

# Combining Iris, Sclera and Pupil Features for Biometric Authentication System on Smartphone Devices

Dr. Mrunal Pathak<sup>1</sup>, Shivani Pokharkar<sup>2</sup>, Madhura Anilrao Belsare<sup>3</sup>, Shweta Manoj Hol<sup>4</sup>

Faculty, Department of Information Technology Engineering<sup>1</sup>

Students, Department of Information Technology Engineering<sup>2,3,4</sup>

AISSMS Institute of Information Technology, Pune, Maharashtra, India

**Abstract:** Currently, biometric authentication systems are commonly used based on physical and behavioural biometric modalities like iris, face, fingerprints, ear, sclera, DNA, voice, signature, etc. Rather than relying on the standalone or unimodal biometric system, multimodal biometric systems are secure and provide more accurate results for person identification and verification. This paper introduces the multimodal eye biometric authentication system where iris, pupil and sclera features are extracted using CNN based on entropy values to perform the accurate automatic segmentation for smartphone devices. The eye images used in the proposed approach for training and testing are completely captured by smartphones. The fusion method used to fuse the colour and texture characteristics of iris and pupil with Y-shaped sclera characteristics from eye image based on support value is Feature Level Fusion. As the images are captured in normal environment settings, it is an unconstrained colour eye image database. MATLAB is used for the experimentation and testing of the model. The proposed eye biometric system outperforms in the case of segmentation and recognition accuracy. Recognition accuracy is % for unconstrained eye images achieved for the eye image database captured by smartphones.

**Keywords:** Entropy based Convolutional Neural Network, Multimodal Biometric Systems, Unimodal, Feature Level Fusion

## I. INTRODUCTION

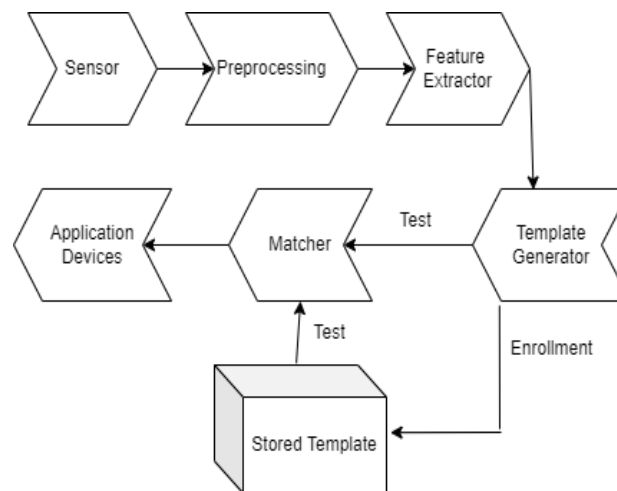
Biometric systems are continuously evolving. Different biometric traits like fingerprint, eyes, face, are employed for authentication. Nowadays, passwords and patterns have become insecure to use because of the threat of being stolen. But, biometric is something that is difficult to steal because it is owned by the person. In this regard, eye recognition has been utilized in many critical applications, such as access control in restricted areas, database access, national ID cards, and financial services, and is considered one of the most reliable and accurate biometric authentication systems. Several studies have demonstrated that the eye biometric traits like iris, pupil sclera has several advantages over other biometric traits (e.g., face, fingerprint, voice), which makes it heartily accepted for application in high reliability and accurate biometric authentication systems. Biometric systems are classified as unimodal and multimodal biometric systems. There are some limitation of unimodal biometric system which can be reduced by using multimodal biometric system. Multimodal biometric system utilizes information which it gathers from multiple modalities by using various processing techniques or both. Therefore, it is better to use multimodal rather than unimodal, it integrates more than one physiological or/and behavioral characteristics for identification. It improves performance as well as reliability.

The unimodal biometric system makes use of a single biometric trait for person identification. The accuracy of the unimodal biometric system is reduced because of the challenges such as noisy data, non-universality, interclass similarity, intraclass variation, spoofing attacks, etc. These challenges are overcome by the multimodal biometric system in which two or more different features of either the same biometric trait or features from different biometric traits are fused together to provide more accurate and reliable recognition results.

