Attendance system essay

Student Attendance System Based On Fingerprint Recognition and One-to-Many Matching A thesis submitted in partial ful? llment of the requirements for the degree of Bachelor of Computer Application in Computer Science by Sachin (Roll no. 107cs016) and Arun Sharma (Roll no. 107cs015) Under the guidance of : Prof. R. C. Tripathi Department of Computer Science and Engineering National Institute of Technology Rourkela Rourkela-769 008, Orissa, India 2 . Dedicated to Our Parents and Indian Scienti? c Community . 3 National Institute of Technology Rourkela Certi? cate

This is to certify that the project entitled, ' Student Attendance System Based On Fingerprint Recognition and One-to-Many Matching' submitted by Rishabh Mishra and Prashant Trivedi is an authentic work carried out by them under my supervision and guidance for the partial ful? llment of the requirements for the award of Bachelor of Technology Degree in Computer Science and Engineering at National Institute of Technology, Rourkela. To the best of my knowledge, the matter embodied in the project has not been submitted to any other University / Institute for the award of any Degree or Diploma.

Date – 9/5/2011 Rourkela (Prof. B. Majhi) Dept. of Computer Science and Engineering 4 Abstract Our project aims at designing an student attendance system which could e? ectively manage attendance of students at institutes like NIT Rourkela. Attendance is marked after student identi? cation. For student identi? cation, a ? ngerprint recognition based identi? cation system is used. Fingerprints are considered to be the best and fastest method for biometric identi? cation. They are secure to use, unique for every person and does not change in one's lifetime. Fingerprint recognition is a mature ? ld

today, but still identifying individual from a set of enrolled ? ngerprints is a time taking process. It was our responsibility to improve the ? ngerprint identi? cation system for implementation on large databases e. g. of an institute or a country etc. In this project, many new algorithms have been used e. g. gender estimation, key based one to many matching, removing boundary minutiae. Using these new algorithms, we have developed an identi? cation system which is faster in implementation than any other available today in the market. Although we are using this ? ngerprint identi? cation system for student identi? ation purpose in our project, the matching results are so good that it could perform very well on large databases like that of a country like India (MNIC Project). This system was implemented in Matlab10, Intel Core2Duo processor and comparison of our one to many identi? cation was done with existing identi? cation technique i. e. one to one identi? cation on same platform. Our matching technique runs in O(n+N) time as compared to the existing O(Nn2). The ? ngerprint identi? cation system was tested on FVC2004 and Veri? nger databases. 5 Acknowledgments We express our profound gratitude and indebtedness to Prof. B.

Majhi, Department of Computer Science and Engineering, NIT, Rourkela for introducing the present topic and for their inspiring intellectual guidance, constructive criticism and valuable suggestion throughout the project work. We are also thankful to Prof. Pankaj Kumar Sa , Ms. Hunny Mehrotra and other sta? s in Department of Computer Science and Engineering for motivating us in improving the algorithms. Finally we would like to thank our parents for their support and permitting us stay for more days to complete

this project. Date – 9/5/2011 Rourkela Rishabh Mishra Prashant Trivedi

Contents

Chapter 1 Introduction 1. 1 Problem Statement Designing a student attendance management system based on ? ngerprint recognition and faster one to many identi? cation that manages records for attendance in institutes like NIT Rourkela. 1. 2 Motivation and Challenges Every organization whether it be an educational institution or business organization, it has to maintain a proper record of attendance of students or employees for e? ective functioning of organization. Designing a better attendance management system for students so that records be maintained with ease and accuracy was an important key behind motivating this project.

This would improve accuracy of attendance records because it will remove all the hassles of roll calling and will save valuable time of the students as well as teachers. Image processing and ? ngerprint recognition are very advanced today in terms of technology. It was our responsibility to improve ? ngerprint identi? cation system. We decreased matching time by partitioning the database to one-tenth and improved matching using key based one to many matching. 13 14 CHAPTER 1. INTRODUCTION 1. 3 Using Biometrics Biometric Identi? cation Systems are widely used for unique identi? cation of humans mainly for veri? cation and identi? ation. Biometrics is used as a form of identity access management and access control. So use of biometrics in student attendance management system is a secure approach. There are many types of biometric systems like ? ngerprint recognition, face recognition, voice recognition, iris recognition, palm recognition etc. In this project, we used ? ngerprint recognition system. 1. 4 What is ? ngerprint? A ? ngerprint is the pattern of ridges and valleys on the surface of a ? ngertip. The endpoints and crossing points of ridges are called minutiae. It is a widely accepted assumption that the minutiae pattern of each ? ger is unique and does not change during one's life. Ridge endings are the points where the ridge curve terminates, and bifurcations are where a ridge splits from a single path to two paths at a Y-junction. Figure 1 illustrates an example of a ridge ending and a bifurcation. In this example, the black pixels correspond to the ridges, and the white pixels correspond to the valleys. Figure 1. 1: Example of a ridge ending and a bifurcation When human ? ngerprint experts determine if two ? ngerprints are from the same ? nger, the matching degree between two minutiae pattern is one of the most important factors.

Thanks to the similarity to the way of human ? ngerprint experts and compactness of templates, the minutiae-based matching method is the most widely studied matching method. 1. 5. WHY USE FINGERPRINTS? 15 1. 5 Why use ? ngerprints? Fingerprints are considered to be the best and fastest method for biometric identi? cation. They are secure to use, unique for every person and does not change in one's lifetime. Besides these, implementation of ? ngerprint recognition system is cheap, easy and accurate up to satis? ability. Fingerprint recognition has been widely used in both forensic and civilian applications.

Compared with other biometrics features , ? ngerprint-based biometrics is the most proven technique and has the largest market shares . Not only it is faster than other techniques but also the energy consumption by such systems is too less. 1. 6 Using ? ngerprint recognition system for attendance management Managing attendance records of students of an institute is a tedious task. It consumes time and paper both. To make all the attendance related work automatic and on-line, we have designed an attendance management system which could be implemented in NIT Rourkela.

