A REAL-TIME FACE TRACKING SYSTEM BASED ON A SINGLE PTZ CAMERA

A Thesis

by

SEOKTAE LEE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Zixiang Xiong
Committee Members,	Erchin Serpedin
	Jyh Liu
	Ulisses Braga-Neto
Head of Department,	Chanan Singh

December 2014

Major Subject: Electrical Engineering

ABSTRACT

It is evident that the effectiveness of education is strengthened from the assistance of distance learning which provides lectures online. The benefits of online education include enabling more distance courses based on local programs and increasing participation and interaction between the students and the instructor. Nowadays, conventional systems commonly implement a combination of static and PTZ cameras. However, such systems are not only costly but also require operators and high computational power in exchange.

Thus, this thesis proposes a real-time face tracking system based on a single PTZ camera as a cost effective solution by minimizing hardware requirements and functioning automatically. The proposed system focuses on the delay possible to occur due to the movement of the PTZ camera and the network delay which varies the video frame rate which alters the performance from a software perspective. The main contributions include the low cost and flexibility regarding installation. Preliminaries are introduced as a basis of the proposed system such that hardware is maintained to be minimal and universal while software is retained to use less computational power. The proposed system minimizes the delays to maintain pace with the subject of interest, provides a smooth and natural movement of the camera as if an actual operator controls the camera, and produces competitive results regarding performance compared to conventional systems.

DEDICATION

To my father, Dr. Sang Ho Lee, and my mother, Sook Yi Kim

ACKNOWLEDGEMENTS

I would like to especially thank my committee chair, Dr. Zixiang Xiong, for his guidance and support throughout the process of this research. It has been an educational and productive journey which required him to be patient with the progress of this project. His method of encouragement has enlightened my vision of interpreting problems and seeking solutions.

Appreciation also goes to my colleagues, Peng Yuan and Bo Wang, for making my time at Texas A&M University a great experience.

In addition, I would like to send my gratitude to my beloved friends in South Korea, Sa-hyun Kim, Woo-soon Hyun, Hyun-gyun Moon, Sang-hyuk Park.

Finally, I would like to professionally send appreciation to my father and mother for their encouragement, support, and belief.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

1. INTRODUCTION

With the assistance of sophisticated but robust networks and the improvement of image processing, the opportunity of education is expanding out from the classroom by providing high quality lectures online to students who are not in the position of attending in person. A survey [7] that has cumulated data over the past 10 years provides that the growth of overall enrollments regarding higher education is relatively lower compared to the number of students taking at least one online course.

The benefits of providing online lectures are not only subject to the number of registered students but also the possible audience regarding instructors. Enabling more distance education courses allows to expand the audience to students at remote locations who intend to acquire the opportunity of accessing equivalent quality lectures at their own convenient time and a reasonable cost. In addition, simultaneous classroom and online delivery tends to lower the barrier of participation and interaction between the instructor and the students [9] which is vital in terms of increasing the effectiveness of education.

However, the limitation associated with these benefits include requirements such as hardware equipment and operators to control the camera or the system [11]. The cost of equipment continues to elevate as systems integrate multiple cameras and computer systems with high computational power due to processing high definition video images. In addition to the limitation of using operators is not only subject to the cost regarding the wage but also requires the operators to be familiar with the camera and the system. Unless properly trained or educated regarding the system, camera operators tend to be unskilled

and do not having a deep understanding of the material presented and thus incorrectly frame the shot when using multiple cameras.

1.1 Related work

Nowadays, a combination of static and PTZ cameras are commonly implemented together allowing the viewer to change the viewing area. However, such systems are not only costly as a result of utilizing multiple cameras and additional operators to control the cameras or maintain the system but also require high computational power in exchange since either complex algorithms or protocols are involved [18].

Before commercial PTZ cameras were introduced, static cameras were set on tripods and the movement was controlled by operators behind the camera to capture the image of an instructor. However, PTZ cameras have become a favorite over static cameras due to the capability of viewing large regions with a single camera whereas several static cameras would need to be involved to cover the same amount of space with an operator manually changing the angle of the camera. Considering a typical lecture room environment, PTZ cameras are not constrained in viewing the entire classroom with the assistance of the panning and tilting features. However, PTZ cameras initially required a distinct operator as well to control the camera by altering the view with a remote controller or a software application. To eliminate the requirement of separate operators controlling the camera, conventional systems combine the object detection technique with the PTZ cameras. It is already known that the object detection technique is an essential element of

intelligent systems actively used in the surveillance field for the purpose of security or the electronic entertainment systems which provide a more realistic virtual experience.

The object detection technique takes part in the tracking operation as a basis to control a PTZ camera following a designated subject. Kang et al. [22] introduces a tracking system that determines the appropriate panning and tilting commands for the PTZ camera using a mosaic technique which detects a certain subject by taking advantage of the intensity values within the displayed image. However, the system requires a constant update or a substitution with pre-defined images of the entire FOV (Field of View) before the subject actually enters within the area regarding the new regions created as the camera moves. Skin color is also a commonly used method to establish a tracking system, as Amnuaykanjanasin et al. [16] proposes. The face of a subject is detected by implementing background subtraction to the image and the 3-dimensional position of the face is estimated referring to the detection result. Data regarding the estimated position of the face is then assigned to a PTZ camera to perform tracking. Yet, the estimation technique does not ensure the tracking accuracy to be identical to the actual movement of the subject. Varcheie and Bilodeau [17] applies motion detection, region sampling, and color appearance all together to detect a subject with a single IP PTZ camera. However, the single IP PTZ camera is only capable of successful tracking with the tradeoff of the frame rate being low and large displacements in the image plane. In other cases, systems [2, 23] introduce real-time face detection and a separate tracking algorithm using a single PTZ camera based on the OpenCV object detector which is a common ground related to the proposed system of this thesis. However, the introduced systems encounter limitations

such as movement delay regarding the camera, network delay which varies the rate of video frames arriving, and lost faces due to a narrow field of view.

1.2 Proposed approach

Among the solutions to the limitations, this thesis focuses on a system that captures the instructor during a lecture. Although viewers demand various views during a lecture such as the instructor, the board, or the lecture slides, the most fundamental view can be considered to be the instructor. Thus, a system capable of reducing the required cost for installation while performing competitively well compared to conventional systems is determined to be a fundamental approach.

This thesis proposes a real-time face tracking system based on a single PTZ camera without in need of an operator to control the camera. The proposed system is considered to be a cost effective solution by minimizing hardware requirements and functioning automatically in real-time. The system reduces the cost by focusing on minimal and universal components regarding hardware components and eliminates the need of an operator to control the camera by establishing the system to function automatically.

The focus of attention regarding the performance of the proposed system are 1) the reaction time of the camera being short enough to continuously keep pace with the targeted subject in real-time, 2) the motion of the camera being at most equivalent with an actual person moving the camera with a smooth and natural motion, 3) and most importantly, the performance of the proposed system to be competitive with conventional systems which projects the possibility of replacing expensive equipment with a cost effective alternative.

1.3 Contribution

The main contributions of this research to the online education field are the low cost and flexibility in installation, namely, mobility. The conventional systems previously mentioned [4, 13, 17, 18, 26] integrate one to multiple cameras, but the cost required for installing such systems for the purpose of research range from 1,000 USD to 7,000 USD whereas the cost of systems utilized by institutions or industries highly exceed the range of systems used for research. Another limitation of the multi-camera system is related to the flexibility of mobilizing the system. With multiple cameras installed in different positions and angles, it is difficult to transport the entire system to another location as intended. However, the proposed system contributes with the flexibility of mobilizing to different locations whenever needed.

Another contribution of the proposed system is to find a middle point of producing competitive results compared to conventional systems with minimum quantity of equipment. Conventional systems produce a rather extreme requirement or result such that systems require the quantity and quality of the equipment to be highly expensive or results produced with minimum quantity and quality of equipment are far away from the results produced with expensive equipment.

1.4 Thesis overview

This thesis initially sets up preliminaries as a basis and extends onto the proposed system. The second chapter introduces preliminaries such that hardware requirement is maintained to be minimal and universal while the software application is retained to be

simple using less computational power. Equipment related specifications are presented and the functionality of the software application is explained in a series of steps in detail. An evaluation test follows to assess the performance of operation regarding the preliminary components while suggesting improvements to adapt to the proposed system.

The third chapter implements the proposed system based on the results from using the preliminary components. Hardware equipment related specifications are initially introduced and continues on to elaborating on the software application with improvements based on the assessment tests performed on the preliminary components. Assessment tests are performed on the proposed system as well in a classroom environment on different subjects to evaluate stability and robustness while maintaining focus on the detection accuracy and the movement regarding tracking.

The last chapter contains the conclusion and introduces possible future work to expand the proposed system.

2. PRELIMINARIES

2.1 Overview

This thesis initially assembles preliminaries to ensure the functionality of real-time tracking and continues on to establishing the proposed system on the basis of the preliminary components introduced. The preliminaries described in this chapter are built on existing methods and algorithms, which extends to establishing the proposed system. The requirement regarding the preliminary components involve a single PTZ camera connected to a computer which performs detection and tracking while ensuring a fluid transition of data between each process within the computer as shown in Figure 1.

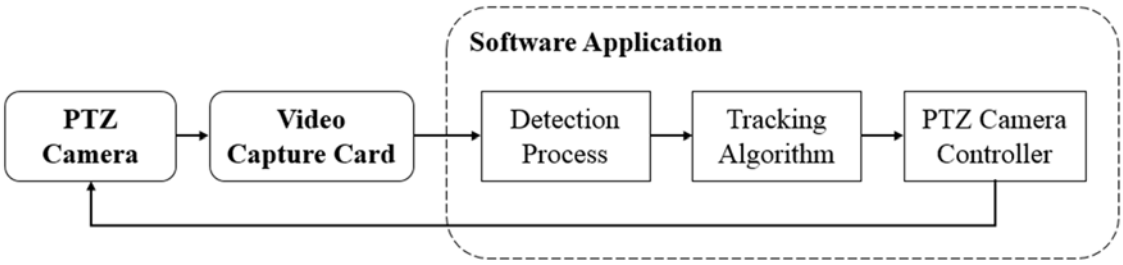

Figure 1: Block diagram of preliminary requirements

As the PTZ camera captures the FOV, the images acquired from the camera are transmitted in the form of a bit stream to a video capture card and stored in the memory. The detection process accesses the bit stream stored in the memory to operate face detection provided by OpenCV [8] searching for a face within each frame. The tracking

algorithm determines the appropriate commands depending on the result of the detection process which targets to maintain the face within the center region of the display screen. From the result of the tracking algorithm, commands are transmitted to the camera utilizing a camera controller which manages transmitting the appropriate commands regarding the movement of the camera.

2.2 Hardware requirements

2.2.1 Camera specifications

A donated UV83 PTZ conference camera holding equivalent specifications with the SONY EVI-D70 PTZ camera is locally connected to a computer. The camera specifications regarding the UV83 PTZ conference camera and the SONY EVI-D70 PTZ camera are compared in Table 1 while the options available for connection and the corresponding cables of the UV83 PTZ conference camera are shown in Figure 2.

Camera	Panning	max(Speed)	Tilting	max(Speed)	Zoom	Focus
UV83	360º	80º/sec	+90º, -30º	80º/sec	18x	Auto
SONY EVI-D70	±170º	100º/sec	+90º, -30º	90º/sec	18x, 12x	Auto

Table 1: Comparison of UV83 and SONY EVI-D70 specifications

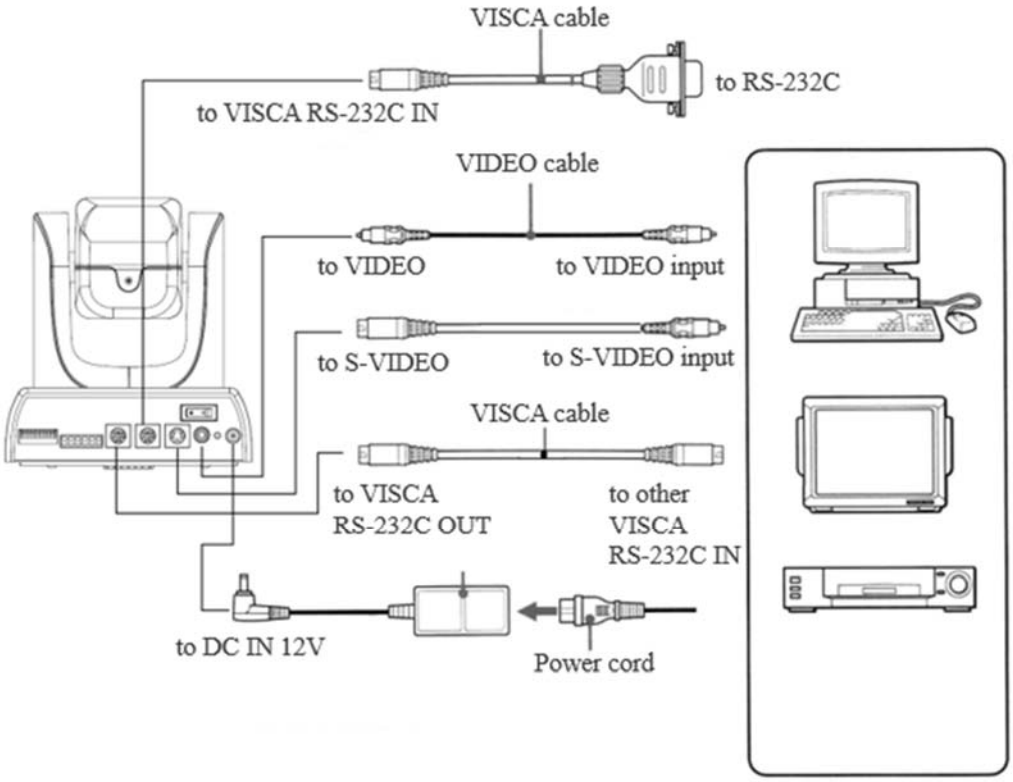

Figure 2: Options for connections and corresponding cables of UV83 PTZ camera

The UV83 PTZ conference camera allows up to 360º of rotation in the horizontal direction. Rotation in the vertical direction changes from the range of -30º to +90º when placed on a flat surface to the range of -90º to +22º when mounted on the ceiling. The control speed of the panning and tilting feature can vary from 0.2º per second to 80º per second while the preset value is set to 80º per second. The hardware and software protocols of the UV83 PTZ conference camera are similar to the commonly used SONY EVI-D70 PTZ camera which allows to use equivalent protocols regarding the connections and the communication method for the purpose of controlling the camera. It is possible to use

RS232, RS485, or RS422 as a control signal interface to transmit commands and allows to connect multiple cameras together with a single computer system. In addition, the camera is capable of outputting CVBS and S-VIDEO as a format of video images. The preliminary system decides to choose CVBS, also known as the composite video format, for the output video format and the VISCA RS232 communication protocol to control the PTZ camera.

