LECTURE CAPTURING SYSTEM: A CASE STUDY

Valuka Rukmali Wijayagunwardene

(IT15081094)

BSc. Honors in Information Technology (Software Engineering)

Sri Lanka Institute of Information Technology

Sri Lanka

11th October 2018

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgment 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 acknowledgment is made in the text.

Also, I hereby grant to Sri Lanka Institute of Information Technology the nonexclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium.

I retain the right to use this content in whole or part in future works.

Name: Wijayagunawardene V.R Date: .....

The above candidate has carried out research for the B.Sc. Dissertation under my Supervision.

Signature of the supervisor: ..... Date: .....

Abstract

Today, with the fast growth of technology and usage of internet services, e-learning has become one of the latest trends in the education sector. As a result, students tend to prefer e-learning than being physically present in a lecture due to multiple reasons. This paper examines an innovative approach to enhance the current e-learning procedures, particularly in universities. The "Lecture Capturing System" is a cloud-based web application which uses enhanced techniques to provide an interactive e-learning experience to users of the system. It uses a facial recognition-based authentication process to allow remote users to login to the system. A Pan-Tilt-Zoom (PTZ) IP camera captures and tracks the lecturer during the lecture session and this is streamed live to remotely logged-in students. The lecturer can also share the computer screen if required. The camera intelligently identifies specific gestures performed by the lecturer to rotate with the aid of gesture analyzing algorithms. Attendance of remote online students is marked automatically during a live-streaming lecture by using multiple facial recognition processes executing on the server. Offline recording of lectures is also supported after which the video is split into a series of chapters/thumbnails and the audio is converted to text; each chapter representing a presentation slide and the relevant text. Bandwidth and quota are managed intelligently to ensure the best possible transmission rate with minimum data consumption in order to avoid filling the link to capacity which would result in network congestion and poor performance of the network. This system is revolutionary and is capable of taking e-learning to the next level as it provides a complete classroom experience and much more to the remote users. It also has the ability to support multiple enterprise customers.

Keywords—PTZ camera control, gesture detection, biometric authentication and attendance, video thumbnails creation, bandwidth and quota management

ACKNOWLEDGEMENT

We would like to acknowledge and pay our esteemed gratitude to the following people for being with us, advising and encouraging us to complete this project with success.

Our supervisor Dr. Malitha Wijesundara, for his greatest dedication and excellent guidance. External Supervisor, Mr. Pramadhi Athapaththu, for his commitment and support, and the Project coordinator, Mr. Jayantha Amararachchi, who provided lecture materials and motivating us all the time under his guidance.

All the members of the team for being hard working throughout the development of “Lecture Capturing System” and for all the people who helped us to complete this project successfully.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

The following gives an introduction to Lecture Capturing System under the topics of Problem to be Addressed, Background Context, Research Gap, and Research Questions.

1.1 Problem to Be Addressed

Today, e-learning platforms are used for knowledge transfer through electronic media. This transfer can address several learning contexts, ranging from conventional classroom delivery to online and offline distance learning tactics.

There are a countless number of e-learning platforms which cover various aspects of learning such as video streaming, capturing the audience in the lecture hall if necessary, and screen sharing. But, the solution we propose is to handle advanced and enhanced features. Biometric recognition of participants to ensure authentication is proposed so that the attendance can be recorded. In addition to this, every lecture will be maintained as an mp4 video with the set of slides used, and the lecturer's voice relevant to each slide, which provides an easy way of reference to the students who missed a particular lecture. Another important aspect considered in this research is the way of interaction between the lecturer and the students who have any doubts to be cleared during the lecture. For this purpose, when the students who are physically present in the lecture hall are concerned, a gesture based system is proposed. Here, when the lecturer notices a specific student with the gesture of asking a question, the lecturer will have to perform a gesture for the camera to turn towards the audience and focus on the specific student who has the doubt by once again detecting the gesture performed by the student. When a remotely logged in user has a question for the lecturer, he/she has to signal the lecturer using a specific command, and then the lecturer has to decide whether to give video and/or audio control over to the specific user. As real time video streaming consumes a lot of quota and bandwidth, the system has to intelligently manage data usage by ensuring the best possible transmission rate with minimum data consumption.

1.2 Background Context

Today, communicating with internet and doing things right from our home becoming more and more practical because it makes day today life very easier. People always try to find a way to use the easiest way to complete their tasks without taking too much time and energy.

The art of learning should evolve over time. Even today with the technology we have, we are still used to physically going to a lecture and listening to the lecturer in real-time and making notes of his/her teachings, what the lecturer sketches on whiteboards or his/her presentation. So, there is a higher chance that students forget what the lecturer said or taught the very next day if they were not properly documented. What if a student gets sick or due to unavoidable circumstances, misses a lecture, how is he/she going to learn the missed session? Something like a live streaming with recording sessions of a lecture would help all the students even if they were present in the lecture itself. The ability to see what they have missed if the student comes to the lecture after it has started, a way to refresh their memories before attending the next lecture would result in a huge academic improvement.

E-Learning helps learning become possible to anyone who wants to learn something right from their home. Main purpose of this system is to give a better effective way to help the students to get learning materials and information from anywhere and to quickly recap any forgotten or absent lectures via the earlier recordings of the sessions.

A system and method for an interactive, Internet-based video conferencing multicast operation which uses a video production studio with a live instructor giving lectures in real-time to multiple participating students. The video conference multicasting permits the students to interact with the instructor and other installations during the course of the lecture and to later browse the recorded session without a hassle.

There are many added such as sharing lecturer's computer screen with students in real time, lecture communicating with students either by voice, video or both on request of the student. In our system we have added smart features such as intelligent bandwidth management to make sure that we use the least possible data bandwidth to transfer the videos to the students.

Web Application development skills, Image processing, Machine learning and techniques are essential. In addition, some basic knowledge on lecturing process and student behavior process are also required throughout the development of the system.

1.3 Research gap

As mentioned above, there are several popular e-learning platforms available currently. To get a clear idea on the research gap we are trying to address with our solution, let us compare the existing systems with our product.

Table 1: Comparision of features

Features	Open Broadcaster Software	Panopto	Kaltura	The Smarter Video Platform	Lecture Capturing System
Support one to many users to concurrently access a Live stream.	✓	✓	✓	✓	✓

Gesture based system to detect the speaker, pan, tilt and zoom.					✓
Thumbnail creation of lectures and voice to text conversion					✓
Screen sharing window for share laptop screen	✓			✓	✓
handle user management, course management, bandwidth and quota management	✓	✓	✓	✓	✓

biometric facial recognition for authenticate users					✓
capture the audience if required and focus on a guest speaker			✓	✓	✓
Intelligently manage data usage for minimum data usage					✓

1.4 Research Questions

Lecture Capturing System is not another web based e-learning applications. It contains unique features which addresses the questions that are there in current systems and uniquely bundle together a well performing e-learning system. Below are some of the main problems that lead to the development of Lecture Capturing System.

