REAL-TIME E-LEARNING PLATFORM

Final Report for Lecturer Tracking Module

Project ID: 19-079

W.A Geeth Sameera
IT16119086

Supervisor – Dr. Malitha Wijesundara

Bachelor of Science (Hons) in Information Technology Specialized in Software Engineering

Department of Software Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

16th September 2019

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. Also, I hereby grant to Sri Lanka Institute of Information Technology the non-exclusive right to reproduce and distribute my dissertation in whole or part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as article or books).

Signature: Date: 16th of September 2019

Signature of the Supervisor: Date: 16th of September 2019

ABSTRACT

Real-time e-learning platform is a WebRTC based live teaching platform which facilitates the e-learning requirements of universities and other institutes. This solution includes lecture live streaming, lecture playback, a vector based interactive whiteboard, chatting and file sharing module and a real-time lecture movement tracking module using a PTZ camera. The system is capable of streaming two simultaneous streams of a 1080p camera and a 720p screen capture seamlessly using a network connection with 256KB/s bandwidth. Lecture playback supports camera video playback in 1080p, 720p and 360p and screen playback in 720p and 360p. This player is integrated with MPEG-DSH technology to support adaptive bitrate streaming. Clients can use this player with an internet connection with minimum of 128 KB/s bandwidth. With these features, this solution is a very cost effective and reliable product compared to existing competitors in the market.

ACKNOWLEDGEMENT

First, I would like to express my gratitude to supervisor Dr. Malitha Wijesundara and co-supervisor Mr. Pramadi Athapaththu for the guidance and mentoring given throughout this period to make this research a success. Also I would like to thank senior lecturer and lecture in-charge for Comprehensive Design and Analysis Projects module, Mr. Jayantha Amararachchi for the valuable assistance given to complete this project.

At last but not least I am also thankful to all my family members and colleagues who have been always helping and encouraging me throughout the research period. I have no valuable words to express my thanks, but my heart is still full of the favors received from every person.

TABLE OF CONTENT

DECLARATION i
ABSTRACT ii
ACKNOWLEDGEMENT iii
TABLE OF CONTENT iv
LIST OF FIGURES v
LIST OF ABBREVIATIONS v
1. INTRODUCTION 1
1.1. Background Literature 1
1.2. Research Gap 2
1.3. Research problem 2
1.4. Research objectives 3
2. METHODOLOGY 4
2.1. Implementation 7
2.2. Commercialization aspects 8
2.2.1. Cloud based revenue model 8
2.2.2. On premises revenue model 8
3. TESTING AND RESULTS 9
4. CONCLUSION 10
5. REFERENCES 11

LIST OF FIGURES

Figure 2.1: High Level Architectural Diagram 4
Figure 2.1.1 Tracking Sample 1 7
Figure 2.1.2 Tracking sample 2 7
Figure 2.2.1 CPU Performance of the tracking algorithm 9
Figure 2.2.2 GPU Performance of the tracking algorithm 9

LIST OF ABBREVIATIONS

PTZ	Pan Tilt Zoom
CPU	Central Processing Unit
GPU	Graphical Processing Unit
Yolo3	You Look Only Once Version 3
SORT	Simple Online and Real-Time Tracking
Deep SORT	Simple Online and Real-Time Tracking with a Deep Association Metric
FPS	Frames Per Second
PyTorch	An open source machine learning library based on the Torch library
OpenCV	Open source computer vision. A library of programming functions mainly aimed at real-time computer vision

1. INTRODUCTION

Real-time e-Learning platform is a live lecture streaming web application which is developed focusing on Universities and other educational institutes. Lecturers can stream lectures from or out of the institute premises remotely while students can join relevant live streams from anywhere. In traditional video conferencing and streaming, video of a moving person is acquired using a human cameraman or a fixed camera. Instead of that PTZ (Pan Tilt Zoom) Camera is used to acquire the video. Camera movements are controlled by an advanced algorithm which keeps track of the lecturer and trying to keep him in the center of every fame. This document will explain the implementation of lecturer tracking component.