2.2.2 Connection

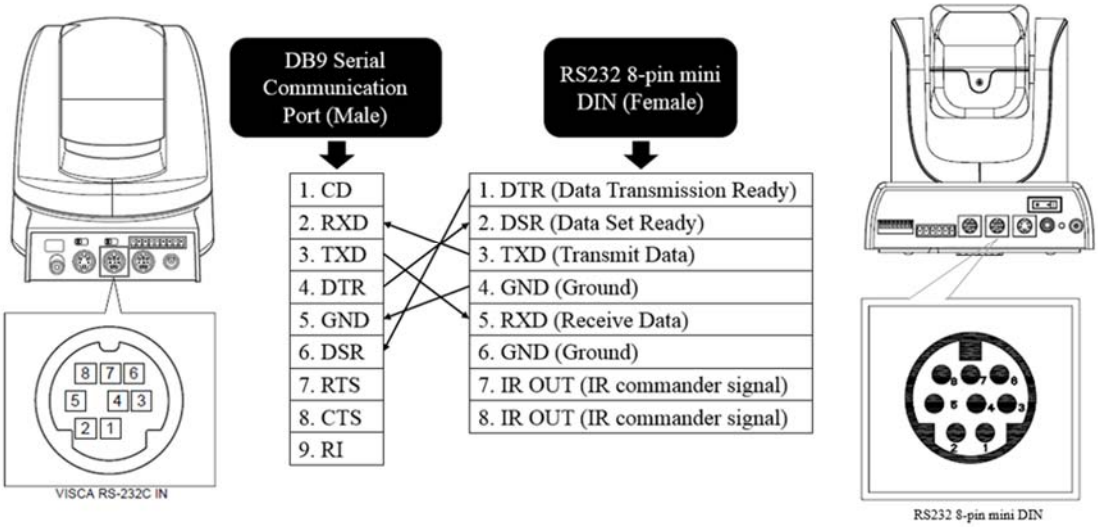

Figure 3: SONY EVI-D70 and UV83 PTZ conference camera pin-out

The SONY EVI-D70 and UV83 PTZ conference camera both adapt the 8-pin mini DIN port as a connection to communicate with the computer as shown in Figure 3. While

adapting identical ports in order to communicate, the pin-out diagram is identical as well which allows connecting to the DB9 serial communication port on the computer.

RS232 is chosen as the communication protocol among the available control signal interfaces including RS485 and RS422 due to the advantage in the communication method. While RS232 operates in full duplex which transmits and receives data simultaneously regarding a device, RS485 and RS422 operates in half duplex which allows to transmit and receive data, but not at the same time. Essentially, RS232 ensures data to flow in both directions simultaneously with separate transmit and receive signal lines.

The DB9 serial communication port on the computer is a standard connection which refers to a common connector type, housing 9 pins for the male connector. Typically, the serial ports are associated with a distinctive number identified on IBM compatible computers as COM (communication) ports which is a type of connection on personal computers used for peripherals. The DB9 serial communication port is a primitive route used in the earlier days of the development of computer systems. However, it is no longer available on hardware systems manufactured these days. Nonetheless, the preliminary system considers using the serial communication port due to the possibility of the connection still existing.

2.2.3 Video capture card

The preliminary system integrates a TW6816 single chip video capture card to receive the CVBS video image format in the form of a bit stream from the PTZ camera which is shown in Figure 4.

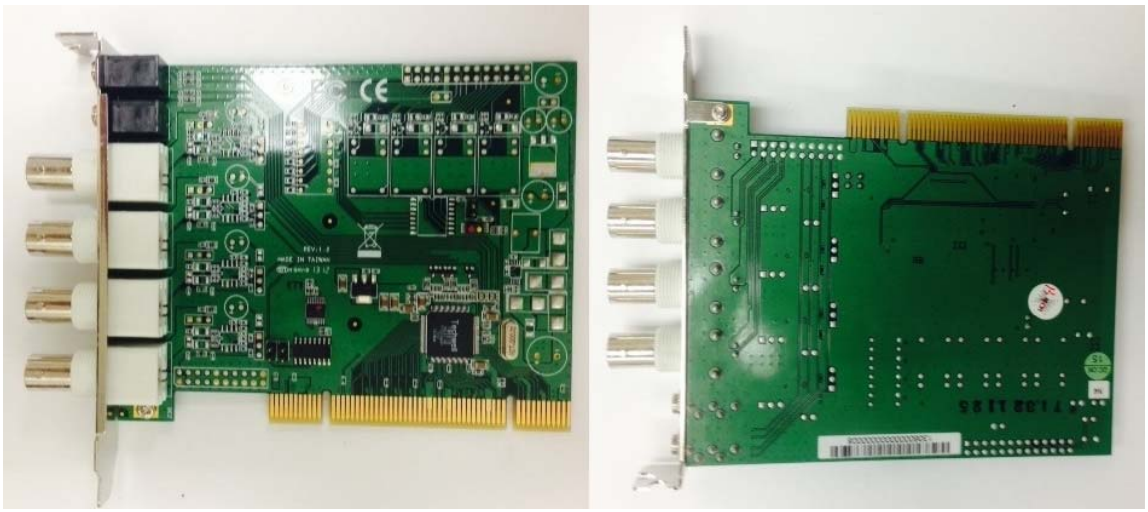

Figure 4: TW6816 single chip video capture card

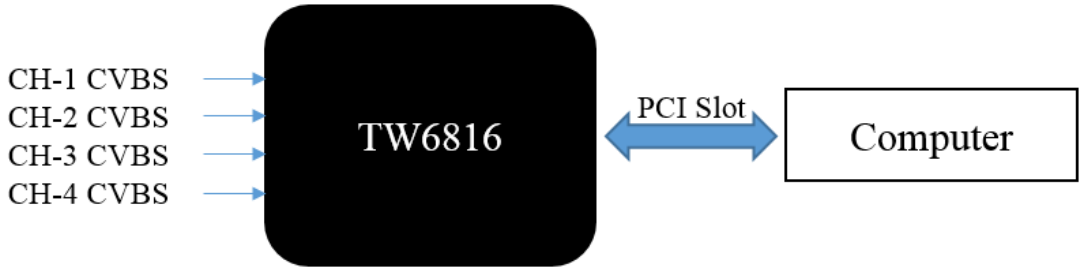

Figure 5: System solution diagram of TW6816 single chip video capture card

The TW6816 chip embedded on the video capture card is a single chip supporting multi-channel real-time video and audio capture via PCI (Peripheral Component Interconnect) interface for computer applications. The video capture card allows the computer to communicate with the PTZ camera using a signal protocol up to 4 channels as shown in Figure 5. The video capture card offers programmable white peak control for CVBS channels, advanced synchronization processing and sync detection for handling non-standard and weak signals, and automatic color control. The essential purpose of using the video capture card is to interact with the PTZ camera by receiving the input video images in a bit stream which is then stored in the memory to perform the detection process.

2.2.4 Video format

The video output connector on the UV83 PTZ conference camera is supported with a RCA phone jack to transmit composite video formats, also known as CVBS. The other end of the RCA cable is in the form of a BNC connector which is fastened to the video capture card on the computer to receive the video images captured by the camera.

The CVBS signal is known to be the lowest quality video format which also suffers from cross color artifacts. However, the composite signal is chosen as a preliminary component regarding video image formats since it is known to be the most commonly used analog video interface. In other words, composite video is the standard that is capable of connecting almost all consumer video equipment. The standard formats of the signal are usually NTSC (National Television Systems Committee), PAL (Phase Alternating

Line), or SECAM (Sequential Couleur Avec Mémorie) which are the three main analog video broadcast standards.

While commonly used, the artifacts of the CVBS signal became more noticeable as broadcast began to use larger and higher quality displays. Normally, CVBS is transmitted over basic composite video cables with male RCA plugs on each end which is the same connector used for standard line level audio connections. Composite video combines the three basic elements of a video picture which are the color information (chroma), the brightness information (luma), and synchronization data into a single combined composite signal.

2.3 Software requirements

2.3.1 Overview

The software application is written in the C++ programming language with Microsoft Visual Studio. The aim is to not only create an application capable of successfully processing the input video images but also to ensure smooth transition regarding data flow between the components.

The software application initially practices a detection process based on the face detector provided by OpenCV which scans the retrieved frames of the input images which is in the form of a bit stream searching for a face to detect. The tracking algorithm determines the appropriate commands referring to the result of the detection process and transmits the commands to the PTZ camera where the commands controlling the camera are based on the command protocol used for the SONY EVI-D70 PTZ camera. The

application is specifically programmed to detect only one subject which indicates that the PTZ camera is determined to stop any movement when there is more than one face detected within the display screen.

2.3.2 Detection process

The proposed tracking system takes advantage of the face detector provided by OpenCV which implements the face detection technique commonly known as the Viola-Jones face detector [19]. OpenCV refers to this detector as the Haar classifier because it uses Haar features that adds and subtracts rectangular image regions before thresholding the result [6]. To allow rapid feature evaluation regarding the Haar features, the detection procedure of Viola-Jones introduces an image representation called the integral image.

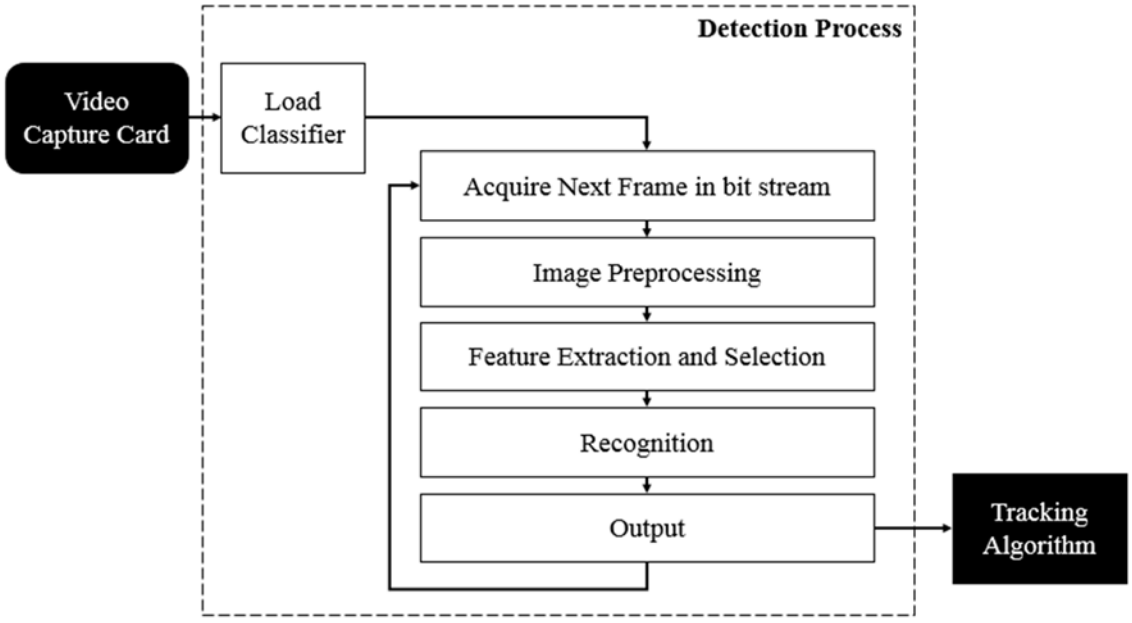

Figure 6: Detection process

The detection process consist of a series of steps performed on each frame to detect a face from the original input video images as shown in Figure 6. Initially, the detection process gains access of the video images from the memory and OpenCV stores each frame in a matrix format. Then, a classifier containing information regarding the features used for detection is loaded into the process. The image preprocessing step transforms the original image in RGB color to a grey scaled version and continues on to histogram equalization which balances the grey scaled image to a uniform distribution.

