
LONG-SHORT TERM CONVOLUTIONAL NEURAL NETWORKS FOR IRIS
RECOGNITION UNDER NON-IDEAL CONDITIONS

Author's Name: Akhila P

Affiliation: Kerala, India

Corresponding Author Name & Email: Akhila P, akhilap255@gmail.com

ABSTRACT

Iris recognition system provides automatic identification of an individual based on unique patterns of the iris in the human eye. These features include the ridges, arches, folds, crypts, freckles, corona, furrows, *etc.* Existing iris recognition systems are heavily dependent on high user cooperation. A major drawback of this stop-and-stare condition is that they lead to low throughput. Also, the user is often required to look at the camera which may not facilitate recognition under covert conditions. In environments where user cooperation is not guaranteed, prevailing segmentation and matching schemes of the iris region are confronted with many problems, such as obstruction by eyelids, eyelashes, specular, non-regular lighting reflections in the eye area, poorly focused images, partial, out-of-iris images, invalid off-angle rotations and motion blurred irises. Currently, there are some works on iris recognition/classification based on convolutional neural network (CNN). They use one of the pre-trained models to extract the features and then support vector machine (SVM) is used for the classification. SVM is a non-parametric model where the complexity grows as the number of training samples increases. To reduce the complexity and improve the recognition accuracy, we propose a novel mechanism for iris classification using long short term memory (LSTM) sequence prediction model. LSTM can extract the long-term dependencies of the data features in the sequence. It is very similar to the recognition by the human visual system. We tested the proposed method on a UBIRIS v2 database which includes iris images under various non-ideal conditions and compared its performance with existing SVM based approach. The proposed method gave an accuracy of 95.25% and outperformed the existing SVM based method.

Keywords: CNN · Super Resolution · Semantic Segmentation · Feature Extraction LSTM.

INTRODUCTION:

A biometric system provides automatic identification of an individual based on unique physiological or behavioral features or characteristics possessed by an individual. Iris recognition is such a technology for identifying humans by capturing and analyzing the unique patterns of the iris in the human eye. Unlike other biometrics such as fingerprints or face, the distinctive aspect of iris comes from randomly distributed features in the iris. Iris recognition is regarded as the most reliable and accurate biometric identification system available at present.

The unique patterns of the iris start to form during the third month of fetal gestation and the structures of the iris are completely formed during the first 2 years of childhood development. The primary purpose of the iris is to control the diameter and the size of the pupil, which help to receive the light rays into the eye.

The iris is composed of two layers, Stroma and Sphincter Muscles. The Sphincter Muscles are responsible for the contraction of the pupil, and another group of muscles known as the Dilator Muscles governs the expansion of the pupil. When we look at our iris in the mirror, we can notice a radiating pattern. This is known specifically as the Trabecular Meshwork. When a Near-Infrared Light (NIR) is flashed onto the iris, many unique features can be observed. These features include the ridges, folds, freckles, furrows, arches, crypts, corona, as well as other patterns that appear in various, discernable fashions. The iris has been deemed to be one of the most stable and unique structures of the human physiology, and in fact, scientific studies have shown that even identical twins have a different iris structure.

Recent advances in science and technology have made it possible to identify individuals through their biometrics. Currently, there are various applications where biometric identification is used. First, it can be used to control access to restricted areas. Second, it can be used in passenger control at airports, as well as in border control. Third, it is useful in database access and financial services, where it has provided simplicity to customers and, at the same time, improved their security. Banking services and payments can also be made via biometrics.

