

Performances of the Likelihood-ratio Classifier based on Different Data Modelings

C. Chen, R.N.J. Veldhuis
Signals and Systems Group, Electrical Engineering
University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands
{c.chen, r.n.j.veldhuis}@utwente.nl

Abstract—The classical likelihood ratio classifier easily collapses in many biometric applications especially with independent training-test subjects. The reason lies in the inaccurate estimation of the underlying user-specific feature density. Firstly, the feature density estimation suffers from insufficient number of user-specific samples during the enrollment phase. Even if more enrollment samples are available, it is most likely that they are not reliable enough. Furthermore, it may happen that enrolled samples do not obey the Gaussian density model. Therefore, it is crucial to properly estimate the underlying user-specific feature density in the above situations. In this paper, we give an overview of several data modeling methods. Furthermore, we propose a discretized density based data model. Experimental results on FRGC face data set has shown reasonably good performance with our proposed model.

Index Terms—likelihood-ratio classifier, density estimation, quantization

I. INTRODUCTION

The statistical pattern recognition technologies for biometric applications fall into two major categories: density-probability driven and data-criterion driven. The density-probability driven approaches, such as the Bayesian decision, hidden Markov Model (HMM) and higher-order statistics rely on the estimation of the probability over samples, by using maximum likelihood method, whereas data-criterion driven approaches such as linear discriminant function (LDF), support vector machine (SVM) aim to find a function with a specified structure or a hyperplane to minimize a criterion, without the knowledge of the underlying probability. As a Bayesian decision classifier with equal prior probability, the likelihood-ratio classifier [1], is theoretically optimal in the Neyman-Pearson sense. Unfortunately, its performances tend to turn down in many practical applications, especially with independent training-test subjects. Usually the reason lies in the inaccurate estimation of the underlying feature densities. To solve this problem, some specified data modelings have been proposed. In this paper, we give an overview of some existing data modeling methods. Moreover, we propose a model based on discretized density. Experimental results on FRGC face data set show reasonably good performance with our proposed model.

This paper proceeds as follows. In Section II, we give a description of several existing models, together with our proposed model. In Section III, we present the experimental results on FRGC face data set with some discussions, and the conclusions are drawn in Section IV.

II. METHOD

The design of a likelihood-ratio based biometric verification system usually includes three steps: training, enrollment and verification. Prior to modeling the data and constructing the classifier, which are carried out in the enrollment and verification steps, a common training step is applied. With a training data set \mathcal{D}_t , the goal of the training step is to extract the “right” features with a reduced dimensionality from the raw measurements. Several leading feature extraction methods are Principle Component Analysis (PCA) [2] [3], Independent Component Analysis (ICA) [4] and Linear Discriminant Analysis (LDA) [5], [6]. In some applications, a combined PCA/LDA method is used [7]. With the reduced dimensionality, we can build a model of the classifier. In the context of a likelihood-ratio classifier, the genuine user and the background density need to be estimated from an enrollment data set \mathcal{D}_e , and subsequently a discriminant function is calculated. Eventually, the verification decision is made on the discriminant values from a verification data set \mathcal{D}_v .

Essentially, the design of such a system involves C two-category classifications, where C denotes the total number of users. This means that for every user ω_i , a covariance matrix $\Sigma_i, i = 1, \dots, C$ needs to be estimated. Nevertheless, it often happens that there are not enough user-specific enrollment samples to accurately estimate the covariance matrix. To solve this problem, some specified covariance matrix models are adopted. Two common primary assumptions of these models are (1) the features are statistically independent, which means the covariance matrix Σ_i is diagonal; (2) the background densities are identical for all features. In this chapter, we give a description of some popular models. Additionally, we propose a new model by discretizing the continuous features and further building the likelihood-ratio classifier on the discrete feature density.

A. Model 1: $\Sigma_i = \sigma_i^2 \mathbf{I}$, Gaussian density

Let $\mathbf{x} = (x_1, \dots, x_d)^t$ be the d -dimensional feature vector. In this model we assume that the genuine user ω_i density $p_g(\mathbf{x}|\omega_i)$ is Gaussian $p_g(\mathbf{x}|\omega_i) \sim N(\boldsymbol{\mu}_i, \Sigma_i)$ with mean $\boldsymbol{\mu}_i$ and covariance matrix Σ_i . Similarly, we assume that the background density is normalized as a Normal distribution $p_b(\mathbf{x}|\bar{\omega}_i) \sim N(0, 1)$. Therefore, the discriminant function of

user ω_i equals to:

$$g_i(\mathbf{x}) = \sum_{j=1}^d \ln \frac{p_g(x_j|\omega_i)}{p_b(x_j|\bar{\omega}_i)}. \quad (1)$$

The verification decision is then made by giving a threshold T to $g_i(\mathbf{x})$.