The proposed system puts forth the eye biometric authentication system based on the content-based image retrieval approach for the fusion of iris, sclera, and pupil features to improve the authentication for non-ideal color eye images

captured by smartphone cameras. Entropy values are estimated based on best quality features extracted from images such as color, brightness, and texture to reduce the computational cost and time. These entropy values are used to classify the iris, sclera, and pupil region using a convolutional neural network (CNN), so the CNN is called E-CNN. Using E-CNN produces better segmentation results and also deducts the segmentation error rate and time enormously. Multi-algorithmic feature extraction is applied to extract prominent features from the segmented images which are then combined together based on a fusion method called feature level fusion to calculate the support value. Authentication of the eye image is performed by comparing the support value match score against the template stored value in the database employing the Euclidean distance method.

This paper is divided into sections. First all the processes which are required for accurate feature identification are stated here and briefly explained. Before that, recent work done by reputed researchers and different techniques regarding biometric models and algorithms is studied in the related work section. At last, results, future work and conclusion sections would be there.



**Figure 1:** General Block Diagram for Biometric System

## II. RELATED WORK

We have studied several research papers which encouraged us for further study and research.

- Multimodal Eye Biometric System Based on Contour Based E-CNN and Multi Algorithmic Feature Extraction Using SVBF Matching: It is published in 2019 and talks about Color Histogram, Log Gabour filter and SVBF fusion (Support value based fusion). It has shown in the paper that how SVBF is better than other fusion techniques. Performance of recognition was improved by feature level fusion of these features based on SVBF matching. To achieve accuracy up to 93.33% and 97% for unconstrained images are taken from database MMU and UBIRIS.v2 respectively. Results for this paper is as follows:  
Segmentation Accuracy - 97.145% || Computation time - 0.9 seconds || FAR - 7.69% || FRR - 1.35% || GAR - 98.64
- Multi-Algorithmic Texture Feature Extraction by Fusing Iris and Sclera Features for Unconstrained Images: This paper is published in 2018. Researchers have used sclera and iris for feature extraction. They have used unconstrained images. Match score level fusion is used with GLCM and wavelet algorithms. They have taken values for the same and got Equal Error Rate as follows: EER: GLCM: 07% || GLCM + Wavelet: 1.21%.
- Multimodal Biometric System Iris and Fingerprint Recognition Based on Fusion Technique: Is published in 2018 has used fingerprint and iris in the stated model. They have used decision fusion techniques for fusion and KNN algorithms. The model proposed in this paper has shown 90% accurate results but it has not used features of pupil and sclera. Also, researchers have used particular ranged camera to capture images of iris which is against of our model.
- Effective segmentation of sclera, iris and pupil in eye images: Is published in 2019 taken iris, sclera and pupil

features for authentication. They have used MMU and UBIRIS database of images. Color Histogram, E-CNN and Log Gabour filter is used in their model. The paper excessively focuses on the segmentation of iris, sclera and pupil, instead of other factors and states that if segmentation is not done perfectly, it can result in poor authentication. We have referred this paper for calculation and results. Results are as follows:

**Table 1:** Results of MMU and UBIRIS image database







Parameters	MMU	UBIRIS
Accuracy:	98.275 %	99.423%
PPV	98.314%	99.926%
FNR	0.042%	0.504 %
FDR	1.685%	10.073%

- *Deep-Learning Based Joint Iris and Sclera Recognition with YOLO Network for Identity Identification:* Is published in 2021 which has proposed a model for two biometric traits that is iris and sclera. YOLOv2, R-CNN has been used and this model doesn't need Sclera & Iris segmentation. Deep-learning based design (YOLOv2) has given Accuracy= Up to 99% (mAP). Other CNN models and the low-complexity YOLO based models were not studied which can enhance real-time performance at the edge device.