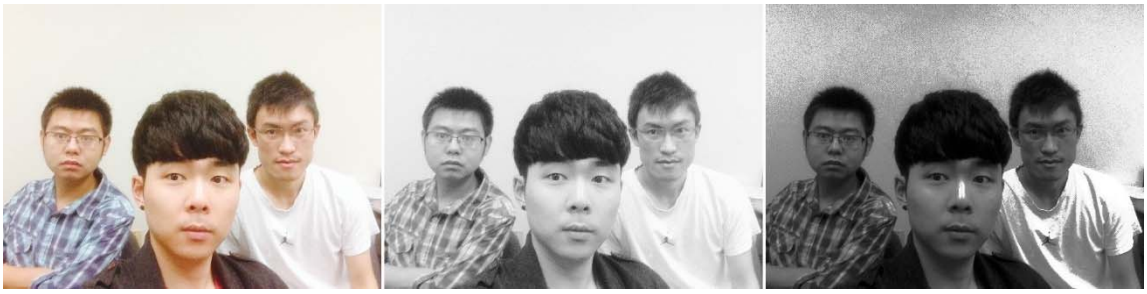

Figure 7: Result after image preprocessing

Image preprocessing is essential in terms of improving the accuracy before actually implementing face detection. The outcome of each preprocessing task applied to the original image is shown in Figure 7. The center figure shows the grey level transformation performed on the original image which is a process of reducing the impact of unconstrained illumination and computation. The step of transforming the original image into a grey scaled version does not only promise to reduce miscalculation and unwanted results but it also allows the classifier to work on the image due to the fact that

classifiers are known to work on grey scaled images. The last figure is the result after the histogram equalization step which maps the initial distribution of the intensity values to a distribution holding a wider and ideally, uniform distribution of intensity values to create a balanced histogram.

After the image preprocessing step, OpenCV initiates feature extraction and selection which refers to applying the classifier and identifying the most likely match of the face among the images that are sent from the camera. The classifier is in the form of an XML file containing a node with sub-nodes which are used to define the shape of the Haar features. Essentially, Haar features are black and white rectangles used together and each black and white patch represents a feature that the algorithm looks for within each frame of the input video image as shown in Figure 8.

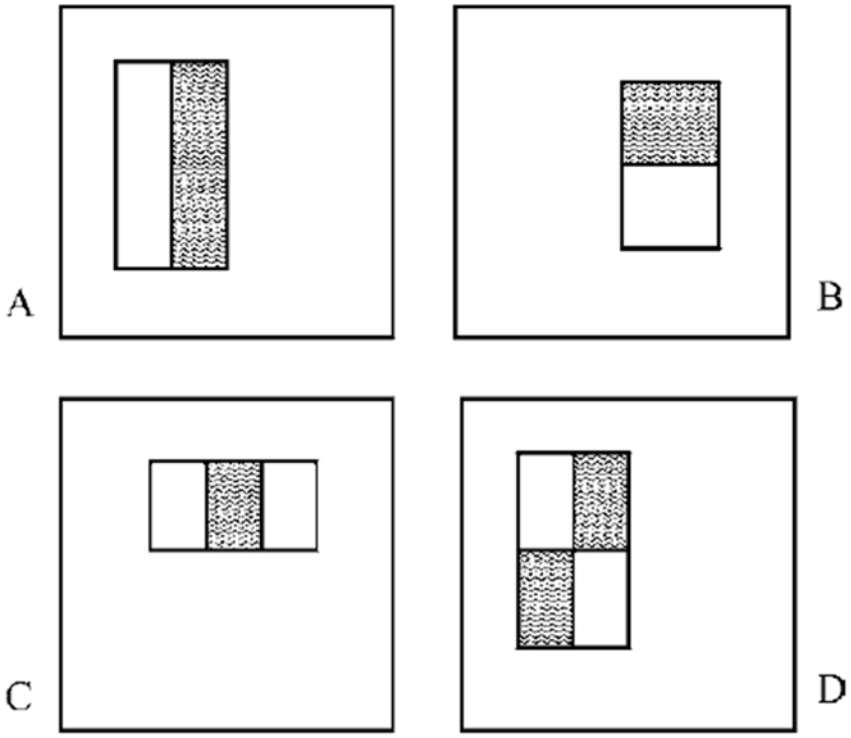

Figure 8: Example of rectangle features

The features of A and B shown in Figure 8 represent the two rectangle features which calculates the difference between the sums of the pixels with two rectangle regions. The following features in C and D represent the three and four rectangle features respectively.

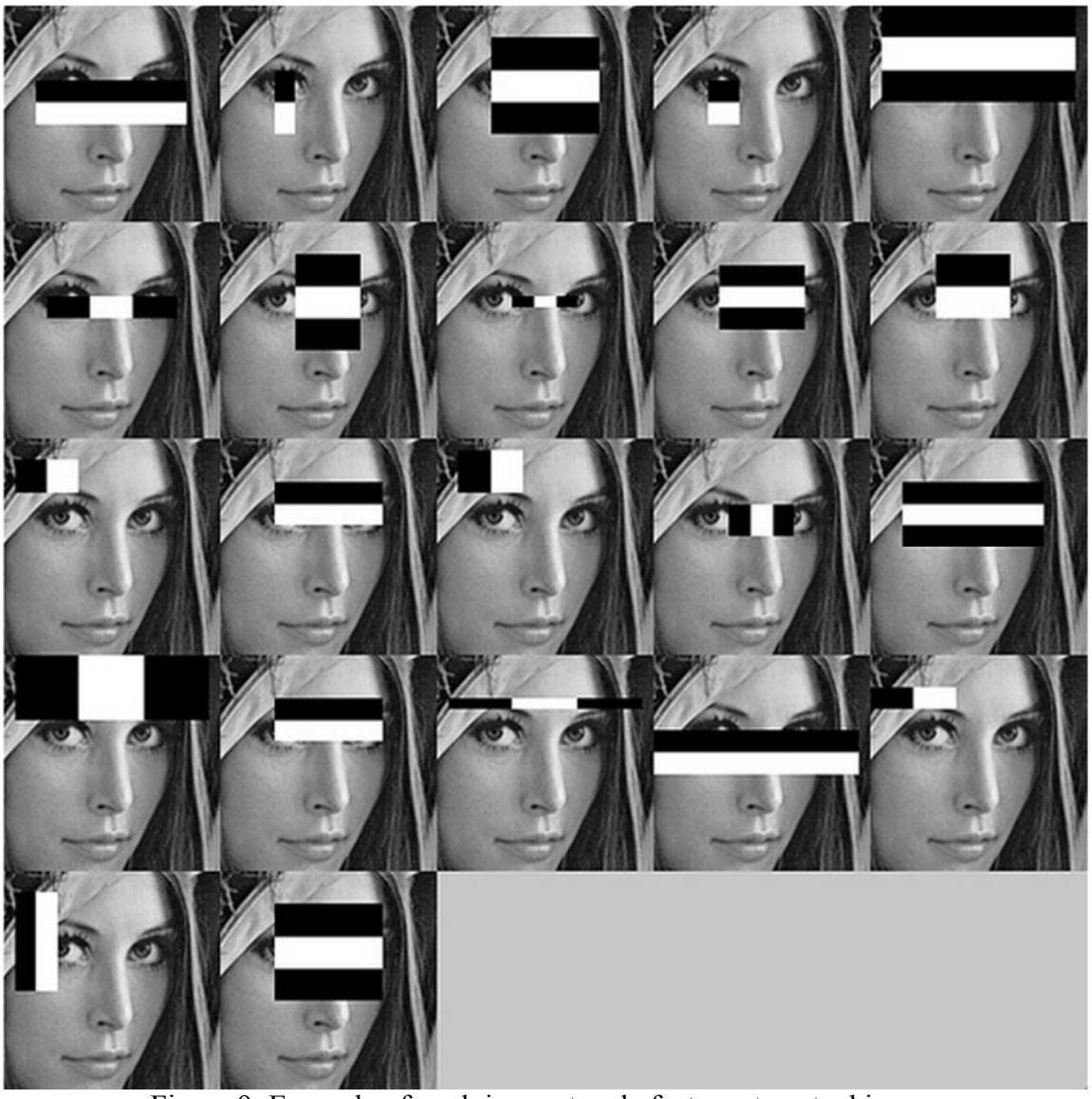

Figure 9: Example of applying rectangle features to actual image

Figure 9 shows an image with the features applied repetitively where each patch of the black and white rectangles is used to compare particular regions within the face to determine if the region corresponds to an actual face or not.

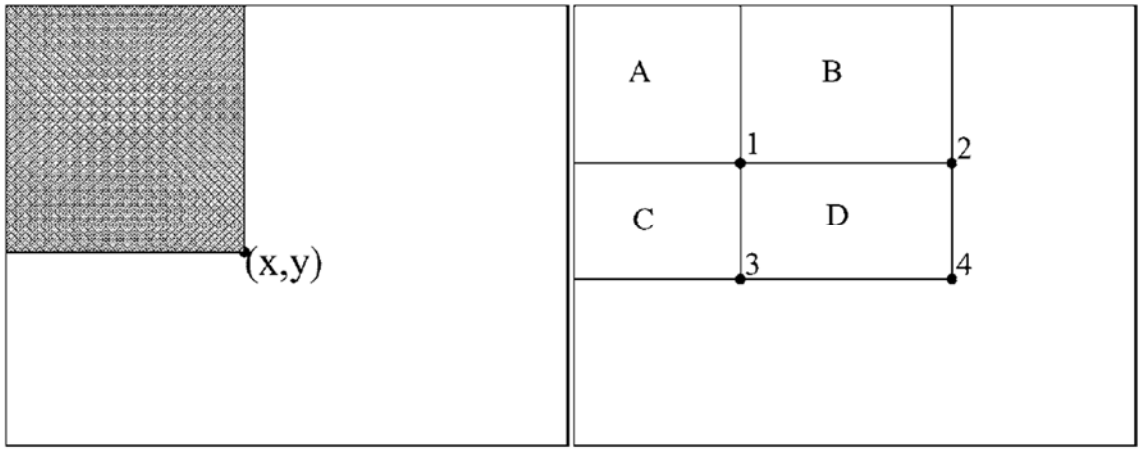

Figure 10: Representation of the integral image

To ensure rapid feature computation with respect to the rectangle features, Viola and Jones introduced the integral image representation [19] which is shown in Figure 10. The left figure depicts the basic concept where the integral image at a location in the 2-dimensional coordinate system is the sum of the pixels above and to the left of point x and y. Using the integral image representation, any rectangular sum can be computed in four array references. With rectangle D as the target of interest, the value of the integral image at location 1 is to be the sum of the pixels in rectangle A. The value at location 2 is the sum of A and B, location 3 is the sum of A and C, and location 4 is the sum of all four rectangles A, B, C, and D together. Thus, the sum within rectangle D would be 4 + 1 – (2 + 3). In other words, if the area of interest within an input image is rectangle D, the pixels within the rectangle are summed together with the image converted into a grey scaled version. Then, the sum of the grey scale values of the white region and the sum of the grey scale values of the black region are taken into comparison.

OpenCV implements a cascade classifier which combines more complex classifiers together in a cascade structure to increase the speed of the detector by focusing attention on promising regions of the image. Choosing the appropriate classifier is guaranteed to increase the possibility of successful face detection where the criterion is focused on the processing speed and the accuracy rate which is a critical requirement for real-time applications. Among the classifiers distributed by the OpenCV library, research [14] points out that the *haarcascade_frontalface_alt2* classifier produces the most promising result regarding face detection. To improve the detection rate, exercising multiple classifiers together promises a better result such that the combination of face and facial features used for detection can improve face detection by adding reliability to the traditional face detection approach. However, the approach requires more computational power due to the fact that additional processing is performed within each frame of the video image. Thus, the processing time is predicted to take longer if higher computational power is not available.

Face detection is performed by finding objects in different sizes within the input video images and the detected objects are returned as a list of rectangles which are affected by the parameters set in the actual detection function. The face detector provided by OpenCV allows to change parameters such as the scaling factor, minimum neighbors of rectangles needed for detection, and minimum/maximum size to ignore. In addition, background environments such as the lighting conditions, shadows on the wall, and obstacles are factors that additionally give effect to the performance of the detection process.

2.3.3 Tracking algorithm

The strategy of the tracking algorithm is to 1) keep the subject in the center region of the display screen with the panning and tilting features, and to 2) keep the projection of the subject at a proper size with the zooming feature as shown in Figure 11.