It uses a ? ngerprint identi? cation system developed in this project. This ? ngerprint identi? cation system uses existing as well as new techniques in ? ngerprint recognition and matching. A new one to many matching algorithm for large databases has been introduced in this identi? cation system. 1. 7 Organization of the thesis This thesis has been organized into ten chapters. Chapter 1 introduces with our project. Chapter 2 explains the proposed design of attendance management system. Chapter 3 explains the ? ngerprint identi? cation system used in this project.

Chapter 4 explains enhancement techniques, Chapter 5 explains feature extraction methods, Chapter 6 explains our database partitioning approach . Chapter 7 explains matching technique. Chapter 8 explains experimental work done and performance analysis. Chapter 9 includes conclusions and Chapter 10 introduces proposed future work. Chapter 2 Attendance Management Framework Manual attendance taking and report generation has its limitations. It is well enough for 30-60 students but when it comes to taking attendance of students large in number, it is di? cult. For taking attendance for a lecture, a conference, etc. oll calling and manual attendance system is a failure. Time waste over responses of students, waste of paper etc. are the disadvantages of manual attendance system. Moreover, the attendance report is also not generated on time. Attendance report which is circulated over NITR webmail is two months old. To overcome these non-optimal situations, it is necessary that we should use an automatic on-line attendance management system. So we present an implementable attendance management framework. Student attendance system framework is divided into three parts : Hardware/Software Design, Attendance Management Approach and On-line Report Generation.

Each of these is explained below. 2. 1 Hardware – Software Level Design Required hardware used should be easy to maintain, implement and easily available. Proposed hardware consists following parts: (1)Fingerprint Scanner, (2)LCD/Display Module (optional), (3)Computer 16 2. 2. ATTENDANCE MANAGEMENT APPROACH Table 2. 1: Estimated Budget Device Cost of Number of Total Name One Unit Units Required Unit Budget Scanner 500 100 50000 PC 21000 100 2100000 Total 21, 50, 000 (4)LAN connection

17 Fingerprint scanner will be used to input ? ngerprint of teachers/students into the computer software.

LCD display will be displaying rolls of those whose attendance is marked. Computer Software will be interfacing ? ngerprint scanner and LCD and will be connected to the network. It will input ? ngerprint, will process it and extract features for matching. After matching, it will update database attendance records of the students. Figure 2. 1: Hardware present in classrooms Estimated Budget Estimated cost of the hardware for implementation of this system is shown in the table 2. 1. Total number of classrooms in NIT Rourkela is around 100. So number of units required will be 100. 2. 2 Attendance Management Approach

This part explains how students and teachers will use this attendance management system. Following points will make sure that attendance is marked correctly, without any problem: (1)All the hardware will be inside classroom. So outside interference will be absent. (2)To remove unauthorized access and unwanted attempt to corrupt the hardware by students, all the hardware except ? ngerprint scanner could be put inside a small 18 CHAPTER 2. ATTENDANCE MANAGEMENT FRAMEWORK cabin. As an alternate solution, we can install CCTV cameras to prevent unprivileged activities. (3)When teacher enters the classroom, the attendance marking will start.

Computer software will start the process after inputting ? ngerprint of teacher. It will ? nd the Subject ID, and Current Semester using the ID of the teacher or could be set manually on the software. If teacher doesn't enter

classroom, attendance marking will not start. (4)After some time, say 20 minutes of this process, no attendance will be given because of late entrance. This time period can be increased or decreased as per requirements. Figure 2. 2: Classroom Scenario 2. 3 On-Line Attendance Report Generation Database for attendance would be a table having following ? elds as a combination for primary ? ld: (1)Day,(2)Roll,(3)Subject and following non-primary ? elds: (1)Attendance,(2)Semester. Using this table, all the attendance can be managed for a student. For on-line report generation, a simple website can be hosted on NIT Rourkela servers, 2. 4. NETWORK AND DATABASE MANAGEMENT 19 which will access this table for showing attendance of students. The sql queries will be used for report generation. Following query will give total numbers of classes held in subject CS423: SELECT COUNT(DISTINCT Day) FROM AttendanceTable WHERE SUBJECT = CS423 AND Attendance = 1 For attendance of oll 107CS016, against this subject, following query will be used: SELECT COUNT(Day) FROM AttendanceTable WHERE Roll = 107CS016 AND SUBJECT = CS423 AND Attendance = 1 Now the attendance percent can easily be calculated :

$$\text{Attendance} = \frac{\text{ClassesAttended}}{\text{ClassesHeld}} \times 100 \quad (2. 1)$$

2. 4 Network and Database Management This attendance system will be spread over a wide network from classrooms via intranet to internet. Network diagram is shown in ? g. 2. 3. Using this network, attendance reports will be made available over internet and e-mail. A monthly report will be sent to each student via email and website will show the updated attendance.

Entity relationship diagram for database of students and attendance records is shown in ? g. 2. 4. In ER diagram, primary ? elds are Roll, Date, SubjectID

and TeacherID. Four tables are Student, Attendance, Subject and Teacher. Using this database, attendance could easily be maintained for students. Data? ow is shown in data ? ow diagrams (DFD) shown in ? gures 2. 5, 2. 6 and 2. 7. 2. 5 Using wireless network instead of LAN and bringing portability We are using LAN for communication among servers and hardwares in the classrooms. We can instead use wireless LAN with portable devices.

Portable device will have an embedded ? ngerprint scanner, wireless connection, a microprocessor loaded with a software, memory and a display terminal, see ? gure 2. 5. Size of device could be small like a mobile phone depending upon how well the device is manufactured. 20 CHAPTER 2. ATTENDANCE MANAGEMENT FRAMEWORK Figure 2. 3: Network Diagram 2. 5. USING WIRELESS NETWORK INSTEAD OF LAN AND BRINGING PORTABILITY21 Figure 2. 4: ER Diagram 22 CHAPTER 2. ATTENDANCE MANAGEMENT FRAMEWORK Figure 2. 5: Level 0 DFD Figure 2. 6: Level 1 DFD 2. 5. USING WIRELESS NETWORK INSTEAD OF LAN AND BRINGING PORTABILITY23 Figure 2. : Level 2 DFD 24 CHAPTER 2. ATTENDANCE MANAGEMENT FRAMEWORK This device should have a wireless connection. Using this wireless connection, Figure 2. 8: Portable Device attendance taken would be updated automatically when device is in network of the nodes which are storing the attendance records. Database of enrolled ? ngerprints will be in this portable device. Size of enrolled database was 12. 1 MB when 150 ? ngerprints were enrolled in this project. So for 10000 students, atleast 807 MB or more space would be required for storing enrolled database. For this purpose, a removable memory chip could be used.