B. Model 2: $\Sigma_i = \Sigma = \sigma^2 \mathbf{I}$, Gaussian density

Sometimes even though in Model 1 we reduce the number of parameter estimation by assuming independent features with Gaussian densities, the number of enrollment samples always seems too small to make a reliable estimation of the user-specific covariance matrix Σ_i . To solve this problem, Model 2 is adopted by further assuming that the covariance matrix is user-independent and therefore identical to every user. That is $\Sigma_i = \Sigma, i = 1, \dots, C$, where C denotes the entire populations. Therefore, when the training data is a good representative of the entire populations, the covariance matrix can be approximately calculated from the training data \mathcal{D}_t .

C. Model 3: Arbitrary density

In some cases, the feature vectors do not obey Gaussian density. Therefore, it is necessary to estimate the genuine user and the background density by using non-parametric methods, such as histogram, Parzen windows and k_n -nearest-neighbor [3].

Here we give an example of an equal probability histogram based density estimation. The key factors of a histogram estimation stem from the number and the location of the bins. Hence there comes a variety of methods to determine the bins [8]. In the case of biometrics, due to the lack of enrollment samples, it is difficult to design a histogram estimation with user-specific bins. Therefore, we propose to locate the bins according to the training data \mathcal{D}_t . Note that it is not accessible to the histogram of the entire enrollment data.

Consider a one-dimensional feature component x , let N_t be the number of samples in the training data, K be the number of histogram bins. The locations of the bins are then determined as:

$$n_b = \frac{N_t}{K}, \quad (2)$$

$$f(x, l_k, h_k) = n_b, k = 1, \dots, K, \quad (3)$$

where l_k, h_k indicates the lower and the higher boundaries of the k^{th} bin, and function f counts the number of samples within the bin $[l_k, h_k]$. An illustration is shown in Fig. 1(a) and 1(b). Constructing the bins in this way in fact constraints all the K bins with equal number of samples n_b , which turns the background density into an uniform density $P_{b,k}(x|\bar{\omega}_i) = 1/K$. Once the bins are determined from the entire training data \mathcal{D}_t , with the enrollment samples, the histogram density $P_{g,k}(x|\omega_i)$ of the genuine user ω_i is calculated as:

$$n_{g,i,k} = f(x, l_k, h_k), \quad (4)$$

$$P_{g,k}(x|\omega_i) = \frac{n_{g,i,k}}{n_{g,i}}, k = 1, \dots, K, \quad (5)$$

with $n_{g,i} = \sum_{k=1}^K n_{g,i,k}$ the number of enrollment samples of user ω_i (see Fig. 1(c)).

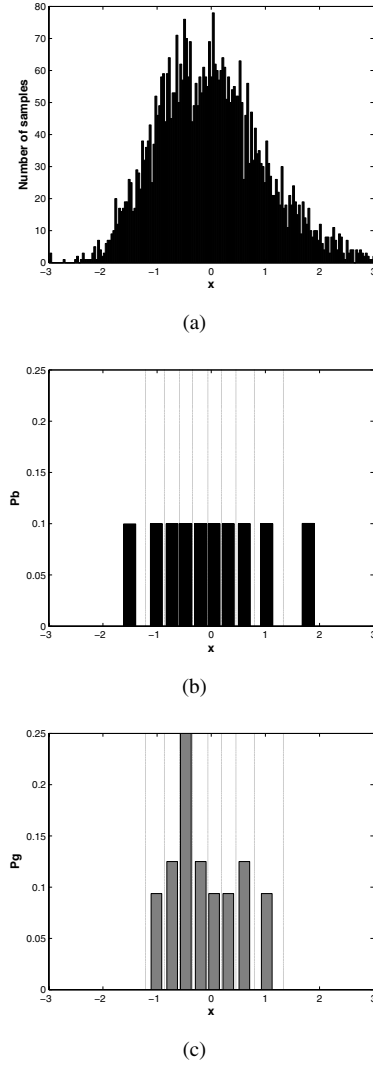


Fig. 1. An example of the histogram based density estimation with $K = 10$. (a) the entire population in the training data; (b) the background density determined by the bins; (c) a genuine user density determined by the bins.

