EYE GESTURE CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK

[1]A. Rajaprabu, [2]Dr. A. Geetha
[1]Research Scholar, [2]Associate Professor
[1, 2] Department of Computer Science and Engineering
[1, 2] Annamalai University, Annamalainagar, Taminadu, India

***Abstract* :** Estimation of eye gesture recognition is useful in various human-computer interaction tasks. We propose a framework for the classification of eye gestures. In the first stage, the algorithm detects face, and from the detected face, eye regions are detected using Haar-cascaded classifiers. A Convolutional Neural Network is employed in this work for the classification of eye gestures. The detected eye-gesture frames are recognized as eye-close, eye-left and eye-right. This proposed work will be usefull in developing hands-free computer interfaces to replace the traditional input devices to create an intelligent Human-Computer Interaction and also usefull in Medical applications,Gaming and Robotic research. The proposed system has been tested and performance analysis is carried out for the model.

***Keywords* – Human Computer Interaction, Eyegestures, Haar-cascadedclassifiers, Convolutional Neural Network, Robotic research.**

I. INTRODUCTION

As the computer technology[1] is growing up, the importance of Human Computer Interaction is rapidly increasing. Most of the mobile devices and laptops are using touch screen technology. But this technology is still not cheap enough to be used on desktop systems. Creating a virtual human computer interactive module such as mouse or keyboard, can be an alternative way for the touch screen. The motivation is to create an object tracking application to interact with the computer, and develop a virtual human computer interaction device. Internet and telecommunications are becoming increasingly integrated processing, storage, access, information distribution and content management system. This convergence is based on the rapid development of digital technology and Internet diffusion concept.

In this proposed work, eyegestures are classified as eye-close, eye-ball left, and eye-ball right. These three states of eye gestures are mapped to Human Computer Interaction tasks. For example mouse and keyboard event handling.
The highlights of the proposed work is shown below:

- A framework is proposed for eye gesture classification.
- Convolutional Neural Network is used as eye gesture classifier, which is robust against eye localization errors.
- The proposed approach outperforms state of the art algorithms in eye gesture classification.

1.1 Deep learning

Deep learning is a sub-field of machine learning dealing with algorithms inspired by the structure and function of the brain called artificial neural networks. In other words, it mirrors the functioning of our brains. Deep learning algorithms are similar to how nervous system structured where each neuron connected each other and passing information. Deep learning models work in layers and a typical model atleast have three layers. Each layer accepts the information from previous and pass it on to the next one. Deep learning models tend to perform well with amount of data wheras old machine learning models stops mproving after a saturation point.

1.2 Convolutional Neural Network

Convolutional Neural Networks (CNN) is shown in figure1 is a class of deep, Feed-Forward (not recurrent) Artificial Neural Networks that are applied to analyzing visual imagery. The process is a 2D convolution on the inputs. The dot products between weights and inputs are integrated across channels. Filter weights are shared across receptive fields. The filter has same number of layers as input volume channels, and output volume has same depth as the number of filters. TheArchitecture of Convolutional Neural Network consists of input layer, convoltional layer, pooling layer, flattening and fully connected layers. Here the raw images taken as input and then these inputs are passed on to Convolutional layer. Dot products between inputs and their corresponding weights are taken place here. RELU(Rectified Linear Unit) is used to increase non-linearity of the network without affecting receptive fields of convolutional layers. Pooling layer do the work of down sampling according to its pool size. Fully connected layer do the work of classication with the help of a special kind of activation layer called as softmax.