Existing iris recognition systems are heavily dependent on specific conditions such as the distance of image acquisition and the stop-and-stare environment, which require significant user cooperation. Sometime recognition becomes difficult to recognize a user directly as noise may be present in iris images. Iris is a reflective mirror and is located behind the cornea. The iris images or templates are disturbed by most common noise factors that result of non-cooperative image capturing processes. There are nine causes that are considered for noise: the iris obstruction by eyelids, eyelashes, specular, lighting reflections, poor focused images, partial, out-of iris images, off-angle iris and motion blurred irises. During feature extraction phase, the uniqueness and discriminative level of the characteristics will determine the reliability of the recognition system. Therefore, unnecessary information must be discarded. A quantifiable set of features may be assigned to each of iris pattern obtained in this step, which will allow the computation of similarity measure between two iris patterns. The matching performance of a recognition system may also be affected by intra-class and interclass variation and improper user interaction. The problem found in the above mentioned noises are shown in the Fig. 1. However, iris recognition systems, which depend on random texture information in the iris,

usually present high failure-to-acquire rate. Iris recognition under visible spectrum (VIS) is a challenging task because of the risk in segmentation process. In this work we propose a novel approach for improving the segmentation and iris recognition in non-cooperative environments and visible wavelength, which exploits a semantic context cues to improve the localization accuracy of iris images. To this end, we take an image as input and start the segmentation process from a noiseless image generated from applying the Very Deep Super Resolution (VDSR) method [2] to the input image. Following, later it correctly classifies the person whether the eye images of her/him is present in the dataset. Our main goal in this research is focusing to give an authentication of a person by detecting the features in their iris, even if he/she is too far from the camera, and iris on the move. *etc.*

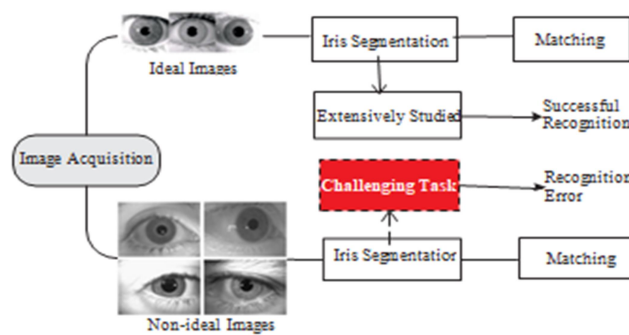


Fig. 1: Iris Recognition Challenges.

Specifically, we consider the following five kinds of steps involved in iris recognition. First, eye images are acquired under a visible wavelength camera; second, we perform a super resolution technique to enhance the quality of iris images for further process; third, we design a semantic context feature to describe the iris pixels and non-iris pixels; Fourth, the features are extracted using the AlexNet model; finally, we design and train a LSTM classifier to predict. Fig.2 illustrates the overview of our approach.

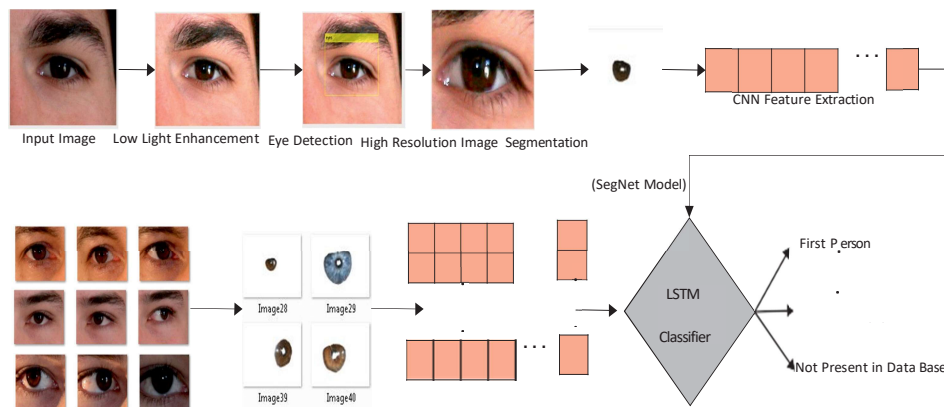


Fig. 2: Schematic Diagram of the Major Steps in Proposed Iris Recognition System.

