

Iris Database Classification And Indexing

Thesis submitted in partial fulfillment of the requirements for the degree of

Master of Technology

in

Computer Science and Engineering

by

Jobin Joseph



Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela, Orissa, 769 008, India

May 2011

**Iris Database
Classification And
Indexing**

Thesis submitted in partial fulfillment of the requirements for the degree of

Master of Technology

in

Computer Science and Engineering

by

Jobin Joseph
(Roll- 209CS1069)

under the guidance of

Dr.Pankaj Sa



Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela, Orissa, 769 008, India

May 2011



Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela-769 008, Orissa, India.

Pankaj Sa
Assistant Professor

May 30, 2011

Certificate

This is to certify that the work in the thesis entitled *Iris Database Classification And Indexing* by JOBIN JOSEPH bearing *Roll No- 209CS1069* is a bonafide record of research work carried out by him under my supervision and guidance in partial fulfillment of the requirements for the award of the degree of Master of Technology in Computer Science and Engineering in the department of Computer Science and Engineering, National Institute of Technology Rourkela. Neither this thesis nor any part of it has been submitted for any degree or academic award elsewhere.

Pankaj Sa

Acknowledgement

First of all I would like to thank, the Almighty God, without whose blessings I wouldn't have been writing this "thesis".

I owe deep gratitude to the ones who have contributed greatly in completion of this thesis.

Foremost, I would like to express my sincere gratitude to my supervisor, Dr. Pankaj Sa for providing me with a platform to work on challenging areas of Computer Vision and Image Processing. I express my heartfelt thanks to him, for his guidance, support, and encouragement during the course of my Master study at the National Institute of Technology, Rourkela. I am especially indebted to him for teaching me both research and writing skills, which have been proven beneficial for my current research and future career. It has been a great honor and pleasure for me to do research under Prof.(Dr.) Pankaj Sa

I am very much indebted to Hunny Mehrotra for his endless support during my work.

I am also grateful to Dr. A. K. Turuk, Dr. S. K. Rath, Dr. Banshidhar Majhi, Dr. S. K. Jena, Dr. R. Baliarsingh, Dr. D. P. Mohapatra, Dr. P. M. Khilar, Dr. B. D. Sahoo, Prof. S. Chinara, Prof. K. S. Babu, Prof. M. N. Sahoo, Prof. S. Mohanty for giving encouragement during my thesis work.

I thank all the members of the Department of Computer Science and Engineering, and the Institute, who helped me by providing the necessary resources, and in various other ways, in the completion of my work.

I would like to thank all my friends and members of Image Processing Laboratory for all the thoughtful and mind stimulating discussions we had, which prompted us to think beyond the obvious.

Finally, I thank my parents and all my family members for their unlimited support and strength. Without their dedication and dependability, I could not have pursued my M.Tech. degree at the National Institute of Technology, Rourkela.

Jobin Joseph

Abstract

For increasing threat to the security systems biometric has been using widely for many applications. Biometric recognition is the recognition of individuals based in their physiological or behavioral characteristics. Examples of biometrics are face, iris, fingerprints, voice, palms, hand geometry, retina, handwriting, gait etc.

The performance of the biometric system depends on the search time and the error rate. these two factors are depends upon the size of the database. So here proposing one method to index the database within minimum time and search the minimum area of the database. The error rates of a biometric identification system are dramatically increasing with the size of database. Here used a method to index the iris data base using dct and the reordering of the DCT coefficients. Here proposed three new methods to extract the features from the iris strip. Among the three partitioning method discussing the efficient searching method the 10x10 square windowing gives a penetration rate of 9.8 percentage L-slicing method given penetration rate of 5.5 and the 8x8 slicing has given a penetration rate of .2 percentage of the total database,. In this the experiment has been done on CASIA Iris database.

Contents

Certificate	ii
Acknowledgement	iii
Abstract	iv
List of Figures	viii
List of Tables	x
1 Introduction	1
1.1 Biometric system	1
1.1.1 Verification:	2
1.1.2 Identification:	2
1.1.3 Physiological	3
1.1.4 Behavioral:	3
1.2 The steps in the biometric system are	3
1.2.1 Biometric data acquisition	3
1.2.2 Feature extraction	3
1.2.3 Biometric database	3
1.2.4 Matching	4
1.3 Iris Anatomy	4
1.4 Types of iris Textures	6
1.4.1 Stream Iris	6
1.4.2 Jewel Iris	6
1.4.3 Shaker iris	6
1.4.4 Flower iris	7
1.5 Performance measures of Biometric system	7

1.5.1	False Acceptance Rate (FAR)	9
1.5.2	False Rejection Rate (FRR)	9
1.5.3	Equal Error Rate (EER)	9
1.5.4	Genuine Acceptance Rate (GAR)	9
1.5.5	Penetration Coefficient (PR)	9
1.5.6	Bin Miss Rate (BM)	10
1.5.7	Cumulative Match Characteristic (CMC) Curve	10
1.6	Iris databases used in this Research	10
1.6.1	Cooperative Databases	10
1.7	Problem Definition	11
2		13
2.1	Literature survey	13
2.2	Indexing Biometric Databases using Pyramid Technique	13
2.2.1	Analysis	15
2.2.2	Result	16
2.3	Indexing Iris Biometric Database Using Energy Histogram of DCT Subbands	17
2.3.1	Preprocessing and feature Extraction	17
2.3.2	Key Generation	18
2.3.3	Database creation and searching	18
2.3.4	Experimental Results	19
2.4	Indexing technic using B+tree	19
2.4.1	indexing method	20
2.4.2	Insertion in modified Tree	20
2.4.3	Searching in Modified B+ tree	21
2.5	Experimental Result	22
3		24
3.1	Proposed Method	24
3.1.1	Feature Extraction using DCT	25
3.1.2	Square windowing	25
3.1.3	L-slice method	26

3.1.4	8x8 block slicing	26
3.2	Indexing using key	27
3.3	Key generation	27
3.4	Database Creation and Searching	28
3.5	Tree Structure	28
3.5.1	Algorithm1 (Tree indexing)	29
3.5.2	Algorithm2 (matching)	29
3.6	Algorithm2 (Experimental Evaluation and Results)	30
4		33
4.1	Conclusion And Future work	33
Bibliography		34

List of Figures

1.1	Various forms of authentication. Traditional methods of authentication using token based and knowledge based approaches (left). Use of biometrics to claim identity (right)	2
1.2	Examples of different biometric characteristics.	4
1.3	Variability is inherent in all signal readings, whether from (a) signature, (b) face, or (c) fingerprint.	5
1.4	Image 02463d1276 from the Iris Challenge Evaluation Dataset. Elements seen in a typical iris image are labeled here.	6
1.5	a)Stream Iris, b)Jewl Iris c)Shaker Iris d) Flower Iris.	7
1.6	Genuine and imposter matching score distribution of biometric database showing various performance measures	8
2.1	Genuine and imposter matching score distribution of biometric database showing various performance measures	14
2.2	Comparison of FRR v/s Penetration Rate for Indexing and Binning Systems	16
2.3	Comparison of FRR v/s Penetration Rate for Indexing and Binning Systems	17
2.4	5Energy Histogram of S10 region	18
2.5	Bin miss rate for change in number of classes	20
2.6	The Generalised structure of B+tree	21
2.7	FRR vs Penetration rate Graph for Bath database for varrious error rate	23
2.8	FRR vs Penetration rate Graph for IITK database for varrious error rate	23

3.1	square windowing	26
3.2	10x10 L slicing	27
3.3	8x8 slicing	28
3.4	Tree structure	29
3.5	square windowing	30
3.6	8x8 window	31
3.7	L-slice	31
3.8	Tabular representation of penetration rate and number of classes	32

List of Tables

Chapter 1

Introduction

1.1 Biometric system

For increasing threat to the security systems biometric has been using widely for many applications. Biometric recognition is the recognition of individuals based in their physiological or behavioral characteristics. Examples of biometrics are face, iris, fingerprints, voice, palms, hand geometry, retina, handwriting, gait etc. The various application using biometrics are passports, driving licenses, banking, refraining imposters from hacking into networks, stealing mails etc. The traditional security systems are Token based system, in this fakers are prevented from accessing protected resources using ID cards, smart cards etc, Knowledge based systems, in this identity is proving by using information like User id and password associated with the system. In some system using both the above mentioned approach. The main disadvantages of token based systems are ID cards can be lost, forged, or misplaced .And in the case of knowledge based system passwords can be forgotten or compromised. If a system using combinations of both knowledge based and token based system, it is not always possible to achieve the security requirements [1]. Biometric system is composed if a pattern recognition system which captures the biometric data of individuals and extracting features from the acquired data, then it comparing the features against the templates stored previously in the database. Biometric system giving a robust solution to the identity management by using fully or partially automated schemes to recognize Individual identity. Depending on the application mode biometric system operates on verification or identification mode [2]. Main advan-

tages of biometric system over conventional approach is the reliability, it cannot be stolen or misplaced.

In a biometric system various biometric features are extracting after capturing the biometric images of the user. And authenticating individual by checking against the templates previously stored in the database. How an individual to be authenticated is depending upon application of the biometric system is used. The types of operating modes of biometric system are verification and identification



Figure 1.1: Various forms of authentication. Traditional methods of authentication using token based and knowledge based approaches (left). Use of biometrics to claim identity (right)

1.1.1 Verification:

In verification system the captured templates are comparing with his pre stored templates and if it is matching then the individual become authenticated by the system. Here actually a 1:1 comparison is doing for knowing whether the identity of person is genuine or not.

