w,i}^2[j] = \sigma_w^2[j] \forall i$. 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 $p_{\mu_b[j]} \sim \mathcal{N}(\mu_b[j], \sigma_b^2[j])$. 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 $f_R[j]$ across the whole population and is also Gaussian with $p_{t[j]} \sim \mathcal{N}(\mu_t[j], \sigma_t^2[j])$, where $\mu_t[j] = \mu_b[j]$ and $\sigma_t^2[j] = \sigma_w^2[j] + \sigma_b^2[j]$. For simplicity, but without loss of generality, we consider $\mu_t[j] = \mu_b[j] = 0$.

As shown in Fig. 3, in both the enrollment and verification phase, the PGC adds random noise \mathbf{w}^e and \mathbf{w}^v with the same probability density to μ_i , resulting in \mathbf{f}_R^e and \mathbf{f}_R^v , respectively. Thus, μ_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 a bit-extraction method, we use the

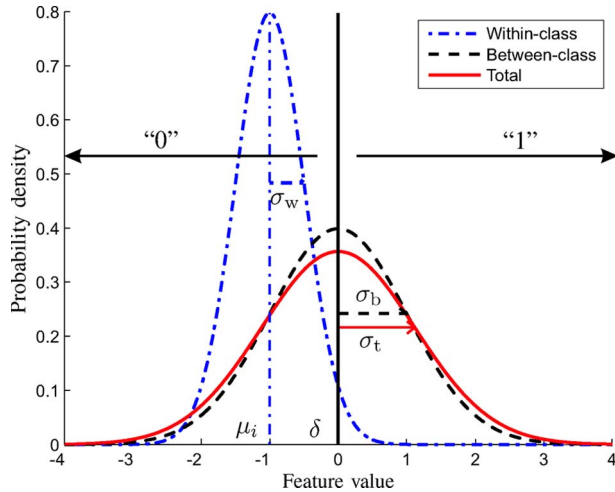


Fig. 4. Modeling of a single feature component of the real-valued biometric.

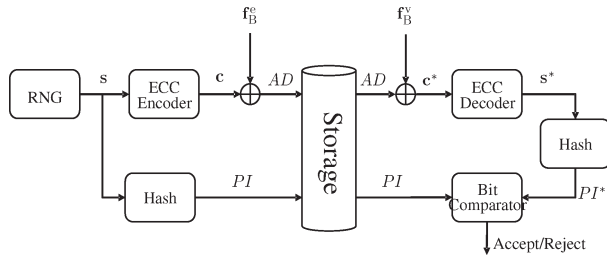


Fig. 5. Fuzzy commitment scheme.

thresholding version used in [2] and [3], where a single bit is extracted from each feature component. Hence, the obtained binary vector $\mathbf{f}_B \in \{0, 1\}^{N_F}$ has the same dimension as \mathbf{f}_R . Furthermore, the binarization threshold for each component $\delta[j]$ is set equal to the mean of the between-class pdf $\mu_b[j]$; if the value of $f_R[j]$ 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}_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 or key out of different biometric measurements of the same subject. However, the binary input vector \mathbf{f}_B itself cannot be used as the key because it is most likely not exactly the same in both the enrollment and verification phase ($\mathbf{f}_B^e \neq \mathbf{f}_B^v$) due to measurement noise and biometric variability that lead to *bit errors*. The number of bit errors is also referred to as the HD $d_H(\mathbf{f}_B^e, \mathbf{f}_B^v)$. Therefore, ECCs 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 s is randomly generated by the *Random-Number-Generator* module. The security level of the system is higher at larger

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_c/n_c

n_c	k_c	t_c	BER = t_c/n_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%

secret lengths. A codeword \mathbf{c} of an ECC is obtained by encoding s in the *ECC-Encoder* module. The codeword is XORed with \mathbf{f}_B^e in order to obtain the auxiliary data AD . Furthermore, the hash of s is taken in order to obtain the pseudoidentity PI . For the sake of coherence, we use the terminology proposed in [16] and [17].

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

In order to illustrate our framework with practical parameter values, we choose the linear block-type “Bose, Ray-Chaudhuri, Hocquenghem” (BCH) encoder/decoder as an example ECC. While more sophisticated ECCs can be used, the BCH accommodates our framework due to its HD classifier property. For example, if we would 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 modeled, which is out of the scope of the framework described in this paper. The ECC 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 $[n_c, k_c, t_c]$. 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}_B^e and \mathbf{f}_B^v is smaller or equal to the error-correcting capability t_c . Thus, there is a match when the HD is smaller than t_c , $d_H(\mathbf{f}_B^e, \mathbf{f}_B^v) = \|\mathbf{f}_B^e \oplus \mathbf{f}_B^v\|_1 \leq t_c$, and the Bit-Protection module can be modeled as an HD classifier with threshold t_c . Some $[n_c, k_c, t_c]$ settings of the BCH code are given in Table I. Note that the maximum number of bits that can be corrected lies between 20% and 25% of the binary vector.

D. Modeling Summary

The following is a summary of the modeling choices and assumptions that we have made.

- **Acquisition and Feature-Extraction Block \mathbf{f}_R**
 - Modeled as a PGC, where each feature component is defined by:
 - Within-class pdf $\sim \mathcal{N}(0, \sigma_w^2[j])$
 - Describes the genuine biometric variability and measurement noise;
 - Homogeneous variance across subjects $\sigma_{w,i}^2[j] = \sigma_w^2[j] \forall i$
 - Noise is independent across channels, measurements, and subjects

- Between-class pdf $\sim \mathcal{N}(0, \sigma_b^2[j])$
 - Characterizes the $\mu_i[j]$ variability across the population
 - Feature components are independent
- Total pdf $\sim \mathcal{N}(0, \sigma_t^2[j])$
 - Defines $f_R[j]$ across the population
- **Bit-Extraction Block f_B**
 - Single bit extraction method, with binarization threshold $\delta[j] = \mu_b[j]$
- **Bit-Protection and Bit-Comparator Block**
 - HD classifier with the ECC settings defining its decision boundary.

III. ANALYTICAL ESTIMATION OF BIT-ERROR PROBABILITIES, FRR, AND FAR

The goal of this paper 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 an HD classifier, the goal is to analytically estimate the expected genuine and impostor HD pmfs ϕ_{ge} and ϕ_{im} , respectively (see Fig. 1). With these pmfs, we can compute the FRR β and the FAR α , where β is the probability that a genuine subject is incorrectly rejected and α is the probability that an impostor is incorrectly accepted by the biometric system.

The HD 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 HD \bar{d}_H between \mathbf{f}_B^e and \mathbf{f}_B^v is

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

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

$$\phi(k) \stackrel{\text{def}}{=} \mathcal{P}\{d_H(\mathbf{f}_B^e, \mathbf{f}_B^v) = k\} = (P_1 * P_2 * \dots * P_{N_F})(k) \quad (2)$$

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

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

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

which are the probabilities of having zero or N_F errors, respectively. The FRR corresponding to an HD threshold T $\beta(T)$ is

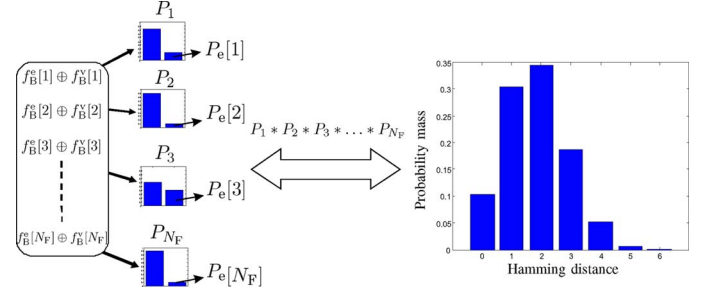


Fig. 6. Toy example of the convolution method given by (2).

the probability that the HD for a genuine comparison is greater than T , therefore

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

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

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

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 impostor 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 sample is taken prior to the binarization process.

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

For the impostor 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, impostor 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

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

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

$$= \binom{N_F}{k} 2^{-N_F} \quad (9)$$

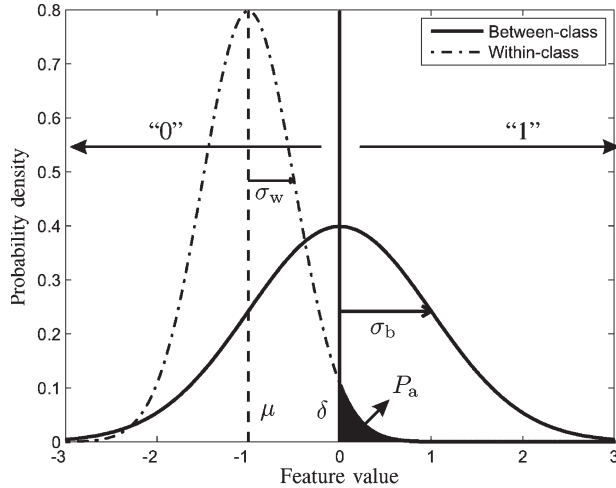


Fig. 7. Measurement error P_a .