We evaluate our method on the UBIRIS v2 dataset [3], one of the large-scale publicly

available dataset acquired using visible imaging in unconstrained imaging environments was utilized. We show that our method improves the recognition rate significantly and achieves the state-of-the-art performance. Our main contributions are summarized as follows: 1) Detect the exact iris region on eye images acquired under VIS; 2) We design and train LSTM network for prediction purpose; 3) We systematically evaluate our method on the UBIRIS v2 dataset and achieve the state-of-the-art recognition rate.

RELATED WORKS

The pioneering work in the iris biometrics is that of Daugman in 1994 [4] and early publications became a standard reference model. Integro-differential operators were used to detect the center and diameter of the iris. The image is converted from cartesian coordinates to polar coordinates and a rectangular representation of the region of the interest is generated. Feature extraction algorithm uses a 2D Gabor wavelets to generate the iris templates which were then matched using Hamming distance. However, this algorithm not succeeds where there is noise in the eye image, such as from reflections, since it works only on a local scale.

A Hough transform proposed by K.Popplewell et. al. [18] considers a set of edge points and finds the circle that best fits most of the edge points. In matching two irises, Daugmans approach involves computation of the normalized Hamming distance between the iris codes, whereas Wildes [5] proposed a Laplacian of Gaussian filter at multiple scales to produce a template. Matching is accomplished via an application of normalized correlation and Fishers linear discriminant as a similarity measure. Wildes briefly describes the results of two experimental evaluations of the approach, involving images from several hundreds of irises. The limitation of this work is that this mechanism does not perform well in the occurrence of bad lighting, occlusion by eyelids, noises or inappropriate eye positioning.

M. Vatsa et. al. [6] proposed a novel iris verification algorithm which uses textural and topological features of the iris image. A 1D log Gabor wavelet is used to extract the textural information and Euler numbers are used to extract the topological information from the iris image. They used Hamming distance algorithm and proposed difference matching algorithm to match the textural and topological information. Based on this algorithm, matching strategy is presented to reduce the false rejection while unaffected false acceptance rates. Li. Ma [7] used vertical, horizontal projections and Hough transform for localization followed by dyadic wavelet for feature vector generation.

These existing iris recognition systems are heavily dependent on specific conditions such as the image acquisition by using macro lens NIR iris camera and the stop-and-stare environment, which require significant user cooperation. Sometimes recognition becomes difficult to recognize a user directly because of the noises present in iris images.

Several works have proposed to use of the Convolutional Neural Network (CNN) for iris recognition. Minaee et al [8] proposed an iris recognition system where they used a pertained VGG-Net to extract the deep features. Then they used a multi-class SVM algorithm for classification. They tested their system on IIT iris dataset [9], and CASIA 1000 Iris dataset [10]. They obtained a recognition accuracy of 99.4%.

Liu et al. [11] proposed a DeepIris network of 9 layers consisting of one pairwise filter layer, one convolutional layer, two pooling layers, two normalization layers, two local layers

and one fully connected layer. The average segmentation errors obtained by MFCNs are 25.62% and 13.24% on UBIRIS.v2 [3] and CASIA.v4-distance [10] datasets, respectively.

G Alaslani et.al [17] proposed an iris recognition system where the features are extracted from a pretrained CNN Alex-Net model, and for the classification task, a multi-class SVM is used. They investigated the performance of their system in two situations: 1) extracting features directly from the segmented iris image and 2) extracting features from the normalized iris pattern. They tested their system on four public datasets namely IITD iris databases [17], CASIA-Iris-V1 [10], CASIA-Iris-thousand [10] and CASIA-Iris- V3 Interval [10] under different non-ideal conditions. Their system achieved excellent results with high accuracy rate of 100%, 98%, 98% and 89% respectively.