1.1.2 Identification:

Identification can be in the form of verification, authenticating a claimed identity (verifying the claim, "I am Joe Smith") or recognition, recognition, determining

the identity of a person from a database of known persons (determining who I am without knowledge of my name). In identification the matching would be one to many types. The Fig 1.1 shows the diagrammatic representations of verification and identification. The various traits are face, iris, fingerprint etc. according to their nature these belong to two categories.

1.1.3 Physiological

This is also termed as passive biometrics. These are based on the direct measurement of the body. Examples of this kind of biometrics are face, iris, retina, hand geometry etc.

1.1.4 Behavioral:

This is also termed as active biometrics. These are mainly by the actions taken by the person. It takes the measures of the actions taken by the person. Examples of these are voice, keystroke speed and signature etc.

1.2 The steps in the biometric system are

1.2.1 Biometric data acquisition

This involves the procedure to acquire biometric data from individuals using various types of sensors.

1.2.2 Feature extraction

The feature sets are taken from the acquired images, like corners of line in iris, or intensity etc.

1.2.3 Biometric database

It contains biometric feature sets enrolled. These will be after quality checking.

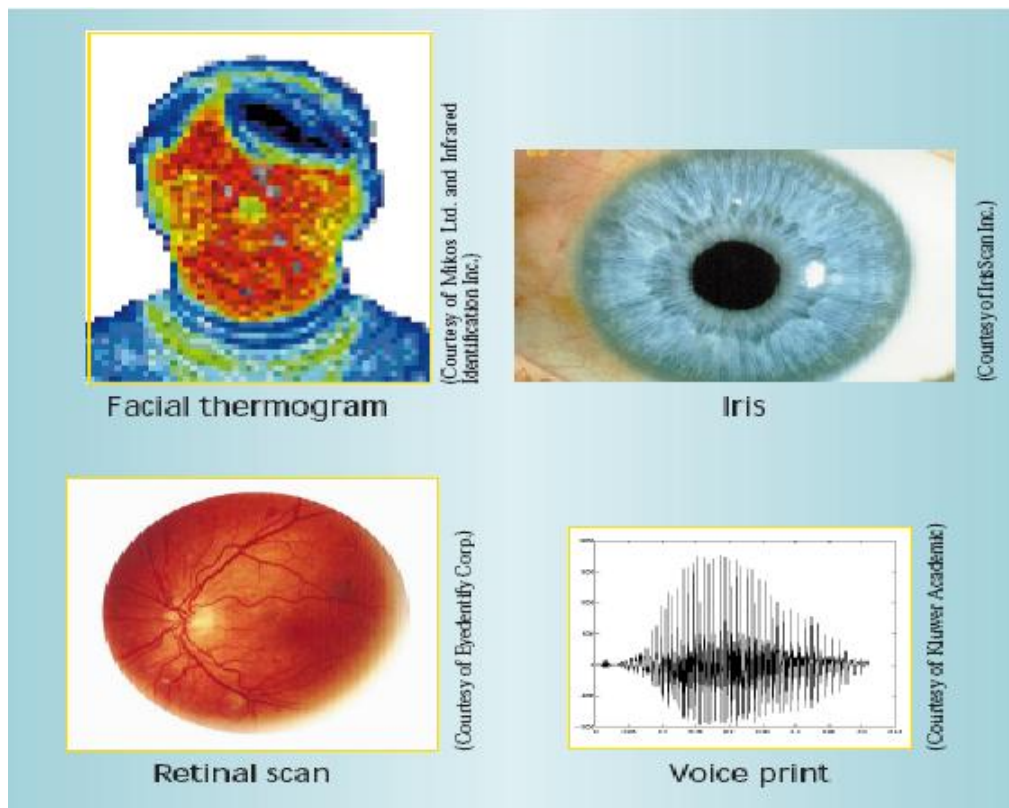


Figure 1.2: Examples of different biometric characteristics.

1.2.4 Matching

The feature sets (templates) extracted from the query templates are matching against the templates in the database and generating matching score.

1.3 Iris Anatomy

Reliability of biometrics is depends on the characteristics to acquire the features and it can be captured in an unchanged manner over the passage of time. Biometric s has it s own strength and weakness. The features of finger print will remain unchanged with the passage of time. But in the case of face the feature values may change with the passage if time. For acquiring the finger print the sensor needs direct contact with the finger.

The iris is the "colored ring of tissue around the pupil through which light...enters the interior of the eye." [3]. Two muscles, the dilator and the sphincter muscles, control the size of the iris to adjust the amount of light entering the

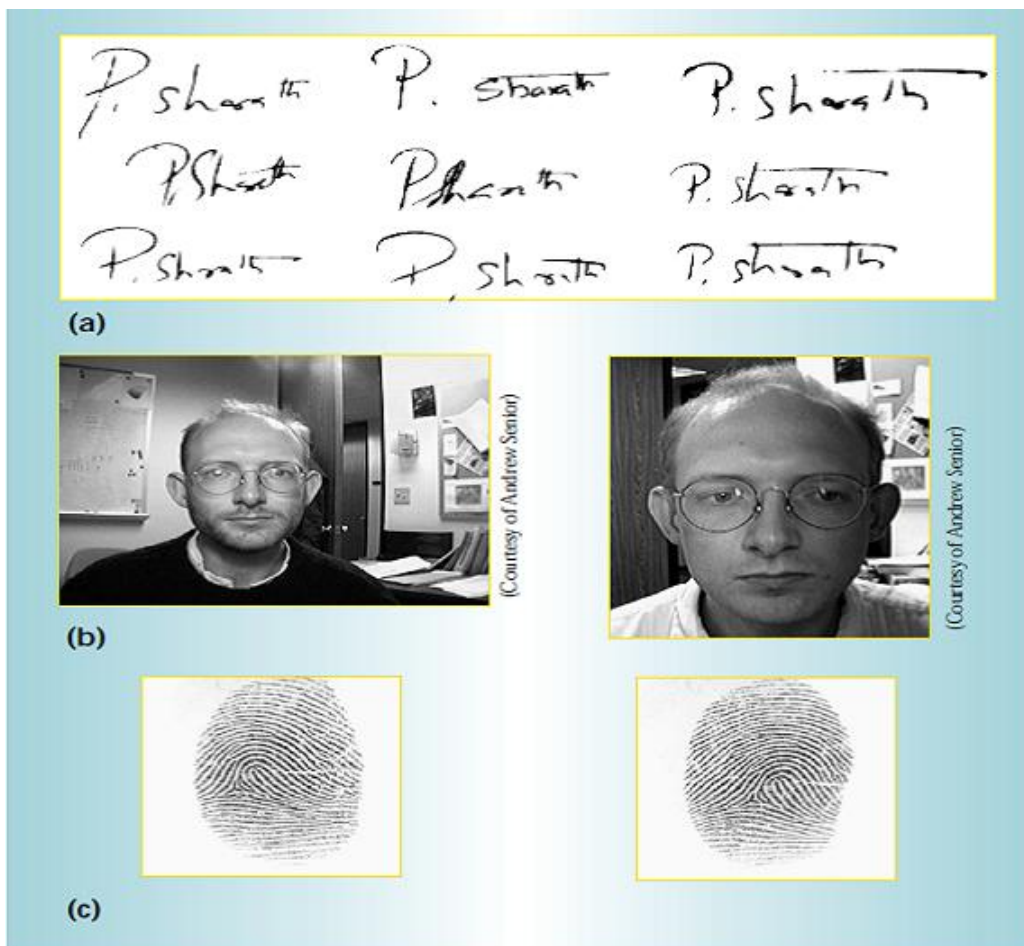


Figure 1.3: Variability is inherent in all signal readings, whether from (a) signature, (b) face, or (c) fingerprint.

pupil. Figure 3 shows an example image acquired by a commercial iris bio-metrics system. The sclera, a white region of connective tissue and blood vessels, surrounds the iris. A clear covering called the cornea covers the iris and the pupil. The pupil region generally appears darker than the iris. However, the pupil may have specular Highlights and cataracts can lighten the pupil. The iris typically has a rich pattern of furrows, ridges, and pigment spots. The surface of the iris is composed of two regions, the central papillary zone and the outer ciliary zone. The collarette is termed s the boundary of these two regions. Iris is different from person to person, left to right iris, even the iris between twins are totally different.

