

2005

Performance analysis of iris based recognition system at the matching score level

Manasi V. Ketkar
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Ketkar, Manasi V., "Performance analysis of iris based recognition system at the matching score level" (2005). *Graduate Theses, Dissertations, and Problem Reports*. 1633.
<https://researchrepository.wvu.edu/etd/1633>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Performance Analysis of Iris Based Recognition System at the Matching Score Level

by

Manasi V. Ketkar

Thesis submitted to the
College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Master of Science
in
Electrical Engineering

Natalia A. Schmid, D.Sc., Chair
Bojan Cukic, Ph.D.
Lawrence Hornak, Ph.D.
Xin Li, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2005

Keywords: Iris Recognition, Performance Analysis, Hypothesis testing, Minimum Probability of Error, Large Deviations, Chernoff Bound

Copyright 2005 Manasi V. Ketkar

Abstract

Performance Analysis of Iris Based Recognition System at the Matching Score Level

by

Manasi V. Ketkar

Master of Science in Electrical Engineering

West Virginia University

Natalia A. Schmid, D.SC., Chair

Over the past three years, iris based personal identification has gained considerable attention both from research groups and government organizations. Public acceptance of this biometric grew substantially too. Modern cameras used for iris acquisition are less intrusive compared to earlier iris scanning devices and public awareness of system reliability is slowly developing. A typical iris system consists of four major subsystems: (i) image acquisition, (ii) preprocessing, (iii) encoding, (iv) decision making. Most current research is focused on redesigning preprocessing and encoding techniques for iris systems. However, a framework for comprehensive analysis of iris recognition systems or a study on how various preprocessing steps influence performance of iris-based identification system does not exist. In this thesis, we propose a methodology to predict performance of a large-scale iris recognition system based on a small testing database available, using information theoretic approach.

In this work, we consider a practical setting where only matching scores are accessible for collecting data. We assume that multiple scans from the same iris are available. We model the matching scores, a sequence of Hamming distances, as realizations of a random process with a number of unknown parameters. These parameters are evaluated empirically. We then design two decision test statistics for the given matching scores. The problems of verification and identification are stated as a binary and $(M+1)$ -ary hypothesis testing problems, respectively. Here M is the individual number of iris classes to be identified. The proposed models are then applied to predict the performance of a large scale iris based recognition system from a small amount of available data. We use empirical approach, Chernoff bound and, Large Deviations approximation to predict the performance.

Acknowledgments

A few lines of acknowledgement do not fully express my gratitude and appreciation for those who guided and supported me through these last two years. I have been fortunate to be surrounded by excellent teachers, family and friends.

I have been privileged to have Dr. Natalia A. Schmid as my advisor. Her attention to detail, quest for excellence, love for perfection, ingenious ideas and understanding has always motivated me to give my best time and again. She has helped me develop and tune my academic skills including research, presentation and writing skills. I am greatly indebted to her not only for making my masters a memorable experience but also for the principles and values I gathered from her.

I am grateful to Dr. Bojan Cukic, Dr. Lawrence Hornak and Dr. Xin Li for their expert advice and support. I would also like to thank the late Dr. Harashinder Singh for his guidance, without whom this work would not have achieved this level of success.

I thank my lab mates for their support and technical discussions which often gave me ideas for my research. I wish all of them success in their future endeavors. I also thank my roommates and friends here in Morgantown for their support during hard times. My heartfelt thanks to my grandparents, my parents Dr. Vijay and Madhavi Ketkar, my sisters Gauri and Janhavi and my relatives for their support, efforts and sacrifice without which I would not have come this far. Last but surely not the least I thank my fiancé, Sagar, for understanding me, standing by me and sharing my joys and frustrations.

To all of you I dedicate this work.

Contents

Acknowledgments	iii
List of Figures	vi
List of Tables	vii
Notation	viii
1 Introduction	1
1.1 Biometric Technology	1
1.2 Iris Recognition	2
1.3 Contribution of the thesis	5
1.4 Organization of the thesis	5
2 Proposed Matching Score Model	6
2.1 Proposed Model for the Matching Scores	6
2.1.1 Verification Case	7
2.1.2 Identification Case	9
2.2 Summary	11
3 Proposed Decision Rules	12
3.1 Proposed Rule 1: The Average Hamming Distance ($D\text{-Bar}$)	12
3.2 Proposed Rule 2: Plug-in Log Likelihood Ratio	13
3.2.1 Verification Case	13
3.2.2 Identification Case	13
4 Performance Analysis	15
4.1 Background	15
4.2 Verification Case	15
4.2.1 Empirical Evaluation	16
4.2.2 Chernoff Bound	16
4.2.3 Large Deviations Approximation	18
4.3 Identification Case	20
4.3.1 The Worst Case Upper Bound	20
4.3.2 Union of Chernoff Bounds	21

5	Results	24
5.1	Results for CASIA Database	25
5.1.1	Generating Matching Scores	25
5.1.2	Maximum Likelihood parameter estimates	26
5.1.3	Shapiro-Wilk normality test results	27
5.1.4	Performance Analysis: Verification Case	27
5.1.5	Performance Analysis: Identification Case	31
5.2	Results for WVU Database	33
5.2.1	Maximum Likelihood parameter estimates	33
5.2.2	Shapiro-Wilk normality test results	34
5.2.3	Performance Analysis: Verification Case	34
5.2.4	Performance Analysis: Identification Case	35
6	Conclusions	38
7	Future Work	40
	References	42
A	Derivations for Rate Functions: Verification Case	45
A.1	Normalized Log-Moment Generating Function under Imposter Hypothesis	47
A.2	Threshold (γ) under Imposter Hypothesis	51
A.3	Normalized Log-Moment Generating Function under Genuine Hypothesis	52
B	Derivation for Chernoff Rate Function: Identification Case	53
C	Asymptotic Expected Values of Likelihood Function	56
C.1	Under Imposter Hypothesis	56
C.2	Under Genuine Hypothesis	57

List of Figures

1.1	Traditional Daugman's Iris Recognition System	3
2.1	Our interpretation of J. Daugman's Iris Recognition System	6
5.1	Example Iris Images and corresponding IrisCodes from CASIA Database	24
5.2	Example Iris Images and corresponding IrisCodes from WVU Database	25
5.3	Pairing IrisCodes to get Imposter Matching Scores	26
5.4	Pairing IrisCodes to get Genuine Matching Scores	26
5.5	ROC curves obtained using empirical approach (no rotation compensation)	28
5.6	ROC curves obtained using empirical approach (with rotation compensation)	29
5.7	Rate functions for the bound and approximation	30
5.8	ROC curves using the bound and approximation	30
5.9	Dependence of the error exponent $E(R)$ on the recognition rate R	32
5.10	Upper Bound on Probability of error	34
5.11	ROC using empirical approach (no rotation compensation)	35
5.12	ROC using empirical approach (with rotation compensation)	35
5.13	ROC curves obtained using the bound and approximation	36
5.14	Dependence of the error exponent $E(R)$ on the recognition rate R	36
5.15	Upper Bound on the total probability of error	37

List of Tables

5.1	Maximum Likelihood Parameter Estimates: CASIA Database	27
5.2	p-values for Shapiro-wilk normality test: CASIA Database	27
5.3	Chernoff bound: CASIA database (No Rotation Compensation)	33
5.4	Chernoff bound: CASIA database (With Rotation Compensation)	33
5.5	Maximum Likelihood Parameter Estimates: WVU Database	33
5.6	p-values for Shapiro-wilk normality test: WVU Database	34
5.7	Chernoff bound: WVU database (No Rotation Compensation)	36
5.8	Chernoff bound: WVU database (With Rotation Compensation)	37

Notation

We use the following notation and symbols throughout this thesis

$(\cdot)^T$: matrix transpose;
$tr(\cdot)$: trace of a matrix;
$\hat{\cdot}$: Maximum Likelihood Estimate of a parameter;
$\langle \cdot, \cdot \rangle$: Euclidean inner product
\otimes	: XOR operator
\cap	: AND operator
$\ \cdot\ $: Euclidean norm
d	: Hamming Distance (Matching Score)
K	: No. of copies (templates) per iris image
M	: No. of individual iris classes
\sim	: Statistical Distribution of a vector of data
γ	: Threshold
GH	: Genuine Hypothesis
IH	: Imposter Hypothesis
MLE	: Maximum Likelihood Estimation
$P(\text{error})$: Minimum Probability of Error
FAR	: False Accept Rate
FRR	: False Reject Rate
ROC	: Receiver Operating Characteristics
LD	: Large Deviations Approximation
CB	: Chernoff Bound

Chapter 1

Introduction

1.1 Biometric Technology

Over the years, user authentication has become an inseparable part of all transactions involving human computer interaction. Most conventional modes of authentication are based on ‘what we know’ (e.g. passwords) and/or ‘what we have’ (e.g. ID card). Biometrics brings in stronger authentication capabilities by adding a third factor - ‘who we are’ (based on our inherent physiological or behavioral characteristics). Secure personal authentication and identification are challenging problems for modern society. With the rapid development of new information technologies, it is also essential to develop new security systems to prevent unauthorized access and abuse. Biometric technologies are hence becoming the foundation of an extensive array of highly protected identification and personal verification systems.

Biometrics are automated methods of recognizing a person based on a physiological or behavioral characteristic. In practice, this means capturing and processing an image of a unique feature of an individual, and comparing it with a processed image captured previously. Among the characteristics measured are face, fingerprint, hand geometry, iris, and voice. Biometric identification systems all rely upon forms of random variation among persons based on these features. More complex is the randomness, the more pronounced features for identification; because more dimensions of independent variation produce signatures having greater uniqueness. As in all pattern recognition problems, the key issue here is the relation between inter-class (between two or more classes/users) and intra-class (within a single class/user) variability [1]. Objects can be reliably classified only if the variability among different instances of a given class is less than the

variability between different classes. A good biometric is characterized by use of a feature that is

1. *highly unique* - so that the chance of any two people having the same characteristic will be minimal,
2. *stable* - so that the feature does not change over time, and
3. *be easily captured* - in order to provide convenience to the user and prevent misrepresentation of the feature.

1.2 Iris Recognition

The use of a biometric for identification purposes requires that the particular biometric factor be unique for each individual, that it can be readily measured and that it is invariant over time. Biometrics such as signatures, fingerprints, voiceprints all have significant drawbacks. Although signatures are cheap to acquire and easy to store, they are often impossible to identify automatically with high assurance, and are easily forged. Electronically recorded voiceprints are susceptible to changes in a person's voice. Fingerprints or handprints require physical contact. They also can be counterfeited and marred by artifacts.

Human iris on the other hand possesses great biometric advantages. The iris is part of the middle coat of the eye. It is a thin diaphragm stretching across the anterior portion of the eye, supported by the lens. The cornea lies in front of the iris and provides a transparent protective coating. As an internal (yet externally visible) organ of the eye, the iris is well protected from the environment. The evidences from clinical observations and developmental biology claim that the structure of iris is unique to an individual and its epigenetic pattern remains stable throughout adult life [2]. The iris also has a great advantage from mathematical stand of view. Its pattern variability among different persons is enormous. From image processing stand of view, iris is a complex texture. As almost a planer object, its image is relatively insensitive to small changes in viewing angle. Finally, the ease of localizing eyes in faces, and the distinctive annular shape of the iris, facilitate reliable and precise isolation of this feature and the creation of a size-invariant representation. All these characteristics make iris very attractive for use as a biometric for iden-

tifying individuals.

Image processing techniques can be employed to extract the unique iris representation from an image of the eye and encode it into a biometric template, which can be stored in a database. This biometric template contains a mathematical representation of the unique information contained in the iris and is used to alleviate the automatic identification.

The person's identity can be resolved either by verification or identification. Verification involves confirming or denying a person's claimed identity. In identification, one has to establish a person's identity. When a candidate wishes to be identified by an iris based recognition system, his/her iris is first imaged and then converted to a template. This template is then compared with each template stored in the database until either a matching template is found and the subject is identified, or no match is found and the subject remains unidentified.

Traditional Iris Recognition System

Although prototype Iris recognition systems had been proposed earlier, it was not until the early nineties that Cambridge researcher, John Daugman, implemented a working prototype of an automated iris recognition system [3][4]. The major processing steps of this system are summarized in Fig. 1.1.

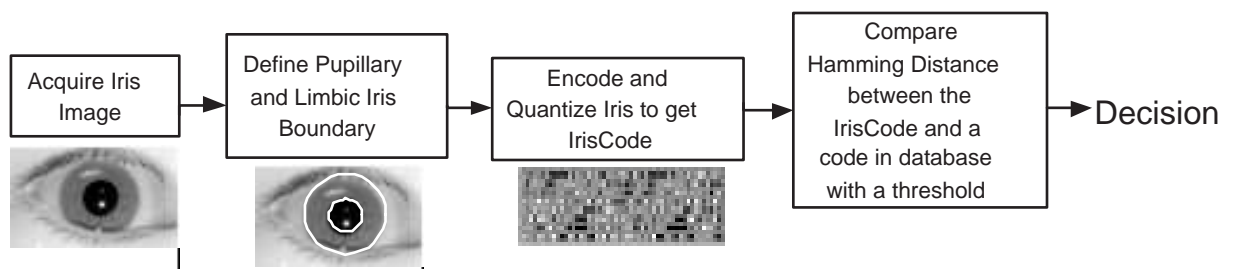


Fig. 1.1: Traditional Daugman's Iris Recognition System

During the encoding step, each isolated iris pattern is first demodulated to extract its phase information using quadrature 2D Gabor Wavelets, generating complex-valued coefficients. The real and imaginary parts specify the coordinates of the phasor in the complex plane. According to J. Daugman [4], only phase information is useful for discrimination because amplitude information is not very reliable. It depends upon extraneous factors such as imaging contrast,

illumination, and camera gain. The angle of each phasor is quantized to one of the four quadrants, setting two bits of phase information. This process is repeated all across the iris with many wavelet sizes, frequencies, and orientations, to extract 2048 bits forming an IrisCode. An equal number of masking bits are also computed to signify whether any iris region is obscured by eyelids, contains any eyelash occlusions, specular reflections, or poor signal to noise ratio and thus should be ignored in the demodulation code as an artifact.

The key to iris recognition is the failure of test of statistical independence, which involves a large number of degrees of freedom. This test virtually is guaranteed to pass whenever the phase codes (IrisCodes) for two different eyes are compared. It uniquely fails when any iris's phase code is compared with another version of itself. Most existing systems evaluate the discrepancy between two IrisCodes using Hamming distances (HD). It is defined as the number of bits at which two IrisCodes differ from each other. After the candidate image is converted to an IrisCode (say, codeA), it is compared with the template IrisCode in the database (say, codeB) to get the Hamming distance as follows

$$HD = \frac{\| (codeA \otimes codeB) \cap maskA \cap maskB \|}{\| maskA \cap maskB \|} \quad (1.1)$$

where maskA and maskB denote the masking bits generated for codeA and codeB respectively, \otimes is the XOR operator and \cap is the AND operator. The L2 norm ($\| \|$) is used to calculate fractional HD as a measure of dissimilarity between the two IrisCodes, with the value of 0 for a perfect match. The HD is then compared with a threshold to decide whether the candidate user is a Genuine or an Imposter. Mathematically, this decision problem can be stated as a hypothesis testing problem. With this biometric recognition problem formulated within the frameworks of signal processing and statistical decision theory, the performance of the system is evaluated by constructing a Receiver Operating Characteristics (ROC) curve. ROC is defined as a plot of the False Reject Rate (FRR), which denotes the measure of likelihood that the system will wrongly reject access to an authorized user, vs. the False Accept Rate (FAR), which indicates the probability that the candidate will be safely accepted when it should have been rejected, parameterized by a decision rule threshold.

1.3 Contribution of the thesis

Since large databases are not available in practice, biometric systems are often tested on a small amount of data. To extrapolate the performance of a small database, we setup the framework for comprehensive analysis of the performance of a large scale iris-based recognition system. This work is directed towards solving the scaling problem to bound performance of large scale iris systems. We propose a methodology to predict performance of a large-scale iris system based on a small testing database available using information theoretic approach.

At this stage, we consider a practical setting where matching scores in the form of HDs are accessible for collecting the data. We assume that multiple scans from the same iris are available. We model these matching scores, a sequence of HDs, as realizations of a Gaussian random process with a number of unknown parameters. These unknown parameters are estimated empirically.

The problems of verification and identification are further stated as a binary and $(M+1)$ -ary hypothesis testing, respectively. Here M is the individual number of iris classes to be identified and an additional ‘Imposter’ hypothesis. We design two decision test statistics for the matching scores based on the multiple templates availability assumption. The proposed models are then applied to predict the performance of a large-scale iris based identification systems from a small amount of available data. We use empirical approach, Chernoff bound, and Large Deviations asymptotic approach to predict the performance of the iris system.

1.4 Organization of the thesis

In Chapter 2, we describe the proposed model for HDs. Chapter 3 explains the proposed decision rules for the matching scores. Chapter 4 focuses on the performance analysis of a large scale iris recognition system. The numerical results for iris-based verification and identification systems are presented in Chapter 5. Conclusions and future work are provided in Chapter 6 and Chapter 7.

Chapter 2

Proposed Matching Score Model

In Daugman’s iris-based recognition system described in Chapter 1, only a single image/template is assumed available to characterize an iris class. The IrisCode of a candidate image is compared against the claimed IrisCode from the database by means of calculating the average HD. The problem of deciding if an input iris image belongs to the claimed identity is often stated as a hypothesis testing problem, where the HD, d , plays the role of a test statistic. Given a threshold γ , the Imposter hypothesis is accepted if $d \geq \gamma$. Otherwise, the Imposter hypothesis is rejected. In the following sections, we propose a model for HDs, state a set of simplifying assumptions, and propose decision rules based on the described model for the data in Chapter 3.

2.1 Proposed Model for the Matching Scores

Iris images are first preprocessed and encoded using our Gabor filter based encoding technique (our interpretation of J. Daugman’s algorithm), to get the IrisCodes. The steps carried out can be briefly summarized as follows.

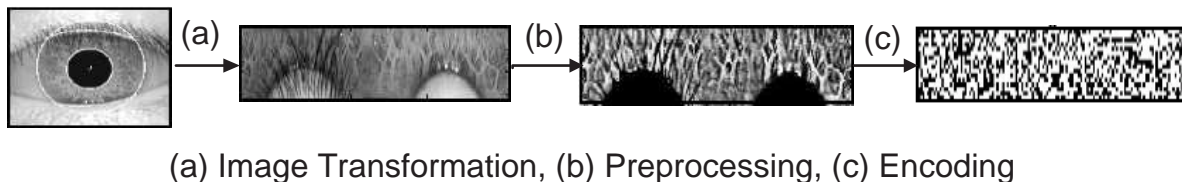


Fig. 2.1: Our interpretation of J. Daugman’s Iris Recognition System

1. An incoming image is first enhanced and transformed into a pseudo polar representation.

2. It is encoded using Gabor filters.
3. The normalized and encoded image is quantized component-by-component to two levels (zero/one) based on the sign of the corresponding filtered image entry. The quantized sequence of components form IrisCode.

The IrisCode of the candidate to be recognized is then compared with a claimed template in the database to get matching scores in the form of HD. In the following, we consider two scenarios of verification and identification separately.

2.1.1 Verification Case

Verification is a one-to-one matching procedure. A verification problem can also be stated as a binary hypothesis testing problem, where only two hypotheses (Genuine/Imposter) are considered. In Daugman's iris verification system the test/candidate is declared to be Imposter if the matching score $d \geq \gamma$. Otherwise the Imposter hypothesis is rejected.

Modern cameras are capable of capturing more than a single snapshot over a small amount of time during adjustment period (approx. 5 to 10 images over 2 sec. period). In this work we propose to use all collected data of a good quality. For our analysis we make the following practically feasible assumptions:

1. Each individual iris class in the database and the incoming candidate user are represented by K iris scans converted into IrisCodes. The practical value of K can be $2 \leq K \leq 10$. However, for the purpose of analysis we assume that K can become large.
2. The matching scores in the form of HD are calculated for arbitrary cross-coupled sets of K IrisCodes such that no same IrisCode is involved twice. This requirement reduces additional dependencies among the matching scores.