1.4.1 To a student who is enrolled in a lecture

- No way of knowing what has happened on a day if he/she unable to attend the lecture physically.
- If the lecture room is too crowded not being able to see the whiteboard or the lecturer clearly.
- No way to revive what was the last lesson the lecturer did if they have not taken down notes.
- No way to view what lecturer did using his computer later on. (eg: typing a partial code, annotating a PowerPoint slide)
- Not being able to interact with the lecturer to solve a question if the lecture room is crowded.

1.4.2 To a lecturer who is teaching a course

- Unable to do a lecture from home and upload it to a server so that all the students can watch it.
- Tracking the real attendance of the students is not fully accurate as friends of the students mark the attendance for absent people.
- Sometimes sharing computer screen reveals all the files in the computer as well as open programs and browser tabs. There is a privacy issue.

2 LITERATURE AND METHODOLOGY

2.1 Addressing the Literature

In recent years, there have been number of research efforts done to address the needs a smart e-learning management system. Below are some of the software functionalities and technologies that has been done prior to our research. Undertaking a Literature Survey helps us on finding and come up with the followings.

Regardless of the enormous growth of e-learning (electronic learning) in education and its perceived benefits, the efficiency of such e-learning systems will not be fully utilized if the students are not inclined to accept and use the system.

As a result, successful implementation of e-learning tools depends on whether the students are willing to adopt and accept the technology. Thus, it has become imperative for e-learning system developers to understand the factors affecting the user acceptance of web-based learning systems in order to enrich the students' learning experience and to create a better product to fulfill the necessary requirements of the student.

"Use of E-Learning", a research was done to find University students' purpose to use e-learning [1]. In this research, Teknologi Malaysia University's students try to apply and use the theory of technology acceptance model (TAM). They have employed structural equation modeling (SEM) approach with a SmartPLS software to investigate students' adoption process. Discoveries indicate that the content of e-learning and self-efficacy have a positive impact and substantially associated with perceived usefulness and student satisfaction, which impact university students' purpose to use e-learning. Although e-learning has expanded acceptance in universities around the world, the study of the intention to use e-learning is still essentially unexplored in Malaysia. The developed model is employed to explain the university student's intention to use e-learning. The study concludes that university students in Malaysia have positive perceptions towards e-learning and intend to practice it for educational purposes.

E-Learning is reflected as an innovative approach to education delivery via electronic forms of information. Multiple researches have been done to find the best way to use the technology and to better fit the students' necessities [1], [2]. The main obstacles that need to be addressed are the insufficient financial support, inadequate training programs, lack of ICT infrastructure, equivocal policies and objectives, and lack of awareness, interest, and motivation toward e-learning

technology are considered as the main obstacles to enhance e-learning in Iraqi universities. The lack of training programs and inadequate ICT infrastructure are considered as the key issues which obstruct advancing of the e-learning process in Iraq [2].
Online body tracking by a PTZ camera has been done before to automatically track a single person and focus on that person [3], [4]. Online human body tracking method by an IP PTZ camera based on fuzzy-feature scoring was done. At every frame, candidate targets are detected by extracting moving targets using optical flow, a sampling, and appearance. The target is determined among samples using a fuzzy classifier. Results show that the system has a good target detection precision (> 88%), and the target is almost always localized within 1/4th of the image diagonal from the image center [3]. Autonomous lecture recording with a PTZ camera [4]. This reaches the same viewing experience while watching lectures recorded by an automated system. To accomplish this, they have developed an automatic cameraman (PTZ camera-unit) that is able to, Detect and track a single person/lecturer, Change between different types of shots, Listen to high-level instructions from a virtual or human director, and Take cinematographic rules into account
By tracking the lecturer, he is framed well in the picture at any moment and viewers can't be distracted. Takes cinematographic rules into account, which ensures that the viewer remains focused and the viewing experience is aesthetically more interesting. The action axis is determined by calculating the direction of movement and the gaze orientation. A PID control loop ensures smooth movement of the camera. Because of the speed of the algorithm, it will be easy to downscale for embedded hardware and still perform the calculations real-time.
Remote controlling of the PTZ camera system for lecture rooms [5]. This consist of a simple and inexpensive software solution for remote management of PTZ camera systems. This provides the ability for users to remotely control the PTZ camera system from one place with the simultaneous image capturing ability. Users of this application are able to choose a number of presets and this functionality is provided programmatically by sending queries to CCTV system, which then responds back to the controller. All of the responses are processed immediately into the desired form and stored in a selected row in SQLite database. This method of data storage doesn´t require the installation of SQL database, which makes the solution easier to apply. The program was created using the programming language C#. But this software solution does not support real-time tracking of a person, just several predefined presets so that feature can be improved to real-time operation in Lecture Capturing System. OpenTrack - Automated Camera Control for Lecture

Recordings [6] records lecture sessions automatically without the need for a human camera person. A Tabletop Lecture Recording System [7]. This research presents a lecture recording system that employs gestures and digital cameras to facilitate remote distance teaching. Virtual Cameraman [8] uses two PTZ cameras having different utilities. One is named full-shot PTZ camera and the other is movement PTZ camera. To get camera movement information, the system first obtains continuous images from full-shot PTZ camera and theSpn extracts four fuzzified movement features which can represent four characters of audiences' motions respectively. On the other hand, an automatic camera movement model (ACMM) is constructed by recording photographers' habit of CM styles and shot types. The proposed system can select suitable CM styles and shot types by inputting the fuzzified motion features into the ACMM. After that, the system chooses the main target in the input frames obtained from the full-shot PTZ camera by using five aesthetic criteria. Finally, the system operates the movement PTZ camera to finish recording. Real-time tracking of a non-rigid target with a moving pan-tilt-zoom (PTZ) camera. The tracking of the object and control of the camera is handled by one computer in real time. The main contribution of the paper [9] is method for target representation, localization and detection, which takes into account both foreground and background properties, and is more discriminative than the common color histogram based back-projection. A Bayesian hypothesis test is used to decide whether each pixel is occupied by the target or not. The target representation is suitable for use with a Continuously Adaptive Mean Shift (CAMSHIFT) tracker. Experiments show that this leads to a tracking system that is efficient and accurate enough to guide a PTZ camera to follow a moving target in real time, despite the presence of background clutter and partial occlusion.

Human tracking applications with a Global Positioning System (GPS) can get the user's position in real time on the open area. But if the user goes into the room or building, then the human tracking application cannot get the user's position as the GPS signal cannot get through the wall. By using RFID, human tracking application can perform tracking in a certain closed area by providing room's or building's information that is entered. The IP camera as part of the system will send the images as visualization inside the room. In this research, human tracking system is built on a specific area that has lots of room or building such as a theme park or sports club. System is designed using an RFID reader, RFID tags, IP Camera, the database server and Android smartphones. The research was done inside Syahdan Campus, Bina Nusantara University – Jakarta.