We cannot use wireless LAN here because fetching data using wireless LAN will not be possible because of less range of wireless devices. So enrolled data would be on chip itself. Attendance results will be updated when portable device will be in the range of nodes which are storing attendance reports. We may update all the records online via the mobile network provided by di? erent companies. Today 3G network provides su? cient throughput which can be used for updating attendance records automatically without going near nodes. In such case, 2. 6. COMPARISON WITH OTHER STUDENT ATTENDANCE SYSTEMS 25 he need of database inside memory chip will not be mandatory. It will be fetched by using 3G mobile network from central database repository. The design of such a portable device is the task of embedded system engineers. 2. 5. 1 Using Portable Device In this section, we suggest the working of portable device and the method of using it for marking attendance. The device may either be having touchscreen input/display or buttons with lcd display. A software specially designed for the device will be running on it. Teachers will verify his/her ? ngerprint on the device before giving it to students for marking attendance.

After verifying the teacher's identity, software will ask for course and and other required information about the class which he or she is going to teach. Software will ask teacher the time after which device will not mark any attendance. This time can vary depending on the teacher's mood but our suggested value is 25 minutes. This is done to prevent late entrance of students. This step will hardly take few seconds. Then students will be given device for their ? ngerprint identi? cation and attendance marking. In the continuation, teacher will start his/her lecture.

Students will hand over the device to other students whose attendance is not marked. After 25 minutes or the time decided by teacher, device will not input any attendance. After the class is over, teacher will take device and will end the lecture. The main function of software running on the device will be ? ngerprint identi? cation of students followed by report generation and sending reports to servers using 3G network. Other functions will be downloading and updating the database available on the device from central database repository. 2. 6 Comparison with other student attendance systems

There are various other kind of student attendance management systems available like RFID based student attendance system and GSM-GPRS based student attendance system. These systems have their own pros and cons. Our system is better because ? rst it saves time that could be used for teaching. Second is portability. Portability 26 CHAPTER 2. ATTENDANCE MANAGEMENT FRAMEWORK has its own advantage because the device could be taken to any class wherever it is scheduled. While GSM-GPRS based systems use position of class for attendance marking which is not dynamic and if schedule or location of the class changes, wrong attendance might be marked.

Problem with RFID based systems is that students have to carry RFID cards and also the RFID detectors are needed to be installed. Nonetheless, students may give proxies easily using friend's RFID card. These problems are not in our system. We used ? ngerprints as recognition criteria so proxies cannot be given. If portable devices are used, attendance marking will be done at any place and any time. So our student attendance system is far better to be implemented at NITR. Chapter 3 Fingerprint Identi? cation

System An identi? cation system is one which helps in identifying an individual among any people when detailed information is not available. It may involve matching available features of candidate like ? ngerprints with those already enrolled in database. 3. 1 How Fingerprint Recognition works? Fingerprint images that are found or scanned are not of optimum quality. So we remove noises and enhance their quality. We extract features like minutiae and others for matching. If the sets of minutiae are matched with those in the database, we call it an identi? ed ? ngerprint. After matching, we perform post-matching steps which may include showing details of identi? ed candidate, marking attendance etc.

A brief ? owchart is shown in next section. 3. 2 Fingerprint Identi? cation System Flowchart A brief methodology of our Fingerprint Identi? cation System is shown here in following ? owchart. Each of these are explained in the later chapters. 27 28 CHAPTER 3. FINGERPRINT IDENTIFICATION SYSTEM Figure 3. 1: Fingerprint Identi? cation System Flowchart Chapter 4 Fingerprint Enhancement The image acquired from scanner is sometimes not of perfect quality . It gets corrupted due to irregularities and non-uniformity in the impression taken and due to variations in the skin and the presence of the scars, humidity, irt etc. To overcome these problems , to reduce noise and enhance the de? nition of ridges against valleys, various techniques are applied as following. 4. 1 Segmentation Image segmentation [1] separates the foreground regions and the background regions in the image. The foreground regions refers to the clear ? ngerprint area which contains the ridges and valleys. This is the area of interest. The background regions refers

to the regions which is outside the borders of the main ? ngerprint area, which does not contain any important or valid ? ngerprint information.

The extraction of noisy and false minutiae can be done by applying minutiae extraction algorithm to the background regions of the image. Thus, segmentation is a process by which we can discard these background regions, which results in more reliable extraction of minutiae points. We are going to use a method based on variance thresholding . The background regions exhibit a very low grey – scale variance value , whereas the foreground regions have a very high variance . Firstly , the image is divided into blocks and the grey-scale variance is calculated for each block in the image .

If the variance is less than the global threshold , then the block is assigned to be part of background region or else 29 30 CHAPTER 4. FINGERPRINT ENHANCEMENT it is part of foreground . The grey – level variance for a block of size S x S can be calculated as : 1 V ar(k) = 2 S S? 1 S? 1 (G(i, j) ? M (k))2 i= 0 j= 0 (4. 1) where Var(k) is the grey – level variance for the block k , G(i, j) is the grey – level value at pixel (i, j) , and M(k) denotes the mean grey – level value for the corresponding block k . 4. 2 Normalization Image normalization is the next step in ? ngerprint enhancement process.