In the d -dimensional feature vector case $\mathbf{x} = (x_1, \dots, x_d)^t$. Let $\tilde{k}_j, j = 1, \dots, d$ be the index of the bin where x_j is located, the discriminant function becomes:

$$g_i(\mathbf{x}) = \sum_{j=1}^d \ln \frac{P_{g,\tilde{k}_j}(x_j|\omega_i)}{P_{b,\tilde{k}_j}(x_j|\bar{\omega}_i)} = \sum_{j=1}^d \ln(K \cdot P_{g,\tilde{k}_j}(x_j|\omega_i)). \quad (6)$$

It is noticeable that the discriminant function is discrete, since the non-parametric density estimation relies on empirical data.

D. Model 4: Discretized density

Quantization technology has been widely used in signal processing as a lossy data compression process [9]. The core idea is that by converting a range of continuous signals into discrete symbols, the most important information of the signal

is maintained. In biometric applications, the estimated continuous feature density is often inaccurate due to the unreliable samples. Therefore, it is possible to apply quantization, with some loss of the information, to yield a reasonable “guess” of the ground truth density. To apply a likelihood-ratio classifier on a discretized density involves two steps: (1) Determine quantization bins; (2) Calculate the genuine user probability and the background probability within each quantization bin.

So far, there has been some quantization methods designed for biometric data [10], [11], [12], [13]. These works are originally motivated for the protection of the biometric information, and the classification is conducted in the binary domain. The bins can be either globally designed [10], [11], or be user-specific [12], [13]. Note that the histogram estimation we proposed in Model 3 can be seen as an empirical quantization method with globally determined bins. For a likelihood-ratio classifier, in addition to the bin design, we need to calculate the probability of both the genuine user and the background within the bins. For this purpose, we can either empirically count the number of samples falling into the bins, or resort to some models (e.g. Gaussian).

Here we present an example of modeling the discretized density from the quantization method in [12]. Consider a one-dimensional feature component x from user ω_i , given the number of quantization bins K , the genuine feature mean μ and standard deviation σ can be calculated from the user enrollment samples. Hence the quantization intervals are determined as:

$$K_1 = \lfloor \frac{K+1}{2} \rfloor, \quad (7)$$

$$I_{i,k} = \left[\mu - [2(K_1 - k) + 1]r\sigma, \mu - [2(K_1 - k) - 1]r\sigma \right), \quad k = 2, \dots, K-1, \quad (8)$$

where $I_{i,k}$ indicates the location of the bins, with $I_{i,1} = (-\infty, \mu - (2K_1 - 3)r\sigma)$, $I_{i,K} = [\mu - [2(K_1 - K) + 1]r\sigma, \infty)$ as the left and the right tails. The parameter r determines the width of the intervals, which are all fixed to $2r\sigma$, with the exception that the left and the right tails are extended to infinity. To calculate the genuine user and the background probability, we employ a Normal $N(x, 0, 1)$ and a Gaussian $N(x, \mu, \sigma)$ density model, respectively. That is:

$$P_{b,k}(x|\bar{\omega}_i) = \int_{I_{i,k}} N(x, 0, 1), \quad (9)$$

$$P_{g,k}(x|\omega_i) = \int_{I_{i,k}} N(x, \mu, \sigma), k = 1, \dots, K. \quad (10)$$

Note that the genuine user probability $P_{g,k}(x|\omega_i)$ is symmetric around the mean. An example can be seen in Fig. 2.