The result shows that the application is working with 100% tapping and mapping accuracy [10].

IP-based physical security products like IP network camera (IPNC) are becoming essential in construction of a modern security system. In order to guarantee interoperability among them, ONVIF has been a de facto standard communication framework. In addition to core specification, ONVIF defines many services. ONVIF Event service is supposed to provide notification messages to registered clients when events happen, which is an essential mechanism to be support to make IPNC intelligent. In this paper, we report our efforts to implement ONVIF Event service for the smart IPNC. First, we design S/W architecture, necessary data structures, and workflow of ONVIF Event service according to ONVIF Event service specification. Then, we implement the design about ONVIF Event service by extending TI's IPNC reference RDK S/W package [11]. Testing via an Open source ONVIF client verifies our implementation works properly. Tracking people or moving objects across a PTZ camera and maintaining a track within a camera is a challenging task in applications of video surveillance. In this paper we propose a novel object positioning tracking framework, PTZ camera based position tracking system, which is also known as PCTS, can help estimate the position of an object by using camera parameters, pan and tilt. From object motion vector, the relevant information such as time, background and geographic parameters can be recorded in database as well [12]. In the experiment, the change of a person' position is recorded by the PTZ camera in real time. The PCTS provides a feasible solution for position analysis and security surveillance services in future. Real-time person tracking has been implemented before, but not quite as what we have planned on doing; real-time broadcasting of the footage without any delay. The complete package of having the lecture capturing along with audience if necessary, screen sharing, Face Recognition based remote login and attendance marking for online participants, viewing the lecture in real-time with added features such as reading what the lecturer has told and intelligently generating chapters on the video according to the lecture slides played alongside with this makes Lecture Capturing System a perfect complete package of e-learning. A thorough research related to e-learning systems has led to the identification of some of the most influential factors used in the field of information systems research. More specifically, characteristics as well as the limitations, weaknesses, and strengths of web-based learning systems. Student variables, such as technical issues and adapting to the new ways are important variables that influence student learning, especially in a

collaborative e-learning environment. Understanding these variables is now helpful for developers to design eloquent educational activities to promote student knowledge construction and make learning more effective and appealing. In particular, this research helps to better understand the characteristics of students and to comprehend what the students expect from the learning management systems. This can help the developers achieve the most effective deployment of such systems and also helps them improve their strategic decision making about technology in the future, they can decide on the best approach that fit their students before implementing any new technology.

Several work exist in facial recognition based authentication, attendance marking and bandwidth management areas. Haar classifier for face detection, Eigen face algorithm for face recognition, image normalization and histogram normalization for image and contrast enhancement is proposed. [25] The system uses IP Camera mounted in front of a classroom which continuously capture image of the entire class at set interval, throughout the period of a lecture and sends the images over the internet to a cloud server for processing. The server processes the images by detecting the human faces contained, extract the faces and matches them with the enrolled faces of the students stored on the database. Before the images are used for detection and recognition purposes, the images are normalized or converted to gray scale which enhances the accuracy of face detection and recognition. Histogram normalization is then proposed for contrast enhancement. [25]

Sujata G. Bhele and V. H. Mankar has described wide range of methods used for face recognition which includes Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA) and Gabor wavelet soft computing tool like Artificial Neural Networks (ANN) for recognition and various hybrid of this techniques. In addition, the methods that challenges face recognition such as pose variation, illumination and facial expressions are described further. [26]

In terms of accuracy and speed of face detection and recognition processes, a new approach is proposed by evaluating various face detection and recognition methods such as Haar-like features, LBP, SVM, Adaboost algorithm, Gabor features and HOG by performing tests on face rich databases in terms of subjects, pose, emotions, race and light. [27] AdaBoost classifier is used with

Haar and Local Binary Pattern (LBP) features whereas Support Vector Machine (SVM) classifier is used with Histogram of Oriented Gradients (HOG) features for face detection evaluation in [27]. Haar-like features is the proposed face detection solution in [27] which has reported relatively well than LBP method. In terms of recognition, Gabor method is the proposed solution in [27] as its qualities overcome datasets complexity.

Traditionally in live streaming, data is served by putting a huge network load on the servers. A new approach is proposed which introduce two new WebRTC-based communication protocols [29] known as WebPeer and CodedWebPeer designed especially for browser based P2P streaming. [28] Two network metrics called network health to measure overall data saturation and network stability to show if the peers in the network are able to serve each other without the help of the server are proposed. The study [28] shows through measurements that by applying network coding, network health is increased by up to 100% without changing the cache size or number of peers. Furthermore, network stability is achieved using up to half the cache compared to the other approaches, without increasing the servers load. A caching system is proposed to reduce bandwidth costs when live streaming. [28]

2.1.1 Existing Applications

1. Panopto - Lecture Capture Software

Panopto is an easy-to-use video platform for training, presenting, and communicating that enables users to record videos and rich media presentations and push out to subscribers in many different formats. Panopto is built with the flexibility to record any combination of video sources, in any configuration, in classrooms of any size. And Panopto scales with ease to meet institution's needs from small departmental deployments to campus-wide installations [20].

1. BigBlueButton

BigBlueButton is an open-source web collaboration software utilized by education organizations for e-learning and training. The software offers numerous options for customization and integration as per requirements of the users. BigBlueButton enables users to conduct web-conferencing and

share documents, audio and video files for online learning. The software's "whiteboard" feature allows presenters to mark valuable topics in the presentation. In addition, its "polling" feature engages learners and helps the presenter to receive feedback. BigBlueButton's "desktop sharing" feature extends beyond slides and allows moderators to share their screen with the audience enabling a better understanding of topics. BigBlueButton supports multiple users in a video conference with no cap on numbers of active webcams. The software also supports voice conferencing via Voice Over IP (VOIP) without additional hardware requirements [21].

1. Echo360

Echo360 combines video management with lecture capture and active learning to increase student success.

Echo360 keeps notes linked to class presentations and videos so that students can jump straight from their own words to those of the instructor and replay the entire learning experience.

High-quality live streaming supports remote learners and classroom overflow situations. The streaming experience leverages Echo360's engagement tools so students can engage with classmates and the instructor no matter where they are.

Built on a scalable cloud architecture, the platform allocates the resources needed to process peak loads automatically. Videos are uploaded and processed in real time so the optimized version is available as soon as class is over [22].

4. Kaltura

The Kaltura Video Player SDK includes a rich set of APIs for player embedding, customization, white-labeling and integration via JavaScript or ActionScript 3.0. By leveraging the Kaltura Player you can create your own custom players with less effort and at no cost [23].