Normalization [1] is a process of standardizing the intensity values in an image so that these intensity values lies within a certain desired range. It can be done by adjusting the range of grey-level values in the image. Let G(i, j) denotes the grey-level value at pixel (i, j), and N(i, j) represent the normalized grey-level value at pixel (i, j). Then the normalized image can de?

ned as: ? ? M + 0 N (i, j) = ? M ? 0 V0 (G(i, j)? M)2 V V0 (G(i, j)? M)2 V , if I(i, j) > M , otherwise where M0 and V0 are the estimated mean and variance of I(i, j), respectively . 4. 3 Orientation estimation The orientation ? eld of a ? ngerprint image de? es the local orientation of the ridges contained in the ? ngerprint . The orientation estimation is a fundamental step in the enhancement process as the subsequent Gabor ? ltering stage relies on the local orientation in order to e? ectively enhance the ? ngerprint image. The least mean square estimation method used by Raymond Thai [1] is used to compute the orientation image. However, instead of estimating the orientation block-wise, we have chosen to extend their method into a pixel-wise scheme, which produces a ? ner and more accurate estimation of the orientation ? eld. The steps for calculating the orientation at pixel i, j) are as follows: 4. 3. ORIENTATION ESTIMATION 31 1. Firstly , a block of size W x W is centered at pixel (i, j) in the normalized ? ngerprint image. 2. For each pixel in the block, compute the gradients dx (i, j) and dy (i, j), which are the gradient magnitudes in the x and y directions, respectively. The horizontal Sobel operator[6] is used to compute dx(i, j) : [1 0 -1; 2 0 -2; 1 0 -1] Figure 4. 1: Orientation Estimation 3. The local orientation at pixel (i; j) can then be estimated using the following equations: i+ W 2 j+ W 2 Vx (i, j) = u= i? W 2 i+ W 2 v= j? W 2 j+ W 2 2? x (u, v)? y (u, v) (4. 2) Vy (i, j) = u= i? W v= j? W 2 2 2 2 ? (u, v) ? ? y (u, v), (4. 3) ?(i, j) = 1 Vy (i, j) tan? 1 , 2 Vx (i, j) (4. 4) where ? (i, j) is the least square estimate of the local orientation at the block centered at pixel (i, j). 4. Smooth the orientation ? eld in a local neighborhood using a Gaussian ? lter. The orientation image is ? rstly converted into a continuous vector ? eld, which is de? ned as: ? x (i, j) = cos 2? (i, j), ? y (i, j) = sin 2? (i, j), (4. 5) (4. 6) where ? x and ? y are the x and y

components of the vector ? eld, respectively. After 32 CHAPTER 4. FINGERPRINT ENHANCEMENT the vector ? eld has been computed, Gaussian smoothing is then performed as follows: w? w? 2 ? x (i, j) = w? u=? 2 w? v=? 2 G(u, v)? x (i ? uw, j ? vw), (4. 7) w? 2 w? 2 ? y (i, j) = w? u=? 2 w? v=? 2 G(u, v)? y (i ? uw, j ? vw), (4. 8) where G is a Gaussian low-pass ? lter of size w? x w? . 5. The ? nal smoothed orientation ? eld O at pixel (i, j) is de? ned as: O(i, j) = ? y (i, j) 1 tan? 1 2 ? x (i, j) (4. 9) 4. 4 Ridge Frequency Estimation Another important parameter, in addition to the orientation image, that can be used in the construction of the Gabor ? lter is the local ridge frequency. The local frequency of the ridges in a ? ngerprint is represented by the frequency image. The ? st step is to divide the image into blocks of size W x W. In the next step we project the greylevel values of each pixels located inside each block along a direction perpendicular to the local ridge orientation. This projection results in an almost sinusoidal-shape wave with the local minimum points denoting the ridges in the ? ngerprint. It involves smoothing the projected waveform using a Gaussian lowpass ? lter of size W x W which helps in reducing the e? ect of noise in the projection. The ridge spacing S(i, j) is then calculated by counting the median number of pixels between the consecutive minima points in the projected waveform.

The ridge frequency F(i, j) for a block centered at pixel (i, j) is de? ned as: F (i, j) = 1 S(i, j) (4. 10) 4. 5. GABOR FILTER 33 4. 5 Gabor ? lter Gabor ? lters [1] are used because they have orientation-selective and frequencyselective properties. Gabor ? lters are called the mother of all other ? lters as other ? lter can be derived using this ? lter. Therefore, applying a properly tuned Gabor ? lter can preserve the ridge structures while reducing noise. An even-

symmetric Gabor ? lter in the spatial domain is de? ned as : 1 x2 y2 G(x, y, ? , f) = exp{? [? + ?]} cos 2? f x? , 2 2 2 ? x ? y (4. 11) x? = x cos ? + y sin ? , (4. 12) y? ? x sin ? + y cos ? , (4. 13) where ? is the orientation of the Gabor ? lter, f is the frequency of the cosine wave, ? x and ? y are the standard deviations of the Gaussian envelope along the x and y axes, respectively, and x? and y? de? ne the x and y axes of the ? lter coordinate frame respectively. The Gabor Filter is applied to the ? ngerprint image by spatially convolving the image with the ? lter. The convolution of a pixel (i, j) in the image requires the corresponding orientation value O(i, j) and the ridge frequency value F(i, j) of that pixel . wy 2 wx 2 E(i, j) = u=? wx 2 w v=? 2y G(u, v, O(i, j), F (i, j))N (i ? u, j ? v), (4. 4) where O is the orientation image, F is the ridge frequency image, N is the normalized ? ngerprint image, and wx and wy are the width and height of the Gabor ? lter mask, respectively.

34 CHAPTER 4. FINGERPRINT ENHANCEMENT 4. 6 Binarisation Most minutiae extraction algorithms operate on basically binary images where there are only two levels of interest: the black pixels represent ridges, and the white pixels represent valleys. Binarisation [1] converts a greylevel image into a binary image. This helps in improving the contrast between the ridges and valleys in a ? ngerprint image, and consequently facilitates the extraction of minutiae.

One very useful property of the Gabor ? lter is that it contains a DC component of zero, which indicates that the resulting ? ltered image has a zero mean pixel value. Hence, binarisation of the image can be done by using a global threshold of zero. Binarisation involves examining the grey-level value of every pixel in the enhanced image, and, if the grey-level value

is greater than the prede? ned global threshold, then the pixel value is set to value one; else, it is set to zero. The outcome of binarisation is a binary image which contains two levels of information, the background valleys and the foreground ridges. . 7 Thinning Thinning is a morphological operation which is used to remove selected foreground pixels from the binary images. A standard thinning algorithm from [1] is used, which performs this operation using two subiterations. The algorithm can be accessed by a software MATLAB via the ' thin' operation of the bwmorph function. Each subiteration starts by examining the neighborhood of every pixel in the binary image, and on the basis of a particular set of pixel-deletion criteria, it decides whether the pixel can be removed or not. These subiterations goes on until no more pixels can be removed.