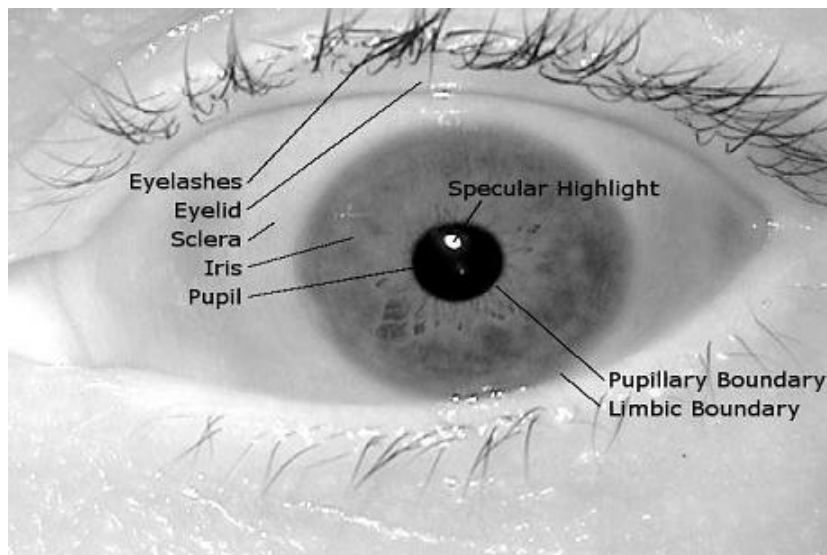


Figure 1.4: Image 02463d1276 from the Iris Challenge Evaluation Dataset. Elements seen in a typical iris image are labeled here.

1.4 Types of iris Textures

1.4.1 Stream Iris

It contains a uniform fiber structure with subtle variations or streaks of color as shown in figure 14.c. The structure of the iris is determined by the arrangement of the white fibers radiating from the center of the iris .

1.4.2 Jewel Iris

It contains dot-like pigments in the iris. The jewel iris can be recognized by the presence of pigmentation or colored dots on top of the fibers as shown in figure 1.4. The dots (or jewels) can vary in color from light orange through black [4]

1.4.3 Shaker iris

It contains dot-like pigments and rounded openings. The shaker iris is identified by the presence of both flower like petals in the fiber arrangement and pigment

dots or jewels is shown in figure 1.4

1.4.4 Flower iris

It contains distinctly curved or rounded openings in the iris. In a flower iris the fibers radiating from the center are distorted (in one or more places) to produce the effect of petals (hence the name flower) . In this image one can notice that they are neither regular nor uniform.

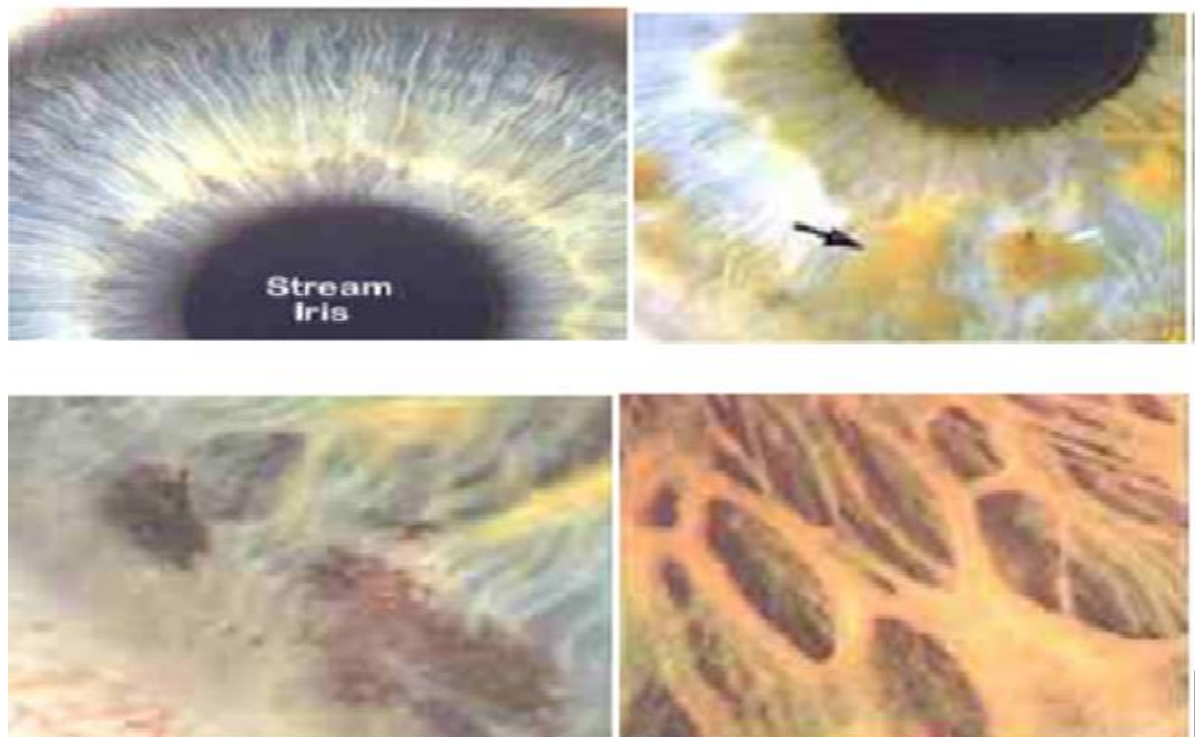


Figure 1.5: a)Stream Iris, b)Jewel Iris c)Shaker Iris d) Flower Iris.

1.5 Performance measures of Biometric system

When comparing passwords the matching is finding by a perfect match with the same set of alphanumeric characters. But in the case of biometrics comparison with the same set of templates is occurring very rarely because of scanning conditions, change in acquiring conditions, change due to aging etc. When two different biometric templates originating from same individual are different then it is known

as intra-class variations. However, variations that occur between templates originating from two different individuals are known as inter-class variations [5].

If two biometric templates are comparing to find out the intra-class variations then the scores obtaining is called as similarity scores or genuine scores. And if two biometric templates are comparing to find out the inter-class similarity, then the score is known as imposter scores. when the scores are exceeding a predefined threshold value(T), then the result is false acceptance. If the genuine scores that laying below the threshold value then it is false rejection. The Fig:1.6 shows the representation of various performance measures. The commonly used biometric measures are given below

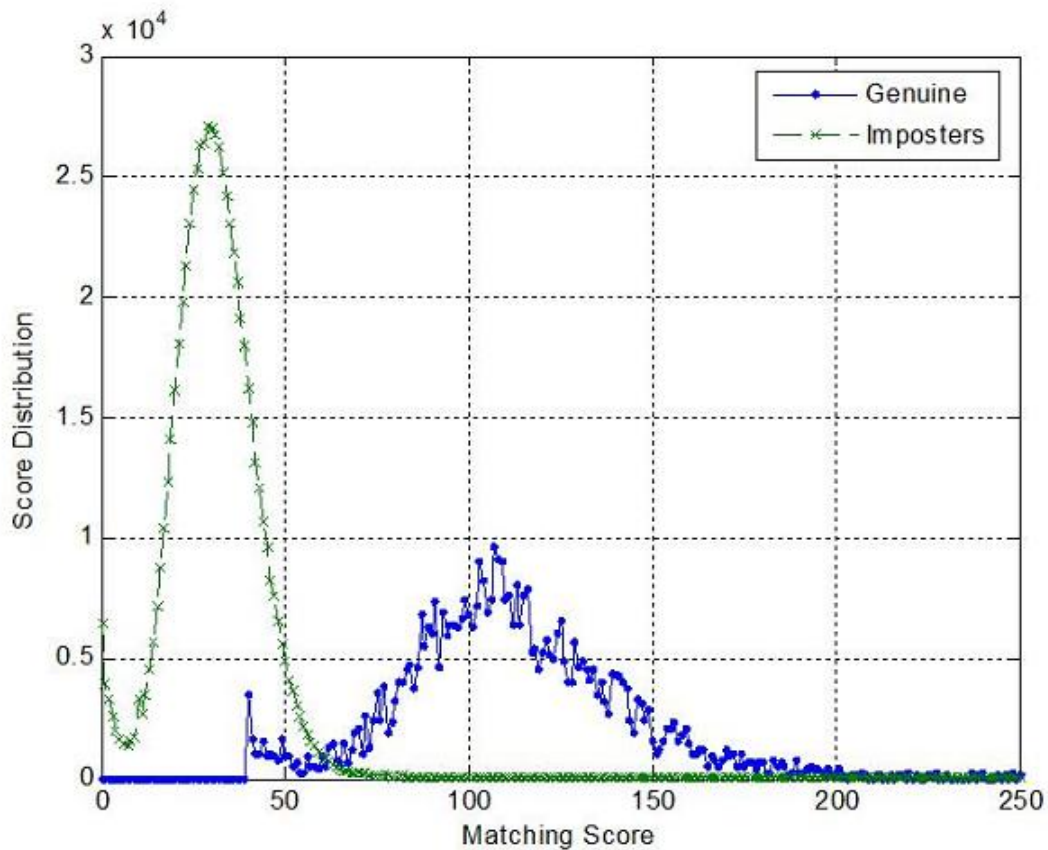


Figure 1.6: Genuine and imposter matching score distribution of biometric database showing various performance measures

1.5.1 False Acceptance Rate (FAR)

FAR is the frequency of fraudulent access to imposters claiming identity [6]. This statistic is used to measure biometric performance when operating in the verification mode. A false accepts occurs when an individual is incorrectly matched to another individual's existing biometric template.

1.5.2 False Rejection Rate (FRR)

FRR is the frequency in rejection of peoples those who are valid user that to the system. This property will check when the biometric system is working on verification mode. A false reject occurs when a person is not matched correctly to his/her own existing biometric template.