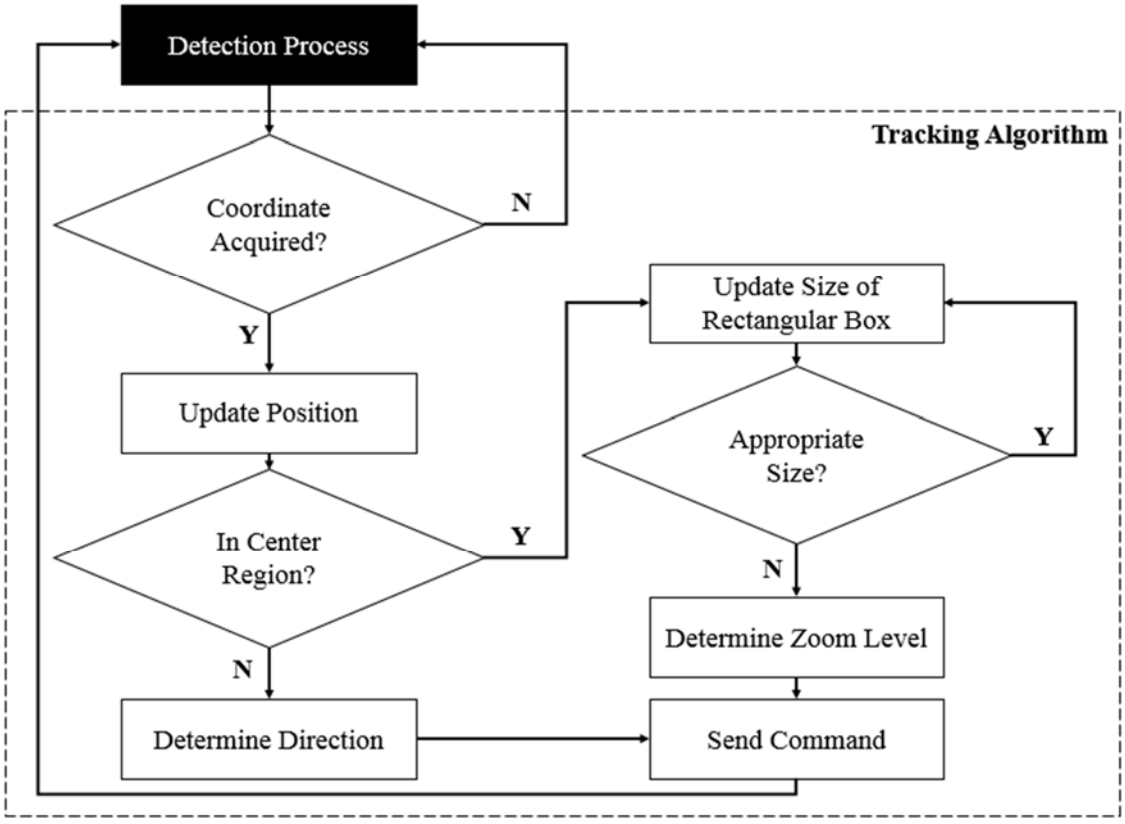

Figure 11: Structure of tracking algorithm

The basic structure of the tracking algorithm is set up to follow the center point of the face which is obtained through the detection process. This center point is determined by using half the width and half the height of the rectangular box placed over the face.

However, with the classifier of the face detector simultaneously identifying multiple faces, the camera is most likely to confuse which face to follow when there is more than one center point identified which leads to continuously switching the view between two or more faces. Thus, the algorithm is specifically set up to initiate tracking only when there is one center point which is equivalent with one face detected within each frame. In any other situation, the camera is demanded to stop movements such as panning, tilting, and zooming.

A straightforward method is used which accesses the coordinates of the center point in the 2-dimensional coordinate system. If the x and y coordinates of the center point is out of the defined center region, that is to say, if a face moves away from the middle of the display screen, the tracking algorithm determines that the camera needs to alter the angle using the panning and tilting features which brings the subject into the center region of the display screen. After the face of the subject is appropriately placed in the center region, the zooming feature is initiated. The tracking algorithm defines a range for the size of the projection of the detected face as a reference to control the level of the zooming feature. The zooming feature is activated by zooming in or zooming out when the size of the rectangular box projecting the face of the subject is smaller or larger than the desired size.

However, as the horizontal and vertical movement of the camera become sensitive while zoomed in, it is rather easy to lose a subject even with the same speed if the subject is in motion. If the camera is zoomed in too much and loses the subject during the tracking process, the PTZ camera intentionally zooms out progressively to search the premises for

the original subject. In other words, when the detection process fails, namely, with no values indicated in the x and y coordinates, the tracking algorithm commands to zoom out gradually from the current state to search for the initial subject. In most cases, this situation occurs when the subject suddenly moves in a certain direction with the camera already zoomed in. Depending on the status of zooming, the camera continues to zoom out from the current state until it finds a subject to track. Eventually, the zooming feature returns to the default state if the target has moved out of the entire FOV or, in this case, a class room environment.

2.3.4 PTZ camera controller

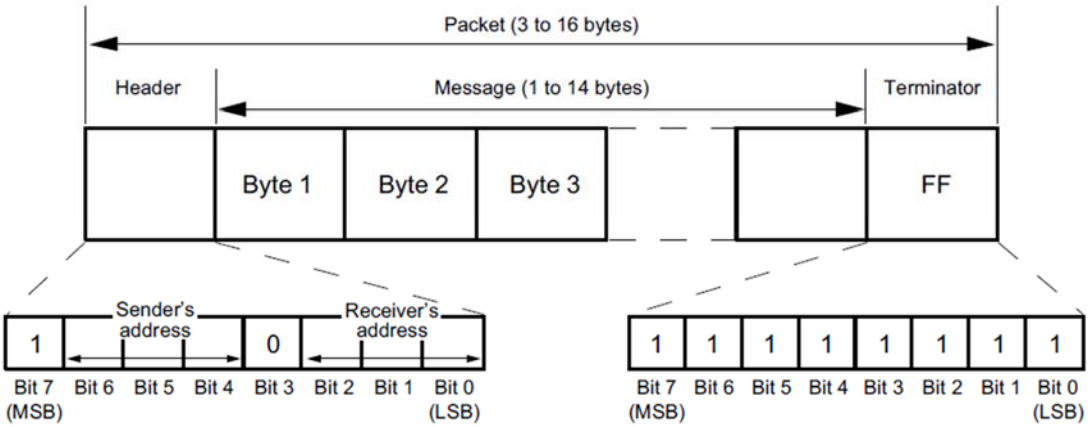

Figure 12: Communication protocol of UV83 and SONY EVI-D70 PTZ camera

Each command regarding the movement of the PTZ camera is in the form of a single packet consisting 3 to 16 bytes as shown in Figure 12. Each command is initiated with a header byte comprising the address of the sender and the address of the receiver

followed by the message bytes and finished by a terminator byte. The commands can be roughly classified into categories to control the camera or to inquire information from the camera. Focus of attention regarding the commands sent from the controller to the camera is concentrated in the message bytes which holds information related to the speed and direction of the camera movement.

Movement	Command (Message)
Up	8x 01 06 01 VV WW 03 01 FF
Down	8x 01 06 01 VV WW 03 02 FF
Left	8x 01 06 01 VV WW 01 03 FF
Right	8x 01 06 01 VV WW 02 03 FF

Table 2: Command examples of the UV83 PTZ conference camera

An example of the commands corresponding to the actual movements of the PTZ camera is shown in Table 2. The byte '8x' corresponds to the header byte where x is distinguished as the address of the receiver. VV indicates the speed of the panning feature which can be chosen between 01 and 18. Similarly, WW appoints the speed of the tilting feature which has the option to choose between 01 and 17. Byte 6 controls the panning direction where 01 indicates to move left, 02 to move right, and 03 to halt horizontal movements. Similar to the panning feature, byte 7 controls the tilting movement where 01 commands to move up, 02 to move down, and 03 to stop vertical movements.

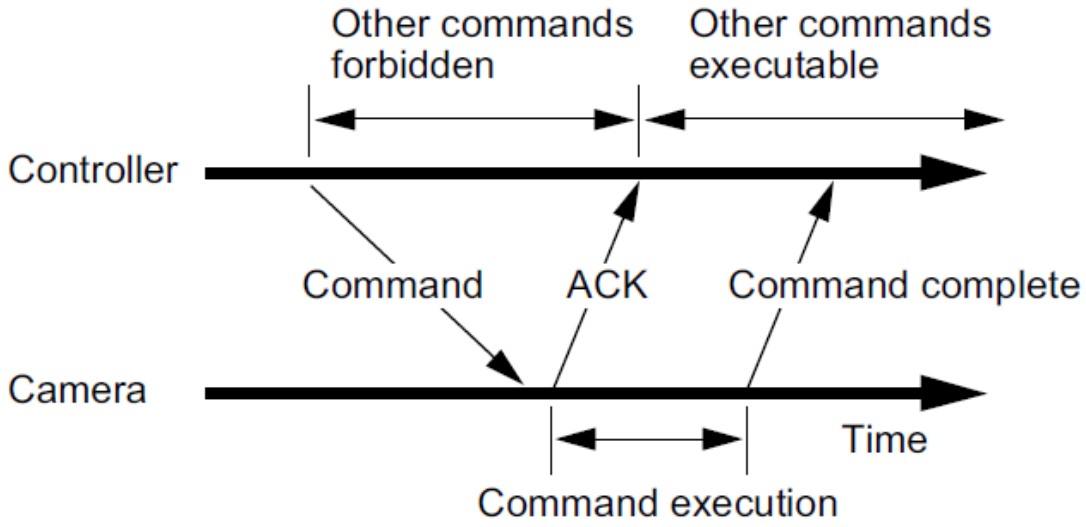

Figure 13: Flow of information transmission for a single command

For each command received by the camera, an ACK (Acknowledgement) signal is transmitted to the controller, or the computer, to ensure the command has been transmitted properly as shown in Figure 13. From the point of a command sent from the controller to the camera, any other command is forbidden to interfere until the ACK signal is transmitted back to the controller. In the case of executing more than two commands, the command queue is possible to overflow and eventually ignore later commands unless the commands are promptly executed which refers to the ACK signal being transmitted back to the controller in a very instant interval [25]. This is an essential requirement related to the performance of the camera regarding the delay of the camera movements to ensure continuous tracking with a smooth movement while tracking a subject.

2.4 Preliminary test results

An evaluation test was conducted on a desktop with an Intel Core 2 Duo 3.16 GHz processor and a 4 GB RAM with the UV83 PTZ conference camera locally connected to assess the functionalities of the preliminary components. Figure 14 shows the system in operation with the preliminary components implemented where the system utilized approximately 40% of the CPU throughout the evaluation test.

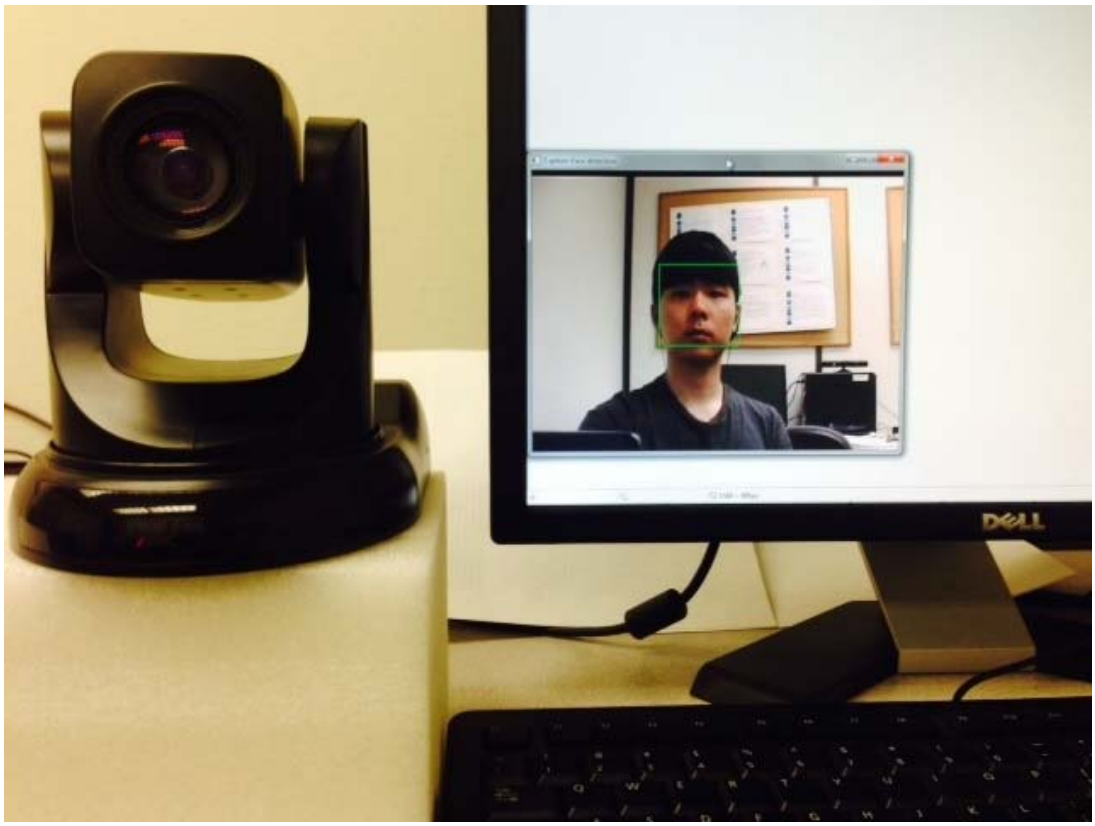

Figure 14: Evaluation test result of system in operation with preliminary requirements

The system operated at an average frame rate of 22 frames per second with frames of 720x576 pixels. Operation for each frame of the detection process took an average of 40 ms while a single tracking command regarding the panning, tilting, or zooming feature each took an average of 50 ms to execute.