- Endless flexibility for creating your own custom design and playback experiences
- Automatically switch between HTML5 video and Flash, maintaining a unified look & feel
- Work with any type of streaming protocols, from adaptive streaming, HTTP streaming, and DRM
- Increase engagement with smart and dynamic playlists, related, and more

- Optimize SEO using Kaltura's SEO best practices embedding guidelines.

5. Open Broadcaster Software (OBS) (Software Tool)

OBS is a free and open-source streaming and recording program maintained by the OBS Project. The program has support for Windows 7 and later, OS X 10.10 and later, and Ubuntu 14.04 and later.

OBS is a free and open-source software suite for recording and live streaming. Written in C and C++, OBS provides real-time source and device capture, scene composition, encoding, recording, and broadcasting. Transmission of data is primarily done via the Real Time Messaging Protocol (RTMP) and can be sent to any RTMP supporting destination, including many presets for streaming websites [24].

2.2 Methodology

2.2.1 System Functionality

The main functionality addressed under this section are gesture based camera control and audio and video streaming.

2.2.1.1 Audio and Video Streaming using Web Real-Time Communication (WebRTC)

An IP camera will be used to track the lecturer's movement and gestures in front of the camera and produce the necessary PTZ signals to pan, tilt and zoom accordingly thus ensuring that the lecturer's actions are always recorded without missing any detail. This video recording will be immediately compressed on the fly in order to reduce its file size, then streamed live and also saved in the database for backup purposes and viewing later. Therefore, the students have the

choice of attending the lecture via the live stream or listening to the lecture later. This is very beneficial to students since they can also attend lectures without being physically present (remotely) at the lecture. During the live streaming session, the system will decide which video resolution (e.g. 480p, 720p, 1080p) to use for the playback at the student's end depending on the speed of his/her internet connection.

Live streaming is achieved via Kurento which is a WebRTC (Web Real-Time Communications) media server and a set of client APIs. During the live streaming session, the lecturer also has the ability to share his/her entire computer screen with the participating students if required, making certain that not even the most minute detail is not missed. In the case of having low bandwidth to support this feature, the lecturer has the option to disable the IP cameras in order to save bandwidth. This mode of lecturing provides better participation and interaction between the lecturers and students.

WebRTC is a set of protocols, mechanisms, and APIs that provide browsers and mobile applications with Real-Time communication capabilities over peer-to-peer connections. [18] WebRTC has proven to be a technology that allows well defined communication between browsers without the mediation of any kind of intermediate software as shown in the figure below.

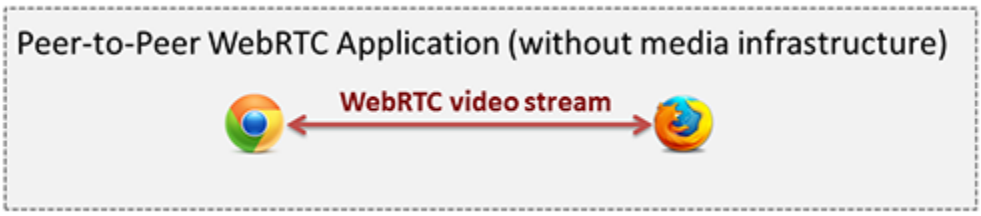


Figure 1: *WebRTC Application without media server.*

S***ource:*** *https://doc-kurento.readthedocs.io/en/stable/user/about.html*

This method of using WebRTC protocol directly can easily be used with simple web applications.

Lecture capturing system requires features such as live video and audio streaming, video recording, and audio/video conferencing. For this purpose there is a need of a WebRTC Media Server to mediate the communication between browsers as shown in the figure below.

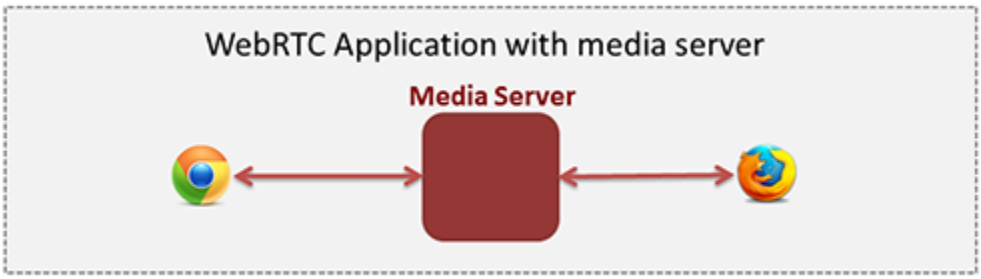


Figure 2: WebRTC Application with media server

S***ource:*** *https://doc-kurento.readthedocs.io/en/stable/user/about.html*

Kurento is a WebRTC Media Server and a set of client APIs that simplify the development of advanced video applications for web and smartphone platforms. Its features include group communications, transcoding, recording, mixing, broadcasting and routing of audiovisual flows.

With Kurento, it's an easy task to add third-party media processing algorithms to any WebRTC application, like integrating Computer Vision, Augmented Reality, video indexing, and speech analysis.

Kurento Media Server is ideal to be used with the Lecture Capturing system because of all the features that are mentioned above. Furthermore, Kurento Media Server is an open source software which provides all of the following advantages:

- Kurento Media Server and applications can be collocated, escalated, or distributed among different machines.
- Developers do not need to be aware of internal complexities of Kurento Media Server.
- Ability to support multiple video and audio sources.
- Streaming from sources such as PTZ cameras.

- End-to-end communication capability.

The main component that our Lecture Capturing System uses is the ability of Kurento Meida Server to obtain a live stream from an IP camera.

The default video source supported by Kurento Media server is the webcam installed in the computer. However, this application requires to capture the lecturer's live video using an IP camera, and stream it to all remotely logged users.

The camera used in Lecture Capturing System uses the Real Time Streaming Protocol (RSTP) for establishing a Transport Protocol for Real Time Applications(RTPA) media session.

To integrate such an IP camera with a webRTC application, it is first required to achieve media interoperability.[17]

Kurento Media Server provides a set of endpoints that are designed for this purpose.

-The Principal component which is responsible for capturing a video stream from an IP camera is the PlayerEndpoint.

- Kurento Media Server WebRTCEndpoint facilitates publishing media streams to WebRTC browsers with full termination of RTSP feedback.

- Kurento Media Server agnostic media capability performs, transparently for the developer, all appropriate trans-codifications when two incompatible media elements are connected, just by connecting the PlayerEndPoint to the WebRTCEndpoint.

2.2.1.2 Gesture based camera control

When it comes to learning, there are obvious moments of direct questioning from the lecturer, by the members of the audience. In such situations, although the crowd present in the classroom can see the person who is asking the question, the remote users will have no clue, unless the person is focused by the camera. For this purpose, when a student indicates that he/she has a doubt, the lecturer will perform a predefined gesture towards the camera

which will recognize it and turn towards the audience. The main objective of having gesture based recognition is ensure remote users get the feeling of being in a real classroom, other than getting the feeling of watching a conventional video. To get the camera turn back to the lecturer, he/she has to perform a predefined gesture towards the webcam in the laptop.