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

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

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

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

where $P_e^{\text{ge}}(\mu)$ 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 . Consequently, (11) becomes

$$\begin{aligned} P_e^{\text{ge}} &= 2 \int_{-\infty}^0 p_b(\mu) P_e^{\text{ge}}(\mu) d\mu \\ &= 2 \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}\sigma_b} e^{-\left(\frac{\mu}{\sqrt{2}\sigma_b}\right)^2} P_e^{\text{ge}}(\mu) d\mu \\ &= \frac{2\lambda}{\sqrt{\pi}} \int_{-\infty}^0 e^{-(\lambda\mu)^2} P_e^{\text{ge}}(\mu) d\mu \end{aligned} \quad (12)$$

where $\lambda = 1/\sqrt{2}\sigma_b$.

We define the measurement or acquisition-error probability P_a , shown 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_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_e)

and verification (N_v) samples are considered, P_a also depends on the number of samples N , given as

$$P_a(\mu; N) = \int_0^{\infty} \frac{\sqrt{N}}{\sqrt{2\pi}\sigma_w} e^{-\left(\frac{\sqrt{N}(x-\mu)}{\sqrt{2}\sigma_w}\right)^2} dx \quad (13)$$

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

$$\sigma_{w,N}^2 = \frac{\sigma_w^2}{N} \Rightarrow \sigma_{w,N} = \frac{\sigma_w}{\sqrt{N}}. \quad (14)$$

With the use of the error function

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \quad (15)$$

and by defining $\eta = (\sqrt{N}/\sqrt{2}\sigma_w)$, $P_a(\mu; N)$ can be rewritten as

$$\begin{aligned} P_a(\mu; N) &= \frac{\eta}{\sqrt{\pi}} \int_0^{\infty} e^{-\eta(x-\mu)^2} dx \\ &= \frac{1}{\sqrt{\pi}} \int_{-\eta\mu}^{\infty} e^{-z^2} dz, \quad \text{with } z = \eta(x-\mu) \\ &= \frac{1}{\sqrt{\pi}} \left[\int_0^{\infty} e^{-z^2} dz - \int_0^{-\eta\mu} e^{-z^2} dz \right], \quad \text{for } \mu \leq 0 \\ &= \frac{1}{\sqrt{\pi}} \left[\frac{\sqrt{\pi}}{2} - \frac{\sqrt{\pi}}{2} \text{erf}(-\eta\mu) \right] \\ &= \frac{1}{2} [1 - \text{erf}(-\eta\mu)] \end{aligned} \quad (16)$$

where we used the well-known result $\int_0^{\infty} \lambda e^{-(\lambda\mu)^2} d\mu = \sqrt{\pi}/2$. 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, $P_e(\mu)$ of (12) becomes

$$\begin{aligned} P_e^{\text{ge}}(\mu; N_e, N_v) &= (1 - P_a(\mu; N_e)) P_a(\mu; N_v) \\ &\quad + P_a(\mu; N_e) (1 - P_a(\mu; N_v)) \\ &= \frac{1}{4} [(1 + \text{erf}(-\eta_e\mu)) (1 - \text{erf}(-\eta_v\mu)) \\ &\quad + (1 - \text{erf}(-\eta_e\mu)) (1 + \text{erf}(-\eta_v\mu))] \\ &= \frac{1}{2} [1 - \text{erf}(-\eta_e\mu)\text{erf}(-\eta_v\mu)] \end{aligned} \quad (17)$$

where $\eta_e = \sqrt{N_e}/\sqrt{2}\sigma_w$ and $\eta_v = \sqrt{N_v}/\sqrt{2}\sigma_w$. By substituting (17) into (12), we obtain

$$\begin{aligned} P_e^{\text{ge}}(N_e, N_v) &= \frac{\lambda}{\sqrt{\pi}} \int_{-\infty}^0 e^{-(\lambda\mu)^2} [1 - \text{erf}(-\eta_e\mu)\text{erf}(-\eta_v\mu)] d\mu \\ &= \frac{\lambda}{\sqrt{\pi}} \int_0^{\infty} e^{-(\lambda\mu)^2} [1 - \text{erf}(\eta_e\mu)\text{erf}(\eta_v\mu)] d\mu \\ &= \frac{1}{2} - \frac{\lambda}{\sqrt{\pi}} \int_0^{\infty} e^{-\lambda^2\mu^2} \text{erf}(\eta_e\mu)\text{erf}(\eta_v\mu) d\mu. \end{aligned} \quad (18)$$

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

$$\int_0^{\infty} e^{-\gamma x^2} \operatorname{erf}(ax) \operatorname{erf}(bx) dx = \frac{\arctan\left(\frac{ab}{\sqrt{\gamma(a^2+b^2+\gamma)}}\right)}{\sqrt{\gamma\pi}}. \quad (19)$$

Thus, (18) can be solved by using (19) with $\gamma = \lambda^2$, $a = \eta_e$, and $b = \eta_v$ as

$$\begin{aligned} P_e^{\text{ge}}(N_e, N_v, \sigma_w, \sigma_b) &= \frac{1}{2} - \frac{\lambda}{\sqrt{\pi}} \frac{\arctan\left(\frac{\eta_e \eta_v}{\sqrt{\lambda^2(\eta_e^2 + \eta_v^2 + \lambda^2)}}\right)}{\lambda \sqrt{\pi}} \\ &= \frac{1}{2} - \frac{1}{\pi} \arctan\left(\frac{\eta \sqrt{N_e N_v}}{\lambda \sqrt{N_e + N_v + \left(\frac{\lambda}{\eta}\right)^2}}\right) \\ &= \frac{1}{2} - \frac{1}{\pi} \arctan\left(\frac{\sigma_b \sqrt{N_e N_v}}{\sigma_w \sqrt{N_e + N_v + \left(\frac{\sigma_b}{\sigma_w}\right)^{-2}}}\right) \end{aligned} \quad (20)$$

where we also included σ_w and σ_b as an argument of the estimation function. As can be observed, P_e^{ge} is dependent on the σ_b/σ_w ratio, N_e , and N_v .

C. Summary

We have presented the analytic expressions of the genuine (ϕ_{ge}) and impostor (ϕ_{im}) HD pmfs and the corresponding FRR ($\beta(T)$) and FAR ($\alpha(T)$) curves. Because of the choice of the binarization scheme, the impostor bit-error probability $P_e^{\text{im}}[j]$ 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 $P_e^{\text{ge}}[j]$ has to be estimated using the analytic expression in (20). Therefore, in the remainder of this paper, we only need to estimate $P_e^{\text{ge}}[j]$, 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 means to generate uncorrelated features in Section IV-E. We conclude by portraying the experimental

TABLE II
OVERVIEW OF THE BIOMETRIC DATABASES

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

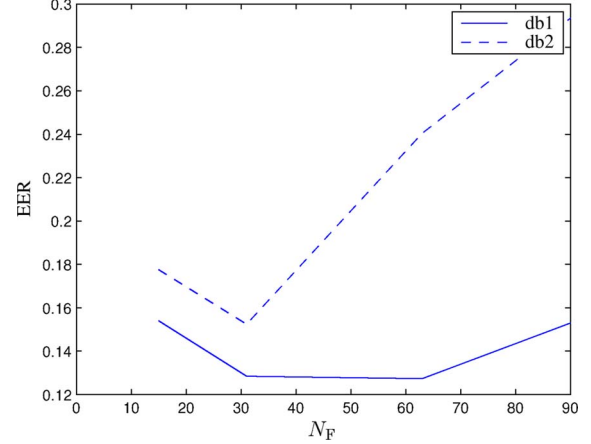


Fig. 8. EER of the training set after applying PCA for different reduced number of features N_F .

$\phi_{\text{ge}}(k)$, $\phi_{\text{im}}(k)$, $\beta(T)$, $\alpha(T)$, and detection error tradeoff (DET) curves in Section IV-F.

A. Biometric Databases and Feature Extraction

The first database (db1) consists of 3-D face images from the FRGC v2 data set [12], where we used the shape-based 3-D face recognizer of [20] to extract feature vectors of dimension $N_{\text{orig}} = 696$. Subjects with at least eight 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 $N_i = 14$ samples per subject. The second database (db2) consists of fingerprint images from the database 2 of FVC2000 [13] and uses a feature-extraction algorithm based on Gabor filters and directional fields [21], resulting in 1536 features ($N_{\text{orig}} = 1536$). 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 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. 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 an HD 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]

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	$\sum_{i=1}^{N_s} \sum_{j=1}^{N_i} (f_{i,j} - \hat{\mu})^2$	$N_t - 1$	

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 focused 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 j th 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 analysis of variance models and describes the 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 derived from this table are that: 1) the total sum of squares is equal to the sum of the within-class and between-class sum of squares $SST = SSW + SSB$ and 2) the total number of *degrees of freedom* (d.f.) is equal to the sum of the between-class and the within-class d.f. 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

$$\hat{\sigma}_w^2 = \frac{1}{N_t - N_s} \sum_{i=1}^{N_s} \sum_{j=1}^{N_i} (f_{i,j} - \hat{\mu}_i)^2 \quad (21)$$

$$\hat{\sigma}_b^2 = \frac{1}{N_i(N_s - 1)} \sum_{i=1}^{N_s} N_i (\hat{\mu}_i - \hat{\mu})^2, \quad \text{with } \bar{N}_i = \frac{N_t}{N_s} \quad (22)$$

$$\hat{\sigma}_t^2 = \frac{1}{N_t - 1} \sum_{i=1}^{N_s} \sum_{j=1}^{N_i} (f_{i,j} - \hat{\mu})^2 \quad (23)$$

with the exception of $\hat{\sigma}_b^2$, which is also divided by the average number of samples per subject \bar{N}_i . Notice that $\hat{\sigma}_w^2$ is calculated as the variance of the aggregated zero-mean samples of the subjects, while taking into account that N_s d.f. are lost because of the need to estimate the mean of each subject $\hat{\mu}_i$. Furthermore,

