Study of Biometric Data Processing by Hadoop

Madhavi Vaidya
VES college
Mumbai

Swati Sherekar
P. G. Deptt. of Computer Science
SGB Amravati University

ABSTRACT
UID is being implemented in India providing valuable information of citizens. Various facilities are to be provided on the basis of that information to the masses. It will also reveal how it can be used for genuine identity of a person and thus improving security for various e-governance applications. In this paper idea of UID and authentication request types such as biometric have been studied. The use of Hadoop Architecture is suggested which can be used for studying the authentication request types such as biometric technique, as it is one of the strongest method of authentication. In Aadhaar Card System, Residents Photograph, Finger Prints and Iris recognition are collected. We have suggested how Hadoop, a distributed file system architecture can be used to develop this model. Hadoop implements Map/Reduce framework. Map/Reduce makes an easy task to process large amount of data on cloud or from various nodes where data is scattered.

1. INTRODUCTION
Biometrics is a rapidly evolving technology that has been widely used in forensics, such as criminal identification and prison security, and has the potential to be widely adopted in a very broad range of civilian applications:

- Banking security, such as electronic fund transfers, ATM security, check cashing, and credit card transactions;
- Physical access control, such as airport access control;
- Information system security, such as access to data bases via login privileges;
- Government benefits distribution, such as welfare disbursement programs;
- Customs and immigration, such as the Immigration and Naturalization Service Passenger Accelerated Service System (INSPASS) which permits faster immigration procedures based on hand geometry;
- National ID systems, which provide a unique ID to the citizens and integrate different government services;
- Voter and driver registration, providing registration facilities for voters and drivers.
- Currently, there are mainly nine different biometric techniques that are either widely used, including face, fingerprint, hand geometry, hand vein, iris, retinal pattern, signature, voice print and facial thermograms

This paper is structured as follows: in section 2, an overview of the work has done by Aadhaar and how in this system the biometric system is used. In section 3, the introduction of biometric system is discussed. In section 4, the various approaches in biometrics are stated and explained. Section 5 describes the HDFS architecture and how a new framework combines Hadoop and Biometric system is explained. Finally conclusions are drawn and future work is outlined in section 6.

2. OVERVIEW OF AADHAAR
Aadhaar will empower poor and underprivileged residents in accessing services such as the formal banking system and give them the opportunity to easily avail various other services provided by the Government and the private sector. The centralized technology infrastructure of the UIDAI will enable 'anytime, anywhere, anyhow' authentication. A key requirement of the Aadhaar is to minimize/eliminate duplicate identity to improve the efficacy of the service delivery. Biometrics features are selected to be the primary mechanism for ensuring uniqueness.

Technology systems will play a major role across the UIDAI infrastructure. The Aadhaar database will be stored on a central server. Enrolment of the residents will be computerized, and information exchange between Registrars and the CIDR will take place over a network. Authentication of the residents will be online. The Authority will also put systems in place for the security and safety of information.[1] Security is seen as a great resource to a company as well as to a government organization. (See Fig 1) If proper security provisions are made at hardware and software level then customers utilizing the resources would feel proud. There is a need of having different security levels for different kinds of task.[2]

In Aadhaar Card System, for authentication, biometric system has been used. Biometrics means statistical study of biological data or the identification of somebody through an electronic system. Originated from the two Greek words- *bios* means life and *metros* means measure[3], biometrics consists of methods for uniquely recognizing humans based upon one or more intrinsic physical or behavioral traits. Residents Photograph (Face Matcher), Finger Prints and Iris recognition are studied.

3. INTRODUCTION TO BIOMETRIC SYSTEM
Various authentication request types such as demographic, biometric, simple or advanced authentications are supported by this application. The Aadhaar submitted is used for 1:1 match for the resident's record. The inputs are then matched against the resident information found in the biometric database.

3.1 Importance of Biometrics
It does the following functionality:-

1. It enables identifying /authenticating individuals based on "credentials" that are hard to forge

2. It has many useful applications where establishing identity is important
3. Banks and Financial Services companies are using biometrics to prevent banking and identity fraud
4. National governments are creating biometric databases for law enforcement and security reasons:-
   a. Assist with criminal investigations (i.e. crime scene fingerprints)
   b. Identify individuals entering and leaving the country
   c. Surveillance

A unique identification means savings of huge public resources as it avoids duplication. Also, having a unique ID would make life easier for people as they could use it for different uses like in Public Distribution system (PDS), interaction with[4] The UID project entails collection of basic information such as name, date of birth, gender, father/guardians' name, and address, apart from ten fingerprints, photograph and an iris scan. The Biometric Standards Committee [5] has already prescribed certain standards (ISO 19794 series of standards) and formats with regards to biometric information.

