PERFORMANCE ANALYSIS OF IRIS BASED IDENTIFICATION SYSTEM AT THE MATCHING SCORE LEVEL

N. A. Schmid, B. Cukic, M. Ketkar
Lane Department of Computer Science and Electrical Engineering
West Virginia University, Morgantown WV 26506
{natalias,cukic,manasik}@csee.wvu.edu

H. Singh
Department of Statistics
West Virginia University, Morgantown WV 26506
hsingh@stat.wvu.edu

ABSTRACT

We analyze the performance of an iris based recognition system. We consider a practical setting where matching scores are accessible for collecting data. We assume that multiple scans from the same iris are available and design the decision rules based on this assumption. We show that vectors of matching scores are described by a Gaussian model with dependent components both under the Genuine and Imposter hypotheses. Two test statistics: the average Hamming distance and the log-likelihood ratio are designed. We show that the log-likelihood ratio with well estimated maximum likelihood parameters in it outperforms the first test statistic. We further use empirical approach, Chernoff bound, and Large Deviations approach to predict the performance of the recognition system.

1. INTRODUCTION

Iris-based identification is gaining considerable attention from the research community in parallel with its public acceptance. Modern cameras used for iris acquisition are less intrusive compared to earlier iris scanning devices and public awareness of system reliability is developing. This resulted in the appearance of a number of new efficient encoding and preprocessing techniques for iris [1,2]. A typical iris system consists of four major subsystems: (i) image acquisition, (ii) preprocessing, (iii) encoding, and (iv) decision making. Current research is mostly focused on redesigning the second and the third subsystems. However, a framework for comprehensive analysis of iris systems or a study on how various processing steps influence performance of iris-based identification system does not exist. We believe that our work is the first of this kind. At this stage, we consider a practical setting where matching scores are accessible for collecting data. We model matching scores, a sequence of Hamming distances, as realizations of a random process with a number of unknown parameters. These unknown parameters are evaluated empirically. The problems of verification and recognition are stated as a binary and (M+1)-ary hypothesis testing, respectively. Here M is the number of individual classes/irises to be identified. The models are then applied to predict the performance of large-scale iris-based identification systems from a small amount of available data. We use empirical approach, Chernoff bound [3], and Large Deviations approach [4] to predict the performance.

2. PROPOSED MODEL

We assume that all IrisCodes are obtained as a result of processing iris images with J. Daugman's algorithm [5]. The major processing blocks of the system are displayed in Fig. 1. An incoming image (i) is enhanced and

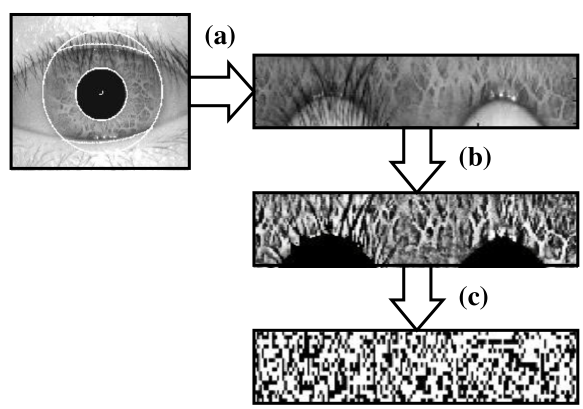


Fig. 1 Block-diagram ((a) Image transformation, (b) Preprocessing, (c) Encoding)

transformed into a pseudo polar representation, (ii) encoded using Gabor filters, and (iii) component-by-component quantized to two levels (zero/one) based on the sign of the corresponding filtered image entry. The quantized sequence of components is called IrisCode.

Most existing systems evaluate the discrepancy (matching scores) between two IrisCodes using the Hamming distance. The problem of deciding if an input iris belongs to the claimed identity is often stated as a hypothesis testing problem, where the Hamming distance, d plays the role of the test statistic. Given a threshold γ, the imposter hypothesis (IH) is accepted if $d \geq \gamma$. Otherwise, the imposter hypothesis is rejected. A typical measure of recognition performance is the average probability of error or, in the verification case, the ROC curve formed as a combination of the False Accept Rates (FAR) and False Reject Rates (FRR) for different values of γ.

