International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT195269

Automatic Learner Focus Detection System Over A Live Feed

Poorani G.[1], Sudharsan V.[2], Moses Mano Raj J.[2]

[1]Assistant Professor, Department of Computer Science, Sri Krishna College of Technology Kovaipudur, Coimbatore, Tamil Nadu, India

[2]Department of Computer Science, Sri Krishna College of Technology Kovaipudur, Coimbatore, Tamil Nadu, India

ABSTRACT

The world is turning into a technological sphere. The technologies are evolving every second. With everything available in the Internet, people have started preferring to take their lectures and classes online. And so have the seminars turned into webinars, seminars over web. This system mainly concentrates on determining if a viewer attending the webinar is focused or not towards it. The main aim of this system is to recognize the eye gaze of a viewer over a live session with the help of some computer technologies and letting the presenter know if the viewer is focused on the session or not. By this, the presenter will know how well he/she is capable of keeping his viewers focused and he/she can work on his skills if the viewers are not focused. The access to the viewer's webcam is sufficient to achieve the goal.

Keywords : Webinars, Computer Technologies, Eye Gaze Prediction, Focus.

I. INTRODUCTION

Advancement in technology never stops. The world is fast turning into a place of great technological hub and the people are migrating towards more technologically advanced stuff. With everything available online , people prefer doing stuff at their own place and this also proves a great time saving method. Education and seminars are also now available online which are taken by a lot of people.

A webinar which can be said as a seminar over the web is a live meeting that takes place online. The webinar can be a discussion, presentation, demonstration, or an instructional session that is held over a web connection in a remote location. Participants use their computers and can view documents and applications , while audio sharing allows for presentation and discussion. People these days prefer webinars over the seminars as it saves them a lot of time and effort. Also, a large number of crowd can be addressed over a webinar than in a seminar.

Educational presentations are also given by professionals on topics related to their businesses and are able to connect with their audiences in a much closer and interactive way. It could be a webinar where a person simply hosts a seminar or lecture to teach something, it could be a promotional presentation for selling a product, or it could also be both. This can be done through a live video session where the presenter and the viewers are connected over the internet. The effectiveness of a webinar can be determined only by how the viewers are focused to the session and there are no effective means till now to determine the focus of a viewer in a webinar. This system aims in determining how focused a viewer is by using certain modern technologies.

Modern technologies have always been fascinating and one such technology is used in this system to achieve it's goal . Eye gaze prediction is one of the modern technologies that make the people go in awe. The prediction of the eye gaze is done by locating the Iris of the viewer through their webcam and reading the direction of gaze where the viewer see. This concept of eye gaze prediction is used to determine where the viewer focus is and can be a solution to determine the viewer focus detection.

II. LITERATURE SURVEY

In the previous research , boosted cascade of simple Haar-like feature to detect the eye which was initially proposed by Paul Viola and was improved by Rainer Lienhart. First, the classifiers were trained with thousands of positive and negative image (image include object and image non object) based on simple features (called Haar-like feature). There was a large of number of features in a sub-window of image 24x24 pixels, a number far larger than the number of pixels. The algorithm of training was AdaBoost which trains classifiers and in the same time select a small set of features which are the best separates the positive and negative examples. After training, only small set of features were selected and these features were combined to form an effective classifier. After the classifiers were trained, a cascade of classifiers was constructed to increase the detection performance. The cascades of classifiers were constructed to reject many of the negative sub-windows while detecting almost all possible instances. Simple classifiers were used to reject the majority of sub windows.

To track the eye movement the algorithm of Lucas Kanade was used. This algorithm was used to track objects in real time. It was used to find if two images are similar or not by constructing a pyramid of data points using certain formula and the pyramid data points was used to track the eye movement.

The gaze of the eye was then predicted using the Gaussian function. The properties of the function at a finite number of point inferences in the Gaussian process will gave them the same answer as taken all other infinitely points. With the help of that function the prediction points which are used to find the gaze point was made. Then the system was trained again from a finite number of values to all the possible values that could be obtained.

In this proposed system, firstly a video call connection is established between a presenter and a number of users using Kurento framework[1] and over the video stream that flows between the peer systems, we use the Fabian-Timm algorithm[2] to find the accurate location of the center of the eye by using the mean of the gradients . Then we use various algorithms to detect the head position of the viewer ,so that the eye can detected correctly even at different head positions [3][4][5][6].

After detecting the eye position , the gaze of the eye is detected and also the direction in which the viewer gaze is also determined[7][8]. This gaze direction is used to determine if the viewer stays focused or not through the session .

III. PROPOSED SYSTEM

Architectural Design:

The architecture of the proposed system consists of 2 modules:

1) The video conference
2) Processing of the live Stream

The video conference

The video conference module aims in connecting the presenters and the viewers over a video session. This is achieved by the use of Kurento Media Framework (KMF). Kurento is one of the Web RTC [9] media servers and it contains a set of client APIs that helps the making of the development of advanced video

applications for WWW and smart phone platforms. The features of Kurento media server include trans-coding, group communications, recording, broadcasting, mixing and routing of the audiovisual flows.

As a differential feature from other media servers, Kurento Media Server also provides some advanced media processing capabilities that involves computer vision(CV), augmented reality, video indexing and analysis of speech. Kurento modular architecture integrates with the third party media processing algorithms like sentiment analysis, speech recognition, face recognition simple, application developers used transparently like how that use the rest of Kurento built-in features.

Using the features of Kurento, the presenter and the viewers are connected with each other over group communication . The connection is done by creating a Web RTC endpoints at both the ends of the connection. Media pipelines are also created at the endpoints for the transfer of media between the connected peers . This sums up the first module of the system.

