

# Classification Performance Comparison of a Continuous and Binary Classifier under Gaussian Assumption

E.J.C. Kelkboom and J. Breebaart

Philips Research,  
The Netherlands

Emile.Kelkboom@philips.com  
Jeroen.Breebaart@philips.com

R.N.J. Veldhuis

University of Twente, Fac. EEMCS,  
The Netherlands

R.N.J.Veldhuis@utwente.nl

## Abstract

Template protection techniques are privacy and security enhancing techniques of biometric reference data within a biometric system. Several of the template protection schemes known in the literature require the extraction of a binary representation from the real-valued biometric sample, which raises the question whether the bit extraction method reduces the classification performance. In this work we provide the theoretical performance of the optimal log likelihood ratio continuous classifier and compare it with the theoretical performance of a binary Hamming distance classifier with a single bit extraction scheme as known from the literature. We assume biometric data modeled by a Gaussian between-class and within-class probability density with independent feature components and we also include the effect of averaging multiple enrolment and verification samples.

## 1 Introduction

The introduction of the ePassport with fingerprint raised some question marks on the privacy of the users and the security of the stored biometric data, especially when the Dutch government decided to store the fingerprint samples in a centralized database [1]. The security and privacy risks related to the storage of biometric data are (i) *identity theft* where an adversary steals the stored reference template and impersonates the genuine user of the system by some spoofing mechanism, (ii) *limited-renewability* implying the limited capability to renew a compromised reference template due to the limited number of biometric instances (for example we only have ten fingers, two irises or retinas, and a single face), (iii) *cross-matching* or linking reference templates of the same subject across databases of different applications, and (iv) derivation of *sensitive medical information* where it is known that biometric data may reveal the presence of certain diseases.

The field of template protection aims at mitigating these privacy and security risks by developing techniques that provide (i) *irreversibility* implying that it is impossible or at least very difficult to retrieve the original biometric sample from the reference template, (ii) *renewability* where it is possible to renew the reference template when necessary, and (iii) *unlinkability* which prevents cross-matching. In the literature, numerous template protection methods such as the *Fuzzy Commitment Scheme* (FCS) [2], *Helper Data System* (HDS) [3, 4, 5], *Fuzzy Extractors* [6, 7], *Fuzzy Vault* [8, 9] and *Cancellable Biometrics* [10] have been proposed.

In general, the extracted feature vector from the biometric sample is real-valued, while several of the proposed template protection schemes depend on the extraction of a binary representation from the biometric sample. The classification performance of the template protection scheme thus depends on the combination of the bit extraction process and the binary classifier. Yet, an unanswered question is what the difference is between the theoretical classification performance at binary level (after the bit extraction) and the performance at the continuous level (before the bit extraction). A potential performance loss after the bit

extraction process may represent the penalty for the requirement to extract a binary representation from the biometric sample. In [11], the performance of a single bit extraction process with a Hamming distance classifier has been theoretically determined under the assumption that the biometric data is Gaussian distributed. In this work we first discuss the theoretical performance of the optimal likelihood-ratio continuous classifier, under the assumption that the biometric data is Gaussian distributed. In [12], the theoretical performance has been derived where the reference template is the average of  $N_e$  enrolment samples with a single verification sample. We extend this analysis by including the averaging of  $N_v$  verification samples. Lastly, we compare the theoretical performance difference between the continuous and binary classifier and study the influence of the number of feature components and the number of enrolment and verification samples.

The outline of this paper is as follows. In Section 2 we briefly describe the model of the biometric data under Gaussian assumption including the averaging of multiple enrolment and verification samples. The theoretical performance estimation for the continuous classifier is derived in Section 3 and Section 4 briefly describes the theoretical performance for the binary classifier known from the literature. The theoretical performance comparison between the two classifiers and the effect of averaging multiple enrolment and verification samples is studied in Section 5. We conclude with our final remarks in Section 6.