1.5.3 Equal Error Rate (EER)

ERR is the point where FAR is equal to FRR. In general, the lower the equal error rate value, the higher the accuracy of the biometric system. Note, however, that most operational systems are not set to operate at the equal error rate, so the measure's true usefulness is limited to comparing biometric system performance. EER is sometimes referred to as the Crossover Error Rate.

1.5.4 Genuine Acceptance Rate (GAR)

GAR is the fraction of genuine scores exceeding the threshold T . It is defined as

$$GAR = 1 - FRR \quad (1.1)$$

1.5.5 Penetration Coefficient (PR)

In identification, the query fature set is compared to all the other templates in the database. Search efficiency can be achieved by partitioning the database based on some criteria. Thus, during identification, the query template is compared to only select templates in the appropriate partitions. The portion of total database to be scanned on an average for each search is called penetration coefficient PR, which can be defined by

$$PR = E/N \quad (1.2)$$

Where E is the expected number of comparisons required for single input and N is the total number of comparisons. On encountering the match, search does not stop but continues through the entire partition [6].

1.5.6 Bin Miss Rate (BM)

A bin error occurs when an attempt is placed in a bin which is not compared with the correct bin for the biometric entity used, and hence will fail to match. The error occurs due to misplacing of biometric template in the wrong bin during identification

1.5.7 Cumulative Match Characteristic (CMC) Curve

The rank- k identification indicates the number of correct identities that occur in top k matches. Let R_k is the number of elements of probe set in top k , then the probability of identification (I) is given by $I = R_k/N$. CMC curve represents the Probability of identification I at various ranks K [7].

1.6 Iris databases used in this Research

To measure the performance of iris biometric system, various experiments have been carried out at various levels. Here we are discussing the various types of databases used in experiment. Experimental results are obtained on various available datasets such as BATH [18], CASIA version 3 [1] and Indian Institute of Technology Kanpur (IITK) to take all possible factors into consideration like rotation, illumination, scaling and noise. These databases are classified into cooperative and non-cooperative categories based upon the restrictions imposed on the user while capturing images.

1.6.1 Cooperative Databases

These databases are acquired from the user in idle conditions with fewer impositions. BATH and CASUA databases are falling under this category.

BATH Databasee

Database available from BATH University includes images from 50 subjects. For each subject, both left and right iris images are obtained, each containing 20 images of the respective eyes.

CASIA Database

CASIA version 3 (CASIAV3) is acquired in an indoor environment. Most of the images have been captured in two sessions with an interval of at least one month. The database comprises 249 subjects with total of 2655 images from left and right eyes. CASIAV3 is a superset of CASIAV1. The pupil regions of all iris images in CASIAV1 were automatically detected and replaced with a circular region of constant intensity to mask out the specular reflections.

1.7 Problem Definition

The error rates of a biometric identification system are dramatically increasing with the size of database. If FAR and FRR indicates the False Acceptance Rate and False Rejection Rates respectively in verification mode, then FARN and FRRN are the false acceptance and false rejection rates in the authentication mode is given by

$$FARN = 1 - (1 - FAR)N \quad (1.3)$$

$$\approx NFAR \quad (1.4)$$

$$FARN = FRR \quad (1.5)$$

$$\text{Number of false acceptance} = N(FARN) \approx N^2FAR \quad (1.6)$$

N is the number of enrolled user in the database. There are two ways to reduce the error in verification mode. 1. By reducing the FAR. 2. By reducing the search space (N). The FAR is depends upon the matching algorithm we are using and the speed and accuracy is depends upon how many templates we are checking

for matching. If we calculating the search space by a fraction of N 'sp' then the calculated FAR is given by:

$$FAR_{spN} = 1 - (1 - FAR)^{spN} \quad (1.7)$$

$$\approx (spN) FAR \quad (1.8)$$

$$FRR_{spN} = FRR \quad (1.9)$$

This minimizes the number of records against which search has to be performed this in turn reduces FAR during identification. So more emphasis is required to develop an indexing scheme for retrieving the query image with less time [5].

Chapter 2

2.1 Literature survey

The idea of automated biometrics system was proposed in 1987 by Flom and Safir [5]. The first operational iris biometric system has been developed at University of Cambridge by Daugman [5]. The image of iris captured using near infrared light source because it can control the illumination. And it will not unaffected the users badly. The next process is to find out the iris from captured image. A deformable template is trained with some parameters and shape of the eye to guide the detection process [7]. Daugman presumed iris and pupil boundaries to be circular thus the boundary of circle can be described with three parameters: radius r , center of the circle x_0, y_0 [16]. Here in this section we are discussing the previous studies and methods used to solve the biometric database indexing.

2.2 Indexing Biometric Databases using Pyramid Technique

To reduce the search space, we need a kind of classification, partitioning or indexing of the database. There exist well-established procedures such as Henry classification system to classify fingerprint records based on the ridge patterns such as 'whorl', 'loop', 'arches' and 'tented arch'. However, the problem with the existing.

Pyramid Technique

The pyramid technique normalizing the data values lies in between zero and one and then dividing the d dimensional data space into 2^d number of pyramids and

each has a common tip point with $(d-1)$ dimensional base. For example in the case of 2 dimensional data space the number of tip point will be one and the number of base of the pyramid will be four. The figure given below showing the above mentioned example. In this if we want to enroll a person in the database obtaining the five samples of needed biometric template and find out the average. The pyramids are numbered in a logical manner, such that a pyramid is numbered 'i' if all the points within the pyramid are farthest from the tip-point along dimension I than any other dimension [8]. Further, we check if the points within the pyramid have their i th coordinate less than or equal to 0.5, in which case, the pyramid is labeled as i , else it is labeled as $(i+d)$, d being the dimensionality of data. For instance, in the figure- 1 shown above, all points within pyramids $p1$ and $p3$ are farthest from the tip point along dimension $d1$ than along dimension $d0$ and points in $p3$ have $d1$ values greater than 0.5[17].

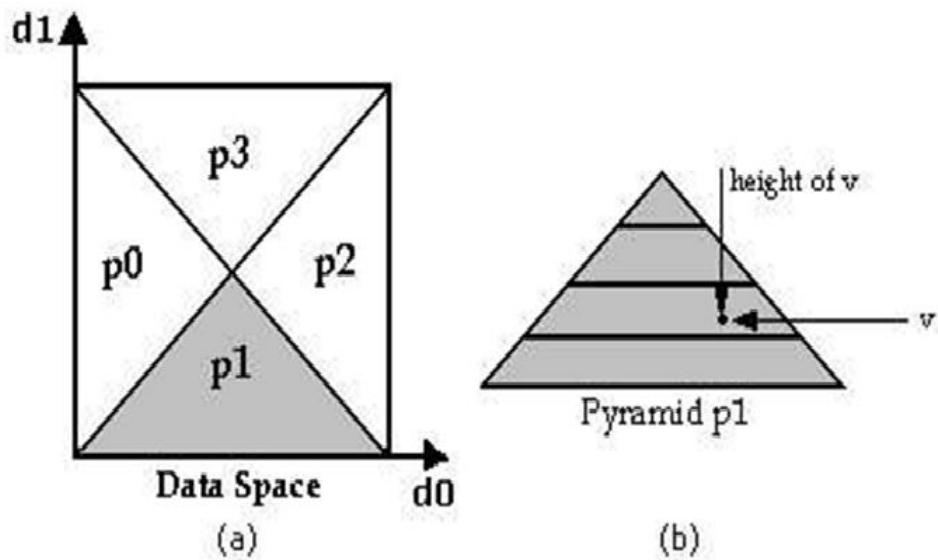


Figure 2.1: Genuine and imposter matching score distribution of biometric database showing various performance measures

The height of the pyramid is the distance from the tip point from the $(i \bmod d)$ direction. All points lying within pyramids $p1$ and $p3$ above, have the height defined as the distance from the tip-point along dimension $d1$. The $(\text{height} + \text{pyramid number})$ forms the key (pyramid value) for every template and is used as the indexing key in the B+ Tree. On receiving an input template for the

identification task, we bound this template by a bounding box to allow a level of tolerance for the feature values of the template by determining the lower and upper bounds[17].

$$Lowerbound_i = F_i - tol_{avgstd} \quad Upperbound_i = F_i + tol_{avgstd} \quad (2.1)$$

F_i : Value of feature- i of the test template

$avgstd_i$: average intra-user standard deviation for feature i tol : *Tolerance – scale factor* tol_{avgstd} *determines the tolerance for each feature.*

By performing the above mentioned operation on each of the dimensions will get a hyper rectangle. To find the candidate templates, we first determine the pyramids that this bounding box intersects. For every intersecting pyramid, we obtain the templates that lie within the range of the bounding box within the pyramid, by performing a range query. These templates form our candidate set for 1:1 matching to be performed by the final stage using another more reliable biometric.

2.2.1 Analysis

A theoretical analysis of the number of comparisons done to find all candidate templates lying within the bounding box of a given test sample yields the following:

$$Number\ of\ Comparisons = \Sigma(2 \log_2(Nm)) + (2D) \approx O(\log_2(N/m)) \quad (2.2)$$

where,

D: dimensionality of the data, 35 in our case

N: total number of templates in the database

m: minimum number of templates to be held per intermediate node

(2xD): comparisons needed to determine the intersecting pyramids

$\log_2(N/m)$: the height of the B+ Tree

$2x\log_2(N/m)$: number of comparisons to get the leaf nodes representing the lower and upper bounds.