### III. PROPOSED SYSTEM

Before starting towards processes and algorithms for the proposed model, let's take a look at how we collect images for our database.

**Table 2:** Image database created by us

Sr no.	Eye and Gaze	Eye Image
1.	Right Eye, Front Gaze	
2.	Right Eye, Left Gaze	
3.	Right Eye, Right Gaze	
4.	Left Eye, Front Gaze	
5.	Left Eye, Left Gaze	
6.	Left Eye, Right Gaze	

As shown in the table 2, we have collected images in constrained environment by taking images in indoor and outdoor conditions. For each user 30 indoor and 30 outdoor images are clicked by 3 smartphone cameras which are 12 to 13 Mps.



While capturing image, the distance between user and camera was approx. 10 cm. For further processing the image dimension is fixed upto 256x256 pixels.

Generally multimodal biometric system acts in two phases:

1. Enrollment phase
2. Authentication Phase

In the enrollment phase biometric traits of a user are stored and captured as a template image which will be required for further processing. In the next authentication phase the image captured is processed and we get the result by processing in this phase thus verification of the person takes place by comparing captured data with templates. All the steps are mentioned below through which image has to go.

### 3.1 Preprocessing

In order to work on the image, the image needs to be preprocessed. Preprocessing generally means enhancing the image in which important features of the image are extracted and unwanted details like noise and distortion is eliminated. The input image is hence intensified. In the proposed system two steps are involved in preprocessing:

#### A. Normalization

The eye image is adjusted in such a way that it fits our requirements. Formula for normalization is

$$N' = \frac{\tilde{Y} - \tilde{Y}_{\min}}{\tilde{Y}_{\max} - \tilde{Y}_{\min}} \tag{1}$$

In above equation,  $N'$  is the normalized image for input image  $\tilde{Y}$  where  $\tilde{Y}_{\max}$  and  $\tilde{Y}_{\min}$  are the maximum and minimum values in image  $\tilde{Y}$ .

After performing normalization, a bilateral filter is used for smoothing the image in low color discrepancy by preserving the edges with the help of equation 2.

#### B. Bilateral Filtering

Quality of the input image is enhanced. It is useful for removal of noise. Brightness and contrast enhancement is also done.

$$\hat{I}(x) = \frac{1}{C'} \sum_{y' \in N(x')} e^{-\frac{\|y'-x\|^2}{2\sigma_d^2}} e^{-\frac{|I(y')-I(x')|^2}{2\sigma_r^2}} I(y') \tag{2}$$

Where,  $\sigma_d$  and  $\sigma_r$  are parameters controlling the fall-off of weights in spatial and intensity domains, individually,  $N(x')$  is a spatial neighborhood of pixel  $I(x')$ , and  $C'$  is the normalization constant.

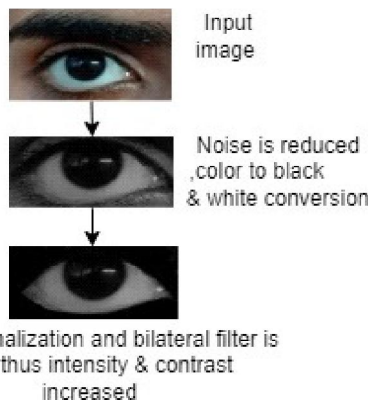


Figure 2: Image sample after preprocessing

In preprocessing input images of any size (dimensions) are taken. All the input images are resized to uniform dimensions using the `resize()`. Then the resulting image is converted to gray image (RGB to Gray) in order to reduce the computational complexity of the image. After that Bilateral Filtering is carried out which is followed by Min max Normalization. The output image of required size which is the gray image is obtained and can be used for further calculations.

#### IV. SEGMENTATION

Segmentation of an image generally means partitioning an image into meaningful subregions. In the proposed method we used segmentation for partitioning the eye into subregions like Iris, Sclera and Pupil. Image Segmentation should be precise in order to have good recognition results.