Figure 4. 2: (a)Original Image, (b)Enhanced Image, (c)Binarised Image, (d)Thinned Image Chapter 5 Feature Extraction After improving quality of the ? ngerprint image we extract features from binarised and thinned images. We extract reference point, minutiae and key(used for one to many matching). 5. 1 Finding the Reference Point Reference point is very important feature in advanced matching algorithms because it provides the location of origin for marking minutiae. We ? nd the reference point using the algorithm as in [2]. Then we ? nd the relative position of minutiae and estimate the orientation ? ld of the reference point or the singular point. The technique is to extract core and delta points using Poincare Index. The value of Poincare index is 180o , ? 180o and 0o for a core, a delta and an ordinary point respectively. Complex ? lters are used to produce blur at di? erent resolutions. Singular point (SP) or reference point is the point of maximum ?

lter response of these ? lters applied on image. Complex ? lters , $\exp(im?)$, of order m (= 1 and -1) are used to produce ? lter response. Four level resolutions are used here: level 0, level 1, level 2, level 3.

Level 3 is lowest resolution and level 0 is highest resolution. Only ? lters of ? rst order are used : $h = (x + iy)m\, g(x, y)$ where $g(x, y)$ is a gaussian de? ned as $g(x, y) = \exp? ((x2 + y 2)/2? 2)$ and m = 1, ? 1. Filters are applied to the complex valued orientation tensor ? eld image $z(x, y) = (fx + ify)2$ and not directly to the image. Here f x is the derivative of the original image in the x-direction and f y is the derivative in the y-direction. To ? nd the position of a possible 35 36 CHAPTER 5. FEATURE EXTRACTION Figure 5. 1: Row 1: ? lter response $c1k$, k = 3, 2, and 1. Row 2: ? ter response $c2k$, k = 3, 2, and 1. SP in a ? ngerprint the maximum ? lter response is extracted in image c13 and in c23 (i. e. ? lter response at m = 1 and level 3 (c13) and at m = ? 1 and level 3 (c23)). The search is done in a window computed in the previous higher level (low resolution). The ? lter response at lower level (high resolution) is used for ? nding response at higher level (low resolution). At a certain resolution (level k), if $cnk (xj , yj)$ is higher than a threshold an SP is found and its position (xj , yj) and the complex ? lter response $cnk (xj , yj)$ are noted. 5. 2 5. 2. 1

Minutiae Extraction and Post-Processing Minutiae Extraction The most commonly employed method of minutiae extraction is the Crossing Number (CN) concept [1] . This method involves the use of the skeleton image where the ridge ? ow pattern is eight-connected. The minutiae are extracted by scanning the local neighborhood of each ridge pixel in the image using a 3 x 3 window. The CN value is then computed, which is de? ned as half the sum

of the di? erences between pairs of adjacent pixels in the eight-neighborhood. Using the properties of the CN as shown in ? gure 5, the ridge pixel can then be classi? d as a ridge ending, bifurcation or non-minutiae point. For example, a ridge pixel with a CN of one corresponds to a ridge ending, and a CN of three corresponds to a bifurcation. 5. 2. MINUTIAE EXTRACTION AND POST-PROCESSING Table 5. 1: Properties of Crossing Number CN Property 0 Isolated Point 1 Ridge Ending Point 2 Continuing Ridge Point 3 Bifurcation Point 4 Crossing Point 37 Figure 5. 2: Examples of (a)ridge-ending (CN= 1) and (b)bifurcation pixel (CN= 3) 5. 2. 2 Post-Processing False minutiae may be introduced into the image due to factors such as noisy images, and image artefacts created by the thinning process.

Hence, after the minutiae are extracted, it is necessary to employ a post-processing [1] stage in order to validate the minutiae. Figure 5. 3 illustrates some examples of false minutiae structures, which include the spur, hole, triangle and spike structures . It can be seen that the spur structure generates false ridge endings, where as both the hole and triangle structures generate false bifurcations. The spike structure creates a false bifurcation and a false ridge ending point. Figure 5. 3: Examples of typical false minutiae structures : (c)Triangle, (d)Spike (a)Spur, (b)Hole, 38 CHAPTER 5.

FEATURE EXTRACTION 5. 2. 3 Removing Boundary Minutiae For removing boundary minutiae, we used pixel-density approach. Any point on the boundary will have less white pixel density in a window centered at it, as compared to inner minutiae. We calculated the limit, which indicated that pixel density less than that means it is a boundary minutiae. We calculated it according to following formula: limit = (w w ? (ridgedensity)) ? Wf req 2 (5.

1) where w is the window size, Wf req is the window size used to compute ridge density. Figure 5. 4: Skeleton of window centered at boundary minutiae

Figure 5. 5: Matrix Representation of boundary minutiae Now, in thinned image, we sum all the pixels in the window of size w centered at the boundary minutiae. If sum is less than limit, the minutiae is considered as boundary minutiae and is discarded. 5. 3. EXTRACTION OF THE KEY 39 5. 3 5. 3. 1 Extraction of the key What is key? Key is used as a hashing tool in this project. Key is small set of few minutiae closest to reference point. We match minutiae sets, if the keys of sample and query ? ngerprints matches. Keys are stored along with minutiae sets in the database.

Advantage of using key is that, we do not perform full matching every time for non-matching minutiae sets, as it would be time consuming. For large databases, if we go on matching full minutiae set for every enrolled ? ngerprint, it would waste time unnecessarily. Two types of keys are proposed – simple and complex. Simple key has been used in this project. Figure 5. 6: Key Representation Simple Key This type of key has been used in this project. Minutiae which constitute this key are ten minutiae closest to the reference point or centroid of all minutiae, in sorted 40 CHAPTER 5. FEATURE EXTRACTION order. Five ? lds are stored for each key value i. e. (x, y, ? , t, r). (x, y) is the location of minutiae, ? is the value of orientation of ridge related to minutia with respect to orientation of reference point, t is type of minutiae, and r is distance of minutiae from origin. Due to inaccuracy and imperfection of reference point detection algorithm, we used centroid of all minutiae for construction of key. Complex Key The complex key stores more information and is structurally more complex. It stores vector of minutiae in

which next minutiae is closest to previous minutiae, starting with reference point or centroid of all minutiae.

