ISSN: 0976-9102(ONLINE)
DOI: 10.21917/ijivp.2013.0080

AN APPROACH TO REDUCE THE STORAGE REQUIREMENT FOR BIOMETRIC DATA IN AADHAR PROJECT

T. Sivakumar[1], G. Chithira Rakshmi[2] and A. Ummu Salma[3]
Department of Information Technology, PSG College of Technology, India
E-mail: [1]sk@ity.psgtech.ac.in, [2]chithira.abi @gmail.com and [3]abd_salma@yahoo.com

Abstract
AADHAR is an Indian Government Project to provide unique identification to each Citizen of India. The objective of the project is to collect all the personal details and the biometric traits from each individual. Biometric traits such as iris, face and fingerprint are being collected for authentication. All the information will be stored in a centralized data repository. Considering about the storage requirement for the biometric data of the entire population of India, approximately 20,218 TB of storage space will be required. Since 10 fingerprint data are stored, fingerprint details will take most of the space. In this paper, the storage requirement for the biometric data in the AADHAR project is analyzed and a method is proposed to reduce the storage by cropping the original biometric image before storing. This method can reduce the storage space of the biometric data drastically. All the measurements given in this paper are approximate only.

Keywords:
Biometrics, Image Processing, Image Storage

1. INTRODUCTION

AADHAR is a project initiative by Government of India to provide unique identification number for each resident in India. The identification number is a 12 digit unique number issued by the Unique Identification Authority of India (UIDAI) on behalf of the Government of India. UIDAI tries to enable a universal identity infrastructure by replacing all other identity cards such as ration card, PAN card etc. AADHAR number will help to access various services like banking, mobile phone connections and other government and non government services [1]. The two major steps in the AADHAR authentication process are (a) enrollment and (b) verification.

Each individual should be first enrolled into the system using his personal and biometric details for verification. These enrollment data are stored in a centralized data repository called 'Central ID Repository' (CIDR). At the time of verification, AADHAR number along with other attributes, including biometrics are submitted online to the CIDR. The AADHAR authentication system responds with only a 'yes/no' and no personal identity information is returned as part of the response[2].

Following are the some major features of the AADHAR system [1].

1) It will only provide identity, not rights, benefits or entitlements
2) Envisions full enrollment of the residents, with a focus on poor and underprivileged communities to improve service delivery to the poor
3) Helps to provide proper verification of identity
4) Leverages the existing infrastructure of government and private agencies across India
5) Provides a flexible model for Registrars for their processes including issuing cards, pricing, collecting data of residents and in authentication
6) Ensures no duplication and provides online authentication
7) It will not share resident data and provides data transparency

In the conventional systems, to receive a service from any agency requires the resident to prove his/her identity by presenting some credentials. These credentials could be physical documents such as an identity card issued by agency, passbook issued by a bank or something similar and pin number, password etc.

AADHAR authentication enables agencies to verify identity of residents using an online and electronic means. It instantly verifies the identity by providing the AADHAR number and different identifiers collected by the agency based on the needs of the service.

The paper is organized as 4 major sections; section 2 explains the working of biometric solution model in the AADHAR project, section 3 contains the analysis of storage requirement for biometric data, section 4 gives the proposed method for reducing the storage requirement, section 5 shows the experimental results and section 6 contains the conclusion and the future work.

1.1 BIOMETRIC SOLUTION MODEL

Combination of fingerprint, face and iris biometric traits of the residents is used to provide authentication. ISO 19794 series of Biometric standards for fingerprint, face and iris are used. The Biometric Solution Provider (BSP) is an entity in the AADHAR architecture which will design, supply, install, configure commission, maintain and support biometric components of the UIDAI system [1].

The biometric verification module which is constructed using Software Development Kit (SDK) provides verification within the authentication server. The templates will be maintained in memory resident database by the UIDAI authentication server application. If the incoming requests contain a biometric image, the authentication server will use SDK to extract the feature and also to generate comparison score of the sample. Fig.1 shows the authentication server architecture used in AADHAR project [1].