Modeling Matching Scores

By Daugman [5], two arbitrary selected IrisCodes are strongly correlated and so are the HDs. We model K - dimensional vectors of HDs as realizations of K - dimensional Gaussian random vectors with correlated entries. Under Imposter hypothesis, H_0 , the vector \mathbf{d} is the Gaussian

distributed with common unknown mean for all entries μ_0 and unknown covariance matrix \mathbf{R}_0 . Under Genuine hypothesis, the vector \mathbf{d} is the Gaussian distributed with common unknown mean for all entries μ_1 and unknown covariance matrix \mathbf{R}_1 (\mathbf{R}_0 and \mathbf{R}_1 are distinct). Mathematically,

$$\begin{aligned} \text{Under Imposter Hypothesis: } \mathbf{d} &\sim N(\mu_0 \cdot \mathbf{1}, \mathbf{R}_0) \\ \text{Under Genuine Hypothesis: } \mathbf{d} &\sim N(\mu_1 \cdot \mathbf{1}, \mathbf{R}_1) \end{aligned} \quad (2.1)$$

We assume that the matrices \mathbf{R}_0 and \mathbf{R}_1 are cyclic,

$$\mathbf{R}_i = \sigma_i^2 \cdot \begin{bmatrix} 1 & \rho_i & \rho_i & \dots & \rho_i \\ \rho_i & 1 & \rho_i & \dots & \rho_i \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_i & \rho_i & \rho_i & \dots & 1 \end{bmatrix}_{K \times K}, \quad i = 0, 1 \quad (2.2)$$

Since the parameters of the models, (μ_i , σ_i^2 , and ρ_i , $i = 0, 1$) are unknown, we estimate them empirically using available data.

Parameter Estimation

We assume that N independent copies of the K -dimensional vector \mathbf{d} are available under both Genuine and Imposter hypotheses. Using this training data, we apply the Maximum Likelihood estimation (MLE) method to estimate the parameters of the model [6]. The ML estimates are given by,

$$\hat{\mu}_i = \frac{1}{KN} \sum_{l=1}^N \sum_{k=1}^K d_{k,l}, \quad \hat{\sigma}_i^2 = \frac{tr(\mathbf{A}_i)}{KN} \quad (2.3)$$

and

$$\hat{\rho}_i = \frac{\mathbf{1}^T \mathbf{A}_i \mathbf{1} - tr(\mathbf{A}_i)}{(K-1)tr(\mathbf{A}_i)}, \quad i = 0, 1 \quad (2.4)$$

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

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

Model Validation

To validate the proposed model for HDs, we apply a multivariate Shapiro-Wilk test for normality [7]. The test provides the p-value statistic which denotes the goodness of fit for the

data to the model. The null hypothesis used by the test is that the data comes from the normal distribution. When p-value falls above the critical p-value (p_{crit})= 0.05, then the null hypothesis is accepted, and the matching scores are declared to have normal distribution. Otherwise the null hypothesis is rejected.

2.1.2 Identification Case

Biometric identification is a process of determining person's identity by performing matches against multiple templates in the database. These systems are designed to determine candidate's identity based solely on biometric information. Identification problem is one to many matching procedure and can be stated as (M+1)-ary hypothesis testing problem. Here M is the number of individual iris classes to be identified and the additional Imposter hypothesis.

Suppose that the IrisCodes from M individual irises, each of length n, are collected and stored in the database. Similar to the verification case, assume that K distinct copies of the IrisCode are available from each iris template in the database. Denote by $\mathbf{X}(l)$, $l = 1, \dots, M$, and \mathbf{Y} random vectors, underlying the IrisCodes of the l^{th} individual and a candidate that submits his/her iris for identification, respectively. Assume that a candidate is also represented by K IrisCodes. Similar to the verification case, the K-dimensional vectors $\mathbf{d}(l)$ are formed by arbitrary coupling K realizations of $\mathbf{X}(l)$ with K realizations of \mathbf{Y} and calculating corresponding K normalized Hamming distances. These individual K-dimensional vectors of the normalized HDs are then concatenated to form a long (M×K)-dimensional vector \mathbf{d} . Identification problem is a multi-hypothesis testing problem. Hence we extend the model used in verification case (2.1) to get the matching score model for identification case.

If the candidate signature is assumed to have come from the i^{th} iris class, then the i^{th} vector component $\mathbf{d}(i)$ in the concatenated vector \mathbf{d} is the HD between two signatures containing the IrisCode of the same i^{th} iris-class and thus can be modeled as drawn from the Genuine distribution. The remaining vector components $\mathbf{d}(j)$, $i \neq j$, $j = 1, \dots, M$, are the HDs between distinct IrisCodes and thus can be modeled as drawn from the Imposter distribution. The IrisCodes forming $\mathbf{d}(i)$ and $\mathbf{d}(j)$ are correlated, partially due to the nature of the random process underlying iris formation and partially due to the operation of filtering during encoding of iris

images [5].

Hence under H_i (i^{th} user hypothesis), we assume that the vector \mathbf{d} is Gaussian distributed with the mean vector \mathbf{m}_i and covariance matrix Σ_i . The mean vector \mathbf{m}_i is a vector column of size MK given by following

$$\mathbf{m}_i = \left[\begin{array}{cccc} \mu_0 \cdot \mathbf{1} & \cdots & \mu_1 \cdot \mathbf{1} & \cdots & \mu_0 \cdot \mathbf{1} \end{array} \right]_{MK \times 1}^T \quad (2.5)$$

Note that the vector has entry $\mu_1 \cdot \mathbf{1}$ on the i^{th} position, $\mathbf{1}$ is K - dimensional vector column of all ones.

The block cyclic covariance matrix Σ_i is given by

$$\Sigma_i = \left[\begin{array}{cccccccc} \mathbf{R}_0 & \tilde{\mathbf{R}}_0 & \cdots & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \cdots & \tilde{\mathbf{R}}_0 \\ \tilde{\mathbf{R}}_0 & \mathbf{R}_0 & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \cdots & \tilde{\mathbf{R}}_0 \\ \tilde{\mathbf{R}}_0 & \cdots & \tilde{\mathbf{R}}_0 & \mathbf{R}_0 & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \cdots & \tilde{\mathbf{R}}_0 \\ \tilde{\mathbf{R}}_0 & \cdots & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \mathbf{R}_1 & \tilde{\mathbf{R}}_0 & \cdots & \tilde{\mathbf{R}}_0 \\ \tilde{\mathbf{R}}_0 & \cdots & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \mathbf{R}_0 & \cdots & \tilde{\mathbf{R}}_0 \\ \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{R}}_0 & \cdots & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \cdots & \mathbf{R}_0 \end{array} \right]_{MK \times MK} \quad (2.6)$$

where the block \mathbf{R}_1 stands on the intersection of the i^{th} block-column and the i^{th} block-row. The diagonal is composed of blocks \mathbf{R}_0 besides the block on the $(i, i)^{th}$ position. The rest of the matrix Σ_i is composed of $\tilde{\mathbf{R}}_0$, where $\tilde{\mathbf{R}}_0 = \sigma_0^2 \rho_0 \cdot (\mathbf{1} \times \mathbf{1}^T)$. The matrices \mathbf{R}_0 and \mathbf{R}_1 have the form given in (2.2).

Similarly, under H_0 (Imposter hypothesis), the vector \mathbf{d} is modeled to be Gaussian distributed with the mean vector \mathbf{m}_0 and covariance matrix Σ_0 . The mean vector is given as

$$\mathbf{m}_0 = \left[\begin{array}{cccc} \mu_0 \cdot \mathbf{1} & \cdots & \mu_0 \cdot \mathbf{1} & \cdots & \mu_0 \cdot \mathbf{1} \end{array} \right]_{MK \times 1}^T \quad (2.7)$$

with dependence only on μ_0 . The block cyclic covariance matrix Σ_0 consists of all the diagonal

blocks \mathbf{R}_0 and remaining block elements $\tilde{\mathbf{R}}_0$. Σ_0 is given by

$$\Sigma_0 = \begin{bmatrix} \mathbf{R}_0 & \tilde{\mathbf{R}}_0 & \dots & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \dots & \tilde{\mathbf{R}}_0 \\ \tilde{\mathbf{R}}_0 & \mathbf{R}_0 & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \ddots & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \dots & \tilde{\mathbf{R}}_0 \\ \tilde{\mathbf{R}}_0 & \dots & \tilde{\mathbf{R}}_0 & \mathbf{R}_0 & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \dots & \tilde{\mathbf{R}}_0 \\ \tilde{\mathbf{R}}_0 & \dots & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \mathbf{R}_0 & \tilde{\mathbf{R}}_0 & \dots & \tilde{\mathbf{R}}_0 \\ \tilde{\mathbf{R}}_0 & \dots & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \mathbf{R}_0 & \dots & \tilde{\mathbf{R}}_0 \\ \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{R}}_0 & \dots & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \tilde{\mathbf{R}}_0 & \dots & \mathbf{R}_0 \end{bmatrix}_{MK \times MK} \quad (2.8)$$

where \mathbf{R}_0 has the same form as in (2.2) and $\tilde{\mathbf{R}}_0 = \sigma_0^2 \rho_0 \cdot (\mathbf{1} \times \mathbf{1}^T)$.

2.2 Summary

To summarize, the proposed models for the matching scores in Verification and Identification cases are described as follows.

In Verification mode, the binary hypothesis testing problem is stated as

$$\begin{aligned} \text{Under the Imposter Hypothesis: } \mathbf{d} &\sim N(\hat{\mu}_0 \cdot \mathbf{1}, \hat{\mathbf{R}}_0) \\ \text{Under the Genuine Hypothesis: } \mathbf{d} &\sim N(\hat{\mu}_1 \cdot \mathbf{1}, \hat{\mathbf{R}}_1) \end{aligned} \quad (2.9)$$

where $\hat{\mathbf{R}}_0$ and $\hat{\mathbf{R}}_1$ have the form as in (2.2) with estimated parameters from (2.3), (2.4) substituted in it.

In Identification mode,

$$\begin{aligned} \text{Under the Imposter Hypothesis: } \mathbf{d} &\sim N(\hat{\mathbf{m}}_0, \hat{\Sigma}_0) \\ \text{Under } i^{\text{th}} \text{ Genuine user Hypothesis: } \mathbf{d} &\sim N(\hat{\mathbf{m}}_i, \hat{\Sigma}_i), \quad i = 1, \dots, M \end{aligned} \quad (2.10)$$

where $\hat{\mathbf{m}}_0$ and $\hat{\mathbf{m}}_i$ have the form as in (2.7) and (2.5), respectively. And $\hat{\Sigma}_0$ and $\hat{\Sigma}_i$ have the same form as in (2.8) and (2.6) respectively, with estimated parameters from (2.3), (2.4) substituted in it.

Chapter 3

Proposed Decision Rules

In this work, we assume that each individual iris class in the database and the incoming candidate user are represented by K iris scans converted into IrisCodes. Also the matching scores in the form of HDs are calculated for arbitrary cross-coupled sets of K independent IrisCodes. Based on these practically feasible assumptions and the proposed models for the matching scores, we design two test statistics.

1. The Average HD
2. Plug-in Log Likelihood Ratio

We describe the design of each of the test statistics in detail.

3.1 Proposed Rule 1:

The Average Hamming Distance (D-Bar)

Let $\mathbf{d} = [d_1, d_2, \dots, d_K]$ be a vector of K HDs formed according to the assumptions above. The first decision rule averages K HDs, that is,

$$\bar{d} = \frac{1}{K} \sum_{k=1}^K d_k \quad (3.1)$$

To make a decision about the origin of the data, \bar{d} is compared with a decision threshold γ , a design parameter. This leads to the following decision rule:

Imposter hypothesis is accepted if, $\bar{d} \geq \gamma$.
Otherwise the Imposter hypothesis is rejected.

The average HD is an intuitive test statistic. In signal processing, averaging is often used to reduce the noise and thus to improve performance.

3.2 Proposed Rule 2: Plug-in Log Likelihood Ratio

In this subsection, we design a plug-in log-likelihood ratio test. In ideal case, when the distribution of the data is completely specified, log-likelihood ratio is the optimal test statistic in the minimum $P(\text{error})$ or Neyman-Pearson sense. If the distribution of the data are not completely specified, the unknown parameters of the distribution are estimated from the available data. When the estimated parameters are substituted in the log-likelihood ratio in place of true unknown parameters, the test becomes suboptimal, but still close to optimal. This is the main motivation behind using log-likelihood ratio as a decision test statistic.

3.2.1 Verification Case

The plug-in log-likelihood ratio for the proposed model (2.9) is given by,

$$\begin{aligned}\Lambda_K &= \frac{1}{K} \log \frac{\hat{p}(\mathbf{d}|GH)}{\hat{p}(\mathbf{d}|IH)} \\ &= -\frac{1}{2K} (\mathbf{d} - \hat{\mu}_1 \mathbf{1})^T \hat{\mathbf{R}}_1^{-1} (\mathbf{d} - \hat{\mu}_1 \mathbf{1}) + \frac{1}{2K} (\mathbf{d} - \hat{\mu}_0 \mathbf{1})^T \hat{\mathbf{R}}_0^{-1} (\mathbf{d} - \hat{\mu}_0 \mathbf{1}) \\ &\quad - \frac{1}{2K} \log \det (\hat{\mathbf{R}}_1 \hat{\mathbf{R}}_0^{-1})\end{aligned}\quad (3.2)$$

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

To make a decision about the origin of the data, Λ_K is compared with a threshold $\gamma = \frac{1}{K} \log \frac{\pi_1}{\pi_0}$, where π_1 and π_0 are prior probabilities of Genuine and Imposter hypothesis respectively. This leads to the following decision rule:

$$\begin{aligned}\text{User is declared as Imposter if, } & \Lambda_K \leq \gamma, \\ \text{User is declared as Genuine if, } & \Lambda_K > \gamma.\end{aligned}$$

3.2.2 Identification Case

The verification case described above can be easily extended to the case of identification. In this work, identification problem is stated as (M+1)-ary hypothesis testing problem. Here M is

the number of individual iris classes to be identified. Let H_i and H_0 be the i^{th} Genuine user and Imposter hypotheses, respectively. The plug-in log likelihood test statistic in the identification mode is a vector of log-likelihood ratios given by

$$\mathbf{\Lambda}_K = \left[\Lambda_K(1), \quad \Lambda_K(2), \quad \dots, \quad \Lambda_K(M) \right]_{M \times 1}^T, \quad (3.3)$$

with the i^{th} entry given by,

$$\Lambda_K(i) = \frac{1}{K} \log \frac{\hat{p}(\mathbf{d}|H_i)}{\hat{p}(\mathbf{d}|H_0)}.$$

To make a decision about the origin of the candidate IrisCodes, each log-likelihood ratio in (3.3) is compared against a common threshold γ .

The candidate IrisCode is recognized as being from the i^{th} iris class if

$$\Lambda_K(i) \geq \Lambda_K(j) \text{ and } \Lambda_K(i) > \gamma, \quad \forall j = 1, \dots, M, j \neq i.$$

The IrisCode is rejected (not recognized as being from any iris class in the database) if all components in (3.3) satisfy

$$\Lambda_K(i) \leq \gamma, \quad \forall i = 1, \dots, M.$$

Chapter 4

Performance Analysis

4.1 Background

Performance evaluation is an important step in designing a decision making system. In the selection of biometrics for use in a recognition system and in the subsequent design of the system, the predicted performance is a key consideration. 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 very few results on performance analysis of iris-based recognition systems [2], [5], [8]. These papers perform analysis under considerably simplified conditions and mostly focus on evaluation of FAR. In this work, we avoid simplifying the model and consider the average probability of error as a measure of performance taking both FAR and FRR into consideration. Since the expressions for the FAR and FRR are often hard to evaluate directly, one can appeal to bounds and approximations. In this work, we use Chernoff Bound that is related to a more restrictive asymptotic approach called Large Deviations [9], [10], [11] to predict the performance of a large scale iris-based recognition system. We consider performance evaluation for verification and identification problem separately.

4.2 Verification Case

Let H_0 and H_1 be the Imposter and Genuine hypotheses, respectively and π_0 and π_1 be their respective prior probabilities, where $\pi_0 + \pi_1 = 1$. The total average probability of error in

verification case can be written as

$$P(\text{error}) = \pi_1 \cdot P(\text{Decide } H_0 | H_1 \text{ is true}) + \pi_0 \cdot P(\text{Decide } H_1 | H_0 \text{ is true}) \quad (4.1)$$

4.2.1 Empirical Evaluation

Using the proposed decision test statistics D-bar (3.1) and Log Likelihood ratio (3.2), we empirically find the probability of error. We evaluate the performance for the following data sets:

1. Simulated data,
2. Bootstrap data.

In the first case, the matching scores are simulated using the proposed model for matching scores (2.9) under both Genuine and Imposter hypothesis. According to the D-bar test statistic, given any threshold γ ,

Imposter hypothesis is accepted if, $\bar{d} \geq \gamma$.

Otherwise the Imposter hypothesis is rejected.

Similarly, according to the log-likelihood test statistic, for the threshold $\gamma = \frac{1}{K} \log \frac{\pi_1}{\pi_0}$,

User is declared as Imposter if, $\Lambda_K \leq \gamma$,

User is accepted as Genuine if, $\Lambda_K > \gamma$.

where π_1 and π_0 are prior probabilities on the Genuine and Imposter hypotheses, respectively. We empirically calculate FAR and FRR for the two decision rules and plot ROC curves as a combination of FAR and FRR. We also evaluate the total probability of error using (4.1).

The bootstrap data in the second case are the data drawn randomly with replacement from a real database [12]. Using the same D-bar and log-likelihood decision rules, we empirically calculate the FAR and FRR for a given range of thresholds and plot a set of ROC curves.

