Biometric Eye Recognition using Alex-Net

Rishina B

M Tech Student, Computer Science and Engineering, AWH College of Engineering, Calicut, India

---

Abstract - *In recent years the concept of iris recognition has achieved big success on biometric identification. Iris recognition performance is affected if iris is captured at a distance. Further, images captured in visible spectrum are more susceptible to noise than if captured in near infrared spectrum. Difficulties in the collection of iris images with high resolution and in the segmentation of valid regions prevent it from applying to large-scale practical applications. In this paper, an eye recognition framework is considered which is based on deep learning, which relaxes the data collection procedure, improves the anti-fake quality, and promotes the performance of biometric identification. The recognition framework, takes the whole eye image, rather than the segmented and normalized iris ones, as a kind of biometrics. Such system unburdens us the demanding on segmentation and relaxes the acquisition requirements. The framework is performed by extracting learned features from a pre-trained Alex-Net Convolutional Neural Network Model followed by a multi-class Support Vector Machine (SVM) algorithm to perform classification. The proposed eye recognition system is tested on CASIA-Iris-IntervalV4 and the UBIRIS.v2 datasets. The system achieved excellent results with very high accuracy rate.*

Key Words: Eye, Biometrics, Recognition, Deep learning, Alex-Net, Feature extraction(FE), Support Vector Machine(SVM)

1. INTRODUCTION

Biometrics is the technical term for body measurements and calculations which refers to metrics related to human characteristics. Biometrics authentication (or realistic authentication) is used in computer science as a form of identification and access control. Biometric identifiers are the distinctive, measurable characteristics used to label and describe individuals. Biometric identifiers are often categorized as physiological versus behavioral characteristics. Physiological characteristics are related to the shape of the body. Examples include, but are not limited to fingerprint, palm veins, face recognition, DNA, palm print, hand geometry, iris recognition, retina and odour/scent. Behavioral characteristics are related to the pattern of behavior of a person, including but not limited to typing rhythm, gait, and voice. However, the collection of biometric identifiers raises privacy concerns about the ultimate use of this information.

Many researches achieved notable progresses in the recognition of iris. The foremost algorithms has reached the accuracy of more than 99%. But, iris recognition has certain disadvantages which gives negative results. Daugman etc. suggested that if we want to get a satisfied accuracy of iris recognition, the radius of iris should be at least 80-130 pixels. The internal factors such as eyelid and eyelash, and external factors, such as eyeglasses, can decrease the quality of the iris images. The iris has to be precisely segmented from the eye/face image with robust segmentation algorithms. When pupil is not strictly a circle it makes the segmentation and the advancing procedures, such as normalization, even more difficult.

Considering all these factors, an eye recognition framework is proposed with Alex-Net Convolutional neural network followed by a Support Vector Machine (SVM) classifier. Compared with iris, the whole eye region consisting of the eyelash, pupil, iris, sclera, eyelid, skin, even with wrinkles and eyeglasses makes gathering easier. The components of eye image is shown in Fig:1.

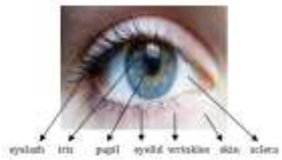

Fig -1: Components of eye image

The eyelash, eyelid, skin, pupil, iris, sclera, wrinkles and eyeglasses was the negative factors in iris recognition which is included in the eye region, are transferred into positive factors in the proposed method. So that new features can provides helpful information for recognition purpose. In this paper, with these concerns, proposes a recognition framework, that takes the whole eye image, rather than the segmented and normalized iris ones, as a kind of biometrics and can employ the strong capability of multi-level feature learning in deep neural network. Such system unburdens us the demanding on segmentation and relaxes the acquisition requirements.

The main contributions of this study are as follows.

(1)Alex-Net based recognition framework utilizes the raw data instead of using local descriptors for feature extraction. The features are self learned by the Alex-Net framework. Using this neural network can improve the accuracy.
(2) The eye recognition framework breaks the traditional iris recognition system with high efficiency, reliability and practicability. The new framework make the recognition task more appropriate.