Gesture recognition can be addressed as a challenging task to computer vision due to the segmentation of the foreground object (in this case, the hand gesture), from a cluttered background. Here, the most obvious reason is because of the semantic gap involves when a human looks at an image, and when a computer or a camera looks at the same image. The human eye can easily identify separately the components of an image, but, a computer or a camera sees an image as a 3 Dimensional matrix. In this scenario, it is necessary to mimic mammalian vision using a PTZ camera to recognize a hand gesture performed by the lecturer in order to turn it towards the audience by 180º.

Currently, there are two types of approaches in gesture recognition that are used in computer vision. They are namely; “Data-Glove based” and “Vision based” approaches. The first approach uses sensor devices for digitizing hand and finger motions into multi-parametric data. The extra sensors make it easy to collect hand configuration and movement. In vision based gesture recognition, only a camera is required to capture a gesture. [15]

Because “Data-glove based” approach is complex and requires more sophisticated equipment, using this method would be cumbersome especially when it comes to the lecture capturing system. In this system, “Vision based” gesture capturing approach is used as it is the most suitable way to overcome the problem we are trying to address in this thesis document.

Under “Vision based” gesture capturing, there are two subcategories to be considered.

1. 3D hand model based approaches :
2. Appearance based approach

In this solution, appearance based approach is mainly used as it is less complex, easier to handle, efficient. The two main technologies used here are OpenCV and Python.

The methodology can be split into two major steps:

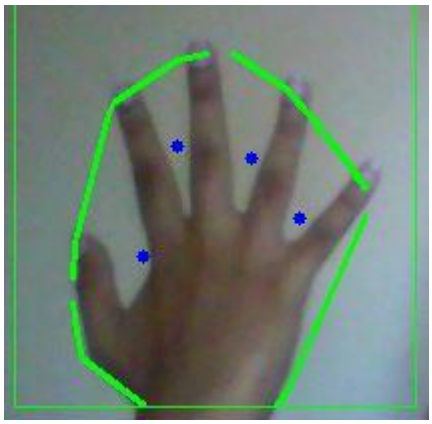
1. Identification and segmentation of the hand region from the video sequence.
2. Counting the number of fingers from the segmented hand region.

The hand region can be referred to as the foreground component and everything else exist in the background of the video sequence considered. This is done by running averages for 30 frames of the video sequence. In that way, the system will know the background of the video, and anything else which comes into it later on will be considered as the foreground object/s. Here, attention should be paid to the way of analyzing the foreground object/s alone and recognizing what a hand object is.

Background subtraction is the method that is used to separately identify the hand gesture from the background. After the background is identified through running averages, the absolute difference between the background and the current frame (which is updated over time) is calculated. This results in an image which holds the newly added foreground image, which is the hand. The resulting image is known as the difference image.

Detecting the hand region from this difference image is done by applying motion detection and thresholding. Here, only the hand region is visible, while all other unwanted regions are painted in black.

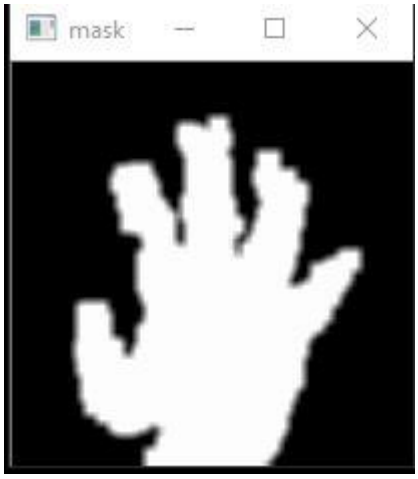
As an example, here is how thresholding is applied to a hand showing 5 fingers towards the camera.



Current Frame

Background

Application of **background subtraction, motion detection** and **thresholding**



Foreground mask

Figure 3: Application of background subtraction, motion detection and thresholding to the current frame and background.

The next step in detecting the hand area is contour extraction. Contour is the outline boundary of an object located in an image.

The human hand is a complex anatomical structure consisting of many connected parts and joints, involving complex relations between them, providing a total of roughly 27 degrees of freedom.

In the figure above, after identifying the skin color pixels, a convex hull is drawn around the region that was identified as the hand in green color, after which the number of defects are calculated. Defects refer to the number of angles that are formed in between the fingers. This way, the algorithm calculates the number of fingers held towards the camera using the following formula.

Number of fingers = Number of defects + 1

In the above example, I have taken 5 fingers. The blue dots indicate the number of defects formed within the convex hull/ hand area. i.e. 4 defects.

The above algorithm is developed to recognize any number of fingers from 1 to 5 according to convenience.

After the predefined gesture is performed at the PTZ camera, it turns 180º towards the audience. The node-onvif is a Node.js module which allows you to communicate with the network camera which supports the ONVIF specifications.
The ONVIF (Open Network Video Interface) is an open industry forum promoting and developing global standards for interfaces of IP-based physical security products such as network cameras. The ONVIF specifications are available in their web site.
Recently, most of network cameras for business support the ONVIF standard. Furthermore, some network cameras for home support it though the implementation is partial. The node-onvif allows you to control network cameras which implement the ONVIF standard.

With this there are separate commands to move the camera movement left, right, up or down at selected speeds and can freeze movement according to the requirements.

2.2.2 Use Case Scenarios

Table 2: Use case scenario 01- Gesture based camera control

Use Case No	01
Use Case Name	Gesture Based audience Capturing
Actors	Lecturer , Members of Audience
Pre-Condition	The lecturer should be on the stage and system should be running.

Main Success Scenario	1. The Lecturer has to log into the system with correct credentials. 2. After login to the system the Lecturer has to start a new session. 3. A member of the audience has a question to ask. 4. The lecturer will perform a predefined gesture at the PTZ camera. 5. The camera will process and identify the gesture and turn towards the audience. 6. The student will ask the question from the lecturer. 7. Once the student finishes asking the question, the lecturer will perform a predefined gesture at the webcam of the laptop.
	9. After recognition of the gesture, the PTZ camera will turn back towards the lecturer.

Post Conditions	1. Camera will successfully focus on the lecturer again.
Extensions	1.1 If the user name or password is incorrect lecturer has to enter again and log into the system. 1.2 If the password or username has forgotten lecturer have to click forgot your password link. 5.1. The camera will not identify the gesture, and the lecturer will have to perform it again.
	9.1. The webcam will not recognize the lecturer's gesture, and he/she will have to perform it again.

2.2.3 User Characteristics

There are three types of users that interact with the system: Student, lecturer and administrator. Each of these users has different use of the system therefore each of them has their own requirements.