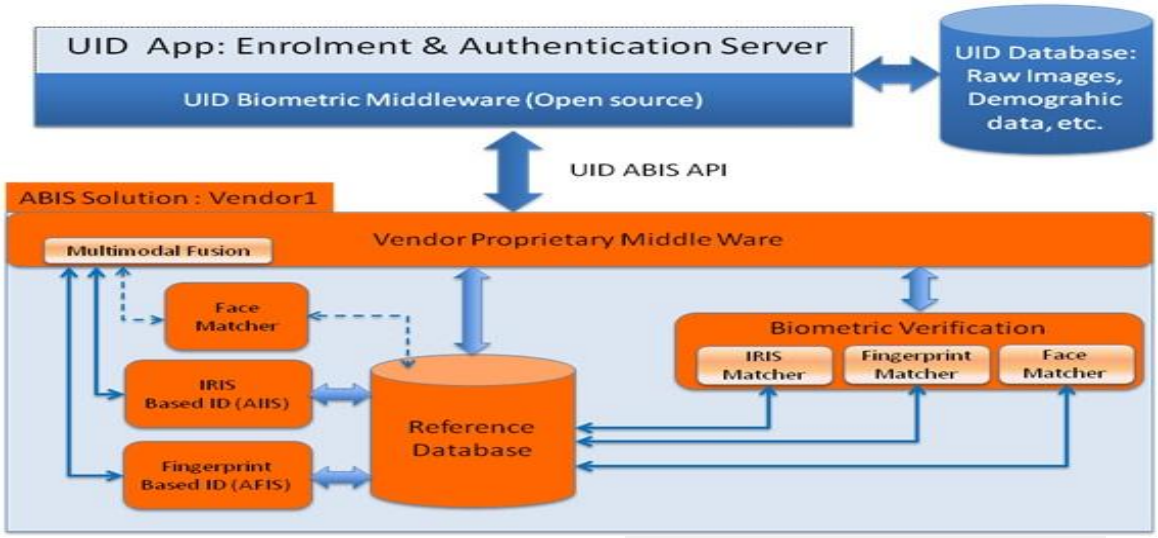

Fig 1 : Working of Biometric at Aadhaar

4. VARIOUS APPROACHES OF BIOMETRICS

4.1 Face Matcher:

The Face Matcher performs facial template matching in 1-to-1 (verification) and 1-to-many (identification) modes. Also the Face Matcher component includes **fused** matching algorithm that allows increasing template matching reliability by matching templates that contain facial image format.

A fingerprint is thus defined by the uniqueness of the local ridge characteristics and their relationships. Minutiae points are these local ridge characteristics that occur either at a ridge ending or a ridge bifurcation. A ridge ending is defined as a point where the ridge ends abruptly and the ridge bifurcation is the point where the ridge splits into two or more branches. Automatic Minutiae detection becomes a difficult task in low quality fingerprint images where noise and contrast deficiency result in pixel configuration similar to that of Minutiae.[2,3]

4.2 Iris Recognition

Iris recognition technology provides positive identification of an individual, at extremely high confidence levels. Iris scan has been developing an identification/verification system. It is useful for identifying and verifying the identity of the individual. It uses the unique patterns of the human iris, shows promise of overcoming previous shortcomings.

4.3 Finger Prints Recognition

Fingerprint identification is widely used in personal identification as it works well in most cases. However, it is difficult to acquire fingerprint features *i.e*. minutiae, for some class of persons such as manual laborers, elderly people, *etc*.. Fingerprint image enhancement and Minutiae extraction A fingerprint is the pattern of ridges and valleys.

The task of the authentication module is to authenticate the identity of the person who intends to access the system. The person to be authenticated indicates his identity and places his finger on the fingerprint scanner; a digital image of his fingerprint is captured; and a minutiae pattern is extracted from the captured fingerprint image and fed to a matching algorithm, which matches it against the person's minutiae templates stored in the system data base to establish the identity.[12]

A typical biometric system works in three distinct stages:-

The Enrollment stage,
The Verification stage and
The Identification stage.