Processing of the live stream

The second module of the system involves processing of the streams that are generated while the communications of the peers take place. This processing of the media streams involves facial expressions of the viewers and produce results that say if a viewer is focused towards the session. This is achieved with the help of Computer Vision. Open CV[10][11] is used in this System to process the video streams and produce results. Open Source Computer Vision Library which is abbreviated as Open CV is an open source machine learning and computer vision software library. It is a library that has more than 2000 optimized algorithms, which includes a comprehensive set of both machine learning and classic and state-of-the-art computer vision algorithms. These algorithms can be used to identify objects, detect and recognize faces, classify human actions in videos, track moving objects, track camera movements, extract 3D models of objects, produce 3D point clouds from stereo cameras,etc. This system uses Open CV's ability to recognize eyes which are the objects of focus, to determine if the user stays focused.

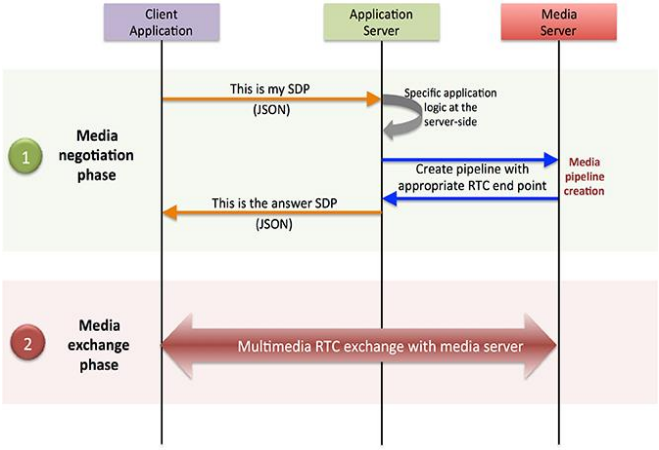

Fig 4.1 Illustration of data exchange in Kurento

IV. METHODOLOGY

The Video session:

Establishing the video connection between the presenter and the viewers is necessary for the transfer of media elements which is done with the help of Kurento Media Server. For the establishment of the connection ,the TURN server is to be set up which helps in the connection of the peer machines over a network .Once TURN server is set up ,the Kurento Media Server will be able to share the media elements between peers.

When a viewer joins a webinar room by entering the credentials of the webinar room, the process of registration and connection of the viewer with the presenter and other

viewers starts. For the registration to be successful and for the connection to be established between the peers, the SDP offers and the ICE candidates are exchanged between the peer systems .The SDP offer

holds the various possible communication methods that can be offered by a browser.

The connection of the peers is established over a common SDP offer that is offered by both the peers. Every end system in the network consists of a Web RTC endpoint which transfers these SDP offers to other system. Once the registration is done and the connection is established, media pipelines are created between the peer systems. The stream data are transferred through these media pipelines that serve as a media transfer connection between the connected peers. When a user leaves the room, the Web RTC connection between that user and the others is closed.

The Focus Detection:

The focus detection is the module where the main process takes place. The detection of a viewer focus is done with the help of a series of algorithms that are supported by Open CV.

The algorithms are used to implement eye detection, pupil detection and gaze direction detection. The video streams that are transferred between the peers are sent to the Open CV focus detector which makes use of the various algorithms in it to process the video stream. The video stream is processed and the eyes, pupil and the gaze directions are predicted. The gaze direction that is predicted is the main component of determining the viewer focus. Based on the direction of the viewer gaze, the Open CV determines if the viewer sees the video display on his screen or not. When the Viewer's gaze direction is out of the video display for a specific duration of time, the presenter is notified that the particular viewer has started to lose focus on the webinar.

Fig 4.2 Gaze Direction prediction using OpenCV

V. FUTURE ENHANCEMENT

The paper provides a solution to determine the focus of a viewer over a live feed by using various algorithms so that the presenter would know if he is good enough to keep the viewers interested. This project involves only the analysis of the viewer data over a video format and the accuracy of this lies between 70% - 80% .The future plans of enhancement are to improve the accuracy of the system as well add other features to the webinar room like chat between the peers, screen sharing, etc.

VI. REFERENCES

[1]. http://www.kurento.org/documentation

[2]. Fabian Timm and Erhardt Barth - “Accurate Eye Centre Localisation by means of Gradients”

[3]. Roberto Valenti, Nicu Sebe , and Theo Gevers - “Combining Head Pose and Eye Location Information for Gaze Estimation”

[4]. Xuehan Xiong ,Fernando De la Torre - “Supervised Descent Method and its Applications to Face Alignment”

[5]. Hiyam Hatem , Zou Beiji , Raed Majeed , Jumana Waleed Mohammed Lutf - “Head Pose Estimation Based On Detecting Facial Features “

[6]. Oliver Jesorsky, Klaus J. Kirchberg, and Robert W. Frischholz - “Robust Face Detection”

[7]. Tom Heyman , Vincent Spruyt , Alessandro Ledda - 3D Face “Tracking and Gaze Estimation Using a Monocular Camera”
[8]. Luke Allen and Adam Jensen - “Webcam-based Gaze Estimation”
[9]. Dan Ristic - “Learning Web RTC ”
[10]. Adrian Kaehler and Gary Bradski-”Learning Open CV”
[11]. https://docs.Open CV.org/2.4.9/

Cite this article as :

Poorani G., Sudharsan V., Moses Mano Raj J., "Automatic Learner Focus Detection System Over A Live Feed", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 2, pp. 325-329, March-April 2019.
Available at doi :
https://doi.org/10.32628/CSEIT195269
Journal URL : http://ijsrcseit.com/CSEIT195269