Behavioural Anomaly Detection Using Eye Movements

Joe Ninan George[1], Affien Panicker[2]
[1]*Software Engineer, Full Creative, Thiruvananthapuram, Kerala, India*
[2]*B.Tech Graduate, Thiruvananthapuram, Kerala, India*

***Abstract* - In a variety of human-computer interface tasks, estimating eye gaze direction is beneficial. Knowledge of the user's gaze direction can provide useful information about the user's focus. Certain eye movements, according to proponents of Neuro-Linguistic Programming (NLP), provides an educated guess about whether or not the person matches the description with the standard set of directions. Despite widespread acceptance of this concept, no previous research has been conducted to test its validity. The classification of eye gaze direction is done using a convolutional neural network.**

***Index Terms* – eye gaze, eye movements, Neuro-Linguistic Programming, convolutional neural network.**

I.INTRODUCTION

In order to create a natural human-computer interaction (HCI) system, it is necessary to first understand human emotions and cognitive states. Human interaction can be made more natural by systems that can detect human emotive and cognitive states. Understanding mental processes can aid computer systems in interacting intelligently with people. Recently, many works have been reported investigating the use of facial expression in HCI. The eyes of a person give a wealth of information about their cognitive processes and emotions. The movement patterns of the eyes reveal information about exhaustion, illnesses, and so forth. Pupil dilation has also been utilised as an indication in cognitive processes research. Each person's eye motions are distinct. Several research have recently advocated using eye movement pattern as a biometric. The majority of efforts on face expression are restricted to desktop settings. However, with the advancement of wearable technologies such as Google Glass and other augmented reality goggles, there are greater chances for employing eye analysis techniques to comprehend users' affective and cognitive processes. The patterns in which the eyes move when humans access their memories is known as eye accessing cues (EAC). The patterns in this non-visual gaze directions have been reported to contain information regarding mental processes. In Neuro-Linguistic Programming (NLP) theory, eye accessing cues gives information about the mental processes from the direction of eye gaze. These movements are reported to be related to the neural pathways which deal with memory and sensory information. The direction of the iris in the socket can give information regarding various cognitive processes. Each direction of non-visual gaze is associated with different cognitive processes.

Neuro-Linguistic Programming (NLP) consists of a diverse collection of psychological techniques that aim to enhance peoples' lives. An important aspect of the work involves attempting to improve people's communication skills by teaching them about an alleged relationship between eye- movements and thought. According to this work, when right handed people look up to their right they are likely to be visualising a 'constructed' (i.e., imagined) event, whilst when they look up to their left they likely to be visualising a 'remembered' memory (i.e., an event that has actually happened to them). In contrast, when they look to their right they are likely to be thinking about a 'constructed' sound, and when they look to their left they are likely to be thinking of a 'remembered' sound. These alleged relationships are frequently taught in NLP training courses, and are ubiquitous on the internet. Indeed, a Google search on the terms 'neuro-linguistic programming' reveals thousands of sites describing the alleged relationship, and two well-known YouTube videos encouraging lie detectors to adopt this approach have received 30,000 and 60,000 views respectively.

In this study, we present a real-time framework for detecting eye gaze direction in desktops utilising off-the-shelf, low-cost cameras. The estimation of gaze position from a webcam sometimes necessitates a time-consuming calibration method. This paper describes previous research on desktop gaze tracking as well as the operation of our suggested system. It also includes a concise overview of the system's development process.

II.LITERATURE REVIEW

The use of a Convolutional Neural Network in a real-time framework for eye gaze direction classification is proposed [1], with average frame rates of 24 frames per second in desktop environments.

An excellent review of gaze tracking methods in desktop environments using state-of-the-art works related to eye gaze direction estimation can be found in [2].

Vrânceanu et al. introduced a methodology [3] for automatic EAC classification. In their methodology, they make use of colour space data. The visual accessing cues are classified based on the relative position of the iris and sclera in the eye bounding box. Another approach proposed by Vrânceanu et al. [4] for detecting EAC uses iris centre detection and facial landmark detection. They used an isophote curvature-based approach to locate the iris centre.

For gaze estimation, Zhang et al. [5] used a convolutional neural network. Using a CNN model, they integrated the data from the face pose estimator and the ocular region. In the output layer, they've trained a regression model.