4.2.2 Chernoff Bound

The Chernoff bound is known as a tight upper bound on the probability of error [13] and can be found for arbitrary selected parameters K and a threshold γ . The FAR and FRR can be

upper bounded as

$$\begin{aligned} FAR(\gamma) &= P(\Lambda_K > \gamma | IH) \leq e^{-K \cdot I_0(K, \gamma)}, \\ FRR(\gamma) &= P(\Lambda_K \leq \gamma | GH) \leq e^{-K \cdot I_1(K, \gamma)} \end{aligned} \quad (4.2)$$

where $I_0(K, \gamma)$, $I_1(K, \gamma)$ are the Chernoff rate functions under the Imposter and Genuine hypotheses, respectively. Note that in this case, the rate functions are functions of both K and γ . The Chernoff rate functions under Imposter hypothesis is given by

$$I_0(K, \gamma) = \sup_s [s\gamma - \bar{\varphi}_0(s, K)] \quad (4.3)$$

where $\bar{\varphi}_0(s, K)$ is the log-moment generating function of the test statistic in (3.2) under the Imposter hypothesis (see appendix A for details), defined as

$$\begin{aligned} \bar{\varphi}_0(s, K) &= \frac{1}{K} \log (E_0 [e^{sK\Lambda_K}]) \\ &= \frac{(\hat{\mu}_1 - \hat{\mu}_0)^2}{2K} \cdot \frac{1}{\lambda_1(1)} \cdot \left[\frac{s^2 \lambda_0(1)}{s\lambda_0(1) - (s-1)\lambda_1(1)} - s \right] \\ &\quad - \frac{s}{2K} \cdot \left[\log \frac{\lambda_1(1)}{\lambda_0(1)} + (K-1) \log \frac{\lambda_1(2)}{\lambda_0(2)} \right] \\ &\quad + \frac{1}{2K} \cdot \left[\log \frac{\lambda_1(1)}{s\lambda_0(1) - (s-1)\lambda_1(1)} + (K-1) \log \frac{\lambda_1(2)}{s\lambda_0(2) - (s-1)\lambda_1(2)} \right], \text{ where} \end{aligned} \quad (4.4)$$

$$\lambda_0(1) = \hat{\sigma}_0^2 + (K-1)\hat{\sigma}_0^2\hat{\rho}_0$$

$$\lambda_0(2) = \hat{\sigma}_0^2 - \hat{\sigma}_0^2\hat{\rho}_0$$

$$\lambda_1(1) = \hat{\sigma}_1^2 + (K-1)\hat{\sigma}_1^2\hat{\rho}_1$$

$$\lambda_1(2) = \hat{\sigma}_1^2 - \hat{\sigma}_1^2\hat{\rho}_1$$

(4.5)

are the eigenvalues of the covariance matrices \mathbf{R}_0 and \mathbf{R}_1

And the threshold γ that gives the optimum solution for Imposter rate function is

$$\begin{aligned} \gamma &= \frac{(\hat{\mu}_1 - \hat{\mu}_0)^2}{2K} \cdot \frac{1}{\lambda_1(1)} \cdot \left[\frac{s^2 \lambda_0(1)^2 + 2s\lambda_0(1)\lambda_1(1) - s^2 \lambda_0(1)\lambda_1(1)}{[s\lambda_0(1) - (s-1)\lambda_1(1)]^2} - 1 \right] \\ &\quad - \frac{1}{2K} \log \frac{\lambda_1(1)}{\lambda_0(1)} - \frac{(K-1)}{2K} \log \frac{\lambda_1(2)}{\lambda_0(2)} \\ &\quad - \frac{1}{2} \cdot \frac{1}{s\lambda_0(1) - (s-1)\lambda_1(1)} \cdot \left[\frac{\lambda_0(1) - \lambda_1(1)}{K} \right] \\ &\quad - \frac{(K-1)}{2} \cdot \frac{1}{s\lambda_0(2) - (s-1)\lambda_1(2)} \cdot \left[\frac{\lambda_0(2) - \lambda_1(2)}{K} \right] \end{aligned} \quad (4.6)$$

Similarly, the Chernoff rate function under the Genuine hypothesis is given by

$$I_1(K, \gamma) = \sup_s [s\gamma - \bar{\varphi}_1(s, K)] \quad (4.7)$$

where $\bar{\varphi}_1(s, K)$ is the log moment generating function under Genuine hypothesis defined as

$$\begin{aligned} \bar{\varphi}_1(s, K) &= \frac{1}{K} \log (E_1 [e^{sK\Lambda_K}]) \\ &= \frac{(\hat{\mu}_1 - \hat{\mu}_0)^2}{2K} \cdot \frac{1}{\lambda_0(1)} \cdot \left[\frac{s^2 \lambda_1(1)}{(s+1)\lambda_0(1) - s\lambda_1(1)} + s \right] \\ &\quad - \frac{s}{2K} \cdot \left[\log \frac{\lambda_1(1)}{\lambda_0(1)} + (K-1) \log \frac{\lambda_1(2)}{\lambda_0(2)} \right] \\ &\quad + \frac{1}{2K} \cdot \left[\log \frac{\lambda_0(1)}{(s+1)\lambda_0(1) - s\lambda_1(1)} + (K-1) \log \frac{\lambda_0(2)}{(s+1)\lambda_0(2) - s\lambda_1(2)} \right] \end{aligned} \quad (4.8)$$

The threshold γ that gives the optimum solution for Genuine rate function is given by

$$\begin{aligned} \gamma &= \frac{(\hat{\mu}_1 - \hat{\mu}_0)^2}{2\lambda_0(1)K} \cdot \left[\frac{s^2 \lambda_0(1)\lambda_1(1) + 2s\lambda_0(1)\lambda_1(1) - s^2 \lambda_1(1)^2}{[(s+1)\lambda_0(1) - s\lambda_1(1)]^2} + 1 \right] \\ &\quad - \frac{1}{2K} \log \frac{\lambda_1(1)}{\lambda_0(1)} - \frac{(K-1)}{2K} \log \frac{\lambda_1(2)}{\lambda_0(2)} \\ &\quad - \frac{1}{2} \frac{1}{(s+1)\lambda_0(1) - s\lambda_1(1)} \left[\frac{\lambda_0(1) - \lambda_1(1)}{K} \right] \\ &\quad - \frac{(K-1)}{2} \frac{1}{(s+1)\lambda_0(2) - s\lambda_1(2)} \left[\frac{\lambda_0(2) - \lambda_1(2)}{K} \right] \end{aligned}$$

4.2.3 Large Deviations Approximation

Using the theory of large deviations, the average probability of error in (4.1) can be asymptotically approximated by a linear combination of two exponential functions with explicit dependence on K . If the Large deviations conditions are satisfied [10], the FAR and FRR can be approximated as follows

$$\begin{aligned} FAR(\gamma) &= P(\Lambda_K > \gamma | IH) \approx G(K, \gamma) e^{-K \cdot I_0(\gamma)}, \\ FRR(\gamma) &= P(\Lambda_K \leq \gamma | GH) \approx G(K, \gamma) e^{-K \cdot I_1(\gamma)} \end{aligned} \quad (4.9)$$

where $G(K, \gamma)$ is a slowly varying function of K and γ (often omitted in analysis) and $I_0(\gamma)$, $I_1(\gamma)$ are the Large Deviations rate functions under the Imposter and Genuine hypotheses, respectively. As $K \rightarrow +\infty$, Λ_K converges to its asymptotic expected value under the two hypotheses. The rate of convergence in our case is exponential under either hypothesis, with exponent determined by the LD rate function [9], [10], [11]. Using LD, we calculate the worst exponent once.

Then we can predict performance of the system for an arbitrary K without recomputing the expression in the exponent for every new value of K .

The large deviations rate function is defined as

$$I_0(\gamma) = \sup_s [s\gamma - \bar{\varphi}_0(s)], \quad (4.10)$$

where $\bar{\varphi}_0(s)$ is the asymptotic normalized log-moment generating function under the Imposter hypothesis given by

$$\bar{\varphi}_0(s) = \lim_{K \rightarrow +\infty} \frac{1}{K} \log (E_0 [e^{sK\Lambda_K}]). \quad (4.11)$$

Substituting expression for log-likelihood function (3.2) into (4.11) and evaluating the expectation and the limit (see Appendix A for details), we obtain

$$\bar{\varphi}_0(s) = -\frac{s-1}{2} \log \frac{\lambda_1(2)}{\lambda_0(2)} + \frac{1}{2} \log \left[\frac{\lambda_0(2)}{s\lambda_0(2) - (s-1)\lambda_1(2)} \right] \quad (4.12)$$

And the threshold γ which gives optimum solution for Imposter rate function is given by,

$$\gamma = -\frac{1}{2} \log \frac{\lambda_1(2)}{\lambda_0(2)} - \frac{1}{2} \left[\frac{\lambda_0(2) - \lambda_1(2)}{s\lambda_0(2) - (s-1)\lambda_1(2)} \right] \quad (4.13)$$

The parameters $\hat{\sigma}_i^2$ and $\hat{\rho}_i$, $i = 0, 1$ are estimated by (2.3), and (2.4).

Similarly, the rate function under Genuine hypothesis is given by

$$I_1(\gamma) = \sup_s [s\gamma - \bar{\varphi}_1(s)], \text{ where} \quad (4.14)$$

$\bar{\varphi}_1(s)$ is the asymptotic normalized log-moment generating function defined as

$$\begin{aligned} \bar{\varphi}_1(s) &= \lim_{K \rightarrow +\infty} \frac{1}{K} \log (E_1 [e^{sK\Lambda_K}]) \\ &= -\frac{s}{2} \log \frac{\lambda_1(2)}{\lambda_0(2)} + \frac{1}{2} \log \left[\frac{\lambda_0(2)}{(s+1)\lambda_0(2) - s\lambda_1(2)} \right] \end{aligned} \quad (4.15)$$

The threshold γ that provides optimum solution for Genuine rate function and the parameters $\lambda_i(j)$, $i = 0, 1, j = 1, 2$ for Genuine hypothesis are same as in (4.13) and (4.5), respectively.

Note that the Large Deviation rate functions are not functions of K .

Therefore given the number of templates per iris class (K), for each value of threshold (γ), the conditional probabilities FAR and FRR can be upper bounded and approximated using (4.2) and (4.9), respectively.

4.3 Identification Case

In the iris-based identification system, the template of a candidate user is compared with templates of all users in the database to find the best match. If the template does not match with any user in the database, it is declared as Imposter. Thus the identification problem is a multi-hypothesis testing problem. We use the following two bounds to analyze the performance of an iris-based identification system.

1. The worst case upper bound on the total probability of error
2. Union of Chernoff Bounds

4.3.1 The Worst Case Upper Bound

Let M be the total number of iris classes. Denote by H_0 and H_i , $i = 1, \dots, M$ an Imposter and the i^{th} user hypothesis, respectively. Denote by π_i , $i = 0, 1, \dots, M$ their corresponding prior probabilities, where $\sum_{i=0}^M \pi_i = 1$. Then the total average probability of error in the system is given by

$$P(\text{error}) = \sum_{k=0}^M \pi_k \sum_{l=0, l \neq k}^M P(\text{decide } H_l \mid H_k \text{ is true}). \quad (4.16)$$

The total probability of error in (4.16) can be upper bounded using conditional probabilities of error for the binary hypothesis testing problem in (4.1). This results in

$$P(\text{error}) \leq \frac{M(M+1)}{2} \max_{k,l=0,\dots,M, k \neq l} \left[\frac{\pi_k}{\pi_k + \pi_l} P(\text{error} \mid H_k^{\text{binary}}) + \frac{\pi_l}{\pi_k + \pi_l} P(\text{error} \mid H_l^{\text{binary}}) \right], \quad (4.17)$$

Details of the derivation can be found in [14]. This bound can be further reduced to

$$P(\text{error}) \leq \frac{M(M+1)}{2} \max_{\alpha, k,l=0,\dots,M, k \neq l} \left[\alpha P(\text{error} \mid H_k^{\text{binary}}) + (1 - \alpha) P(\text{error} \mid H_l^{\text{binary}}) \right], \quad (4.18)$$

where $\alpha = \frac{\pi_k}{\pi_k + \pi_l}$.

Using the verification results, we can further write,

$$P(\text{error}) \leq \frac{M(M+1)}{2} \max_{\alpha} [\alpha P(\text{Decide } H_1 \mid H_0 \text{ is true}) + (1 - \alpha) P(\text{Decide } H_0 \mid H_1 \text{ is true})]. \quad (4.19)$$

This is the worst case upper bound on the total probability of error for identification case.

4.3.2 Union of Chernoff Bounds

To obtain a tighter result for $P(\text{error})$, we use Chernoff bounds on conditional error probabilities. The total probability of error for identification case can be rewritten as follows

$$\begin{aligned}
 P(\text{error}) &= \sum_{l=0}^M \pi_l \cdot P(\text{error}|H_l) \\
 &= \pi_0 \cdot \sum_{l=1}^M P(H_l|H_0) + \sum_{k=1}^M \pi_k \cdot P(H_0|H_k) \\
 &\quad + \sum_{k,l=1, k \neq l}^M \pi_k \cdot P(H_l|H_k),
 \end{aligned} \tag{4.20}$$

where $P(H_l|H_k)$ is the conditional probability that the candidate is decided to have come from H_l hypothesis when in reality H_k hypothesis is true. Each conditional probability in (4.20) has an upper bound determined by the corresponding Chernoff rate function. Furthermore, the exponential rate for the Bayes rule in the $(M+1)$ -ary case is determined by the minimum Chernoff information¹ among all distinct pairs of hypotheses including H_0 [14].

To find the upper bound on each conditional probability $P(H_l|H_k)$ in (4.20), we find chernoff information which is a point of intersection of chernoff rate functions under H_l and H_k hypotheses. This point is represented as a threshold (\mathbf{t}) in $(M+1)$ -dimensional decision region. Let L_0 and L_1 be the asymptotic expected values of the log-likelihood functions under Imposter and Genuine hypotheses, respectively. Denote by $I_0(\mathbf{t})$ and $I_l(\mathbf{t})$, $l = 1, \dots, M$ the Chernoff rate function under the Imposter and the l^{th} hypotheses, respectively. In case of $P(H_l|H_0)$, $l = 1, \dots, M$, the point of intersection of rate functions under the Imposter and the l^{th} hypotheses is given by $\mathbf{t} = (0, L_0, \dots, L_0)$ [14]. Therefore, the $P(H_l|H_0)$ is upper bounded as

$$P(H_l|H_0) \leq e^{-K I_0(0, L_0, \dots, L_0)}, \quad l = 1, \dots, M$$

The conditional probability $P(H_l|H_k)$, $l, k = 1, \dots, M$ is upper bounded using Chernoff rate functions under H_l and H_k hypotheses. The corresponding rate functions $I_l(\mathbf{t})$ and $I_k(\mathbf{t})$ intersect at

¹For any distributions P_1 and P_2 , Chernoff Information is the highest achievable exponent for the probability of error. Mathematically it is defined as

$$C(P_1, P_2) = - \min_{0 \leq \lambda \leq 1} \log \left(\sum_x P_1^\lambda(x) P_2^{1-\lambda}(x) \right)$$

the point $(\frac{L_0+L_1}{2}, \frac{L_0+L_1}{2}, L_0, \dots, L_0)$. Also, the rate functions under $I_k(\mathbf{t})$ and $I_0(\mathbf{t})$ are related as

$$I_k(\mathbf{t}) = I_0(\mathbf{t}) - t_k$$

Therefore the conditional probability $P(H_l|H_k)$ can be upper bounded as

$$P(H_l|H_k) \leq e^{-K I_0(\frac{L_0+L_1}{2}, \frac{L_0+L_1}{2}, L_0, \dots, L_0) - \frac{L_0+L_1}{2}}$$

The upper bound on the probability of error in (4.20) is the union of Chernoff bound on each of the conditional probabilities and is given by

$$\begin{aligned} P(\text{error}) &\leq \pi_0 \sum_{i=1}^M e^{-K \cdot I_0(0, L_0, \dots, L_0)} + (1 - \pi_0) \cdot e^{-K \cdot I_0(0, L_0, \dots, L_0)} \\ &\quad + \sum_{j=1}^M \pi_j \cdot \sum_{i=1, i \neq j}^M e^{-K [I_0(\frac{L_0+L_1}{2}, \frac{L_0+L_1}{2}, L_0, \dots, L_0) - \frac{L_0+L_1}{2}]} \end{aligned} \quad (4.21)$$

Extending the formulation in verification problem, the rate function under the Imposter hypothesis for the identification case is given by

$$I_0(K, \mathbf{t}) = \sup_{\mathbf{s}} [\langle \mathbf{s}, \mathbf{t} \rangle - \bar{\varphi}_0(\mathbf{s}, K)] \quad (4.22)$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product², \mathbf{t} is a vector of thresholds and $\bar{\varphi}_0(\mathbf{s}, K)$ is the log-moment generating function under the Imposter hypothesis defined as

$$\bar{\varphi}_0(\mathbf{s}, K) = \frac{1}{K} \log (E_0 [e^{K \langle \mathbf{s}, \mathbf{\Lambda}_K \rangle}])$$

Here $\mathbf{s} = [s_1, s_2, \dots, s_M]^T$ and $\mathbf{\Lambda}_K = [\Lambda_K(1), \Lambda_K(2), \dots, \Lambda_K(M)]^T$.

Solving the above expectation and normalizing by K (see appendix for details),

$$\begin{aligned} \bar{\varphi}_0(\mathbf{s}, K) &= \frac{1}{2K} (\Delta^T \cdot \mathbf{Q}^{-1} \cdot \Delta) \\ &\quad - \frac{1}{2K} \left[\sum_{i=1}^M s_i (\hat{\mathbf{m}}_i^T \cdot \hat{\Sigma}_i^{-1} \cdot \hat{\mathbf{m}}_i - \hat{\mathbf{m}}_0^T \cdot \hat{\Sigma}_0^{-1} \cdot \hat{\mathbf{m}}_0) + \hat{\mathbf{m}}_0^T \cdot \hat{\Sigma}_0^{-1} \cdot \hat{\mathbf{m}}_0 \right] \\ &\quad - \frac{1}{2K} \left[\sum_{i=1}^M s_i \log \det(\hat{\Sigma}_i \hat{\Sigma}_0^{-1}) + \log \det(\hat{\Sigma}_0 \mathbf{Q}^{-1}) \right] \end{aligned} \quad (4.23)$$

²The Euclidean inner product is defined for Discrete-Time Signals as [15] $\langle \mathbf{w}, \mathbf{v} \rangle = \sum_{i=-\infty}^{\infty} w_i \cdot v_i = \mathbf{w}^T \cdot \mathbf{v}$

where

$$\Delta = \mathbf{Q} \left[\sum_{i=1}^M s_i \left(\hat{\Sigma}_i^{-1} \cdot \hat{\mathbf{m}}_i - \hat{\Sigma}_0^{-1} \cdot \hat{\mathbf{m}}_0 \right) + \hat{\Sigma}_0^{-1} \cdot \hat{\mathbf{m}}_0 \right],$$

$$\mathbf{Q} = \left[\sum_{i=1}^M s_i \left(\hat{\Sigma}_i^{-1} - \hat{\Sigma}_0^{-1} \right) + \hat{\Sigma}_0^{-1} \right]^{-1}$$

$\hat{\mathbf{m}}_i$ and $\hat{\Sigma}_i$, $i = 1, \dots, M$ are the mean vectors and covariance matrices under Genuine hypothesis given by (2.5) and (2.6), respectively with the estimated parameters substituted in them. Similarly, $\hat{\mathbf{m}}_0$ and $\hat{\Sigma}_0$ are the mean vector and covariance matrix under Imposter hypothesis given by (2.7) and (2.8), respectively.

To summarize, for the identification case, given the number of users in the database (M) and the number of templates per user (K), (4.19) and (4.21) provide the worst case upper bound and tight Chernoff upper bound on the total probability of error. For a fixed value of the upper bound on the probability of error and a fixed number of templates per iris class, these expressions give the maximum number of users that the iris recognition system can handle, so that the total error probability of the system does not exceed the fixed value. Similarly, the performance of the recognition system for a given value of the total probability of error and a given number of users in the system can be analyzed by varying the number of copies per iris class.

Chapter 5

Results

All experiments were performed on two datasets:

1. the CASIA database provided by the Chinese Academy of Sciences [16]
2. a special dataset composed of infrared and visible light images collected at WVU.

The CASIA database contains frontal view iris images of 108 iris classes with 6 images per class. The WVU database contains 20 iris classes, 4 images per class. We use images from these datasets to generate IrisCodes and form sets of matching scores of two types:

1. without compensation for rotation, and
2. with compensation for rotation¹.

The sample images from the CASIA database together with the corresponding IrisCodes are shown in Fig. 5.1.

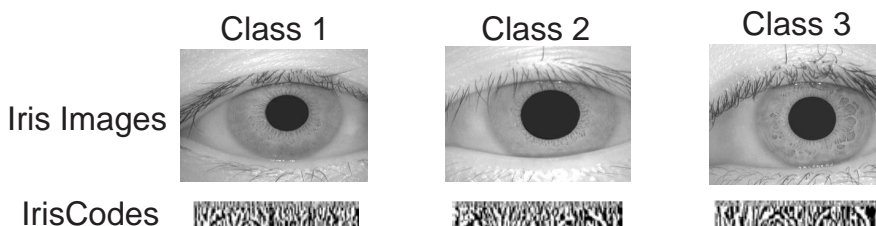


Fig. 5.1: Example Iris Images and corresponding IrisCodes from CASIA Database

Sample images from WVU dataset together with the corresponding IrisCodes are displayed in Fig.5.2.