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

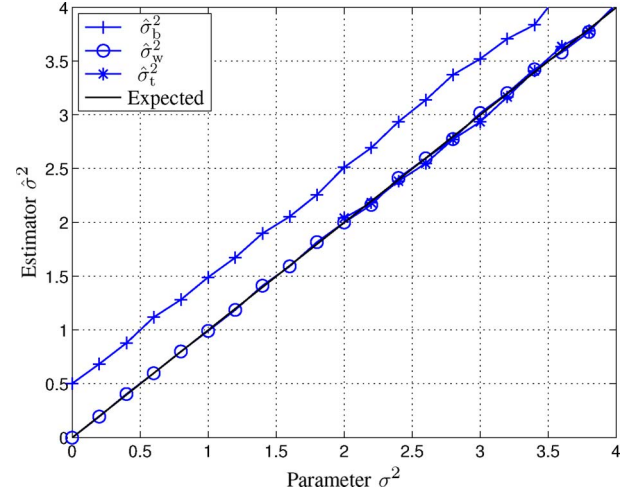


Fig. 9. Within-class, between-class, and total variance estimation for different settings of $\{\sigma_w^2, \sigma_b^2\}$.

$\hat{\sigma}_w^2$ is also equal to the weighted average of the variance of each subject because (21) can also be written as

$$\begin{aligned} \hat{\sigma}_w^2 &= \frac{1}{N_t - N_s} \sum_{i=1}^{N_s} (N_i - 1) \hat{\sigma}_{w,i}^2 \\ &= \frac{1}{N_s (\bar{N}_i - 1)} \sum_{i=1}^{N_s} (N_i - 1) \hat{\sigma}_{w,i}^2 \\ &= \frac{1}{\frac{1}{N_s} \sum_{i=1}^{N_s} (N_i - 1)} \frac{1}{N_s} \sum_{i=1}^{N_s} (N_i - 1) \hat{\sigma}_{w,i}^2, \quad \text{with} \\ \hat{\sigma}_{w,i}^2 &= \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (f_{i,j} - \hat{\mu}_i)^2 \end{aligned} \quad (24)$$

which turns into $\hat{\sigma}_w^2 = (1/N_s) \sum_{i=1}^{N_s} \hat{\sigma}_{w,i}^2$ when N_i is equal for each subject.

The variance estimators are validated using a synthetically generated database of $N_s = 1000$ subjects with $N_i = 4$ samples each. The parameters $\{\sigma_w^2, \sigma_b^2\}$ are used during the synthesis, and we estimated $\{\hat{\sigma}_w^2, \hat{\sigma}_b^2, \hat{\sigma}_t^2\}$ using (21), (22), and (23), respectively. The synthesis and estimation processes are performed ten times (tenfold), and the average of the result is taken. Fig. 9 shows the estimation results of $\hat{\sigma}_w^2$ for different values of σ_w^2 with $\sigma_b^2 = 2$, and both $\hat{\sigma}_b^2$ and $\hat{\sigma}_t^2$ for different values of σ_b^2 with $\sigma_w^2 = 2$. We can conclude that the $\hat{\sigma}_w^2$ and $\hat{\sigma}_t^2$ estimators give values that closely resemble the underlying model parameters σ_w^2 and σ_t^2 , but we observe a constant estimation error for the $\hat{\sigma}_b^2$ estimator. This estimation error is examined for different values of σ_w^2 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 $\hat{\sigma}_b^2$ 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 of the sample mean $\hat{\mu}_i$ is given by

$$\sigma_{\hat{\mu}_i}^2 = \frac{\sigma_{w,i}^2}{N_i}. \quad (25)$$

If more samples are taken to estimate the sample mean, the estimation variance decreases. This implies that the estimation

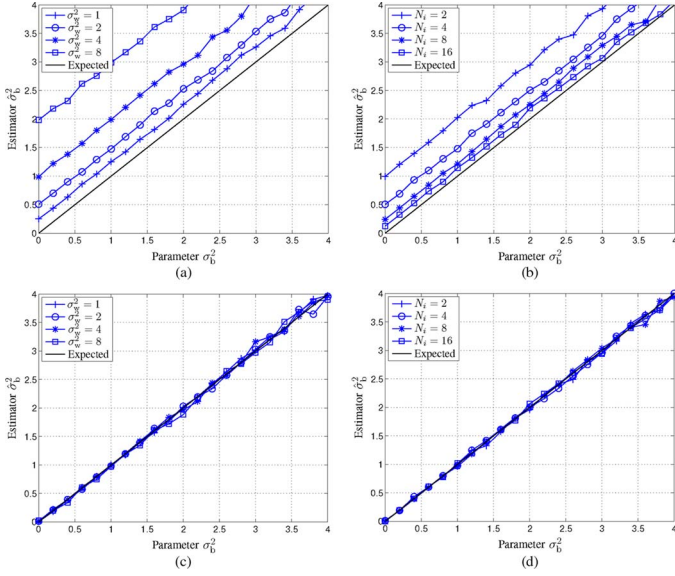


Fig. 10. Between-class estimation of (22) at (a) different values of σ_w^2 with $N_i = 2$ and (b) different values of N_i with $\sigma_w^2 = 2$, with its corrected version (27) in (c) and (d), respectively.

$\hat{\sigma}_b^2$ of (22) is in fact

$$\hat{\sigma}_b^2 = EST(\sigma_b^2 + \sigma_\mu^2) = EST\left(\sigma_b^2 + \frac{\sigma_w^2}{\sigma_i}\right) \quad (26)$$

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

$$\check{\sigma}_b^2 = \hat{\sigma}_b^2 - \frac{\hat{\sigma}_w^2}{N_i}. \quad (27)$$

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

C. Boundaries of Tolerated Estimation Errors

When estimating $P_e[j]$ 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 LB tolerance bounds for the estimation errors. Such an example is shown in Fig. 11 for a synthetic data set of similar size as db2 ($N_s = 110$ and $N_i = 8$) but with $N_F = 500$ and $\sigma_w^2[j] = 1$, with $\sigma_b^2[j]$ randomly drawn from the uniform distribution $U(0, 16)$ with minimum and maximum values of 0 and 16, respectively. Fig. 11 compares the estimated bit-error probability of the synthetic data set $\hat{P}_e^{\text{sy}}[j]$ with the corresponding analytically obtained $P_e^{\text{ge}}[j]$, which stands for $P_e^{\text{ge}}(N_e, N_v, \hat{\sigma}_w[j], \check{\sigma}_b[j])$ of (20), where $\hat{\sigma}_w[j]$ and $\check{\sigma}_b[j]$ are estimated using (21) and (27), respectively. $\hat{P}_e^{\text{sy}}[j]$ is reported by a circle (“o”) at its estimated $\hat{\sigma}_b[j]/\hat{\sigma}_w[j]$ ratio, and its analytic estimation is the value of the solid line at the same $\hat{\sigma}_b[j]/\hat{\sigma}_w[j]$ ratio. A greater vertical distance implies a greater analytical estimation error.

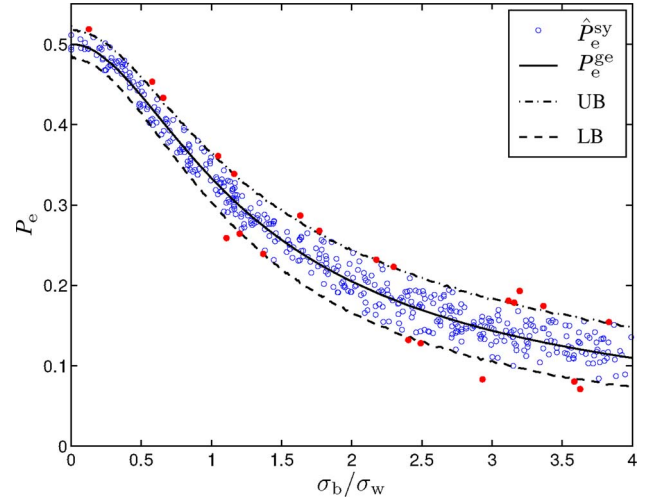


Fig. 11. Random estimation errors due to the random nature and the UB and LB boundaries.