## 2 Preliminaries

Random variables are underlined. Let  $\underline{x}_i \simeq N(\underline{\mu}_e, \sigma_w^2)$ ,  $i = 1, \dots, N_e$  denote the enrolment samples (features, in fact) and  $\underline{y}_i \simeq N(\underline{\mu}_v, \sigma_w^2)$ ,  $i = 1, \dots, N_v$  the verification samples with  $\sigma_w^2$  being the within-class variance. We assume that for a given class mean  $\mu$  the samples drawn from that class are i.i.d. The enrolment and verification class means are also Gaussian random variables, in particular  $\underline{\mu}_e, \underline{\mu}_v \simeq N(0, \sigma_b^2)$  with  $\sigma_b^2$  being the between-class variance. The reference template  $\underline{r}$  and the verification template  $\underline{v}$  are sample means, i.e.

$$\underline{r} = \frac{1}{N_e} \sum_{i=1}^{N_e} \underline{x}_i \quad (1)$$

$$\underline{v} = \frac{1}{N_v} \sum_{i=1}^{N_v} \underline{y}_i. \quad (2)$$

Because the samples are assumed to be independent we obtain  $\underline{r} \simeq N(\underline{\mu}_e, \frac{\sigma_w^2}{N_e})$  and  $\underline{v} \simeq N(\underline{\mu}_v, \frac{\sigma_w^2}{N_v})$ .

In the genuine case, the features originate from the same, unknown, mean, i.e.  $\underline{\mu}_e = \underline{\mu}_v = \underline{\mu}$ . In the impostor case the features originate from arbitrary means drawn from the between-class density. The purpose of the classifier is to discriminate between genuine and impostor comparisons.

## 3 Continuous Classifier Performance

### 3.1 The Log Likelihood Ratio Comparison Score

Let  $p_{\underline{r}, \underline{v}}(r, v|\text{gen})$ ,  $p_{\underline{r}, \underline{v}}(r, v|\text{imp})$  denote the joint probability densities of  $\underline{r}$  and  $\underline{v}$  in the genuine and impostor cases, respectively. The likelihood ratio in this case is defined by

$$l(r, v) = \frac{p_{\underline{r}, \underline{v}}(r, v|\text{gen})}{p_{\underline{r}, \underline{v}}(r, v|\text{imp})}. \quad (3)$$

We conveniently arrange  $\underline{r}$  and  $\underline{v}$  in a column vector  $\mathbf{z} = (\underline{r}, \underline{v})^T$ . We write

$$p_{\underline{r}, \underline{v}|\text{gen}}(r, v|\text{gen}) = \frac{1}{2\pi\sqrt{|C_{\text{gen}}|}} e^{-\frac{\mathbf{z}^T C_{\text{gen}}^{-1} \mathbf{z}}{2}} \quad (4)$$

$$p_{\underline{r}, \underline{v}|\text{imp}}(r, v|\text{imp}) = \frac{1}{2\pi\sqrt{|C_{\text{imp}}|}} e^{-\frac{\mathbf{z}^T C_{\text{imp}}^{-1} \mathbf{z}}{2}}, \quad (5)$$

where  $C_{\text{gen}}$  and  $C_{\text{imp}}$  are the co-variance matrices for the genuine and imposter comparisons, respectively. For  $p_{\underline{r}, \underline{v}|\text{gen}}(r, v|\text{gen})$ , we can write

$$p_{\underline{r}, \underline{v}|\text{gen}}(r, v|\text{gen}) = \int_{-\infty}^{\infty} p_{\underline{r}|\underline{\mu}}(r|\mu) p_{\underline{v}|\underline{\mu}}(v|\mu) p_{\underline{\mu}}(\mu) d\mu. \quad (6)$$

Using this we obtain  $E\{\underline{r}|\text{gen}\} = E\{\underline{v}|\text{gen}\} = 0$ ,  $E\{r^2|\text{gen}\} = \sigma_b^2 + \frac{1}{N_e}\sigma_w^2$ ,  $E\{v^2|\text{gen}\} = \sigma_b^2 + \frac{1}{N_v}\sigma_w^2$ , and  $E\{rv|\text{gen}\} = \sigma_b^2$ , therefore,

$$C_{\text{gen}} = \begin{pmatrix} \sigma_b^2 + \frac{1}{N_e}\sigma_w^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \frac{1}{N_v}\sigma_w^2 \end{pmatrix}. \quad (7)$$