¹When matching a pair of IrisCodes, one of two IrisCodes was circularly shifted upto 20 pixels (20 degrees)

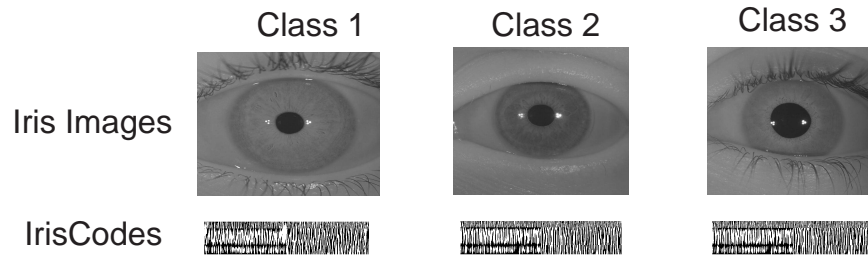


Fig. 5.2: Example Iris Images and corresponding IrisCodes from WVU Database

5.1 Results for CASIA Database

To convert iris images into IrisCodes, we encode images using Gabor filters and quantize the phase of the filtered outputs to two levels. To be more precise, the following steps are implemented:

1. *Localization of region of interest* - The pupil, sclera, and eyelids are segmented.
2. *Normalization* - The localized iris region is transformed from Cartesian co-ordinates to doubly dimensionless polar co-ordinates.
3. *Encoding* - 2D Gabor wavelets are used to encode the image content that is then quantized to two levels based on phase information of the output. The result of encoding step is presented as a binary template called 'IrisCode'.

5.1.1 Generating Matching Scores

Given IrisCodes for all the users and the corresponding templates, we calculate the Genuine and Imposter matching scores. The matching scores are calculated using arbitrary cross coupled sets of IrisCodes such that no same Iriscode is involved twice.

Imposter Matching Scores

Since there are 108 classes in the CASIA database, we compare respective templates of every two users to calculate the Imposter HD as shown in Fig. 5.3. We form 54 vectors of HDs each of size 6. ML estimates for Imposter distribution are then calculated using (2.3), (2.4) with $K = 6$ and $N = 54$.

to the right and to the left. The rotated IrisCode that resulted in the smallest value of the Hamming distance between two IrisCodes is a 'compensated for rotation' IrisCode

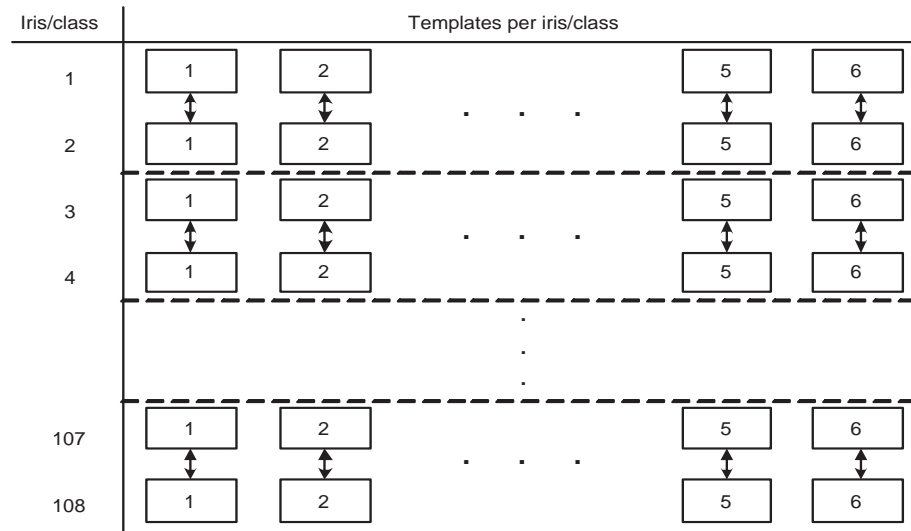


Fig. 5.3: Pairing IrisCodes to get Imposter Matching Scores

Genuine Matching Scores

The Genuine matching scores are calculated in a similar way. For a given class, we pair neighboring templates to obtain a vector of matching scores of length 3. The coupling procedure is as shown in Fig. 5.4.

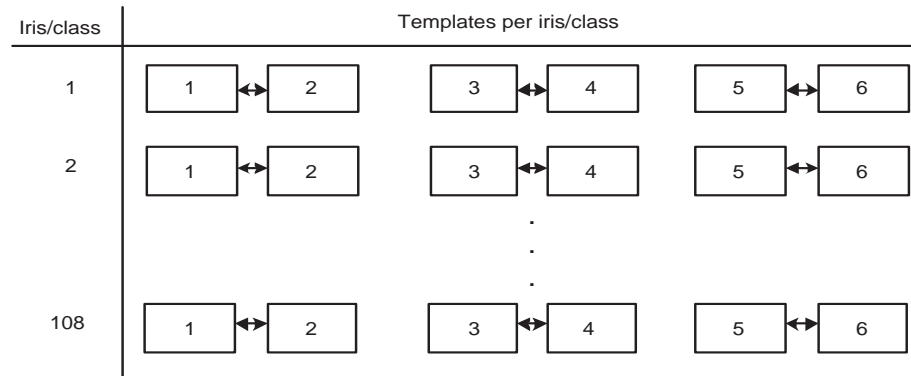


Fig. 5.4: Pairing IrisCodes to get Genuine Matching Scores

This results in 108 vectors of HDs each of size 3. The ML estimates of the parameters for Genuine matching score model are calculated using (2.3), (2.4) with $K = 3$ and $N = 108$.

5.1.2 Maximum Likelihood parameter estimates

The unknown parameters in the proposed Gaussian model for Genuine and Imposter matching scores are estimated using MLE procedure. The values of the ML estimates for parameters $\hat{\mu}_i, \hat{\sigma}_i^2$

and $\hat{\rho}_i$, $i = 0, 1$ under Imposter and Genuine hypothesis are shown in Table 5.1. Note that the

Table 5.1: Maximum Likelihood Parameter Estimates: CASIA Database

(No Rotation Compensation)		(With Rotation Compensation)	
Genuine	Imposter	Genuine	Imposter
$\hat{\mu}_1 = 0.3832$	$\hat{\mu}_0 = 0.4613$	$\hat{\mu}_1 = 0.3236$	$\hat{\mu}_0 = 0.4398$
$\hat{\sigma}_1^2 = 0.004718$	$\hat{\sigma}_0^2 = 2.4505 * 10^{-4}$	$\hat{\sigma}_1^2 = 0.0030$	$\hat{\sigma}_0^2 = 2.1481 * 10^{-4}$
$\hat{\rho}_1 = 0.1126$	$\hat{\rho}_0 = 0.3832$	$\hat{\rho}_1 = 0.1979$	$\hat{\rho}_0 = 0.2888$

mean $\hat{\mu}_i$, $i = 0, 1$ and variance $\hat{\sigma}_i^2$, $i = 0, 1$ for the matching score with rotation compensation is lower than that of with no compensation for rotation.

5.1.3 Shapiro-Wilk normality test results

Shapiro-Wilk test checks the normal distribution assumption about the data by constructing a p-value statistic. The null hypothesis for this test is that the data are normally distributed. For the alpha level of 0.05, if the p-value of the test is less than 0.05, then the null hypothesis that the data are normally distributed is rejected. If the p-value is greater than 0.05, then the null hypothesis has not been rejected. To validate the fit of the Gaussian model with the estimated parameters to the data, we apply Shapiro-Wilk test for normality [7] under Genuine and Imposter hypothesis, separately. Table 5.2 shows the results for p-value statistic for the Shapiro-Wilk normality test.

Table 5.2: p-values for Shapiro-wilk normality test: CASIA Database

No. of Samples	(No Rotation Compensation)		(With Rotation Compensation)	
	Genuine	Imposter	Genuine	Imposter
K=1	0.52	0.46	0.0309	0.3234
K=3	0.2218	0.0722	0.0045	0.3046

The Critical p-value is 0.05. The results indicate that the proposed Gaussian model for the CASIA data with no rotation compensation is not rejected and thus may provide a good fit. The test rejects Gaussian model for genuine scores with compensation for rotation.

5.1.4 Performance Analysis: Verification Case

The performance of the iris based verification system is analyzed numerically using simulated and bootstrapped data. Assuming the Gaussian model described in (2.9), Chernoff bound and

Large Deviations approximation on the total probability of error are obtained.

Empirical Evaluation

Using the proposed decision rules, D-bar (3.1) and Log Likelihood ratio (3.2) with the estimated parameters substituted in them, we empirically calculate the FAR, FRR, and total probability of error. We evaluate the performance using the following testing datasets:

1. Simulated data - the matching scores are simulated using the proposed Gaussian model under the Genuine and Imposter hypotheses.
2. Bootstrap data - the matching scores are randomly selected from the training dataset.

We plot a set of ROC curves. Fig. 5.5 shows the results for $K = 1, 3$ obtained using simulated and bootstrapped data when no compensation for rotation is performed.

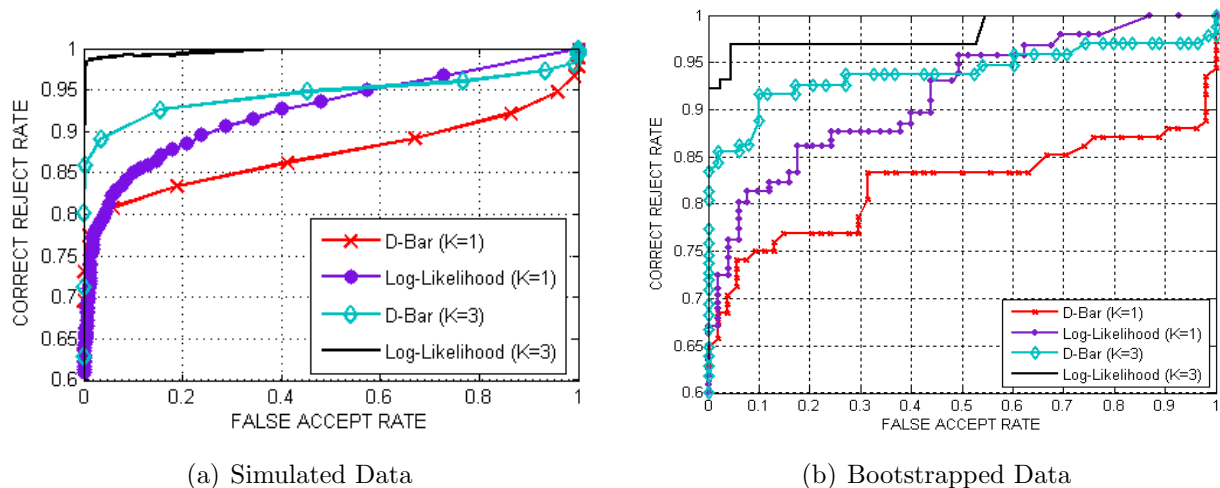


Fig. 5.5: ROC curves obtained using empirical approach (no rotation compensation)

Fig. 5.6 shows the results of empirical evaluation using data compensated for rotation.

The following conclusions can be made from analyzing the results in Fig. 5.5 and 5.6:

1. In most cases, Log-Likelihood decision rule outperforms the D-bar statistic. The exclusion is the case for $K = 1$ when D-bar outperforms Log-likelihood rule in some region.
2. Performance from bootstrapped data is similar to performance from simulated data. This confirms that our model for the matching scores provides a reasonable fit to the data.
3. As the value of K increases, the recognition performance of the system improves.

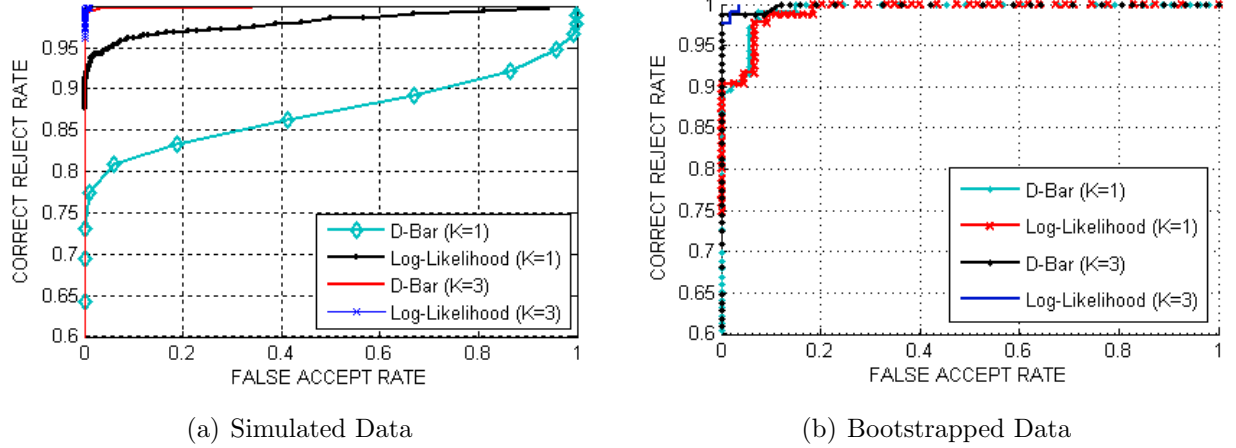


Fig. 5.6: ROC curves obtained using empirical approach (with rotation compensation)

Chernoff Bound and Large Deviations Approximation

The upper bound on FAR and FRR is given by (4.2)

$$FAR(\gamma) = P(\Lambda_K > \gamma | IH) \leq e^{-K \cdot I_0(\gamma, K)},$$

$$FRR(\gamma) = P(\Lambda_K \leq \gamma | GH) \leq e^{-K \cdot I_1(\gamma, K)}$$

where $I_0(K, \gamma)$, $I_1(K, \gamma)$ are the Chernoff rate functions for the bound under the Imposter and Genuine hypotheses, respectively.

Using the theory of Large Deviations, the average probability of error given by (4.1) is asymptotically approximated by an exponential function with explicit dependence on the estimated parameters of the system. Under the Large Deviation approximation, the FAR and FRR can be expressed as

$$FAR(\gamma) = P(\Lambda_K > \gamma | IH) \approx G(K, \gamma) e^{-K \cdot I_0(\gamma)},$$

$$FRR(\gamma) = P(\Lambda_K \leq \gamma | GH) \approx G(K, \gamma) e^{-K \cdot I_1(\gamma)}$$

where $G(K, \gamma)$ is a slowly varying function of K and $I_0(\gamma)$, $I_1(\gamma)$ are the Large Deviations rate functions under the Imposter and Genuine hypotheses, respectively.

The Chernoff bound and Large Deviations approximation rate functions under the Genuine and Imposter hypotheses are plotted in Fig. 5.7.

The value of the rate function at the intersection of the Genuine and Imposter rate functions (encircled point) is called *Chernoff Information*. It is defined as the best achievable exponent for

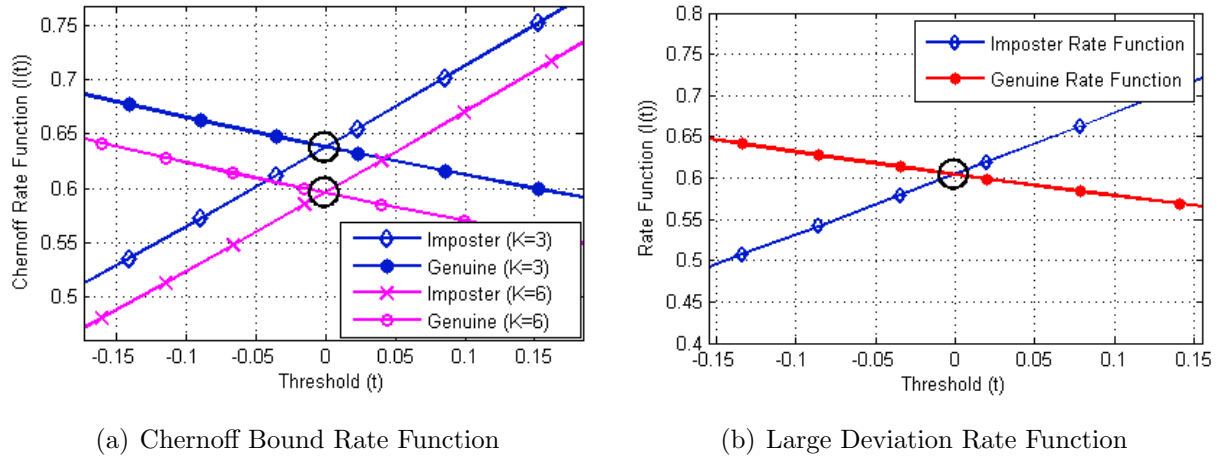


Fig. 5.7: Rate functions for the bound and approximation

the approximation on the total probability of error.

Using these rate functions we find Chernoff bound and the Large Deviations approximations on the FAR and FRR for various values of K and plot them in the form of ROC curves in Fig. 5.8.

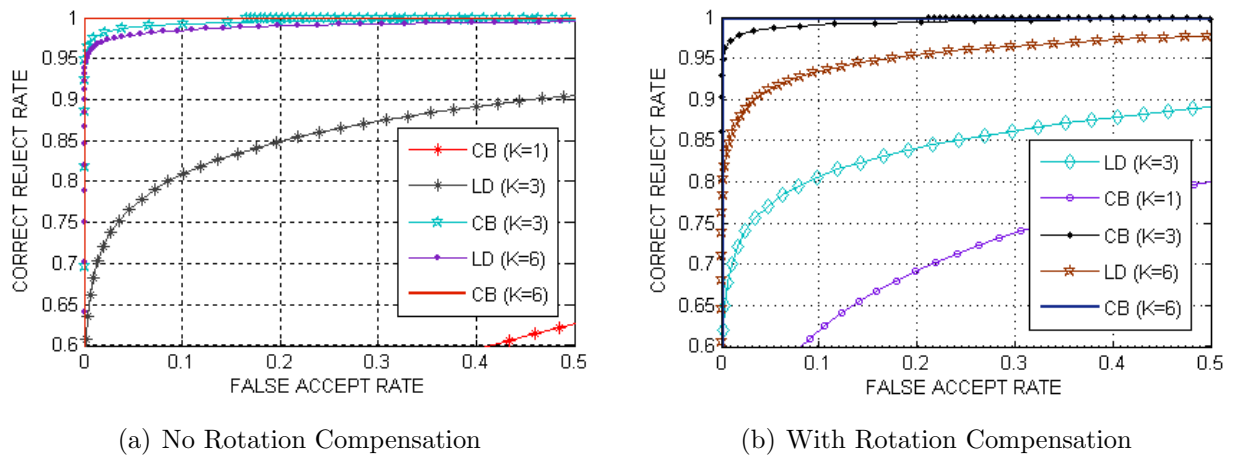


Fig. 5.8: ROC curves using the bound and approximation

The following conclusions can be drawn from the results in Fig.5.8:

1. ROC curves generated from Chernoff Bounds are tight.
2. ROC curves based on the Large Deviations approximation provide a loose fit for a small value of K ($K = 3$) and improve as K increases ($K \geq 6$). This approximation is useful when a quick estimate of an error order has to be obtained.

5.1.5 Performance Analysis: Identification Case

Worst Case Upper Bound

Iris-based identification problem is often stated as an $(M+1)$ -ary hypothesis testing problem. Here M is the total number of individual iris classes in the database. The worst case upper bound on the total probability of error is given by

$$\begin{aligned} P(\text{error}) &\leq \frac{M(M+1)}{2} \max_{k,l=0,\dots,M, k \neq l} \left[\frac{\pi_k}{\pi_k + \pi_l} P(\text{error}|H_k^{\text{binary}}) + \frac{\pi_l}{\pi_k + \pi_l} P(\text{error}|H_l^{\text{binary}}) \right] \\ &= \frac{M(M+1)}{2} \max_{\alpha, k,l=0,\dots,M, k \neq l} \left[\alpha P(\text{error}|H_k^{\text{binary}}) + (1-\alpha) P(\text{error}|H_l^{\text{binary}}) \right], \end{aligned}$$

where $\alpha = \frac{\pi_k}{\pi_k + \pi_l}$.

The results for the verification case (binary hypothesis testing problem) are extended to get the upper bound on the total probability of error for the identification case.

$$P(\text{error}) \leq \frac{M(M+1)}{2} \max_{\alpha} [\alpha P(H_1|H_0) + (1-\alpha) P(H_0|H_1)], \quad (5.1)$$

This is the worst case upper bound on the total probability of error for the identification problem. Using terminology from [17], we introduce the recognition rate

$$R = \frac{\log M}{K} \quad (5.2)$$

and the error exponent for the bound on $P(\text{error})$

$$E(R) = -\frac{\log(\text{Bound on } P(\text{error}))}{K} \quad (5.3)$$

where the bound is given by (5.1). Fig. 5.9 presents the results for the worst case upper bound plotted as the error exponent $E(R)$ vs. recognition rate R for various values of K .