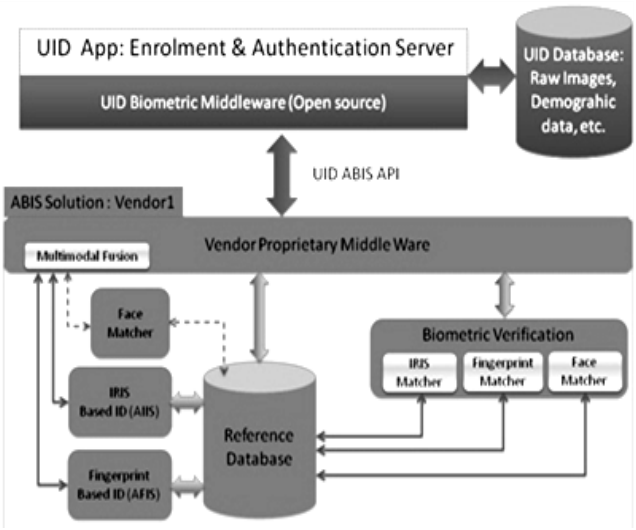


Fig.1. Architecture of Biometric Solution Provider in AADHAR

1.1.1 Fingerprint Image:

Ten good quality plain fingerprint image of each residents are collected and stored in CIDR to provide authentication using any finger and also to provide more accuracy. Uncompressed image storage is recommended by the UIDAI committee on Biometrics. In uncompressed mode, the total storage required for the entire population is 10,000 TB. For legacy reasons, lossless JPEG 2000 or WSQ compression is accepted. Storage format as per ISO section 8.3, minutiae format as per ISO 19794-2 and application dependent multi-finger fusion algorithms are used [3].

1.1.2 Face Image:

Full frontal 24-bit color image with minimum 90 pixels of inter-eye distance of at least 120 pixels optimum quality face image is required. For de-duplication and authentication of individuals who do not have fingerprints, three samples of face image may be used. In such cases, one should be frontal image and other two should be left and right side image. For verification, images with JPEG 2000 compression ratio are used. The image size after compression should not be less than 11 KB [3].

1.1.3 Iris Image:

The two eye images are captured simultaneously to assure correct assignment of left and right eyes and for more accurate estimation of roll angle. In order to obtain good quality template, the iris diameter should be minimum 140 native pixels. The iris images should be stored in ISO standard format using either JPEG 2000 or PNG lossless compression. It is expected that each enrollee will require 150 KB of storage space, thus requiring total storage space of 200 TB for the entire population [3].

2. STORAGE REQUIREMENTS ANALYSIS

According to the Census conducted in the year 2011, the population in India is 1,210,193,422. According to UIDAI Biometrics Committee Report, the storage requirement for fingerprint image data will take 7.5 MB per each individual, and for uncompressed minutiae details will take 10,000 TB for the entire population, for Iris image before compression 150 KB per each individual, for Iris image after compression 2-10 KB per each individual [3].

2.1 FINGERPRINT IMAGE

Since ten fingerprint images and uncompressed fingerprint minutiae details of each individual are stored in the database, approximately 17.5 MB storage is required for each person. For the entire population, approximately 20000 TB storage is required. Table.1 shows the requirement of storage space for fingerprint data.

Table.1. Storage Requirements of Fingerprint Data

Type of information	Storage size per subject	Storage size for entire population
Fingerprint image	7.5 MB	8000-10000 TB
Uncompressed fingerprint minutiae details	10 MB	10000 TB

2.2 FACE IMAGE

Three samples of face image, one frontal image, one left side and one right side image of each individual requires 4 KB storage space for uncompressed images and 2 KB for compressed images. For the entire population, approximately 8 TB of storage is required to store the face biometric data. Table.2 shows the storage requirement for face data.

Table.2. Storage Requirements of Face Data

Type of information	Storage size per subject	Storage size for entire population
Face image	4 KB	5 TB
Face image after compression	2 KB	3 TB

2.3 IRIS IMAGE

Two iris images of each individual require 150 KB memory per subject before compression and 2-10 KB after compression. Approximately a maximum of 210 TB of memory is needed to store the iris data. Table.3 shows the storage requirement for iris data.