Lectures streamed from institute premises will be equipped with a PTZ camera to get a third person perspective camera feed of the lecturer. Lecturers can use their web camera instead of PTZ camera if they are streaming remotely outside of the institute premises. Apart from the camera feed, Lecturer's Computer screen is captured and streamed. Lecturer can select between which screen, which application to share. Also audio feed from a mic attached to the lecturer's PC is broadcasted. The application is capable of handling multiple simultaneous lecture sessions without any interruptions to each other. This component is also responsible for recording live streams to facilitate the lecture playback where students are given the capability to rewind and playback already streamed lectures for future reference.

1.1. Background Literature

Virtual education is an emerging concept around the world. Students are starting to adapt to learn more productively through internet than traditional classroom due to many reasons. One of the main reasons is that every student has a different pace of catching-up with the teaching where some of them can be considered as fast learners while others may be slow. In traditional classroom scenarios, there is no solution to this issue. Another reason is the schedule flexibility. Through virtual learning, students can access their courses at any time anywhere with internet giving students the full control of their schedule of the day. Supporting the facts, universities provide students with the access to course materials via internet. Going beyond this, Real-time e-Learning Platform is a solution for the shortcomings in conventional education system mainly with universities. [1][2]

This proposed system is a web-based application where students can watch live lectures and interact with the lecturer and also playback previous lectures online. A tracking camera will be tracking the lecturer movements while recording and both lecturer's video feed and the lecturer's computer screen feed will be streamed. Also this system contains features to interact with the lecture. Student can ask questions from the lecturer remotely via a webcam in real-time and also can use a virtual whiteboard that is provided with the system to describe the question.

1.2. Research Gap

Compared to existing competitors of this field, most of them uses a fixed camera/cameras to take the feed of the lecturer moving in the lecture hall. If it's one camera, then lecturer may not be visible in some parts of the lecture hall based on the positioning of the camera. To address that issue, cameras with wide view is used to covers the entire lecture hall. Then the students can't see the lecturer properly. In a multi camera setups, some program is used to switch camera feeds according to the lecturer position. That will add extra cost without improving the student experience significantly. Both these approaches have major impact on the student experience on using an e-learning platform.

1.3. Research problem

The most common teaching and learning practice adopted by many enterprises has always been a classroom with one or more instructors and learners meeting physically and in real-time. But in this teaching model, there are several drawbacks which will be addressed as problems within this research.

A classroom based learning experience means the class schedule is predetermined and not subject to change. Students must shape their personal schedules around school instead of the other way around. If plans unexpectedly change or an emergency comes up, the student cannot adjust the class schedule to turn in the work at a different time. This is one of the main problems that this research will find solutions for.

Content non-reusability is another problem that can be found in classroom learning. Memorizing or writing all the necessary content while listening to a lecture is difficult. Therefore, students might miss many important points that the lecturer is pointing out during a lecture.

1.4. Research objectives

1. **Real-time lecturer tracking**

   Lecturer must be detected and tracked accurately in real time to generate camera control commands. Algorithm should perform efficiently even in low light environments. False detections must be reduced to improve the accuracy of the algorithm

2. **Handling multiple person detections (Person Re-Identification)**

   If lecturer take a student to the stage or if someone passes across the lecture hall, multiple detections will be there. Since the camera only moves according to the lecturer, algorithm must be able to quickly differentiate between lecturer and the other detections and generate commands bases on that detections throughout the session

3. **Camera movements control**

   Based on the detections, a PTZ camera has to be moved smoothly and accurately keeping lecturer in the center of the video output. If the detection is failed at some point, camera control module should have a predefined way to find the lecturer again.