2.1. Verification Case

In this study, we make the following practically feasible assumptions. (i) Each class/iris in the database and a new input are represented by K iris scans converted into IrisCodes. The practical value of K is $K \leq 10$. However, for the purpose of analysis we will assume that K can become large. (ii) The Hamming distance is calculated for arbitrary cross-coupled sets of K IrisCodes such that no same IrisCode is involved twice. This requirement reduces additional dependences among Hamming distances. Let $\mathbf{d} = [d_1, d_2, \ldots, d_K]$ be a vector of K Hamming distances formed according to the assumptions above. Consider the following two test statistics: the average of K Hamming distances

$$\bar{d} = \sum_{k=1}^{K} d_k / K \tag{1}$$

and the plug-in log-likelihood ratio test. The average Hamming distance, $\bar{d}$, is an intuitive statistic. Averaging is typically used to reduce the noise and thus to improve performance [9].

By Daugman [5], two arbitrary selected IrisCodes are strongly correlated, and so are Hamming distances. We model K-dimensional vectors of Hamming distances as realizations of K-dimensional Gaussian random vectors with correlated entries. To form the second test statistic, consider the following hypothesis testing problem. Under the imposter hypothesis, H_0, the vector $\mathbf{d}$ is Gaussian distributed with common unknown mean for all entries $m_0 \cdot \mathbf{1}$ and unknown covariance matrix $\mathbf{R}_0$. Under the genuine hypothesis, H_1, the vector $\mathbf{d}$ is Gaussian distributed with common unknown mean $m_1 \cdot \mathbf{1}$ (m_0 and m_1 are distinct) and unknown covariance matrix $\mathbf{R}_1$ ($\mathbf{R}_0$ and $\mathbf{R}_1$ are distinct). The matrices $\mathbf{R}_1$ and $\mathbf{R}_0$ are cyclic with the first row given by $\sigma_i^2[1, \rho_i, \rho_i, \cdots, \rho_i]$, $i = 0,1$. Since the parameters of the models are unknown, we assume the availability of training data (N independent copies of the vector $\mathbf{d}$) and apply the maximum likelihood (ML) estimation method to estimate the parameters. The ML estimates are given by

$$\hat{m}_i = \frac{1}{KN}(\sum_{i=1}^{N}\sum_{j=1}^{K} d_{i,j}), \qquad \hat{\sigma}_i^2 = \frac{tr(\mathbf{A}_i)}{KN},$$

and

$$\hat{\rho}_i = \frac{\mathbf{1}^{\mathrm{T}}\mathbf{A}_i\mathbf{1} - tr(\mathbf{A}_i)}{(K-1)tr(\mathbf{A}_i)}, \quad i = 0,1,$$

where $\mathbf{A}_i$ is the Wishart distributed $K \times K$ dimensional matrix with the (p,q) entry given by

$$\mathbf{A}_i(p,q) = \sum_{n=1}^{N}(d_{n,p} - \hat{m}_i)(d_{n,q} - \hat{m}_i),$$

$tr(\mathbf{A}_i)$ denotes its trace, $\mathbf{1}$ is a vector column of all ones, and N is the number of independent data realizations.

The plug-in log-likelihood ratio for this model is given by:

$$l_K \equiv \frac{1}{K}\log\frac{\hat{p}(\mathbf{d} \mid GH)}{\hat{p}(\mathbf{d} \mid IH)} = -\frac{1}{2K}(\mathbf{d} - \hat{m}_1\mathbf{1})^{\mathrm{T}}\hat{\mathbf{R}}_1^{-1}(\mathbf{d} - \hat{m}_1\mathbf{1}) + \frac{1}{2K}(\mathbf{d} - \hat{m}_0\mathbf{1})^{\mathrm{T}}\hat{\mathbf{R}}_0^{-1}(\mathbf{d} - \hat{m}_0\mathbf{1}) - \frac{1}{2K}\log\det(\hat{\mathbf{R}}_1\hat{\mathbf{R}}_0^{-1}) \tag{2}$$

where $\hat{m}_i$, $\hat{\sigma}_i^2$, and $\hat{\rho}_i$, $i = 0,1$ are the ML estimated parameters.