In the impostor case,  $\underline{r}$  and  $\underline{v}$  are independent and

$$C_{\text{imp}} = \begin{pmatrix} \sigma_b^2 + \frac{1}{N_e}\sigma_w^2 & 0 \\ 0 & \sigma_b^2 + \frac{1}{N_v}\sigma_w^2 \end{pmatrix}. \quad (8)$$

Instead of the likelihood ratio we compute a comparison score based on the log likelihood ratio, from which constant terms and factors have been removed:

$$s(r, v; N_e, N_v) = -\mathbf{z}^T C_{\text{gen}}^{-1} \mathbf{z} + \mathbf{z}^T C_{\text{imp}}^{-1} \mathbf{z}. \quad (9)$$

On substitution of (7) and (8) into (9) and after simplification and elimination of constants we obtain the following expression for the comparison score

$$s(r, v; N_e, N_v) = -\frac{r^2}{\sigma_b^2 + \frac{1}{N_e}\sigma_w^2} - \frac{v^2}{\sigma_b^2 + \frac{1}{N_v}\sigma_w^2} + 2\frac{rv}{\sigma_b^2}, \quad (10)$$

in which we included the number of enrolment  $N_e$  and verification  $N_v$  samples as parameters. Examples of  $s(r, v; N_e, N_v)$  are portrayed by contour plots in Fig. 1 for different number of enrolment  $N_e$  or verification  $N_v$  samples with within-class and between class variance  $\sigma_w^2 = \sigma_b^2 = 1$ . Positive comparisons scores are obtained when the  $\{r, v\}$ -pair is close the  $r = v$ -axis (the positive diagonal line) and being further away from the origin increases the comparison score. Negative comparisons scores are obtained when the  $\{r, v\}$ -pair is closer the  $-r = v$ -axis (the negative diagonal line) and increases when further away from the origin. Increasing both the number of enrolment and verification samples shifts the zero-contour lines closer to the  $r = v$ -axis, because the expected uncertainty has decreased due to the reduction of the within-class variance by averaging multiple samples. Hence, a similar behavior can be expected when decreasing the within-class variance directly. Increasing only the number of enrolment (verification) samples mainly shifts the horizontal (vertical) zero-contour line closer to the  $r = v$ -axis.

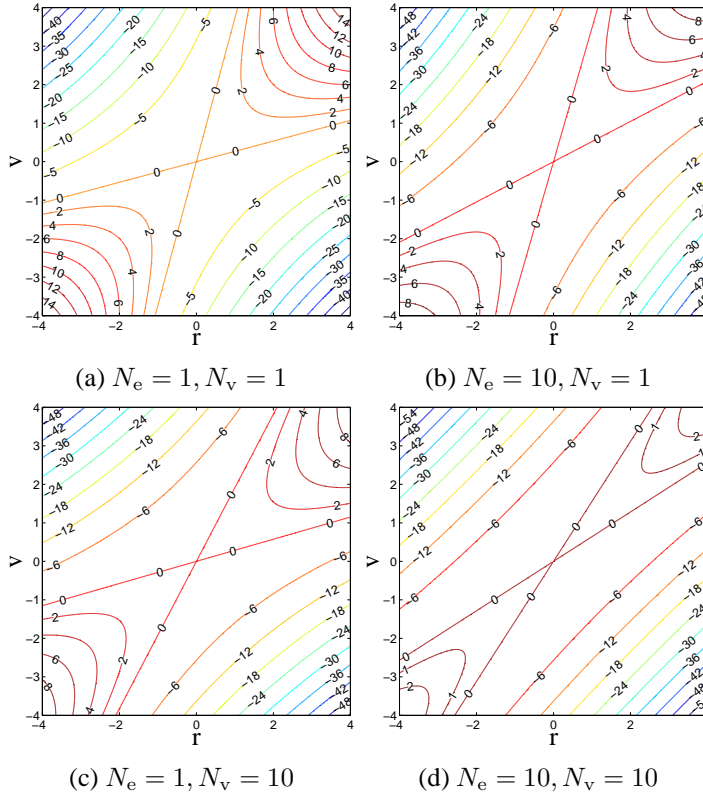


Figure 1: Contour plot of the log likelihood ratio comparison score  $s(r, v; N_e, N_v)$  from (10) with within-class and between class variance  $\sigma_w^2 = \sigma_b^2 = 1$  for different number of enrolment  $N_e$  or verification  $N_v$  samples.