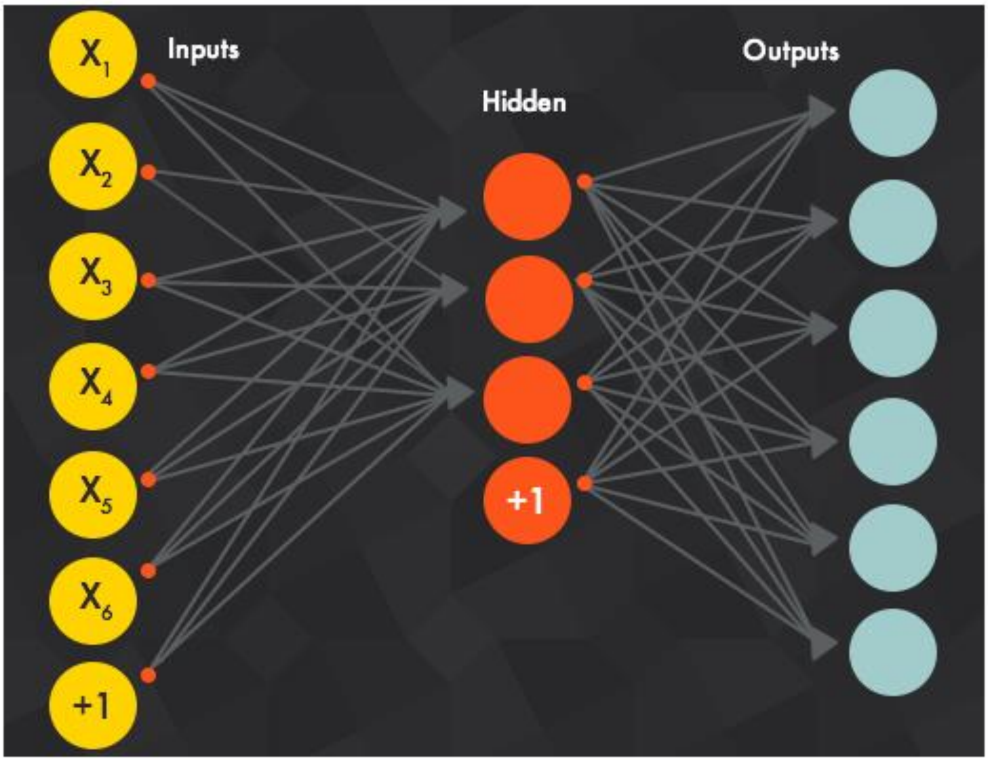

Figure1:Architecture of Deep Neural Network

II. RELATED WORKS

2.1 Efficient CNN Implementation for Eye-Gaze Estimation on Low-Power/Low-Quality Consumer Imaging Systems.

In this paper[2] a new hardware friendly, convolutional neural network model with minimal computational requirements is introduced and assessed for efficient appearance-based gaze estimation. The model is tested and compared against existing appearance based CNN approaches, achieving better eye gaze accuracy with significantly fewer computational requirements.

2.2 Strabismus Recognition Using Eye-tracking Data.

Strabismus[3] is a common ophthalmic disease that can lead to weak 3D perception, amblyopia, or even blindness if it is not timely diagnosed and well treated. Here Strabismus recognition done by using eye-tracking data and Convolutional Neural Network.

2.3 Real time Eye gaze direction classification using Convolutional Neural Network.

In this paper[4] Real time frame work was developped to recognize Eye Gaze direction and also estimate Eye Accessing Cues. From the Estimated EAC human behaviours were studied. Here Convolutional Neural Network is used to recognize the Eye gaze direction.

2.4 Deep Neural Network Ensemple Architecture for Eye movements Classification.

In this paper[5] Visual feature extraction done by using Deep Convolutional Neural Network and Temporal information gathering done by Recurrent layers. Both DNN(Deep Neural Network) and RNN(Recuurent Neural Network) run parallely to finish the task.

2.5 A Novel Approach for Human Computer Interface Based on Eye Movements for Disabled People.

In this paper[6], image based eye tracking technique is used for interaction. The aim of this paper is to help the disabled people to use the computer efficiently. It is based on controlling cursor movements on the screen using only the eyes. The cursor is moved in the desired direction on the screen based on the estimated eye gaze.