#### V. ENTROPY BASED CNN

Entropy means joint probability of occurrence of pixels in an image by removing redundant features. E-CNN segmentation is used to separate Iris, Sclera, Pupil regions. It is used to classify images based on textures in an image. Images are classified according to color, texture and brightness features.

Initially entropy values of the contour image are obtained and an entropy image is formed from it. This now becomes an input to CNN. The entropy image pixel values are further grouped together based on the similarity of characteristics like color, texture which are called superpixels. This results in improving the computational efficiency. Lastly, classified regions of Iris, Sclera and Pupil are obtained based on the superpixels.

Shannon's entropy is used to find structures and patterns in the data. Equation for entropy of  $i$ th superpixel is:

$$E_y^i = \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} P(i, j) (-\log_2(P(i, j))) \quad (1)$$

For N dimensional co-occurrence matrix,  $P(i, j)$  represent elements in matrix at coordinates  $(i, j)$

##### 5.1 Steps for E-CNN

The E-CNN Model consists of 3 convolutional layers, 2 max pooling layers, 1 Fully connected layer and softmax function. Convolutional Layers of 5X5 filter size are used to obtain local features. Max pooling has a filter size 2X2 and is used to perform local max operation to reduce the parameters and get local invariant features. The output of the last pooling layer is taken as input in a fully connected layer and a certain vector value probability is obtained to guess the trait to be iris, sclera or pupil after applying softmax function.

$$P_i = \frac{e^{x_i}}{\sum_1^k e^{x_i}}$$

Where  $x$  is the system input and output clusters of CNN are iris, sclera or pupil segments.

#### VI. CONTOUR BASED FEATURE EXTRACTION

Contour represents characteristics of visual pattern available in image which is derived based on the features like color, texture, and brightness. Therefore the set of contours represents the subset of features that help to perform the segmentation of the eye images accurately with fewer features which in turn reduces the computational cost. The texture is analysed using textons, which are small sets of prototype response vectors. The texture value of the local region around the pixel is estimated by comparing the texton distribution on either side of the pixel relative to its dominant orientation. The texture lies between 0 and 1. The objects in the image which are silent in colour can be black or white, this helps to describe the characteristics of brightness. Brightness can be measured by comparing the intensity value of a pixel with its neighbouring pixel. For a uniform image, the brightness value is zero. Colour features are one of the most widely used attributes to represent the characteristics of images. Colour features are estimated based on the occurrences of every colour index in an image with different intensities.

#### VII. ENTROPY BASED FEATURE EXTRACTION

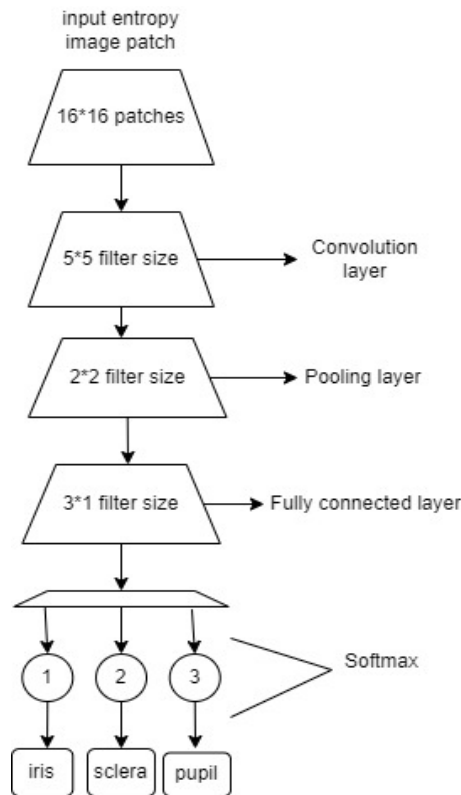
Entropy is nothing but a measurement of the degree which shows uncertainty that might exist in a system. To evaluate



structures and patterns in the data which can be used to characterize texture in the image is known as Shannon’s entropy. The entropy extraction for the effective segmentation should be done for iris, sclera and pupil regions. The proposed system calculates the entropy values to distinguish iris, sclera and pupil region and it is based on available texture in the contoured image.