### 3.2 Comparison Score Density and the Classification Performance

In order to estimate the performance, first we have to derive the density of the log likelihood comparison score  $s(r, v; N_e, N_v)$  from (10), denoted as  $p_{s_j|\text{gen}}(s|\text{gen})$  for the genuine case and  $p_{s_j|\text{imp}}(s|\text{imp})$  for the imposter case. By combining (10) with the joint probability density  $p_{r,v|\text{gen}}(r, v|\text{gen})$  from (4) for the genuine and  $p_{r,v|\text{imp}}(r, v|\text{imp})$  from (5) for the imposter case, respectively, we approximate the score density by means of numerical integration of the joint probability density along the score contour. Because  $s(r, v; N_e, N_v)$  from (10) is derived for the univariate case, thus the score densities  $p_{s_j|\text{gen}}(s|\text{gen})$  and  $p_{s_j|\text{imp}}(s|\text{imp})$  are for the univariate case as denoted by the  $j$  subscript.

For the multivariate case, when there are  $n$  independent feature components, the likelihood ratio equals the product of the likelihood ratio of each component. Because we use the log likelihood ratio as the comparison score, the multivariate comparison score equals the sum of the  $n$  univariate scores defined in (10). Hence, the multivariate comparison score density for the genuine  $p_{\underline{s}|\text{gen}}(s|\text{gen})$  and imposter case  $p_{\underline{s}|\text{imp}}(s|\text{imp})$  becomes the convolution of the univariate score density  $p_{s_j|\text{gen}}(s|\text{gen})$  and  $p_{s_j|\text{imp}}(s|\text{imp})$ , respectively, namely

$$p_{\underline{s}}(s) \stackrel{\text{def}}{=} (p_{s_1} * p_{s_2} * \dots * p_{s_n})(s). \quad (11)$$

Because the log likelihood comparison score is a similarity score, a match is returned only when the comparison score is larger than or equal to the operating point  $T$ . The two error types are a match obtained at an imposter comparison known as a false match and a non-match at a genuine comparison known as a false non-match. As the performance measures,

we use the false non-match rate (FNMR)  $\beta(T)$  and the false match rate (FMR)  $\alpha(T)$  at the operating point  $T$ . With the multivariate score density we can compute the FNMR and FMR as

$$\beta(T) = \int_{-\infty}^T p_{\underline{s}|\text{gen}}(s|\text{gen})ds, \quad (12)$$

$$\alpha(T) = \int_T^{\infty} p_{\underline{s}|\text{imp}}(s|\text{imp})ds. \quad (13)$$

### 3.3 Results

Fig. 2 illustrates several examples of the approximated score density at (a) genuine and (b) imposter comparisons for the univariate case for different number number of enrolment and verification samples with  $\sigma_b^2 = \sigma_w^2 = 1$ , and (c) their corresponding receiver operating characteristics (ROC) curves. Similarly for the multivariate case in (d), (e) and (f), respectively, but for different dimensions  $n$  with  $\sigma_b^2 = \sigma_w^2 = N_e = N_v = 1$ . Note that the genuine score density is symmetric at a score of zero, while the imposter density is skewed towards the negative scores. Averaging multiple enrolment and verification samples has the effect of concentrating the genuine score density closer to zero, while skewing the imposter score density further towards the negative values. Both effects improve the performance as observed by the ROC curves. For the multivariate case, when increasing the number of components  $n$  the imposter score density significantly skews and shifts to the negative values while the genuine density becomes broader but remains symmetric. Overall, both effect combined improve the performance as illustrated by the ROC curves.