The Figure 2 shows the overall system design and implementation. At first Eye movement video is read from a specified location using opencv video read function, then each frame of the video converted in to gray scale image and resized with 150 pixels width and height. The pre processed each frame of video stored in to a batch array. Then this batch array fit in to CNN Model. Here prediction is done with the help of the model with labeling their corresponding class name.

III. SYSTEM DESIGN

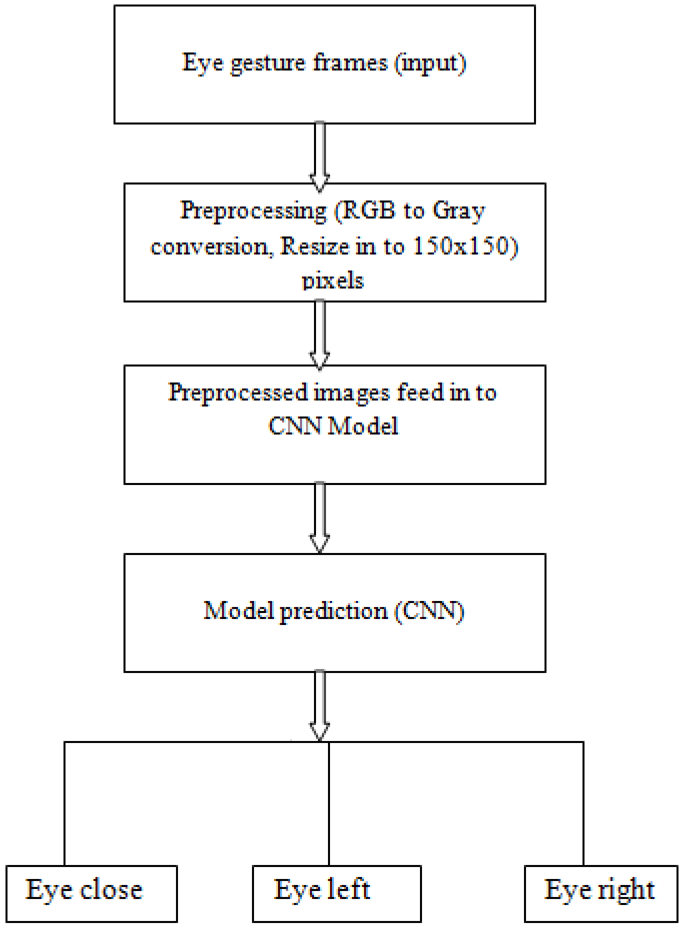

Figure 2: Overall Flow Diagram of the Proposed Work

IV. TRAINING, VALIDATION AND TESTING USING CNN ALGORITHM

Convolutional Neural Network algorithm[7] is a multilayer perceptron that is the special design for identification of two-dimensional image information. The key technology of CNN is the local receptive field, sharing of weights, sub sampling by time or space, so as to extract feature and reduce the size of the training parameters. The advantage of CNN algorithm is to avoid the explicit feature extraction, and to learn implicitly from the training data. A CNN consists of an input and output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consists of convolutional layers, pooling layers, fully connected layers and Normalization layers.

Convolutional layers apply a convolution operation to the input, passing the result to the next layer. The convolution emulates the response of an individual neuron to visual stimuli. Each convolutional neuron processes data only for its receptive field. Pooling layers combines the outputs of the neuron clusters at one layer in to a single neuron in the next layer. For example max pooling uses the maximum value from each of a cluster of neurons at the prior layer. Like that average pooling which uses the average value from each of a cluster of neurons at the prior layer. Fully connected layers connect every neuron in one layer to every neuron in another layer. It is in principle the same as the traditional multilayer perceptron Neural Network.

V. SYSTEM IMPLEMENTATION AND TESTING

Deeplearning enviornment has been set by by installing python 3.5 version along with anaconda library. Opencv(version 3.3.0) library is successfully linked with python 3.5 interpreter. Using conda install command required libraries such as numpy, PIL, scikitlearn, scipy, sklearn, pyautogui, glob, pyautogui are imported.