Figure 15 shows the result of a demonstration regarding the detection process and the tracking algorithm using the preliminary system. The camera enables the panning and tilting features to bring the detected face into the center region as the subject is initially detected from the side of the FOV. As the subject starts moving further away in distance, the zooming feature initiates since the size of the projection of the subject is determined to be smaller than the intended size. The PTZ camera zooms into the face to enlarge the projection size to a proper size regarding the rectangular bounding box over the detected face. The cabinet in the background is considered as a criterion determining the level of the zooming feature since the size of the face is intended to maintain a certain size throughout the process. The subject then progressively walks toward the camera which increases the size of the rectangular box over the face. The system immediately acknowledges the situation and gradually zooms out of the current state to bring the projection of the face back to a proper size.

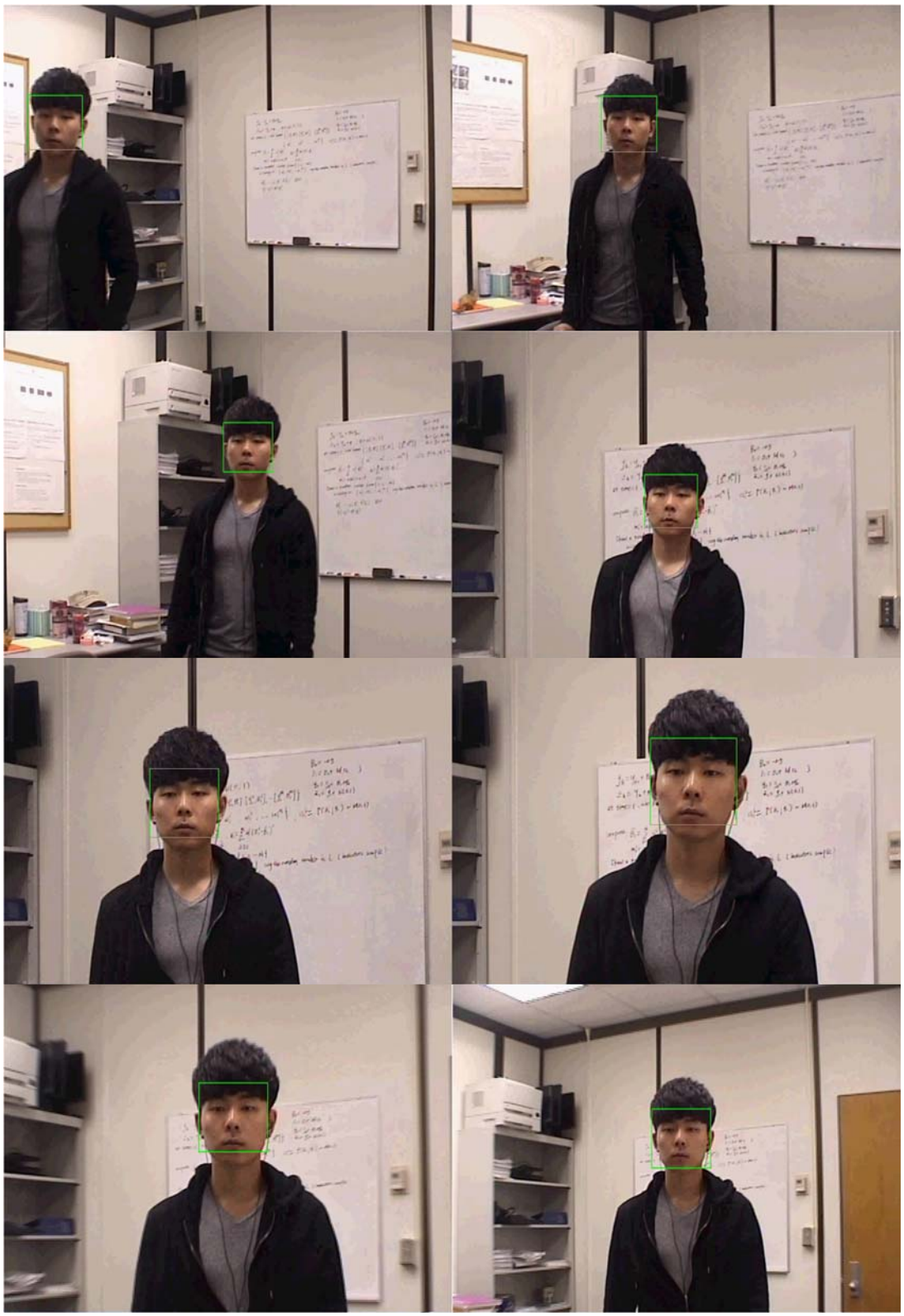

Figure 15: Demonstration of the detection process and the tracking algorithm

2.5 Discussion

The main objective of the evaluation test was to realize existing methods and algorithms related to real-time tracking based on the face detection technique with a single PTZ camera. With wired connections and primitive hardware components, the preliminary system successfully established a basis for the proposed system. The system was capable of functioning automatically without in need of an operator to control the movement of the camera.

The focus of attention pointed out in the introduction has been satisfied such that the reaction time of the camera being approximately at 50 ms ensured to continuously keep pace with the subject in real-time while the combination of the features provided by the PTZ camera showed that the transition of the movement was smooth and natural.

However, it is necessary to seek improvements before implementing the proposed system. The development of computer systems manufactured nowadays needs to be taken into consideration since the preliminaries are subject to hardware equipment manufactured in the earlier days. Initially, the proposed system needs to consider expanding the options regarding hardware components such as the connection availability since the DB9 serial communication port is no longer available on computer systems manufactured nowadays. In addition, as broadcast is using larger and higher quality displays and considering the artifacts of the CVBS signal, the software application of the proposed system needs to deliberate on the capability of processing video image formats in higher definition compared to the composite video format.

Aside from improvement suggestions, a noticed limitation related to the performance of the preliminary system was the effect of luminance on the detection process. The brightness condition is an important factor since the classifier loaded into the detection process works on data with grey scaled images which implies that the image is processed according to the brightness. In environments with imperfect lighting conditions, the false positive detection rate is possible to increase which indicates the possibility of the face detector understanding a non-face object, or a shadow, as a face. Thus, a method to improve the detection rate in such environments with inconsistent or imperfect lighting conditions is in need.

3. PROPOSED SYSTEM CONFIGURATION

3.1 Overview

To assist online education platforms, the proposed system is implemented as a part of the platform as shown in Figure 16. The basic structure is maintained to be based on a single PTZ camera with a computer unit capable of processing the software application. While retaining the framework established with the preliminary components, the proposed system introduces an encoder to transfer high definition images between the PTZ camera and the computer as it is shown in Figure 17 and Figure 18 respectively. Fundamentally, the primary objective is to not only verify the functionality of each process in the software application interacting with the hardware components but also validate firm transition respect to the flow of data to ensure smooth operation.

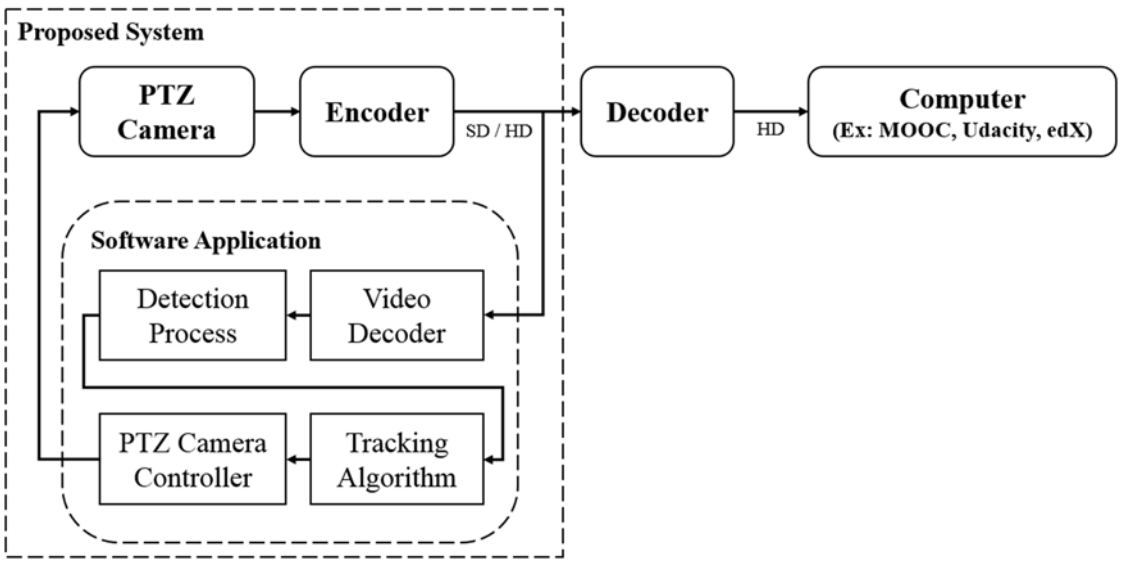

Figure 16: Proposed system implemented in online education platforms

Figure 17: Proposed system in operation

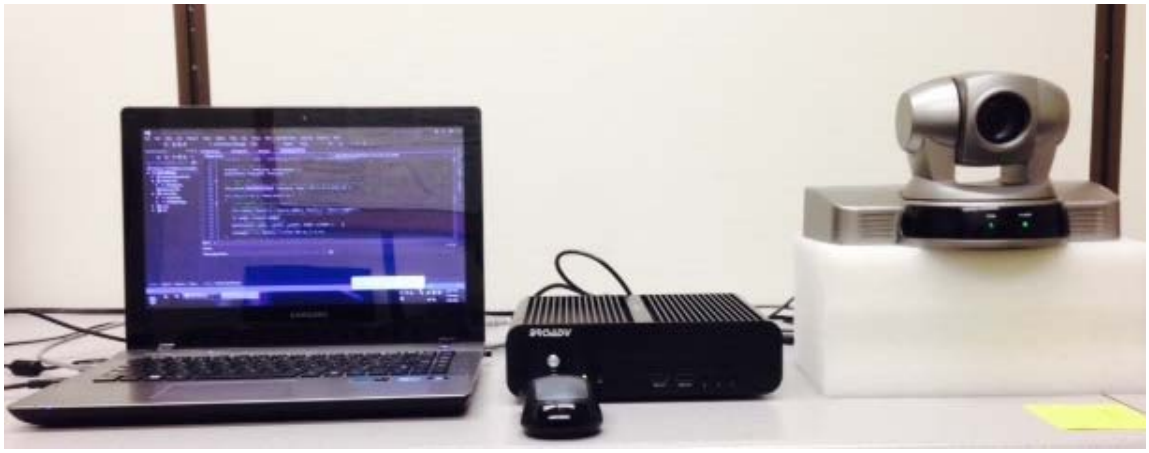

Figure 18: Frontal view of proposed system

The PTZ camera initially passes on the video images in an uncompressed format to the encoder for the purpose of compression before transmitting to the computer unit. The video decoder within the software application retrieves the images in a compressed format and initiates the detection process after decompressing the video images back to the original format. The process of the tracking algorithm and the PTZ camera controller are identical to those described in the preliminary system. The tracking algorithm determines the appropriate commands regarding the movement of the PTZ camera referring to the information received from the detection process and the camera controller transmits the commands to the PTZ camera. The fundamental task is to maintain the subject within the center region of the display screen during operation.

3.2 Hardware implementation

3.2.1 Camera specifications

The proposed system introduces a donated BLM-500BH IP PTZ camera to the system allowing the availability to distribute access through the internet. The BLM-500BH IP PTZ camera holds specifications similar to the SONY EVI-HD1 PTZ conference camera where the specifications of the two cameras are shown in Table 3. As the major features function similar to each other, the difference in the specifications are predicted to merely give an effect to the performance.