During the enrollment stage, a user's biometric data is acquired and processed to extract a feature set that is stored in the database. The stored feature set is labeled with the

user's identity, is referred to as a template. In order to account for variations in the biometric data of a user,

multiple templates corresponding to each user may be stored. (1: N) . During verification and Identification stage users biometric data is required and extracted feature set is matched again the templates stored in database in order to identify a previously enrolled individual or to validate a claimed identity. [6] As I stated earlier there are different three ways of the same:-

1. **Enrollment:** add an identity and associated biometric data to the database if they do not already exist **(See Fig 2)**

2. **Verification:** lookup the biometric template for a single individual and determine whether it matches a captured biometric measurement (1 to 1 match) **(see Fig 3)**

3. **Identification:** determine the identity of an individual given some biometric measurements (1 to N match) **(See Fig4)**[11]

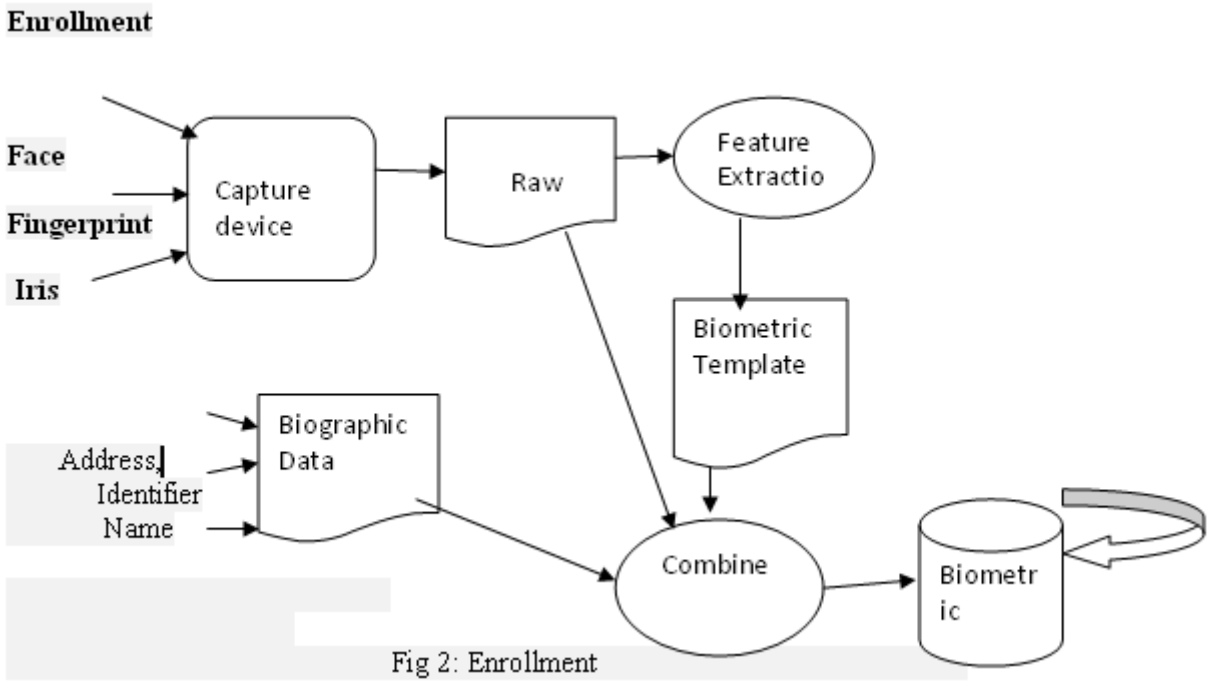

Fig 2: Enrollment

The above figure (Fig 2) states that the information such as name, address and the identifier is collected from an individual. Capture biometric data in raw form. The transformation of raw biometric data into the encoded biometric template is done. Then the information has to be stored in biometric database.