## 4 Binary Classifier Performance

The theoretical performance of a binary classifier when using a bit extraction method based on a single threshold at the background mean has been studied in [11]. For the genuine comparisons, the average bit-error probability of component  $j$  is analytically determined to be equal to

$$P_e^{\text{ge}}[j] = \frac{1}{2} - \frac{1}{\pi} \arctan \left( \frac{\sigma_b[j]}{\sigma_w[j]} \frac{\sqrt{N_e N_v}}{\sqrt{N_e + N_v + \left(\frac{\sigma_b[j]}{\sigma_w[j]}\right)^{-2}}} \right). \quad (14)$$

The bit-error probability determines the number of bit errors or Hamming distance  $\epsilon$  between the binary vectors extracted in the enrolment and verification phase. Under the assumption of having independent components, the probability mass function (pmf) of  $\epsilon$  is the following convolution

$$p_{\underline{\epsilon}}(\epsilon) \stackrel{\text{def}}{=} (P_1 * P_2 * \dots * P_{n_c})(\epsilon), \quad (15)$$

where  $P_j = [1 - P_e[j], P_e[j]]$  is the marginal pmf of the single bit extracted from component  $j$ . Note that the number of bit errors  $\epsilon$  is a distance score and a match is obtained when  $\epsilon$  is smaller or equal to the operating point  $T$ . Thus, the FNMR  $\beta(T)$  and FMR  $\alpha(T)$  at the operating point  $T$  are defined as

$$\begin{aligned} \beta(T) &= \sum_{\epsilon=T+1}^n p_{\underline{\epsilon}|\text{gen}}(\epsilon|\text{gen}), \\ \alpha(T) &= \sum_{\epsilon=0}^T p_{\underline{\epsilon}|\text{imp}}(\epsilon|\text{imp}), \end{aligned} \quad (16)$$

where the bit-error probability  $P_e^{\text{ge}}$  from (14) is used for the genuine case and  $P_e^{\text{im}} = 0.5$  for the imposter case.