Table.3. Storage Requirements of Iris Data

Type of information	Storage size per subject	Storage size for entire population
Iris image	150 KB	170-200 TB
Iris image after compression	2-10 KB	10 TB

3. PROPOSED RECOMMENDATION

The analysis shows that the biometric data requires an approximate storage depository of 20,218 TB for the entire population to store the three biometric traits. If the enrolled images are cropped to the exact region of interest that is needed for the verification process, the storage requirement can be reduced. For a fingerprint image, portions around the core point will be sufficient for the verification process. So, it is not necessary to store the whole fingerprint image in CIDR. The overall working model of the proposed model is shown in Fig.2.

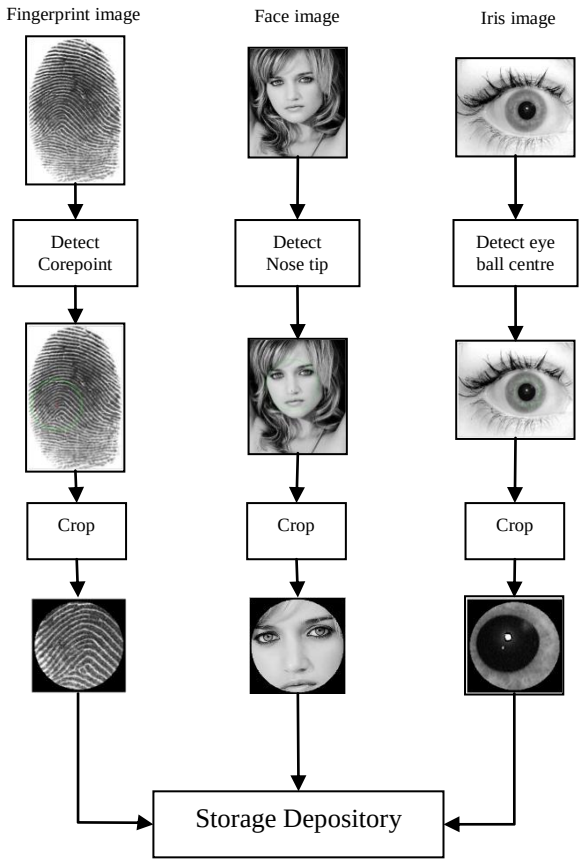


Fig.2. Proposed System Model

ALGORITHM 1

Algorithm for Fingerprint Processing

Step 1: Read the input fingerprint image of dimension 388 × 374

Step 2: Apply Normalization process for image enhancement

Step 3: Find the orientation of ridges in the fingerprint

Step 4: Smoothen the orientation field of ridges and calculate the sine component

Step 5: Find the maximum curvature in concave ridges and assign it as the core point

Step 6: Crop the original image in the dimension of 250 × 250 pixels or in the radius of 5

3.1 FINGERPRINT

According to the storage study, it clearly says that storage of fingerprint image and fingerprint minutiae details takes more size than the others. So in this paper, we are suggesting a fingerprint matching mechanism which will match the preprocessed cropped fingerprint images. At a global level the fingerprint pattern exhibits the area that ridge lines assume distinctive shapes [6]. The minutiae points around the core point of the fingerprint image will be sufficient for identification and matching.

In this method, first the core point of the fingerprint image is detected and is cropped in a fixed dimension. For the core point detection, first the preprocessing techniques would be applied and the singularity point is located. Image enhancement will try to improve the quality of the fingerprint image to increase the success rate of the system and noise will be removed.

Normalization is the first process in preprocessing. The normalized image is represented as in Eq.1 [7].

$$N(i,j)=\begin{cases} M_0+\sqrt{\dfrac{V_0\left(I(i,j)-M_i\right)^2}{V_i}} & \text{if } I(i,j)>M \\ M_0-\sqrt{\dfrac{V_0\left(I(i,j)-M_i\right)^2}{V_i}} & \text{Otherwise} \end{cases} \quad (1)$$

where, $I(i, j)$ denotes the gray level value in the original image, M_i and V_i denotes mean and variant of the original image and $N(i, j)$ denotes the normalized gray value of the image.