However, these existing paradigm is based on feed forward neural network architecture, with fixed network depth and receptive field size. Such a paradigm has several limitations: Fixed receptive field limits its recognition capacity when the input iris image is very large; it is quite different from the human visual system for recognition, which involves both feed forward and recurrent pre-processing. Recent researchers in the field are faced with the challenges of iris recognition in less constrained imaging conditions. These include processing and encoding of a non-ideally captured iris images. Non-ideal conditions are defined as off-angle, occluded, blurred and noisy images, and iris captured at a distance *etc.* Under these non-ideal situation, traditional iris recognition systems would not work well.

Nowadays most of the researchers use one of the pre-trained network for feature extraction and SVM for the matching. To the best of our knowledge, there are no CNN- LSTM approaches specifically designed for iris recognition. An important property and advantage of proposed CNN-LSTM is the efficiency in fast and easy prediction of person-authentication. The prediction is done by matching the feature values which is specially stored in a memory cell of LSTM. The proposed paradigm has the ability to significantly increase the range of the receptive field of the neural network. Aiming to reduce the computational complexity, incorporating the human visual perception system and achieve higher recognition accuracy, we propose to use CNN for feature extraction and an LSTMs to support sequence prediction.

1 CONVOLUTIONAL-LSTM IRIS RECOGNITION

We take as our system input a poor quality iris image and aim to check the authenticity of person. Our approach consists of four steps, as illustrated in Fig. 1. We first generate a set of high resolution (HR) images from the low resolution (LR) images, we then segment the iris region from the HR eye image, then encode the iris texture into some useful readable bits. Finally, verifying whether the iris patterns derived in the previous step matches with the one stored in the database *i.e.* intends to check the authenticity of person. We now introduce the details of each step of our pipeline, focusing on the automatic segmentation of iris region and classification process.

1.1 Image Acquisition and Image Pre-processing

In the image acquisition stage, the iris images are captured under visible wavelength. For improving the recognition accuracy we perform the following pre-processing steps on the acquired iris images.

1. Enhancement of the poor light images using Low-Light Illumination Map Estimation method.
2. Locating the eye portion using the Viola-Jones object detection mechanism.
3. Improving the resolution for image using VDSR method. The pre-processing steps are given in Algorithm 1.

Algorithm 1: Pre-processing of Iris Images

Input: Low Resolution Eye Images

Output: High Resolution Images

- 1: Enhance the low-illumination images using low-light illumination map estimation method.
- 2: Eye is detected from the image using Viola-Jones method,
- 3: The image is subjected to super resolution using the following steps.
 - a: Cropped low resolution image is interpolated as Interpolated Low Resolution (ILR) image and input to the network.
 - b: ILR image pass through 19 times of Conv. and Rectified Linear Unit (ReLu) layer (These are 64 filters with the size of 3 ×3 for each Conv.layer).
 - c: Compute the following loss function to minimize the error between output and input using,

$$\|r - f(x)\|_1 \quad (1)$$
 where x is the ILR image, y is the output image, $r=y-x$ is the residual image and f is the network prediction
- 4: The residual output is added with the ILR image.

PROPOSED SEGMENTATION TECHNIQUE BASED ON ENCODER-DECODER NETWORK

Segmentation is the process of extracting iris region from the eye image for the identification purpose. So, a proper segmentation technique is required for accurate iris recognition, otherwise recognition error rate will be high. The pixels present in the iris are not the same as that of the pixels which are present in the surrounding regions of the eye image. Based on this idea we propose a pixel level classification method for segmentation process. To perform the segmentation of iris automatically, the images has to be labeled. The pixels of the iris region are given the label 'iris 'and rest of the portions including pupil, eyebrows, eyelash, eyelid and periocular regions are assigned another label 'non-iris'. Training dataset consists a collection of labeled images and its corresponding unlabeled images. A SegNet deeplearning model is trained with this training set. Fig.3 shows the proposed architecture for segmentation.