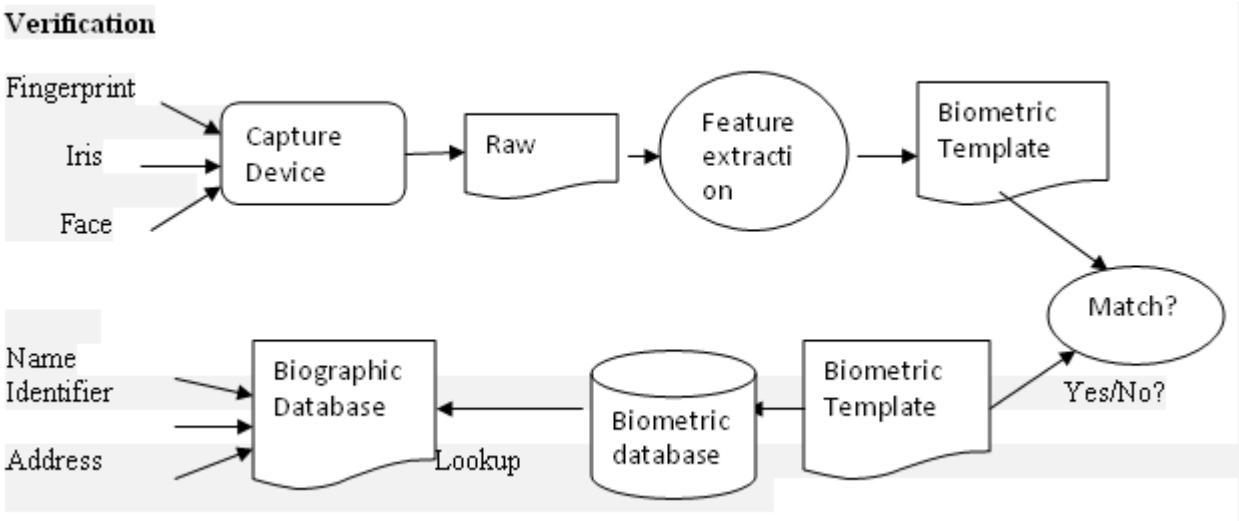

Fig 3: Verification

This figure(Fig 3) states that there is a lookup in the biometric template for a particular individual.

The matching or the verification of the stored template and recently captured template are matched.

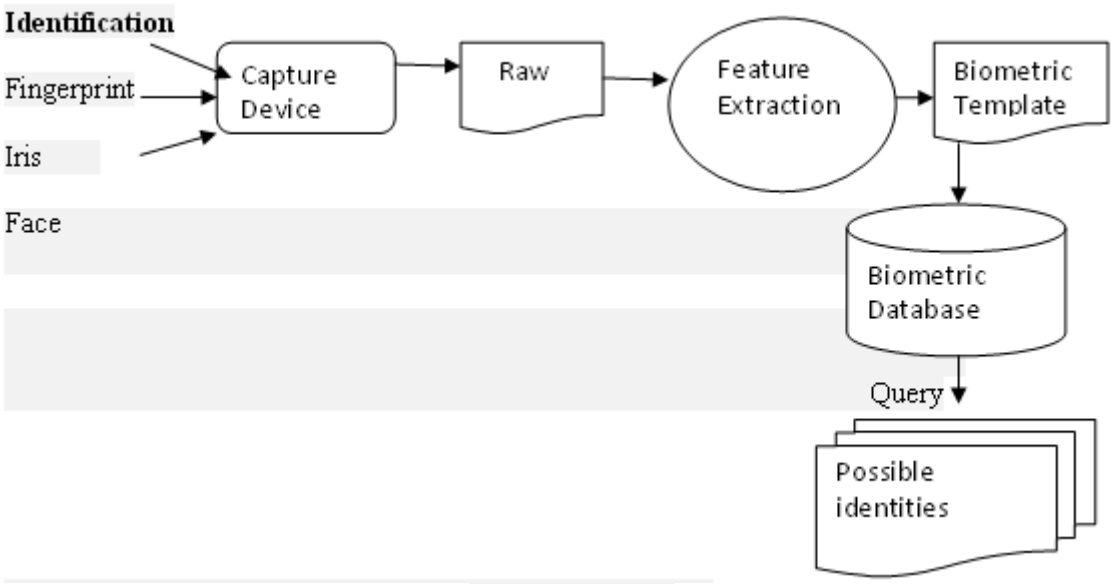

Fig 4: Identification

Some number of raw biometric measures is captured; they can be converted into biometric templates.