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

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

D. Validation of the Analytic Expression P_e^{ge}

In this section, we experimentally validate the analytic expression of the bit-error probability P_e^{ge} . In the previous section, we have discussed the use of PCA for decorrelating the feature components and for reducing the dimension to $N_F = 31$. 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 $\hat{P}_e^{\text{db}}[j]$ with the corresponding analytically obtained $P_e^{\text{ge}}[j]$. The same test protocol is used as discussed in Section IV-C. The experimental results for db1 and db2 for different values of N_e and N_v are shown in Figs. 12 and 13, respectively. The circles (disks) correspond to cases where $\hat{P}_e^{\text{db}}[j]$ is within (outside) the 95 percentile boundaries. We refer to the number of disks as the estimation error ϵ_{P_e} . If all the assumptions hold, then we expect the relative ϵ_{P_e} to be around 5%. Table IV reports the absolute and

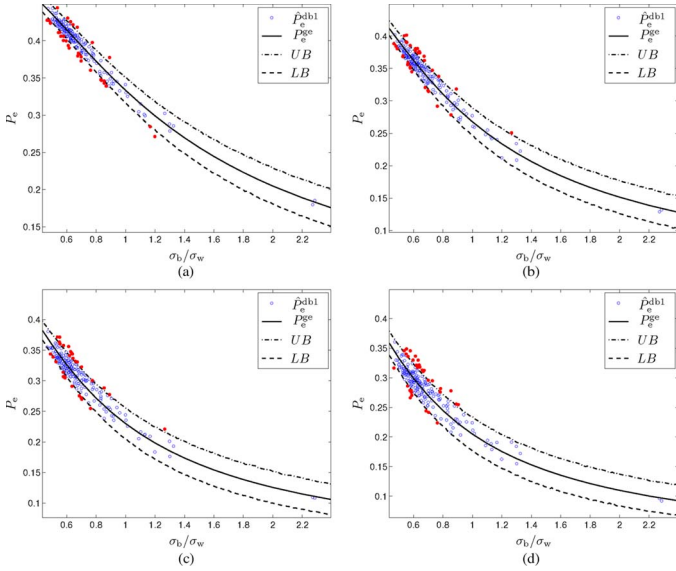


Fig. 12. Comparison between $P_e^{ge}[j]$ and $\hat{P}_e^{db1}[j]$ for different settings. (a) $N_e = N_v = 1$. (b) $N_e = N_v = 2$. (c) $N_e = N_v = 3$. (d) $N_e = N_v = 4$. The circles (disks) correspond to cases where $\hat{P}_e^{db1}[j]$ falls within (outside) the boundaries.

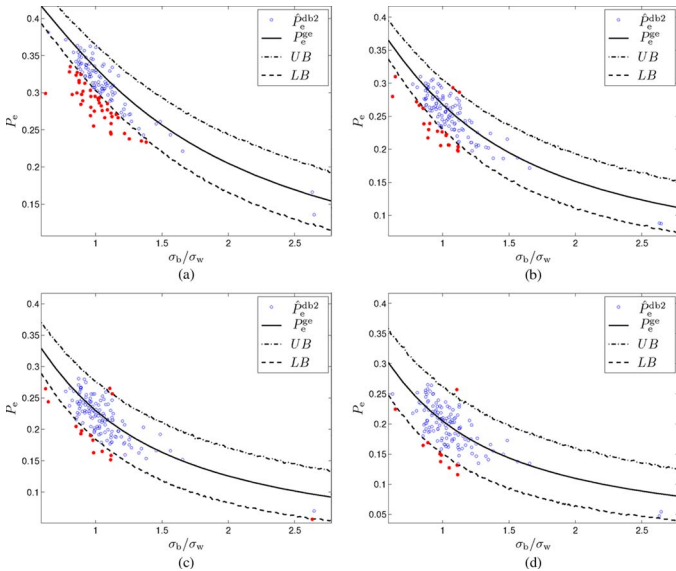


Fig. 13. Comparison between $P_e^{ge}[j]$ and $\hat{P}_e^{db2}[j]$ for different settings. (a) $N_e = N_v = 1$. (b) $N_e = N_v = 2$. (c) $N_e = N_v = 3$. (d) $N_e = N_v = 4$. The circles (disks) correspond to cases where $\hat{P}_e^{db2}[j]$ falls within (outside) the boundaries.

relative ϵ_{P_e} . Because ϵ_{P_e} is noisy due to the random selection of N_e and N_v samples within the test protocol, we repeat the estimation 20 times and report its mean. For db1, ϵ_{P_e} is 16.7% for $N_e = N_v = 1$ and decreases to 13% for $N_e = N_v = 4$. In the case of db2, ϵ_{P_e} is very large; 27.3% for $N_e = N_v = 1$ but decreases significantly when both N_e and N_v are increased, reaching 6.3% when $N_e = N_v = 4$. 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]. In addition, note that many $\hat{P}_e^{db1}[j]$ estimations of db1 are very close to the 95 percentile boundaries, hence, small estimation errors

TABLE IV
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

Setting	db1		db2	
	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%

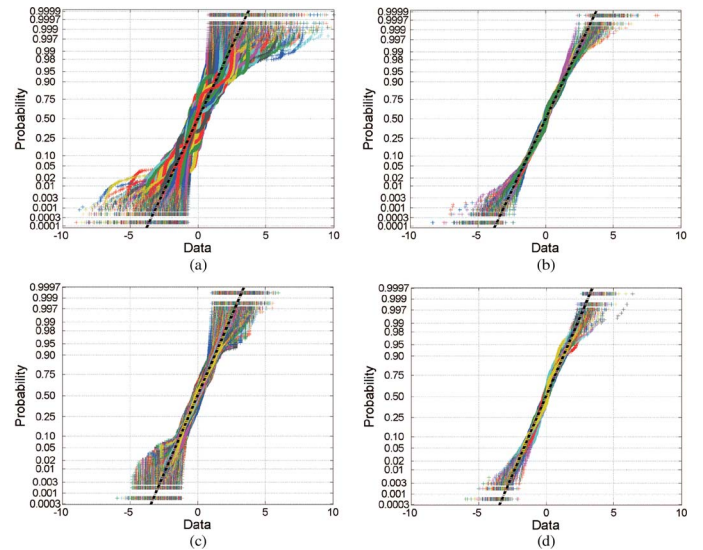


Fig. 14. Normal probability plot of each feature-vector component of db1 and db2 before and after applying PCA. (a) db1 before PCA. (b) db1 after PCA. (c) db2 before PCA. (d) db2 after PCA.

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. Effect of PCA on the Gaussian Assumption

As described in Section II, the analytic framework is based on the Gaussian model assumption. Fig. 14(a) and (c) shows 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 data set 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. Fig. 14(b) and (d) shows the normal probability plot for each of the 696 components of db1 and 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

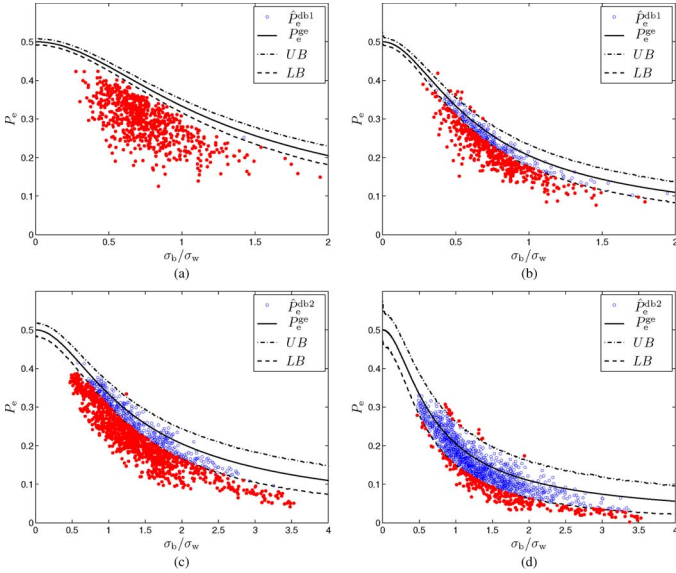


Fig. 15. $\hat{P}_e^{\text{dbx}}[j]$ at different settings of N_e and N_v for both db1 and db2 before applying the PCA transform. (a) db1 with $N_e = N_v = 1$. (b) db1 with $N_e = N_v = 4$. (c) db2 with $N_e = N_v = 1$. (d) db2 with $N_e = N_v = 4$.