Once obtained the leaf nodes corresponding to the bounds, simply have to do a linear scan through the chain of the leaf nodes of the B+ Tree for getting the

final candidate set for further investigation. Thus we show theoretically that the comparisons needed to determine the candidate set for the final stage, that is, the total time for pruning, is logarithmic in the total number of samples N in the database. Now we experimentally show that the size of the pruned candidate set determined by the indexing system is much smaller than the original size of the database.

2.2.2 Result

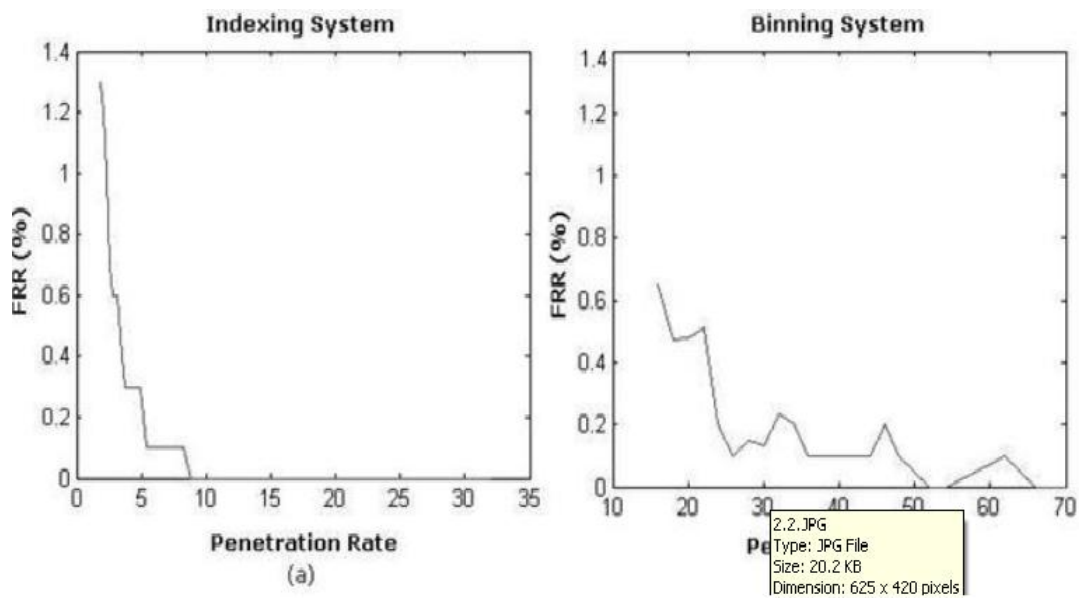


Figure 2.2: Comparison of FRR v/s Penetration Rate for Indexing and Binning Systems

In this method they presented a framework for implementing indexing in biometric databases and to reduce the search space for the identification tasks. We have showed that on using the Pyramid technique for indexing biometric databases, it can prune the database to 8.86

2.3 Indexing Iris Biometric Database Using Energy Histogram of DCT Subbands

2.3.1 Preprocessing and feature Extraction

During preprocessing the iris is transformed into rectangular strip. Features are extracted using DCT of each 8x8 block. It has been observed that multiresolution decomposition provides useful discrimination between texture. Each block of the computed DCT coefficients has to be reordered to form subbands like 3 level wavelet decomposition. The block of size 88 is reordered to transform coefficients into multiresolution form. For a coefficient $D(u, v)$ of the block, ordering is done and stored in S_i where i is defined by [18]

$$i=0 \text{ for } m=0$$

$$(m-1) \cdot 3 + (a/m) \cdot 2 + (b/m) \text{ otherwise}$$

Let $m = \max(a, b)$ for $2^{a-1} \leq u \leq 2^a$ and $2^{b-1} \leq v \leq 2^b$ and , a and b are the integer values and i ranges from 1 to 10. After reordering, the coefficients $D(1, 1)$, $D(1, 2)$, $D(2, 1)$ and $D(2, 2)$ are stored in subband S_1 , S_2 , S_3 and S_4 respectively. The multiresolution subband ordering for 8 8 block is shown in Fig 2.4

	1	2	3	4	5	6	7	8
1	S_1	S_2	S_5		S_8			
2	S_3	S_4	S_5					
3	S_6		S_7		S_8			
4								
5	S_9				S_{10}			
6								
7	S_9				S_{10}			
8								

Figure 2.3: Comparison of FRR v/s Penetration Rate for Indexing and Binning Systems

2.3.2 Key Generation

The energy histogram (H_i) is build for each subband (S_i) using all the images in the database. This presents the distribution of energy for each subband. Dividing the histogram generated from the subbands into bins. The texture details having the same energy values will be in the same bin for getting accurate match. Images are falling under the each bin is represented on the histogram. The size of the bin may be fixed or variable here it is used as fixed. This bin number is used to form a global key for indexing. Image key consists of bin number corresponding to each subband. The bin numbers for each subband recombined together using Morton order traversal which places low-frequency coefficients before high-frequency coefficients. For example the image I using Morton order forms the key as (3-5-7-8-2-1-4-5-6-7). Similarly all the images in the database obtains keys.

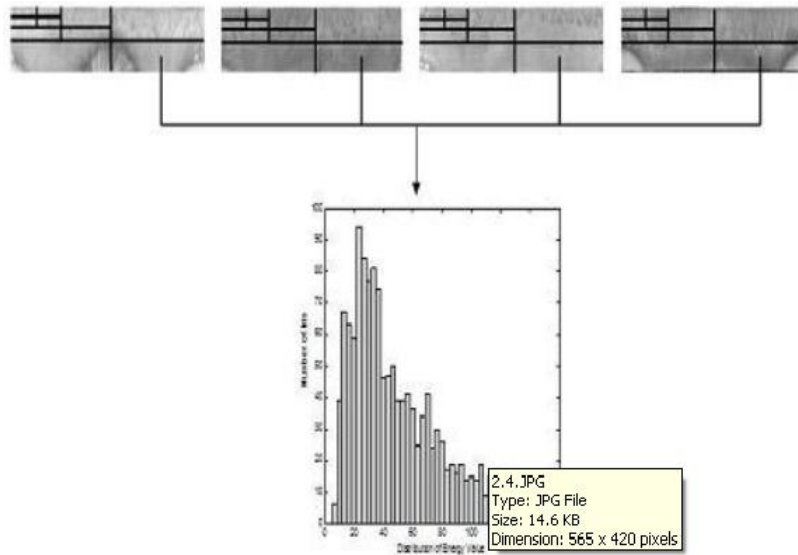


Figure 2.4: 5Energy Histogram of S10 region

2.3.3 Database creation and searching

The key is used for inserting an image in the database at the time of enrollment. The data structure using to store an iris template is B tree. The degree of the tree is total number of bins that has been constructed for each subband. The height of tree is equal to the number of subband s_i that has been taken into consideration.

The root node of the tree represents subband S0 with bins as children that are formed from energy histogram. The leftmost branch represents the first bin and then the next branch represents the second bin and so on. Each node in the second level of the tree corresponds to the immediate following subband. To insert a template in the database, B tree is traversed using the image . After reaching at the leaf node the template is inserted in the database. Each leaf node in the tree is denoted as a class that contains iris templates. The tree structure used for indexing is given in Fig. Thus more the number of classes lesser will be the retrieval time. The correct match for query strip is finding by searching the database using the full or partial key. Each block of the image is partitioned into subbands using multiresolution reordering of coefficients. The coefficients of each subband is used to compute the energy values. The query image is compared with the retrieved images to find a suitable match. However if the complete key is used for traversing the tree then the probability of finding exact match becomes less[18].

2.3.4 Experimental Results

The performance of an identification system is measured in terms of bin miss rate and penetration rate. Bin miss rate is obtained by counting the number of genuine biometric samples that has been miss-placed in a wrong class [6]. Penetration rate is defined as the percentage of total database to be scanned on an average for each search. The lower the penetration rate, more efficient the system. A comparative study on performance rates is done by changing the number of subbands. The below figure 2.6 shows the penetration rate of the system.

2.4 Indexing technic using B+tree

In this paper an efficient indexing technic of the identification system with large biometric database has been proposed. This using B+tree for redusing the retrival time. Here the multidimensional feature vector will conert to single dimensional value then it is using to index the b+tree.

Subband(#)	CASIA			BATH			IITK		
	Class(#)	BM%	PR%	Class(#)	BM%	PR%	Class(#)	BM%	PR%
1	2	0.00	99.69	5	04	26.14	5	1.5	41.44
2	5	1.60	35.96	23	12	7.69	19	5.0	17.21
3	16	3.60	22.70	66	26	3.04	46	5.5	9.24
4	39	13.2	10.23	130	36	1.42	93	10.0	4.77
5	82	24.0	6.12	197	38	0.92	148	12.5	3.25
6	158	30.8	3.46	313	56	0.49	252	15.5	1.56
7	233	35.6	2.63	399	60	0.30	396	20.5	0.92
8	304	40.0	1.77	492	70	0.16	584	29.0	0.50
9	387	42.0	1.22	583	72	0.09	744	37.5	0.27
10	519	43.6	0.63	648	72	0.06	856	44.0	0.20

Figure 2.5: Bin miss rate for change in number of classes

2.4.1 indexing method

Consider there are N feature vectors F1,F2,F3..FN and each feature vector having m dimensions.

$$F1=[f11,f12f1m]$$

$$F2=[f21,f22f2m]$$

.