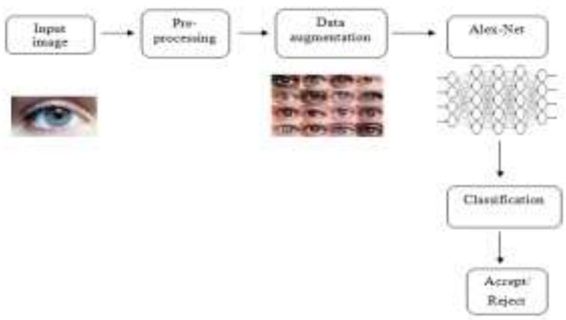

Fig -2: The Framework

The rest of this paper is organized as follows: Section 2, provides the works presented by other authors. Section 3, introduces a description of the proposed eye recognition system. Section 4, the experimental results is presented. Finally, the Conclusion is given in section 5.

3. PROPOSED EYE RECOGNITION SYSTEM

The total framework and the Alex-Net architecture of eye recognition system are shown in Figure 2 and Figure 5 respectively. The architecture describes overall process details which includes pre-processing, data augmentation, Alex-Net, classification.

3.1 Pre-processing

In image pre-processing, image data recorded restrain errors related to geometry and brightness values of the pixels. These errors are corrected using appropriate mathematical models which are either definite or statistical models. Image enhancement is the modification of image by changing the pixel brightness values to improve its visual impact. Image enhancement involves a collection of techniques that are used to improve the visual appearance of an image, or to convert the image to a form which is better suited for human or machine interpretation. Sometimes images obtained from conventional and digital cameras lack in contrast and brightness because of the limitations of imaging sub systems and illumination conditions while capturing image. Images may have different types of noise. In image enhancement, the goal is to accentuate certain image features for subsequent analysis or for image display. Enhancement algorithms are generally interactive and application dependent. Some of the enhancement techniques are:

3.1.1 Contrast Stretching

Some images are homogeneous i.e., they do not have much change in their levels. In terms of histogram representation, they are characterized as the occurrence of very narrow peaks. The homogeneity can also be due to the incorrect illumination of the scene. Ultimately the images hence obtained are not easily interpretable due to poor human perceptibility. This is because there exists only a narrow range of gray-levels in the image having provision for wider range of gray-levels. The contrast stretching methods are designed exclusively for frequently encountered situations. Different stretching techniques have been developed to stretch the narrow range to the whole of the available dynamic range.

3.1.2 Noise Filtering

Noise Filtering is used to filter the unnecessary information from an image. It is also used to remove various types of noises from the images. Mostly this feature is interactive. Various filters like low pass, high pass, mean, median etc., are available

3.1.3 Histogram modification

Histogram has a lot of importance in image enhancement. It reflects the characteristics of image. By modifying the histogram, image characteristics can be modified. One such example is Histogram Equalization. Histogram equalization is a nonlinear stretch that redistributes pixel values so that there is approximately the same number of pixels with each value within a range. The result approximates a flat histogram. Therefore, contrast is increased at the peaks and lessened at the tails

3.1.a Image segmentation(only performed if there is a segmentation problem)

Segmentation is one of the key problems in image processing. Image segmentation is the process that subdivides an image into its constituent parts or objects. The level to which this subdivision is carried out depends on the problem being solved, i.e., the segmentation should stop when the objects of interest in an application have been isolated. Image thresholding techniques are used for image segmentation.

After thresholding a binary image is formed where all object pixels have one gray level and all background pixels have another - generally the object pixels are 'black' and the background is 'white'. The best threshold is the one that selects all the object pixels and maps them to 'black'. Various approaches for the automatic selection of the threshold have been proposed. Thresholding can be defined as mapping of the gray scale into the binary set {0, 1}.

When the histogram has two pronounced maxima, which reflect gray levels of object and background, it is possible to select a single threshold for the entire image. Sometimes gray level histograms have only one maximum. This can be caused, e.g., by inhomogeneous illumination of various regions of the image. In such case it is impossible to select a single thresholding value for the entire image and a local binarization technique must be applied. Segmentation of images involves sometimes not only the discrimination between objects and the background, but also separation between different regions. One method for such separation is known as watershed segmentation.

3.2 Data augmentation

To show the neural net with different variation of the same image helps to prevent over fitting.

3.2.1 Data augmentation by mirroring

If we have an image of an eye in the training set, its mirror image is also a valid image of a eye. The figure below for an example. Can double the size of the training dataset by simply flipping the image about the vertical axis.

