



Graduate Theses, Dissertations, and Problem Reports

2006

Recognition capacity of biometric-based systems

Francesco P. Nicolo
West Virginia University

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

Recommended Citation

Nicolo, Francesco P., "Recognition capacity of biometric-based systems" (2006). *Graduate Theses, Dissertations, and Problem Reports*. 1780.
<https://researchrepository.wvu.edu/etd/1780>

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.

Recognition Capacity of Biometric-Based Systems

Francesco P. Nicolò

A 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
Matthew C. Valenti, Ph.D.
Bojan Cukic, Ph.D.

Lane Department of Computer Science
and Electrical Engineering

Morgantown, West Virginia
2006

Keywords: Biometrics, Information theory, Stochastic Model, Goodness of Fit,
Capacity

© Copyright 2006 by Francesco P. Nicolò
All Rights Reserved

ABSTRACT

Recognition Capacity of Biometric-Based Systems

Francesco P. Nicolò

Performance of biometrics-based recognition systems depends on various factors: database quality, image preprocessing, encoding techniques, etc. Given a biometric database and a selected encoding method, the capability of a recognition system is limited by the relationship between the number of classes that the recognition system can encode and the length of encoded data describing the template at a specific level of distortion. In this work, we evaluate constrained recognition capacity of biometric systems under the constraint of two global encoding techniques: Principal Component Analysis and Independent Component Analysis. The developed methodology is applied to predict capacity of different recognition channels formed during acquisition of different iris and face databases. The proposed approach relies on data modeling and involves classical detection and information theories. The major contribution is in providing a guideline on how to evaluate capabilities of large-scale biometric recognition systems in practice. Recognition capacity can also be promoted as a global quality measure of biometric databases.

Acknowledgements

Firstly, I would like to say thank you to my advisor and committee chair, *Dr. Natalia Schmid*. Her support and guidance from the first day I joined the lab has been inestimable. I learned so much from her. Her dedication for work, her brilliant ideas but especially her values and principles make her an exemplar mentor.

My graduate Committee members, *Dr. Bojan Cukic* and *Matthew C. Valenti* for their expert advice and support.

To my father and my mother for always supporting my choices.

To *Zio Enrico*, *Zia Judy* my family in the USA, without their help I couldn't have been here.

To *Xiaohan*, *Jinyu*, *Nathan*, *Shan Shan* for their support and discussion which helped me so much on my research.

To all my close friends, their support during difficult time have been precious.

Contents

Acknowledgements	iii
1 Introduction	1
1.1 Motivation	1
1.2 Performance Analysis of Recognition Systems	2
1.3 Contributions	3
1.4 Organization	3
2 Recognition Capacity	4
2.1 Biometric-Based Recognition	4
2.2 Recognition Capacity	5
3 PCA Representation Rate	9
3.1 PCA Encoding	9
3.2 Asymptotic Behavior of Eigenvalues	10
3.3 Model for PCA encoded data	11
3.3.1 Gaussian Model	11
3.3.2 Database Descriptions	13
3.3.3 Model Verification: Shapiro-Wilk Test	14
3.4 PCA-Based Recognition Capacity	15
3.5 Recognition Capacity vs Resolution	21
4 ICA Representation rate	23
4.1 ICA based encoding	23

4.2	Model for ICA encoded data	24
4.2.1	Bessel K Model	24
4.2.2	Model Verification: K-S and K-L Test	27
4.3	ICA-Based Recognition Capacity	35
4.4	ICA-Based Recognition Capacity vs PCA-Based Recognition Capacity	36
5	Conclusion	37
6	Future Work	39
	Appendices	40
A	Goodness of fit	40
A.1	Hypothesis Testing and Test Statistics	40
A.2	Critical value and p value	41
B	Bessel K Distribution	42
	Bibliography	44

List of Tables

3.1	Shapiro-Wilk test results for Biometric Databases	15
3.2	Recognition Capacity of Biometric Databases using PCA encoded data	21
4.1	Estimated parameters from empirical data	31
4.2	K-S test results for the biometrics databases	33
4.3	K-L test result for the biometrics databases	34
4.4	Recognition Capacity of Biometric Databases using ICA encoded data	35
4.5	Recognition Capacity of Biometric Databases using ICA and PCA en- coded data	36

List of Figures

2.1	A block-diagram of a biometrics-based recognition system	6
2.2	Constellation of the templates in White Gaussian Noise Channel	7
3.1	Dependency of n -th Eigenvalue on M . The results are obtained using CASIA1 iris database	11
3.2	Samples of Iris Biometrics from the CASIA1, WVU, BATH, CASIA3 databases	14
3.3	Samples of Face Biometrics from FRGC database	14
3.4	n -th Eigenvalue vs the number of training images (that is the number of users M) for CASIA1 DATABASE (linear scale)	16
3.5	n -th Eigenvalue vs the number of training images for CASIA1 DATABASE (log scale)	16
3.6	n -th Eigenvalue vs the number of training images for WVU DATABASE (linear scale)	17
3.7	n -th Eigenvalue vs the number of training images for WVU DATABASE (log scale)	17
3.8	n -th Eigenvalue vs the number of training images for BATH DATABASE (linear scale)	18
3.9	n -th Eigenvalue vs the number of training images for BATH DATABASE (log scale)	18
3.10	n -th Eigenvalue vs the number of training for CASIA3 (DEV1) DATABASE (linear scale)	19
3.11	n -th Eigenvalue vs the number of training images for CASIA3 (DEV1) DATABASE (log scale)	19

3.12	n-th Eigenvalue vs the number of training images for FRGC DATABASE (linear scale)	20
3.13	n-th Eigenvalue vs the number of training images for FRGC DATABASE (log scale)	20
3.14	Capacity vs reciprocal of SNR ($\bar{\gamma}^{-1}$) for CASIA1 Iris Database . . .	22
4.1	Empirical pdf of ICA encoded Iris Images from CASIA-1 Database and the parametric distribution providing the best fit (Kurtosis of empirical pdf = 83.7712)	28
4.2	Empirical pdf of ICA encoded Iris Images from WVU Database and the parametric distribution providing the best fit (Kurtosis of empirical pdf = 91.7742)	29
4.3	Empirical pdf of ICA encoded Iris Images from BATH Database and the parametric distribution providing the best fit (Kurtosis of empirical pdf = 37.1320)	29
4.4	Empirical pdf of ICA encoded Iris Images from CASIA3 (DEV1) Database and the parametric distribution providing the best fit (Kurtosis of em- pirical pdf = 51.8308)	30
4.5	Empirical pdf of ICA encoded Face Images from FRGC Database and the parametric distribution providing the best fit (Kurtosis of empirical pdf = 98.8771)	30
4.6	KS Test for CASIA IRIS DATABASE	32
4.7	KS Test for WVU IRIS DATABASE	32
4.8	KS Test for BATH IRIS DATABASE	32
4.9	KS Test for CASIA3 (DEV1) IRIS DATABASE	32
4.10	KS Test for FRGC FACE DATABASE	33
4.11	ICA-Based and PCA-Based Recognition Capacity of Iris Databases .	36

Chapter 1

Introduction

1.1 Motivation

In many large scale biometrics-based recognition problems, knowledge of the limiting capabilities of underlying recognition systems is critical. These limits, however, are determined by a variety of factors including source coding techniques used to process data, quality, complexity, and variability of the collected data. Given an encoding technique, the remaining factors can be attributed to a recognition channel introduced and characterized by O’Sullivan and Schmid in [1] and further analyzed by Westover and O’Sullivan [2]. Similar to a communication channel, a recognition channel is characterized by its capacity, with the difference being recognition capacity. In a biometrics-based recognition problem, recognition capacity can be thought as being the maximum number of classes that can be successfully recognized asymptotically with probability close to zero when the number of informative samples gets large. Thus, capacity can be viewed as a measure of performance that can be used to evaluate capabilities of large scale recognition systems. Also, since the maximum number of biometric classes that can be successfully recognized is directly related to distortions and noise present in the images or signals submitted for recognition, we propose to treat capacity as a measure of overall quality of data in a given database. In this thesis, we briefly summarize the results by Schmid and O’Sullivan on recognition capacity, introduce a concept of constrained capacity, and evaluate

the constrained capacity for specific iris and face-based recognition problems using 5 public databases. To evaluate capacity of a biometric-based recognition channel we first formulate stochastic model that describe data and then involve fundamentals of classical detection and information theories. Since the computation of the capacity of a biometric channel requires knowledge of a model characterizing encoded data, before computing capacity we validate the fit of proposed models by using statistical goodness tests.

1.2 Performance Analysis of Recognition Systems

An early general approach for prediction of performance of Recognition systems was introduced by O’Sullivan and Schmid [3]. In this work the concepts of physical signature like random process was introduced. An unknown physical signature is authenticated if it is statistically correlated with a set of known signatures. The problem of authentication is stated as a binary hypothesis testing. The prediction of performance is based on application of Chernoff information and Large Deviation.

In a successive work of Schmid and O’Sullivan [1], a recognition system is compared to a maximum likelihood decoder that recovers the original signal transmitted into a noisy Gaussian channel. Classical information theory concepts are used to derive an upperbound to the error recognition exponent and the concept of constrained capacity is introduced as well. The value of capacity provides the asymptotic recognition limits of the system under practical constrains.

The work by Westover and O’Sullivan [2] continues the preceding work analyzing the tradeoff between the complexity of a signature representation and the complexity of the environment for reliable communication of that signature. It is shown that the tradeoff can be characterized in terms of three rates related to the number of bits available for representing memory, sensory data, and the number of patterns populating a given statistical environment.

In ATR applications, other examples are given by Shusterman et al. [4] that proposes a rate-distortion framework based on Hilbert-Schmidt norm for evaluation of the codebook (target library) size. The analyzed data are in the form of computer-aided design model.

Other few results on performance analysis for various ATR systems are given by Srivastava et al. [5] that suggests use of Laplace approximation method for solving an integral for Bayesian probability. The asymptotic analysis relies on vanishing value of noise variance, which results into overoptimistic limits.

1.3 Contributions

The main contributions of this work are: (1) in designing a methodology to predict capabilities of biometrics-based recognition system by developing and evaluating models characterizing encoded biometric data and (2) in introducing a global quality measure of biometrics databases.

1.4 Organization

In Chapter 2 we introduce the concepts of recognition capacity and channel. In Chapter 3 we describe the model for PCA encoding of iris and face biometric; numerical results for capacity are also provided. Chapter 4 will be focused on BesselK model and its validation for ICA encoded data. Numerical results for the capacity are produced as well. Conclusions and Future Work are provided in Chapter 5 and Chapter 6.

Chapter 2

Recognition Capacity

2.1 Biometric-Based Recognition

Biometrics-based recognition systems are a special case of a more general recognition systems. Biometric recognition systems operate in two modes: enrollment and matching. During enrollment mode, an image or signal containing information about an individual's biometrics is acquired, encoded and stored in a database.

During matching mode, an image or signal containing information about a biometrics is encoded and compared against each encoded biometrics stored in the database. Encoded biometrics-based information is called a template. A template from the database that has most similarity with the template submitted for recognition represents the identified biometric class.

In the following, both images, signals, and templates containing information about a biometrics are considered to be realizations of random processes underlying generation of the biometrics.

Consider $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_M$ n -dimensional templates of M (M is the number of individuals in the database) distinct biometric vectors contained in a biometric database. Assume that the underlying biometric vectors are independent and identically distributed.

Denote by \underline{Y} an n -dimensional biometric template submitted for recognition. We state the problem of identification as $(M + 1)$ -hypothesis testing problem [6].

We assume that \underline{Y} is recognized as coming from class k , $k = 1, \dots, M$ if the template \underline{X}_k and \underline{Y} have a biometric signature in common and thus a joint probability distribution, $P_{\underline{X}_k, \underline{Y}}$. The remaining templates do not have signatures in common and thus are described by a product distribution $P_{\underline{X}_m} \times P_{\underline{Y}}$.