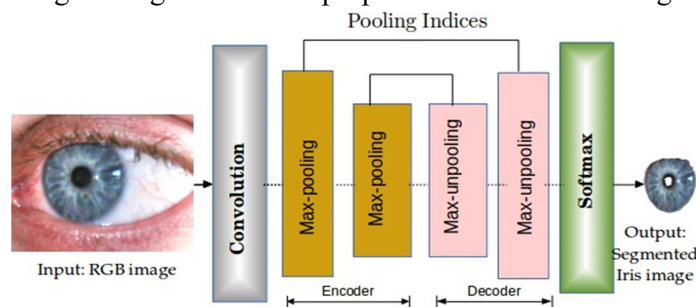


Fig. 3: Schematic Diagram of the Proposed Segmentation Process

Semantic segmentation for the proposed method (SegNet) is very similar to the series of encoder network and a corresponding decoder network, followed by a final pixel wise classification layer. The proposed scheme consists of convolutional operation, max- pooling (downsampling) operations and un-pooling (upsampling) operations. Input RGB image is passed through the convolutional layer. Resultant feature map will be given as the input for the max-pooling operation. max-pooling is a form of down-sampling use to identify the most important features. Later, the output from the previous step is given as the input for the un-pooling (upsampling) process to get the same dimensionality of the input image. For each pooling layer, the max locations are stored. These locations are then used in the un-pooling layer. Finally, a softmax function is used for the prediction for classifying whether the selected pixel is an iris or non-iris pixel. The proposed semantic segmentation is given in Algorithm 2.

Algorithm 2: Proposed Segmentation Method

Input: High resolution iris image.

Output: Segmented iris region from the input image.

- 1: Analyze training data for semantic segmentation.
 - a. Create image data store for training and testing
 - b. Create data store of pixel labeled images: A pixel labeled image is an image where every pixel value represents the categorical label of that pixel.
- 2: Import pre-trained VGG-16 model and modify to SegNet.
- 3: Train the SegNet network.
- 4: Input the test image for the network.

FEATURE EXTRACTION BASED ON CNN

Feature extraction is the process of encoding iris texture into some useful readable bits. For this feature extraction task, a pre-trained CNN model called AlexNet is used for training the images to get the features. Then, this feature map is given as the input to an LSTM network for the classification process as shown in the Fig. 4. Because of the shortage of images per class, the proposed method uses a pretrained AlexNet model to extract the features. The AlexNet architecture is composed of 25 layers with 5 convolutional layers followed by 3 fully connected layers. This pretrained model shows good result for 1000 classes of image data. In the proposed model, we modified the final fully connected layer into 28 classes which is used for the training and testing process. The whole process can be understood from the Algorithm 3.

Proposed Classification Based on LSTM Model

The features extracted from the CNN have the invariance of translation, rotation, and scaling. However, the output layer of the traditional CNN is a fully connected layer with a hidden layer. This feature blend method which takes all outputs of the convolutional. layer is far too simple for the purpose of the proposed model. Difficulties with this method include unsuitable kernels, multiple kernels extracting the same information, and unnecessary information extracted by kernels.