VIII. IMAGE SEGMENTATION USING DEEP LEARNING

Deep learning based clustering approach is used to segment iris, sclera and pupil region .This is done by passing the entropy image through convolution neural network which trends to extract global features available from image. CNN is composed of multiple layers and is one of the variations of feed forward artificial neural networks. CNN has three main layers: convolution layers, pooling layers and fully connected layers.



Every layer has its own functions like in convolution layers, filters are applied on input images so that the system can generate the feature map. There are many sub-layers in each basic layer and every sub-layer can have multiple series of filters known as kernels. In the proposed system 5\*5 filter size is used at each convolution layer to extract the local features which is referred as weight sharing. After the completion of the convolution layer, a non-linear activation function Rectified Linear Unit (ReLU) to obtain the output of the convolution layer accelerates the convergence of CNN. Second layer is the pooling layer which is known as the down sampling layer. This layer, when applied on the output of the convolution layer, reduces the dimensions and number of parameters of CNN. Fully connected layers take output from the pooling layer as an input to flatten into a single vector of values, each representing a probability that a certain feature belongs to a either iris, sclera or pupil region which can be rectified after applying the softmax function.

8.1 Feature Extraction

For correct authentication in eye biometric system color, texture and shape these are the prominent and reliable features for iris, sclera and pupil. These features are used to generate the feature vector. The system which is proposed here will extract the color and texture information from segmented iris and pupil region. The segmented iris and pupil region is combined with the Y-shaped features extracted from sclera. To extract the color features color histogram algorithm is used



and to extract texture features log Gabor filter is used. The unique pattern of sclera for every person forms the stable blood vessel pattern.

8.2 Fusion

The fusion method used for the fusion of features extracted from iris, sclera, and pupil biometric traits is feature level fusion. These features are combined to calculate a joint feature vector known as the support value. Following is the formula used for calculation

S\_tilde\_value = (G'(f) + R' + D'\_LCH(I')) / (G'(f) \* R' \* D'\_LCH(I')) (2)

Where, G'(f) is the log Gabor feature, R' is the sclera

Y-shaped feature, D'\_LCH(I') is the color histogram feature. The score measure dependent on the support value, minimum and the maximum value of all features is portrayed as,

S' = S\_tilde\_value + W\_max + W\_min (3)

W\_max = max{all features} (4)

W\_min = min{all features} (5)

8.3 Matching

We use the Euclidean distance method to compare the support match score value of test data with estimated values stored in trained enrollment database. Euclidean distance is applied for pixel wise comparison of images. If the calculated distance is lesser than threshold value then data classified as recognized otherwise it is rejected. Next section is about results and scores that we got after performing code on matlab.

Table 3: Local Indoor Dataset Results

Table with 11 columns: Indoor, Accuracy, Sensitivity, Specificity, fmeasure, PPV, NPV, MCC, GAR, FAR, FFR. Rows include Left Front Test, Right Front Test, Left Left Test, Right Right Test, Left Right Test, Right Left Test.

From Table 3: we have observed that Accuracy of all the gazes is same and the system is trained accordingly so that if any gaze image is taken correct authentication is done.

IX. RESULT AND DISCUSSION

The proposed eye biometric system is implemented on the working platform MATLAB. Experimentations are on freely available research VISOB 2.0 Dataset ICIP2016 Challenge Version.

We used the 3200 images in total from the dataset which were acquired with the help of Samsung and Oppo devices capturing bursts of still images at 1080p resolution using pixel binning from 8 to 12 inches from subject faces. For unconstrained eye images we created our database which contains 900 images acquired at a distance of about 10 cm. It represents 2 different types of images, one captured indoor and other outdoor. The database was collected from a total of 15 subjects. Each user has 30 indoor and 30 outdoor images, which were further classified as front eye images, right gaze eye images and left gaze eye images. 3200 and 630 images from VISOB and ad-hoc databases are used respectively for training and testing purposes in ratio 80:10.