2. METHODOLOGY

In order to develop this real-time application which is usable from multiple devices as laptops, tablets and smartphones we have chosen WebRTC technology. WebRTC is an open source project developed by google to make web browsers support peer-to-peer communications. On top of that, to implement our solution, we used Janus as the media server. Janus is a general purpose WebRTC server which is capable of implementing WebRTC media communications with browsers [3]. Backend application server is developed with Node.JS while frontend application is implemented using React.JS. MPEG-DASH technology is adhered in development of the playback player in order to support adaptive bitrate streaming [4] [5].

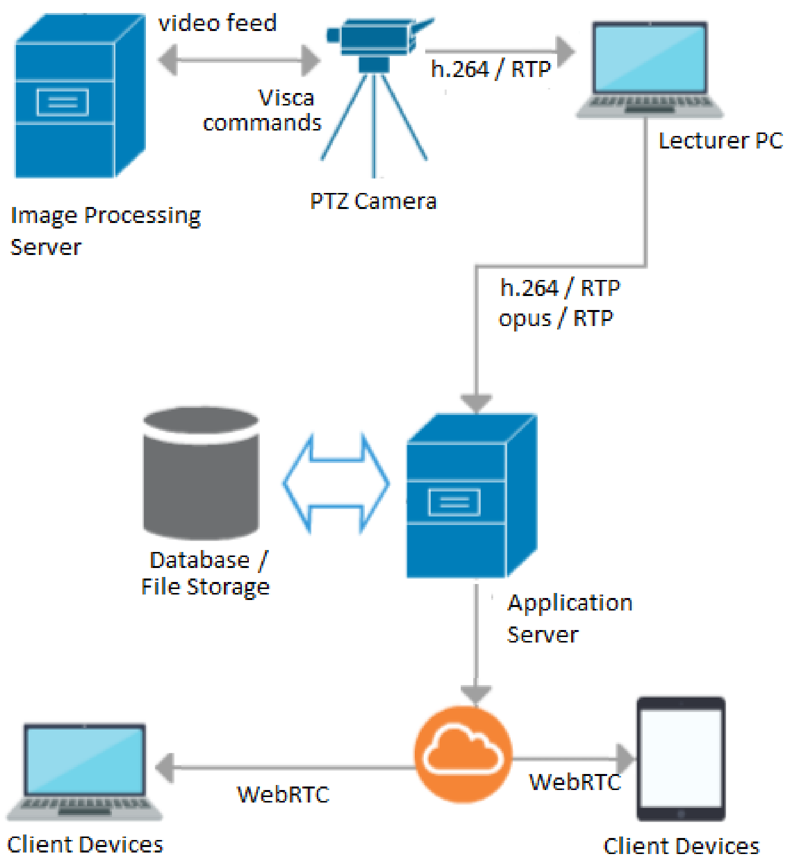

Figure 2.1: High Level Architectural Diagram

Main objective of lecturer tracking component is to automate a special camera (PTZ) using an advance algorithm to do lecturer tracking process throughout the lecture. Using the tracker, set of commands will be generated to control a pan tilt zoom (PTZ) camera. Output video will then be streamed as the lecturer's video alongside with lecturer's pc video. Tracking algorithm consists of mainly three components

1. Person Detector and Tracker
2. Person Re Identification Module
3. Camera Controller

1 Person Detector and Tracker

Person detector is developed using GPU implementation of the Yolo version 3. Yolo is an extremely fast real time object detection model. It gives the best balance between the accuracy and the speed. Yolo stands for "You look only once". It applies one neural network to entire frame/image and come up with bounding boxes with the probabilities for each of them. Logistic regression is used to come up with the probabilities for detections and each bounding box is weighted by the associated probabilities. openCV is used to preprocess every frame before passing in to the neural network. Deep neural network of openCV does not support nvidia GPU so the neural network module of Pytorch is used to implement yolo v3 in a nvidia GPU [4][5][6]. Yolo is a general purpose object detector with 80 classes but the tracking algorithm only requires to detect person class. So the other classes are filtered out and bounding boxes are only drawn for person detections.