- The student can log in to the web application using facial recognition process and follow the live streaming lecture. The student also can view the attendance updated to the web application by the automated attendance monitoring process in the system or manually by the lecturer and administrator. The student also can access the lecture recordings uploaded by the lecturer.

- The lecturer can log in to the web application and open a new lecture session to live stream the lecture and share the PC screen to other remote students logged in to the web application. The lecturer also can update the attendance of the students manually if the automated attendance process has done a mistake and also, the lecturer can record the presentations offline and upload to the server. Finally, the lecturer can manage courses and manage the lecture time schedule to automatically activate the camera to mark the attendance of the students.

The administrator logs in to the web application and register users with the image data for facial recognition-based login. The admin is responsible for monitoring bandwidth and managing quota for the students. The admin can also update attendance of the students. This person is also responsible for user management.

2.2.4 Constraints

- Time is the major constraint of this system. Team members that have more speed than the others can finish their tasks quickly within a short period of time. So when allocating tasks better team members come to the top and those who have a less speed going to the bottom of the list. Those who are finishing early can imagine their feedback also in a good state. Maintaining a peer reviewing like this can save the time

of the supervisor. Because he can analyze all the good and bad reviews of the team members and can update his team member's list to allocate tasks automatically.

- Quality is also a good constraint when giving the feedback to a team member. If a team member finished his work early, but the work is not in a good quality that is not good. If it is like that others are not giving a compliment about him.

  Therefore, team members must concern about his or her quality too.

- The internet connection should have a higher bandwidth as it needs to transfer a large set of data.

- There is the secondary memory, which is where your files are stored, and a computer can use a Hard Disk to store memory, or a Solid State Drive (As well as other kinds of Flash Memory). Modern Hard Disks (HDs) can store from around 500 GB to 2TB of data. Hard Drive writing and reading speed is also a factor since have to deal with high amount of data when encoding, and writing video files to both cache of the local computer and to the main server storage.

- There is also primary memory, which is the memory that stores information that you are manipulating with immediately, when the computer is ON. To maintain an efficient smooth usage of the system there should be adequate amount of RAM in the local computer.

2.2.5 Assumptions and Dependencies

- Web Server is up and running 24x7.
- If the computer does not have enough hardware resources available for the application, there may be scenarios where the application does not work as intended or even at all.
- Computer is connected to the same network as of the necessary cameras to run the system.
- Computer is connected to a reliable internet connection with a high speed bandwidth connection.

2.2.6 External User Interface Requirements

2.2.6.1 User Interfaces

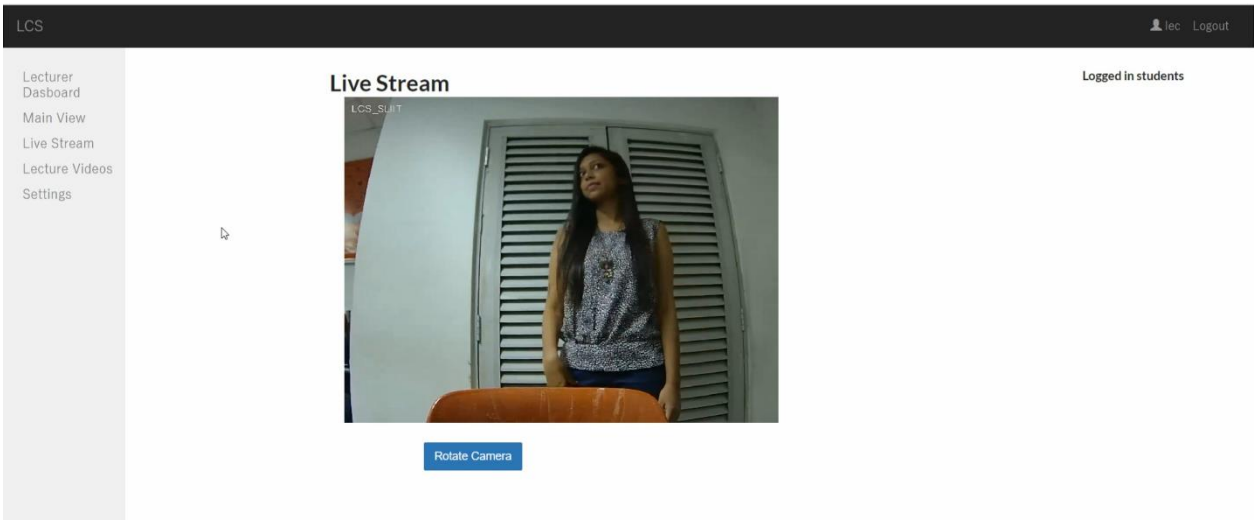


Figure 4: Main Interface with Live Stream

2.2.6.2 Hardware Interfaces

A PTZ camera is used to capture the lecturer and the classroom.

2.2.6.3 Software Interfaces

The communication between the different parts of the system is important since they depend on each other. Following software, interfaces are required to Lecture Capturing System.

- MongoDB database
- WebStorm IDE
- Open Broadcaster Software
- A latest updated web browser (Mozilla Firefox version 59 or above, Google Chrome etc.)
- Browser WebRTC Extension for Screen Sharing Studio3T MongoDB GUI and IDE.

- Kurento Media Server should be running on Ubuntu 14.04 or 16.04.

2.2.6.4 Communication Interfaces

Following communication, interfaces are required operate Lecture Capturing System.

- Internet connection with fairly higher bandwidth will be required to run the system efficient since it deals with large about of video files.
- Database connection interface will be needed to connect to the database.

2.2.7 Software System Attributes

The requirements in this section specify the required reliability, availability, security and maintainability of the software system.

2.2.7.1 Reliability

The system should be reliable in its operations which it provides for all users. Information present in the system should be consistent created in the system will be retained for a number of years without the data being changed by the system

2.2.7.2 Availability

- The system functionalities should be available at any given time to users (24/7 accessibility).
- There should be no downtime or lag in the system when there are many concurrent users accessing the system.

2.2.7.3Security

User access to the system has to be controlled so that only authorized users can gain access to the university's content and resources.

- Users have to be authenticated by their valid username and password to access the web application.
- Passwords must be hashed before storing them in the database instead of plain text so that no unauthorized persons can access another user's account.
- User authorization verifies that users have access to information depending on their access roles/rights (e.g. Students cannot access resource materials of lecturers).
- Biometric authentication in the form of facial recognition would be used to identify remotely logged in students to the system. This further ensures that the university's resources are not exposed to unauthorized people.