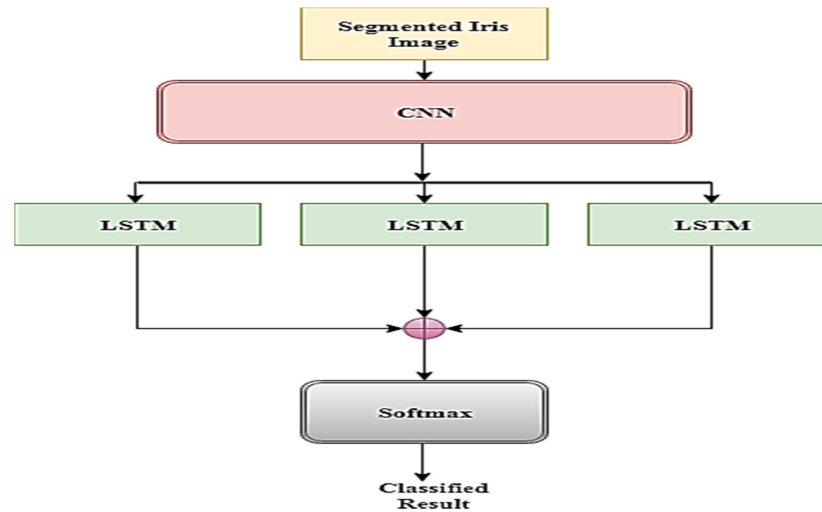


Fig. 4: Convolutional Neural Network and Long Short-Term Memory Network (CNNLSTM) Overall Architecture.

It is possible to extract deeper image features and improve recognition accuracy by increasing the number of convolutional kernels, convolutional layers, and pooling layers. But it will undoubtedly lead to a huge network, thereby increasing the cost of computation and the risk of overfitting. As a time recurrent neural network, LSTM is suitable for handling the sequence problem with time dependence. The input feature vector is selectively forgotten, input and output through three threshold structures. It can filter and fuse the empty input, similar information, and unnecessary information extracted by the convolutional kernels, so that the effective feature information can be stored in the state cell for a long time. Therefore, an algorithm combining CNN and LSTM is proposed in this work for achieving better classification of iris images.

Proposed model can overcome the reliance of feature extraction and data reconstruction relying on human experience and subjective consciousness in traditional recognition algorithms. It uses multiple convolutional kernels to scan the entire input image to obtain redundant features of all objects as candidates. In the feature hybrid stage, the three-dimensional feature vector output from the last layer of CNN is firstly stretched into a one-dimensional feature vector. As mentioned earlier, this vector has all feature information extracted by convolutional kernels, which includes some blank information,

Algorithm 3: CNN-Feature Extraction

Input :Segmented iris image

Output :Feature vector

- 1: Segmented RGB image is resized into 227×227 dimension.
- 2: Perform convolutional operation to the input with 96 kernels of size $11 \times 11 \times 3$. The first two Convolutional layers are followed by the Overlapping Max Pooling layers. The third, fourth and fifth convolutional layers are connected directly. The fifth convolutional layer is followed by an Overlapping Max Pooling layer.

3: Output from the previous step goes into a series of two fully connected layers. The second fully connected layer feeds into a softmax classifier with 28 class labels. ReLU nonlinearity is applied after all the convolution and fully connected layers.

4: The ReLU nonlinearity of the first and second convolution layers are followed by a local normalization step as in (2) before doing the pooling.

$$f(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

5: Apply activation function in the last fully connected layer to extract the features for the classification purpose.

similar information, unnecessary information, and so on. Then the feature vector is mapped to two-dimensional vector as the input of LSTM. Each row of the feature matrix is considered as a basic unit to be hybridized. In each time steps feature information is read from a column in the feature matrix. In this way, a single iris image is converted into sequential data. Steps of the proposed classification model is given in Algorithm 4.

Algorithm 4: Proposed LSTM-Classification

Input: 4096×28 Feature Map.

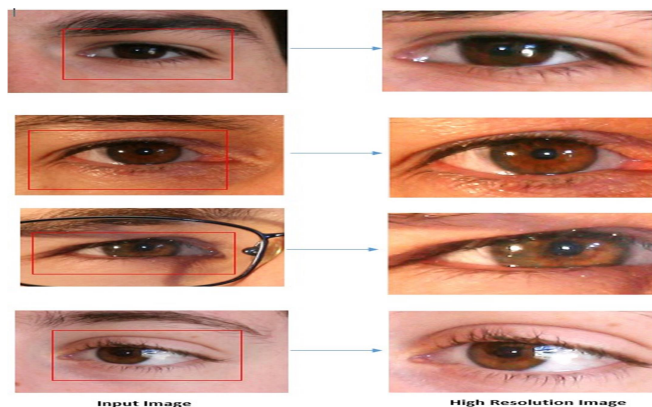
Output: Classification of an iris image.