It stores < x, y, ? , t, r, d, ? >. Here x, y, t, r,? are same, d is distance from previous minutiae entry and ? is di? erence in ridge orientation from previous minutiae. Data: minutiaelist = Minutiae Set, refx = x-cordinate of centroid, refy = y-cordinate of centroid Result: Key d(10)= null; for j = 1 to 10 do for i = 1 to rows(minutiaelist) do d(i) Chapter 6 Partitioning of Database Before we partition the database, we perform gender estimation and classi? cation. 6. 1 Gender Estimation In [3], study on 100 males and 100 females revealed that signi? cant sex di? erences occur in the ? ngerprint ridge density.

Henceforth, gender of the candidate can be estimated on the basis of given ? ngerprint data. Henceforth, gender of the candidate can be estimated on the basis of given ? ngerprint data. Based on this estimation, searching for a record in the database can be made faster. Method for ? nding mean ridge density and estimated gender: The highest and lowest values for male and female ridge densities will be searched. If ridge density of query ? ngerprint is less than the lowest ridge density value of females, the query ? ngerprint is obviously of a male. Similarly, if it is higher than highest ridge density value of males, the query ? gerprint is of a female. So the searching will be carried out in male or female domains. If the value is between these values, we search on the basis of whether the mean of these values is less than the density of query image or higher. 41 42 CHAPTER 6. PARTITIONING OF DATABASE Figure 6. 1: Gender Estimation 6. 1. GENDER ESTIMATION Data: Size of Database = N; Ridge Density of query ? ngerprint = s Result: Estimated Gender i. e. male or female maleupperlimit= 0; femalelowerlimit=

20; mean= 0; for image < femalelowerlimit then femalelowerlimit 43 if s < maleupperlimit then estimatedgender 44 CHAPTER 6.

PARTITIONING OF DATABASE 6. 2 Classi? cation of Fingerprint We divide ? ngerprint into ? ve classes – arch or tented arch, left loop, right loop, whorl and unclassi? ed. The algorithm for classi? cation [4] is used in this project. They used a ridge classi? cation algorithm that involves three categories of ridge structures: nonrecurring ridges, type I recurring ridges and type II recurring ridges. N1 and N2 represent number of type I recurring ridges and type II recurring ridges respectively. Nc and Nd are number of core and delta in the ? ngerprint. To ? nd core and delta, separate 135o blocks from orientation image. 35o blocks are shown in following ? gures. Figure 6. 2: 135o blocks of a ? ngerprint Based on number of such blocks and their relative positions, the core and delta are found using Poincare index method. After these, classi? cation is done as following: 1. If (N2 > 0) and (Nc = 2) and (Nd = 2), then a whorl is identi? ed. 2. If (N1 = 0) and (N2 = 0) and (Nc = 0) and (Nd = 0), then an arch is identi? ed. 3. If (N1 > 0) and (N2 = 0) and (Nc = 1) and (Nd = 1), then classify the input using the core and delta assessment algorithm[4]. 4. If (N2 > T2) and (Nc > 0), then a whorl is identi? ed. 5.

If (N1 > T1) and (N2 = 0) and (Nc = 1) then classify the input using the core and delta assessment algorithm[4]. 6. If (Nc = 2), then a whorl is identi? ed. 7. If (Nc = 1) and (Nd = 1), then classify the input using the core and delta assessment algorithm[4]. 8. If (N1 > 0) and (Nc = 1), then classify the input using the core and delta assessment algorithm. 6. 3. PARTITIONING 9. If (Nc = 0) and (Nd = 0), then an arch is identi? ed. 10. If none of the above

conditions is satis? ed, then reject the ? ngerprint. 45 Figure 6. 3: Fingerprint Classes (a)Left Loop, (b)Right Loop, (c)Whorl, (d1)Arch, (d2)Tented Arch . 3 Partitioning After we estimate gender and ? nd the class of ? ngerprint, we know which ? ngerprints to be searched in the database. We roughly divide database into one-tenth using the above parameters. This would roughly reduce identi? cation time to one-tenth. 46 CHAPTER 6. PARTITIONING OF DATABASE Figure 6. 4: Partitioning Database Chapter 7 Matching Matching means ? nding most appropriate similar ? ngerprint to query ? ngerprint. Fingerprints are matched by matching set of minutiae extracted. Minutiae sets never match completely, so we compute match score of matching. If match score satis? s accuracy needs, we call it successful matching. We used a new key based one to many matching intended for large databases. 7. 1 Alignment Before we go for matching, minutiae set need to be aligned(registered) with each other. For alignment problems, we used hough transform based registration technique similar to one used by Ratha et al[5]. Minutiae alignment is done in two steps minutiae registration and pairing. Minutiae registration involves aligning minutiae using parameters $< \Delta x, \Delta y, \theta >$ which range within speci? ed limits. $(\Delta x, \Delta y)$ are translational parameters and θ is rotational parameter.

Using these parameters, minutiae sets are rotated and translated within parameters limits. Then we ? nd pairing scores of each transformation and transformation giving maximum score is registered as alignment transformation. Using this transformation $< \Delta x, \Delta y, \theta >$, we align query minutiae set with the database minutiae set. Algorithm is same as in [5] but we have excluded factor ? s i. e. the scaling parameter because it does not

a? ect much the alignment process. ? lies from -20 degrees to 20 degrees in steps of 1 or 2 generalized as < ? 1 , ? 2 , ? 3 ...? k > where k is number of rotations applied.