When \underline{Y} is submitted for identification, to conclude on similarities with templates we test the following $(M + 1)$ hypotheses:

$$\left\{ \begin{array}{l} H_o : \underline{X}_k, k = 1, \dots, M \text{ and } \underline{Y} \text{ are independent.} \\ H_1 : \underline{X}_1 \text{ and } \underline{Y} \text{ have signature in common;} \\ \quad \underline{Y} \text{ is independent of all } \underline{X}_k \neq \underline{X}_1. \\ \vdots \\ H_M : \underline{X}_M \text{ and } \underline{Y} \text{ have signature in common;} \\ \quad \underline{X}_i, i = 1, \dots, M - 1 \text{ and } \underline{Y} \text{ are independent.} \end{array} \right.$$

The test statistic for the recognition problem above is a vector of loglikelihood ratios, which in this specific case is called information density [7]:

$$i_n^M = \frac{1}{n} \begin{bmatrix} \log \frac{P_{\underline{X}_1, \underline{Y}}}{P_{\underline{X}_1} \times P_{\underline{Y}}} \\ \vdots \\ \log \frac{P_{\underline{X}_M, \underline{Y}}}{P_{\underline{X}_M} \times P_{\underline{Y}}} \end{bmatrix} \quad (2.1)$$

The decision is made in favor of hypothesis that \underline{X}_M and \underline{Y} have signal in common if the m -th entry results in the maximum value and exceeds a decision threshold, γ . If all entries in (2.1) are below γ , then \underline{Y} is recognized as an imposter.

2.2 Recognition Capacity

In this section, we find the operational capacity for a biometrics-based identification system. This requires both the number of biometric classes in the database and the length of templates grow. The number of classes has to grow exponentially with the length of the observed vector.

The results in this section are based on a straightforward analogy between a

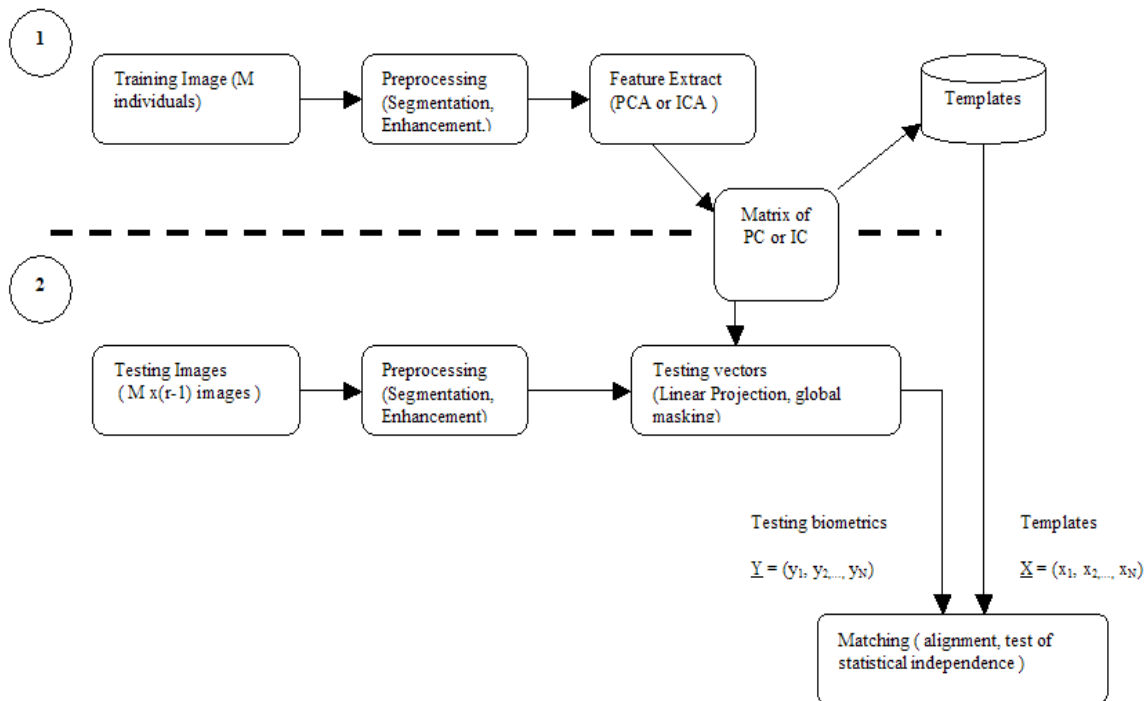


Figure 2.1: A block-diagram of a biometrics-based recognition system

recognition system and a communication system that uses random coding [8]. The templates in the recognition problem correspond to codewords in the communication problem. Suppose that $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_M$, with $M = 2^{nR}$, templates (random codewords) are drawn i.i.d. from the distribution $P_{\underline{X}}$. Here R is the rate of the random code. Given that one of these random codewords is randomly selected as the truth, assume that an observation \underline{Y} is drawn from the corresponding conditional distribution determined by $P_{\underline{X}, \underline{Y}}$, and that this distribution satisfies the assumptions from the previous sections. The observation \underline{Y} is directly analogous to the output of a communication system. The observation \underline{Y} must be matched against each of the 2^{nR} templates stored in the database. This matching process defines the recognition output and is analogous to channel decoding.

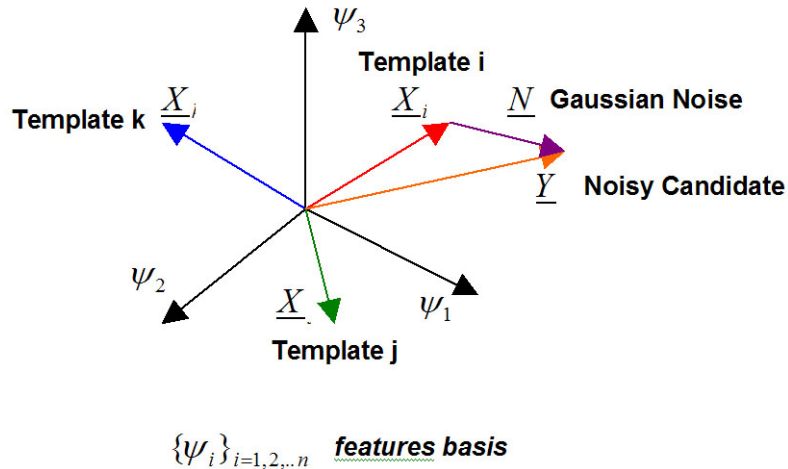


Figure 2.2: Constellation of the templates in White Gaussian Noise Channel

The probability of error is

$$P_e = \frac{1}{M} \sum_{m=1}^M P_{e|\underline{X}_m}, \tag{2.2}$$

where $P_{e|\underline{X}_m}$ is the conditional probability of error given that the codeword \underline{X}_m is sent over the channel. The minimum probability of error decision rule is the same as the maximum likelihood decision rule: select the most likely observation given \underline{X}_m .

A rate R is achievable if a sequence of recognition systems can be defined with

rates $R_n \leq R$ so that the probability of error goes to zero as n goes to infinity. Define the recognition capacity to be the supremum of achievable rates. The direct analogy to communication systems using random codes immediately yields the recognition capacity.

Theorem 1 *Under the assumptions above, the recognition capacity equals the supremum mutual information rate $\bar{I}(X;Y)$. That is, all rates less than $C = \bar{I}(X;Y)$ are achievable (the probability of error using the minimum probability of error decoding rule tends to zero as n tends to infinity); conversely, if rate R is achievable then $R \leq C$.*

In the next two chapters we will provide two examples (capacity of biometric recognition channels under constraints of PCA and ICA data encoding) to illustrate the above theory.

Chapter 3

PCA Representation Rate

In this chapter we will derive an expression for the constrained capacity of noisy biometrics-based recognition channel under the constraint of PCA-encoded data.

3.1 PCA Encoding

Principal Component Analysis is a global coding algorithm that in the field of biometrics is largely applied to face recognition problem [9]. In our experiments we will use PCA to encode both iris and face biometrics. Consider a biometric database with M users and r images per user. M preprocessed images (one per user) $\underline{B}_1, \underline{B}_2, \dots, \underline{B}_M$ are available for training a biometric recognition system that uses PCA encoded data for recognition. These images are encoded and stored in the database in a form of templates.

PCA algorithms requires training, that is estimation of a set of parameters including scatter matrix and its eigenvalues and eigenvectors. The estimate of scatter matrix is given by :

$$\Sigma_Z = \frac{1}{m-1} \sum_{i=1}^m (\underline{B}_i - \overline{\underline{B}})(\underline{B}_i - \overline{\underline{B}}) = \underline{Z}^T \underline{Z}$$

where $\overline{\underline{B}}$ is the sample mean. The matrix Σ_Z is decomposed using an eigenvalue decomposition [6]:

$$\Sigma_Z = Q\Lambda Q^T \quad ,$$

where Λ is the matrix of eigenvalues. Q is the orthogonal matrix with columns composed of eigenvectors of Σ_Z . In practice only the n largest eigenvalues are chosen from the total M . Hence, a new matrix \tilde{Q} with vector columns corresponding to the essential eigenvalues is formed.

The feature vector \underline{X}_i of the i -th user is obtained by projecting image $(\underline{B}_i - \overline{\underline{B}})$ onto columns of matrix \tilde{Q}

$$\underline{X}_i = \tilde{Q}^T (\underline{B}_i - \overline{\underline{B}}) \quad (3.1)$$

Each column of \tilde{Q} represents a vector of a set of bases $\{\tilde{q}_j\}_{j=1,2,\dots,n}$ is assumed to be statistically uncorrelated.

3.2 Asymptotic Behavior of Eigenvalues

The n eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ of the matrix Σ_X have been obtained from the largest n eigenvalues of the matrix Λ such that the sum of all n eigenvalues is about 99 percent of the sum of all M eigenvalues. Experimental tests on our Iris and Face Database show that the values of the eigenvalues decrease as M increases. M indicates the number of images used to train our PCA encoding system. The dependency of $\lambda_1(M), \lambda_2(M), \dots, \lambda_n(M)$ on M is shown in Fig. 3.1:

Note that the eigenvalues λ_i exists only if $M > i$.

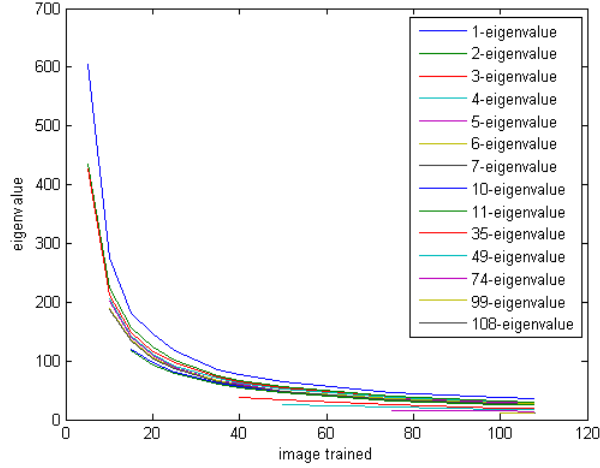


Figure 3.1: Dependency of n-th Eigenvalue on M . The results are obtained using CASIA1 iris database

3.3 Model for PCA encoded data

In this section we propose and validate a Gaussian model for PCA-encoded iris and face data.

3.3.1 Gaussian Model

In Chapter 2 we have introduced the concept of information density and explained how it can be applied to the problem biometric-based recognition. To apply this model to PCA-encoded data we assume that PCA templates \underline{X}_i , $i = 1, 2, \dots, M$ stored in the database are realizations of i.i.d. vector processes. The processes are Gaussian with zero mean and covariance matrix:

$$\Sigma_X = \begin{pmatrix} \lambda_1 & & 0 \\ & \dots & \\ 0 & & \lambda_n \end{pmatrix},$$

where $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of the scatter matrix Σ_Z .