2.2.7.4 Maintainability

The system may need regular corrective and preventive maintenance to make certain that any hindrances and limitations are removed at the earliest time of detection. This is importance to ensure a smoothly functioning system

2.2.8 Research Findings

When addressing gesture detection using the appearance based approach, it is proven to be the most efficient algorithm in real-time gesture recognition. It is simpler than the 3D hand model algorithm as the appearance based algorithm applies cosine rule to the hand region present in the area of interest and calculates the number of fingers in real time. When compared with 3D model based approach, appearance based approach used in this system is more content and comprehensive. In this manner, without the use of sophisticated calculations as in the 3D model approach, it detects the gesture efficiently and commands the camera to turn towards the audience within 1 second of recognition. Although the said approach is the most suitable for Lecture Capturing System, it is necessary to have optimal lighting conditions and no skin coloured objects other than the lecturer's hand present in the background.

Media streaming is the most important part in Lecture Capturing System. Kurento Media Server was selected to perform live media streaming because, unlike other media servers like Jitsi, Janus, and Ant Media server, Kurento is free and open source with a set of flexible functionality which can be used by the Lecture Capturing System. For optimum results, in is advisable to run Kurento on a separate server machine with Ubuntu 14.04 or 16.04 installed.

3 RESULTS AND DISCUSSION

3.1 Evidence

Figure 5: Before gesture detection

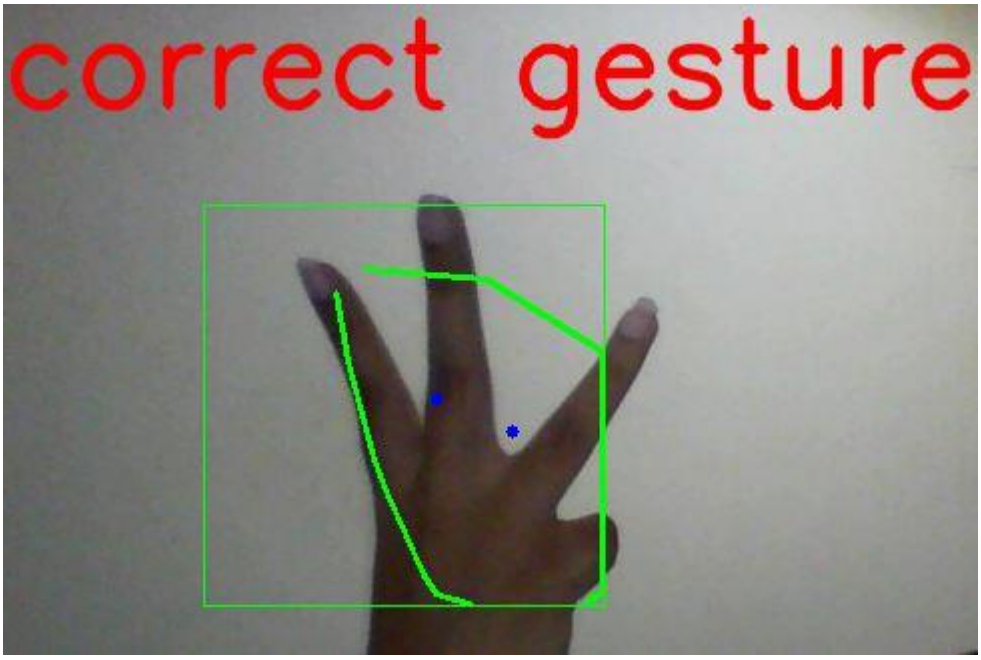


Figure 6: After identification of the correct gesture

3.2 Discussion

In the scenario of gesture based camera control, the appearance based approach described under methodology gives an accuracy of 93.25% based on the set of input gestures given. As the number of defects is calculated by applying the cosine rule to all the triangles identified in the hand region, it is highly important that there should be optimum lighting level in the lecture hall. Furthermore, there should be no other skin coloured objects in the area of interest where the gesture is performed (i.e. the green coloured rectangular box region).This is because skin tone thresholding is involved by pixel selection. [16]

When obtaining the live stream using Kurento Media Server, it is necessary for Kurento Media Server to be up and running on a Linux based operating System. Kurento Media Server exclusively supports Ubuntu 14.04 and Ubuntu 16.04 versions. After installing and running Kurento Media Server as mentioned in Kurento documentation [19], it can be tested and verified whether Kurento Media Server is up and running as shown in the figure below.

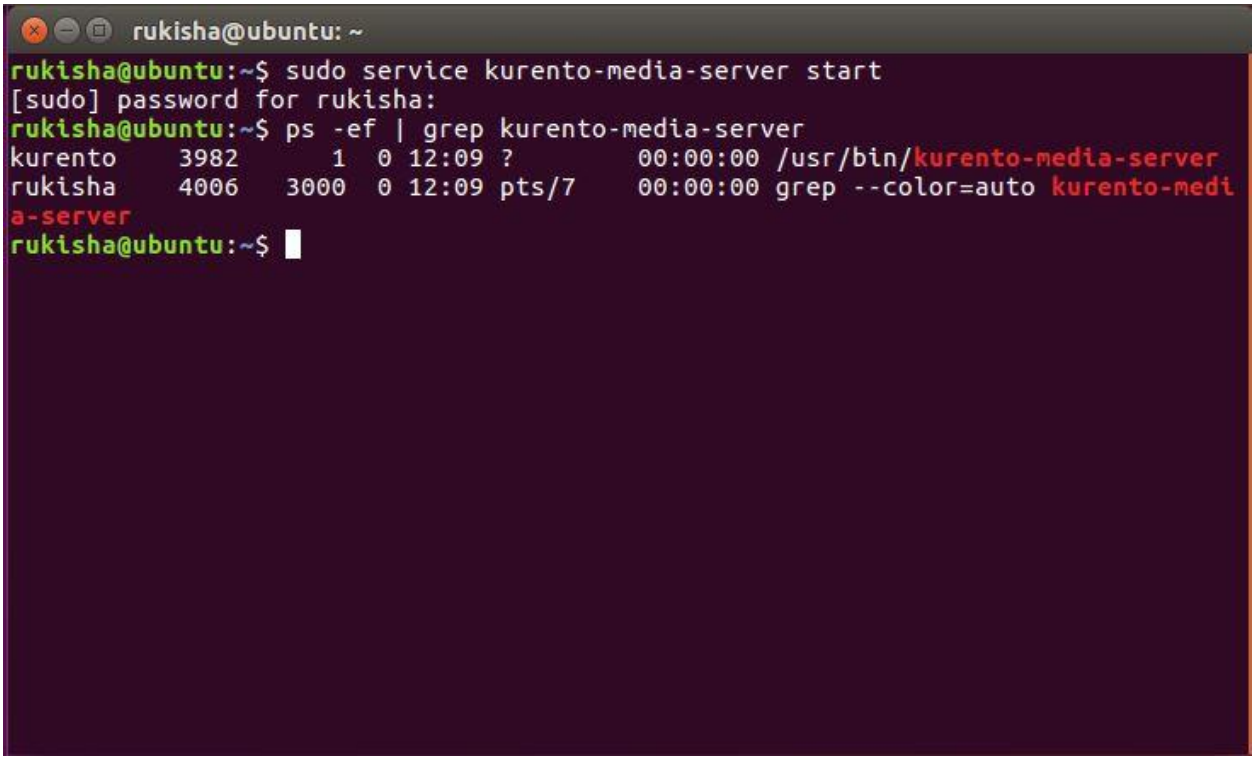


Figure 7: Output of the processes of Kurento Media Server when it is up and running

To verify further the port that Kurento Media Server Listens on, the following can be done.