Fig -3: Example for image augmentation by mirroring

3.2.2 Data augmentation by various random crops

In addition, cropping the original image randomly will also lead to additional data that is just a shifted version of the original data. The input given to AlexNet extracts random crops of size 227×227 from inside the 256×256 image boundary to use as the network's inputs. The size of the data is increased by a factor of 2048.

Fig -4: Random crops of eye image

Notice the four randomly cropped images look very similar but they are not exactly the same. This teaches the Neural Network that minor shifting of pixels does not change the fact that the image is still that of a eye. Without data augmentation, cannot be able to use such a large network because it would have suffered from substantial over fitting.

3.3 AlexNet Model

AlexNet is a Convolutional neural network model. Alex-Net was much larger than the previous CNNs used for computer vision tasks There are 60 million parameters and 650,000 neurons. Alex-Net took five to six days to train on two GTX 580 3GB GPUs. Normally an AlexNet consists of 5 Convolutional Layers and 3 Fully Connected Layers.

Multiple Convolutional layers will extract the important features in an image. In a single convolutional layer, there are usually many kernels of the same size. As an example, 96 kernels of size 11x11x3 is contained in the first convolutional layer of Alex-Net. The width and height of the kernel are usually the same and the depth is the same as the number of channels.

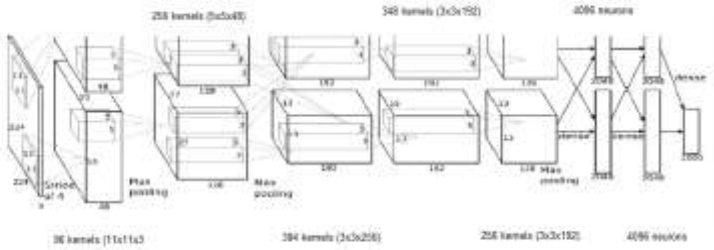

Fig -5: Alex-Net architecture

When we consider the Alex-Net architecture, the first two Alex-Net convolutional layers are followed by the max pooling layers which are overlapped. The third, fourth and fifth layers of convolutional are connected directly and the fifth layer is then followed by the max pooling layer which are overlapped. The output will be sented to the series of two fully connected layers and the second fully connected layers will be given to softmax classifier with more labels.

Then comes the ReLU activation function. ReLU nonlinearity function is applied after all the convolution and fully connected layers. The first and second convolution layers ReLU nonlinearity are followed by a local normalization step before doing pooling. Many researches doesn't find normalisation as useful.

Max pooling layer

To down sample the width and height of the tensors max pooling layers are used. Max pooling keeps the depth as same always. Overlapping of the max pool layers are similar to the max pool layers, except the adjacent windows over which the max is computed overlap each other. Many of the authors used 3 x 3 size pooling windows with a stride of 2 between the adjacent windows. That overlapping nature of pooling will help to reduce the top-1 error rate by 0.4% and top-5 error rate by 0.3% respectively when compared to using non-overlapping pooling windows of size 2×2 with a stride of 2 that would give same output dimensions.

ReLU Nonlinearity

An important feature of the AlexNet is the use of ReLU(Rectified Linear Unit) Nonlinearity. Tanh or sigmoid activation functions are widely used to train a neural network model.
Traditional activation functions sigmoid Equation is

$$f(x) = \frac{1}{1+e^{-x}} \quad (1)$$

Tanh activation function equation is

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2)$$

The ReLU function is suitable for forward/back propagation because of its concise form as shown in Eq. 3, compared with the complex operations of sigmoid and tanh[1].

$$f(x) = \max(0, x) \quad (3)$$

Rectified Linear Units (ReLU) becomes more and more popular on large datasets since convolutional neural network has been created in recent years. AlexNet showed that using ReLU nonlinearity, deep CNNs could be trained much faster than using the saturating activation functions like tanh or sigmoid. The figure below shows that using ReLUs(solid curve), AlexNet could achieve a 25% training error rate six times faster than an equivalent network using tanh(dotted curve). This was tested on the CIFAR-10 dataset.