According to the introduced model a random source signal \underline{X}_Y is transmitted

through a white gaussian noise channel. Thus the template submitted for identification is given by:

$$\underline{Y} = \underline{X}_Y + \underline{N},$$

where:

$$\underline{N} \sim \mathcal{N}(0, \sigma^2 I),$$

with σ^2 being noise variance.

In our case the noisy candidate \underline{Y} is one of the images from the testing set. Note that training and testing data do not overlap. The noisy template will be compared against all the templates \underline{X}_i , $i = 1, 2, \dots, M$ in the database.

Note that we assume that \underline{X}_i are i.i.d. The assumption of independency is derived from the fact that each signature belongs to a different individual.

To find constrained recognition capacity we form joint and marginal distribution for concatenated vector $[\underline{X}_i, \underline{Y}]$ and arrive to the following expression for information density:

$$i_n = -\frac{1}{2n} [\underline{X}_i, \underline{Y}] (\mathbf{R}_1^{-1} - \mathbf{R}_0^{-1}) [\underline{X}_i, \underline{Y}]^T - \frac{1}{2n} \log \det(\mathbf{R}_1 \mathbf{R}_0^{-1}) \quad (3.2)$$

where

$$\mathbf{R}_0 = \begin{pmatrix} \Sigma_X & 0 \\ 0 & \Sigma_X + \sigma^2 I \end{pmatrix}$$

is the covariance matrix of the vector $[\underline{X}_i, \underline{Y}]$ when the signature X_i and the noisy candidate \underline{Y} are independent and

$$\mathbf{R}_1 = \begin{pmatrix} \Sigma_X & \Sigma_X \\ \Sigma_X & \Sigma_X + \sigma^2 I \end{pmatrix}$$

is the covariance matrix of the vector $[\underline{X}_i, \underline{Y}]$ when the signature X_i and noise candidate \underline{Y} have a signal in common.

The information rate is the average of the information density under the joint

distribution of $\underline{X}_i, \underline{Y}$:

$$I_n(M) = \frac{1}{n} E_{\underline{X}_i, \underline{Y}} \{i_n(M)\} = \frac{1}{2n} \sum_{k=1}^n \log \left(1 + \frac{\lambda_k(M)}{\sigma^2(M)} \right) \quad (3.3)$$

where $\sigma^2(M)$ is an entry of the diagonal covariance matrix of \underline{N} .

For very large database the number of users $M \rightarrow \infty$ and number of principal components $n \rightarrow \infty$. In this case we have the *PCA-based recognition capacity* or *PCA representation rate* (asymptotic information rate) :

$$\bar{I} = \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{2n} \sum_{k=1}^n \log \left(1 + \frac{\lambda_k(M)}{\sigma^2(M)} \right) \quad (3.4)$$

3.3.2 Database Descriptions

All experiments have been performed on five datasets:

1. CASIA version 1 iris database provided by the Chinese Academy of Science [10]
2. An iris database of images collected at WVU
3. Bath iris database provided by University of Bath [11]
4. CASIA version 3 iris database
5. FRGC version 1 face database used in the Face Recognition Grand Challenge [12]

The CASIA v.1 contains 108 iris classes with 6 images per class. The WVU dataset we used contains 108 iris classes with 5 images per class. The Bath dataset we used contains 50 iris classes with 20 images per class. The CASIA v.3 we used contains 59 iris classes with 20 images per class. The FRGC dataset we used contains 108 face classes with 6 images per class. In figures 3.2, 3.3 samples of biometrics images from the considered databases are shown:

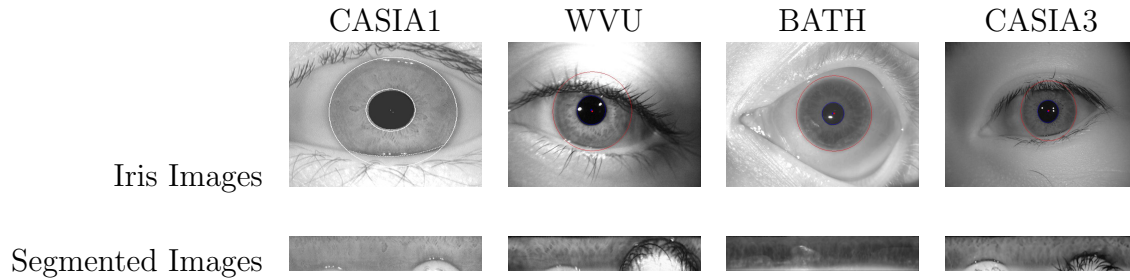


Figure 3.2: Samples of Iris Biometrics from the CASIA1, WVU, BATH, CASIA3 databases

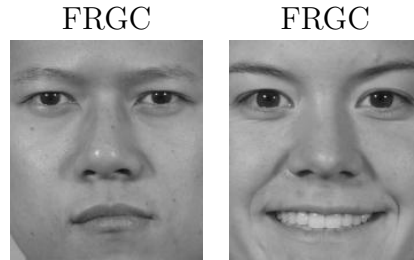


Figure 3.3: Samples of Face Biometrics from FRGC database

3.3.3 Model Verification: Shapiro-Wilk Test

The Shapiro-Wilk [13] test is a statistical method to determine if a sample set \mathcal{S} follows a hypothesis of normality. The test produces the W statistics based on the set \mathcal{S} of size k . The W statistics is here defined:

$$W = \frac{\left(\sum_{i=1}^k w_i x'_i\right)^2}{\sum_{i=1}^k (x_i - \bar{x})^2} \quad (3.5)$$

where x'_i , x_i , \bar{x} are respectively the original data, the ordered data and the sample mean of the set S and w_i are constants derived from the means, variances and covariances of the order statistics of a sample of size k from a standard normal distribution. Small values of W indicate departures of S from normality. The probability of the W statistic being at least as extreme as the one observed, given that the hypothesis of normality is true, is called p-value (see appendix A). A value $p < p_{crit}$ ($p_{crit} = 0.05$)

rejects the hypothesis of normality. For each vector \underline{X}_i , we normalize the n components by the relative n eigenvalues. A set of 1 – 3 samples for each template \underline{X}_i is randomly selected. We concatenate all the selected samples to create a testing set \mathcal{S} of size $k \geq M$. The results of applying Shapiro-Wilk test to each database are reported in the Table 3.1.

Iris Database	p	k (samples/templates)	M (templates)
CASIA	0.8469	1	108
	0.4714	3	108
WVU	0.3880	1	108
	0.1609	3	108
BATH	0.8682	1	50
	0.5300	3	50
CASIA3	0.6294	1	59
	0.1444	3	59
Face Database	p	k (samples/templates)	M (templates)
FRGC	0.7777	1	108
	0.1128	3	108

Table 3.1: Shapiro-Wilk test results for Biometric Databases

For all the considered databases the hypothesis of normality is never rejected.

3.4 PCA-Based Recognition Capacity of Iris and Face Biometric Channels

To find the limits (3.4) we empirically evaluate the ratio of variances. The following plots indicate the dependency of $\lambda_i(M)$ and $\sigma(M)$ on M . The assumption that we make is that the noise \underline{N} is identical distributed for each principal component. The noise is estimated for $r \geq 5$ images per class. From the graph we can note that $\lambda_i(M)$ and $\sigma(M)$ have the same kind of dependency on M and the ratio $\frac{\lambda(M)}{\sigma(M)} = \gamma(M)$ (Signal to Noise Ratio) converges to a fixed value $\bar{\gamma}$ as $M \rightarrow \infty$. This imply that the sum (3.4) converges to an asymptotic value $\bar{I} = \frac{1}{2} \log(1 + \bar{\gamma})$. \bar{I} will be the obtained recognition capacity for PCA encoded data of the considered database. Fig 3.4-3.13 show the

trend of eigenvalues and noise variances as M increases. Note in all considered cases, the limiting value of Signal to Noise Ratio is clearly observed.

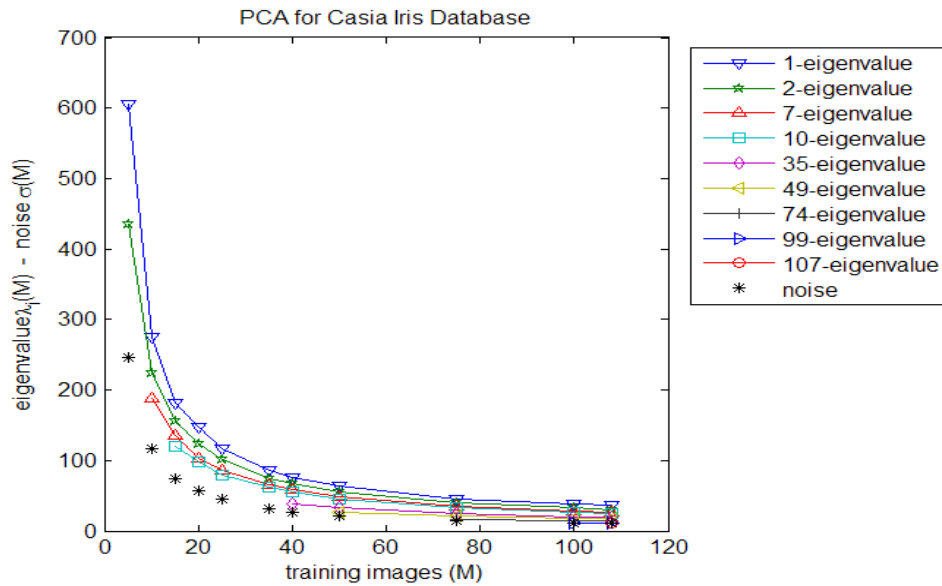


Figure 3.4: n -th Eigenvalue vs the number of training images (that is the number of users M) for CASIA1 DATABASE (linear scale)

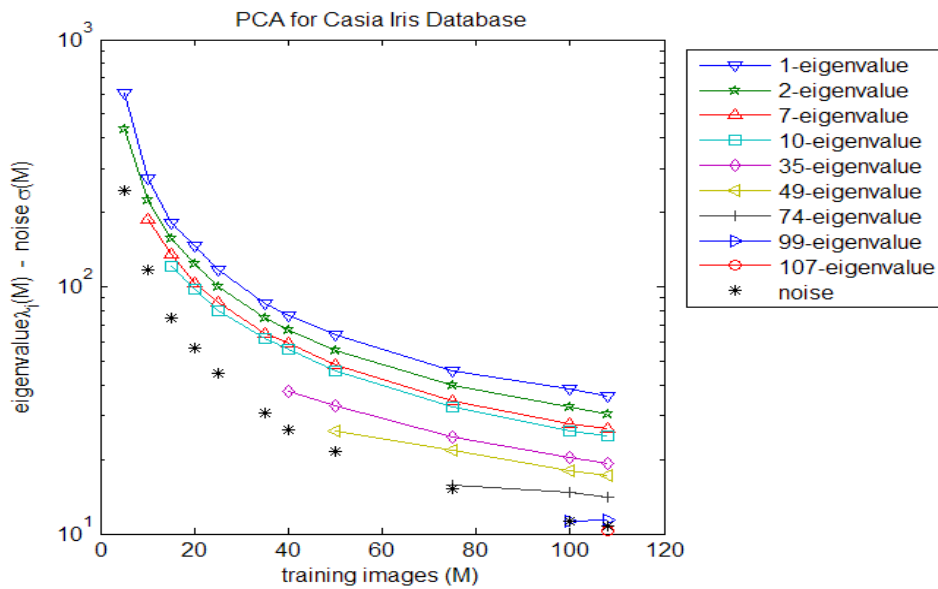


Figure 3.5: n -th Eigenvalue vs the number of training images for CASIA1 DATABASE (log scale)