Next step is to find the direction or orientation of the ridges (Eq.2, Eq.3 and Eq.4) in the fingerprint image by dividing the image into $w \times w$ non-overlapping windows which is defined as [7],

$$\theta(i,j)=\frac{1}{2}\tan^{-1}\left(\frac{V_y(i,j)}{V_x(i,j)}\right) \quad (2)$$

$$V_x(i,j)=\sum_{u=i-w/2}^{i+w/2}\sum_{v=j-\frac{w}{2}}^{j+\frac{w}{2}}2\partial_x(u,v)\partial_y(u,v) \quad (3)$$

$$V_y(i,j)=\sum_{u=i-w/2}^{i+w/2}\sum_{v=j-\frac{w}{2}}^{j+\frac{w}{2}}\partial_x{}^2(u,v)\partial_y{}^2(u,v) \quad (4)$$

where, $\theta(i, j)$ represents the ridge orientation of the block and V_x and V_y represents the gradients of the centre pixel in the block.

For the core point detection, the orientation field is smoothened (Eq.5) and the sine component (Eq.6) is calculated using the following equations,

$$\theta'(i,j)=\frac{1}{2}\tan^{-1}\left(\frac{\varnothing'_y(i,j)}{\varnothing'_x(i,j)}\right) \quad (5)$$

$$\varepsilon(i,j)=\sin(\theta'(i,j)) \quad (6)$$

A label image B is initialized and each pixel is assigned the value of the difference in integrated pixel intensity of the regions R_1 and R_2 (Eq.7) which are determined empirically and are designed to capture the maximum curvature in concave ridges [6]. The maximum value in A and its coordinates are assigned as the core point.

$$A(i,j) = \sum_{R_1} \varepsilon(i,j) - \sum_{R_2} \varepsilon(i,j) \quad (7)$$

After detecting the core point, the original image will be cropped in a fixed dimension of 250 × 250 to obtain the surrounding regions of the core point. Fig.3 represents the block diagram which shows the sequence of operations to be performed to process the fingerprint image.

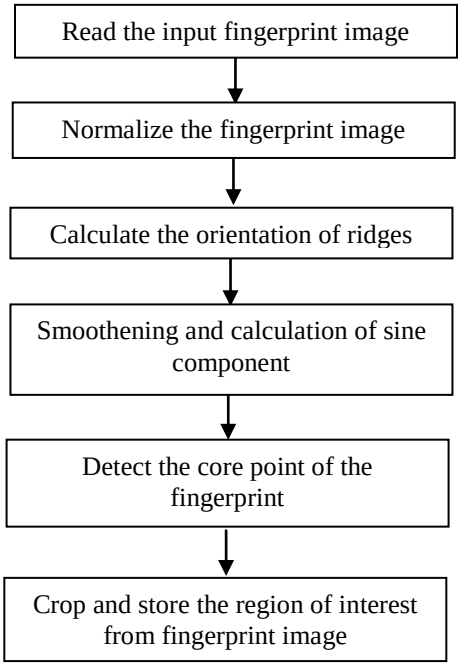


Fig.3. Flowchart for fingerprint processing

These cropped images are used for the matching purpose using minutiae matching technique. It gives almost same percentage of matching rate as for the original images.

3.2 FACE IMAGE

AADHAR Biometrics Committee strongly recommends storing the uncompressed image to preserve the quality of image [3]. The three sample face images are stored for authentication in AADHAR, frontal image and left and right side images. Uncompressed image will take more storage space than the compressed image but will degrade the performance due to the lack of quality. Storage space can be reduced by storing only that part of the face image where more facial features for matching are available. Fig.4 shows the sequence of operations to be performed for processing the face image.