5. LIMITATIONS OF BIOMETRICS

1. The main reason for introducing biometric systems is to increase overall security.
2. Biometric identification is not perfect it is never 100% certain, it is vulnerable to errors and it can be 'spoofed'
3. Fraudulent reproduction of biometric data is possible; this depends heavily on the modality, application and resources being considered and availability of the data to be reproduced. [7]

The data is large in nature and to process the same from the various nodes where it has taken from citizens of India the biometric system has limitations. This data is distributed across machines i.e. it is distributed in nature. Hence we have suggested the HDFS architecture which can be used for processing the data from many nodes which is described in section 5.

6. HADOOP ON DISTRIBUTED FILE SYSTEMS

HDFS, the Hadoop Distributed File System, is a distributed file system designed to hold very large amounts of data (terabytes or even petabytes), and provide high-throughput access to this information. Files are stored in a redundant fashion across multiple machines to ensure their durability to failure and high availability to very parallel applications. Hadoop Map/Reduce is a software framework that process large amount of data, in parallel, in a fault tolerant manner.

MapReduce is highly efficient and scalable, and thus can be used to process huge datasets. HDFS has a master /slave architecture. An HDFS cluster consists of a single NameNode, a master server that manages the file system namespace and regulates access to files by clients.

In addition, there are a number of DataNodes, usually one per node in the cluster, which manages storage attached to the nodes that they run on. MapReduce programming consists of writing two functions, a map function, and a reduce function. The map function takes a key, value pair and outputs a list of intermediate values with the key. The map function is written in such a way that multiple map functions can be executed at once, so it's the part of the program that divides up tasks. The reduce function then takes the output of the map functions, and does some process on them, usually combining values, to generate the desired result in an output file.

6.1 Fingerprint Recognition

A fingerprint is the pattern of ridges and valleys on the fingertip. A fingerprint is thus defined by the uniqueness of the local ridge characteristics and their relationships. Following steps can be implemented to process the data.

6.1.1 Description of the steps of the algorithm

1. Pre-processing

A major problem [2] in AADHAAR is that more than 600 million Indians who work in agriculture, construction and other manual works have worn out fingers due to life time of hard labor. The input image is segmented from the background which ensures the removal of noise. Removal of noise is verified and checked out for the whole image. The image obtained from the above step is then normalized to get the desired given image.

2. Minutiae Extraction

The next step after enhancement of the image is the extraction of the minutiae. The enhanced image is binarized first in this step. The skeleton of the image is then formed.

The binary image is thinned as a result of which a ridge is only one pixel wide. Thus the minutiae points are those which have a pixel value of one (ridge ending) as their

neighbor or more than two (ridge bifurcations) in their neighborhood. This ends the process of extraction of minutiae points.

3. Post Processing

The Minutiae points obtained in the above step may contain spurious minutiae. This may occur due to the presence of ridge breaks in the given figure itself which could not be improved even after enhancement. This results in false

Minutiae points which needs to be removed. These unwanted Minutiae points are removed in the post processing stage. False minutiae points will be obtained at the borders as the image ends abruptly.

This processing of the minute data can be done by using Hadoop. As the data of so many residents of India will be gathered and then processed for Pre- Processing, Extraction and Post Processing, this can be used by Map Reduce algorithm. The data which is collected will be mapped and then reduced to the resultant value if the person's finger prints get matched with the already present in the database.

6.2 Iris Recognition

The iris is the colored portion of the eye surrounding the pupil. Among several physiological biometric characteristics, iris patterns have a wonderful and abundant structure and full of complex textures. The iris texture is unique from one person to other ones and can't be stolen. Also iris image can be easily captured from Distance without physical contact .These features are stable during the life.

Having localized the region of an acquired image that corresponds to the iris, the final task is to decide if this pattern matches a previously stored iris pattern. This matter of pattern matching can be decomposed into four parts:

1) Bringing the newly acquired iris pattern into spatial alignment with a candidate data base entry;

2) Choosing a representation of the aligned iris patterns that makes their distinctive patterns apparent;

3) Evaluating the goodness of match between the newly acquired and data base representations;

4) Deciding if the newly acquired data and the data base entry were derived from the same iris based on the goodness of match. [8][9]

A typical Iris recognition process consists of following main sub processes:

1) **Image Acquisition**- In image acquisition process eye image is captured for processing. The image can be acquired from live video camera or can be used already stored one in memory.