.

$$FN=[fN1,fN2..fNm]$$

$$\text{Feature vector } FCi=[f1i,f2ifNi]$$

These vectors are considered as the key for indexing. The generalized structure of modified b=tree is given in fig: These feature vectors FC1,FC2..FCm are considered as keys of indexing. The generalized structure of Modified B+tree shown in figure6,where FE is the feature value as range Ri and IDs are the identifiers as multiple data pointers[14].

2.4.2 Insertion in modified Tree

In insertion procedure thea key is inserting into the modified B+ tree . Modified B+ tree T is formed for every feature value of FC1,FC2FCm.

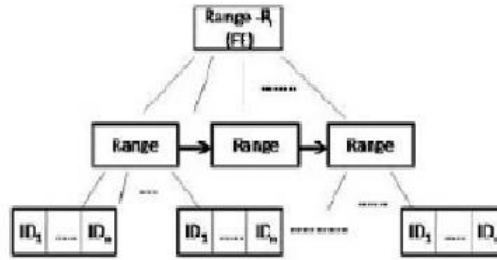


Figure 2.6: The Generalised structure of B+tree

For the given feature value, a range computed and then traversing the pto finding the appropriate leaf node. . defined B+ tree

Algorithm 1 insert (feature value v): In modified B+ tree

1. Compute range R_i for feature value v .
2. Determine the node containing the range r_i .
3. If the range node r_i is found then
4. Insert only ID of v in the range node and Insert ID of v
5. else
6. create node for the range R_i Insert ID of v .
7. end if

2.4.3 Searching in Modified B+ tree

Given a Modified B+ tree consisting of n range values as a key, whose each range value is feature values of N feature vectors, The problem is to search the smallest range in which a given query vector Q of m dimension $[q_1, q_2, \dots, q_m]$ lies [9]. For every feature value q_i of query template Q , it computes the range R_i , By traversing the tree we end with leaf node having the range R_i . A set of IDs is extracted corresponding to the feature value q_i of the query template.

Algorithm 2 search (feature value q): in Modified B+ tree

1. Compute range R_i for the query feature value q .
2. R_i be the input search range and RANGE be the range stored in the nodes.
3. Start the searching at the root.
4. If we encounter an internal node v then search for R_i among the RANGE stored at v .
5. search for R_i among the RANGE stored at v .
6. If $R_i \leq \text{RANGE min at } v$ then
7. Follow the left child pointer.
8. End if
9. If $\text{RANGE}_i \leq R_i \leq \text{RANGE}_{i+1}$ for two consecutive RANGE_i and RANGE_{i+1} for two consecutive RANGE_i and RANGE_{i+1} at v then
10. Follow the left child pointer of RANGE_{i+1}
11. End if
12. If $R_i \leq \text{RANGE}_m$ at v then
13. Follow the right pointer of RANGE_m
14. End if
15. End if
16. If we encounter a leaf node l then stored RANGE.
17. End if

2.5 Experimental Result

This indexing technique is tested for iris biometric database. Features are extracted from the iris image using localization of inner and outer iris boundaries. In the proposed strategy acquired iris image is localized using circular through transformation [9]. Input iris image is used to find the edge image. This method reduces the search space of the database by 56.4 percentage of the total database with zero percentage FRR in Bath database for dimension 88 and database 52.7 percentage with zero FRR in 20 dimension. And for IITK database reduces the search space of the database by 45.6 percentage of the total database with 6 percentage FRR in Bath database for dimension 88 and database 66.3 percentage with 6.8 FRR in 40 dimension.

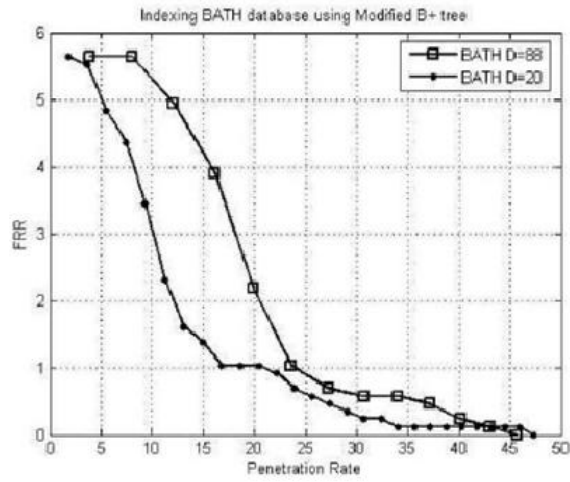


Figure 2.7: FRR vs Penetration rate Graph for Bath database for varrious error rate

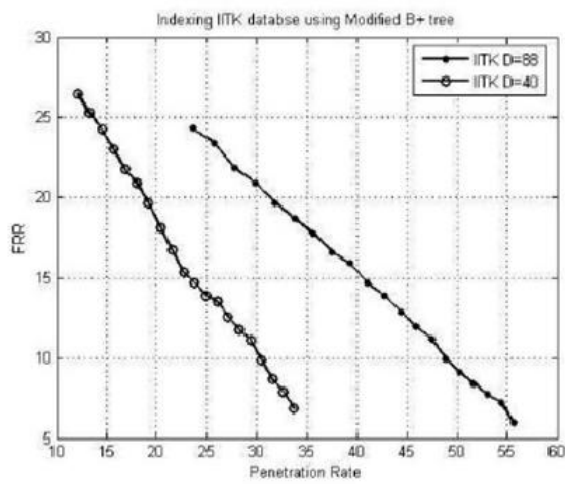


Figure 2.8: FRR vs Penetration rate Graph for IITK database for varrious error rate

Chapter 3

3.1 Proposed Method

Here the DCT is used for extracting the features of the iris strip. Here we are proposed three kinds of division of blocks and three methods of extracting features from that. In identification mode the biometric system recognizes individuals by searching all templates entered previously in the database. Here a one to many matching is doing for detecting the identity of individual. The two main challenges related to the identification system are one is high delay due to large number of comparisons. As the size of database increases the time required to declare an individual's identity increases significantly [10]. Secondly, it has been observed mathematically that the number of false positives (FAR) also increases geometrically with increase in the database size [11]. Two ways to improve the performance of biometric system are one is to reduce the false positives and the second is to reduce the search space. Here proposing a method to reduce the search space of the biometric database. In the proposed method using a DCT based approach to extract the feature vectors of the iris. Here iris strips are using for extracting the feature vectors. The reason behind using DCT that it has strong energy compaction. The strong capability of the DCT to compress energy makes the DCT a good candidate for pattern recognition applications. The DCT decomposes a signal into its elementary frequency components. When applied to an $M \times N$ image/matrix, the DCT compresses all the energy/information of the image and concentrates it in a few coefficients located in the upper-left corner of the resulting real-valued $M \times N$ DCT/frequency matrix [7].

3.1.1 Feature Extraction using DCT

Features are extracted from fixed size normalized strip using Discrete Cosine Transformation (DCT) [12]. Raw iris image needs to be preprocessed. It involves detection of inner and outer iris boundary using Circular Hough Transform [11]. The annular region lying between the two boundaries is transformed into a rectangular block [13]. The transformed rectangular region is enhanced to improve the texture details and make it illumination invariant [14]. The intensity variation along the whole image is computed by finding the mean of 1616 block. This is for finding the back ground illumination. The mean image is then expanded to the size of the original image by using bi-cubic Interpolation. At last the background illumination is subtracted from the original image. These images then corrected using histogram equalization method. The features are extracted from the preprocessed image using multiresolution DCT. DCT has strong energy compaction property and its coefficients represent some dominant gray level variations of the image. The reason behind using block based DCT approach is that it extracts local texture details of an image. It has been observed that multiresolution decomposition provides useful discrimination between textures. Here we are discussing three methods to divide the input iris strip

1. 10x10 block.
2. 8x8 block.

After finding the DCT then reordering the coefficients and it is in three methods. One is slicing of 8x8block and the other two are square windowing and L-slice method.

3.1.2 Square windowing

In the square window method, let a_{mn} designate the coefficient in the DCT matrix located in the m th row and n th column. Then a 1x1 window, generates the vector $W_{1 \times 1} = [a_{11}]$. Similarly, a 2x2 window generates the vector $W_{2 \times 2} = [a_{11} \ a_{12} \ a_{21} \ a_{22}]$ and a 3x3 window produces the vector $W_{3 \times 3} = [a_{11} \ a_{12} \ a_{13} \ a_{21} \ a_{22} \ a_{23} \ a_{31} \ a_{32} \ a_{33}]$. After that obtaining the energy value of each 10 window by finding the sum of squares of each window. The feature vector consists of different energy

values from ten windows. The logical diagram shown in figure:1

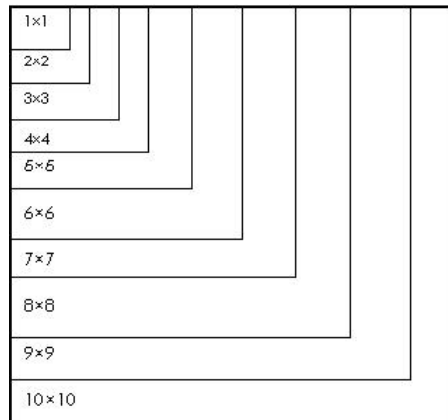


Figure 3.1: square windowing

3.1.3 L-slice method

In this method the 10x10 DCT coefficients located in the m th row and n th column. slice1 generates the vector $W_{1 \times 1} = [a_{11}]$. Similarly, a slice2(SL2)generates the vector , $[a_{21} \ a_{22} \ a_{23}]$ and a slice3 (SL3)window produces the vector $[a_{31} \ a_{32} \ a_{33} \ a_{34}]$,and so on. After extracting each L-slice the energy values of all windows corresponds to the window.