Given the number of templates per iris class and the total error probability, Fig. 5.9 and (5.3) specify the maximum number of iris classes that an iris recognition system may contain such that the total probability of error does not exceed the specified value.

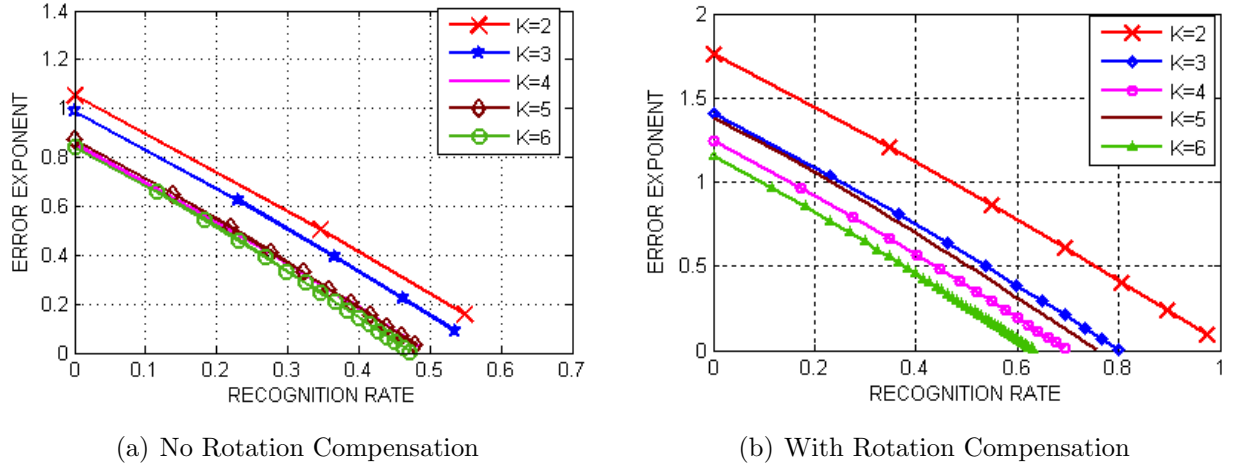


Fig. 5.9: Dependence of the error exponent $E(R)$ on the recognition rate R

Union of Chernoff Bounds

In this section, we use the Chernoff bounds on conditional probability of error, to find the upper bound on the total $P(\text{error})$. Using the Chernoff bound rate function formulated in (4.21), we evaluate the Chernoff bound on the total probability of error.

$$P(\text{error}) \leq \pi_0 \sum_{i=1}^M e^{-K \cdot I_0(0, L_0, \dots, L_0)} + (1 - \pi_0) \cdot e^{-K \cdot I_0(0, L_0, \dots, L_0)} + \sum_{j=1}^M \pi_j \cdot \sum_{i=1, i \neq j}^M e^{-K [I_0(\frac{L_0+L_1}{2}, \frac{L_0+L_1}{2}, L_0, \dots, L_0) - \frac{L_0+L_1}{2}]}$$

For the case of $M = 2$ the bound is given by

$$P(\text{error}) \leq \pi_0 \sum_{i=1}^2 e^{-K \cdot I_0(0, L_0)} + (1 - \pi_0) \cdot e^{-K \cdot I_0(0, L_0)} + \sum_{j=1}^2 \pi_j \cdot \sum_{i=1, i \neq j}^2 e^{-K [I_0(\frac{L_0+L_1}{2}, \frac{L_0+L_1}{2}) - \frac{L_0+L_1}{2}]}$$

Assuming equal prior probabilities for all the hypotheses ($\pi_j = \frac{1}{M+1}, \forall j = 0, \dots, M$),

$$P(\text{error}) \leq (\pi_0 + 1) \cdot e^{-K \cdot I_0(0, L_0)} + (1 - \pi_0) \cdot e^{-K [I_0(\frac{L_0+L_1}{2}, \frac{L_0+L_1}{2}) - \frac{L_0+L_1}{2}]}$$

The upper bound on the total $P(\text{error})$ as a function of K for $M = 2$ is shown in Fig. 5.10. Solving the log-likelihood ratios L_0, L_1 (see appendix for derivations of L_0 and L_1) and corresponding rate function, we calculate the upper bound on the probability of error. Table 5.3 and 5.4 summarize the numerical results for the CASIA database with $M = 2$.

Table 5.3: Chernoff bound: CASIA database (No Rotation Compensation)

K	L_0	L_1	$I_0(0, L_0)$	$I_0(\frac{L_0+L_1}{2}, \frac{L_0+L_1}{2})$	$P(\text{error})$
1	-1.7281	23.7308	0.9671	17.3250	0.5069
2	-1.7731	23.5381	0.9585	15.8552	0.1955
3	-1.7959	23.4448	0.94	15.1388	0.0793
4	-1.8096	23.3897	0.9291	14.7174	0.0323
5	-1.8188	23.3533	0.9192	14.4373	0.01342
6	-1.8254	23.3275	0.9123	14.2447	5.57×10^{-3}
10	-1.8399	23.2716	0.8979	13.8197	1.6758×10^{-4}
20	-1.8520	23.2255	0.8813	13.4703	2.94422×10^{-8}

Table 5.4: Chernoff bound: CASIA database (With Rotation Compensation)

K	L_0	L_1	$I_0(0, L_0)$	$I_0(\frac{L_0+L_1}{2}, \frac{L_0+L_1}{2})$	$P(\text{error})$
1	-3.1558	40.0026	1.8409	33.2755	0.21094
2	-3.168	36.8288	1.8182	28.1746	0.03503
3	-3.174	35.153	1.7811	25.5976	6.3574×10^{-3}
4	-3.178	34.1185	1.7463	24.0506	1.23×10^{-3}
5	-3.1801	33.4153	1.7195	23.0277	2.4546×10^{-4}
6	-3.1816	32.9065	1.698	22.2731	5.0×10^{-5}
10	-3.1846	31.7768	1.6519	20.6743	8.89×10^{-8}
20	-3.1868	30.8143	1.6121	19.3679	1.322×10^{-15}

5.2 Results for WVU Database

Similar to encoding the data from CASIA database, we encode the iris images from WVU database using a Gabor filter-based method, our interpretation of Daugman's algorithm.

5.2.1 Maximum Likelihood parameter estimates

Table 5.5 summarizes the ML estimates of the unknown parameters for Genuine and Imposter matching score models.

Table 5.5: Maximum Likelihood Parameter Estimates: WVU Database

(No Rotation Compensation)		(With Rotation Compensation)	
Genuine	Imposter	Genuine	Imposter
$\hat{\mu}_1 = 0.3912$	$\hat{\mu}_0 = 0.4610$	$\hat{\mu}_1 = 0.3279$	$\hat{\mu}_0 = 0.4402$
$\hat{\sigma}_1^2 = 0.006457$	$\hat{\sigma}_0^2 = 2.1621 * 10^{-4}$	$\hat{\sigma}_1^2 = 0.0041$	$\hat{\sigma}_0^2 = 6.3415 * 10^{-5}$
$\hat{\rho}_1 = 0.2117$	$\hat{\rho}_0 = 0.1976$	$\hat{\rho}_1 = 0.2257$	$\hat{\rho}_0 = 0.2933$

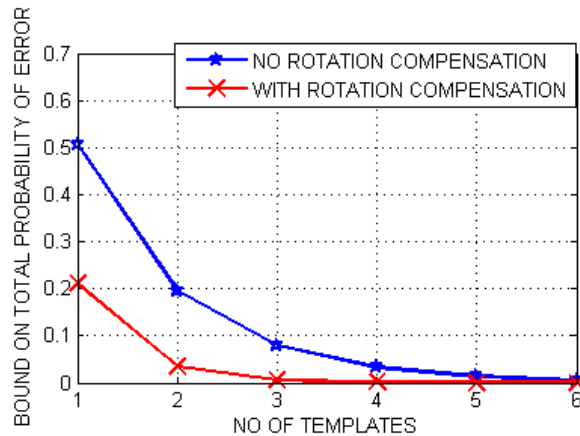


Fig. 5.10: Upper Bound on Probability of error

5.2.2 Shapiro-Wilk normality test results

Table 5.6 shows the p-value statistic for the Shapiro-Wilk normality test for the Genuine and Imposter matching scores. Note that for the case of $K = 1$ the Gaussian model is rejected when

Table 5.6: p-values for Shapiro-wilk normality test: WVU Database

No. of Samples	(No Rotation Compensation)		(With Rotation Compensation)	
	Genuine	Imposter	Genuine	Imposter
K=1	0.5496	0.3393	0.697	0.01908
K=3	0.00544	0.0407	0.01337	0.1144

fit of Imposter data is tested. For $K = 3$, the Gaussian model is rejected.

5.2.3 Performance Analysis: Verification Case

We analyze the performance of our models and test statistics using empirical approach (simulated and bootstrapped data) and using the Chernoff bounds and Large Deviations approximations.

Empirical Evaluation

Fig. 5.11 compares the performance of two decision rules: the D-bar and Log-likelihood ratio when $K = 1$ (a single image) and $K = 3$ (3 snapshots of the same iris) are used. The left panel shows ROC's for simulated data. the right panel shows the results for bootstrapped data. Fig. 5.12 shows similar result for the case when data are compensated for rotation.

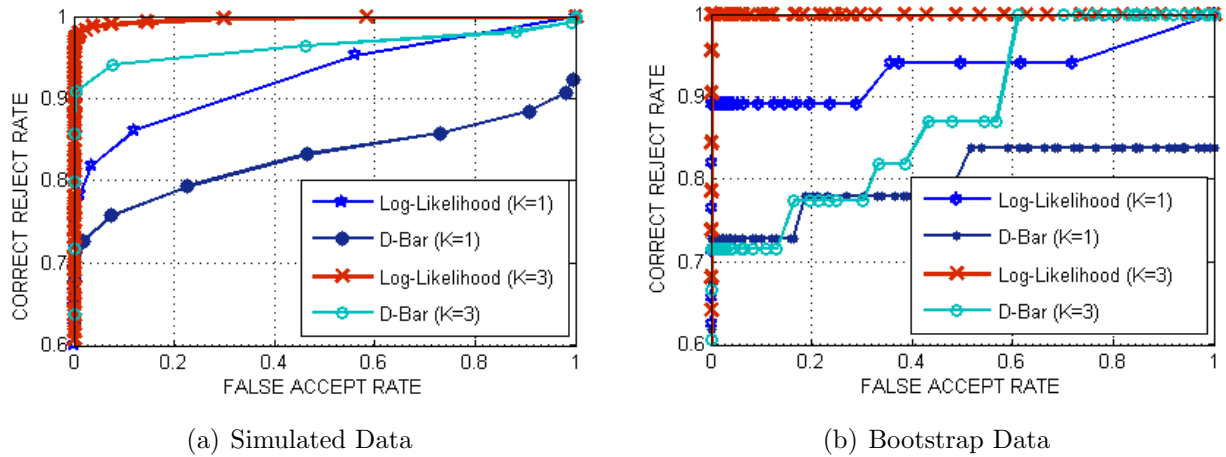


Fig. 5.11: ROC using empirical approach (no rotation compensation)

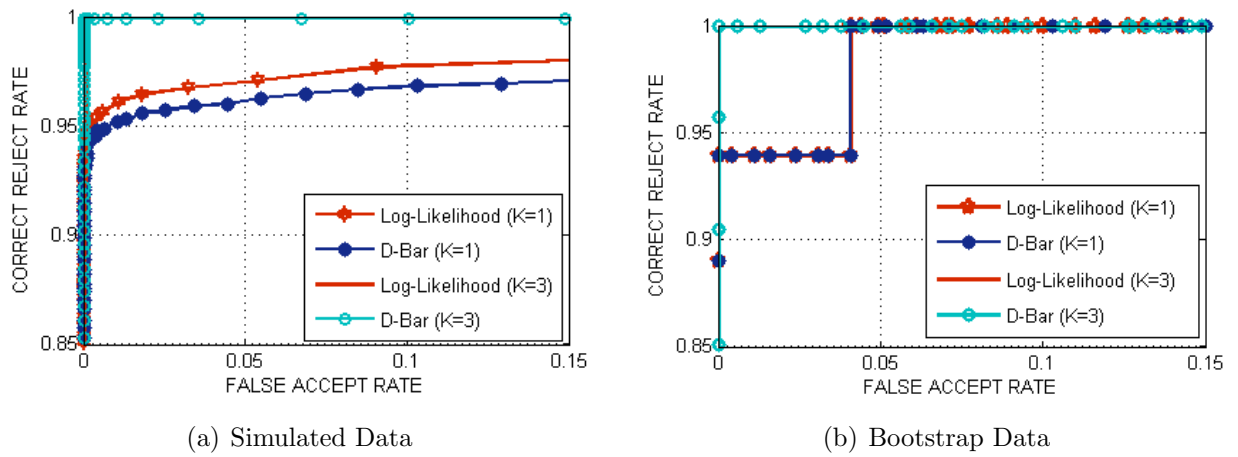


Fig. 5.12: ROC using empirical approach (with rotation compensation)

Chernoff Bound and Large Deviations Approximation

The FAR and FRR are upper bounded using Chernoff bound and approximated using Large Deviation approximation. The resulting ROC curves are shown in Fig. 5.13.

5.2.4 Performance Analysis: Identification Case

The Worst Case Upper Bound

We further find the recognition rate and the upper bound on the error exponent. Fig. 5.14 shows the plots of the error exponent as a function of the recognition rate parameterized by K .

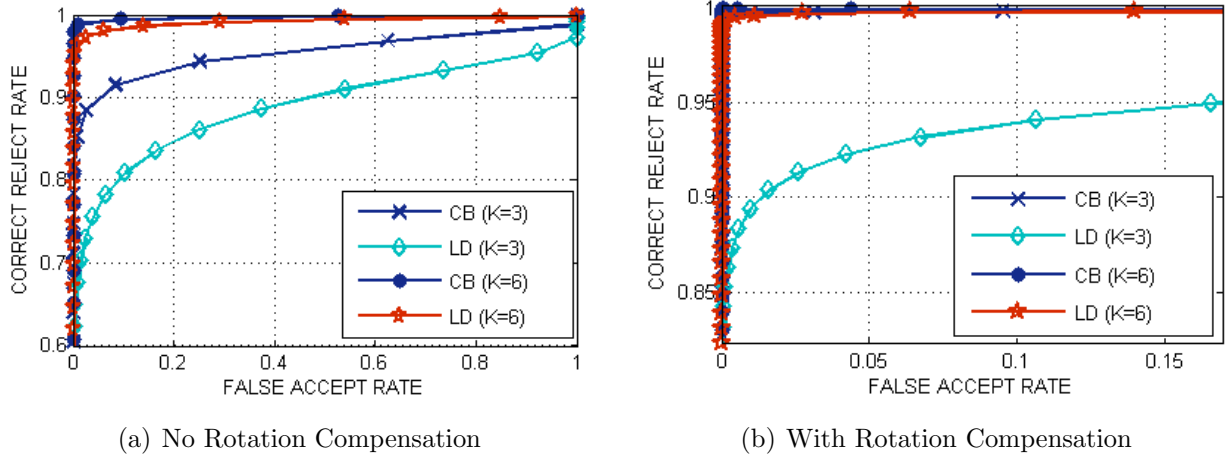


Fig. 5.13: ROC curves obtained using the bound and approximation

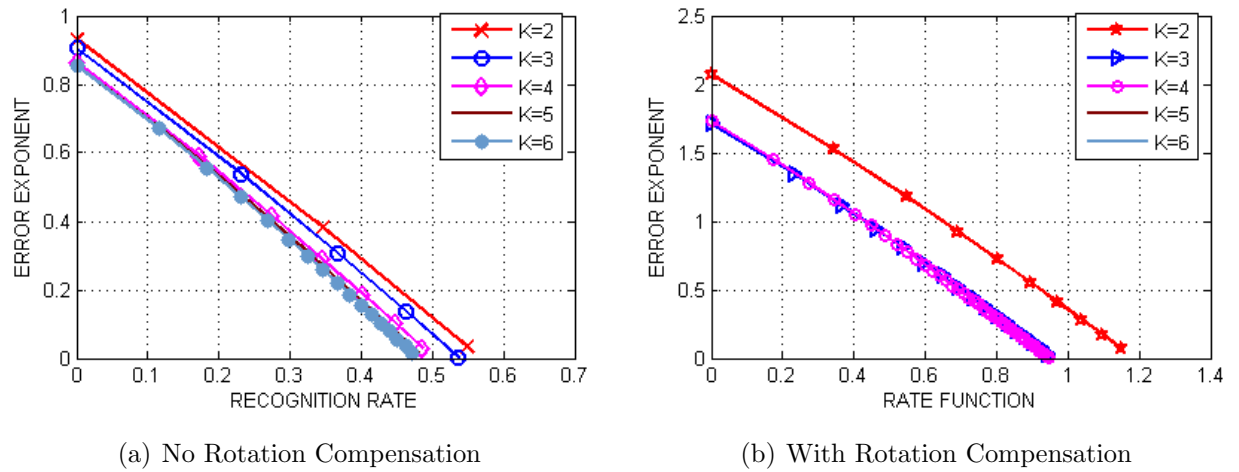


Fig. 5.14: Dependence of the error exponent $E(R)$ on the recognition rate R

Union of Chernoff Bounds

Table 5.7 and 5.8 summarize the numerical results for L_0 , L_1 and the corresponding rate function with $M = 2$.

Table 5.7: Chernoff bound: WVU database (No Rotation Compensation)

K	L_0	L_1	$I_0(0, L_0)$	$I_0(\frac{L_0+L_1}{2}, \frac{L_0+L_1}{2})$	$P(\text{error})$
1	-1.61148	25.027	0.9015	20.8124	0.5399
2	-1.6281	24.5777	0.9107	20.3484	0.2152
3	-1.6388	24.3063	0.9081	20.2344	0.08889
4	-1.6463	24.1244	0.899	20.1925	0.03648
5	-1.6518	23.9941	0.8874	20.1697	0.01573
6	-1.6560	23.8960	0.8759	20.1614	6.9417×10^{-3}

Table 5.8: Chernoff bound: WVU database (With Rotation Compensation)

K	L_0	L_1	$I_0(0, L_0)$	$I_0(\frac{L_0+L_1}{2}, \frac{L_0+L_1}{2})$	$P(\text{error})$
1	-3.1703	141.6742	2.0679	110.7071	0.1681
2	-3.2020	132.8483	1.9526	109.37	0.026733
3	-3.2199	128.2190	1.8705	108.9243	4.8618×10^{-3}
4	-3.2313	125.3680	1.8209	108.7015	9.1331×10^{-4}
5	-3.2393	123.4356	1.7785	108.5678	1.8276×10^{-4}
6	-3.2452	122.0396	1.7455	108.4787	3.7626×10^{-5}

The upper bound on the total probability of error is plotted vs. K.

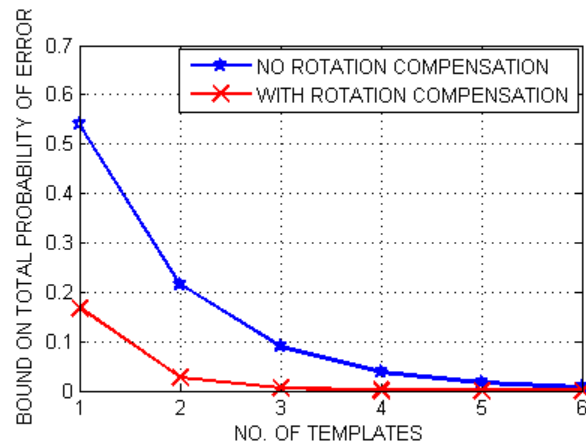


Fig. 5.15: Upper Bound on the total probability of error

Chapter 6

Conclusions

In a traditional iris-based recognition system, a single image for each iris class is used for recognition. In this thesis, we exploited the possibility of using multiple copies of the same iris. Modern cameras are capable of capturing more than a single snapshot over a small amount of time during adjustment period (approx. 5 to 10 images over 2 sec. period). Therefore our assumption of availability of multiple images from the same iris is practically feasible.