The CNN architecture(CNN-Convolutional Neural Network) of the proposed system is designed with a simple stack of 3 convolution layers followed by relu activation function and followed by maxpooling layers. Here the filter kernel size is set with 3x3. Single neuron is set for outputlayer. Here the problem consists of two classes only. So when the model is compiled, categorical_cross entropy is used as parameter in model.compile() function. The other two parameters are rmsprop as an optimizer and accuracy as metrics.

5.1 Training

In training phase,following parameters were set.

1.No.of epochs is 5.

2.No.of training samples for each class consist of 250 eye close frames,250 eye-left frames and 250 eye-right frames.

The model.fit_generator() function trains the frame samples and feature extraction takes place automatically.The eye-close,eye-left and eye-right image frames are stored in a stack-array and resized to 150x150. The array contents are stacked in to a batch and passed on to the model for training to classify the above three eye gestures. The model passes through many epochs and classifies the images with class indices 0^{th} position for eye-close and first position for eye-left and second position for eyeright.

5.2 Evaluation on the Validation Set

The model.evaluate_generator (validation_generator,validation_samples) function evalautes the validation samples to calculate the accuracy of the proposed system. In our problem totally 150 eye gesture frames are used for validation. The validation directory consists of 50 eye-close frames, 50 eye-left frames and 50 eye-right frames. After applying data augumentation technique in the model the accuracy is lightly improved. This above technique reduces the overfitting and allows better generalization capability to the Network.

5.3 Testing

In testing phase, 100 mixed eye-close(28), eyeleft(33) and eyeright(39) image frames are taken. These frames were stored in a stack-array and resized in to 150 x150 pixels size. The array contents are stacked in to a batch and passed on to the already trained CNN model for testing. Here prediction takesplace with their corosponding label names. Miss classification also takesplace according to model accuracy.

5.4 Output Screen Shots

The classified eye gesture classification output is shown in the following Figures.