Energy (E_i) = sum of squares of a_{ij} . From i th slice of all blocks corresponds to an iris strip.

3.1.4 8x8 block slicing

In this obtaining 15 slices from each 8x8 block. Like did in previous methods finding the energy value of each slice. Slice1 generates the vector $[a_{11}]$. Similarly, a slice2 generates the vector, $[a_{21} \ a_{22}]$ and a slice3 window produces the vector $[a_{31} \ a_{32} \ a_{33}]$ and so on Energy (E_i) = sum of squares of a_{ij} . From i th slice of all blocks belongs to an iris strip.

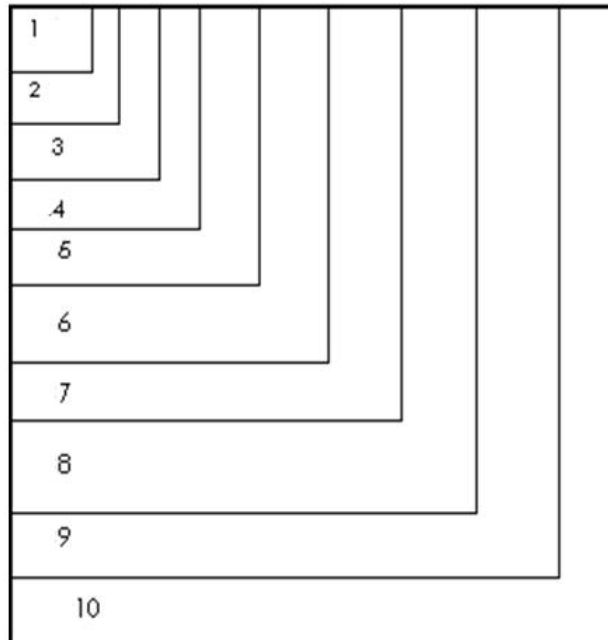


Figure 3.2: 10x10 L slicing

3.2 Indexing using key

In database the query time should be depends on the number of templates matching with the query template and not to the maximum size of the database. The partitioning of the database is like that the images having the similar texture patterns are indexed together

3.3 Key generation

The feature vector obtained from the image consists of the following number of energy values. In the case of 8x8 slice 15 and 10 each in the case of square windowing and L-slicing respectively. The obtained energy values are grouping into bins in such a way that the possible maximum energy value is dividing into ten equal units, so the number of bins will be 10 and the energy value of the window or slice is within the range of a bin are giving the bin number. Here the size of bin is fixed (10) for experimental work. So the key consists of the bin number and is using for indexing the database. For example the image key consists of the bin numbers (1-2-8-5-6-9-3-2-10-7). In the case of slicing method the size of the key would be 15.

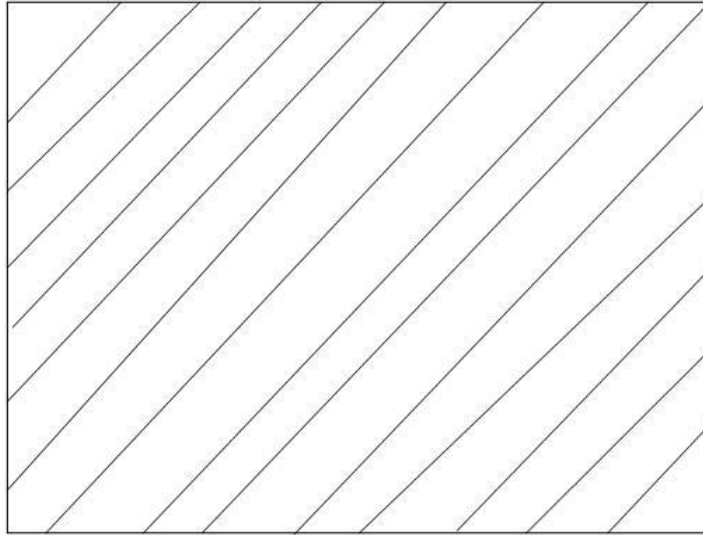


Figure 3.3: 8x8 slicing

3.4 Database Creation and Searching

The key obtained is using for the enrollment of the image in the database. Here a tree structure is using for storing the images. The degree of the node consists of the number of bins used. The height of the database is the number of window or slice we are using. The root node represents the window1 or slice1 with bins as children that are formed using bin number checking. The leftmost branch represents the first bin and the next branch represents the second bin and so on. Each node in the second level of the tree corresponds to the immediate following window or slice. To insert a template in the database, tree is traversed using the image key generated. After traversing up to the leaf node the template is inserted in the database. Each leaf node in the tree is represented as a class that contains iris templates. The tree structure used for indexing is given in Figure. Thus, the number of classes and the retrieval time are inversely proportional.

3.5 Tree Structure

The tree structure is B tree format. In this the number of levels of the tree is equal to the size of the key. The number of entries or the number of child of a node is equal to the partitioning of the total maximum energy values. The leaf

node of the tree storing the corresponding iris template.

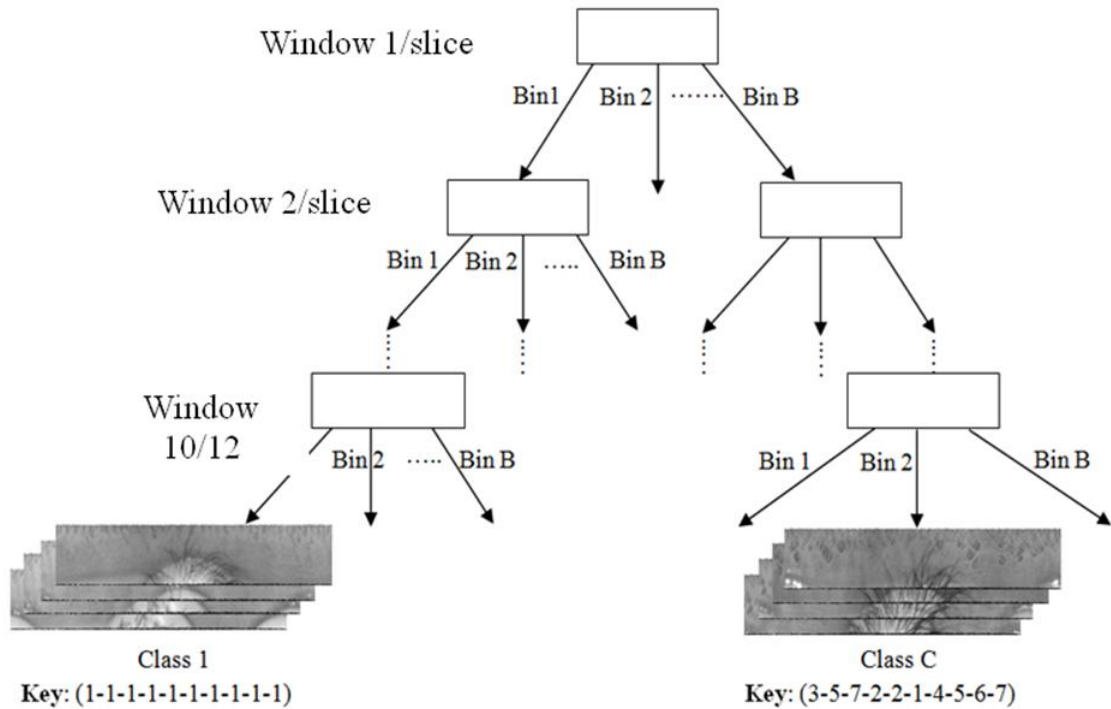


Figure 3.4: Tree structure

3.5.1 Algorithm1 (Tree indexing)

- 1) For every image I in the database.
- 2) Find the DCT of 10x10 blocks.
- 3) Rearranging DCT coefficients into windows/slices.
- 4) Find the total energy value of the each window/slice.
- 5) Grouping windows into Bins.
- 6) Obtain the key using window/slice number.
- 7) Traverse the tree.

3.5.2 Algorithm2 (matching)

- 1) Find the DCT of 10x10 blocks.
- 2) Rearranging DCT coefficients into windows/slices.
- 3) Find the total energy value of the each windows/slices.
- 4) Find the bin number (keys) of each window/slice.

- 5) Traverse the tree using full or partial key.
- 6) Retrieve the candidate sets of templates and compare.
- 7) Find the match.

3.6 Algorithm2 (Experimental Evaluation and Results)

The performance of an identification system is measured in terms of bin miss rate and penetration rate. Bin miss rate is obtained by counting the number of genuine biometric samples that has been mis-placed in a wrong class. Penetration rate is defined as the percentage of total database to be scanned on an average for each search. The lower the penetration rate, more efficient the system. In estimating penetration rate it is assumed that the search does not stop on finding the match but continues through the entire partition the evaluation has been done on CASIA database.