Performance of segmentation is analysed for proposed entropy feature-based deep learning segmentation method using CNN. From the contour image generated based on the colour, texture, and brightness features, we calculate the entropy value which minimises the feature set by avoiding the redundant and poor quality features. This helps to reduce the computational

complexity as well as the time required for segmentation.

We are able to achieve classification accuracy up to 95%, and sensitivity of around 98%. But we received a considerably high FAR(False Acceptance Rate), and FRR(False Rejection Rate). That is possibly because we used unconstrained settings for acquiring eye images without any help of expertise as an experiment to procure data. Another probable reason for it can be that the algorithms and the approach used in this experiment have been previously employed for images taken by automated or semi automated cameras in specific sessions, arrangements and with the help of expertise in procuring and managing databases and, not just with smartphone cameras .

### 9.1 Local Dataset Outdoor

**Table 4:** Local Outdoor Dataset Results

Indoor	Accuracy	Sensitivity	Specificity	fmeasure	PPV	NPV	MCC	GAR	FAR	FFR
Left Front Test	94.5	98.6014	84.21053	94	96	15.78947	1.398601	96.24573	86.33118	98.6014
Right Front Test	94.5	98.6014	84.21053	94	96	15.78947	1.398601	96.24573	86.33118	98.6014
Left Left Test	94.5	98.6014	84.21053	94	96	15.78947	1.398601	96.24573	86.33118	98.6014
Right Right Test	94.5	98.6014	84.21053	94	96	15.78947	1.398601	96.24573	86.33118	98.6014
Left Right Test	94.5	98.6014	84.21053	94	96	15.78947	1.398601	96.24573	86.33118	98.6014
Right Left Test	94.5	98.6014	84.21053	94	96	15.78947	1.398601	96.24573	86.33118	98.6014

Table 4 has same observation values as Table 3 and hence if input image is given for authentication in any light conditions i.e., dim or bright will result in correct authentication

### 9.2 Indoor + Outdoor Results

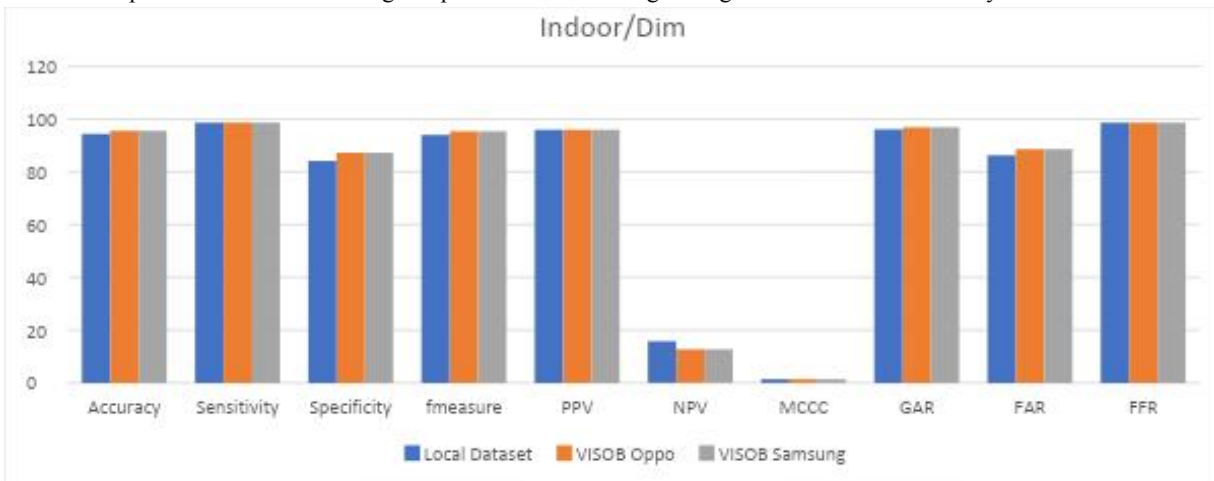
**Table 5:** Local Outdoor Dataset Results

Accuracy	Sensitivity	Specificity	fmeasure	PPV	NPV	MCC	GAR	FAR	FFR
94.5	98.6014	84.21053	94	96	15.78947	1.398601	96.24573	86.33118	98.6014

Table 5 observations suggest that there is no change in authentication if the systems indoor and outdoor features are combined and suggests system is trained properly and can give correct results.