In the d -dimensional feature vector case, let $\tilde{k}_j, j = 1, \dots, d$ be the bin where x_j is located, the discriminant function becomes:

$$g_i(\mathbf{x}) = \sum_{j=1}^d \ln \frac{P_{g,\tilde{k}_j}(x_j|\omega_i)}{P_{b,\tilde{k}_j}(x_j|\bar{\omega}_i)}. \quad (11)$$

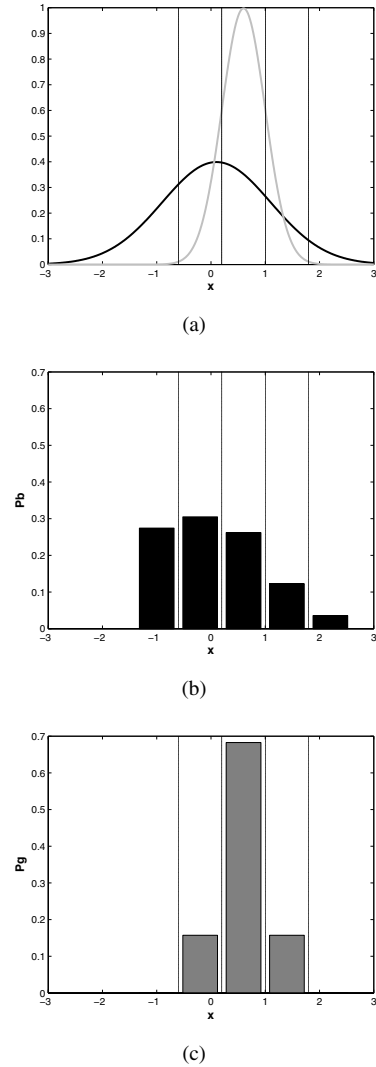


Fig. 2. An example of the discretized density estimation with $K = 5$, $r = 1$. (a) the Gaussian model, background (black); genuine user (gray); (b) the background density determined by the bins; (c) a genuine user density determined by the bins.

E. Model comparison

Here we compare the above models by investigating the properties of their discriminant functions $g(x)$ in a one-dimensional case, with an example in Fig. 3.

Model 1 and Model 2, by fully employing a theoretical density (e.g. Gaussian in Fig. 3(a)), yield a continuous discriminant function (Fig. 3(b)). The likelihood-ratio classifier built on such discriminant function is optimal in the Neyman-Pearson sense, if the underlying density strictly fits the theoretical model. However, once the samples do not fit the model that we employed, for instance, the data is not Gaussian, or the mean and standard deviation are not correct, or even the features are not independent, the likelihood-ratio classifier collapses.

Model 3 and Model 4, with less or no reliance on a theoretical model, yield a discrete discriminant function (Fig. 3(c)). Such discriminant function has the characteristic that

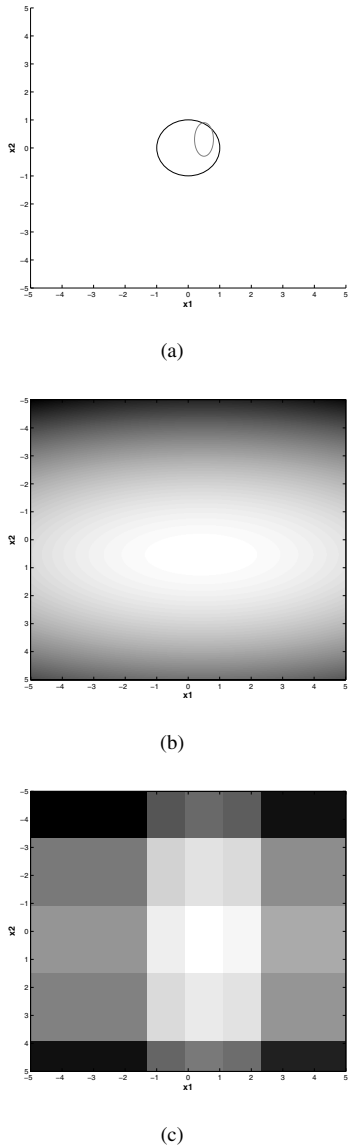


Fig. 3. An illustration of samples drawn from two-dimensional distributions. (a) the theoretical Gaussian density, the background density (black), and the genuine user density (gray); (b) the discriminant function $g(\mathbf{x})$ in model 2; (c) discriminant function $g(\mathbf{x})$ in Model 4, with $K = 5$, $r = 2$.

even though the $g(\mathbf{x})$ calculated in (11) is based on a model assumption in Fig. 3(a), the samples which fall in the same quantization cell share the same $g(\mathbf{x})$ value, which is less relevant to the employed model within the cell and consequently less sensitive to the density variation within one cell, as compared to the continuous discriminant function in Fig. 3(b). This can be seen as a way to use the theoretical model to determine a discriminant value at a larger scale (between cells) while ignoring the model details at a smaller scale (within one cell). Such discriminant function might bring benefits to samples which are so unreliable that we can not fully trust the theoretical model that we employ. However, the disadvantages of this model is that there is a shape cut of the discriminant

values on the cell boundaries. Moreover, when the features are not independent, the likelihood-ratio classifier collapses.