In [6] Radlak et al. presented a method for gaze direction estimation method in static images. They used an ellipse detector with a Support Vector based verifier. The bounding box is obtained using the hybrid projection functions. Finally, the gaze direction is classified using Support Vector Machine (SVM) and random forests.

Haar Cascade classifiers are an effective way for object detection. This method was proposed by Paul Viola and Michael Jones in their paper [7].

III.OBJECTIVE

One of many obvious-once-it’s-pointed-out discoveries the NLP came up with was the famous, and sometimes infamous, NLP Eye Assessing Cues method. Used with skill this enables us to know how a person is thinking from watching how they move their eyes. Estimating the direction of one's gaze is an efficient way to learn one's Eye Accessing Cues. Most of the approaches for gaze estimation uses active infrared based methods which require expensive hardware. And estimation of gaze location from webcam often requires cumbersome calibration procedure. The proposed gaze direction classification is multi-class classification problem, avoiding the need for calibration. The eye directions obtained can be used to find the EAC and thereby infer the user’s cognitive process. The information obtained can be useful in the analysis of interrogation videos, human-computer interaction, information retrieval, effective communication etc.

The predicted output is displayed in the interface and stored externally with respective timestamps for further processing in future works.

IV.METHODOLOGY

We develop a real-time framework which can detect eye gaze direction using off-the-shelf, low-cost cameras in desktops and other smart devices. Estimation of gaze location from webcam often requires cumbersome calibration procedure. The suggested gaze direction classification is a multi-class proposed problem that does not require calibration. The acquired eye gaze directions can be utilised to locate the EAC and so deduce the user's cognitive process. The collected information is valuable in the analysis of questionnaire videos, interaction between people and computers, information retrievals and so on.

A. System Design

This application would be implemented as a computer program, with an interface that presents questions designed to stimulate the senses of sight, hearing, and emotion, as well as tasks demanding both memory and mental creation. The user answers while eye movements are monitored by the camera and appropriate output indicated thought process is displayed.

The application uses a computer webcam or an external camera attachment to capture live video feed of the subject. This live feed is then processed into individual image frames using an existing Computer Vision library called OpenCV as shown in figure 3. Using facial landmark localization and OpenCV the pupils of both eyes are detected from the frame as shown in figure 4. And then is fed into the eye gaze direction classifier which processes the data to predict the movement of the eye.

B. Eye Accessing Cues

Eye Accessing Cues in NLP are movements of the eyes in certain directions that suggest visual, auditory, or kinesthetic thinking. According to Neuro Linguistic Programming, people make eye movements that signal the representational system they are using. It is believed that we all go within and access information by eye movement, and people store knowledge in a certain way so that they can find it visually, auditorily, or kinesthetically. When we ask someone a question, we might have seen their eyes move. Processing these micro-facial expressions, we can tell which

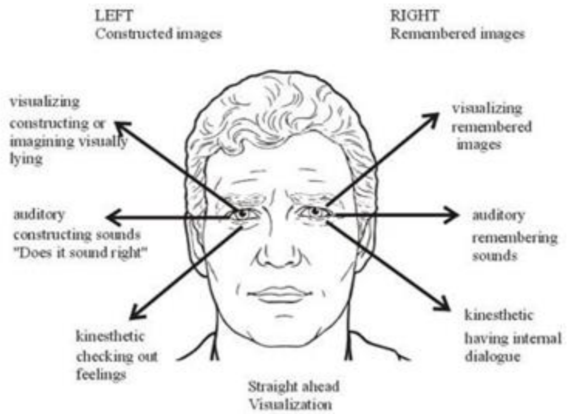

Fig. 1. Eye Positions as looking at another person representational system a person is using based on how their eyes move.

C.Desktop Application

In the desktop application, a user-friendly interface is designed which displays question meant to stimulate a cognitive response from the user. The camera module of the system picks monitors the movement and position of both the eyes. This process of detection is done by using the methods of facial landmark localization and eye gaze direction classification. After detection of the pupils in the eyes, the movement or the gaze direction is found out by calculating the distances of the detected pupil and its respective facial landmark.