### 9.3 Indoor Comparison Graph

The below Graph 1 shows the comparison of 3 datasets viz. Local dataset (Dataset prepared by us using VIVO Y71, VIVO Y17, OPPO mobile phone camera), VISOB Dataset (using OPPO camera), VISOB Dataset (Using Samsung camera) indoor images. The results of three datasets are about the same. As VISOB Datasets have more images in it and as they have used certain atmospheric conditions for image acquisitions the results give slight increase in the accuracy.

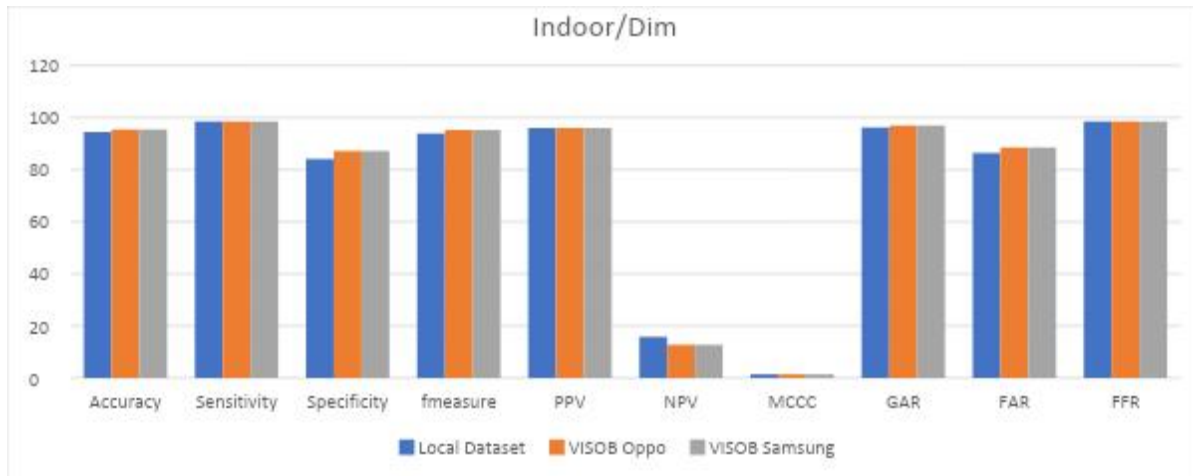


**Graph 1:** Indoor Comparison Graph





Outdoor Comparison Graph



Graph2: Outdoor Comparison Graph

Graph 2 shows comparison of outdoor images of three datasets viz. Local dataset (Dataset prepared by us using VIVO Y71, VIVO Y17, OPPO mobile phone camera), VISOB Dataset (using OPPO camera), VISOB Dataset (Using Samsung camera). The results observed from Graph 1 and Graph 2 are the same because all the three datasets give correct authentication in any light conditions i.e., dim or bright.

X. CONCLUSION

Present unimodal biometric systems are not 100% reliable, they also suffer from spoofing attacks due to lack of invariant representation and noisy input. Therefore multiple biometric traits are fused together to achieve better performance of biometric authentication. Multimodal biometric system helps to overcome these issues. We went a step further, and created a multi biometric system for eye images taken in an unconstrained environment, without any use of iris recognition platforms available like LG IrisAccess and under no expertise. Proposed eye biometric system combines the prominent features of iris, sclera and pupil for the images acquired by smartphone cameras in an unconstrained environment. Entropy based feature selection proves that it reduces time required for segmentation because of selection of optimal set of feature selection. Classification accuracy increases with increase in entropy values for input image. But we did face high FAR, and FRR because no special arrangement was done for image procurement, and algorithms need to be tuned specifically to work for images captured by smartphone cameras. In future work, we will attempt to lower down the FAR and FRR by adapting the algorithms according to the needs of eye images captured by smartphones.