After drawing bounding boxes in the frame it will be used to determine the movements of the detections. Before determining the movements, detections are filtered out again using re-identification module leaving only the lecturer's bounding box in the frame. Position of that bounding box is kept as a reference to following frames. By analyzing 10 such frames, movement of the bounding box is determined with the direction of the movement. 10 frames are used to minimize the noise occurred due to small movements of the bounding box. Those measures will be used by the camera control module to generate control commands to the camera.

2 Person Re-Identification Module

This module will handle the multiple person detections and keeps tracking only the lecturer. Deep sort, an extension of sort algorithm is used as a base to identify multiple detections and to calculate the similarity scores of them [8]. These scores will be generated using a set of features extracted from the bounding box. At the start of the lecture, only the lecturer must be there in the camera view. His object id will be 1. After starting the lecture, tracking algorithm will only consider the movements of bounding box with object id 1. If the detection is lost at some stage of the tracking process, then the camera will first zoom out to see if the lecturer is visible in the full zoom out view. If lecturer is unavailable in full zoom out view, then camera return to a predefined home position where the lecturer will re appear again. Then the same tracking process will continue from the beginning by issuing object id as 1 to the lecturer again.

3 Camera Controller

This module will control the PTZ camera based on the detections produced by above two components. Tracking module will store the centroid coordinates of bounding box in a numpy array. Based on the last 10 coordinates, movements and the direction of the movements are generated using the tracking module. Based on those data, camera control commands are generated and send to the camera. VISCA protocol is used to send control signals to the camera over a socket connection. Camera is connected to the image processing server via an Ethernet cable for sending control signals. USB connection from the same camera is used to obtain the video feed of the camera.

If the bounding box is stable for a pre-defined time without any movement, camera zoom function is triggered and the bounding box is zoomed in. when he starts moving again camera will return to initial zoom and keep tracking the lecturer again. Camera will be moved in all x(pan), y(tilt) and z(zoom) axis. Speed of the camera controls will be dynamic according to lecturer movements.

2.1. Implementation

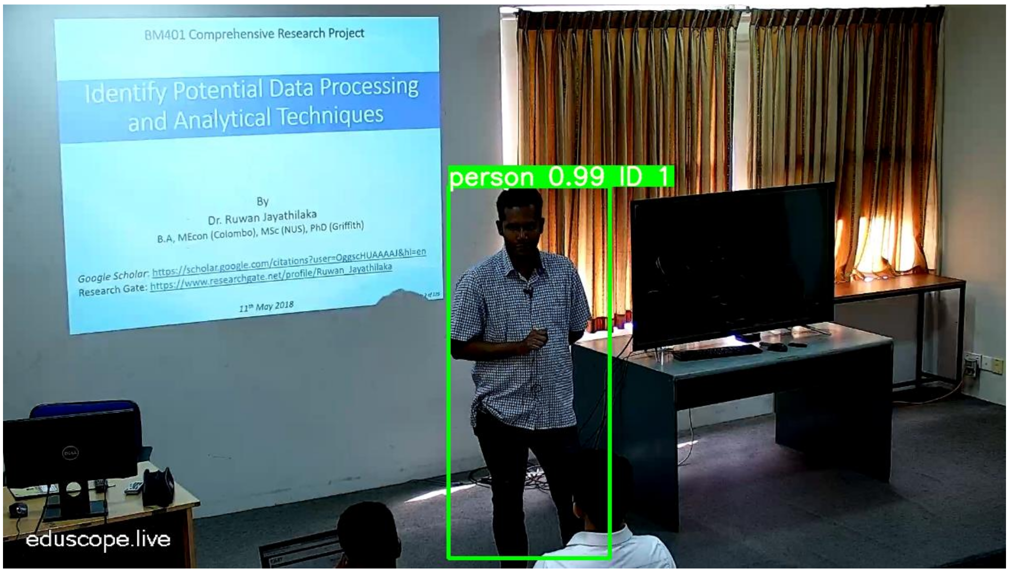

Figure 2.1.1 Tracking Sample 1

Figure 2.1.2 Tracking sample 2

2.2. Commercialization aspects

Key stakeholders of this system are:

- University / Educational Institutes / Schools
- Lecturers / Teachers
- Students

Our main goal of the product in the aspect of commercialization is reduce the cost for the usage of the product than the competitors in the current market.