- 1: Define a LSTM network.
- 2: Pass each sequence of feature vector into LSTM input layer.
- 3: Train the network.
- 4: Effective features extracted from the CNN are stored in cell state .
- 5: Test the network on the dataset for classification.

Experimental Results and Discussion

In order to ascertain the performance of the proposed iris recognition mechanism, UBIRIS.v2 as being one of the publicly available database acquired using visible imaging in unconstrained imaging environments was utilized. This database contains severely noisy images of 171 subjects. The size of the images are 400 × 300 pixels. The images of the iris were acquired from people walking 4 8 meters away from a high-resolution visible light camera in visible light illumination. We have taken 1260 iris images corresponding to 28 subjects for training and testing.

Result of Super Resolution Using VDSR



To improve the resolution of iris images we applied a super resolution technique. Results of an implemented method on sample image are shown in Fig. 5.

Table 1: PSNR of the High Resolution Images

	PSNR
Image 1	41.3652
Image 2	40.0221
Image 3	40.0323
Image 4	40.3670

Result of Segmentation Using SegNet Model

The proposed semantic segmentation method can successfully segment the iris region from the unconstrained images as shown in the Fig. 6.

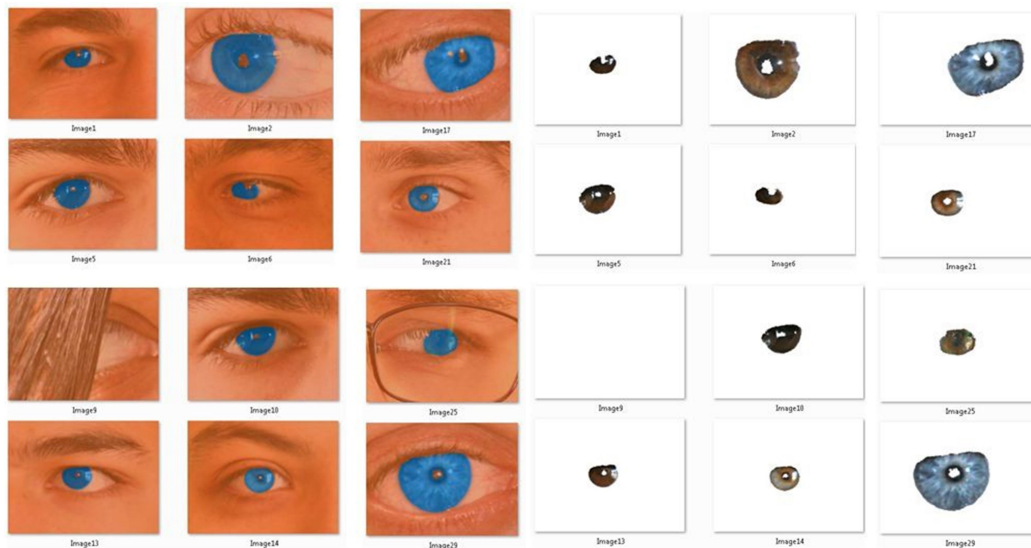


Table 2 shows the confusion matrix for 300 test images. The confusion matrix is a two by two table that contains four outcomes produced by a binary classifier. Table 2 shows that the proposed segmentation method works well for non-ideal iris images. Table 3 gives various measures such as accuracy, Intersection-Over-Union (IoU) Metric, and BF (Boundary F1) Score derived from the confusion matrix to test the algorithm

		Predicted Class	
		Iris	Non-iris
Observed Class	Iris	98.89 % (TPR)	1.107 % (FNR)
	Non-iris	1.729 % (FPR)	98.27 % (TNR)

Table 2: Confusion Matrix of Segmentation Process.