For every query minutiae i we check if ? k + ? i = ? j where ? i and ? j are orientation 47 48 CHAPTER 7. MATCHING parameters of ith minutia of query minutiae set and j th minutia of database minutiae set. If condition is satis? ed, A(i, j, k) is ? agged as 1 else 0. For all these ? agged values, (? x, ? y) is calculated using following formula: ? (? x , ? y) = qj ? ? cos? sin? ? ? ? pi , (7. 1) ? sin? cos? where qj and pi are the coordinates of j th minutiae of database minutiae set and ith minutiae of query minutiae set respectively. Using these < ? x, ? y, ? k > values, whole query minutiae set is aligned.

This aligned minutiae set is used to compute pairing score. Two minutiae are said to be paired only when they lie in same bounding box and have same orientation. Pairing score is (number of paired minutiae)/(total number of minutiae). The i, j, k values which have highest pairing score are ? nally used to align minutiae set. Co-ordinates of aligned minutiae are found using the formula: ? qj = ? cos? sin? ? ? ? pi + (? x , ? y), (7. 2) ? sin? cos? After alignment, minutiae are stored in sorted order of their distance from their centroid or core. 7. 2 Existing Matching Techniques

Most popular matching technique of today is the simple minded n2 matching where n is number of minutiae. In this matching each minutiae of query ? ngerprint is matched with n minutiae of sample ? ngerprint giving total number of n2 comparisons. This matching is very orthodox and gives headache when identi? cation is done on large databases. 7. 3 One to Many

matching Few algorithms are proposed by many researchers around the world which are better than normal n2 matching. But all of them are one to one veri? cation or one to one identi? cation matching types. We developed a one to many matching technique which uses key as the hashing tool.

Initially, we do not match minutiae sets instead we per- 7. 3. ONE TO MANY MATCHING 49 form key matching with many keys of database. Those database ? ngerprints whose keys match with key of query ? ngerprint, are allowed for full minutiae matching. Key matching and full matching are performed using k*n matching algorithm discussed in later section. Following section gives method for one to many matching. Data: Query Fingerprint; Result: Matching Results; Acquire Fingerprint, Perform Enhancement, Find Fingerprint Class, Extract Minutiae, Remove Spurious and Boundary Minutiae, Extract Key, Estimate Gender; M . 3. 1 Method of One to Many Matching The matching algorithm will be involving matching the key of the query ? ngerprint with the many(M) keys of the database. Those which matches , their full matching will be processed, else the query key will be matched with next M keys and so on. 50 Data: Gender, Class, i; Result: Matching Results; egender CHAPTER 7. MATCHING if keymatchstatus = success then eminutiae 7. 4 Performing key match and full matching Both key matching and full matching are performed using our k*n matching technique. Here k is a constant(recommended value is 15) chosen by us.

In this method, we match ith minutiae of query set with k unmatched minutiae of sample set. Both the query sets and sample sets must be in sorted order of distance from reference point or centroid. ith minutia of query minutiae list is matched with top k unmatched minutiae of database

minutiae set. This type of matching reduces matching time of n2 to k*n. If minutiae are 80 in number and we chose k to be 15, the total number of comparisons will reduce from 80*80= 6400 to 80*15= 1200. And this means our matching will be k/n times faster than n2 matching. 7. 5. TIME COMPLEXITY OF THIS MATCHING TECHNIQUE 51 Figure 7. : One to Many Matching 7. 5 Time Complexity of this matching technique Let s = size of the key, n = number of minutiae, N = number of ? ngerprints matched till successful identi? cation, k = constant (see previous section). There would be N-1 unsuccessful key matches, one successful key match, one successful full match. Time for N-1 unsuccessful key matches is (N-1)*s*k (in worst case), for successful full match is s*k and for full match is n*k. Total time is (N-1)*s*k+n*k+s*k = k(s*N+n). Here s= 10 and we have reduced database to be searched to 1/10th , so N matching technique, it would have been O(Nn2).

For large databases, our matching technique is best to use. Averaging for every ? ngerprint, we have O(1+n/N) in this identi? cation process which comes to O(1) when N >> n. So we can say that our identi? cation system has constant average matching time when database size is millions. Chapter 8 Experimental Analysis 8. 1 Implementation Environment We tested our algorithm on several databases like FVC2004, FVC2000 and Veri? nger databases. We used a computer with 2GB RAM and 1. 83 GHz Intel Core2Duo processor and softwares like Matlab10 and MSAccess10. 8. 2 8. 2. 1 Fingerprint Enhancement Segmentation and Normalization

Segmentation was performed and it generated a mask matrix which has values as 1 for ridges and 0 for background . Normalization was done with

mean = 0 and variance = 1 (? g 8. 1). Figure 8. 1: Normalized Image 52 8. 2. FINGERPRINT ENHANCEMENT 53 8. 2. 2 Orientation Estimation In orientation estimation, we used block size = 3*3. Orientations are shown in ? gure 8. 2. Figure 8. 2: Orientation Image 8. 2. 3 Ridge Frequency Estimation Ridge density and mean ridge density were calculated. Darker blocks indicated low ridge density and vice-versa. Ridge frequencies are shown in ? gure 8. 3. Figure 8. 3: Ridge Frequency Image 8. 2. 4

Gabor Filters Gabor ? lters were employed to enhance quality of image. Orientation estimation and ridge frequency images are requirements for implementing gabor ? lters. ? x and ? y are taken 0. 5 in Raymond Thai, but we used ? x = 0. 7 and ? y = 0. 7. Based on these values , we got results which were satis? able and are shown in ? gure 8. 4. 54 CHAPTER 8. EXPERIMENTAL ANALYSIS Figure 8. 4: Left-Original Image, Right-Enhanced Image 8. 2. 5 Binarisation and Thinning After the ? ngerprint image is enhanced, it is then converted to binary form, and submitted to the thinning algorithm which reduces the ridge thickness to one pixel wide.