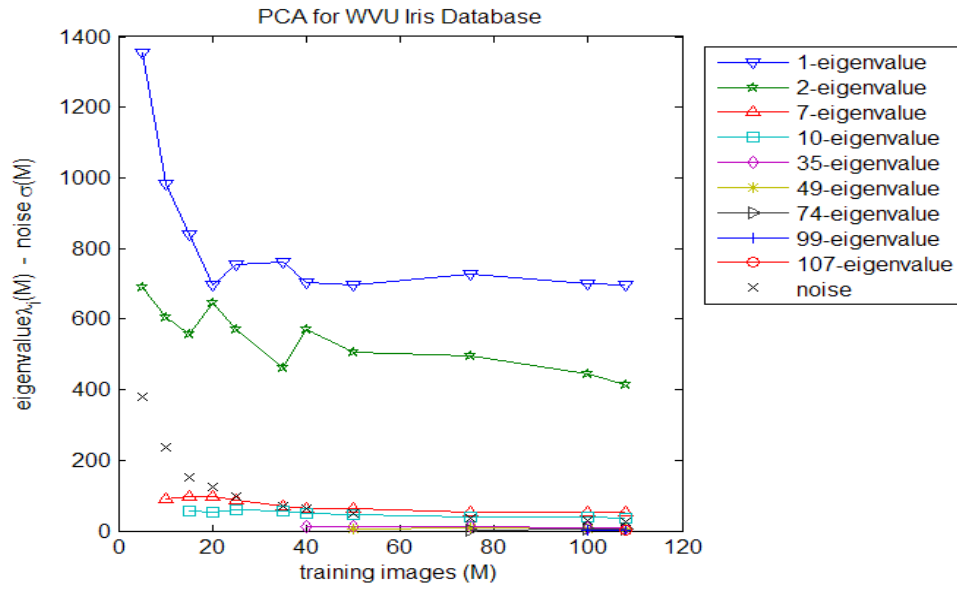


Figure 3.6: n-th Eigenvalue vs the number of training images for WVU DATABASE (linear scale)

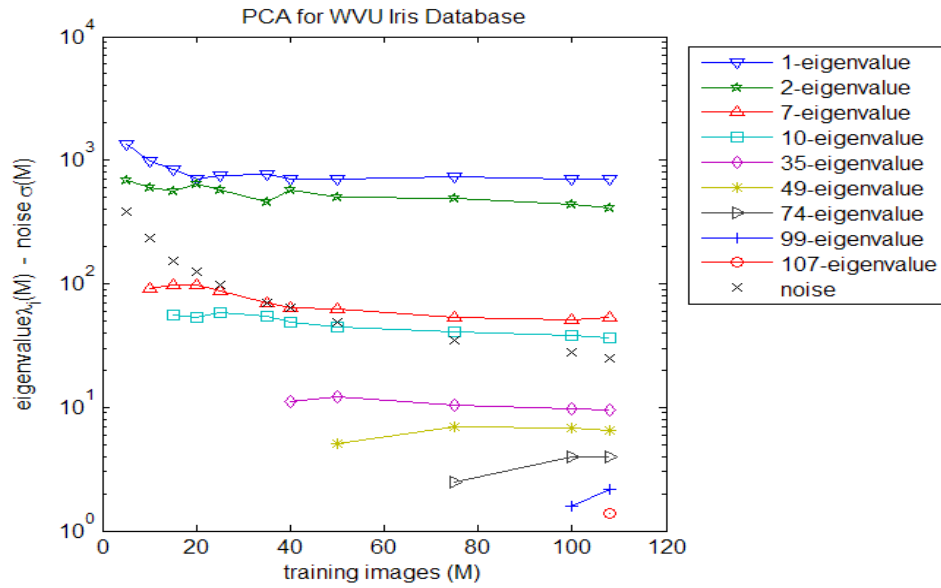


Figure 3.7: n-th Eigenvalue vs the number of training images for WVU DATABASE (log scale)

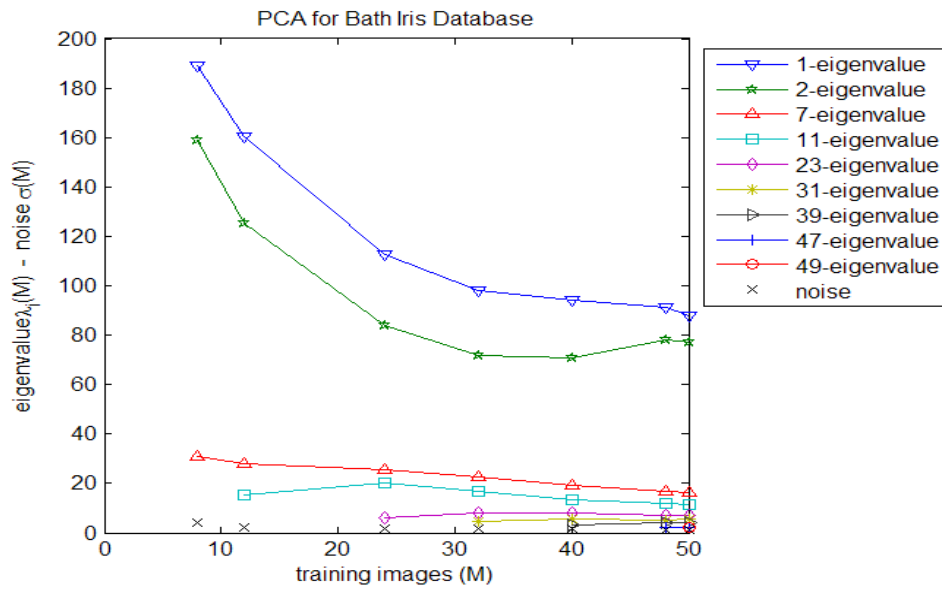


Figure 3.8: n -th Eigenvalue vs the number of training images for BATH DATABASE (linear scale)

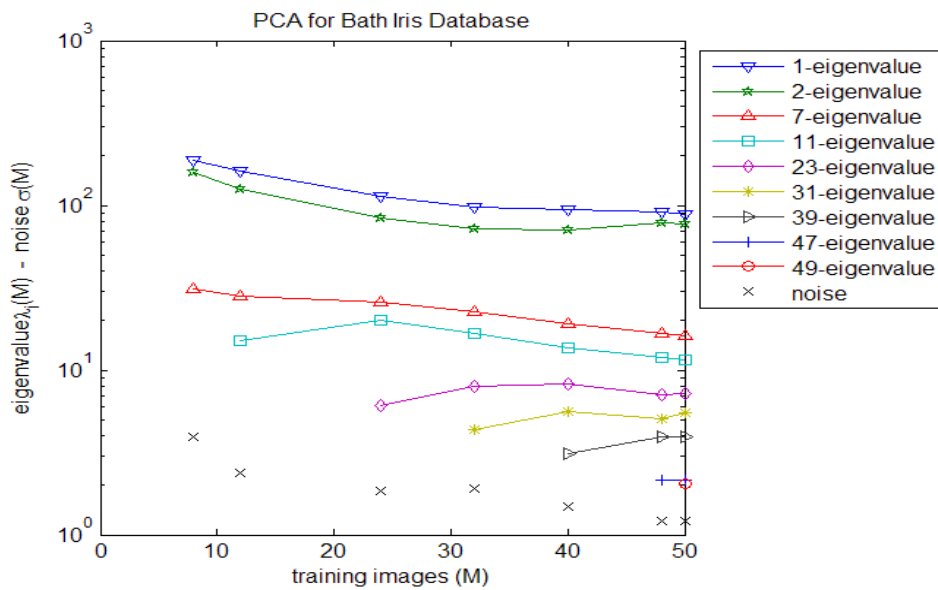


Figure 3.9: n -th Eigenvalue vs the number of training images for BATH DATABASE (log scale)

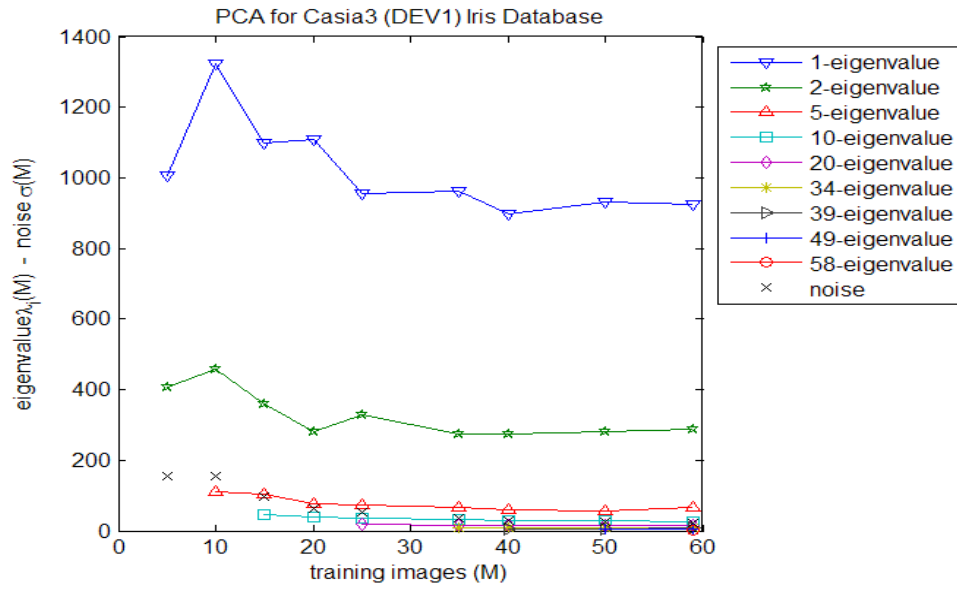


Figure 3.10: n-th Eigenvalue vs the number of training for CASIA3 (DEV1) DATABASE (linear scale)

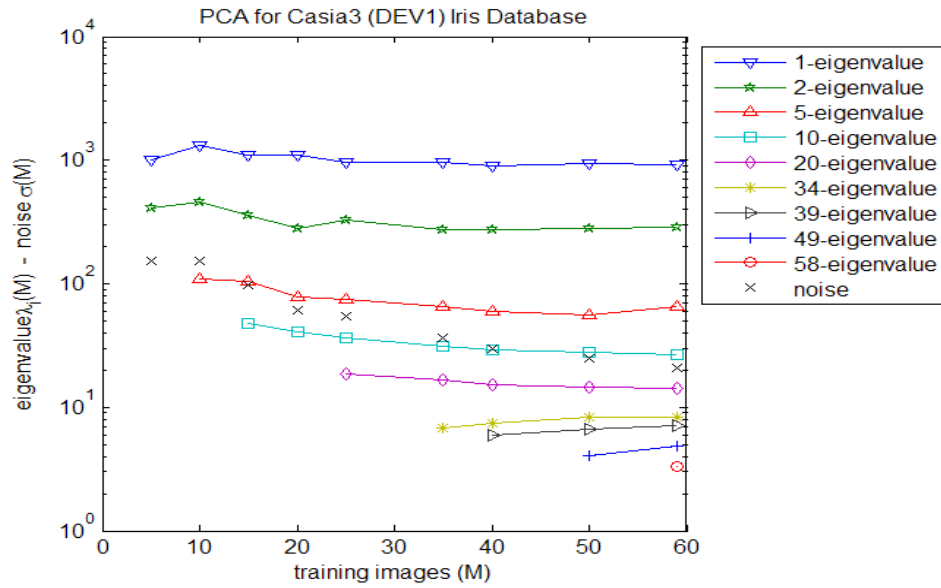


Figure 3.11: n-th Eigenvalue vs the number of training images for CASIA3 (DEV1) DATABASE (log scale)