REFERENCES

- [1]. Multimodal Biometric System Iris and Fingerprint Recognition Based on Fusion Technique Abdullah Ahmed, Ahmed Shamil Mustafa International Journal of Advanced Science and Technology
- [2]. Multimodal Biometric Authentication for VR/AR using EEG and Eye Tracking Vrishab Krishna Research Mentorship Program Santa Barbara, California, USA
- [3]. Multi-Algorithmic Texture Feature Extraction by Fusing Iris and Sclera Features for Unconstrained Images Mrunal Pathak\*, Department of CSE, KL University, Guntur, A.P., India. E-mail: mrunalpathak@gmail.com Dr. Vinayak Bairagi, Department of E&TC AISSMS, IOIT, S.P. Pune University, Pune, Maharashtra, India.
- [4]. Support Value Based Fusion Matching using Iris, Sclera Features for Person Authentication in Unconstrained Environment.
- [5]. Feature Level Fusion of Iris and Sclera using Entropy Based CNN Features to Improve the Performance of Biometric Authentication.
- [6]. Robust Multimodal Biometric Authentication Integration Iris, Face and Palmprint Fenghua Wang, Jiuqiang Han School of Electronics and Information Engineering, Xi'an Jiaotong University 710049 Xi'an, China.
- [7]. A.K.Jain, A.Ross S. Prabhakar: An Introduction to biometric recognition, IEEE Trans., Circuit systems and Video

- Technol., 14(1)(2004),pp 4-20
- [8]. Sharma ,R. Pavlovic, V. I. , Huang, T.S.: Towards multimodal Human computer interface. In Proceeding IEEE, 86(5),pp.853-860(1998).
  - [9]. Ross , A. Jain : Information fusion in biometrics, Journal of pattern recognition letters, vol 24, no.13,pp. 2115-2125,Sep 2003.
  - [10]. Pradeep Atrey, Anwar Houssain,Abdulmotaleb Saddik, Mohan kankanhalli , Multimodal fusion for multimedia analysis: A survey, Springer trans. multimedia systems (2010), 16: 345-379 [11] A. A. RossRGovindarajan: Feature level fusion using hand and face biometrics, Proc SPIE, Vol. 5779.,pp .
  - [11]. L. Hong, A.Jain : Integrating face and fingerprints for personal identification, IEEE Trans., Pattern Anal. Match. Intell, vol. 20, no. 1, pp. 1295-1307, Dec 1998 .
  - [12]. Giorgio Giacinto , Fabio Roli ,Et Dynamic Classifier Selection based on Multiple Classifier Behaviour. Pattern recognition. 34(1),179-181(2001).
  - [13]. Giorgio Giacinto , Fabio Roli ,Et Dynamic Classifier Selection based on Multiple Classifier Behaviour. Pattern recognition. 34(1),179-181(2001)
  - [14]. Aguilar, J.F., Garcia, J.O., Romero, D.G., Rodriguez, J.G.: A comparative evaluation of fusion strategies for multimodal biometric verification. In International conference on video based biometric person authentication, pp., 830-837,Guildford(2003)
  - [15]. Y. Zheng and A. Elmaghraby, "A Brief Survey on Multispectral IFace Recognition and Multimodal Score Fusion," in Proc. of IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Bilbao, pp. 543-550, 14-17 Dec 2011.
  - [16]. D. Kisku, A. Rattani, P. Gupta and J. Sing, Biometric Sensor Image Fusion for Identity Verification: A Case Study with Wavelet-Based Fusion Rules Graph Matching, in Proc. of IEEE Conference on Technologies for Homeland Security, HST '09, Boston, pp. 433-439, 11 -12 May 2009.