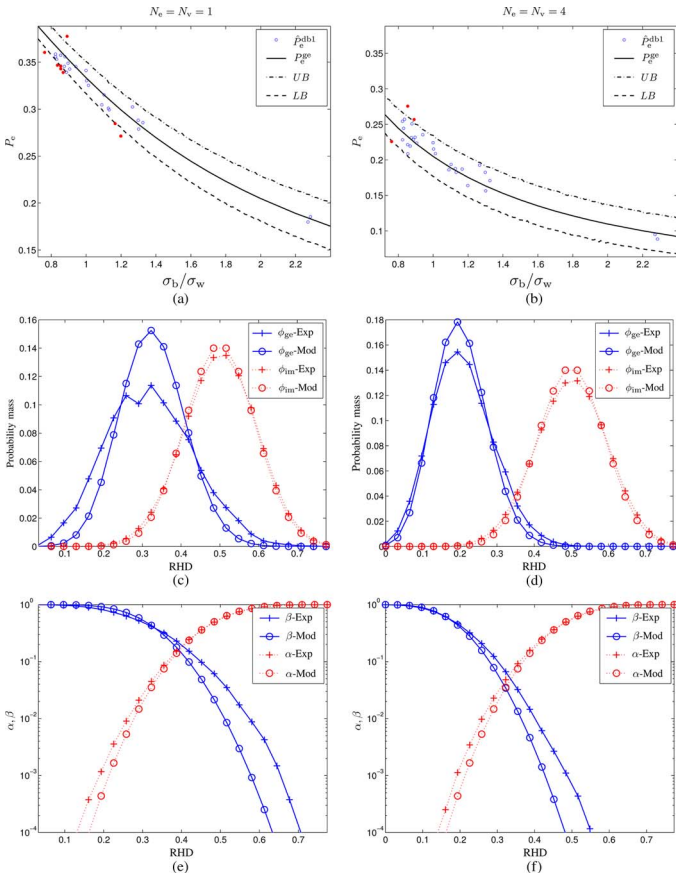


Fig. 16. Results for db1 with $N_F = 31$. (a) and (b) \hat{P}_e^{db1} and the analytical estimation of P_e^{ge} . (c) and (d) $\phi_{\text{ge}}(k)$ and $\phi_{\text{im}}(k)$ pmfs. (e) and (f) the $\alpha(T)$ and $\beta(T)$ curves. The graphs on the left (right) correspond to $N_e = N_v = 1$ ($N_e = N_v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

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

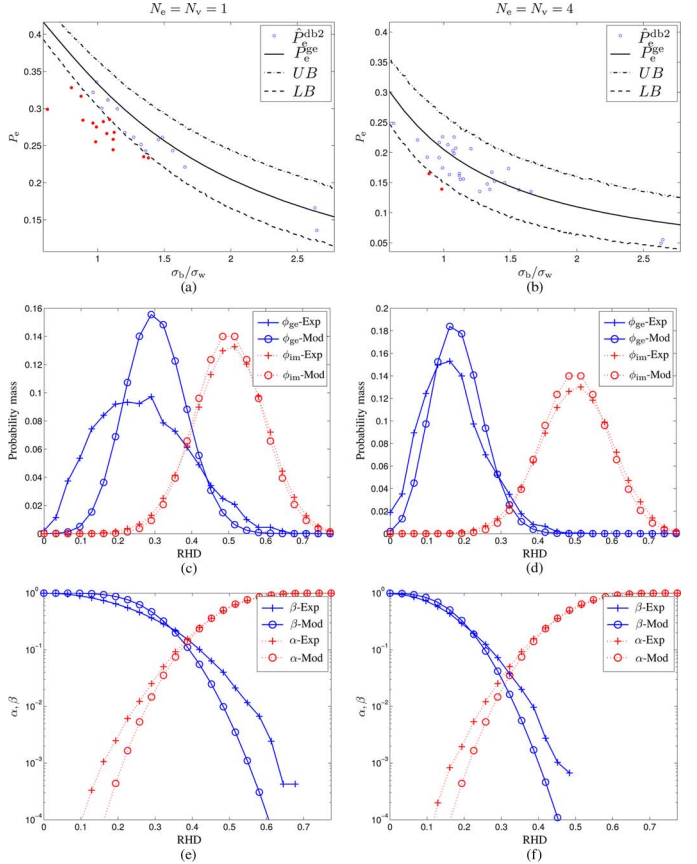


Fig. 17. Results for db2 with $N_F = 31$. (a) and (b) \hat{P}_e^{db2} and the analytical estimation of P_e^{ge} . (c) and (d) $\phi_{\text{ge}}(k)$ and $\phi_{\text{im}}(k)$ pmfs. (e) and (f) the $\alpha(T)$ and $\beta(T)$ curves. The graphs on the left (right) correspond to $N_e = N_v = 1$ ($N_e = N_v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

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 $\phi_{\text{ge}}(k)$ and impostor $\phi_{\text{im}}(k)$ HD pmfs, and the $\beta(T)$ and $\alpha(T)$ curves. The results are shown in Figs. 16 and 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 HD 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 Figs. 16 and 17 indicate that there is a good agreement between $\phi_{\text{im}}(k)$ -Exp and $\phi_{\text{im}}(k)$ -Mod. Large differences are observed between $\phi_{\text{ge}}(k)$ -Exp and $\phi_{\text{ge}}(k)$ -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 the previous section. It is interesting to note the differences between the estimation errors of $\phi_{\text{ge}}(k)$ of db1 and db2. For db1, the centers of gravity of $\phi_{\text{ge}}(k)$ -Exp and $\phi_{\text{ge}}(k)$ -Mod practically coincide. The only difference is the width of the pmfs since the

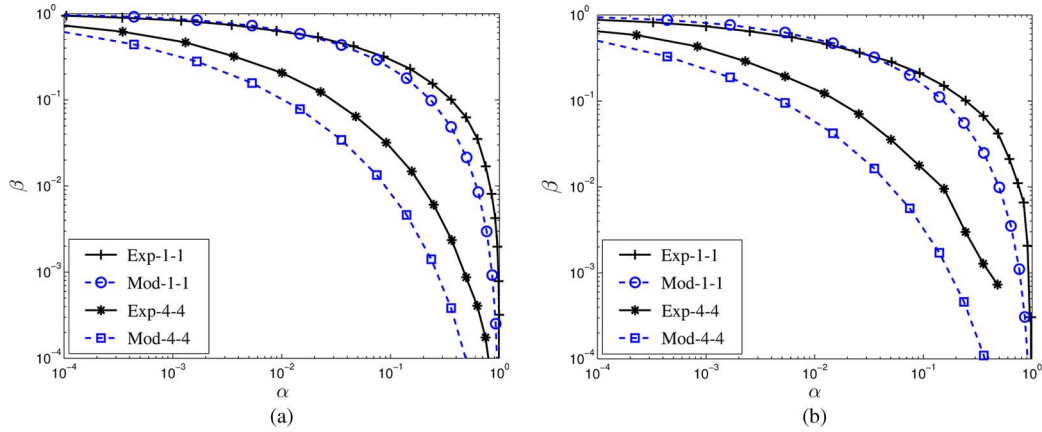


Fig. 18. DET curves for both db1 and db2 for $N_F = 31$ with different values of N_e , and N_v . The values N_e and N_v are indicated in the legend in the subsequent order. The experimentally obtained curves are denoted by Exp, while the analytical by Mod. (a) db1 with $N_F = 31$. (b) db2 with $N_F =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

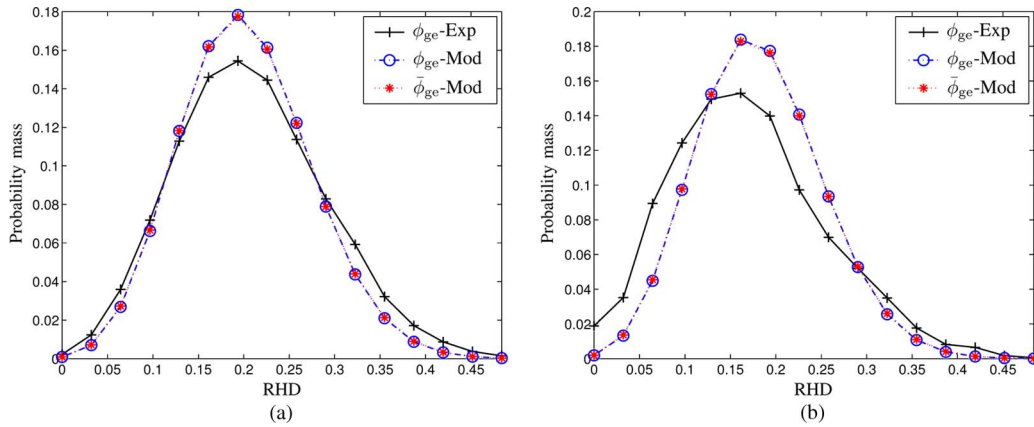


Fig. 19. Approximation of the genuine HD pmf as binomial with \bar{P}_e [(26)] for the $N_e = N_v = 4$ case with $N_F = 31$. (a) db1. (b) db2.

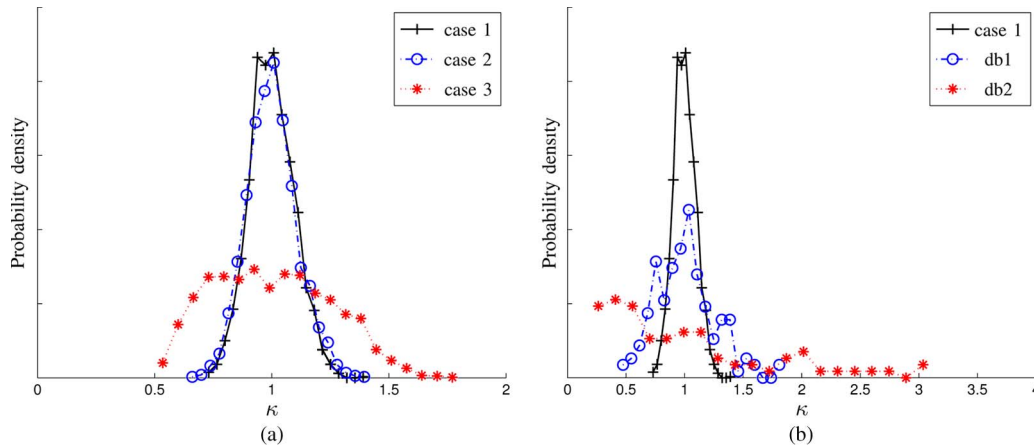


Fig. 20. Empirical estimated probability density p_{κ_i} using synthetic databases (a) of 2000 subjects with $N_F = 31$, $N_i = 8$, $\sigma_b^2[j] = 1$, where for case 1, every subject has the same $\sigma_{w,i}^2[j] = 1$; in case 2, $\sigma_{w,i}^2[j] = 1 + \nu_i[j]$; and for case 3, $\sigma_{w,i}^2[j] = 1 + \nu_i$, 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.