Real-time e-learning platform is commercialized based on two business revenue models.

- Cloud based revenue model
- On premises revenue model

2.2.1. Cloud based revenue model

This revenue model is used if the client agrees to install their application setup in an amazon cloud environment. Amazon will charge for monthly resource usage. Clients are charged with an additional service charge which will be a multiple of the entire usage bill from cloud service.

Total usage bill = Amazon cloud bill x billing factor

2.2.2. On premises revenue model

If the client does not want to use a 3rd party service to host their application and wants to setup the system on premises, this revenue model will be used. Here client will be billed monthly based on the user accounts created in the system.

3. TESTING AND RESULTS

Tracking algorithm was tested on a core i5 3.7 GHz 8th generation CPU using 8GB Ram and in a GPU (NVidia GTX 1050) with the same CPU configuration. CPU performance was around 6-8 FPS. In GPU, FPS was around 20. GPU performance is sufficient to run the algorithm in real time.

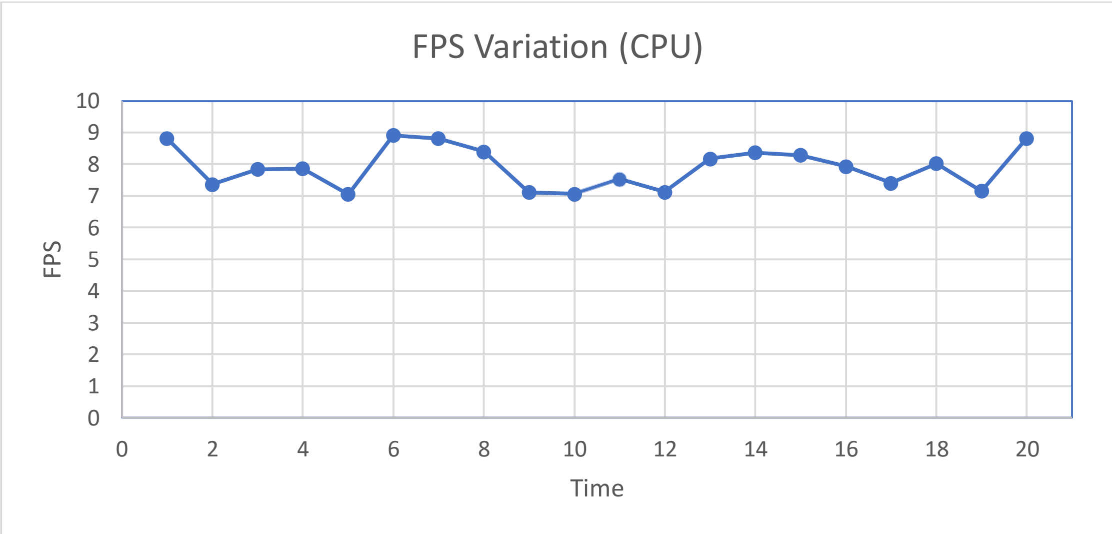

Figure 2.2.1 CPU Performance of the tracking algorithm

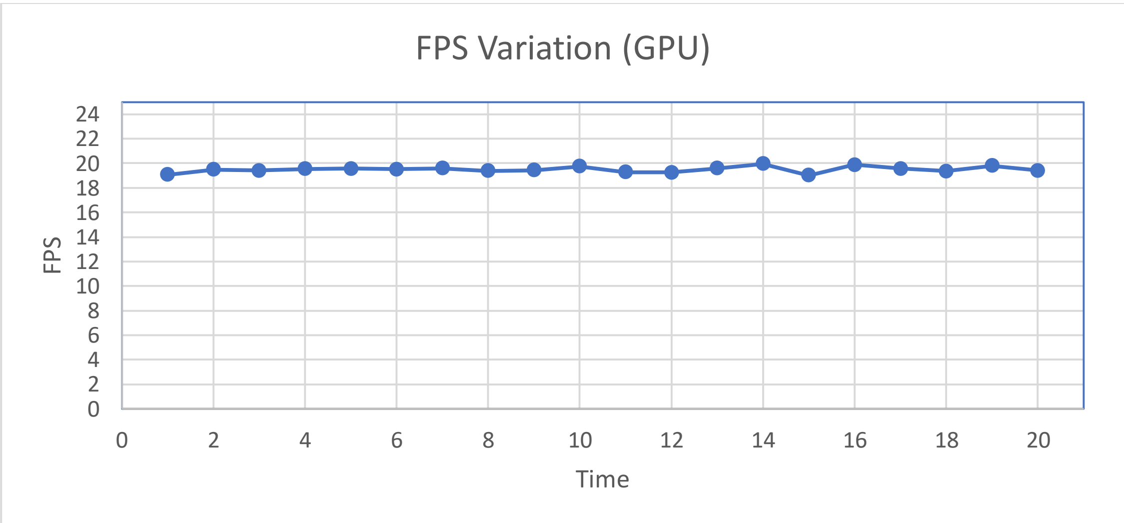

Figure 2.2.2 GPU Performance of the tracking algorithm

4. CONCLUSION

Using automated PTZ camera improves student experience on e-learning platforms. yolo3 can continue tracking even in low light conditions with high accuracy. Deep sort and yolo3 both implemented using pytorch and it works well with a frame rate around 20. Camera controller smoothly drives the PTZ camera to obtain a good video feed of the lecturer.