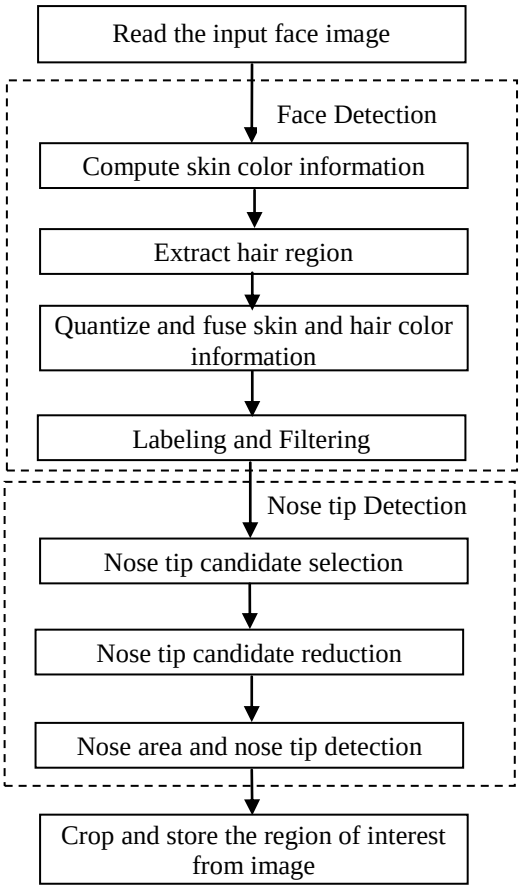


Fig.4. Flowchart for face image processing

ALGORITHM 2

Algorithm for Face Image Processing

Step 1: Read the input face image of dimension 428 × 498

Step 2: Compute the skin color information

Step 3: Extract hair region using the intensity element of the image

Step 4: Quantize and fuse skin color information and hair color information

Step 5: Apply labeling and filtering to remove noise components

Step 6: Select nose tip candidates using effective energy

Step 7: Reduction of nose tip candidates using mean & variance

Step 8: Nose area detection

Step 9: Detect nose tip from nose tip candidates which are within the nose area

Step 10: Crop the original image into 102 × 102 or in a radius of 5 using the nose tip as centre

Yao-Jiunn Chen et al [5] presented a simple face detection algorithm by fusing the quantized skin color information and hair color information. The skin color is distributed among the red and green colors in the defined range of $r = 0.33$ and $g = 0.33$ [6]. So the upper limit and the lower limit of the skin color is expressed in Eq.8 and the Eq.9 gives the computation of the skin color [6].

$$\begin{aligned} F_1(r) &= -1.376r^2 + 1.0743r + 0.2 \\ F_2(r) &= -0.776r^2 + 0.5601r + 0.18 \end{aligned} \quad (8)$$

$$Skin = \begin{cases} 1 & if\,(g < F_1(r) \cap g > F_2(r) \cap w > 0.001) \\ 0 & Otherwise \end{cases} \quad (9)$$

where, $w = (r\text{-}0.33)^2+(g\text{-}0.33)^2 > 0.001$.

Using the intensity element of the image, hair region is extracted which is given by Eq.10,

$$Hair = \begin{cases} 1 & if\,\left(I < 80 \cap \left(B-G<15 \cup B-R<15\right)\right) \\ & \quad \cup \left(20 < H \le 40\right) \\ 0 & Otherwise \end{cases} \quad (10)$$

where, $I = \frac{1}{3}(R+G+B)$

The quantized skin color information and hair color information are fused together and is labeled using the 8-connectivity method and a filter is applied to remove the noise components.

After detecting the face, the nose tip is detected to make it as a centre point for cropping the image. A method proposed by Wei Jen Chew et al. [8] for nose tip detection on a three dimensional face range image can be used. The nose tip candidates are detected by calculating the effective energy of each neighboring pixels P_i of each pixel P using the Eq.11.

$$Effective\ Energy = ||P_i - P||\ cos\ \theta \quad (11)$$

A pixel with a negative value for the effective energy of all the neighboring pixels (d_i) is selected as a potential nose tip candidate. The mean (μ) and variance (σ^2) values of the neighboring effective energy are used to reduce the unnecessary nose tip candidates using the Eq.12 and Eq.13,

$$\mu = \frac{1}{n}\sum_{i=1}^{n} d_i \quad (12)$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(d_i - \mu)^2 \quad (13)$$

The nose area is detected by forming a triangle after detecting the two eye blobs and mouth. The three pixels which have the densest amount of nose tip candidate neighbors and within the nose area are assumed as the nose tip. After detecting the nose tip, the face image is cropped in circular or rectangular shape using nose tip as the centre. The original facial images used for experiment was of size 82.9 KB and is reduced to 4.48 KB using rectangular cropping and 8.31 KB using circular cropping.

3.3 IRIS IMAGE

Instead of storing the image of whole eye region, it is sufficient to store only the iris part of the eye. For that, first the eyeball region within the eye will be detected and is cropped to only that information. Fig.5 shows the sequence of operations to be performed to process the iris image.