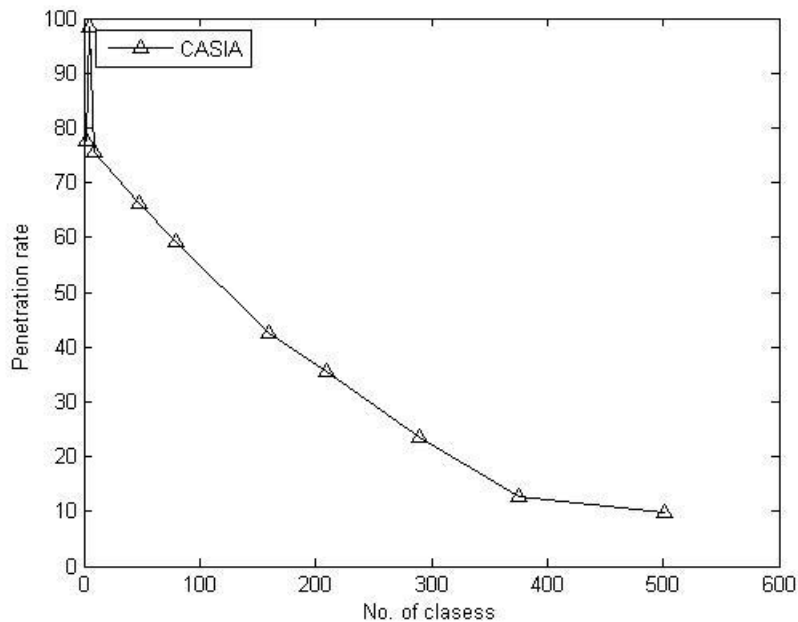


Figure 3.5: suare windowing

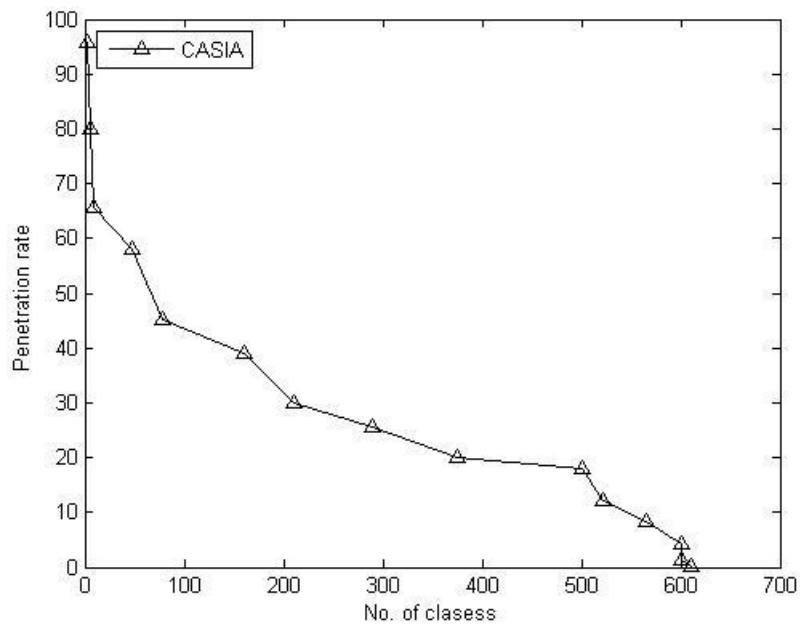


Figure 3.6: 8x8 window

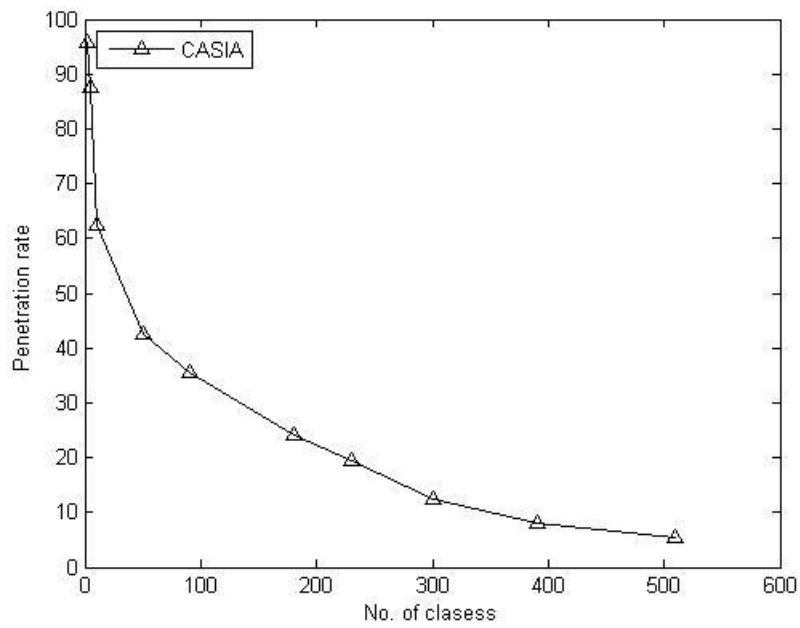


Figure 3.7: L-slice

No. of Keys	10x10 Window squaring		L-slicing		8x8 slicing	
	NO OF CLASSES	PENETRATION RATE	NO OF CLASSES	PENETRATION RATE	NO OF CLASSES	PENETRATION RATE
1	2	97.7	2	95.7	2	95.7
2	5	78.5	5	87.5	5	80
3	9	75.5	10	62.5	9	65.5
4	48	66.3	50	42.5	48	58
5	79	59.1	90	35.5	79	45.1
6	160	42.6	180	24	160	39
7	210	35.5	230	19.5	210	30.0
8	290	23.5	300	12.5	290	25.5
9	315	12.58	390	8	75	20
10	501	9.8	510	5.5	501	18
11		-----	-----	-----	521	12
12		-----	-----	-----	565	8.2
13		-----	-----	-----	600	4.3
14		-----	-----	-----	601	1.2
15		-----	-----	-----	611	.2

Figure 3.8: Tabular representation of penetration rate and number of classes

Chapter 4

4.1 Conclusion And Future work

Among the three partitioning method we have discussed the efficient searching method the 10x10 square windowing gives a penetration rate of 9.8 percentage L-slicing method given penetration rate of 5.5 and the 8x8 slicing has given a penetration rate of .2 percentage of the total database. The length of key is depending upon the application context and level of security. From the results we can understand that the system can be deployed for filtering the database using partial keys. This reduces the penetration rate of the biometric system by grouping the irises with similar texture information. Further, an efficient matching strategy can be applied on the filtered database for finding the exact match. In future the performance of proposed indexing scheme can be extended in the context of invariance to scale and rotation.

Bibliography

- [1] MA Norwell. Personal identification in networked society. Kluwer Academic Publishers, USA, 1998.
- [2] Anil K. Jain, Patrick Flynn, and Arun A. Ross. *Handbook of Biometrics*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.
- [3] Clyde Oyster. *The Human Eye Structure and Function*. Sinauer Associates, 1999.
- [4] J.L. Wayman. Error rate equations for the general biometric system. *Robotics Automation Magazine, IEEE*, 6:35 –48.
- [5] Manesh Kokare, B. N. Chatterji, and P. K. Biswas. Cosine-modulated wavelet based texture features for content-based image retrieval. *Pattern Recognition Letters*, 25(4):391 – 398, 2004.
- [6] M. Golfarelli, D. Maio, and D. Malton. On the error-reject trade-off in biometric verification systems. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):786 –796, jul 1997.
- [7] Ahmad M. Sarhan. Iris recognition using discrete cosine transform and artificial neural networks. *Journal of Computer Science*, 25(4), may.
- [8] Hans-Peter Kriegel Stefan Berchtold, Christian Bohm. In *The Pyramid-Technique: Towards Breaking the Curse of Dimensionality*, Washington, 1998. Proc. Int. Conf. on Management of Data, ACM SIGMOD,.
- [9] Devdatt and Phalguni Gupta Umarnani Jayaraman, Surya Prakash. International conference on wavelet analysis and pattern recognition. In *An Indexing Technique For Biometric Database*, pages 758–763, Hong Kong, Aug 2008.
- [10] Ruud Bolle and Sharath Pankanti. *Biometrics, Personal Identification in Networked Society: Personal Identification in Networked Society*. Kluwer Academic Publishers, Norwell, MA, USA, 1998.

- [11] S. Chikkerur A. Mhatre and V. Govindaraju. In international conference on audio and video based biometric person authentication. *Indexing biometric databases using pyramid technique.*, pages 841–849, jul 2005.
- [12] S. A. Khayam. The discrete cosine transform (dct): Theory and application. Technical report, Michigan State University, 2003.
- [13] J. Daugman. How iris recognition works. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I–33 – I–36 vol.1, 2002.
- [14] Li Ma, Tieniu Tan, Yunhong Wang, and Dexin Zhang. Efficient iris recognition by characterizing key local variations. *Image Processing, IEEE Transactions on*, 13(6):739 –750, june 2004.
- [15] Atherton T.J Kerbyson, D.J. Circle detection using hough transform filter. In *Fifth International Conference on Image Processing and its Applications*, pages 370–374, 1995.