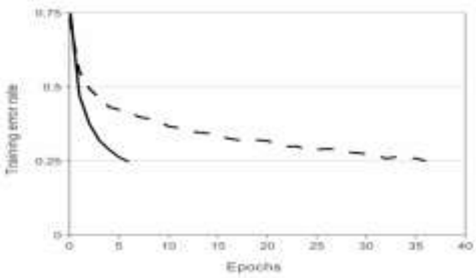

Fig-6: Solid curve of Relu

The ReLU function is given in equation (3)
The tanh function saturates at very high or very low values of z. At these regions, the slope of the function goes very close to zero. This can slow down gradient descent. On the other hand the ReLU function's slope is not close to zero for higher positive values of z. This helps the optimization to converge faster. For negative values of z, the slope is still zero, but most of the neurons in a neural network usually end up having positive values. ReLU wins over the sigmoid function too for the same reason.

3.4 The classification stage

There are different types of classifiers used for this task, for example, Support Vector Machine, Softmax Regression, and Neural Network are some of these. The classifier is used after feature extraction is to find the corresponding label for every test image. In this work, a Support Vector Machine classifier has been used.

SVM is described as follows:

Assuming we have the set of training data (x1,y1),(x2,y2),…(xn,yn) and we want to classify the set into two classes where xi belongs to Rd is the feature vector and yi belongs to {-1, +1} is the label class. The two classes are linearly separable with a hyperplane w.x+b=0. With no other previous knowledge about the data, SVM can find the optimal hyperplane as the one with the maximum margin [3]. The multi-class SVM can be implemented for a set of data with M classes, we can train M binary classifiers that can distinguish each class against all other classes, then select the class that classifies the test sample with the greatest margin [2].

Algorithm 1: Extracting the features of images using a pre-trained Alex-Net convolutional neural network and classifying the features using SVM algorithm.

//Input: The input images
//Output: The recognition accuracy
1: Load the input images and its labels.
2: Split each category into the similar number of images.
3: Load the pre-trained CNN (Alex Net model)
4: Pre-process the images For Alex Net model
5: Split the sets of the images into training and testing data.
6: Extract features from the deeper layers of Alex-Net model.
7: Get training labels from the training set.
8: Use the training features to train a multiclass SVM classifier.
9: Extract the features from test set.
10: Use the trained classifier to predict the label for test set.
11: Get the known labels for test set.
12: Results are tabulated by a confusion matrix.
13: Convert confusion matrix into percentage form.
14: Display the mean accuracy.

4. EXPERIMENTAL RESULTS

a) Setup

The system and its stages are implemented using MATLAB 2018b on a laptop with Core i7 CPU running at 2.8GHz. The two databases used are (a) The CASIA-Iris-IntervalV4 database, (b) The UBIRIS.v2 database.

b. Evaluation on Data Augmentation

By doing data augmentation the system performance are investigated by the two datasets. Without data augmentation, the performance is quite unsatisfied which are 45.83% and 12.64%, respectively. Many kind of data augmentation strategies are implemented the accuracies are gradually increasing. By looking at the images in the dataset we can understand that images are seldom captured with various degrees of rotation, but on the other hand, a lot of them are centered at different locations, which more or less explain why cropping works much better than rotation. An important thing to consider is the cropping size and interval. These also have an influence in the performance. It is sure that different cropping sizes and intervals perform differently on different datasets. Cropping a large image by using a window is to retain most of the useful information from image, but is useless for increasing the performance. The cropping parameters in UBIRIS.v2 are also three times of those in CASIA-Iris-IntervalV4, perfectly explains the simulation result.

c. Performance comparison with classic approaches

Many approaches are there for performing the recognition. To compare the performance between our proposed approach with classic ones, several traditional classifiers are used, which includes the kNN, and plain convolutional neural network (CNN) on the CASIA-Iris- IntervalV4 dataset. To control the training time in a tolerable range, we take a data augmentation strategy that limits the number of training samples less than 200,000. The number of images in the training set for this experiment is 110382. Table 1 presents the performance of each method.

Table -1: Performance comparison with classic approaches.

Approach	Accuracy
k-Nearest Neighbor	85.23%
CNN	91.32%
Support Vector Machine(SVM)	93%
Naïve Bayes	87.56%

Table 1 results that Support Vector Machine approach perform much better than traditional classifiers. Specifically, the proposed method with Alex-Net overcomes the one with other convolutional neural network models.

d. Performance evaluation with different architectures

Different types of architectures are used in deep learning. Here the verification is performed by comparing the performance with other architectures. The results shown in Table 2 indicate that the proposed Alex-Net achieves the highest accuracy on the CASIA-Iris-IntervalV4 dataset which is with the same data augmentation strategy.