ALGORITHM 3

Algorithm for Iris Image Processing

Step 1: Read the input Iris image

Step 2: Make invert of the input image

Step 3: Convert it to gray scale

Step 4: Apply binary filter with threshold value 220

Step 5: Find the biggest object in the binary filtered image

Step 6: Find the centre point and height of the object

Step 7: Crop the image in circular or rectangular shape from that point taking radius as 2.5 multiply of height

The Eq.14 is applied to all the pixels in the input eye image to obtain the invert of the image and is converted to gray scale. A binary filter with threshold value 220 is applied to get the corresponding binary image and is labeled to detect the biggest object in the region and the centre point and height is calculated.

$$Pi = 255 - Pi \quad (14)$$

The method is experimented using the original eye image of size 106 KB and the resultant image after rectangular cropping have the size of 5.57 KB and circular cropping have the size of 9.55 KB.

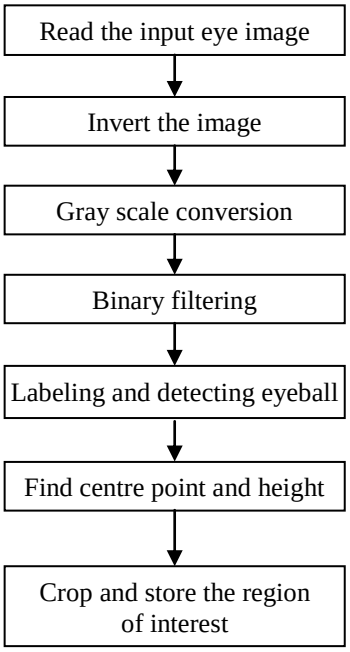


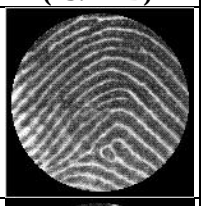
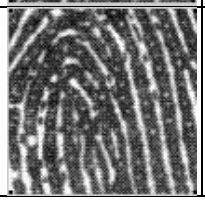
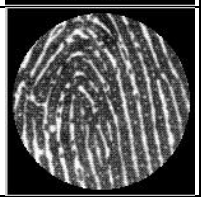
Fig.5. Flowchart for iris image processing

4. EXPERIMENTAL RESULTS

4.1 FINGERPRINT IMAGE

This method was verified using a database of which contains fingerprint images of dimension 388 × 374 acquired through the optical sensor. Each original image in the database of 142 KB size is reduced to 14.1 KB which is 10 times smaller than the original image by applying the above method. Table.4 shows the result of the cropping process, and Table.5 gives the storage analysis of cropped fingerprint image.

Table.4. Result for Fingerprint image cropping

Original Image (142 KB Size)	Rectangular cropped Image (14.1 KB)	Circular cropped image (15.1 KB)
		
		
		

Since the original image size of 142 KB is reduced to 14.1 KB for each fingerprint image, the storage size of image 768 KB which is used in the AADHAR project can also be approximately reduced to 10 times, 76.8 KB. Then the storage requirement for the entire population will be approximately a maximum of 90 TB.

Table.5. Storage Analysis of cropped fingerprint image

Type of Information	Storage Size Per Subject Approximately	Storage Size for Entire Population Approximately
Fingerprint Image with original size	7.5 MB	8000-10000 TB
Rectangular cropped fingerprint image	80 KB	800-1000 TB
Circular cropped fingerprint image	85 KB	880-1100 TB

4.2 FACE IMAGE

The original face image of size 82.9 KB was reduced to 4.48 KB in rectangle cropping process and to 8.29 KB in circular cropping process. Approximately 20% storage size of the original image was reduced using the rectangular cropping method and 10% using the circular cropping method.

Table.6. Result for Fingerprint image cropping

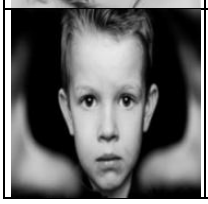
Original Image (82.9 KB)	Rectangle cropped image (4.48 KB)	Circular cropped image (8.29 KB)
		
		
		

Table.6 shows the experimental results of the face image cropping and Table.7 shows the storage analysis for the cropped face image for the entire population. The storage size of face image 5 TB can be reduced to 0.25 TB using the rectangular cropping process approximately.