2.2. Identification Case

Suppose now that IrisCodes, each of length n, from M individual irises are collected and stored in the database. Assume K copies of IrisCodes are available from the same iris. Denote by $\mathbf{X}(k), k = 1, \ldots, M$ and $\mathbf{Y}$ random vectors underlying the IrisCodes of the kth individual and a candidate that submits his/her iris for identification, respectively. Assume that a candidate is also represented by K IrisCodes.

We further state the identification problem as a multi-hypothesis testing problem. The test statistic in this case is a vector of log-likelihood ratios

$$\mathbf{l} = \left[l(1), l(2), \ldots, l(M)\right]^{\mathrm{T}},$$

where the kth entry is

$$l(k) = \frac{1}{K}\log\frac{p(\mathbf{d}(k) : GH)}{p(\mathbf{d}(k) : IH)}.$$

To identify the candidate the following test is performed

Decide H_i : $l(i) > l(j)$ for all $j \neq i$ and $l(i) > \gamma$,

Decide H_0 : otherwise,

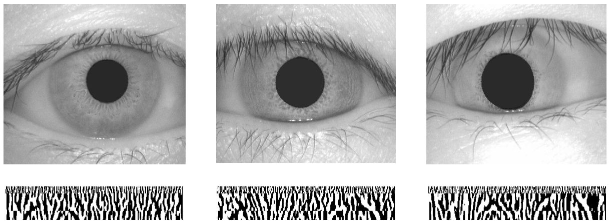


Fig. 2 Iris Images and corresponding IrisCodes.

where γ is a threshold and $i = 1,\ldots,M$.

3. PEFORMANCE ANALYSIS

Performance evaluation is an important step in designing a decision making system. In practical setting, it is of interest to derive a single analytical expression that can be used to predict performance of a large scale system based on a small amount of available data. The literature contains a few results on performance analyses for iris-based recognition systems [7,8]. Both papers perform analysis under dramatically simplified conditions and mostly focus on evaluation of the False Accept Rate, the probability that no intruder is able to access the iris system. In this work, we avoid simplifying the model and consider the entire average probability of error as a measure of performance. Since the expressions for the False Accept Rate (FAR) and False Reject Rate (FRR) are often hard to evaluate directly, one can appeal to bounds and approximations. In this work, we use the Chernoff bound that is related to a more restrictive asymptotic approach called Large Deviations [4]. While the Chernoff bound is a valid tight upper bound on a probability of error for arbitrary selected parameters K and γ, the Large Deviations approach requires K be large to produce tight results. If the Large Deviations conditions are satisfied, the FAR can be approximated as $FAR(\gamma) = \Pr\left[l_K > \gamma \mid IH\right] \approx G(K,\gamma)\exp(-KI_0(\gamma))$,

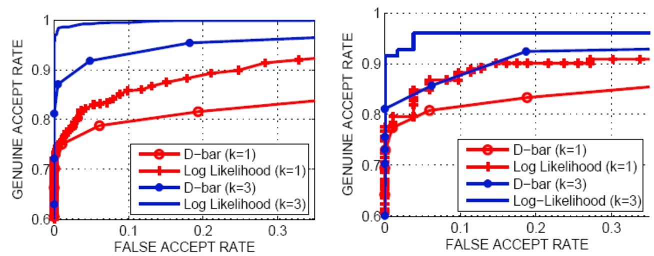
where $G(K,\gamma)$ is a slowly varying function of K and γ (often omitted in analysis) and $I_0(\gamma)$ is the Large Deviations rate function under imposter distribution. If the log-likelihood ratio is used as a test statistic, the rate function is given by

$$I_0(\gamma) = \sup_s \left[s\gamma - \frac{s}{2}\log\hat{f} + \frac{1}{2}\log\left(s\hat{f} - s + 1\right)\right],$$

where $\hat{f} = \dfrac{(1-\hat{\rho}_0)\hat{\sigma}_0^2}{(1-\hat{\rho}_1)\hat{\sigma}_1^2}$.