D.Facial Landmark Estimation

Identifying important areas on a face, such as the tip of the nose and the centre of the eye, is known as facial landmark estimation. Depending on the amount of facial landmark points, several estimate models exist. The simplest model is the 5- point model, which detects simply the edges of each eye and the bottom of the nose.

However, we employ a 68-point landmark model in our suggested system for simplicity of future modification and a larger range of application.

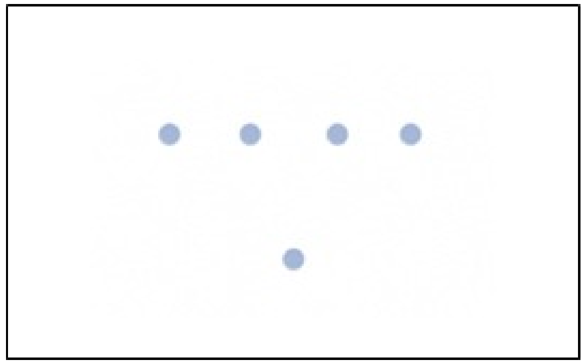

Fig. 2. 5-point Model

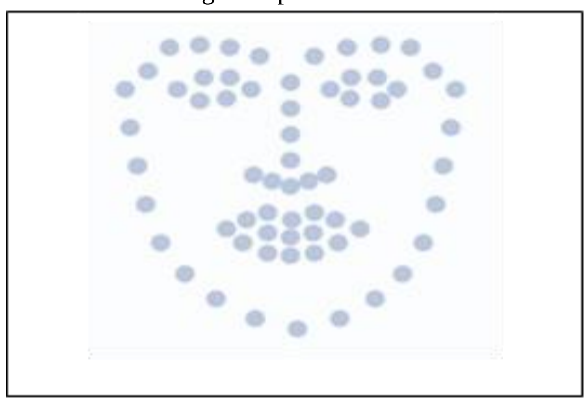

Fig. 3. 68-point model

E.Basic working of the proposed system

In this approach we have used a CNN model with three convolution layers. This stage is followed by a rectifier linear unit (ReLU) and a max pool layer. ReLU layer introduces a non-linearity to the activations and the max pooling is used for performs a spatial sub-sampling of each output images. Aft er the convolutional, ReLU, and max-pooling layers in the third convolutional layer, the outputs from all the activations are joined in a fully connected layer. The number of output nodes corresponds to the number of classes in the particular application. The softmax loss is used over classes as the error measure. Cross entropy loss is minimized in the training. Two CNN networks are trained independently for left and right eyes. The scores from both the networks are used to obtain the class labels. The class can be found out as the label with maximum probability.

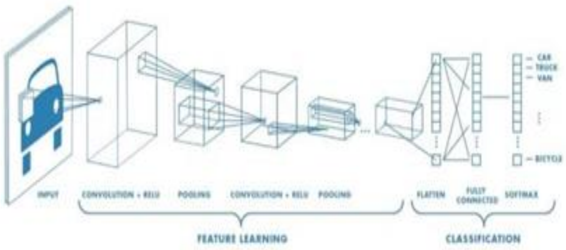

Fig. 4. Convolutional Neural Network Architecture

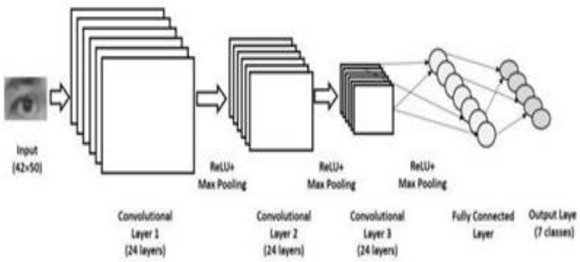

Fig. 5. Architecture of CNN used here

F. Schematic diagram of the proposed system

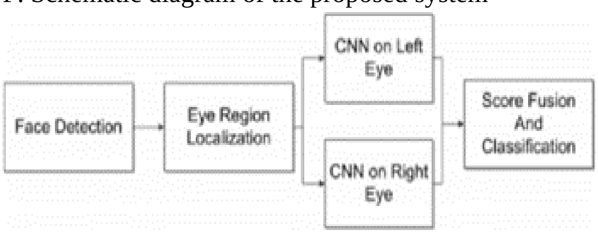

Fig. 6. Schematic Diagram

G. Training And Data Preprocessing

A dataset consisting of 6 subsets which contain images of positions of eye with respect to the sclera in different positions and orientations i.e. centre, left, right, top-right, top-left, bottom left, and bottom right. This is accessed from disk and trained on the model