2) **Image Preprocessing**- Preprocessing includes image filtering and enhancement, image iris localization, iris normalization. Image is converted into gray scale image if it is colored one technique is applied here.

3) **Feature extraction**- Iris structure has complex and plentiful textures which can be extracted as features for iris coding. The wrapped iris pattern is converted into unwrapped rectangular iris template. Feature vector is extracted that is compared with the already stored iris templates. Integer feature vector for comparison is been used.

4) **Feature Matching** - is the final step of iris recognition. The unknown iris template and stored template is matched. This matching is performed by computing the distance between two templates.

6.3 Use of Map Reduce for Fingerprint and Iris Recognition

In this section the elaboration given on finger print and iris matching process on Map/Reduce framework. Map/Reduce framework works on <key, value> format. It works on the basis of parallel method when data can be processed in chunks from all nodes in parallel manner. Hadoop is a potential solution to the problem. As Hadoop works on the record format, so entire database is stored in a text format and uploaded on HDFS. Here the key is nothing but the image and value to check whether it is present in the database or if not present then newly can be added. Mappers take its input from HDFS and reducer stores the final result on HDFS. The format of record is as follows:

<String name, width, height, pixel values>. String name is used as key. [10]

Pixel values are the Intensity values of the image either taken from iris detection or fingerprint recognition. Width and height parameter determines how many values to be read against one key.

The input image as given above for the finger print image and iris images can be scanned and then it can be matched which is there in data base if already present , is scanned, template is extracted and then probability is calculated of each stored template against input image. The entire database is divided into chunks and distributed to all mappers. The output of each mapper is<string name, probability value>. These intermediate outputs are collected from all mappers, sorting and shuffling for all data is implemented.

The data can be put on different machines and it can be distributed across various nodes then the map reduce architecture can be used. The data can be extracted by mappers when distributed in nature and reducers are applied for the data which is scattered and then processed to verify resultant value and entered value is same.

Even the data which is scattered across nodes , from which the noise can be removed by using preprocessing and extraction techniques. The data which is cleaned now can be processed by map reduce framework in HDFS.

7. CONCLUSION

There are different ways of biometrics by taking the current system Aadhaar have been studied in this paper. Enrollment, Verification and Identification are the three different techniques at the base by which biometric system functions. The enormous data is generated from this Indian Government project i.e. UIDAI. Hadoop is a great platform for solving all sorts of such big data which gets generated from distributed nodes. The processing of the data is difficult and it can even be the noisy data. Secondly, to avoid the noise from the same, the data mining techniques can be applied and then it can be processed by Map Reduce Framework. In the future, we have planned to consider the data which is generated from mining to be taken on HDFS and to be processed by map reduce technique to improve processing performance. To keep this huge data safe the security aspect on Hadoop can also be studied.

8. REFERENCES

[1] uidai.gov.in/

[2] UNIQUE IDENTIFICATION NUMBER AND EGOVERNANCE SECURITY - Subash ChanderGovernment College, Karnal, Haryana, IndiaInternational Journal of Computing and Business Research (IJCBR) Volume 1, N. 1 December – 2010

[3] INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND MANAGEMENT (IJCSM) ISSUE 1 VOL1 Application of Fingerprint Recognition in User Identification

[4] www.uidaicards.com

[5] www.wikipedia.com

[6] Biometric template selection and update: a case study in fingerprints, the Journal of Pattern Recognition Society , Umut Uludag, Arun Ross & Anil Jain13th Nov 2003

[7] Authentication Technologies: http://biometrics.pbworks.com/w/page/14811351/Authentication%20technologies#FactorsOfAuthentication

[8] Iris Recognition: An Emerging Biometric Technology PROCEEDINGS OF THE IEEE, VOL. 85, NO. 9, SEPTEMBER 1997

[9] US Patent - Biometric Personal Identification System based on Iris Analysis, Patent No 5291560 March 1 1994

[10] Iris recognition on hadoop: A biometrics system implementation on cloud computing, Shelly, N.S. Raghava proceedings of IEEE CCIS2011

[11] Figures from Biometric Hadoop Summit 2010

[12] An Identity-Authentication System

[13] Using Fingerprints Anil K. Jain, Fellow, IEEE, Lin Hong, Sharath Pankanti, Associate Member, IEEE and Ruud Bolle, Fellow, IEEE Proceedings Of IEEE Vol. 85, No. 9, September 1997