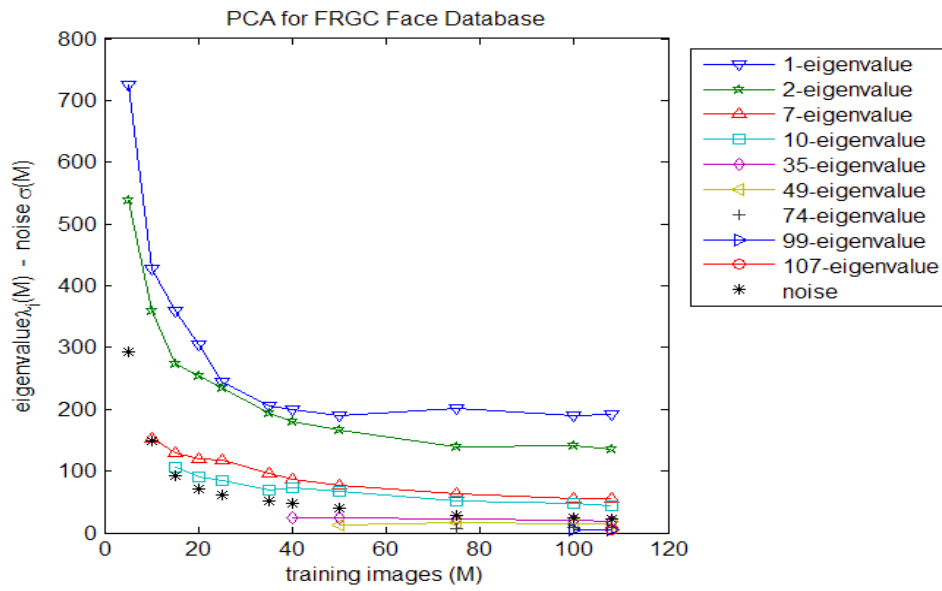


Figure 3.12: n -th Eigenvalue vs the number of training images for FRGC DATABASE (linear scale)

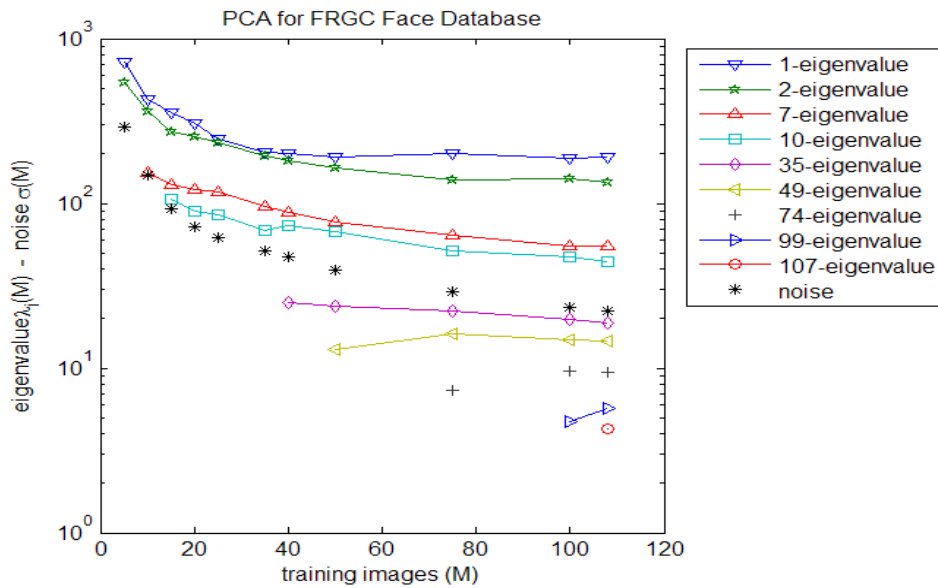


Figure 3.13: n -th Eigenvalue vs the number of training images for FRGC DATABASE (log scale)

The values of PCA-based recognition capacity \bar{I} (bits per principal component) as a characteristic of the overall quality of database for different databases are summarized in Table 3.2.

Iris Database	\bar{I} (bits/pc)	$\bar{\gamma}$	classes (M)
CASIA	0.4842	0.9552	108
WVU	0.0387	0.0551	108
BATH	0.7128	1.6863	50
CASIA3 (DEV1)	0.1068	0.1596	59
Face Database	\bar{I} (bits/pc)	$\bar{\gamma}$	classes (M)
FRGC	0.1274	0.1931	108

Table 3.2: Recognition Capacity of Biometric Databases using PCA encoded data

The rate indicates the constrained relationship between the number of classes that the PCA system can encode and the length of encoded data describing the template at a specific level of distortion. Based on the values of constrained capacity we can evaluate the overall quality of biometric database. High values of recognition rate indicate good quality database, low values of recognition rate indicate low quality database. Note, because we are evaluating biometric images in the feature space, the additive noise in this space corresponds to the effect of various factors in the image space; in particular for iris many factors like occlusion, specular reflection, resolution, blurring, lighting and resolution and lighting for face. In the next section we will illustrate how one of the listed factor (resolution) influences the value of recognition capacity.

3.5 Recognition Capacity vs Resolution

The following experiment shows how PCA-based recognition capacity depends on the resolution of the iris images. We illustrate this on the CASIA-1 iris database. The dataset is composed of 108 iris classes with 6 image per class. We evaluate capacity of the recognition channel at six different resolutions. The original images (360x64 pixels) were downsampled. The dependence of constrained capacity on the

reciprocal of SNR ($\bar{\gamma}^{-1}$) is shown in Fig. 3.14. The recognition capacity decreases as $\bar{\gamma}^{-1}$ increases.

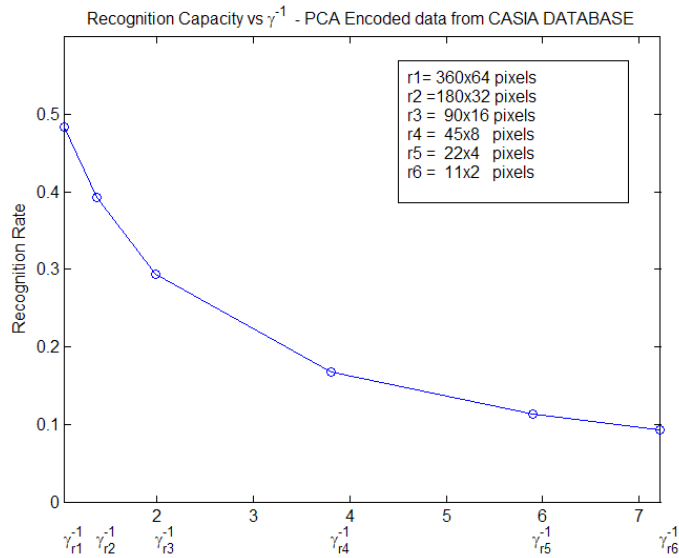


Figure 3.14: Capacity vs reciprocal of SNR ($\bar{\gamma}^{-1}$) for CASIA1 Iris Database

Note that in this case that $\bar{\gamma}^{-1}$ increases as the image resolution decreases. This result is analogous to the rate-distortion curve in the case of lossy coding.

Chapter 4

ICA Representation rate

In this chapter we will derive an expression for the constrained capacity of noisy biometrics-based recognition channel under the constraint of ICA-encoded data.

4.1 ICA based encoding

Consider a biometric database composed of images characterizing M classes, r images per class. The images are encoded following a two-step procedure: (1) images are first subject to principal component analysis and then (2) independent component analysis is applied to PCA templates. As shown in [14], PCA templates do not possess high discriminating power when applied to iris biometric, since PCA algorithm extracts global features. To improve recognition performance, the PCA templates are further refined using ICA algorithm.

Similar to PCA algorithm the ICA encoding method projects images into a new coordinate system. While PCA components in a PCA template are empirically uncorrelated, the ICA components in an ICA template are empirically independent.

Let \tilde{I} be a matrix with vector columns given by $\tilde{I}_1, \tilde{I}_2, \dots, \tilde{I}_M$, a sequence of pre-processed, normalized, and whitened images indexed by their class. Assume that each class is represented by a single iris image.

ICA is a blind source-separation method. It assumes that observed data can be represented as a linear combination of a number of independent signals. The

unknowns are the mixing coefficients and the independent input signals. Let S be a matrix composed of unknown independent input signals arranged in columns. Then the linear forward model that ICA assumes is $I = AS$ where A is the unknown mixing matrix. As argued in [15], the results of linear mixing of non-Gaussian signals are more Gaussian than the input signals. Then to estimate the mixing matrix A and one of the components of S , one has to define a measure of non-Gaussianity. One of theoretically sound criteria is the maximization of the negentropy given by:

$$J(\underline{X}) = H(A^{-1}I)_{gauss} - H(A^{-1}I) \quad (4.1)$$

where H_{gauss} is the entropy of the data under the assumption that data are Gaussian distributed and under the constraint of the same covariance matrix for the distributions in H and H_{gauss} in (see [15] for more detailed explanation). Once the mixing matrix and one of the input signals are estimated, the remaining input signals can be obtained by invoking the Gram-Schmidt orthogonalization procedure. To deal with empirical case, (4.1) is approximated by expressions involving empirical moments.

4.2 Model for ICA encoded data

In this section we propose and validate a Bessel K model for ICA-encoded iris data.

4.2.1 Bessel K Model

Let $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_M$ be n -dimensional vectors of ICA components. Each vector is a projection of an image from an individual biometric class onto the space formed by the columns of mixing matrix. Suppose that vectors are independent and identically distributed each described by the Bessel K distribution. To be more specific,

$$\underline{X}_m = \sqrt{G_m} \underline{Z}_m + \underline{\mu}, \quad m = 1, 2, \dots, M \quad (4.2)$$

where \underline{Z}_m is the Gaussian distributed vector with zero mean and covariance matrix Σ_z , G_m is a gamma-distributed random variable with parameters α and θ and $\underline{\mu}$ is

the mean vector of \underline{X}_m . The vector \underline{Z}_m and the scalar G_m are independent.

A noisy ICA template presented for identification is modeled as a Bessel K distributed vector augmented with independent Gaussian noise with zero mean and covariance matrix Σ_N :

$$\underline{Y} = \underline{X}_Y + \underline{N}, \quad \underline{N} \sim \mathcal{N}(0, \Sigma_N).$$

Because BesselK distribution is a kind of Scale Gaussian Mixture [16], the conditional distribution of \underline{X}_m given $G_m = g$:

$$\tilde{\underline{X}}_m = (\underline{X}_m | G_m = g) \sim \mathcal{N}(0, g\Sigma_Z), \quad \tilde{\underline{Y}} = (\underline{X}_Y + \underline{N} | G_Y = g).$$

To evaluate the information density we first find the joint and marginal distributions of the concatenated random vector $[\tilde{\underline{X}}_m, \tilde{\underline{Y}}]^T$.

Under hypothesis H_m , the noisy candidate \underline{Y} has the same signature of \underline{X}_m and the conditional vector $[\tilde{\underline{X}}_m, \tilde{\underline{Y}}]^T$ is Gaussian with the following covariance matrix:

$$\mathbf{R}_1 = \begin{pmatrix} g\Sigma_Z & g\Sigma_Z \\ g\Sigma_Z & g\Sigma_Z + \Sigma_N \end{pmatrix}.$$

Under the null hypothesis H_0 , the signature part \underline{X}_Y of the noisy candidate is independent of all signatures \underline{X}_m . The conditional vector $[\tilde{\underline{X}}_m, \tilde{\underline{Y}}]^T$ is Gaussian with the following covariance matrix:

$$\mathbf{R}_0 = \begin{pmatrix} g\Sigma_Z & 0 \\ 0 & g\Sigma_Z + \Sigma_N \end{pmatrix}.$$

The conditional information density for this vector is:

$$\tilde{i}_n = -\frac{1}{2n} \sum_{k=1}^n \left[\frac{\tilde{X}_k^2}{\sigma_{N_k}^2(M)} - 2 \frac{\tilde{X}_k \tilde{Y}_k}{\sigma_{N_k}^2(M)} + \frac{g(M) \sigma_{Z_k}^2(M) \tilde{Y}_k^2}{\sigma_{N_k}^2(M) (\sigma_{N_k}^2(M) + g(M) \sigma_{Z_k}^2(M))} - \log \left(1 + \frac{g(M) \sigma_{Z_k}^2(M)}{\sigma_{N_k}^2(M)} \right) \right],$$

where the matrices Σ_Z, Σ_N are diagonal with the respective k-th entries $\sigma_{Z_k}^2, \sigma_{N_k}^2$ and $\underline{\mu}$ is zero.