The data is then pre-processed. The training data is shuffled randomly. This data is then resized to (244,244,3) as the input to the model has to be strictly in the above-mentioned dimensions. Normalization is performed on the data

H. Eye Gaze Direction Classification

MobileNetv2 is a class of CNN (open-sourced by Google) is used as the deep learning model. It significantly reduces the number of parameters when compared to the network with regular convolutions with the same depth in the nets.

Transfer Learning is a popular method where one model developed for a particular task is again used as the starting point to develop a model for another task i.e. a pre-trained model. We re-use MobileNetV2, freeze the base layers and then add and train the top few layers.

When working with a small dataset, we can take advantage of features learned by a model trained on a larger dataset in the same domain. This is done by instantiating the pre-trained model and adding a fully-connected classifier on top. The pre- trained model is 'frozen' and only the weights of the classifier get updated during training. This in turn helps us achieve better accuracy for our model, even if we have a small dataset and perform only fewer computations for training.

The model is compiled using Adam optimizer with suitable learning rate and a los function of sparse categorical cross entropy

I. Face Detetion And Eye Region Localization

Facial landmark detection is used for detecting the face and the eye. The facial and eye detected data is passed into the model for classification. The model returns labels according to the position of eye with respect to the landmark corresponding the edge of the eye.

V.RESULTS AND DISCUSSION

This project has been successfully completed as intended. The learning rate was changed to reduce the case of overfitting of the model on the training dataset. A small dataset can be used, if so data augmentation procedure can be applied for effective training. An accuracy of 94% was obtained after training the model. This project is implemented with a low- cost camera setup which is enabled to detect face and eye regions on the face. The prediction of the model is displayed as a label on the output screen. The program was able to track and monitor the shifting eye gaze and display the corresponding label. The outputs for the project are provided below. The project was completed in a span of 3-4 months where initial research took almost 1 month and hardware and software implementation took about 4 months to complete.

Fig. 7. Tracking eye and displaying label (right)

VI.CONCLUSION AND FUTURE WORK

In this work, a framework for real-time classification of behaviour anomalies through eye movements is obtained using EAC pattern from the Neuro Linguistic Programming Theory. This project has been successful in detecting various positions of the eye and determine a predicted output based on the EAC cue model followed. There is very minimal computer complexity involved and cost effective as we use low-cost camera modules in desktops and other smart devices.

Future work on this project can include incorporating other facial landmark and analyzing them through which expressions and various other facial cues are also factored, then combine the data from face pose estimator and eye region using a CNN model and train a regression model in the output layer.

REFERENCE

[1] A. George and A. Routray, "Real-time eye gaze direction classification using convolutional neural network," 2016 International Conference on Signal Processing and Communications (SPCOM), 2016, pp. 1-5.

[2] D. W. Hansen and Q. Ji, “In the eye of the beholder: A survey of models for eyes and gaze, ”Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 3, pp. 478–500, 2010.

[3] R. Vranceanu, C. Vertan, R. Condorovici, L. Florea, and C. Florea, “A fast method for detecting eye accessing cues used in neuro-linguistic programming,” in Intelligent Computer Communication and Processing(ICCP), 2011 IEEE International Conference on IEEE, 2011, pp.225–229.

[4] R. Vranceanu, C. Florea, L. Florea, and C. Vertan, “Automatic detection of gaze direction for nlp applications,” in Signals, Circuits and Systems (ISSCS), 2013 International Symposium on. IEEE, 2013, pp. 1–4.

[5] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, “Appearance-based gaze estimation in the wild,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2015.

[6] K. Radlak, M. Kawulok, B. Smolka, and N. Radlak, “Gaze direction estimation from static images,” in Multimedia Signal Processing (MMSP),2014 IEEE 16th International Workshop on. IEEE, 2014, pp. 1–4.

[7] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 2001.