Table 3: Class Wise Performance Matrix for Multiple Test Images

	Accuracy	IoU	BF Score
Iris	0.98893	0.76147	0.54618
Non-iris	0.98271	0.95656	0.85552

Result of Feature Extraction and Classification

The result of the proposed LSTM classification on 504 corresponding image to 28 subjects are given below. The comparison of the performance of the proposed LSTM classification algorithm with SVM on 504 images are given in Table 4. Table shows that proposed one has a better accuracy than SVM classification

Table 4: Comparison of SVM and Proposed LSTM Classification Methods

Method	Accuracy
AlexNet+SVM	0.8325
AlexNet+LSTM (Proposed)	0.9525

The recognition performance of comparing with existing methods are shown in Table 5

Table 5: Comparison of different Methods

Approach	Equal Error Rate	Iris Images
Daugman [14]	0.316	1000
SIFT [14]	0.2521	1000
2D Gabor Filters [15]	0.2614	1227
SVM	0.02	1260
LSTM(Proposed)	0.005	1260

CONCLUSION

In this paper we proposed novel mechanism for segmentation and classification of non-ideal iris images. To improve the quality of captured images we adopted a super resolution technique. The high resolution images are then used for segmentation by using a deep-learning SegNet model, which gives 98% accuracy when compared to existing methods. After segmentation, feature extraction is done by using AlexNet and classification is done using LSTM model. The existing methods used CNN for feature extraction and SVM for classification. In the proposed method we extracted the features using CNN and modified it in the form of a sequence. The sequenced data is then given as an input to LSTM for classification. The proposed method gave an accuracy of 95.25% compared to the existing SVM based approach which gave an accuracy of 83.25% on UBIRIS v2 dataset.

REFERENCES:

1. Alaslani MG, Elrefaei LA. Convolutional neural network-based feature extraction for iris recognition. Int. J. Comp. Sci. Info. Tech.. 2018;10:65-78.
2. Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks. InProceedings of the IEEE conference on computer vision and pattern recognition

-
- 2016 (pp. 1646-1654).
3. H. Proenca et al. The UBIRIS.v2: A Database of Visible Wavelength Images Captured On-The-Move and At-A-Distance. In: IEEE Trans. PAMI 32.8 (Aug. 2010), pp. 1529-1535. DOI: 10.1109/TPAMI.2009.66.
 4. Daugman J. How iris recognition works. In: The essential guide to image processing 2009 Jan 1 (pp. 715-739). Academic Press.
 5. Wildes RP. Iris recognition: an emerging biometric technology. Proceedings of the IEEE. 1997 Sep; 85(9): 1348-63.
 6. Vatsa M, Singh R, Noore A. Improving iris recognition performance using segmentation, quality enhancement, match score fusion, and indexing. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics). 2008 Aug; 38(4): 1021-35.
 7. Ma L, Tan T, Wang Y, Zhang D. Efficient iris recognition by characterizing key local variations. IEEE Transactions on Image processing. 2004 Jun; 13(6): 739-50.
 8. Minaee S, Abdolrashidiy A, Wang Y. An experimental study of deep convolutional features for iris recognition. In: 2016 IEEE signal processing in medicine and biology symposium (SPMB) 2016 Dec 3 (pp. 1-6). IEEE.
 9. Kumar A, Passi A. Comparison and combination of iris matchers for reliable personal authentication. Pattern recognition. 2010 Mar 1; 43(3): 1016-26.
 10. BIT.CASIA Iris Image Database. <http://biometrics.idealtest.org.com>. accessed December 18, 2018.
 11. Liu N, Li H, Zhang M, Liu J, Sun Z, Tan T. Accurate iris segmentation in non-cooperative environments using fully convolutional networks. In: 2016 International Conference on Biometrics (ICB) 2016 Jun 13 (pp. 1-8). IEEE.
 12. Johnson PA, Lopez-Meyer P, Sazonova N, Hua F, Schuckers S. Quality in face and iris research ensemble (Q-FIRE). In: 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS) 2010 Sep 27