Table-2: Performance with other models

Index	Model	Accuracy
1	Convolutional neural Network (CNN)	85.87%
2	ResNet	89.76%
3	VGG16	90.88%
4	Alex-Net	93.24%

The Alex-Net model followed by the Support Vector Machine classifier is the best choice for the eye recognition.

5. CONCLUSIONS

This paper proposes a novel eye recognition framework that overcomes the defects in the traditional iris recognition system. Such framework removes the feature extraction procedures, which increase the learning complexity, and are influenced by a number of issues such as the low resolution of the sample, and the interferences from eyelid, eyeglasses, etc. Compared with iris, the whole eye region consisting of the eyelash, pupil, iris, sclera, eyelid, skin, even with wrinkles and eyeglasses makes gathering easier. To apply the eye recognition problem, a deep learning architecture is used. The extracted learned features from a pre-trained Alex-Net Convolutional Neural Network followed by multi-class SVM algorithm is used to perform iris recognition. Also, we study how different data augmentation strategies have influences on the system performance, and it turns out that such eye recognition framework relies on a proper augmentation strategy to work well.

The experimental results on the CASIA-Iris-IntervalV4 show that this approach achieves highest accuracy within a proper number of training epoch/iteration. With the Alex-Net, the possible accuracy rates on the CASIA-Iris-IntervalV4 and the UBIRIS.v2 datasets are 97.11% and 92.58%, respectively, which not only exceeds the performances of traditional classifiers and other previously proposed deep learning based architectures, but is competitive with the most state-of-art iris recognition algorithms as well.

In the future the combination of whole eye and the periocular region which include eyebrow features can also be added. And also the combination of convolutional neural networks can increase the accuracy rate and performance rate.

REFERENCES

1. Zi Wang, Chengcheng Li, Huiru Shao, Jiande Sun, "Eye Recognition with Mixed Convolutional and Residual Network(MiCoRe-Net)", School of Information Science and Engineering, Shandong Normal University, Jinan 250014, Shandong Province, China. 2018
2. Maram.G Alaslani1 and Lamiaa A. Elrefaei, "Convolutional Neural Network based Feature Extraction for Iris Recognition", International Journal of Computer Science & Information Technology (IJCSIT) Vol 10, No 2, April 2018.
3. S. Zhan, Q. Q. Tao, and X. H. Li, “Face detection using representation learning,” Neurocomputing, vol. 187, pp. 19–26, 2016.
4. S. Minaee, A. Abdolrashidiy, and Y. Wang, "An experimental study of deep convolutional features for iris recognition," in Signal Processing in Medicine and Biology Symposium (SPMB), 2016 IEEE, 2016, pp. 1-6.
5. K. Itqan, A. Syafeeza, F. Gong, N. Mustafa, Y. Wong, and M. Ibrahim, "User identification system based on finger-

vein patterns using Convolutional Neural Network," ARPN Journal of Engineering and Applied Sciences, vol. 11, pp. 3316-3319, 2016.
6. C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation," IEEE Transactions on Multimedia, vol. 17, no. 11, pp. 2049–2058, 2015.
7. Z. Zhou, E. Y. Du, N. L. Thomas, and E. J. Delp, "A comprehensive multimodal eye recognition," Signal Image & Video Processing, vol. 7, no. 4, pp. 619–631, 2013.
8. Z. Zhou, E. Y. Du, C. Belcher, N. L. Thomas, and E. J. Delp, "Quality fusion based multimodal eye recognition," in IEEE International Conference on Systems, Man, and Cybernetics, 2012, pp. 1297–1302.
9. Y. Du and E. J. Delp, "Multimodal eye recognition," Proceedings of SPIE - The International Society for Optical Engineering, vol. 7708, no. 1, pp. 247–247, 2010.

BIOGRAPHY

Rishina B is a student of department of Computer Science and Engineering, AWH Engineering College, Calicut, Kerala, India. Her current interests include deep learning, artificial neural networks.