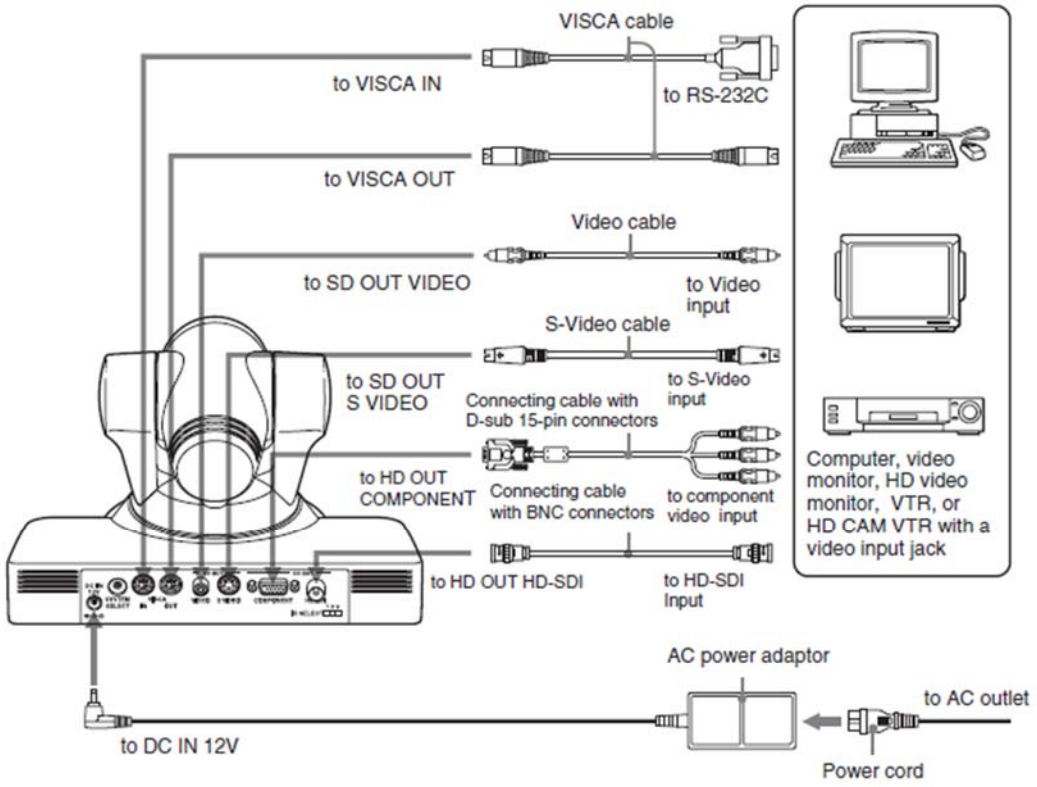

Figure 19: Options of connections and corresponding cables of SONY EVI-HD1 PTZ camera

Camera	Panning	max(Speed)	Tilting	max(Speed)	Zoom	Focus
BLM-500BH	±170º	100º/sec	+90º, -30º	90º/sec	18x, 12x	Auto
SONY EVI-HD1	±100º	300º/sec	±25º	125º/sec	40x, 10x	Auto

Table 3: Comparison of BLM-500BH and SONY EVI-HD1 PTZ specifications

Figure 19 shows the options considering the connections and the corresponding cables of the SONY EVI-HD1 PTZ camera. The BLM-500BH utilizes identical cables corresponding to the connection ports of the SONY EVI-HD1 PTZ camera. The BLM-500BH PTZ camera provides a variety of choices regarding the format of the output image which are the composite video format, HD-SDI, and S-Video. The proposed system approaches the SDI format to improve the quality of the output image displayed and utilizes the RS232 8-pin mini DIN port to maintain the control protocol.

3.2.2 Connection

As the BLM-500BH PTZ conference camera holds similar connection options to the SONY EVI-HD1 PTZ camera, it secures the 8-pin mini DIN port connection to communicate with the computer which is identical to the UV83 PTZ conference camera used in the preliminary system as shown in Figure 20.

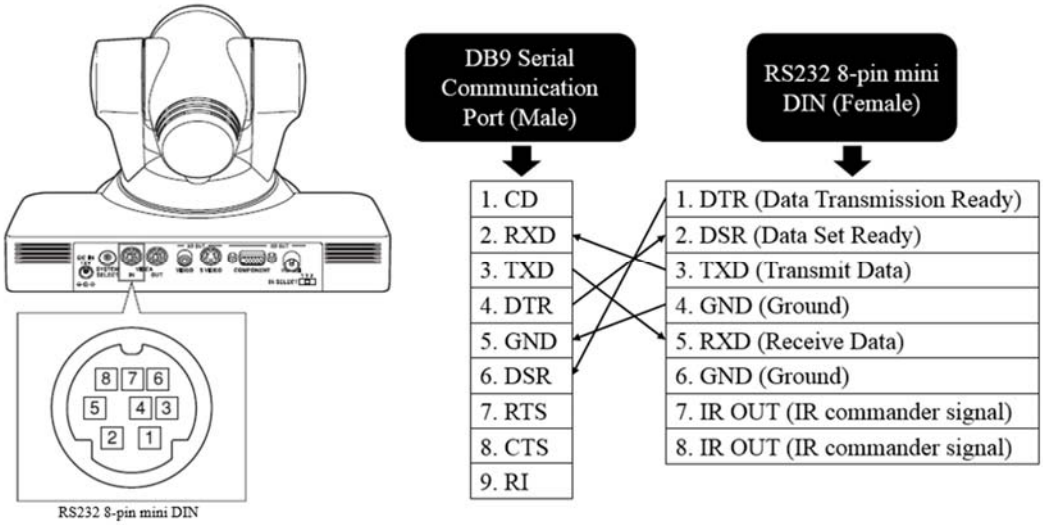

Figure 20: SONY EVI-HD1 PTZ conference camera pin-out

The proposed system takes into consideration the fact that DB9 serial communication ports are no longer commonly used but insists on maintaining the connection protocol. Thus, the proposed system expands the options of the connection availabilities by accessing the USB (Universal Serial Bus) port which is commonly used. The system utilizes a DB9 to USB connection cable to not only preserve the RS232 communication protocol but also to use the USB port as a connection option. Essentially, the DB9 to USB cable is considered to be a convertor capable of connecting to computer systems without the DB9 serial communication port but still function equivalently with the RS232 communication protocol. The advantage of using the USB port is considered to be the low power consumption and being powered from the USB connection which also reduces signal loss and improves the quality of transmission.

3.2.3 Encoder

The proposed system integrates a LeC200 multi-interface encoding video capture system with an embedded IP71 encoder which has been donated. The video capture system supports HD-SDI, CVBS, VGA, and HDMI (High Definition Multimedia Interface) inputs while capable of encoding each signal. Image scaling is performed on the resolution of the output image regarding the signal, resolution, and frame rate of the input image while the UDP (User Datagram Protocol) is supported to transmit the output stream. The actual encoder, IP71, is capable of performing H.264 video encoding at a bit rate in the range of 50 Kbps to 8 Mbps. The encoder also supports AAC (Advanced Audio Coding) audio encoding at a bit rate in the range of 16 Kbps to 64 Kbps.

The proposed system integrates an encoder to compress the high definition video image format and transmits the compressed video bit stream to the computer unit before processing through a video decoder associated with the software application. The encoder is embedded to a video capture system as shown in Figure 21 which is essentially a video compressor which converts digital video images into a format that can be stored or transmitted while taking up less capacity. The compressed bit stream is produced through prediction, transformation, and encoding processes by the encoder embedded to the capture system.

Figure 21: External view of video capture system with an embedded encoder

The IP PTZ camera connects to one of the four SDI-IN ports on the video capture system through a video cable with BNC ports on each end while the Ethernet cable accomplishes the task of transmitting the compressed video image to the computer in accordance with the UDP after compression is processed.

3.2.4 Video format

The proposed system is capable of managing SDI which is an uncompressed digital video format. SDI which is also known as Serial Digital Interface is a family of digital video interfaces capable of sending both video and audio through a single connection. In order to process the stream of high resolution images, an encoder utilizes the LAN (Local Area Network) with an Ethernet cable to transmit data to the computer while a video decoder application runs in real-time. The resolutions possible to display are 640x360 and 1280x720 pixels which can be chosen depending on the UDP address. The signals are uncompressed and are self-synchronizing in between the transmitter and the receiver.

The proposed system mainly focuses on processing images in the resolution of 640x360 pixels considering the possibility of using computer systems with moderate computational power which might not be efficient enough to process high definition images in 1280x720 pixels for each frame.

3.3 Software implementation

3.3.1 Overview

The basic structure of the software application of the proposed system is based on the preliminary software application with an additional detector included in the detection process and additional features in the tracking algorithm. The software application is divided into the detection process and the tracking algorithm which are processed simultaneously as the system initiates. On the basis of the software application established

with the preliminary system, the proposed system includes a video decoder to decompress the video images received from the encoder capture system.

3.3.2 Connection

The proposed system utilizes the UDP with an Ethernet cable to access the input video image sequences in real-time. Typically, the TCP (Transmission Control Protocol) is more frequently used compared to UDP since it is the most commonly used protocol on the internet. TCP and UDP are both considered as layer components providing a connection point for applications to access network services. Both protocols use IP (Internet Protocol), which is a lower-layer best effort delivery service which delivers encapsulated TCP packets and UDP datagrams across the internet. The advantage of UDP is not only using IP to transport information but also adds the capability of delivering to multiple destinations given a host computer. The proposed system takes advantage of UDP which is commonly used for streaming audio and video. Usually, data sent over the internet is affected by collisions and errors are to be present. TCP is normally used to transmit important data since it provides flow control and error correction [12]. Flow control determines when data needs to be sent again, and stops the flow of the data until previous packets are successfully transferred. This method is implemented due to the possibility of a collision occurring when a packet of data is sent. When a collision occurs, the client requests the packet from the server to be transmitted again until the entire packet is complete and identical compared to its original packet. UDP is preferred over TCP for the purpose of streaming since it offers faster transmission speed due to the absence of

error correction and flow control. Even though UDP can be used in networks where TCP is normally used, reliability or correct data sequencing cannot be guaranteed for the same reason of not having error correction and flow control. This indicates that data may not be identical to its original or arrive in a different order due to the possibility of being affected by collisions and errors.

The distinguished UDP address used to access the video images from the video capture system is paired with an IP address to transport the images to the computer through an Ethernet cable. In order to transmit datagram using UDP, the network address of the network device hosting the service and the UDP port number that is used for communication is in need. To carry out this procedure, the system takes advantage of the Teredo Tunneling Pseudo-Interface which is a tunnel adapter for Teredo, a method of transmitting IPv4-encapsulated IPv6 packets across a network address. The link-local address of the Tunnel adapter is in a (Prefix::Server IPv4:Flags:Port:Client IPv4) format. The link-local address is a network address that is only valid for communications within a broadcast domain with the host connected. Teredo enables nodes located behind a number of IPv4 NATs (Network Address Translator) to obtain IPv6 connectivity by tunneling packets over UDP [13]. With IPv6 being introduced as a solution to the depletion of IPv4, research [4] shows that the peak performance of IPv4 is similar to IPv6.

3.3.3 Video decoder

The proposed system integrates the open source VLC player into the system for the purpose of decoding the compressed input video streams received from the video

capture system. The open source VLC player is capable of streaming MPEG-1, MPEG-2, MPEG-4 files, and also live videos on a network with the option of unicast or multi-cast. The main purpose of adapting the VLC player is to receive, decode, and display video streams over a network even under multiple operating systems.

To decode a stream of the input video image, the VLC player library initially demuxes the stream which is a process of reading the container format and separating video, audio, and subtitles. A container format contains streams encoded by codecs which are compression algorithms which intend to reduce the size of each stream. After the stream is read, each are processed through decoders that performs mathematical processing to decompress the streams transmitted from the video capture encoder system. In a video stream, container formats always comprise the video and audio signal together. The demuxers extract the streams from the input video stream and passes it to the decoders.

The video output modules allow the VLC library to display video on the display screen by guessing the most suitable video output module for the system. However, it is also possible to force a specific module to display the video stream. The video filter modules perform modifications on the rendered image including de-interlacing, hue/contrast/saturation adjusting, cropping, etc.

3.3.4 Detection process

As the encoder compresses the input video images in a compact form to increase the speed of transmission, it is in need of a video decoder application to return the input

video images back to the original bit stream before the detection process takes action. The decoder carries out the complementary processes of decode, inverse transform and reconstruction to produce a decoded video sequence. The decoder implements on the basis of the VLC player which is known for processing high definition video images.