A similar expression can be obtained for the FRR. Because of limited space we do not present the expression for the Chernoff bound.

The performance of the multi-hypotheses testing problem is described the total probability of error given by



(a) (b)

Fig. 3 Shown are the plots of the Genuine Reject Rate vs. the False Accept Rate for (a) simulated and (b) bootstrapped data, K=1 and 3.

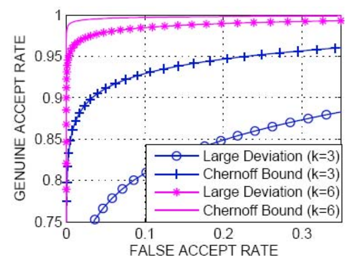


Fig. 4 ROC curves generated using the Chernoff bound and the Large Deviations approximations for K=3 and 6.

$$P(\text{error}) = \sum_{k=0}^{M}\pi_k \sum_{l=0, l\neq k}^{M} P(\text{decide } H_l \mid H_k \text{ is true}).$$

The total probability of error can be upper bounded by the probability of error for a binary hypothesis testing problem as

$$P(\text{error}) \leq \frac{M(M+1)}{2} \max_{k,l=0,\ldots,M,\, k\neq l}\left[\frac{\pi_k}{\pi_k+\pi_l} P(\text{error} \mid H_k^{\text{binary}}) + \frac{\pi_l}{\pi_k+\pi_l} P(\text{error} \mid H_l^{\text{binary}})\right],$$

where π_k are the prior on the hypothesis H_k. This expression can be further reduced to

$$P(\text{error}) \leq \frac{M(M+1)}{2} \max_{\alpha;\, k,l=0,\ldots,M,\, k\neq l}\left[\alpha P(\text{error} \mid H_k^{\text{binary}}) + (1-\alpha) P(\text{error} \mid H_l^{\text{binary}})\right], \quad (3)$$

where $\alpha = \pi_k / (\pi_k + \pi_l)$.

4. RESULTS

All experiments were performed on the CASIA dataset provided by the Chinese Academy of Sciences [6]. The database contains "non-ideal" iris images of 108 irises with 6 images per iris. The examples of images from the CASIA database together with the corresponding IrisCodes are shown in Fig. 2.

To obtain the ML estimates of the unknown parameters in Gaussian models, we formed 54 vectors each of size 6 and 108 vectors each of size 3 samples of genuine and imposter Hamming distances, respectively. The values of the estimated parameters are given by $\hat{m}_1 = 0.3832$, $\hat{\sigma}_1^2 = 0.0047$, $\hat{\rho}_1 = 0.1126$ and $\hat{m}_0 = 0.4613$, $\hat{\sigma}_0^2 = 2.4505\times10^{-4}$, $\hat{\rho}_0 = 0.3832$.

To validate the model fit we applied the multivariate Shapiro-Wilk test for normality. For a single dimension, the test produced the p-value equal to 0.46 for the imposter distribution and 0.52 for the genuine distribution. The critical p-value is $p_{\text{crit}} = 0.05$. For vectors with 3 components, the p-value is equal to 0.0722 for the imposter distribution and 0.2218 for the genuine distribution. The critical value for this case is 0.05. These results confirm that the model provides a reasonable fit especially for a single dimension or for vectors with 2 and 3 components.

The proposed decision statistics (1) and (2) are tested both on a dataset generated using the estimated parameters and on a dataset created by applying bootstrapping technique to the data in the CASIA dataset. The results of the direct testing are demonstrated in Fig. 3.