The conditional information rate is the expected value of the previous expression under the joint distribution of $\underline{X}_m, \underline{Y}$

$$\tilde{I}_n(M) = \frac{1}{2n} \sum_{k=1}^n \log \left(1 + \frac{g(M)\sigma_{Z_k}^2(M)}{\sigma_{N_k}^2(M)} \right). \quad (4.3)$$

This rate depends on the non deterministic scalar $g(M)$. Because of (4.2), for the generic template \underline{X} the covariance matrix can be easily factorized resulting in:

$$\Sigma_X = E_g\{G_m\}\Sigma_Z = \alpha\theta\Sigma_Z.$$

Because the data are pre-whitened, the covariance matrix Σ_X is:

$$\Sigma_X = \frac{1}{M-1} \sum_{i=1}^M (\underline{X}_i - \underline{\mu})(\underline{X}_i - \underline{\mu}) = I,$$

where I is identity matrix. The conditional rate (4.3) becomes:

$$\tilde{I}_n(M) = \frac{1}{2n} \sum_{k=1}^n \log \left(1 + \frac{g(M)}{\sigma_{N_k}^2(M)\alpha(M)\theta(M)} \right).$$

Finally, the information rate of the random vector $[\underline{X}_m, \underline{Y}]^T$ is:

$$I_n(M) = E_g\{\tilde{I}_n(M)\} = \frac{1}{2n} \sum_{k=1}^n E_g \left\{ \log \left(1 + \frac{g(M)}{\sigma_{N_k}^2(M)\alpha(M)\theta(M)} \right) \right\}. \quad (4.4)$$

It is not straightforward to estimate the value of information rate using (4.4).

If $\frac{g}{\sigma_{N_k}^2 \alpha \theta} < 1$ we can apply the Taylor expansion of $\log \left(1 + \frac{g}{\sigma_{N_k}^2 \alpha \theta} \right)$. In this case we have the approximate expression:

$$I_n(M) = E_g\{\tilde{I}_n(M)\} = \frac{1}{2n\alpha(M)\theta(M)} \sum_{i=1}^{\infty} \sum_{k=1}^n \frac{(-1)^{(i-1)}}{i} \frac{E_g\{g^i(M)\}}{\sigma_{N_k}^2(M)}.$$

Taking the limit with respect to n and M results in expression for *ICA-based recognition capacity* or *ICA representation rate*:

$$\bar{I} = \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{2n\alpha(M)\theta(M)} \sum_{i=1}^{\infty} \sum_{k=1}^n \frac{(-1)^{(i-1)} \mathbb{E}\{g^i(M)\}}{i \sigma_{N_k}^2(M)}. \quad (4.5)$$

4.2.2 Model Verification: K-S and K-L Test

Let \underline{X}_i be a vector of ICA components. The components in the vector \underline{X}_i are empirically statistically independent by construction. In our computations we model the mean vector as $m = \mu[1, 1, \dots, 1]$ and covariance matrix as $\Sigma_x = \sigma^2 I$ (I identity matrix). Hence we can consider each single component $X_i^{(k)}$ as a random sample from a univariate distribution with mean μ and variance σ^2 . Because we have also assumed \underline{X}_i for $i = 1, 2, \dots, M$ to be i.i.d, if we consider all the weights $X_i^{(k)}$ for all the users $i = 1, 2, \dots, M$ and for all the components $k = 1, 2, \dots, n$ we have a sample set \mathcal{S} of size $M \times n$ from a single univariate distribution. So, given a number of biometric templates M we can fit the distribution of the sample set \mathcal{S} with a given set of probability density functions (pdf). This is a parametric approach that requires estimation of the parameters of the theoretical fit by empirical data. For our experiment we arbitrary choose three theoretical distributions: BesselK, Laplace, Gaussian. The following list summarizes which parameters need to be specified to completely define the considered distributions:

- BESSELK : $\alpha = \frac{k_4}{3k_2}$, $\theta = \frac{3k_2^2}{k_4}$, $\mu = k_1$
- LAPLACE : $b = \sqrt{\frac{k_2}{2}}$, $\mu = k_1$
- GAUSSIAN : $\mu = k_1$, $\sigma^2 = k_2$,

where k_1, k_2, k_4 are the first, second and the fourth cumulants.

Now, the unbiased estimators for the 1st, the 2nd and 4th order cumulants of the distribution are:

$$k_1 = \mu_1 \quad (4.6)$$

$$k_2 = \frac{n}{n-1} \mu_2 \quad (4.7)$$

$$k_4 = \frac{n^2[(n+1)\mu_4 - 3(n-1)\mu_2^2]}{(n-1)(n-2)(n-3)} \quad (4.8)$$

where μ_1 , μ_2 and μ_4 are the first, second and fourth sample central moments and n is the number of samples in the empirical distributions.

The following figures show the best fit of the distributions for the considered biometric databases:

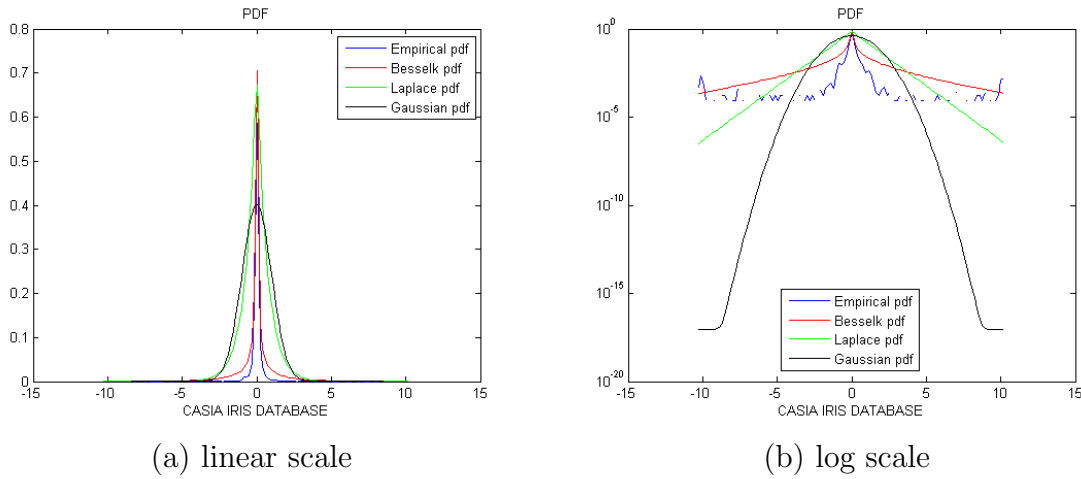
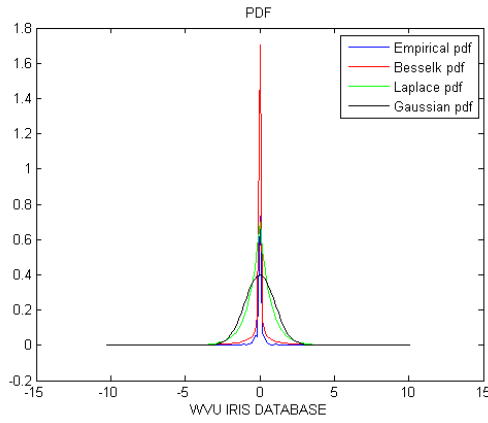
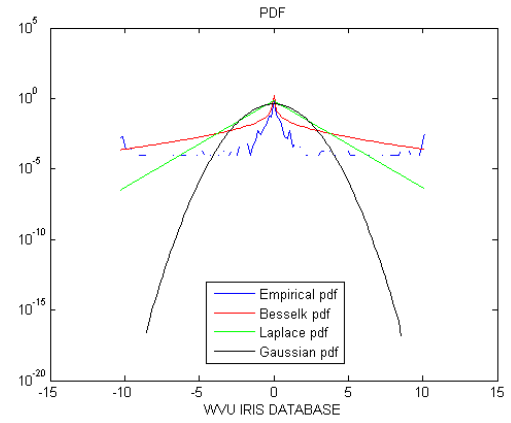


Figure 4.1: Empirical pdf of ICA encoded Iris Images from CASIA-1 Database and the parametric distribution providing the best fit (Kurtosis of empirical pdf = 83.7712)

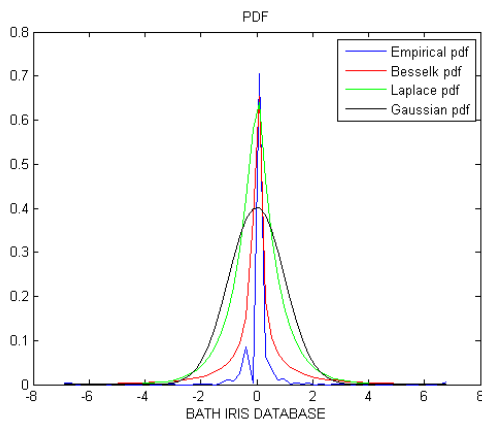


(a) linear scale

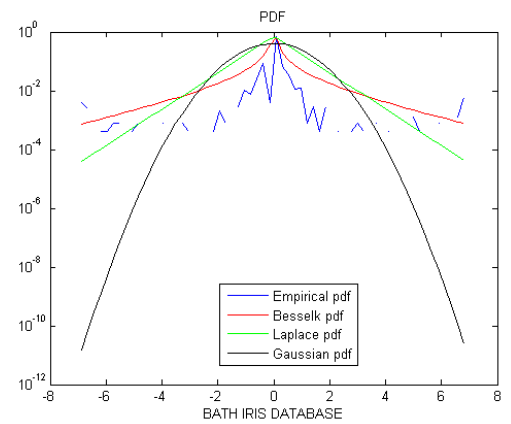


(b) log scale

Figure 4.2: Empirical pdf of ICA encoded Iris Images from WVU Database and the parametric distribution providing the best fit (Kurtosis of empirical pdf = 91.7742)



(a) linear scale



(b) log scale

Figure 4.3: Empirical pdf of ICA encoded Iris Images from BATH Database and the parametric distribution providing the best fit (Kurtosis of empirical pdf = 37.1320)

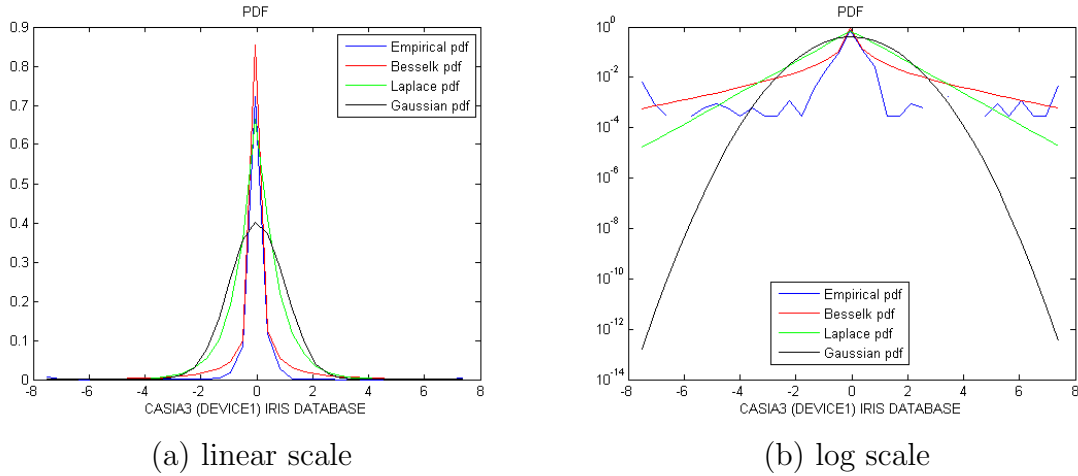


Figure 4.4: Empirical pdf of ICA encoded Iris Images from CASIA3 (DEV1) Database and the parametric distribution providing the best fit (Kurtosis of empirical pdf = 51.8308)