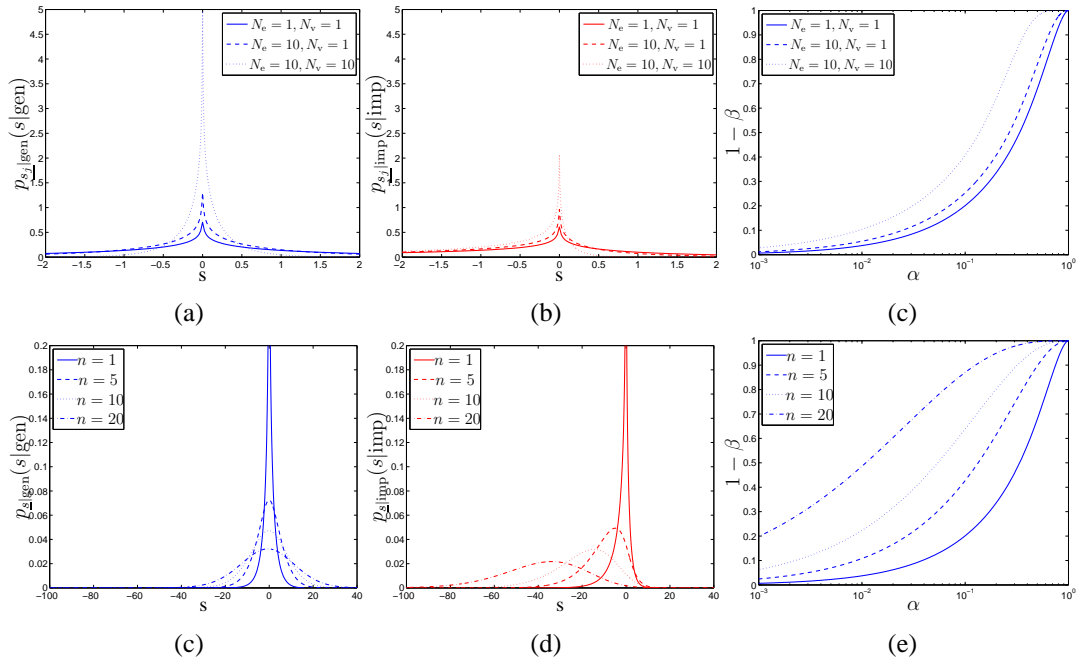


Figure 2: The approximated comparison score density for the univariate case with within-class and between class variance  $\sigma_w^2 = \sigma_b^2 = 1$  for different number of enrolment  $N_e$  or verification  $N_v$  samples is shown (a) for the genuine  $p_{s_j|gen}(s|gen)$  and (b) imposter  $p_{s_j|imp}(s|imp)$  case, and (c) portrays the corresponding ROC curves. Furthermore, for the multivariate case is shown (d)  $p_{s_j|gen}(s|gen)$ , (e)  $p_{s_j|imp}(s|imp)$ , and (f) the ROC curves for different number of components  $n$  with  $\sigma_w^2 = \sigma_b^2 = N_e = N_v = 1$ .

## 5 Performance Comparison

A comparison of the theoretical performances determined in Section 3 for the continuous classifier and Section 4 for the binary classifier is portrayed by the ROC curves in Fig. 3(a) for different feature dimensions  $n$  with  $\sigma_w^2 = \sigma_b^2 = N_e = N_v = 1$ , for different number of enrolment samples  $N_e$  with  $n = 10$  and  $\sigma_w^2 = \sigma_b^2 = N_v = 1$  in Fig. 3(b), and in Fig. 3(c) for different number of enrolment and verification samples  $N_e = N_v$  with  $n = 10$  and  $\sigma_w^2 = \sigma_b^2 = 1$ . The continuous classifier is denoted by the prefix  $C$ , while the binary classifier is denoted by the prefix  $B$ . In all three cases the results clearly show that the continuous classifier outperforms the binary classifier and changing either the dimension  $n$  or the number of enrolment or verification samples has a greater improvement for the continuous classifier. A drawback of the binary classifier is that the binarization process under consideration extracts a single bit by coarsely dividing the feature space of a component in two regions only and therefore discarding essential information. This loss is clearly shown by the ‘ $n=1$ ’ ROC curve in Fig. 3(a), where the continuous classifier ROC curve has an infinite number of operating points and can reach any FMR of FNMR value, while the binary classifier has only two operating points where the smallest FMR is 50%. As observed in Fig. 3(a), this information loss has a snowball effect when increasing the dimension  $n$ , because the performance of the continuous classifier has a greater improvement with increasing  $n$  than the binary classifier performance. Extracting a single bit becomes more disadvantageous when the within-class variance is suppressed by increasing the number of enrolment or verification samples, or similarly having better feature components, i.e. feature components with a larger feature quality ratio  $\frac{\sigma_b}{\sigma_w}$ . When having better feature components it may be better to extract

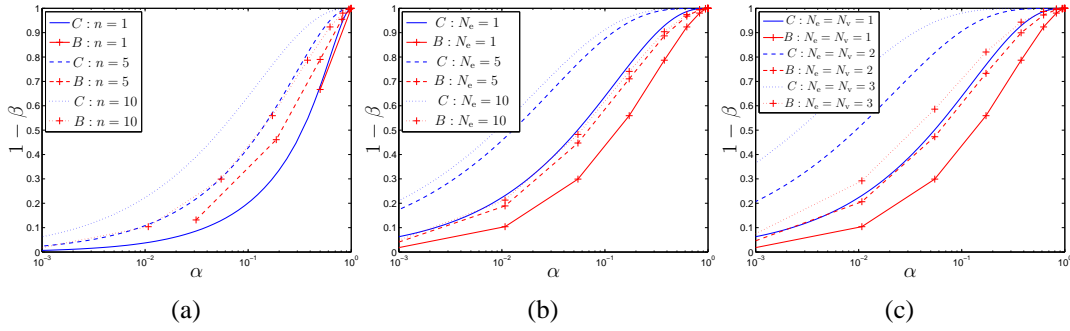


Figure 3: The ROC performance comparison between the continuous (denoted by  $C$ ) and binary classifier (denoted by  $B$ ) for (a) different feature dimensions  $n$  with  $\sigma_w^2 = \sigma_b^2 = N_e = N_v = 1$ , (b) different number of enrolment samples  $N_e$  with  $n = 10$  and  $\sigma_w^2 = \sigma_b^2 = N_v = 1$ , and (c) different number of enrolment and verification samples  $N_e = N_v$  with  $n = 10$  and  $\sigma_w^2 = \sigma_b^2 = 1$ .

more bits instead of one.