Table.7. Storage analysis of cropped face image

Type of Information	Storage Size Per Subject Approximately	Storage Size for Entire Population Approximately
Original face image	4 MB	5 TB
Face image after rectangular cropping	0.2 KB	0.25 TB
Face image after circular cropping	0.4 KB	0.5 TB

4.3 IRIS IMAGE

The eye image of size 106 KB was reduced to 5.57 KB using rectangular cropping process and 9.55 KB using circular cropping processing. Table.8 shows the results for iris image cropping.

Table.9 shows the storage analysis of the cropped iris image for the entire population. The maximum storage size of 200 TB can be reduced to a maximum of 18 TB approximately. For real time implementation region other than pupil can be extracted and stored to improve the performance.

Table.8. Result for iris image cropping

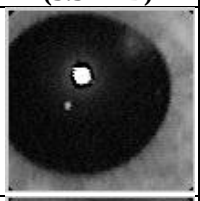
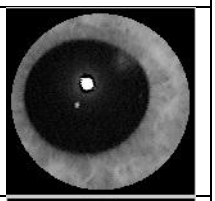
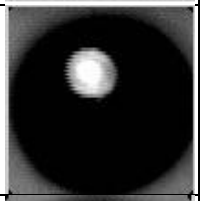
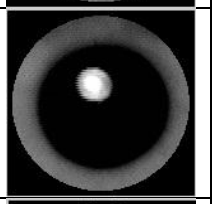
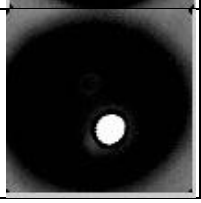
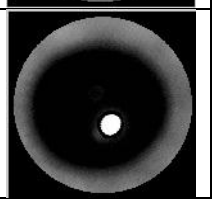
Original Image (106 KB)	Rectangle cropped image (5.57 KB)	Circular cropped image (9.55 KB)
		
		
		

Table.9. Storage analysis of cropped iris image

Type of Information	Storage Size Per Subject Approximately	Storage Size for Entire Population Approximately
Eye image	150 MB	170-200 TB
After rectangular cropping	7.87 KB	8.9-10.5 TB
After circular cropping	13.5 KB	15.3-18 TB

5. CONCLUSION AND FUTURE WORK

In this paper, the various biometric data collected has been studied. The method used to process and store these data in AADHAR has also been described. The storage required for biometric data is presented for both compressed and uncompressed versions. As per the analysis, fingerprint data consumes more storage than the other biometric details. The fingerprint image size per subject in the 'AADHAR' project is 7.5 MB, i.e. approximately 750 KB per image. Using the proposed method, the size of the original image can be decreased to six times. For iris and face images also the storage requirement can be reduced significantly by using the proposed method. Since all the storage analysis is imprecise, for real time implementation the storage requirement may change. Our future work is to analyzing the effectiveness of the matching process between processing the entire biometric data versus processing the partial (rectangle or circle) biometric data. Using such system, the verification phase will consist of matching the compressed versions of these data. This results in less storage and comparison time.

REFERENCES

[1] Atipat Julasayvake and Somsak Choomchuay, "An Algorithm for Fingerprint Core Point Detection", *9th International Symposium on Signal Processing and its Applications*, pp. 1-4, 2007.

[2] Yao-Jiunn Chen and Yen-Chun Lin, "Simple Face-detection Algorithm Based on Minimum Facial Features", *The 33rd Annual Conference of the IEEE Industrial Electronics Society*, pp. 455-460, 2007.

[3] T.K. Leung, M.C. Burl and P. Perona, "Finding Face in Cluttered Scenes using Random Labeled Graph Matching", *Proceedings of 5th IEEE International Conference on Computer Vision*, pp. 637-644, 1995.

[4] Wei Jen Chew, KahPhooiSeng and Li-MinnAng, "Nose Tip Detection on a Three-Dimensional Face Range Image Invariant to Head Pose", *Proceedings of the International Multi-Conference of Engineers and Computer Scientists*, Vol. 1, 2009.

[5] Unique Identification Authority of India, "*AADHAR Authentication API Specification version 1.0*", pp. 5, 2010.

[6] UIDAI Committee on Biometrics, "*Biometrics Design Standards for UID Applications version 1.0*", pp. 32-43, 2009.

[7] Raymond Thai, "Fingerprint Image Enhancement and Minutiae Extraction", Report, 2003.

[8] Unique Identification Authority of India , http://uidai.gov.in