Results of binarisation are shown in ? gure 8. 5 and of thinning are shown in ? gure 8. 6. Figure 8. 5: Binarised Image 8. 3. FEATURE EXTRACTION 55 Figure 8. 6: Thinned Image 8. 3 8. 3. 1 Feature Extraction Minutiae Extraction and Post Processing Minutiae Extraction Using the crossing number method, we extracted minutiae. For this we used skeleton image or the thinned image. Due to low quality of ? ngerprint, a lot of false and boundary minutiae were found. So we moved forward for post-processing step. Results are shown in ? gure 8. 7 and 8. 8. Figure 8. 7: All Extracted Minutiae 56 CHAPTER 8. EXPERIMENTAL ANALYSIS

Figure 8. 8: Composite Image with spurious and boundary minutiae After Removing Spurious and Boundary Minutiae False minutiae were removed using method described in earlier section. For removing boundary minutiae, we employed our algorithm which worked ? ne and minutiae extraction results are shown in table 8. 2. Results are shown in ? gure 8. 9 and 8. 10. Figure 8. 9: Minutiae Image after post-processing As we can see from table 8. 2 that removing boundary minutiae considerably reduced the number of false minutiae from minutiae extraction results. 8. 4. GENDER ESTIMATION AND CLASSIFICATION 57 Figure 8. 0: Composite Image after post-processing Table 8. 1: Average Number of Minutiae before and after post-processing DB After After Removing After Removing Used Extraction Spurious Ones Boundary Minutiae FVC2004DB4 218 186 93 FVC2004DB3 222 196 55 8. 3. 2 Reference Point Detection For reference point extraction we used complex ? lters as described earlier. For a database size of 300, reference point was found with success rate of 67. 66 percent. 8. 4 8. 4. 1 Gender Estimation and Classi? cation Gender Estimation Average ridge density was calculated along with minimum and maximum ridge densities shown in table 8. . Mean ridge density was used to divide the database into two parts. This reduced database size to be searched by half. Based on the information available about the gender of enrolled student, we can apply our gender estimation algorithm which will further increase the speed of identi? cation. 8. 4. 2 Classi? cation Fingerprint classi? cation was performed on both original and enhanced images. Results were more accurate on the enhanced image. We used same algorithm as in sec 6. 2 to classify the ? ngerprint into ? ve classes – arch, left loop, right loop, whorl and 58 CHAPTER 8.

EXPERIMENTAL ANALYSIS Figure 8. 11: Plotted Minutiae with Reference Point(Black Spot) Table 8. 2: Ridge Density Calculation Results Window Minimum Maximum Mean Total Average Size Ridge Ridge Ridge Time Time Taken Density Density Density Taken Taken 36 6. 25 9. 50 7. 87 193. 76 sec 1. 46 sec unclassi? ed. This classi? cation was used to divide the database into ? ve parts which would reduce the database to be searched to one-? fth and ultimately making this identi? cation process ? ve times faster. Results of classi? cation are shown in table 8. 4, 8. 5 and 8. 6. 8. 5 Enrolling

At the time of enrolling personal details like name, semester, gender, age, roll number etc. were asked to input by the user and following features of ? ngerprint were saved in the database (1)Minutiae Set (2)Key (3)Ridge Density (4)Class Total and average time taken for enrolling ? ngerprints in database is shown in table 8. 6. MATCHING Table 8. 3: Classi? cation Results on Original Image Class No. of (1-5) Images 1 2 2 2 3 3 4 4 5 121 Table 8. 4: Classi? cation Results on Enhanced Image Class No. of (1-5) Images 1 8 2 3 3 3 4 6 5 112 59 8. 7. All the personal details were stored in the MS Access database and were modi? d by running sql queries inside matlab. Fingerprint features were stored in txt format inside a separate folder. When txt ? le were used, the process of enrolling was faster as compared to storing the values in MS Access DB. It was due to the overhead of connections, running sql queries for MS Access DB. 8. 6 Matching Fingerprint matching is required by both veri? cation and identi? cation processes. 8. 6. 1 Fingerprint Veri? cation Results Fingerprint veri? cation is the process of matching two ? ngerprints against each other to verify whether they belong to same person or not. When a ? gerprint matches with the ? ngerprint of same individual,

we call it true accept or if it doesn't, we call it false reject. In the same way if the ? ngerprint of di? erent individuals match, we call it a false accept or if it rejects them, it is true reject. False Accept Rate (FAR) and False Reject Rate (FRR) are the error rates which are used to express matching trustability. FAR is de? ned by the formula : 60 CHAPTER 8. EXPERIMENTAL ANALYSIS Table 8. 5: Time taken for Classi? cation Image Average Total Taken Time(sec) Time(sec) Original 0. 5233 69. 07 Enhanced 0. 8891 117. 36 Table 8. : Time taken for Enrolling No. of Storage Average Total Images Type Time(sec) Time(hrs) 294 MS Access DB 24. 55 2. 046 60 MS Access DB 29. 37 0. 49 150 TXT ? les 15. 06 1. 255 F AR = FA ? 100, N (8. 1) FA = Number of False Accepts, N = Total number of veri? cations FRR is de? ned by the formula : FR ? 100, N F RR = (8. 2) FR = Number of False Rejects. FAR and FRR calculated over six templates of Veri? nger DB are shown in table 8. 8. This process took approximately 7 hours. 8. 6. 2 Identi? cation Results and Comparison with Other Matching techniques Fingerprint identi? cation is the process of identifying a query ? gerprint from a set of enrolled ? ngerprints. Identi? cation is usually a slower process because we have to search over a large database. Currently we match minutiae set of query ? ngerprint with the minutiae sets of enrolled ? ngerprints. In this project, we store key in the database at the time of enrolling. This key as explained in sec 5. 3 helps in 8. 6. MATCHING Table 8. 7: Error Rates FAR FRR 4. 56 12. 5 14. 72 4. 02 61 Figure 8. 12: Graph: Time taken for Identi? cation vs Size of Database(key based one to many identi? cation) reducing matching time over non-matching ? ngerprints. For non-matching enrolled ? gerprints, we don't perform full matching, instead a key matching. Among one or many keys which matched in one iteration of one to many matching, we allow full

minutiae set matching. Then if any full matching succeeds, we perform post matching steps. This identi? cation scheme has lesser time complexity as compared to conventional n2 one to one identi? cation. Identi? cation results are shown in table 8. 9. The graph of time versus N is shown in ? gure 8. 13. Here N is the index of ? ngerprint to be identi? ed from a set of enrolled ? ngerprints. Size of database of enrolled ? ngerprints was 150. So N can vary from