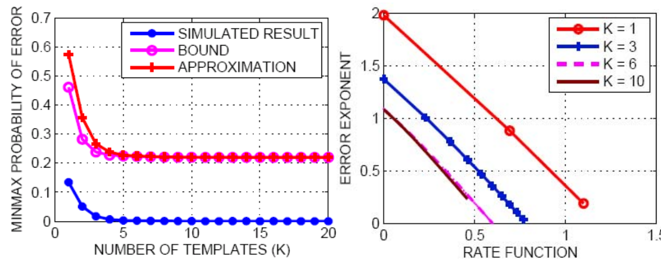
The results confirm a well known fact that the plug-in log-likelihood ratio test with well estimated parameters substituted in place of the true unknown parameters and with a model well fitted into data is almost optimal in the minimum probability of error (or Neyman-Pearson) sense.

The plots of the Genuine Reject Rate vs. False Accept Rate obtained using the Chernoff bounds and the Large Deviations approximations are shown in Fig. 4 (for *K*=3,6). Note that the ROC curves generated from Chernoff bounds are tight (ROC approaches directly computed ROC's from below).

As expected, the ROC curves based on the Large Deviations approximation provide a loose fit for a small value of K and improve as K increases. This approximation is useful when a quick estimate of an error order has to be obtained.

We further evaluate the upper bound on the total probability of error in recognition problem. The functions displayed in Fig. 5(a) are the minimax verification error, the corresponding minimax Chernoff bound, and the minimax Large Deviations approximation displayed as functions of the parameter K. Note that the minimax error is involved in (3). The bound on the total probability of error for the recognition problem is displayed as a function of the total number of classes/irises and the number templates used per iris in Fig. 5(b). Using terminology from [10], we introduce the recognition rate as $R = \log(M)/K$ and the bound on the recognition error exponent as

$$E(R) = -\frac{\log\left(Bound\ on\ P(error)\right)}{K}. \qquad (4)$$



(a) (b)

Fig. 5 (a) Minimax probability of error vs. the number of templates, K. (b) Bound on the recognition error exponent.

Since the minimax verification error is fixed for every given K, then

$$E(R) = -\frac{\log(C(K))}{K} - \frac{\log(M(M+1)/2)}{K},$$

where $C(K)$ is a function of K only. This explains the nature of lines in Fig. 5(b). Given the number of templates per class/iris and the total error probability, Fig. 5(b) and the expression (4) specify the maximum number of classes/irises that an iris recognition system may contain such that the total probability of error does not exceed the specified value.

5. REFERENCES

1. L. Masek, Recognition of Human Iris Patterns for Biometric Identification, M.S. Dissertation, The University if Western Australia, 2003, http://www.csse.uwa.edu.au/~pk/studentprojects/libor/
2. Li Ma, T. Tan, Y. Wang, and D. Zhang, "Personal Identification Based on Iris Texture Analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1519--1533, 2003.
3 H.L. Van Trees, *Detection, Estimation, and Modulation Theory*, Wiley, New York, 2001.
4. J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation*, John Wiley & Sons, Inc., New York, 1990.
5 J. Daugman, "The importance of being random: Statistical principles of iris recognition," *Pattern Recognition*, vol. 36, no. 2, 2001, pp 279-291.
6 CASIA Iris Image Database (ver. 1.0), http://www.sinobiometrics.com/casiairis.htm
7 Daugman J. (1993) "High confidence visual recognition of persons by a test of statistical independence." *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 15(11), pp. 1148-1161.
8 R. P. Wildes, "Automated iris recognition: An emerging biometric technology," *Proceedings of the IEEE 85 (9)*, 1348-1363, 1997.
9 R. C. Gonzales and R. E. Woods, *Digital Image Processing*, Addison-Wesley, 1992.
10. R. E. Blahut, *Principles and Practice of Information Theory,* Addison-Wesley, 1987.