We took a model based approach and designed a suboptimal decision rule that in most cases outperforms the traditional rule based on comparison of a HD value with a threshold. We model the matching scores in the form of Hamming distances as realizations of Gaussian processes with unknown parameters. These parameters are estimated using the ML estimation procedure. The Shapiro-Wilk normality test is used to validate the model fit to the data we model. The p-values for the CASIA and WVU databases show reasonable fit to the Gaussian model.

We designed the plug-in log-likelihood test statistic with ML estimated parameters substituted in place of unknown true parameters and evaluated its performance. In all the experiments the designed plug-in Log likelihood test statistic outperforms \bar{d} decision rule, based on comparison of the average HD value with a threshold. The verification and identification problems are stated as binary and $(M+1)$ -ary hypothesis testing problems, respectively. Here M is the total number of iris classes and an additional Imposter hypothesis. The performance of the iris-based verification system is evaluated using empirical data. Performance of a large scale system is predicted by invoking Chernoff bounds and Large Deviations approximation. Chernoff bound

provides tight results. The Large Deviations theory assumes that a probability of error can be approximated by an exponential function and deals with the asymptotic exponent, called Large Deviations rate function. It provides an asymptotically tight expression for the convergence rate of error probability.

The performance of identification case is analyzed by extending the results for verification case. Influence of the number of iris classes and the number of copies of the same iris on the recognition performance is explored.

Chapter 7

Future Work

1. In this thesis, we assumed that collected data are of good quality. In practice, however, quality of iris images in an acquired video sequence may vary in time. For example, a slight move of the individual whose iris is imaged causes a smear in the image. Camera with automatic focus often goes out of focus when object at a close distance is imaged. The resulting image is de-focused.
In our future work, we propose to model and analyze the case when iris images collected from the same iris class vary in their quality. It should be interesting to find a model that provides a good fit to these data.
2. In this work, we assumed that the number of iris images collected to describe an iris class is fixed. In practical situations, different iris classes may be represented by a different number of iris images. In our future work, we propose to analyze how this factor influences the performance of iris recognition system.
3. Since the Gaussian model proposed in this thesis does not always provide a good fit to the data (especially for the case when data are compensated for rotation), the model needs to be refined. Instead of using a Gaussian model, we propose to consider a mixture of Gaussian distributions to model the matching scores under Genuine and Imposter hypotheses.
4. Currently, only Chernoff bound has been designed for identification case. The Large deviations approximation can also be derived using the same identification problem formulation. This will provide a quick estimate of the probability of error.
5. The analysis performed in this thesis can be further extended to evaluating the performance

of iris-based recognition system implementing Principle Component Analysis (PCA) and Independent Component Analysis (ICA) encoding techniques [18]. It can also be employed to evaluate the performance of a recognition system using synthetic iris database, in which the iris images are artificially generated and used for recognition.

References

- [1] R. Duda, P. Hart, D. Stork, *Pattern Classification*. 2nd edition, A Wiley-Interscience Publication, John Wiley and Sons, 2001.
- [2] R. P. Wildes, “Automated iris recognition: An emerging biometric technology,” in *Proceedings of the IEEE*, vol. 85(9), September 1997, pp. 1348–1363.
- [3] J. Daugman, “How Iris Recognition Works,” in *Proceedings of International Conference on Image Processing*, vol. 1, 2002, pp. 33–36.
- [4] —, “Biometric personal identification system based on iris analysis,” United States Patent (Patent Number: 5,291,560), 1994.
- [5] —, “High confidence visual recognition of persons by a test of statistical independence,” in *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 15(11), November 1993, pp. 1148–1161.
- [6] H. L. Van Trees, *Detection, Estimation and Modulation Theory*. Wiley, New York, 2001.
- [7] S. S. Shapiro, M. B. Wilk, “An Analysis of Variance Test for Normality (Complete Samples),” in *Biometrika*, vol. 52, no. 3/4, December 1965, pp. 591–611.
- [8] J. Wayman., A. Jain, D. Maltoni, D. Maio, *Biometric Systems: Technology, Design and Performance Evaluation*. Springer, 2005.
- [9] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation*. John Wiley and Sons, Inc., New York, 1990.
- [10] S. R. S. Varadhan, *Large Deviations and Applications*. SIAM, Philadelphia, 1984.
- [11] A. Dembo and O. Zeitouni, *Large deviations techniques and applications, second edition. Applications of mathematics, Vol 38*. Springer, New York, 1998.
- [12] M. R. Chernick, *Bootstrap Methods : A Practitioner’s Guide*. Wiley Series in Probability and Statistics, 1999.
- [13] H. Chernoff, “A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations,” in *Annals of Mathematical Statistics*, vol. 23, no. 4, December 1952, pp. 493–507.
- [14] N. A. Schmid, J. A. O’Sullivan, “Performance Prediction Methodology for Biometric Systems Using a Large Deviations Approach,” in *IEEE Transactions on Signal Processing*, vol. 52, no. 10, October 2004, pp. 3036– 3045.

- [15] T. Moon, W. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, Upper Saddle River, NJ, USA., 2000.
- [16] *CASIA Iris Image Database, ver. 1.0*, <http://www.sinobiometrics.com/casiairis.htm>.
- [17] R. E. Blahut, *Principles and Practice of Information Theory*. Addison-Wesley, 1987.
- [18] V. Dorairaj, N. Schmid, G. Fahmy, "Performance Evaluation of Iris Based Recognition System Implementing PCA and ICA Techniques," in *Proc. of the SPIE 2005 Symp. on Defence and Security, Conf. 5779*, March 28-29, Orlando, FL.
- [19] A. K. Jain, R. Bolle, and S. Pankanti, *Introduction to Biometrics, Biometrics Personal Identification in Networked Society*. Kluwer Acad. Publisher, 1999.
- [20] J. A. O'Sullivan, N. A. Schmid, "Performance Analysis of Physical Signature Authentication," in *IEEE Transactions on Information Theory*, vol. 47, no. 7, November 2001, pp. 3034–3039.
- [21] D. W. Stroock, *An Introduction to the Theory of Large Deviations*. Springer, New York, 1984.
- [22] J. D. Deuschel, D. W. Stroock, *Large Deviations, Pure and Applied Math vol. 137*. Academic Press, Boston, 1989.
- [23] L. Masek, *Recognition of Human Iris Patterns for Biometric Identification*, B. S. Dissertation, The University of Western Australia, 2003.
- [24] L. Ma, T. Tan, Y. Wang and D. Zhang, "Personal Identification Based on Iris Texture Analysis," in *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, 2003, pp. 1519–1533.
- [25] J. Daugman, "The importance of being random: Statistical principles of iris recognition," in *Pattern Recognition*, vol. 36, no. 2, 2001, pp. 279–291.
- [26] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Addison-Wesley, 1992.
- [27] J. A. O'Sullivan and N. A. Schmid, "Large Deviations Performance Analysis for Biometrics Recognition," in *in Proc. of 40th Annual Allerton Conference on Communication, Control, and Computing*, October 2002.
- [28] P. Dupuis, R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley Series in Probability and Statistics, John Wiley and Sons, Inc., New York,, 1997.
- [29] T. Cover, J. Thomas, *Elements of Information Theory*. Wiley, New York, 1991.
- [30] F. R. Gantmacher, *The Theory of Matrices, Vol. 1*. Chelsea Publishing Company, New York, 1959.
- [31] A. Mansfield, J. Wayman, "Best Practice Standards for Testing and Reporting on Biometric Device Performance," in *Nat. Physical Lab., Middlesex, U.K.*, 2002.
- [32] C. C. Leang, D. H. Johnson, "On the Asymptotic of M-Hypothesis Bayesian Detection," in *IEEE Transaction on Information Theory*, vol. 43, no. 1, 1997, pp. 280–282.

- [33] V. Dorairaj, N. Schmid, G. Fahmy, "Performance Evaluation of Non-Ideal Iris Based Recognition System Implementing Global ICA Encoding," in *Proc. of ICIP*, to appear, 2005.
- [34] A. S. Abhyankar, L. A. Hornak, S. C. Schuckers, "Biortogonal wavelet-based iris recognition," in *Proc. of the SPIE 2005 Symp. on Defence and Security, Conf. 5779*, March 28-29, Orlando, FL.
- [35] C. Fancourt, L. Bogoni, J. Hanna, Y. Guo, N. Nakahashi, "Iris Recognition at a distance," in *Proc. of the IAPR Conf. on Audio and Video Based Biometric Person Authentication*, to appear, 2005.
- [36] Y. Du, B. Bonney, R. Ives, D. M. Etter, "Analysis of partial iris recognition," in *Proc. of the SPIE 2005 Symp. on Defence and Security, Conf. 5779*, March 28-29, Orlando, FL.
- [37] L. Devroye, L. Gyrfi and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag, New York, 1996.
- [38] A. Leon-Garcia, *Probability and Random Processes for Eletrical Engineering*. 2nd, Ed., Addison Wesley, 1994.

Appendix A

Derivations for Rate Functions: Verification Case

In the verification problem,

$$\begin{aligned} \text{Under Imposter Hypothesis: } \mathbf{d} &\sim N(\hat{\mu}_0 \mathbf{1}, \hat{\mathbf{R}}_0) \\ \text{Under Genuine Hypothesis: } \mathbf{d} &\sim N(\hat{\mu}_1 \mathbf{1}, \hat{\mathbf{R}}_1) \end{aligned} \tag{A.1}$$

where $\hat{\mathbf{R}}_0$ and $\hat{\mathbf{R}}_1$ have the form as in (2.2) with estimated parameters from (2.3), (2.4) substituted in it.

The plug-in log-likelihood ratio for the proposed Gaussian model for matching scores (A.1) is given by,

$$\begin{aligned} \Lambda_K &= \frac{1}{K} \log \frac{\hat{p}(\mathbf{d}|GH)}{\hat{p}(\mathbf{d}|IH)} \\ &= -\frac{1}{2K} (\mathbf{d} - \hat{\mu}_1 \mathbf{1})^T \hat{\mathbf{R}}_1^{-1} (\mathbf{d} - \hat{\mu}_1 \mathbf{1}) + \frac{1}{2K} (\mathbf{d} - \hat{\mu}_0 \mathbf{1})^T \hat{\mathbf{R}}_0^{-1} (\mathbf{d} - \hat{\mu}_0 \mathbf{1}) \\ &\quad - \frac{1}{2K} \log \det (\hat{\mathbf{R}}_1 \hat{\mathbf{R}}_0^{-1}) \end{aligned} \tag{A.2}$$

Using Chernoff bound, the upper bound on FAR and FRR are given by

$$FAR(\gamma, K) = P(\Lambda_K > \gamma | IH) \leq e^{-K \cdot I_0(\gamma, K)},$$

$$FRR(\gamma, K) = P(\Lambda_K \leq \gamma | GH) \leq e^{-K \cdot I_1(\gamma, K)}$$

where $I_0(\gamma, K)$, $I_1(\gamma, K)$ are the Chernoff rate functions under Imposter and Genuine hypothesis, respectively. Under Imposter hypothesis,

$$I_0(\gamma, K) = \sup_s [s\gamma - \bar{\varphi}_0(s, K)] \quad (\text{A.3})$$

where $\bar{\varphi}_0(s, K)$ is the normalized log moment generating function under Imposter hypothesis

$$\bar{\varphi}_0(s, K) = \frac{1}{K} \log (E_0 [e^{sK\Lambda_K}]) \quad (\text{A.4})$$

Under Genuine Hypothesis,

$$I_1(\gamma, K) = \sup_s [s\gamma - \bar{\varphi}_1(s, K)] \quad (\text{A.5})$$

where $\bar{\varphi}_1(s, K)$ is the normalized log moment generating function under Genuine hypothesis

$$\bar{\varphi}_1(s, K) = \frac{1}{K} \log (E_1 [e^{sK\Lambda_K}]) \quad (\text{A.6})$$

Using the theory of large deviations, the average probability of error given by (4.1) can be asymptotically approximated by an exponential function with explicit dependence on the estimated parameters of the system. If the Large deviation conditions are satisfied, the FAR and FRR can be approximated as

$$FAR(\gamma) = P(\Lambda_K > \gamma | IH) \approx G(K, \gamma) e^{-K \cdot I_0(\gamma)},$$

$$FRR(\gamma) = P(\Lambda_K \leq \gamma | GH) \approx G(K, \gamma) e^{-K \cdot I_1(\gamma)}$$

where $G(K, \gamma)$ is a slowly varying function of K and γ (often omitted in analysis), and $I_0(\gamma)$, $I_1(\gamma)$ are the large deviation rate functions under Imposter and Genuine hypothesis, respectively.

Under Imposter hypothesis,

$$I_0(\gamma) = \sup_s [s\gamma - \bar{\varphi}_0(s)] \quad (\text{A.7})$$

where $\bar{\varphi}_0(s)$ is the normalized log moment generating function under Imposter hypothesis

$$\bar{\varphi}_0(s) = \lim_{K \rightarrow \infty} \frac{1}{K} \log (E_0 [e^{sK\Lambda_K}]) \quad (\text{A.8})$$

Under Genuine Hypothesis,

$$I_1(\gamma) = \sup_s [s\gamma - \bar{\varphi}_1(s)] \quad (\text{A.9})$$

where $\bar{\varphi}_1(s)$ is the normalized log moment generating function under Genuine hypothesis

$$\bar{\varphi}_1(s) = \lim_{K \rightarrow \infty} \frac{1}{K} \log (E_1 [e^{sK\Lambda_K}]) \quad (\text{A.10})$$

A.1 Normalized Log-Moment Generating Function under Imposter Hypothesis

$$\begin{aligned}\bar{\varphi}_0(s, K) &= \frac{1}{K} \log (E_0 [e^{sK\Lambda_K}]) \\ &= \frac{1}{K} \log \left[\int_{K\text{-fold}} \dots \int e^{s[-\frac{1}{2}(\mathbf{d}-\hat{\mu}_1\mathbf{1})^T\hat{\mathbf{R}}_1^{-1}(\mathbf{d}-\hat{\mu}_1\mathbf{1})+\frac{1}{2}(\mathbf{d}-\hat{\mu}_0\mathbf{1})^T\hat{\mathbf{R}}_0^{-1}(\mathbf{d}-\hat{\mu}_0\mathbf{1})-\frac{1}{2}\log\det(\hat{\mathbf{R}}_1\hat{\mathbf{R}}_0^{-1})]} \cdot P_0(\mathbf{d}) \cdot d\mathbf{d} \right]\end{aligned}$$

where $P_0(\mathbf{d})$ is the probability density of \mathbf{d} under Imposter hypothesis (H_0) given by

$$P_0(\mathbf{d}) = \int_{K\text{-fold}} \dots \int \frac{e^{[-\frac{1}{2}(\mathbf{d}-\hat{\mu}_0\mathbf{1})^T\hat{\mathbf{R}}_0^{-1}(\mathbf{d}-\hat{\mu}_0\mathbf{1})-\frac{1}{2}\log\det\hat{\mathbf{R}}_0]}}{(2\pi)^{K/2}} \cdot d\mathbf{d}$$

Considering the expression in the exponential under the integral,

$$\begin{aligned}& -\frac{s}{2}(\mathbf{d}-\hat{\mu}_1\mathbf{1})^T\hat{\mathbf{R}}_1^{-1}(\mathbf{d}-\hat{\mu}_1\mathbf{1}) + \frac{s}{2}(\mathbf{d}-\hat{\mu}_0\mathbf{1})^T\hat{\mathbf{R}}_0^{-1}(\mathbf{d}-\hat{\mu}_0\mathbf{1}) - \frac{s}{2}\log\det(\hat{\mathbf{R}}_1\hat{\mathbf{R}}_0^{-1}) \\ & - \frac{1}{2}(\mathbf{d}-\hat{\mu}_0\mathbf{1})^T\hat{\mathbf{R}}_0^{-1}(\mathbf{d}-\hat{\mu}_0\mathbf{1}) - \frac{1}{2}\log\det\hat{\mathbf{R}}_0 \\ &= -\frac{s}{2}(\mathbf{d}-\hat{\mu}_1\mathbf{1})^T \cdot \hat{\mathbf{R}}_1^{-1} \cdot (\mathbf{d}-\hat{\mu}_1\mathbf{1}) + \frac{1}{2}(\mathbf{d}-\hat{\mu}_0\mathbf{1})^T \cdot [s\hat{\mathbf{R}}_0^{-1} - \hat{\mathbf{R}}_0^{-1}] \cdot (\mathbf{d}-\hat{\mu}_0\mathbf{1}) \\ & \quad - \frac{s}{2}\log\det(\hat{\mathbf{R}}_1\hat{\mathbf{R}}_0^{-1}) - \frac{1}{2}\log\det(\hat{\mathbf{R}}_0) \\ &= -\frac{s}{2}(\mathbf{d}-\hat{\mu}_1\mathbf{1})^T \cdot \hat{\mathbf{R}}_1^{-1} \cdot (\mathbf{d}-\hat{\mu}_1\mathbf{1}) + \frac{1}{2}(\mathbf{d}-\hat{\mu}_0\mathbf{1})^T \cdot [(s-1)\hat{\mathbf{R}}_0^{-1}] \cdot (\mathbf{d}-\hat{\mu}_0\mathbf{1}) \\ & \quad - \frac{s}{2}\log\det(\hat{\mathbf{R}}_1\hat{\mathbf{R}}_0^{-1}) - \frac{1}{2}\log\det(\hat{\mathbf{R}}_0) \tag{A.11} \\ &= -\frac{1}{2}[(\mathbf{d}-\hat{\mu}_0\mathbf{1}) + (\mu_0 - \mu_1)\mathbf{1}]^T \cdot s\hat{\mathbf{R}}_1^{-1} \cdot [(\mathbf{d}-\hat{\mu}_0\mathbf{1}) + (\mu_0 - \mu_1)\mathbf{1}] \\ & \quad + \frac{1}{2}(\mathbf{d}-\hat{\mu}_0\mathbf{1})^T \cdot [(s-1)\hat{\mathbf{R}}_0^{-1}] \cdot (\mathbf{d}-\hat{\mu}_0\mathbf{1}) - \frac{s}{2}\log\det(\hat{\mathbf{R}}_1\hat{\mathbf{R}}_0^{-1}) - \frac{1}{2}\log\det(\hat{\mathbf{R}}_0) \\ &= -\frac{1}{2}(\mathbf{d}-\hat{\mu}_0\mathbf{1})^T \cdot [s\hat{\mathbf{R}}_1^{-1} - (s-1)\hat{\mathbf{R}}_0^{-1}] \cdot (\mathbf{d}-\hat{\mu}_0\mathbf{1}) - \frac{1}{2}(\hat{\mu}_0 - \hat{\mu}_1)\mathbf{1}^T \cdot [s\hat{\mathbf{R}}_1^{-1}] \cdot (\mathbf{d}-\hat{\mu}_0\mathbf{1}) \\ & \quad - \frac{1}{2}(\mathbf{d}-\hat{\mu}_0\mathbf{1})^T \cdot [s\hat{\mathbf{R}}_1^{-1}] \cdot (\hat{\mu}_0 - \hat{\mu}_1)\mathbf{1} - \frac{1}{2}(\hat{\mu}_0 - \hat{\mu}_1)\mathbf{1}^T \cdot [s\hat{\mathbf{R}}_1^{-1}] \cdot (\hat{\mu}_0 - \hat{\mu}_1)\mathbf{1} \\ & \quad - \frac{s}{2}\log\det(\hat{\mathbf{R}}_1\hat{\mathbf{R}}_0^{-1}) - \frac{1}{2}\log\det(\hat{\mathbf{R}}_0)\end{aligned}$$