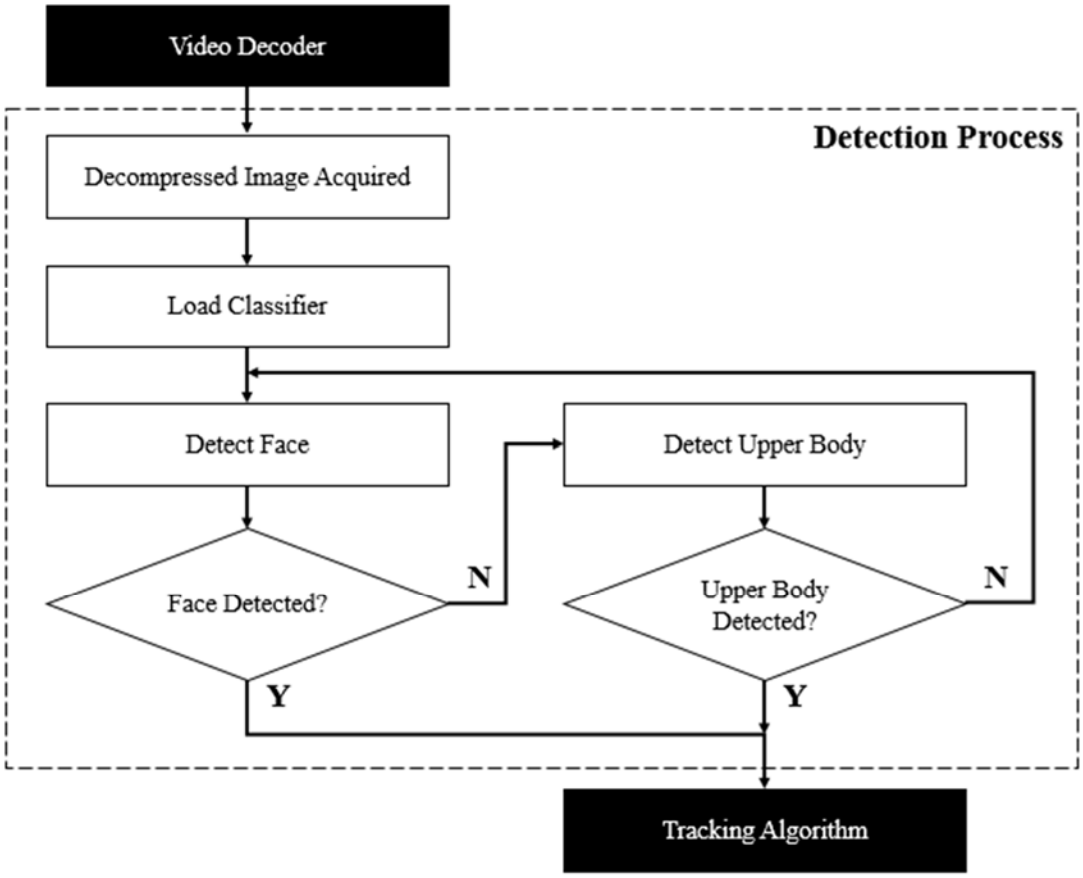

Figure 22: Detection process using face and upper body detector

While maintaining the processing speed of the detection process, the proposed system includes an additional upper body detector with the initial face detector to increase

the detection rate. A detailed description regarding the detection process using two detectors is shown in Figure 22. Since simultaneously processing two classifiers is promised to decrease the speed of detection, the classifiers are used back-to-back which implicates that each detector is initiated only when the other is unsuccessful to detect a subject. In other words, the upper body detector is only in action when the face detector fails to detect a face within the FOV and vice versa. Essentially, the upper body detector is used to extend the possibility of detection even when the subject faces away from the camera which indicates that the face detector is likely to fail without the full frontal view of the face. Thus, the upper body detector is considered to be a substitute for tracking a target when the face is not visualized to the PTZ camera.

3.3.5 Tracking algorithm

The tracking algorithm used for the proposed system takes advantage of the preliminary tracking algorithm. As the preliminary tracking algorithm malfunctions when the display changes from 640x360 pixels to 1280x720 pixels and vice versa with the center region fixed at a particular size, the tracking algorithm of the proposed system considers the possibility of the pixels of the display image changing since the system tested with the preliminary components only considered a fixed number regarding the pixels. Thus, the tracking algorithm improves by allowing the center region in the display screen and the desired range of the projection size to change according to the resolution of the input video image.

In addition, while a default speed was used for the panning and tilting feature in the preliminary tracking algorithm, the speed of the camera for the proposed system is determined by comparing the difference between the current and previous position of the target. The tracking application increases the speed of the panning and tilting feature when the difference of consecutive points are larger than a certain degree and decreases back to the default speed in any other case. Whilst there exists an application [5] implementing eight directions to control a PTZ camera, the proposed system only processes four out of the eight to reduce the complexity which are left, right, up, and down.

3.4 System test results

The proposed system is evaluated on a laptop with an Intel Core i5 2.5 GHz processor and a 6 GB RAM with the BLM-500BH IP PTZ conference camera connected to the system where the system utilized approximately 20% of the CPU during operation. The system operated at an average frame rate of 30 frames per second based on the IP PTZ camera processing a resolution of 640x360 pixels. The processing speed improved to taking approximately 30 ms for face detection and an additional 10 ms for upper body detection. The processing time for each tracking command was found to be significantly faster at 25 ms compared to the preliminary system.

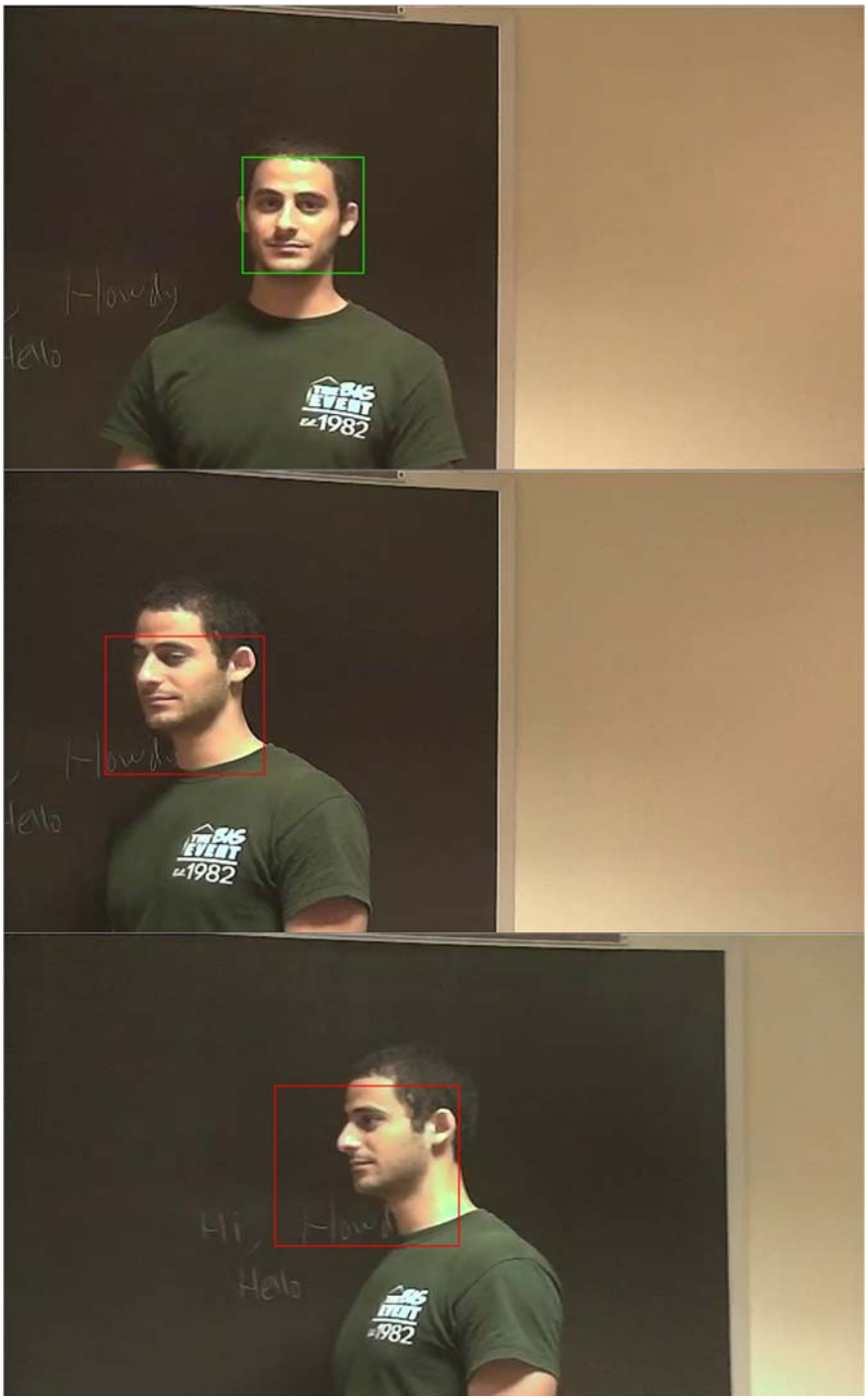

Figure 23: Evaluation test result of face and upper body detection process and tracking algorithm

The result of face and upper body detection with tracking based on the proposed system is shown in Figure 23. The first figure represents the state of the subject being initially detected with the face detector. The subject turns to the right which initiates the upper body detector since the face detector fails to perform without a full frontal view of the face. The upper body detector is distinguished by the rectangular box covering the region starting from the shoulders up to the head whereas the face detector specifically concentrates on the region over the face. At this point, the preliminary detection process is most likely to lose the subject and the rate of false detection is possible to increase. The second image shows the upper body detected from the side of the display screen while the last image represents the moment after the tracking algorithm commanded the camera to move, bringing the detected upper body into the central region of the screen. This example evaluates the robustness of the tracking algorithm implementing an additional upper body detector used together with the original face detector.

The demonstration shown in Figure 24 elaborates on the zooming feature used for the proposed system. The first figure in the upper left corner presents the initial detection respect to the face detector. The tracking algorithm understands that the position of the face is out of the center region and commands the PTZ camera to move. As the position of the face detected approaches the center region of the display screen, the camera begins to zoom in until the size of the rectangular box reaches the range indicated by the tracking algorithm. The current level of the zooming feature is determined referring to the size of the blackboard used as a criterion. The subject then turns away from the camera which indicates that the subject is most likely writing on the blackboard. After the subject faces

away and neither face nor upper body detection is successful, the camera begins to progressively zoom out from the current state of the zooming level understanding that the subject does not exist within the FOV. As the camera fails to detect the original subject neither a new subject while zooming out progressively, the zooming feature eventually returns to the default state while the subject continues to face away from the camera.

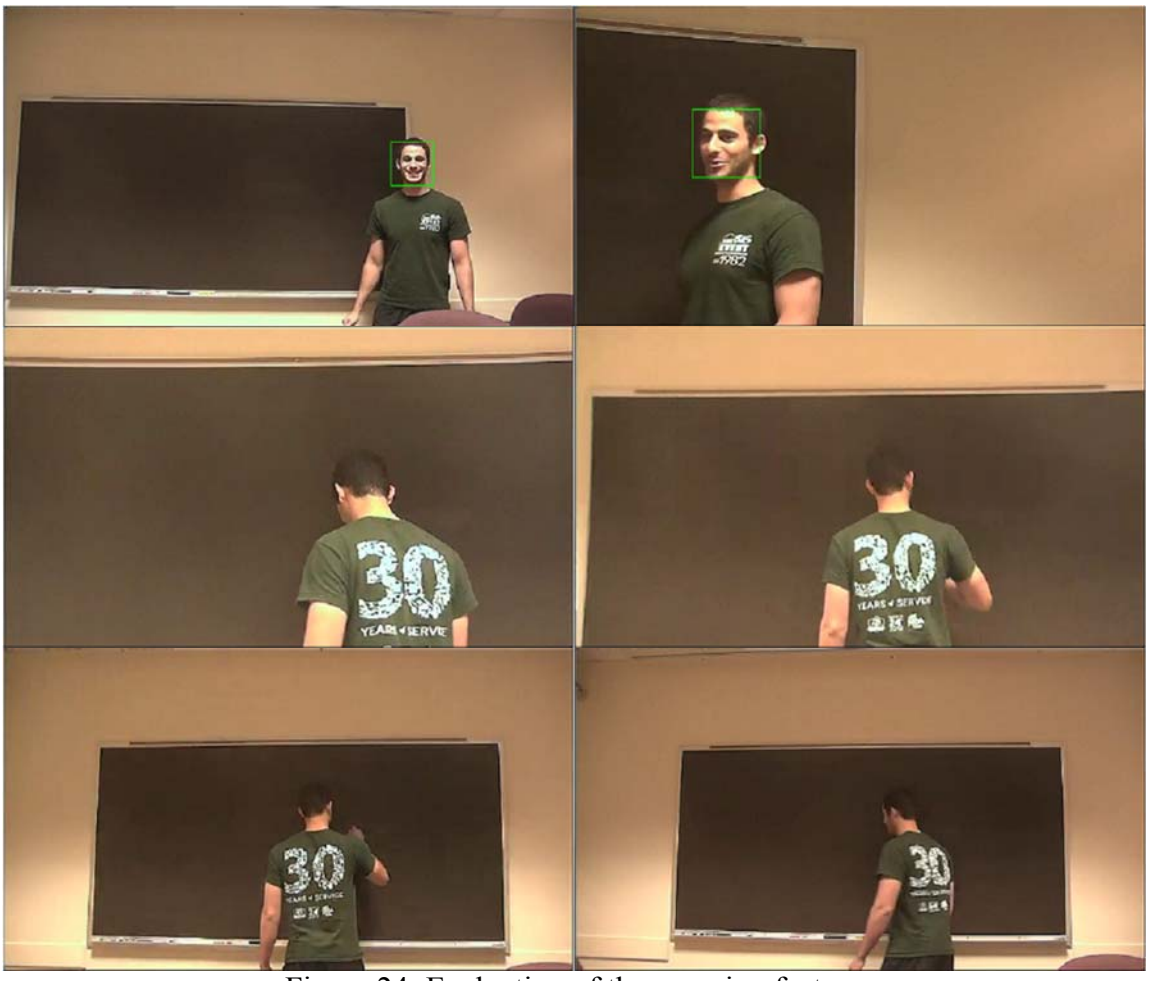

Figure 24: Evaluation of the zooming feature

The example shown in Figure 25 is a demonstration mimicking a lecture in session where the lecturer is in freedom regarding movements such as direction and speed.