Looking ahead, even though yolo3 works well in this scenario it is a general purpose object detector with a variety of 80 classes. Well optimized and specific purpose detection model can be trained using the previously recorded videos of the lecturer sessions by annotating frames as the image dataset to the model.

Yolo3 with deep sort requires a GPU to get higher FPS value. With the advancement of deep learning and the computing devices, if this algorithm can be more optimized in future versions, so that it can fit into a device like a raspberry pie, it will reduce the overall cost by tremendous amount.

5. REFERENCES

[1] Chris Curren, "*Strategies for e-Learning in Universities*", CSHE.7.04,2004

[2] H. M. Truong, "Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities," Computers in Human Behavior, 2015

[3] Meetecho s.r.l, "Meetecho Janus Recordings", https://janus.conf.meetecho.com/docs/recordings.html [Accessed: 2019-08-06]

[4] Rosebrock, A. (2019). YOLO object detection with OpenCV - PyImageSearch. [online] PyImageSearch. Available at: https://www.pyimagesearch.com/2018/11/12/yolo-object-detection-with-opencv/ [Accessed 6 Aug. 2019].

[5] Medium. (2019). Real-time Object Detection with YOLO, YOLOv2 and now YOLOv3. [online] Available at: https://medium.com/@jonathan_hui/real-time-object-detection-with-yolo-yolov2-28b1b93e2088 [Accessed 6 Aug. 2019].

[6] Redmon, J. (2019). YOLO: Real-Time Object Detection. [online] Pjreddie.com. Available at: https://pjreddie.com/darknet/yolo/ [Accessed 6 Aug. 2019].

[7] En.wikipedia.org. (2019). VISCA Protocol. [online] Available at: https://en.wikipedia.org/wiki/VISCA_Protocol [Accessed 6 Aug. 2019].

[8] Medium. (2019). People Tracking using Deep Learning. [online] Available at: https://towardsdatascience.com/people-tracking-using-deep-learning-5c90d43774be [Accessed 16 Aug. 2019].