III. EXPERIMENTS AND DISCUSSIONS

We tested the four data models on a face database FRGC (version 1) [14]:

- **FRGC_T**: This is the total FRGC (version 1) face data set, containing various number of images of 275 users. The images were taken under both controlled and uncontrolled conditions and were aligned using manually labeled landmarks. A normalized region of interest (ROI) was extracted from every 128 by 128 image, resulting in 8762 pixel values as the raw measurement.
- **FRGC_S**: This is a subset of FRGC_T, containing 198 users with at least 2 images per user. The images were taken under uncontrolled conditions.

In the experiment, we randomly selected independent users for training and test (including enrollment and verification), while the enrollment and verification are involved with identical users. To evaluate the error with a cross-validation procedure, we repeated our experiment with a number of 5 partitionings. With n data samples per user, the division of the data is listed in Table I.

TABLE I
TRAINING, ENROLLMENT AND VERIFICATION DATA DIVISION (NUMBER OF USERS \times NUMBER OF SAMPLES PER USER) FOR FRGC_T AND FRGC_S.

	Training	Enrollment	Verification	Partitioning
FRGC _T	$210 \times n$	$65 \times 2n/3$	$65 \times n/3$	5
FRGC _S	$150 \times n$	$48 \times 2n/3$	$48 \times n/3$	5

TABLE II
 EER (%) PERFORMANCES OF THE FOUR MODELS, ON DATA SET (A) FRGC_T AND (B) FRGC_S.

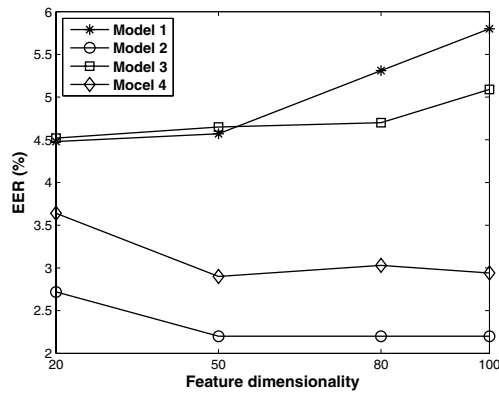
	$d = 20$	$d = 50$	$d = 80$	$d = 100$
Model 1	4.48	4.57	5.31	5.80
Model 2	2.72	2.20	2.20	2.20
Model 3	4.52 ($K = 4$)	4.65 ($K = 2$)	4.70 ($K = 2$)	5.09 ($K = 2$)
Model 4	3.64 ($r = 2$)	2.90 ($r = 2$)	3.03 ($r = 1$)	2.94 ($r = 1$)

(a)

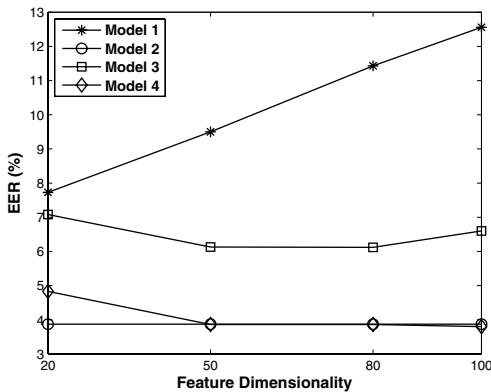
	$d = 20$	$d = 50$	$d = 80$	$d = 100$
Model 1	7.73	9.50	11.43	12.56
Model 2	3.87	3.87	3.87	3.87
Model 3	7.08 ($K = 4$)	6.13 ($K = 2$)	6.12 ($K = 2$)	6.60 ($K = 2$)
Model 4	4.83 ($r = 3$)	3.86 ($r = 1$)	3.86 ($r = 1$)	3.80 ($r = 1$)

(b)