Let $\mathbf{Q}^{-1} = [s\hat{\mathbf{R}}_1^{-1} - (s-1)\hat{\mathbf{R}}_0^{-1}]$. Therefore the expression in the exponential now reduces to

$$\begin{aligned}
 & -\frac{1}{2}(\mathbf{d} - \hat{\mu}_0\mathbf{1})^T \cdot \mathbf{Q}^{-1} \cdot (\mathbf{d} - \hat{\mu}_0\mathbf{1}) - \frac{1}{2}(\hat{\mu}_0 - \hat{\mu}_1)\mathbf{1}^T \cdot [s\hat{\mathbf{R}}_1^{-1}] \cdot (\mathbf{d} - \hat{\mu}_0\mathbf{1}) \\
 & -\frac{1}{2}(\mathbf{d} - \hat{\mu}_0\mathbf{1})^T \cdot (s\hat{\mathbf{R}}_1^{-1}) \cdot (\hat{\mu}_0 - \hat{\mu}_1)\mathbf{1} - \frac{1}{2}(\hat{\mu}_0 - \hat{\mu}_1)\mathbf{1}^T \cdot (s\hat{\mathbf{R}}_1^{-1}) \cdot (\hat{\mu}_0 - \hat{\mu}_1)\mathbf{1} \\
 & -\frac{s}{2} \log \det (\hat{\mathbf{R}}_1\hat{\mathbf{R}}_0^{-1}) - \frac{1}{2} \log \det (\hat{\mathbf{R}}_0) \\
 = & -\frac{1}{2}[\mathbf{d} - \hat{\mu}_0\mathbf{1} - (\hat{\mu}_1 - \hat{\mu}_0)\mathbf{1} \cdot \mathbf{Q} \cdot (s\hat{\mathbf{R}}_1^{-1})\mathbf{1}]^T \cdot \mathbf{Q}^{-1} \cdot [\mathbf{d} - \hat{\mu}_0\mathbf{1} - (\hat{\mu}_1 - \hat{\mu}_0)\mathbf{1} \cdot \mathbf{Q} \cdot (s\hat{\mathbf{R}}_1^{-1})\mathbf{1}] \\
 & -\frac{1}{2}(\hat{\mu}_1 - \hat{\mu}_0)^2 \cdot s \cdot \mathbf{1}^T \cdot \hat{\mathbf{R}}_1^{-1} \cdot \mathbf{1} + \frac{1}{2}(\hat{\mu}_1 - \hat{\mu}_0)^2 \cdot s^2 \cdot \mathbf{1}^T \cdot \hat{\mathbf{R}}_1^{-1} \cdot \mathbf{Q} \cdot \hat{\mathbf{R}}_1^{-1} \cdot \mathbf{1} \\
 & -\frac{s}{2} \log \det (\hat{\mathbf{R}}_1\hat{\mathbf{R}}_0^{-1}) - \frac{1}{2} \log \det \mathbf{Q} + \frac{1}{2} \log \det \mathbf{Q} - \frac{1}{2} \log \det (\hat{\mathbf{R}}_0)
 \end{aligned}$$

Substituting this expression for the exponential under the integral in (A.11), we get the normalized log-moment generating function. Collecting all the terms involving \mathbf{d} , completing the squares and integrating over all values of \mathbf{d} we get,

$$\begin{aligned}
 & \int_{K\text{-fold}} \dots \int \frac{e^{-\frac{1}{2}[\mathbf{d} - \hat{\mu}_0\mathbf{1} - (\hat{\mu}_1 - \hat{\mu}_0)\mathbf{1} \cdot \mathbf{Q} \cdot (s\hat{\mathbf{R}}_1^{-1})\mathbf{1}]^T \cdot \mathbf{Q}^{-1} \cdot [\mathbf{d} - \hat{\mu}_0\mathbf{1} - (\hat{\mu}_1 - \hat{\mu}_0)\mathbf{1} \cdot \mathbf{Q} \cdot (s\hat{\mathbf{R}}_1^{-1})\mathbf{1}] - \frac{1}{2} \log \det \mathbf{Q}}}{(2\pi)^{K/2}} d\mathbf{d} \\
 & = 1
 \end{aligned}$$

Hence the normalized log-moment generating function is given by the the remaining terms (terms independent of \mathbf{d})

$$\begin{aligned}
 \bar{\varphi}_0(s, K) & = \frac{1}{K} \log \left[e^{\frac{s^2}{2}(\hat{\mu}_1 - \hat{\mu}_0)^2 \cdot \mathbf{1}^T \cdot \hat{\mathbf{R}}_1^{-1} \cdot \mathbf{Q} \cdot \hat{\mathbf{R}}_1^{-1} \cdot \mathbf{1} - \frac{s}{2}(\hat{\mu}_1 - \hat{\mu}_0)^2 \cdot \mathbf{1}^T \cdot \hat{\mathbf{R}}_1^{-1} \cdot \mathbf{1} - \frac{s}{2} \log \det (\hat{\mathbf{R}}_1\hat{\mathbf{R}}_0^{-1}) - \frac{1}{2} \log \det (\hat{\mathbf{R}}_0)\mathbf{Q}^{-1}} \right] \\
 & = \frac{s^2}{2K}(\hat{\mu}_1 - \hat{\mu}_0)^2 \cdot \mathbf{1}^T \cdot \hat{\mathbf{R}}_1^{-1} \cdot \mathbf{Q} \cdot \hat{\mathbf{R}}_1^{-1} \cdot \mathbf{1} - \frac{s}{2}(\hat{\mu}_1 - \hat{\mu}_0)^2 \cdot \mathbf{1}^T \cdot \hat{\mathbf{R}}_1^{-1} \cdot \mathbf{1} \\
 & \quad - \frac{s}{2K} \log \det (\hat{\mathbf{R}}_1\hat{\mathbf{R}}_0^{-1}) - \frac{1}{2K} \log \det (\hat{\mathbf{R}}_0\mathbf{Q}^{-1})
 \end{aligned}$$

Let $\lambda_1(l)$, $l = 1, \dots, K$ be the l^{th} eigenvalue of the covariance matrix $\hat{\mathbf{R}}_1$ and $\lambda_0(l)$, $l = 1, \dots, K$ be the l^{th} eigenvalue of the covariance matrix $\hat{\mathbf{R}}_0$.

Recall that the matrix \mathbf{Q}^{-1} in the above equations defined as

$$\mathbf{Q}^{-1} = [s\hat{\mathbf{R}}_1^{-1} - (s-1)\hat{\mathbf{R}}_0^{-1}]$$

Let $q(l)$, $l = 1, \dots, K$ be the eigenvalues of \mathbf{Q}^{-1} . We write the eigenvalues of \mathbf{Q}^{-1} in terms of eigenvalues of $\hat{\mathbf{R}}_0$ and $\hat{\mathbf{R}}_1$ as follows

$$q(l) = \frac{s}{\lambda_1(l)} - \frac{(s-1)}{\lambda_0(l)}$$

Therefore, the eigenvalues of matrix \mathbf{Q} can be written as

$$\tilde{q}(l) = \frac{\lambda_1(l) \cdot \lambda_0(l)}{s\lambda_0(l) - (s-1)\lambda_1(l)}$$

But the covariance matrices $\hat{\mathbf{R}}_0$ and $\hat{\mathbf{R}}_1$ are hermitian symmetric with all the diagonal components as $\hat{\sigma}_i^2$, $i = 0, 1$ and all off-diagonal elements as $\hat{\sigma}_i^2 \hat{\rho}_i$, $i = 0, 1$, respectively. Hence each of them has only two non-zero and unique eigenvalues $\lambda_0(l)$ and $\lambda_1(l)$, $l = 1, 2$ given by

$$\begin{aligned} \lambda_0(1) &= \hat{\sigma}_0^2 + (K-1)\hat{\sigma}_0^2 \hat{\rho}_0 \\ \lambda_0(2) &= \hat{\sigma}_0^2 - \hat{\sigma}_0^2 \hat{\rho}_0 \\ \lambda_1(1) &= \hat{\sigma}_1^2 + (K-1)\hat{\sigma}_1^2 \hat{\rho}_1 \\ \lambda_1(2) &= \hat{\sigma}_1^2 - \hat{\sigma}_1^2 \hat{\rho}_1 \end{aligned} \tag{A.12}$$

Thus solving $\bar{\varphi}_0(s, K)$ further we get

$$\begin{aligned} \bar{\varphi}_0(s, K) &= \frac{s^2}{2K} \cdot (\hat{\mu}_1 - \hat{\mu}_0)^2 \cdot \frac{\lambda_0(1)}{\lambda_1(1)} \cdot \frac{K}{s\lambda_0(1) - (s-1)\lambda_1(1)} - \frac{s}{2K} \cdot (\hat{\mu}_1 - \hat{\mu}_0)^2 \cdot \frac{K}{\lambda_1(1)} \\ &\quad - \frac{s}{2K} \left[\log \frac{\lambda_1(1)}{\lambda_0(1)} + (K-1) \log \frac{\lambda_1(2)}{\lambda_0(2)} \right] \\ &\quad - \frac{1}{2K} \left[\log \frac{s\lambda_0(1) - (s-1)\lambda_1(1)}{\lambda_1(1)} + (K-1) \log \frac{s\lambda_0(2) - (s-1)\lambda_1(2)}{\lambda_1(2)} \right] \end{aligned} \tag{A.13}$$

For Chernoff rate function in (A.3) the normalized log-moment generating function $\bar{\varphi}_0(s, K)$ is given by (A.13). Note that the rate function is function of both γ and K .

Under large deviation approximation, we assume $K \rightarrow \infty$. Therefore the normalized log-moment generating function under Imposter hypothesis (A.13) can be approximated as

$$\begin{aligned}
\bar{\varphi}_0(s) &= \frac{s^2}{2} \cdot (\hat{\mu}_1 - \hat{\mu}_0)^2 \cdot \frac{\lambda_0(1)}{\lambda_1(1)} \cdot \frac{1}{s\lambda_0(1) - (s-1)\lambda_1(1)} - \frac{s}{2} \cdot (\hat{\mu}_1 - \hat{\mu}_0)^2 \cdot \frac{1}{\lambda_1(1)} \\
&\quad - \frac{s}{2} \left[\log \frac{\lambda_1(2)}{\lambda_0(2)} \right] \\
&\quad + \frac{1}{2} \left[\log \frac{\lambda_1(1)\lambda_0(2)}{s\lambda_0(2) - (s-1)\lambda_1(2)} \right] - \frac{1}{2} \log [\lambda_0(2)] \\
&= \frac{(\hat{\mu}_1 - \hat{\mu}_0)^2}{2} \left[\frac{\lambda_0(1)}{\lambda_1(1)} \cdot \frac{s^2}{s\lambda_0(1) - (s-1)\lambda_1(1)} - \frac{s}{\lambda_1(1)} \right] - \frac{s}{2} \log \left[\frac{\lambda_1(2)}{\lambda_0(2)} \right] \\
&\quad + \frac{1}{2} \log \left[\frac{\lambda_0(2)}{s\lambda_0(2) - (s-1)\lambda_1(2)} \right] + \frac{1}{2} \log \lambda_1(2) - \frac{1}{2} \log \lambda_0(2) \\
&= \frac{(\hat{\mu}_1 - \hat{\mu}_0)^2}{2} \left[\frac{s(s-1)}{s\lambda_0(1) - (s-1)\lambda_1(1)} \right] - \frac{s-1}{2} \log \left[\frac{\lambda_1(2)}{\lambda_0(2)} \right] \\
&\quad + \frac{1}{2} \log \left[\frac{\lambda_0(2)}{s\lambda_0(2) - (s-1)\lambda_1(2)} \right]
\end{aligned}$$

But as $K \rightarrow \infty \Rightarrow \lambda_0(1) \rightarrow \infty$ and $\lambda_1(1) \rightarrow \infty$. Therefore the first term

$$\frac{(\hat{\mu}_1 - \hat{\mu}_0)^2}{2} \left[\frac{s(s-1)}{s\lambda_0(1) - (s-1)\lambda_1(1)} \right] \rightarrow 0$$

and can be neglected under Large Deviation approximation. Therefore we write the log-moment generating function under Imposter hypothesis as

$$\bar{\varphi}_0(s) = -\frac{s-1}{2} \log \left[\frac{\lambda_1(2)}{\lambda_0(2)} \right] + \frac{1}{2} \log \left[\frac{\lambda_0(2)}{s\lambda_0(2) - (s-1)\lambda_1(2)} \right]$$

This is the log-moment generating function for LD rate function.

A.2 Threshold (γ) under Imposter Hypothesis

For Chernoff bound, the threshold $\gamma(s, K)$ in (A.3) can be found to get optimum solution for (A.3). It is given by

$$\begin{aligned}
\gamma(s, K) &= \frac{d}{ds} \bar{\varphi}_0(s, K) \\
&= \frac{(\hat{\mu}_1 - \hat{\mu}_0)^2}{2K} \cdot \frac{1}{\lambda_1(1)} \cdot \left[\frac{s^2 \lambda_0(1)^2 + 2s \lambda_0(1) \lambda_1(1) - s^2 \lambda_0(1) \lambda_1(1)}{[s \lambda_0(1) - (s-1) \lambda_1(1)]^2} - 1 \right] \\
&\quad - \frac{1}{2K} \log \frac{\lambda_1(1)}{\lambda_0(1)} - \frac{(K-1)}{2K} \log \frac{\lambda_1(2)}{\lambda_0(2)} \\
&\quad - \frac{1}{2} \cdot \frac{1}{s \lambda_0(1) - (s-1) \lambda_1(1)} \cdot \left[\frac{\lambda_0(1) - \lambda_1(1)}{K} \right] \\
&\quad - \frac{(K-1)}{2} \cdot \frac{1}{s \lambda_0(2) - (s-1) \lambda_1(2)} \cdot \left[\frac{\lambda_0(2) - \lambda_1(2)}{K} \right]
\end{aligned} \tag{A.14}$$

Under LD approximation, the threshold $\gamma(s)$ in (A.7) can be found to get optimum solution for (A.7). It is given by

$$\begin{aligned}
\gamma(s) &= \frac{d}{ds} \bar{\varphi}_0(s) \\
&= \frac{d}{ds} \left(-\frac{s-1}{2} \log \left[\frac{\lambda_1(2)}{\lambda_0(2)} \right] + \frac{1}{2} \log \left[\frac{\lambda_0(2)}{s \lambda_0(2) - (s-1) \lambda_1(2)} \right] \right) \\
&= -\frac{1}{2} \log \left[\frac{\lambda_1(2)}{\lambda_0(2)} \right] + \frac{1}{2} \frac{s \lambda_0(2) - (s-1) \lambda_1(2)}{\lambda_0(2)} \cdot -\frac{\lambda_0(2)(\lambda_0(2) - \lambda_1(2))}{[s \lambda_0(2) - (s-1) \lambda_1(2)]^2} \\
&= -\frac{1}{2} \log \left[\frac{\lambda_1(2)}{\lambda_0(2)} \right] - \frac{1}{2} \left[\frac{(\lambda_0(2) - \lambda_1(2))}{s \lambda_0(2) - (s-1) \lambda_1(2)} \right]
\end{aligned} \tag{A.15}$$

A.3 Normalized Log-Moment Generating Function under Genuine Hypothesis

The normalized log-moment generating function and threshold under Genuine hypothesis is derived similarly.

For Chernoff bound the log-moment generating function under Genuine hypothesis is given by

$$\begin{aligned}
\bar{\varphi}_1(s, K) &= \frac{1}{K} \log (E_1 [e^{sK\Lambda_K}]) \\
&= \frac{(\hat{\mu}_1 - \hat{\mu}_0)^2}{2\lambda_0(1)K} \cdot \left[\frac{s^2\lambda_1(1)}{(s+1)\lambda_0(1) - s\lambda_1(1)} + s \right] \\
&\quad - \frac{s}{2K} \left[\log \frac{\lambda_1(1)}{\lambda_0(1)} + (K-1) \log \frac{\lambda_1(2)}{\lambda_0(2)} \right] \\
&\quad + \frac{1}{2} \left[\log \left(\frac{\lambda_0(1)}{(s+1)\lambda_0(1) - s\lambda_1(1)} \right) + (K-1) \log \left(\frac{\lambda_0(2)}{(s+1)\lambda_0(2) - s\lambda_1(2)} \right) \right]
\end{aligned}$$

The threshold $\gamma(s, K)$ in (A.5) can be derived as follows

$$\begin{aligned}
\gamma &= \frac{d}{ds} \bar{\varphi}_1(s, K) \\
&= \frac{(\hat{\mu}_1 - \hat{\mu}_0)^2}{2\lambda_0(1)K} \cdot \left[\frac{s^2\lambda_0(1)\lambda_1(1) + 2s\lambda_0(1)\lambda_1(1) - s^2\lambda_1(1)^2}{((s+1)\lambda_0(1) - s\lambda_1(1))^2} + 1 \right] \\
&\quad - \frac{1}{2K} \log \frac{\lambda_1(1)}{\lambda_0(1)} - (K-1) \log \frac{\lambda_1(2)}{\lambda_0(2)} \\
&\quad - \frac{1}{2} \frac{1}{((s+1)\lambda_0(1) - s\lambda_1(1))} \left[\frac{\lambda_0(1) - \lambda_1(1)}{K} \right] \\
&\quad - \frac{(K-1)}{2} \frac{1}{((s+1)\lambda_0(2) - s\lambda_1(2))} \left[\frac{\lambda_0(2) - \lambda_1(2)}{K} \right]
\end{aligned}$$

Under LD approximation,

$$\bar{\varphi}_1(s) = -\frac{s-1}{2} \log \left[\frac{\lambda_1(2)}{\lambda_0(2)} \right] + \frac{1}{2} \log \left[\frac{\lambda_0(2)}{(s+1)\lambda_0(2) - s\lambda_1(2)} \right]$$

The threshold γ can be written as

$$\begin{aligned}
\gamma(s) &= \frac{d}{ds} \bar{\varphi}_1(s) \\
&= -\frac{1}{2} \log \left[\frac{\lambda_1(2)}{\lambda_0(2)} \right] - \frac{1}{2} \left[\frac{(\lambda_0(2) - \lambda_1(2))}{(s+1)\lambda_0(2) - s\lambda_1(2)} \right]
\end{aligned}$$

Appendix B

Derivation for Chernoff Rate Function: Identification Case

The rate function under the Imposter hypothesis for identification case is given by

$$I_0(K, \underline{\gamma}) = \sup_{\mathbf{s}} [\langle \mathbf{s}, \underline{\gamma} \rangle - \bar{\varphi}_0(\mathbf{s}, K)]$$

where $\langle \cdot, \cdot \rangle$ denotes the vector inner product and $\bar{\varphi}_0(\mathbf{s}, K)$ is the normalized log-moment generating function under Imposter hypothesis defined as

$$\bar{\varphi}_0(\mathbf{s}, K) = \frac{1}{K} \log (E_0 [e^{K \langle \mathbf{s}, \mathbf{\Lambda}_K \rangle}])$$

Plugging in the vector of log-likelihood ratio $\mathbf{\Lambda}_K = \begin{bmatrix} \Lambda_K(1) & \Lambda_K(2) & \dots & \Lambda_K(M) \end{bmatrix}_{M \times 1}^T$, $\mathbf{s} = \begin{bmatrix} s_1 & s_2 & \dots & s_M \end{bmatrix}^T$, and taking the inner product, we write the normalized log-moment generating function as follows

$$\begin{aligned} \bar{\varphi}_0(\mathbf{s}, K) &= \frac{1}{K} \log [E_0 [e^{K(s_1 \Lambda_K(1) + s_2 \Lambda_K(2) + \dots + s_M \Lambda_K(M))}]] \\ &= \frac{1}{K} \log \left[\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{K(s_1 \Lambda_K(1) + s_2 \Lambda_K(2) + \dots + s_M \Lambda_K(M))} \cdot P_0(\mathbf{d}) \cdot d\mathbf{d} \right] \\ &= \frac{1}{K} \log \left[\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{K(s_1 \Lambda_K(1) + s_2 \Lambda_K(2) + \dots + s_M \Lambda_K(M))} \cdot \frac{e^{-\frac{1}{2}[(\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\mathbf{\Sigma}}_0^{-1} (\mathbf{d} - \hat{\mathbf{m}}_0)]}}{(2\pi)^{M/2} \cdot (\det \hat{\mathbf{\Sigma}}_0)^{\frac{1}{2}}} \cdot d\mathbf{d} \right] \\ &= \frac{1}{K} \log \left[\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{e^{[K(s_1 \Lambda_K(1) + s_2 \Lambda_K(2) + \dots + s_M \Lambda_K(M)) - \frac{1}{2}[(\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\mathbf{\Sigma}}_0^{-1} (\mathbf{d} - \hat{\mathbf{m}}_0)] - \frac{1}{2} \log \det \hat{\mathbf{\Sigma}}_0]} (2\pi)^{M/2}} \cdot d\mathbf{d} \right] \end{aligned}$$