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 this figure, 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 Figs. 16 and 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 as follows: 1) the nonhomogeneous within-class variance; 2) the dependence between features; and 3) the dependence between bit errors. The 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 HDs. In the literature, these subjects are referred to as goats [25], [26]. If the features are dependent, then the HD 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 HD pmf. A right tail is slightly visible for db1 but is clearly present for db2, as shown in Fig. 16(c) and (d) and Fig. 17(c) and (d), respectively.

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

V. RELAXING THE HOMOGENOUS WITHIN-CLASS VARIANCE ASSUMPTION

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

$$\bar{\phi}_{ge}(k) = \binom{N_F}{k} (\bar{P}_e^{ge})^k (1 - \bar{P}_e^{ge})^{N_F - k} \quad (28)$$

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 pmfs $\bar{\phi}_{ge}(k)$ are shown 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 nonhomogeneous 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 HD pmf defined as

$$\bar{\Phi}_{ge}(k) = \int_0^{1/2} p_{\bar{P}_e^{ge}}(\tau) \bar{\phi}_{ge}(k|\tau) d\tau \quad (29)$$

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 (28) as

$$\bar{\phi}_{ge}(k|\tau) = \binom{N_F}{k} (\tau)^k (1 - \tau)^{N_F - k}. \quad (30)$$

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

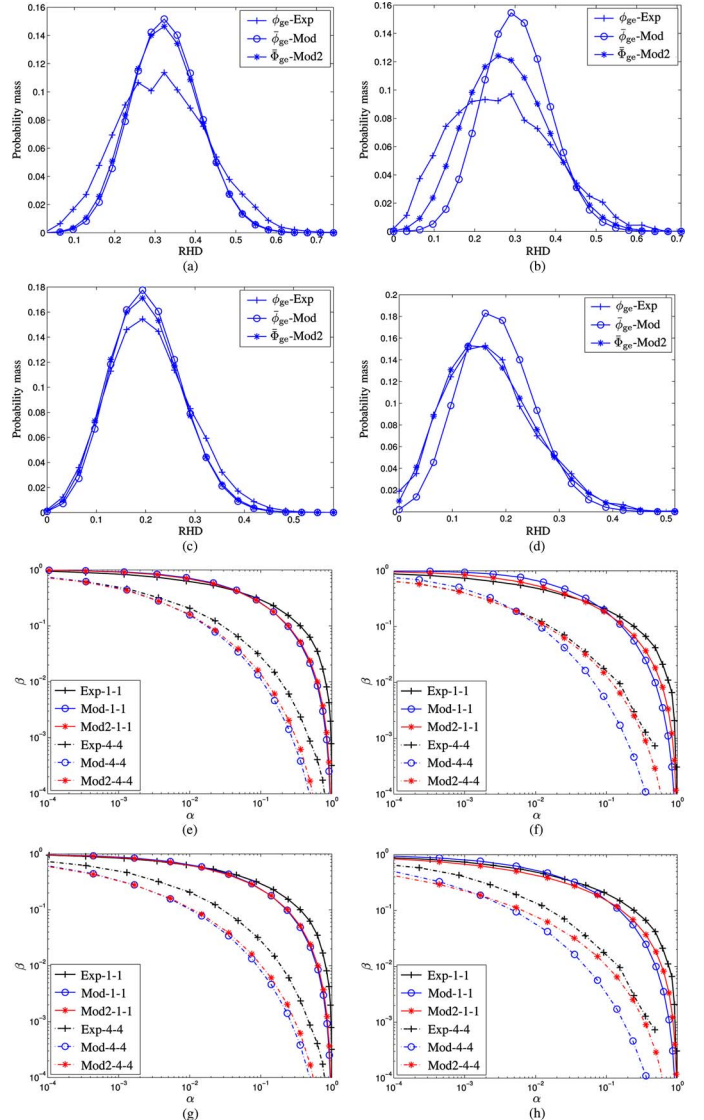


Fig. 21. Results of the proposed method incorporating the nonhomogeneous property of db1 and db2 for the cases $N_e = N_v = 1$ and $N_e = N_v = 4$ with $N_F = 31$. (a)–(d) show the HD pmf estimations, while (e)–(h) show the DET curves estimation, where Mod and Mod2 indicate the modeling method without and with the nonhomogeneous 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.

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 $N_i - 1$ d.f., 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 $\hat{\sigma}_{w,i}^2[j]$. It is well known that the mean associated with a χ^2 distribution is equal to its number of d.f.; thus, by omitting the $(N_i - 1)$ multiplications, it becomes a unit mean.

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

$$\kappa_i = \frac{1}{N_F} \sum_{j=1}^{N_F} \hat{\sigma}_{w,i}^2[j]/\sigma_w^2[j]. \quad (31)$$

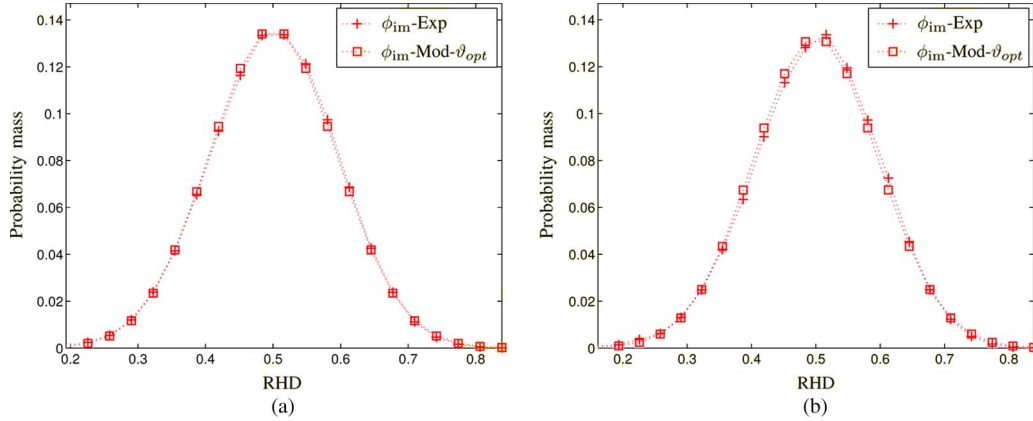


Fig. 22. Results of estimating ϑ_{opt} from (ϕ_{im}) -Exp using (33) for the $N_e = N_v = 1$ case for both databases. The variance-corrected Gaussian approximated curve as described by (32) is shown as (ϕ_{im}) -Mod- ϑ . (a) db1 $\vartheta_{opt} = 1.11$. (b) db2 $\vartheta_{opt} = 1.17$.

We can model the nonhomogeneous 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 “goatlike” 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 synthetically generated databases containing 2000 subjects with $N_F = 31$, $N_i = 8$, and $\sigma_b^2[j] = 1$, where for “case 1,” every subject has the same $\sigma_{w,i}^2[j] = 1$; in “case 2,” $\sigma_{w,i}^2[j] = 1 + \nu_i[j]$; and for “case 3,” $\sigma_{w,i}^2[j] = 1 + \nu_i$, 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 $\sigma_{w,i}^2[j]$ 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 $\sigma_{w,i}^2[j]$ 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 HD pmfs of db2, as shown in Fig. 17.

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

$$\bar{P}_{e,i}^{ge} = \frac{1}{N_F} \sum_{j=1}^{N_F} P_e^{ge} \left(N_e, N_v, \sqrt{\kappa_i \hat{\sigma}_w^2[j]}, \hat{\sigma}_b[j] \right) \quad (32)$$

where we take the average of $P_e^{ge}[j]$ across all features, while using $\hat{\sigma}_b[j]$ and the modified within-class variance estimation $\sqrt{\kappa_i \hat{\sigma}_w^2[j]}$. Because of the nonlinear relationship between

$P_e^{ge}[j]$ and $\hat{\sigma}_w[j]$, we take the average over $P_e^{ge}[j]$ instead of estimating P_e^{ge} , using the average of $\hat{\sigma}_w[j]$.