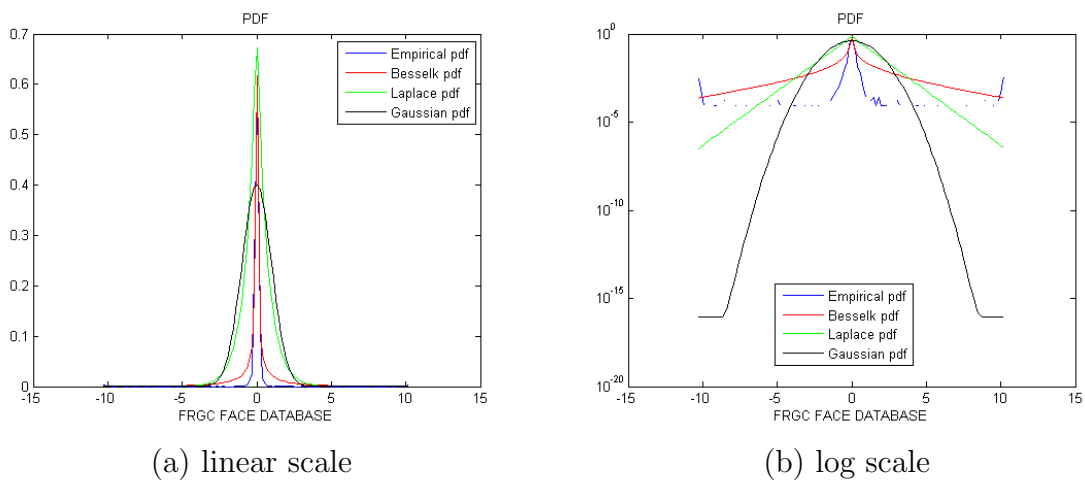


Figure 4.5: Empirical pdf of ICA encoded Face Images from FRGC Database and the parametric distribution providing the best fit (Kurtosis of empirical pdf = 98.8771)

The estimated parameters of the theoretical distribution are summarized in the table 4.1.

	BesselK			Laplace		Gaussian	
	α	θ	μ	β	μ	σ	μ
CASIA	0.0371	26.6767	9.3e-018	0.7039	9.3e-018	0.9908	9.3e-018
WVU	0.0338	29.3199	-2.6e-017	0.7039	-2.6e-017	0.9908	-2.6e-017
BATH	0.0879	11.1543	-4.9e-017	0.7001	-4.9e-017	0.9804	-4.9e-017
CASIA3(1)	0.0614	16.0057	-4.2e-018	0.7012	-4.2e-018	0.9833	-4.2e-018
FRGC	0.0313	31.6659	8.3e-017	0.7039	8.3e-017	0.9908	8.3e-017

Table 4.1: Estimated parameters from empirical data

Since the empirical value of the mean is very small, of order $10^{-17} - 10^{-18}$. We assume it is zero.

Two criteria have been used to verify the goodness of data fits:

- The Kolmogorov-Smirnov test
- K-L divergence.

The Kolmogorov-Smirnov (K-S) test is used to verify that a particular sample \mathcal{S} comes from a population with a specific distribution [17]. The K-S test compares the cumulative probability functions (cdf) $G(x)$ of the empirical data with the cdf $F(x)$ of the theoretical distribution which we want to fit. In particular, the two-sided test statistic is:

$$T = \sup_{x \in \mathcal{S}} |F(x) - G(x)| \quad (4.9)$$

The statistical test will be the following:

$$\begin{cases} H_0 : \text{The sample } \mathcal{S} \text{ follows the theoretical distribution;} \\ H_1 : \text{The sample } \mathcal{S} \text{ doesn't follows the theoretical distribution.} \end{cases}$$

Given that the null hypothesis H_0 is true, the probability of the T statistic being at least as extreme as the one observed is called p-value (see appendix A). A value of $p < p_{crit}$ ($p_{crit} = 0.05$) rejects the hypothesis H_0 .

Empirical and parametric (fitted) distribution are shown in Figures 4.6, 4.7, 4.8, 4.9, 4.10.

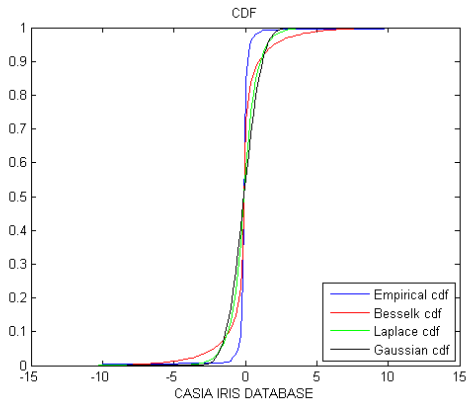


Figure 4.6: KS Test for CASIA IRIS DATABASE

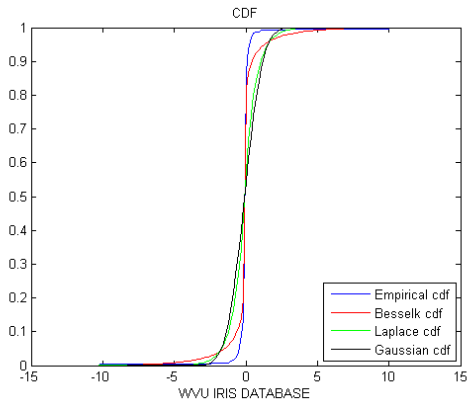


Figure 4.7: KS Test for WVU IRIS DATABASE

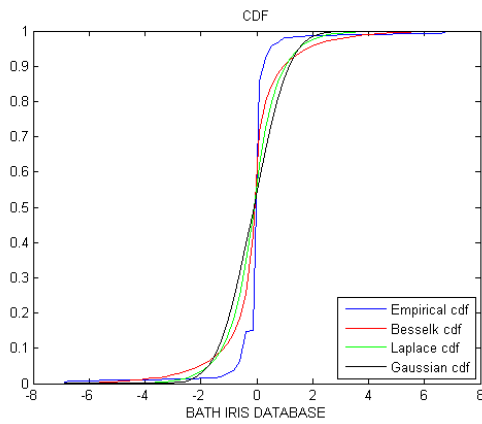


Figure 4.8: KS Test for BATH IRIS DATABASE

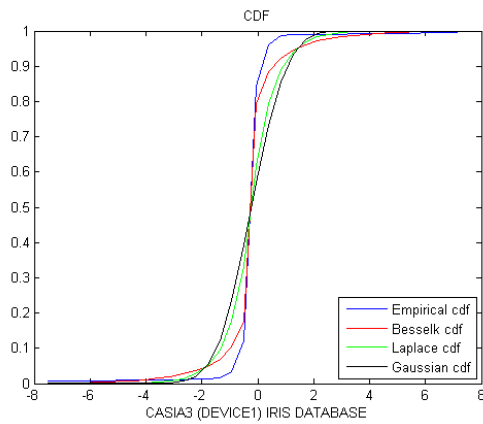


Figure 4.9: KS Test for CASIA3 (DEV1) IRIS DATABASE

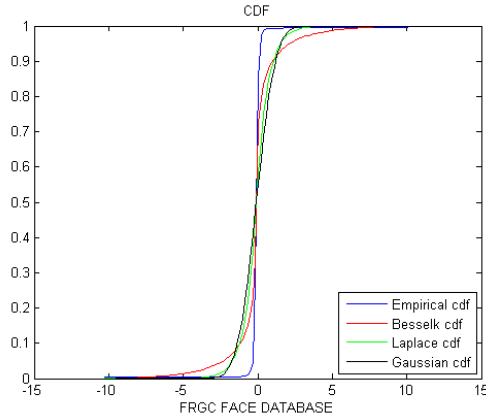


Figure 4.10: KS Test for FRGC FACE DATABASE

The results of the K-S tests are listed in Table 4.2:

	BesselK		Laplace		Gaussian	
	T	p	T	p	T	p
CASIA	0.1302	0.2624	0.2603	5.9901e-004	0.3058	2.7410e-005
WVU	0.0751	0.8885	0.2969	5.2222e-005	0.3264	5.7513e-006
BATH	0.1424	0.5277	0.2373	0.0520	0.2804	0.0123
CASIA3 (DEV1)	0.1039	0.9882	0.2836	0.1008	0.3343	0.0315
FRGC	0.1758	0.0509	0.2937	6.5351e-005	0.3499	8.5718e-007

Table 4.2: K-S test results for the biometrics databases

When the BesselK model is used, the p-values always exceeds the critical p-value ($p_{crit} = 0.05$) and are always greater than the Laplace and Gaussian ones. Note that the Gaussian model is always rejected for all the considered databases. Thus, the BesselK model provides the best fit.

The Kullback-Leibler (K-L) divergence [8] measures the relative entropy [18] between the empirical pdf $f(x)$ and theoretical pdf $g(x)$:

$$D_{KL}(f \parallel g) = \sum_{x \in \mathcal{S}} f(x) \log \left[\frac{f(x)}{g(x)} \right]$$

The smallest values of $D_{KL}(f \parallel g)$ indicates the best fit. Because this measurement is not symmetric, we define and use the symmetric K-L divergence:

$$d(f, g) = \frac{1}{2} \sum_{x \in \mathcal{S}} f(x) \log \left[\frac{f(x)}{g(x)} \right] + \frac{1}{2} \sum_{x \in \mathcal{S}} g(x) \log \left[\frac{g(x)}{f(x)} \right]$$

The value of $d(f, g)$ indicates the distance between the empirical pdf $f(x)$ and theoretical pdf $g(x)$.

The result of applying K-L test to the biometric databases are reported in the Table 4.3:

Database	d (BesselK)	d (Laplace)	d (Gaussian)
CASIA	0.4183	0.9986	1.5191
WVU	0.3109	1.4053	2.0038
BATH	0.3585	0.7650	1.2049
CASIA3 (DEV1)	0.3888	1.0694	1.6996
FRGC	0.8558	1.6603	2.3646

Table 4.3: K-L test result for the biometrics databases

Again, in all cases the BesselK model gives the lowest value of $d(f, g)$ corresponding to the best fit. In conclusion, the K-S and K-L tests have showed that the BesselK model provides a good fit into the empirical distribution of ICA encoded iris and face data.

4.3 ICA-Based Recognition Capacity of Iris and Face Databases

To estimate the ICA-based recognition capacity we use the approximate expression (4.5). The assumption that we make is that the noise \underline{N} is identical distributed for each independent component. Empirical studies on the considered databases about the convergence of (4.5) show that estimation of the first moment of the random variable g in (4.5) is sufficient to obtain the practical values of recognition capacity. In Table 4.4 we summarize the estimated values of ICA-based recognition capacity \bar{I} (bits per independent component) for the considered iris and face databases. These results have been obtained testing the system for a total of 5 images or more per users.

Iris Database	\bar{I} (bits/ic)	classes (M)
CASIA	0.5695	108
WVU	0.3629	108
BATH	1.5761	50
CASIA3 (DEV1)	0.5907	59
Face Database	\bar{I} (bits/ic)	classes (M)
FRGC	0.3510	108

Table 4.4: Recognition Capacity of Biometric Databases using ICA encoded data

Similarly to the PCA case the ICA-based recognition capacity is an index of the overall quality of the database. High values of recognition rate indicate good quality database, low values of recognition rate indicate low quality database.

4.4 ICA-Based Recognition Capacity vs PCA-Based Recognition Capacity

In Table 4.5 we compare the obtained values of ICA-based (bits per independent component) and PCA-based (bits per principal component) recognition capacity for the considered iris and face databases.

Iris Database	\bar{I}_{ICA} (bits/ic)	\bar{I}_{PCA} (bits/pc)	classes (M)
CASIA	0.5695	0.4842	108
WVU	0.3629	0.0387	108
BATH	1.5761	0.7128	50
CASIA3 (DEV1)	0.5907	0.1068	59
Face Database	\bar{I}_{ICA} (bits/ic)	\bar{I}_{PCA} (bits/pc)	classes (M)
FRGC	0.3510	0.1274	108

Table 4.5: Recognition Capacity of Biometric Databases using ICA and PCA encoded data

For all the considered databases the values of Recognition Capacity for ICA encoded data are greater than PCA-based ones.