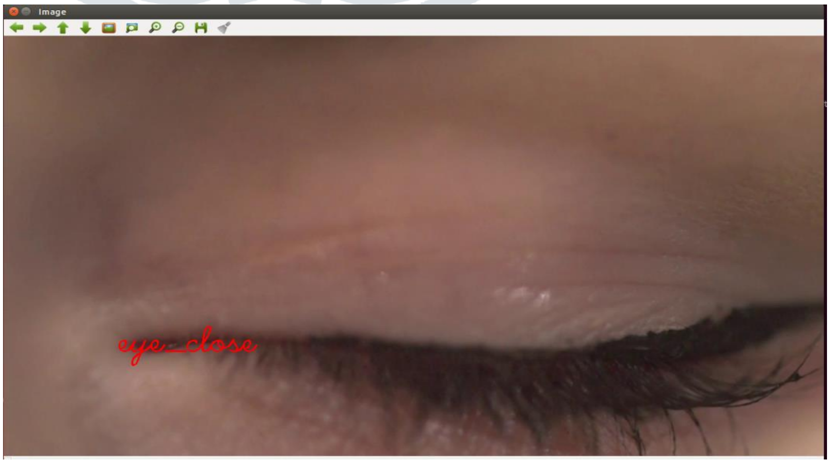

Figure 3: Eye-close gesture recognition

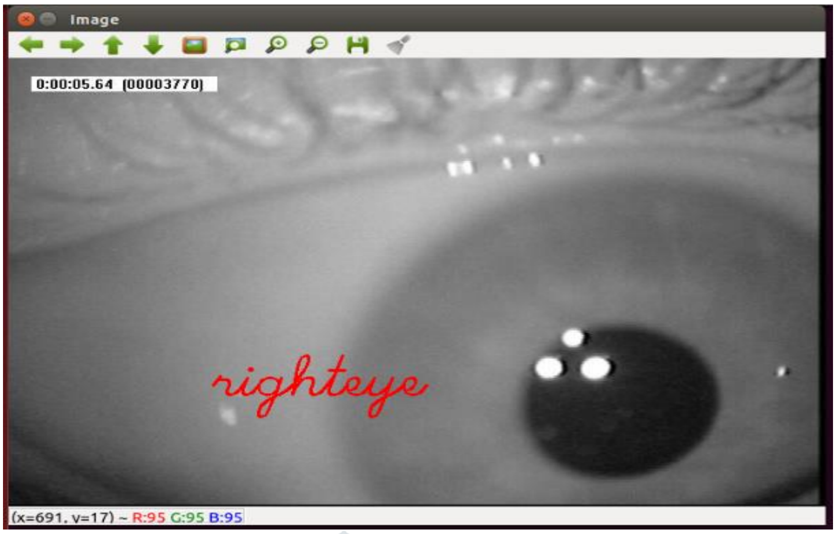

Figure 4 : Eye-right gesture recognition

Figure 5: Eye-left gesture recognition

VI. RESULTS AND DISCUSSION

The performance of the proposed system is analysed using the metrics as below and the results are shown .

6.1 Confusion Matrix for CNN Model

A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

Table 1 Basis of Confusion Matrix

Actual class		**Predicted class**	
		P	N
	P	True positive (TP)	False Negative (FN)
	N	False positive (FP)	True Negative(TN)

Table 2 Calculated Confusion Matrix for CNN

	Eye_close	Eye_left	Eye_right
Eye_close	26	2	0
Eye_left	0	33	0
Eye_right	0	0	39

6.2 Precision, Recall and Accuracy

In pattern recognition ,precision is the fraction of relevant instances among the retrieved instances ,while recall is the fraction of relevant instances that have been retieved over the total amount of instances. Accuracy is the Number of correct predictions among the Total number of predictions. The following formulas are used to calculate Precision,Recall and Accuracy from their respectieve confusion matrices.

Precision =TP/TP+FP
Recall =TP/TP+FN
Accuracy =TP+TN/TP+TN+FP+FN

Where,
TP-True Positive
FP-False Positive
TN-True Negative
FN-False Negative

Table 3 Classification Report for CNN

	Precision(%)	**Recall(%)**	**Accuracy**
Eye-close	92.85	100	98
Eye-left	100	94.28	98
Eye-right	100	100	100
Avg/Total	97.61	98.09	98.66

The calculated accuracy for Convolutional Neural Network is 98.66%.

VII. CONCLUSION AND FUTURE WORK

The proposed system aims at classifying eye gestures with higher accuracy for the purpose of implementing human computer interaction tasks.The system found to produce satisfactory performance with 98.66% accuracy. The developed system is tested with mixed eye frames and performed well. The system will be extended to recognize more directions in eye gaze.

REFERENCES

[1] Ramesh R* and Rishikesh M, “Eye Ball Movement to Control Computer Screen”, Research article, Journal of B iosensors and Bio electronics 2015.

[2] Joseph Lemley,and Anuradha Kar, “Efficient CNN Implementation for Eye-Gaze Estimation on Low-Power/Low-Quality Consumer Imaging Systems”, IEEE, June 2018.

[3] Zenghai Chen, Hong Fu , Wai-Lun Lo, and Zheru Chi ,” Strabismus Recognition Using Eye-Tracking Data and Convolutional Neural Networks”, Journal of Healthcare Engineering , Volume 2018 , Article ID 7692198.

[4] Anjith George, Real-time, ” Eye Gaze Direction Classification Using Convolutional Neural Network” , IEEE, 2018.

[5] Marko Arsenovic, Srdjan Sladojevic, Darko Stefanovic, Andras Anderla, “Deep Neural Network Ensemble Architecture for Eye movements Classification” , IEEE, 2018.

[6] S.S.Deepika and G.Murugesan, “A Novel Approach for Human Computer Interface Based on Eye Movements for Disabled People” , IEEE, 2015.

[7] Tianyi Liu, Shuangsang Fang, Yuehi Zhao, Peng Wang, Jun Zhang,”Implementation of Trainning Convolutional Neural Networks ”.