In practice, we can rewrite (29) as

$$\bar{\Phi}_{ge}(k) = \frac{1}{N_s} \sum_{i=1}^{N_s} \bar{\phi}_{ge} \left(k | \bar{P}_{e,i}^{ge} \right). \quad (33)$$

We applied this new method for estimating $\phi_{ge}(k)$ of db1 and db2, and the results are shown in Fig. 21(a)–(d) for the $N_e = N_v = 1$ and $N_e = N_v = 4$ cases with $N_F = 31$, where $\phi_{ge}(k)$ -Exp is the experimentally obtained pmf, $\phi_{ge}(k)$ -Mod is obtained using (2), and $\bar{\Phi}_{ge}(k)$ -Mod2 with (31). The results show that ϕ_{ge} -Exp is better approximated when using the new method $\bar{\Phi}_{ge}(k)$ -Mod2. In the case of db1, there is a small improvement, but for db2, there is a significant improvement, and even a better estimation is obtained when $N_e = N_v = 4$. Furthermore, Fig. 21(e)–(h) shows the DET curve results. In Fig. 21(e) and (f), the same α is used for each DET curve in order to isolate the estimation errors of $\phi_{ge}(k)$, while in Fig. 21(g) and (h), α -Exp is used for the Exp curves and α -Mod is used for both the Mod and Mod2 curves. With the new method, the DET curve estimation has improved, most significantly for db2. However, the differences between Fig. 21(e) and (f) and Fig. 21(g) and (h) clearly indicate that the remaining estimation errors are caused by the estimation of α . As shown in Fig. 16(c) and (d) and Fig. 17(c) and (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 DEPENDENCES

In the 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 dependences between feature components. We propose to estimate the dependence from the (ϕ_{im}) pmf and apply it to the ϕ_{ge} pmf estimation. Hence, we assume that both pmfs are influenced by the dependence to the same extent.

We estimate the dependence 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 $\mathcal{N}(N_F P_e, N_F P_e(1 - P_e))$, with mean $N_F P_e$ and variance $N_F P_e(1 - P_e)$. For the impostor case, we know that $P_e = 1/2$, from which its mean and variance become $N_F/2$ and $N_F/4$, respectively. Hence, the Gaussian approximation of the (ϕ_{im}) -Exp pmf with the variance parameter ϑ used for fitting becomes

$$\begin{aligned} \phi_{\text{im}}(k)\text{-Mod-}\vartheta &= \frac{1}{\sqrt{2\pi\vartheta\sigma^2}} e^{-\frac{(k-\mu)^2}{2\vartheta\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi\vartheta N_F P_e(1 - P_e)}} e^{-\frac{(k - N_F P_e)^2}{2\vartheta N_F P_e(1 - P_e)}} \\ &= \frac{2}{\sqrt{2\pi\vartheta N_F}} e^{-\frac{(2k - N_F)^2}{2\vartheta N_F}} \end{aligned} \quad (34)$$

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

$$\vartheta_{\text{opt}} = \arg \min_{\vartheta} \sum_{k=0}^{N_F} (\phi_{\text{im}}(k)\text{-Exp} - \phi_{\text{im}}(k)\text{-Mod-}\vartheta)^2. \quad (35)$$

The estimation results of ϑ_{opt} for the $N_e = N_v = 1$ 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 dependences between the feature components is relatively similar for both databases. Furthermore, the (ϕ_{im}) -Exp pmf is better estimated when compared with its first estimation disregarding the feature-component dependences, as shown 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 (33) as

$$\phi_{\text{ge}}(k) = \frac{1}{N_s} \sum_{i=1}^{N_s} \frac{1}{\sqrt{2\pi\sigma_{\text{cor}}^2}} e^{-\frac{(k - \bar{P}_{e,i}^{\text{ge}} N_F)^2}{2\sigma_{\text{cor}}^2}} \quad (36)$$

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

$$\phi'_{\text{ge}}(k) = \frac{1}{\sum_{k=0}^{N_F} \phi_{\text{ge}}(k)} \phi_{\text{ge}}(k). \quad (37)$$

The estimation results using (37) for the cases of $\vartheta = 1$ and $\vartheta = \vartheta_{\text{opt}}$ are shown in Fig. 23. For the $\vartheta = 1$ case, the Gaussian approximation is used without the variance correction. Fig. 23(a)–(d) shows that the $\phi_{\text{ge}}(k)$ pmf estimation has slightly improved. The $\bar{\Phi}'_{\text{ge}}\text{-Mod-}\vartheta_{\text{opt}}$ curve is closer to $\phi_{\text{ge}}(k)\text{-Exp}$ than $\bar{\Phi}'_{\text{ge}}\text{-Mod-}\vartheta_1$. This holds across the whole curve for the $N_e = N_v = 1$ case and mainly for the right tail for the $N_e = N_v = 4$ case. The same conclusions are also shown by the DET curves of Fig. 23(e)–(f), where each DET curve uses the same α curve, namely, the experimentally obtained $\alpha\text{-Exp}$, in order to isolate the $\phi_{\text{ge}}(k)$ pmf estimation errors. The DET curves in Fig. 23(g)–(h) use the actual α curves, thus $\alpha\text{-Mod-}\vartheta_1$ for the DET-Mod- ϑ_1 curves and $\alpha\text{-Mod-}\vartheta_{\text{opt}}$ for the DET-Mod- ϑ_{opt} curves, respectively. The curves show that the DET-Mod- ϑ_{opt} curve

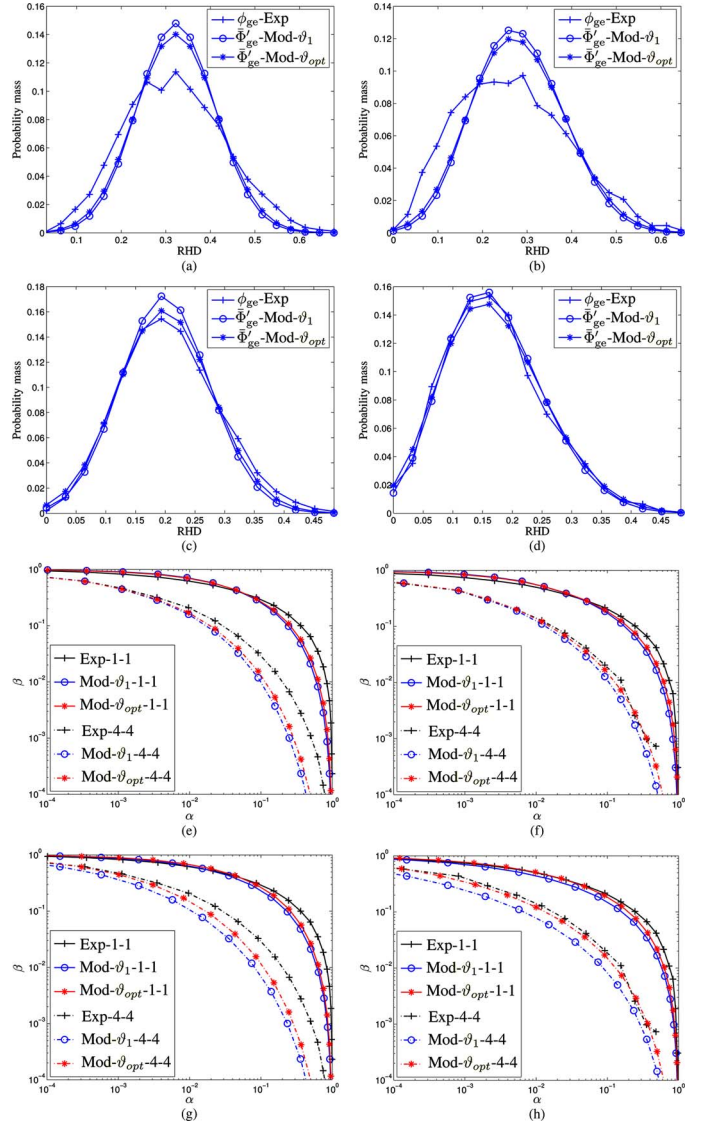


Fig. 23. Results of the proposed method incorporating both the dependence and nonhomogeneous property of db1 and db2 for the cases $N_e = N_v = 1$ and $N_e = N_v = 4$, with $N_F = 31$. (a)–(d) shows the ϕ_{ge} estimations, while (e)–(h) shows the DET curve estimation. The label Mod- ϑ_1 indicates the new modeling method but with $\vartheta = 1$, hence using only the Gaussian approximation of the binomial pmf, including the nonhomogeneous property. The label Mod- ϑ_{opt} indicates the cases where $\vartheta = \vartheta_{\text{opt}}$. In (e) and (f), all the DET curves are plotted using the experimentally obtained $\alpha\text{-Exp}$, while in (g) and (h), we use the $\alpha\text{-Exp}$ for the Exp curves, $\alpha\text{-Mod-}\vartheta_1$ for the Mod- ϑ_1 curves, and $\alpha\text{-Mod-}\vartheta_{\text{opt}}$ for the Mod- ϑ_{opt} curves.

is clearly closer to DET-Exp curve because $\alpha\text{-Mod-}\vartheta_{\text{opt}}$ is a better approximation of $\alpha\text{-Exp}$ as we have shown earlier.

VII. PRACTICAL CONSIDERATIONS

In the 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 LB bounded because of the limited number of bits the ECC can correct. For the setting of $N_F = 31$, which is equal to 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

TABLE V
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 DEPENDENCES (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}	$3.80 \cdot 10^{-3}$	$6.31 \cdot 10^{-1}$	$3.94 \cdot 10^{-3}$	$2.01 \cdot 10^{-1}$

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 $N_e + N_v \leq 5$. 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 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 LB 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 with 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. CONCLUSION

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 PGC 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 nonhomogeneous 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 dependence between the feature components and corresponding bits. The final proposed framework also incorporated feature-component dependence, whose value was derived from the calculated impostor HD pmf of the database. This method resulted in the most optimum estimation of the DET performance curves.