Fig 4.11 shows the histogram of ICA-based and PCA-based recognition capacity for the iris databases.

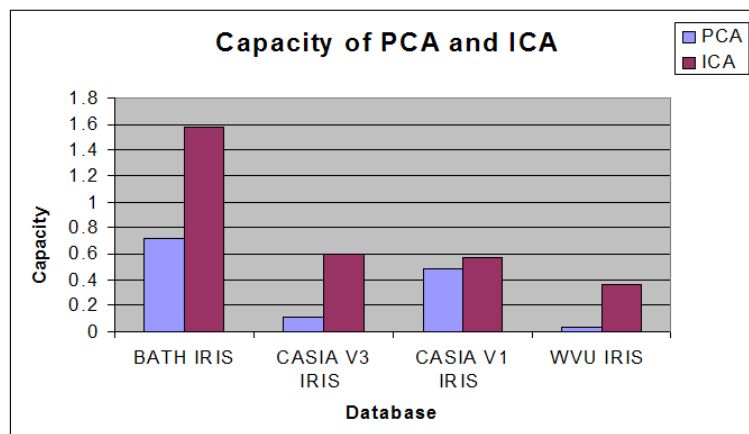


Figure 4.11: ICA-Based and PCA-Based Recognition Capacity of Iris Databases

Chapter 5

Conclusion

In this work we have proposed and validated two theoretical models to calculate the constrained capacity of PCA and ICA encoded data. We have considered four iris databases (CASIA1, WVU, BATH, CASIA3) and one face database (FRGC).

Firstly, we have modeled the PCA feature vectors as realizations of a Gaussian Process. The Shapiro-Wilk test has been used to validate this hypothesis. A graphical representation of eigenvalues and noise of PCA components has been proposed to calculate the asymptotic rate for a large number of users. The values of rate obtained discriminates the global quality of each database. We have further showed how capacity depends on resolution of the images tested.

In the second part of the work, we have represented the Independent Components as BesselK random variables. This model has been compared with the Gaussian and Laplace model. The Kolmogorov-Smirnov test and Kullback-Leibler divergence have been used to verify the goodness of data fit. Both tests have confirmed BesselK distribution as best fit. The parameters of the distribution have been estimated as well. Finally the capacity of ICA encoded data has been calculated for each database. Again, the global quality of the databases can be measured by the values of capacity obtained.

Comparing ICA and PCA capacity tables we can state that ICA capacity is always larger than the PCA capacity for the five databases. This result indicates that the ICA encoding is more efficient in terms of retrieving information for the purpose of recognition than PCA with respect to the considered databases.

Chapter 6

Future Work

1. PCA and ICA are only two encoding techniques used to encode iris and face data. In future work we can calculate capacity for other encoding methods like correlation based or Gabor Filter Based.
2. A parametric approach has been used to model the distribution of the PCA and ICA weights and to calculate the capacity. In future work we can use a non parametric approach to calculate the recognition rate and compare it with our models.
3. In this work we have assumed that the number of images per class of a considered database is the same and equal or larger than five. These assumptions have been made to have a good estimation of the noise. In practice different numbers of images per class are available. In future work, we can explore how this factor influences capacity.
4. In this work we have considered an example of how resolution affects capacity. In future we could study how other factors (blurring, lighting, etc) affect capacity.
5. Recognition capacity is an index of overall quality of database. In future we can compare this methodology with other methodologies for quality assessment.

Appendix A

Goodness of fit

A.1 Hypothesis Testing and Test Statistics

A statistical hypothesis test is a kind of statistical inference to make a quantitative decision about a process. Specifically, it decides whether there is enough evidence to "reject" a supposition or hypothesis about the process based on a sample data set \mathcal{S} . The supposition is called null hypothesis H_0 . A common layout for hypothesis testing is:

$$\begin{cases} H_0 : & \text{The sample } \mathcal{S} \text{ follows the null hypothesis;} \\ H_a : & \text{The sample } \mathcal{S} \text{ follows the alternative hypothesis.} \end{cases}$$

The test statistic T summarizes the information of the sample \mathcal{S} that is relevant to the hypothesis. Thus, the test statistic is defined for a specific hypothesis test. A good test statistic tends to take on certain values when the null hypothesis is true and to take on other values when it is false. That is, a good test statistic is a sensitive indicator of whether the sample data \mathcal{S} agrees or disagrees with the null hypothesis. Given the specific hypothesis, the sample set \mathcal{S} is considered a set of realizations of hypothetical random variables; thus the test statistic T is a function of the random variables.

A.2 Critical value and p value

The critical value or significance level α defines the sensitivity of a statistical test and indicates the maximum probability of rejecting a true null hypothesis (a decision known as type I error). In practice, for many reasons, this value is decided arbitrarily; typical critical values are 0.1, 0.05, and 0.01. The p-value is the probability of a test statistic T being at least as extreme as the one observed t_{obs} given that the null hypothesis is true. Precisely, given the specific null hypothesis as the truth and observed the value t_{obs} of the test statistic T from the sample S , the value p is defined as:

1. $P(T \geq t_{obs})$, for an upper-tailed (one-sided) test
2. $P(T \leq t_{obs})$, for a lower-tailed (one-sided) test
3. $P(T \leq t_{obs}^1) + P(T \geq t_{obs}^2)$, for a double-tailed (two-sided) test
where t_{obs}^1, t_{obs}^2 are two possible values of T for the same observation

The type of test is defined contextually within the test statistic T . For example, if the test statistic T takes small values when the null hypothesis is true, the upper-tail test will be used. Similarly, the appropriate definition of p will be used for the other cases. In general, the null hypothesis is rejected if the p-value is smaller than or equal to the critical value. The interval of probability for which the null hypothesis is rejected is called critical region (or rejection region) of the test statistic; the complementary region is called acceptance region. The distribution of test statistic T , given that the null hypothesis is true, is called Null distribution.

Appendix B

Bessel K Distribution

A Multivariate Bessel K Distribution is a kind of gaussian scale mixture (GSM) defined as:

$$\underline{X} = \sqrt{G}\underline{Z} + \mu$$

where $Z \sim \mathcal{N}(0, \Sigma_Z)$ and $G \sim \Gamma(\alpha, \theta)$.

The peculiarity of any GSM is:

$$(\underline{X} | G = g) \sim \mathcal{N}(\mu, g\Sigma_Z)$$

The general form of the probability density function (pdf) with zero mean ($\mu = 0$) of a multivariate BesselK distribution is:

$$p_X(\underline{X}) = \frac{2}{\Gamma(\alpha)(\pi|\Sigma_Z|)^{\frac{n}{2}}(2\theta)^{\frac{\alpha+n}{2}+\frac{n}{4}}} q(\underline{X})^{(\frac{\alpha-n}{2}-\frac{n}{4})} K_{(\alpha-\frac{n}{2})} \left(\sqrt{\frac{2q(\underline{X})}{\theta}} \right) \quad (\text{B.1})$$

where :

$$q(\underline{X}) = \underline{X}\Sigma_z^{-1}\underline{X}^T$$

is the usual quadratic form of a multivariate distribution and K_λ is the modified Bessel function of third kind of index λ .

The one dimensional version of the (B.1) with unitary variance is:

$$p_X(x) = \frac{2}{\Gamma(\alpha)\sqrt{\pi}(2\theta)^{\frac{\alpha}{2}+\frac{1}{4}}} |x|^{\left(\frac{\alpha}{2}-\frac{1}{4}\right)} K_{(\alpha-\frac{1}{2})} \left(\sqrt{\frac{2}{\theta}} |x| \right) \quad (\text{B.2})$$

The univariate BesselK distribution (BK) is a leptokurtic distribution (kurtosis > 3) with very heavy tail. It's a very special form of the Generalized Hyperbolic Distribution [19].

The odd cumulants of a BK are zero and even ones have the following form:

$$k_{2i} = \alpha \left(\frac{\theta}{2} \right)^i \frac{(2i)!}{i!}, \quad i > 1$$

In particular the second and fourth cumulants are:

$$k_2 = \text{var}(x) = \text{mean}(g) = \alpha\theta$$

$$k_4 = E\{x^4\} - 3k_2^2 = 3E\{g^2\} - 3E\{g\}^2 = 3\text{var}\{g\} = 3\alpha\theta^2$$

from these equations we can determine the two parameters α, θ

$$\alpha = \frac{k_4}{3k_2}$$

$$\theta = \frac{3k_2^2}{k_4}$$

The kurtosis of the BK distribution is:

$$\text{kurt}(x) = \frac{k_4 + 3k_2^2}{k_2^2} = \frac{k_4}{k_2^2} + 3 = \frac{1}{\alpha} + 3$$

Because $\alpha > 0$, the kurtosis is always > 3 (leptokurtic).

Bibliography

- [1] N.A. Schmid and J.A. O’Sullivan, “Performance prediction methodology for biometrics systems using a large deviations approach,” *IEEE Trans. on Signal Proc. Supplement on Secure Media*, vol. 52, no. 10, pp. 3036–3045, 2004.
- [2] M.B. Westover and J.A. O’Sullivan, “Achievable rates for pattern recognition: Binary and gaussian cases,” Adelaide, Australia, 2005, in Proc. of International Symp. on Information Theory (ISIT), pp. 28–32.
- [3] J.A. O’Sullivan and N.A. Schmid, “Performance analysis of physical signature authentication,” *IEEE Trans. on Info. Theory*, vol. 47, no. 7, pp. 3034–3039, 2001.
- [4] E. Shusterman, M.I. Miller, and B. Rimoldi, “Rate-distortion theory applied to automatic object recognition systems,” *IEEE Trans. on Info. Theory*, vol. 46, no. 5, pp. 1921–1927, 2000.
- [5] A. Srivastava, M.I. Miller, and U. Grenander, “Bayesian automated target recognition,” *in Handbook of Video and Image Processing*, pp. 869–881, 2000.
- [6] H.L. Van Trees, *Detection, Estimation and Modulation Theory, Part I*, John Wiley and Sons, New York, 2001.
- [7] R.M. Gray, *Entropy and Information Theory*, Springer-Verlang, New York, 1990.
- [8] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications. New York, 1991.

- [9] M.A. Turk and A.P. Pentland, "Eigenfaces for recognition," *J. of Cognitive Neuroscience*, pp. 71–86, 1991.
- [10] "Casia iris image database," <http://www.sinobiometrics.com/casiairis.htm>.
- [11] "Bath iris image database," <http://www.bath.ac.uk/elect-eng/research/sipg/irisweb/database.htm>.
- [12] "Face recognition grand challenge," <http://www.frvt.org/FRGC>.
- [13] S.S. Shapiro and M.B. Wilk, "An analysis of variance test for normality," *Biometrika*, vol. 52, no. 3-4, pp. 591–611, December 1965.
- [14] V. Dorairaj, N.A. Schmid, and G. Fahmy, "Performance evaluation of iris based recognition system implementing pca and ica techniques," *Proc. of the SPIE 2005 Symp. on Defense and Security*, pp. 51–58, March 2005.
- [15] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley and Sons, 2001.
- [16] M.J. Wainwright and E.P. Simoncelli, "Scale mixtures of gaussian and the statistics of natural images," *NIPS 1999*, vol. 3813, pp. 188–195, July 1999.
- [17] NIST, "Engineering statistics handbook," <http://www.itl.nist.gov/>, 2006.
- [18] R. E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley, New York,.
- [19] E. Eberlein and E.A. v.Hammerstein, "Generalized hyperbolic and inverse gaussian distributions: Limiting cases and approximation of processes," *In Seminar on Stochastic Analysis, Random Fields and Applications IV*, vol. 80, pp. 221–264, 2002.