## 6 Conclusion

The requirement to extract a binary representation from the real-valued biometric sample for several template protection schemes known in the literature raises the question whether the bit extraction method reduces the classification performance. In this work we compared the theoretical performance of the optimal log likelihood ratio continuous classifier with the binary Hamming distance classifier under the assumption of Gaussian biometric data modeled by the between-class and within-class densities with independent feature components and including the averaging of multiple enrolment and verification samples.

In the literature, the theoretical performance for the binary classifier consisting of a single bit extraction method based on thresholding has been studied. Similarly, the theoretical performance of a continuous classifier based on the log likelihood ratio comparison scores has been analyzed, but was limited to the averaging of multiple enrolment samples only. Hence, in this work we extended the analysis by including the averaging of multiple verification samples. We approximated the density of the comparison score for the univariate and multivariate case, from which we computed the corresponding performance curve.

Consequently, we compared the theoretical performance of the continuous and binary classifier and studied the effect of the number of the feature dimension and the number of enrolment and verification samples. In all cases the continuous classifier outperforms the binary classifier, which is expected as the likelihood ratio is the optimal classifier if the class-conditional probability is well-known. In this work we assumed the class-conditional probability to be well defined. In practice, however, the performance advantage of the continuous classifier will be less because it is known to be difficult to have a perfect estimation of the class-conditional probability, especially at high feature dimensions or correlated feature components. A drawback of the binary classifier under consideration is that the bit extraction method coarsely divides the feature space of a component in only two regions in order to extract a single bit and therefore discarding essential information. This drawback is amplified when the within-class noise is suppressed by increasing the number of enrolment or verification samples, where it may be more advantageous to extract more than one bit from each feature component.

As future work, it would be of great interest to derive the theoretical performance of more

advanced bit extraction methods that can extract more robust bits or multiple bits from each component in order to close the gap between the continuous and binary classifier. Furthermore, it is important to investigate the sensitivity of both classifiers with respect to correlated feature components and estimation errors of the class-conditional probability.

## References

- [1] NRC, “Fingerprints in passports can’t be used by the police - yet ,” 18 September 2009. [Online]. Available: [http://www.nrc.nl/international/Features/article2363938.ece/Fingerprints\\_in\\_passports\\_cant\\_be\\_used\\_by\\_the\\_police\\_-\\_yet](http://www.nrc.nl/international/Features/article2363938.ece/Fingerprints_in_passports_cant_be_used_by_the_police_-_yet)
- [2] A. Juels and M. Wattenberg, “A fuzzy commitment scheme,” in *6th ACM Conference on Computer and Communications Security*, November 1999, pp. 28–36.
- [3] 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.
- [4] 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.
- [5] 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, pp. 393 – 402.
- [6] 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.
- [7] 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.
- [8] A. Juels and M. Sudan, “A fuzzy vault scheme,” *Designs, Codes and Cryptography*, vol. 38, no. 2, pp. 237–257, February 2006.
- [9] K. Nandakumar, A. K. Jain, and S. Pankanti, “Fingerprint-based fuzzy vault: Implementation and performance,” in *IEEE Transactions on Information Forensics and Security*, December 2007, pp. 744–757.
- [10] N. K. Ratha, S. Chikkerur, J. H. Connell, and R. M. Bolle, “Generating cancelable fingerprint templates,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 561–572, April 2007.
- [11] E. J. C. Kelkboom, G. Garcia Molina, J. Breebaart, R. N. J. Veldhuis, T. A. M. Kevenaar, and W. Jonker, “Binary biometrics: An analytic framework to estimate the performance curves under gaussian assumption,” *IEEE Transactions on Systems, Man and Cybernetics Part A, Special Issue on Advances in Biometrics: Theory, Applications and Systems (accepted)*, 2010.
- [12] R. N. Veldhuis and A. Bazen, “One-to-template and one-to-one verification in the single and multi-user case,” in *26th Symposium on Information Theory in the Benelux*, Brussels, 2005.