Figure 8: Result indicates that Kurento has opened the port 8888 to send and receive requests using Kurento Protocol

4 CONCLUSION

This report examines an innovative approach that is best suited to develop a lecture capturing system that provides a complete classroom experience and much more to remotely logged in students. This system stands itself unique from other existing products and being as a comprehensive product that includes biometric authentication, gesture detection, live streaming of lectures, automated attendance marking, offline recording of lectures, bandwidth management and desktop screen capturing all in one.

This research work has been developed mainly for addressing the problems in Sri Lankan universities, specifically addressing the lack of interactivity between the lecturer and the students. Even though this research focuses on universities, it definitely has the potential to be used in other fields such as business conferencing. In next stage, in one hand, research team will be focusing on improving the accuracy of the face recognition and gesture detection models by testing other algorithms effectively. Also, research team will focus on minimizing bandwidth costs by testing out bandwidth optimization techniques. It is hoped that for any person who expects to build a similar system or any other real-time system, results of this research will be an aid and will provide insight on the performance, accuracy and reliability level that can be expected with the combination of tools, technologies, programming approach considered in this document.

5 REFERENCES

[1] W., *"Use of E-Learning," Universiti Teknologi*. Malaysia, Johor, Malaysia, 2018.

[2] M. A. Mahmod, "E-learning in Iraqi Universities," in *International Conference on Computing, Engineering, and Design (ICCED), Selangor*, Malaysia, 2017.

[3] P. D. Z. Varcheie, "Online Body Tracking by a PTZ Camera in IP Surveillance System," Department of Computer Engineering and Software Engineering, Station Centre-ville, Montr´eal, (Qu´ebec), Canada, 2009.

[4] T. G, "Dries Hulens, 'Autonomous lecture recording with a PTZ camera,'" in *Canadian Conference on Computer and Robot Vision*, Sint-Katelijne-Waver, Belgium, 2014.

[5] M. M. M. H. R. Jacko, "Remote control of the PTZ camera system for lecture rooms," *Dep. Comput. Inform.*, 2015.

[6] B. Wulff, "OpenTrack - Automated Camera Control for Lecture Recordings," *IEEE Int. Symp. Multimed.*, 2011.

[7] C.-F. C. a. P.-C. S. Yong-Quan CHEN, "A Tabletop Lecture Recording System," in *International Conference on Consumer Electronics-Taiwan*, Taiwan, 2015.

[8] Y.-T. T. S. C. a. S.-W. C, "Chiung-Yao Fang, 'Chiung-Yao Fang, You-Ting Tsai, Shuan Chu, and Sei-Wang Chen,'" Department of Computer Science and Information Engineering, Taiwan, 2015.

[9] P. Kumar and T. S. Sheng, "Real Time Target Tracking with Pan Tilt Zoom Camera," Digital Image Computing, Adelaide, 2009.

[10] D. P. Hutabarat, D. Patria, S. Budijono and R. Saleh, "Human Tracking Application in a Certain Closed Area Using RFID Sensors and IP," Information Tech, Jakarta, 2016.

[11] C. C. Phan, B. T. Nguyen and S.-T. Chung, "Design and Implementation of ONVIF-based Event Service for DM 814x Camera," IEEE, DaNang, 2015.

[12] C.-S. Yang, R.-H. Chen, C.-Y. Lee and S.-J. Lin, "PTZ Camera Based Position Tracking in IP - Surveillance System," IEEE, Tainan, 2008.

[13] Synesis, ONVIF Device Manager (Onvifdm), April 2015. [Online]. Available: http://sourceforge.net/projects/onvifdm

[14] "YOLO: Real-Time Object Detection." [Online]. Available: https://pjreddie.com/darknet/yolo/. [Accessed: 08-Oct-2018].

[15] Pragati Garg, Naveen Aggarwal and Sanjeev Sofat "Vision Based Hand Gesture Recognition" World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:3, No:1, 2009. [Accessed: 04-July-2018]

[16]Swapnil S. Potnis, Asst. Prof. Aditi S. Jahagirdar "Real Time Hand Gesture Recognition for Smart Classroom Environment" International Journal of Computer Trends and Technology (IJCTT) – volume 17 number 2 – Nov 2014 [Accessed: 19-July-2018]

[17] "Interoperating WebRTC and IP Cameras"[Online]. Available: http://www.kurento.org/blog/interoperating-webrtc-and-ip-cameras [Accessed: 08-Aug-2018]

[18] “About Kurento and WebRTC” [Online]. Available: https://doc-kurento.readthedocs.io/en/stable/user/about.html [Accessed: 20-Feb-2018]

[19] “Installation Guide” [Online]. Available: https://doc-kurento.readthedocs.io/en/stable/user/installation.html [Accessed: 7-May-2018]

[20] “Lecture Capture & Recording Software For Education | Panopto,” *Panopto Video Platform*. [Online]. Available: https://www.panopto.com/panopto-for-education/lecture-capture/. [Accessed: 10-Oct-2018].

[21] “Home - BigBlueButton.” [Online]. Available: https://bigbluebutton.org/. [Accessed: 10-Oct-2018].

[22] “Echo360 - The Smarter Video Platform for Higher Ed and Continuing Ed.” [Online]. Available: https://echo360.com/. [Accessed: 10-Oct-2018].

[23] “Home,” *Video Platform | Kaltura*. .

[24] “Open Broadcaster Software | Home.” [Online]. Available: https://obsproject.com/. [Accessed: 10-Oct-2018].

[25] [Online]. Available: http://www.academia.edu/37437099/Automated_Student_Attendance_Management_System_Using_Face_Recognition. [Accessed: 07-Aug-2018]

[26] S. G. B. a. V. H. Mankar. [Online]. Available: https://www.researchgate.net/publication/270583005_A_Review_Paper_on_Face_Recognition_Techniques.[Accessed: 07-Aug-2018]

[27] F. Ahmad, "Image-based Face Detection and Recognition," [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1302/1302.6379.pdf. [Accessed 1-05-2018].

[28] E. Patrick J. Braun. [Online]. Available: https://www.researchgate.net/publication/317452964_On_the_Performance_Boost_for_Peer_To_Peer_WebRTC-based_Video_Streaming_with_Network_Coding. [Accessed 8 4 2018]

[29] [Online]. Available: https://www.w3.org/TR/webrtc/. [Accessed 20 6 2018].