Consider the expression in the exponential

$$\begin{aligned}
& K(s_1\Lambda_K(1) + s_2\Lambda_K(2) + \dots + s_M\Lambda_K(M)) - \frac{1}{2} \left[(\mathbf{d} - \hat{\mathbf{m}}_0)^T \hat{\Sigma}_0^{-1} (\mathbf{d} - \hat{\mathbf{m}}_0) \right] - \frac{1}{2} \log \det \hat{\Sigma}_0 \\
&= -\frac{s_1}{2} (\mathbf{d} - \hat{\mathbf{m}}_1)^T \cdot \hat{\Sigma}_1^{-1} \cdot (\mathbf{d} - \hat{\mathbf{m}}_1) + \frac{s_1}{2} (\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0^{-1} \cdot (\mathbf{d} - \hat{\mathbf{m}}_0) - \frac{s_1}{2} \log \det \hat{\mathbf{m}}_1 \hat{\mathbf{m}}_0^{-1} \\
&\quad - \frac{s_2}{2} (\mathbf{d} - \hat{\mathbf{m}}_2)^T \cdot \hat{\Sigma}_2^{-1} \cdot (\mathbf{d} - \hat{\mathbf{m}}_2) + \frac{s_2}{2} (\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0^{-1} \cdot (\mathbf{d} - \hat{\mathbf{m}}_0) - \frac{s_2}{2} \log \det \hat{\mathbf{m}}_2 \hat{\mathbf{m}}_0^{-1} \\
&\quad - \dots - \frac{s_M}{2} (\mathbf{d} - \hat{\mathbf{m}}_M)^T \cdot \hat{\Sigma}_M^{-1} \cdot (\mathbf{d} - \hat{\mathbf{m}}_M) + \frac{s_M}{2} (\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0^{-1} \cdot (\mathbf{d} - \hat{\mathbf{m}}_0) \\
&\quad - \frac{s_M}{2} \log \det \hat{\mathbf{m}}_M \hat{\mathbf{m}}_0^{-1} - \frac{1}{2} (\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0^{-1} (\mathbf{d} - \hat{\mathbf{m}}_0) - \frac{1}{2} \log \det \hat{\Sigma}_0 \\
&= -\frac{1}{2} \mathbf{d}^T \left[s_1 (\hat{\Sigma}_1^{-1} - \hat{\Sigma}_0^{-1}) + s_2 (\hat{\Sigma}_2^{-1} - \hat{\Sigma}_0^{-1}) + \dots + s_M (\hat{\Sigma}_M^{-1} - \hat{\Sigma}_0^{-1}) + \hat{\Sigma}_0^{-1} \right] \mathbf{d} \\
&\quad + \frac{1}{2} \mathbf{d}^T \left[s_1 (\hat{\Sigma}_1^{-1} \hat{\mathbf{m}}_1 - \hat{\Sigma}_0^{-1} \hat{\mathbf{m}}_0) + \dots + s_M (\hat{\Sigma}_M^{-1} \hat{\mathbf{m}}_M - \hat{\Sigma}_0^{-1} \hat{\mathbf{m}}_0) + \hat{\Sigma}_0^{-1} \hat{\mathbf{m}}_0 \right] \mathbf{d} \tag{B.1} \\
&\quad + \frac{1}{2} \mathbf{d}^T \left[s_1 (\hat{\mathbf{m}}_1^T \hat{\Sigma}_1^{-1} - \hat{\mathbf{m}}_0^T \hat{\Sigma}_0^{-1}) + \dots + s_M (\hat{\mathbf{m}}_M^T \hat{\Sigma}_M^{-1} - \hat{\mathbf{m}}_0^T \hat{\Sigma}_0^{-1}) + \hat{\mathbf{m}}_0^T \hat{\Sigma}_0^{-1} \right] \mathbf{d} \\
&\quad - \frac{1}{2} \left[s_1 \log \det \hat{\Sigma}_1 \hat{\Sigma}_0^{-1} + s_2 \log \det \hat{\Sigma}_2 \hat{\Sigma}_0^{-1} + \dots + s_M \log \det \hat{\Sigma}_M \hat{\Sigma}_0^{-1} + \log \det \hat{\Sigma}_0 \right] \\
&\quad - \frac{1}{2} \left[s_1 (\hat{\mathbf{m}}_1^T \hat{\Sigma}_1^{-1} \hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_0^T \hat{\Sigma}_0^{-1} \hat{\mathbf{m}}_0) + \dots + s_M (\hat{\mathbf{m}}_M^T \hat{\Sigma}_M^{-1} \hat{\mathbf{m}}_M - \hat{\mathbf{m}}_0^T \hat{\Sigma}_0^{-1} \hat{\mathbf{m}}_0) + \hat{\mathbf{m}}_0^T \hat{\Sigma}_0^{-1} \hat{\mathbf{m}}_0 \right] \\
&= -\frac{1}{2} \left[(\mathbf{d} - \Delta)^T \cdot \mathbf{Q}^{-1} \cdot (\mathbf{d} - \Delta) \right] + \frac{1}{2} (\Delta^T \cdot \Delta) \\
&\quad - \frac{1}{2} \left[s_1 (\hat{\mathbf{m}}_1^T \hat{\Sigma}_1^{-1} \hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_0^T \hat{\Sigma}_0^{-1} \hat{\mathbf{m}}_0) + \dots + s_M (\hat{\mathbf{m}}_M^T \hat{\Sigma}_M^{-1} \hat{\mathbf{m}}_M - \hat{\mathbf{m}}_0^T \hat{\Sigma}_0^{-1} \hat{\mathbf{m}}_0) + \hat{\mathbf{m}}_0^T \hat{\Sigma}_0^{-1} \hat{\mathbf{m}}_0 \right] \\
&\quad - \frac{1}{2} \left[s_1 \log \det \hat{\Sigma}_1 \hat{\Sigma}_0^{-1} + s_2 \log \det \hat{\Sigma}_2 \hat{\Sigma}_0^{-1} + \dots + s_M \log \det \hat{\Sigma}_M \hat{\Sigma}_0^{-1} + \log \det \hat{\Sigma}_0 \right]
\end{aligned}$$

where

$$\Delta = \mathbf{Q} \left[s_1 (\hat{\Sigma}_1^{-1} \hat{\mathbf{m}}_1 - \hat{\Sigma}_0^{-1} \hat{\mathbf{m}}_0) + \dots + s_M (\hat{\Sigma}_M^{-1} \hat{\mathbf{m}}_M - \hat{\Sigma}_0^{-1} \hat{\mathbf{m}}_0) + \hat{\Sigma}_0^{-1} \hat{\mathbf{m}}_0 \right]$$

and

$$\mathbf{Q} = \left[s_1 (\hat{\Sigma}_1^{-1} - \hat{\Sigma}_0^{-1}) + s_2 (\hat{\Sigma}_2^{-1} - \hat{\Sigma}_0^{-1}) + \dots + s_M (\hat{\Sigma}_M^{-1} - \hat{\Sigma}_0^{-1}) + \hat{\Sigma}_0^{-1} \right]^{-1}$$

Integrating the above expression in (B.1) over all values of \mathbf{d} , and solving,

$$\begin{aligned}
\bar{\varphi}_0(\mathbf{s}, K) &= \frac{1}{K} \log \left(E_0 \left[e^{K(s_1 \Lambda_K(1) + s_2 \Lambda_K(2) + \dots + s_M \Lambda_K(M))} \right] \right) \\
&= \frac{1}{K} \left[\frac{1}{2} (\boldsymbol{\Delta}^T \cdot \mathbf{Q}^{-1} \cdot \boldsymbol{\Delta}) \right] \\
&\quad - \frac{1}{2K} \left[s_1 (\hat{\mathbf{m}}_1^T \hat{\boldsymbol{\Sigma}}_1^{-1} \hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_0^T \hat{\boldsymbol{\Sigma}}_0^{-1} \hat{\mathbf{m}}_0) + \dots + s_M (\hat{\mathbf{m}}_M^T \hat{\boldsymbol{\Sigma}}_M^{-1} \hat{\mathbf{m}}_M - \hat{\mathbf{m}}_0^T \hat{\boldsymbol{\Sigma}}_0^{-1} \hat{\mathbf{m}}_0) + \hat{\mathbf{m}}_0^T \hat{\boldsymbol{\Sigma}}_0^{-1} \hat{\mathbf{m}}_0 \right] \\
&\quad - \frac{1}{2K} \left[s_1 \log \det \hat{\boldsymbol{\Sigma}}_1 \hat{\boldsymbol{\Sigma}}_0^{-1} + \dots + s_M \log \det \hat{\boldsymbol{\Sigma}}_M \hat{\boldsymbol{\Sigma}}_0^{-1} + \log \det \hat{\boldsymbol{\Sigma}}_0 \mathbf{Q}^{-1} \right] \\
&= \frac{1}{2K} (\boldsymbol{\Delta}^T \cdot \mathbf{Q}^{-1} \cdot \boldsymbol{\Delta}) \\
&\quad - \frac{1}{2K} \left[\sum_{i=1}^M s_i (\hat{\mathbf{m}}_i^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\mathbf{m}}_i - \hat{\mathbf{m}}_0^T \hat{\boldsymbol{\Sigma}}_0^{-1} \hat{\mathbf{m}}_0) + \hat{\mathbf{m}}_0^T \hat{\boldsymbol{\Sigma}}_0^{-1} \hat{\mathbf{m}}_0 \right] \\
&\quad - \frac{1}{2K} \left[\sum_{i=1}^M s_i \log \det (\hat{\boldsymbol{\Sigma}}_i \hat{\boldsymbol{\Sigma}}_0^{-1}) + \log \det (\hat{\boldsymbol{\Sigma}}_0 \mathbf{Q}^{-1}) \right]
\end{aligned}$$

Appendix C

Asymptotic Expected Values of Likelihood Function

C.1 Under Imposter Hypothesis

Denote by L_0 the asymptotic expected value of the likelihood function under Imposter hypothesis. It is defined as

$$L_0 = E_0 [\Lambda_K] \quad (\text{C.1})$$

where E_0 is the expectation under the Imposter hypothesis and Λ_K is the likelihood function in (A.2).

Substituting the likelihood function in (C.1) and solving the expectation we get

$$\begin{aligned} L_0 &= E_0 \left[-\frac{1}{2K} (\mathbf{d} - \hat{\mathbf{m}}_1)^T \cdot \hat{\Sigma}_1 \cdot (\mathbf{d} - \hat{\mathbf{m}}_1) + \frac{1}{2K} (\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0 \cdot (\mathbf{d} - \hat{\mathbf{m}}_0) - \frac{1}{2K} \log \det \hat{\Sigma}_1 \hat{\Sigma}_0^{-1} \right] \\ &= -\frac{1}{2K} \log \det \hat{\Sigma}_1 \hat{\Sigma}_0^{-1} + E_0 \left[-\frac{1}{2K} (\mathbf{d} - \hat{\mathbf{m}}_1)^T \cdot \hat{\Sigma}_1 \cdot (\mathbf{d} - \hat{\mathbf{m}}_1) + \frac{1}{2K} (\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0 \cdot (\mathbf{d} - \hat{\mathbf{m}}_0) \right] \\ &= -\frac{1}{2K} \log \det \hat{\Sigma}_1 \hat{\Sigma}_0^{-1} + E_0 \left[-\frac{1}{2K} (\mathbf{d} - \hat{\mathbf{m}}_1)^T \cdot \hat{\Sigma}_1 \cdot (\mathbf{d} - \hat{\mathbf{m}}_1) \right] + \\ &\quad E_0 \left[\frac{1}{2K} (\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0 \cdot (\mathbf{d} - \hat{\mathbf{m}}_0) \right] \end{aligned}$$

Consider two expectations in (C.2) separately.

$$\begin{aligned}
& E_0 \left[-\frac{1}{2K} (\mathbf{d} - \hat{\mathbf{m}}_1)^T \cdot \hat{\Sigma}_1 \cdot (\mathbf{d} - \hat{\mathbf{m}}_1) \right] \\
&= -\frac{1}{2K} E_0 \left[(\mathbf{d} - \hat{\mathbf{m}}_0 + \hat{\mathbf{m}}_0 - \hat{\mathbf{m}}_1) \cdot \hat{\Sigma}_1 \cdot (\mathbf{d} - \hat{\mathbf{m}}_0 + \hat{\mathbf{m}}_0 - \hat{\mathbf{m}}_1) \right] \\
&= -\frac{1}{2K} \left(E_0 \left[(\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_1 \cdot (\mathbf{d} - \hat{\mathbf{m}}_0) \right] + E_0 \left[(\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_1 \cdot (\hat{\mathbf{m}}_0 - \hat{\mathbf{m}}_1) \right] \right) \\
&\quad - \frac{1}{2K} \left(E_0 \left[(\hat{\mathbf{m}}_0 - \hat{\mathbf{m}}_1)^T \cdot \hat{\Sigma}_1 \cdot (\mathbf{d} - \hat{\mathbf{m}}_0) \right] + E_0 \left[(\hat{\mathbf{m}}_0 - \hat{\mathbf{m}}_1)^T \cdot \hat{\Sigma}_1 \cdot (\hat{\mathbf{m}}_0 - \hat{\mathbf{m}}_1) \right] \right) \\
&= -\frac{1}{2K} \text{tr} \left(\hat{\Sigma}_0 \hat{\Sigma}_1^{-1} \right) - \frac{1}{2K} (\hat{\mathbf{m}}_0 - \hat{\mathbf{m}}_1)^T \cdot \hat{\Sigma}_1 \cdot (\hat{\mathbf{m}}_0 - \hat{\mathbf{m}}_1)
\end{aligned}$$

And

$$\begin{aligned}
& E_0 \left[\frac{1}{2K} (\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0 \cdot (\mathbf{d} - \hat{\mathbf{m}}_0) \right] \\
&= \frac{1}{2K} E_0 \left[(\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0 \cdot (\mathbf{d} - \hat{\mathbf{m}}_0) \right] \\
&= \frac{1}{2K} \text{tr} \left(\hat{\Sigma}_0 \hat{\Sigma}_0^{-1} \right) \\
&= 1
\end{aligned}$$

Therefore the asymptotic expected value of likelihood function under Imposter hypothesis is given by

$$L_0 = -\frac{1}{2K} \log(\det(\hat{\Sigma}_1 \hat{\Sigma}_0^{-1})) - \frac{1}{2K} \text{tr} \left(\hat{\Sigma}_0 \hat{\Sigma}_1^{-1} \right) - \frac{1}{2K} (\hat{\mathbf{m}}_0 - \hat{\mathbf{m}}_1)^T \cdot \hat{\Sigma}_1 \cdot (\hat{\mathbf{m}}_0 - \hat{\mathbf{m}}_1) + 1$$

C.2 Under Genuine Hypothesis

The asymptotic expected value of a likelihood function under Genuine hypothesis is calculated in a similar way as under Imposter hypothesis. Denote by L_1 the asymptotic expected value of the likelihood function under Genuine hypothesis. It is defined as

$$L_1 = E_1 [\Lambda_K] \tag{C.2}$$

where E_1 is the expectation under the Genuine hypothesis and Λ_K is the likelihood function in (A.2).

Substituting the likelihood function in (C.2) and solving the expectation we get

$$\begin{aligned}
L_1 &= E_1 \left[-\frac{1}{2K}(\mathbf{d} - \hat{\mathbf{m}}_1)^T \cdot \hat{\Sigma}_1 \cdot (\mathbf{d} - \hat{\mathbf{m}}_1) + \frac{1}{2K}(\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0 \cdot (\mathbf{d} - \hat{\mathbf{m}}_0) - \frac{1}{2K} \log \det \hat{\Sigma}_1 \hat{\Sigma}_0^{-1} \right] \\
&= -\frac{1}{2K} \log \det \hat{\Sigma}_1 \hat{\Sigma}_0^{-1} + E_1 \left[-\frac{1}{2K}(\mathbf{d} - \hat{\mathbf{m}}_1)^T \cdot \hat{\Sigma}_1 \cdot (\mathbf{d} - \hat{\mathbf{m}}_1) + \frac{1}{2K}(\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0 \cdot (\mathbf{d} - \hat{\mathbf{m}}_0) \right] \\
&= -\frac{1}{2K} \log \det \hat{\Sigma}_1 \hat{\Sigma}_0^{-1} + E_1 \left[-\frac{1}{2K}(\mathbf{d} - \hat{\mathbf{m}}_1)^T \cdot \hat{\Sigma}_1 \cdot (\mathbf{d} - \hat{\mathbf{m}}_1) \right] + \\
&\quad E_1 \left[\frac{1}{2K}(\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0 \cdot (\mathbf{d} - \hat{\mathbf{m}}_0) \right]
\end{aligned}$$

Consider two expectations in (C.3) separately.

$$\begin{aligned}
&E_1 \left[-\frac{1}{2K}(\mathbf{d} - \hat{\mathbf{m}}_1)^T \cdot \hat{\Sigma}_1 \cdot (\mathbf{d} - \hat{\mathbf{m}}_1) \right] \\
&= -\frac{1}{2K} E_1 \left[(\mathbf{d} - \hat{\mathbf{m}}_1)^T \cdot \hat{\Sigma}_1 \cdot (\mathbf{d} - \hat{\mathbf{m}}_1) \right] \\
&= -\frac{1}{2K} \text{tr} \left(\hat{\Sigma}_1 \hat{\Sigma}_0^{-1} \right) \\
&= -1
\end{aligned}$$

And

$$\begin{aligned}
&E_1 \left[\frac{1}{2K}(\mathbf{d} - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0 \cdot (\mathbf{d} - \hat{\mathbf{m}}_0) \right] \\
&= \frac{1}{2K} E_1 \left[(\mathbf{d} - \hat{\mathbf{m}}_1 + \hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0 \cdot (\mathbf{d} - \hat{\mathbf{m}}_1 + \hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_0) \right] \\
&= \frac{1}{2K} \left(E_1 \left[(\mathbf{d} - \hat{\mathbf{m}}_1)^T \cdot \hat{\Sigma}_0 \cdot (\mathbf{d} - \hat{\mathbf{m}}_1) \right] + E_1 \left[(\mathbf{d} - \hat{\mathbf{m}}_1)^T \cdot \hat{\Sigma}_0 \cdot (\hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_0) \right] \right) \\
&\quad + \frac{1}{2K} \left(E_1 \left[(\hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0 \cdot (\mathbf{d} - \hat{\mathbf{m}}_1) \right] + E_1 \left[(\hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0 \cdot (\hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_0) \right] \right) \\
&= \frac{1}{2K} \text{tr} \left(\hat{\Sigma}_1 \hat{\Sigma}_0^{-1} \right) + \frac{1}{2K} (\hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0 \cdot (\hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_0)
\end{aligned}$$

Therefore the asymptotic expected value of likelihood function under Genuine hypothesis is given by

$$L_1 = -\frac{1}{2K} \log(\det(\hat{\Sigma}_1 \hat{\Sigma}_0^{-1})) + \frac{1}{2K} \text{tr} \left(\hat{\Sigma}_1 \hat{\Sigma}_0^{-1} \right) + \frac{1}{2K} (\hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_0)^T \cdot \hat{\Sigma}_0 \cdot (\hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_0) - 1$$