ACKNOWLEDGMENT

The authors would like to thank B. Gökberk and A. Bazen for facilitating the use of their feature vectors on which this paper is based on.

REFERENCES

- [1] A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *Proc. 6th ACM Conf. Comput. Commun. Security*, Nov. 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 *Proc. Int. Conf. Biometrics*, Seoul, Korea, Aug. 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 *Proc. 4th IEEE Workshop AutoID*, Buffalo, NY, Oct. 2005, pp. 21–26.
- [4] J.-P. Linnartz and P. Tuyls, "New shielding functions to enhance privacy and prevent misuse of biometric templates," in *Proc. 4th Int. Conf. AVBPA*, 2003, pp. 393–402.
- [5] E.-C. Chang and S. Roy, "Robust extraction of secret bits from minutiae," in *Proc. Int. Conf. Biometrics*, Seoul, Korea, Aug. 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 *Proc. Adv. Cryptology—Eurocrypt*, 2004, Lecture Notes in Computer Science, vol. 3027, pp. 532–540.
- [7] A. Juels and M. Sudan, "A fuzzy vault scheme," *Designs, Codes Cryptography*, vol. 38, no. 2, pp. 237–257, Feb. 2006.
- [8] K. Nandakumar, A. K. Jain, and S. Pankanti, "Fingerprint-based fuzzy vault: Implementation and performance," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 4, pp. 744–757, Dec. 2007.
- [9] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Syst. J.*, vol. 40, no. 3, pp. 614–634, Mar. 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 *Proc. 2nd IEEE BTAS*, 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, Jan. 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 *Proc. IEEE CVPR*, Jun. 2005, vol. 2, pp. 454–461.
- [13] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "Fvc2000: Fingerprint verification competition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 402–412, Mar. 2002.
- [14] Y. J. Chang, W. Zhang, and T. Chen, "Biometrics-based cryptographic key generation," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jun. 2004, vol. 3, pp. 2203–2206.
- [15] C. Chen, R. Veldhuis, T. Kevenaar, and A. Akkermans, "Multi-bits biometric string generation based on the likelihood ratio," in *Proc. IEEE Conf. Biometrics: Theory, Appl. Syst.*, Washington, DC, Sep. 2007, p. 1–6.
- [16] J. Breebaart, C. Busch, J. Grave, and E. Kindt, "A reference architecture for biometric template protection based on pseudo identities," in *Proc. BIOSIG*, Darmstadt, Germany, Sep. 2008, pp. 25–38.
- [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 identities based on fingerprint characteristics," in *Proc. IEEE 4th IHH-MSP*, 2008, pp. 1063–1068.
- [18] J. Daugman, "The importance of being random: Statistical principles of iris recognition," *Pattern Recognit.*, vol. 36, no. 2, pp. 279–291, Feb. 2003.
- [19] A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev, *Integrals and Series*. 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 Vis. Comput.*, vol. 24, no. 8, pp. 857–869, Aug. 2006.
- [21] A. M. Bazen and R. N. J. Veldhuis, "Likelihood-ratio-based biometric verification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 86–94, Jan. 2004.
- [22] P. Tuyls, A. H. M. Akkermans, T. A. M. Kevenaar, G.-J. Schrijen, A. M. Bazen, and R. N. J. Veldhuis, "Practical biometric authentication with template protection," in *Proc. 5th Int. Conf., AVBPA*, Rye Brook, NY, Jul. 2005, pp. 436–446.

- [23] C. Chatfield, *Statistics for Technology: A Course in Applied Statistics*, 3rd ed. London, U.K.: Chapman & Hall, 1983.
- [24] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, 2nd ed. Reading, MA: Addison-Wesley, 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 *Proc. ICSLP*, 1998, pp. 1351–1354.
- [26] J. Breebaart, T. Akkermans, and E. Kelkboom, "Intersubject differences in false nonmatch rates for a fingerprint-based authentication system," *EURASIP J. Adv. Signal Process.*, vol. 2009, Article ID 896383, 9 p., 2009. DOI:10.1155/2009/896383.
- [27] N. Yager and T. Dunstone, "Worms, chameleons, phantoms and doves: New additions to the biometric menagerie," in *Proc. AutoID*, 2007, pp. 1–6.



Emile J. C. Kelkboom was born in Oranjestad, Aruba, in 1980. He received the M.Sc. degree in electrical engineering from the Delft University of Technology, Delft, the Netherlands, in June 2004. Since August 2006, he has been working toward the Ph.D. degree at the Philips Research Laboratories, Eindhoven, The Netherlands, and in the Department of Electrical Engineering, Mathematics, and Computer Science, University of Twente, Enschede, The Netherlands.

From October 2004 to July 2006, he was an Application Engineer on CD, DVD, and Blu-ray drives with the Storage Engines Department of Philips Semiconductors. His focus was on safeguarding the privacy of the biometric information of users within biometric systems, namely, the field of template protection. His research interests include biometrics, pattern recognition, signal processing, and security.



Gary Garcia Molina received the M.Sc. degree in electrical engineering in 2001 and the Ph.D. degree for his work on electroencephalogram based brain-computer interfaces in July 2004, both from the Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland.

From 2000 to March 2001, he was with Fastcom Technology SA (Lausanne-Switzerland) as a software developer for video surveillance. In April 2001, he joined the doctoral program of the Signal Processing Institute at EPFL. He then became software development leader with the startup company EMITALL (Montreux-Switzerland). He has been with Philips Research Laboratories (Eindhoven-Netherlands) since May 2005. Currently, he leads an activity aimed at developing practical brain-computer interfaces.



Jeroen Breebaart received the M.Sc. degree in biomedical engineering and the Ph.D. degree in auditory psychophysics from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 1997 and 2001, respectively.

From 2001 to 2007, he was with the Digital Signal Processing Group, Philips Research, conducting research in the areas of spatial hearing, parametric audio coding, automatic audio content analysis, and audio effects processing, and where, since 2007, he has been the leader of the biometrics cluster of the

Information and System Security Group, expanding his research scope toward secure and convenient identification. He published more than 50 papers at international conferences and journals. He contributed to the development of audio coding algorithms as recently standardized in MPEG and 3GPP such as HEAAC, MPEG Surround, and the upcoming standard on spatial audio object coding.

Dr. Breebaart is a member of the AES. He also actively participates in the ISO/IEC IT security techniques standardization committee and is significantly involved in several EU-funded projects.



Raymond N. J. Veldhuis received the B.S. degree in electrical engineering from the University of Twente, Enschede, The Netherlands, in 1981 and the Ph.D. degree from Nijmegen University, Nijmegen, The Netherlands, in 1988. His dissertation was titled "Adaptive restoration of lost samples in discrete-time signals and digital images."

From 1982 to 1992, he was a Researcher with Philips Research Laboratories, Eindhoven, The Netherlands, in various areas of digital signal processing, such as audio and video signal restoration and audio source coding. From 1992 to 2001, he was with the Institute of Perception Research, Eindhoven, in speech signal processing and speech synthesis. From 1998 to 2001, he was Program Manager with the Spoken Language Interfaces research program. He is currently an Associate Professor with the University of Twente, working in the fields of biometrics and signal processing. His research interest involves digital signal processing for audio, images and speech, statistical pattern recognition, and biometrics. He has been active in the development of MPEG standards for audio source coding.



Tom A. M. Kevenaer received the M.Sc. and Ph.D. degrees from the Technical University of Eindhoven, Eindhoven, The Netherlands, in 1961, 1986, and 1993, respectively.

As a Postdoctoral Fellow, first with the University of Eindhoven and later with Hitachi Central Research Laboratory, Tokyo, Japan, he worked on several subjects in the field of design automation for analog and RF circuits. In 1996, was with Philips Research where he guided the development of a large variety of analog design tools such as analog/RF circuit simulators, optimization tools, fault simulation tools, etc. In 2001, his attention moved to the field of cryptography with a focus on low-power cryptographic algorithms and standards and slightly later became involved in privacy enhancing technologies for biometric applications. In 2008, he cofounded the priv-ID B.V. startup company, which develops and sells software products and solutions to protect the privacy of biometric information. He has over 40 publications in international conferences and journals and is coeditor of the book *Security with Noisy Data* (Springer 2007).



Willem Jonker studied mathematics and computer science at Groningen University, Groningen, The Netherlands, in 1962. He then joined the Delft University of Technology, Delft, The Netherlands for his Ph.D. research on knowledge-based systems.

In 1989, he was with KPN Research to work on knowledge-based systems, database systems, and distributed systems. In 1992, he was with the European Computer Industry Research Center, Munich, Germany (a joint research laboratory of Bull, ICL, and Siemens), to work on intelligent and federated database systems. Late 1994, he was back with KPN Research to become the head of the database group. In 1999, he founded a new research department of KPN Research at the campus of Twente University, Enschede, The Netherlands. In September 2001, he was with Philips Research to work on secure content management in networked environments, and in April 2004, he became Department Head of the Information and System Security group. In October 2005, he became Sector Head with the Lifestyle Experience Solutions sector. Finally, since 1998, he has been a part-time Full Professor of computer science with University of Twente, Enschede, The Netherlands.