We evaluated the equal error rate (EER) performances at a number of predefined PCA/LDA [7] output feature dimensionalities. Model 3 were tested with various settings of K , and Model 4 (at $K = 3$) were tested with various settings of r . Their best performances, together with the results of Model 1 and 2, are presented in Fig. 4 and Table II. Overall Model 1 results in high EER , the performance deteriorates dramatically with the increasing feature dimensionality. The moderate results of Model 3 suggest that with lower feature dimensionality (e.g. $d = 20$), a larger number of bins yields better performance (e.g. $K = 4$). However, by increasing the



(a)



(b)

Fig. 4. EER (%) performances of the four models, on (a) FRGC_T and (b) FRGC_S.

feature dimensionality with more unreliable features, the best performances merely allow a histogram with 2 bins. The reason for this might be that Model 3 is strongly dependent on the empirical data, which easily leads to the curse of dimensionality in high-dimensional cases. The results of Model 4 suggest that with lower feature dimensionality (e.g. $d = 20, 50$), a larger r achieves better performance ($r = 2$). By contrast, higher feature dimensionality (e.g. $d = 80, 100$) allows small r ($r = 1$). The performances of Model 2 and 4 exhibit stable performances with respect to the increasing dimensionality, which implies that assuming user-independent covariance or employing user-specific quantization might be less prone to unreliable data. Nevertheless, all these data modeling methods are highly data-dependent. Hence it is not possible to conclude which one should be the gold standard.

IV. CONCLUSIONS

In this paper we give an overview of several data modeling methods used in biometric applications for a likelihood-ratio classifier. Furthermore, we propose a discretized density estimation model which relies on a quantization scheme. Experiments on FRGC face data shows that both using user-independent covariance matrix (Model 2), and applying a dis-

cretized density (Model 4) give reasonably good performance and the models are less prone to unreliable data.

REFERENCES

- [1] A.M. Bazen and R.N.J. Veldhuis. Likelihood-ratio-based biometric verification. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1):86–94, 2004.
- [2] I. T. Jolliffe. *Principal Component Analysis*. Springer, 2nd edition, 2002.
- [3] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley-Interscience, 2nd edition, 2000.
- [4] Te-Won Lee. *Independent Component Analysis - Theory and Applications*. Springer, 1st edition, 1998.
- [5] J.H. Friedman. Regularized discriminant analysis. *Journal of the American Statistical Association*, 84(405):165–175, 1989.
- [6] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.R. Muller. Fisher discriminant analysis with kernels. In *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*, 1999.
- [7] R.N.J. Veldhuis, A. Bazen, J. Kauffman, and P. Hartel. Biometric verification based on grip-pattern recognition. *Security, Steganography, and Watermarking of Multimedia Contents VI. Edited by Delp, Edward J., III; Wong, Ping W. Proceedings of the SPIE*, 5306:634–641, 2004.
- [8] D.W. Scott. On optimal and data-based histograms. *Biometrika*, 66(3):605–610, 1979.
- [9] A. Gersho and R.M. Gray. *Vector Quantization and Signal Compression*. Springer, 1st edition, 1991.
- [10] P. Tuyts, A.H.M. Akkermans, T.A.M. Kevenaer, G.J. Schrijen, A.M. Bazen, and R.N.J. Veldhuis. Practical biometric authentication with template protection. In Takeo Kanade, Anil K. Jain, and Nalini K. Ratha, editors, *AVBPA*, volume 3546 of *Lecture Notes in Computer Science*, pages 436–446. Springer-Verlag, 2005.
- [11] T.A.M. Kevenaer, G.J. Schrijen, M. van der Veen, A.H.M. Akkermans, and F. Zuo. Face recognition with renewable and privacy preserving binary templates. In *IEEE Workshop on Automatic Identification Advanced Technologies (AutoID 2005)*, pages 21–26. IEEE Computer Society, 2005.
- [12] Y. Chang, W. Zhang, and T. Chen. Biometrics-based cryptographic key generation. In *ICME*, pages 2203–2206, 2004.
- [13] C. Chen, R.N.J. Veldhuis, T.A.M. Kevenaer, and A.H.M. Akkermans. Multi-bits biometric string generation based on the likelihood ratio. In *IEEE Conference on Biometrics: Theory, Applications and Systems*, 2007.
- [14] P. J. Phillips, P. J. Flynn, W. T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W.J. Worek. Overview of the face recognition grand challenge. In *CVPR (1)*, pages 947–954, 2005.