Figure 25: Performance evaluation of proposed system mimicking a lecture

The lecturer starts out being detected from the side of the black board. The tracking system initiates and continues to follow the subject detected. The camera briefly zooms into the subject due to the size of the box over the detected face being smaller than the desired size. The size of the black board in the captured images performs as a criterion to determine the level of the zooming feature. The subject changes the direction of movement while the camera continues to track but eventually loses the subject. At this point, the tracking algorithm determines to stop the PTZ camera from moving. Without a subject to track, the system starts zooming out gradually from the current state to search the premises. Eventually, the camera detects the face of the initial lecturer while zooming out gradually and continues to perform according to the tracking algorithm.

3.5 Discussion

Maintaining the hardware structure of the system similar to the system in the preliminary test resulted in allowing the two PTZ cameras to interoperate. The proposed system is capable of using the UV83 PTZ conference camera as an alternative to the BLM-500BH PTZ conference camera. The proposed system maintained the method of connection by utilizing the RS232 communication protocol but adapted the USB connection which ensured connecting to systems without the availability of the DB9 serial communication port. The proposed system also expanded the format of video images possible to process from CVBS to SDI which are both commonly used video formats.

The advantage noticed from the proposed system was the improved performance of the detection process by including the additional upper body detector which increased

the probability of detecting an instructor even without the full frontal view of the face available. In addition, the movement of the camera tracking a subject was rapid and smooth even with only four out of eight directions implemented.

The method of the zooming feature gradually zooming out from the current state in the case of losing a subject was proved to be effective to find the initial subject. Even with the objective to keep track of a subject disregarding the movement or speed, the strategy of re-tracking a lost subject was proven to be successful by zooming out progressively and searching the premises.

System	Camera	Image	Frame Rate
Intel Core 2 Duo 3.16 GHz	UV83 conference camera	720x576	≈ 22 frames per second
Intel Core i5 2.5 GHz	IP PTZ camera	640x360 1280x720	≈ 30 frames per second
Pentium Dual Core 3 GHz	SONY EVI-D70P	320x240	≈ 25 frames per second
Intel Quad Core 2.8 GHz	SONY EVI-HD1 SONY EVI-D100	720x360 720x486	≈ 8 frames per second
Intel Xeon 3.0 GHz	SONY SNC-RZ30N	640x480	≈ 15 frames per second
Intel Xeon 5150 2.66 GHz	SONY SNC-RZ50N SONY SNC-RZ25N	640x480	≈ 30 frames per second
Pentium 4 3.0 GHz	SONY EVI-D31	352x288	≈ 25 frames per second
HP Z600 Dual Quad Core 2 GHz	AXIS 214	704x480	≈ 12 frames per second

Table 4: Performance evaluation

Table 4 presents a performance evaluation comparing the system established with the preliminary components and the proposed system to conventional systems [3, 10, 18, 21, 24, 26]. The first two results represent the test performed with the preliminary components and the proposed system respectively. It is noticed that with comparable or, in some cases, less computational power available, the performance of the proposed system is competitive compared to conventional systems. Considering most systems processed video images with a resolution at 640x480, the proposed system is capable of producing competitive results regarding the frame rate even with higher resolution images.

4. CONCLUSION AND FUTURE WORK

4.1 Conclusion

In this thesis, a configuration of an automatic face tracking system functioning in real-time is proposed. The framework consists of a single PTZ camera connected to a computer system with moderate level computational power. The aspect of attention is to continuously detect and track an instructor in a typical classroom environment. Furthermore, the proposed system is intended to be a starting point of developing cost effective online education platforms providing competitive performance compared to conventional systems. However, the main concentration is to establish a software application performing detection and tracking while rotating the data flow among the hardware components and the software application. In addition, minimizing the processing time relating to the delay caused by the camera movements and the network delay has been a concentration point throughout the process of implementing the proposed system.

The proposed system initially established preliminaries to confirm the functionality of a real-time face tracking system. The preliminaries focused on universal hardware components and video formats to ensure operation on moderate level computers. Our test results proved that the system comprised with the preliminary components are capable of performing as well as conventional systems but needed to consider improvement regarding the connection options, video formats possible to process, and false positives detected during the detection process.

The proposed system expanded the connection options by adapting the USB port but continued to maintain the connection method of using the RS232 communication protocol. Essentially, while retaining the method of transmitting commands to the PTZ camera according to the RS232 communication protocol, the proposed system simply introduced a DB9 to USB converter to provide an alternative option as a connection. In addition, the proposed system stretched the video formats possible to process from the composite video format, or CVBS, to SDI which is a higher quality image format. Finally, the limitation of detecting false positives during the detection process has been minimized by using a face detector and upper body detector together based on the OpenCV library. The detection process implemented the two separate classifiers back-to-back rather than simultaneously which extended detection regarding an instructor even when the complete frontal view of the face was not available to the camera. After a subject was detected through the detection process, the tracking algorithm determined the appropriate movement commands for the camera to maintain the position of the subject within the center region of the display screen disregarding the initial position, movement, and speed of the instructor.

As briefly stated in the introduction, one of the major concerns regarding conventional systems was the delay regarding the movement of the PTZ camera since this is a critical factor affecting the performance. A movement delay is possible to occur if the detection time is too long or each command is not promptly executed. If either one of the two possibilities occur, it is most likely to fail to continuously track in real-time. The proposed system operated with a processing speed of approximately 40 ms for face

detection and an additional 10 ms for upper body detection while each tracking command processed at an average speed of 25 ms. The processing speed of the proposed system was compared to the system with the preliminary components which ensured the delay in the movement was minimized. Cheng et al. [20] presented a real-time face tracking system with performance at 30 to 45 frames per second regarding the frame rate and an average of 100 ms regarding the face detection process. Even with a high frame rate, the system requires a large amount of time for the detection process compared to the proposed system. In addition, Iraqui et al. [1] proposed a system for face detection and tracking which processes detection and tracking on an average of 200 ms in total with the frame rate being approximately at 10 frames per second. The system implements a combination of an omnidirectional camera and a PTZ camera together which ensures high quality images as an output but increases the required cost regarding hardware equipment while performing worse compared to the proposed system. It is rather difficult to compare the processing time of every system related to real-time tracking since there are numerous variables that are capable of effecting the result starting from the number of cameras involved to the performance capability of the computer used. However, the proposed system targeted to find the middle point regarding hardware equipment and system performance while minimizing the disadvantages discovered within systems already developed.

In conclusion, the proposed system minimized the reaction time of the camera to continuously maintain pace with the subject of interest, provided a smooth and natural motion regarding the transition of the movement equivalent with an actual operator

controlling the camera, and produced competitive results regarding performance compared to conventional systems as a cost effective solution.

4.2 Future work

The proposed system is considered to be a starting point to expand the usage of a real-time face tracking system based on a single PTZ camera. Regarding hardware properties, combining the system implemented with the preliminary components and the proposed system together would be a prospective approach. Using the preliminary system as a method to initially identify a subject and operate the detection process while performing tracking with the proposed system would distribute the work load and improve the overall performance since the proposed system will only perform tracking regarding a certain subject and output the resulting image from executing the appropriate commands transmitted from the preliminary camera. In addition, additional static cameras or PTZ cameras would be able to provide a variety of choices regarding the view of a lecture in session as the proposed system solely focuses on establishing a tracking system maintaining the interest on the instructor. The usage of multiple cameras will hold the capability of increasing the number of views and also the involvement of participation regarding the students.

In terms of the software properties, the approach of the face detection process introduced can improve by adding additional facial features within the region of the face initially detected or by manipulating the classifier itself. This type of extension shall require improved computational power to handle real-time processing but will increase

the successful detection rate as a result. A modified version of the classifier suited for moderate computational power would be a target to improve detection as well. An alternative method of detection would be focusing on the motion of the individual instead of the face. Establishing a detection process involving motion detection can eliminate the obligation of facing the camera when a face detector is used. Improving the software aspect will ensure the advantage of mobility while enhancing the performance of the system in total.

The most important property concerned would be the delay possible to occur regarding the movement of the camera. Establishing hardware equipment to directly access the bit streams of the input image will most likely improve the detection speed which will minimize the delay of transmitting each frame. With rapid detection speed and prompt execution of the movement commands shall provide a more robust tracking system.

REFERENCES

[1] A. Iraqui, H. Y. Dupuis, R. Boutteau, J. –Y. Ertaud, X. Savatier, "Fusion of omnidirectional and ptz cameras for face detection and tracking," *International Conference on Emerging Security Technologies*, 2010.

[2] A. Mian, "Realtime face detection and tracking using a single pan, tilt, zoom camera," *Proceedings of the Image and Vision Computing New Zealand*, Christchurch, NZ: IEEE, 2008.

[3] B. M. Nair, J. Foytik, R. Tompkins, Y. Diskin, T. Aspiras, V. Asari, "Multi-pose face recognition and tracking system," *International Conference on Computer Vision Theory and Applications*, *VISAPP*, pp. 378-386, 2011.

[4] E. Gamess, N. Morales, "Peak performance of TCP and UDP in IPv4 and IPv6 over Ethernet networks," *JDCTA*, vol. 7, no. 9, pp. 519-528, 2013.

[5] F. Chang, G. Zhang, X. Wang, Z. Chen, "PTZ camera target tracking in large complex scenes," *8th World Congress on Intelligent Control and Automation*, July 2010.

[6] G. Bradski, A. Kaehler, "Learning OpenCV: computer vision with the OpenCV library," O'Reilly Media, Inc., 2008.

[7] I. E. Allen, J. Seaman, "Grade change: tracking online education in the United States," *Sloan Consortium*, 2014.

[8] Intel. Intel Open Source Computer Vision Library, v.2.4.8. http://sourceforget.net/ projects/opencvlibrary/.

[9] J. M. Pullen, “Pros and cons for teaching courses in the classroom and online simultaneously,” *ITiCSE*, pp. 180-185, 2012.

[10] K. Bernardin, F. Camp, R. Stiefelhagen, “Automatic person detection and tracking using fuzzy controlled active cameras,” *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2007.

[11] L. A. Rowe, D. Harley, P. Pletcher, S. Lawrence, “BIBS: A lecture webcasting system,” Tech. rep. Berkeley Multimedia Research Center, U. C. Berkeley, 2001.

[12] L. Fitigau, G. Toderean, “Performance analysis of TCP and UDP using Opnet simulator,” *11th International Conference on Development and Application Systems*, 2012.

[13] M. Aazam, S. A. H. Shah, I. Khan, A. Qayyum, “Deployment and performance evaluation of Teredo and ISATAP over real test-bed setup,” *Management of Emergent Digital EcoSystems*, pp. 229-233, 2010.

[14] M. Castrillón-Santana, O. Déniz-Suárez, L. Antón-Canalis, J. Lorenzo-Navarro, “Face and facial feature detection evaluation,” *International Conference on Computer Vision Theory and Applications*, *VISAPP08*, 2008.

[15] M. Kim, S. Kumar, V. Pavlovic, H. Rowley, “Face tracking and recognition with visual constraints in real-world videos,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, June 2008.

[16] P. Amnuaykanjanasin, S. Aramvith, T. H. Chalidabhongse, “Face tracking using two cooperative static and moving cameras,” *IEEE International Conference on Multimedia and Expo*, ICME 2005.

[17] P. D. Z. Varcheie, G. A. Bilodeau, "People tracking using a network-based PTZ camera," *Machine Vision and Applications*, vol. 22, no. 4, pp. 671-690, 2011.

[18] P. I. Wilson, J. Fernandez, "Facial feature detection using Haar classifiers," *Journal of Computing Sciences in Colleges*, vol. 21, no. 4, pp. 127-133, 2006.

[19] P. Viola, M. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, pp. 137-154, 2004.

[20] S. Cheng, A. Asthana, S. Zafeiriou, J. Shen, M. Pantic, "Real-time generic face tracking in the wild with CUDA," *MMSys*, 2014.

[21] S. Huttunen, J. Heikkila, "An active head tracking system for distance education and videoconferencing applications," *AVVS*, 2006.

[22] S. Kang, J. Paik, A. Koschan, B. Abidi, M. Abidi, "Real-time video tracking using PTZ cameras," *Proc. of SPIE 6th International Conference on Quality Control by Artificial Vision*, vol. 5132, pp. 103-111, May 2003.

[23] S. V. Viraktamath, M. Katti, A. Khatawkar, P. Kulkarni, "Face detection and tracking using OpenCV," *The SIJ Transactions on Computer Networks & Communication Engineering (CNCE)*, vol. 1, no. 3, Aug 2013.

[24] T. B. Dinh, N. Vo, G. Medioni, "High resolution face sequences from a PTZ network camera," *FG*, pp. 531-538, 2011.

[25] T. Dinh, Q. Yu, G. Medioni, "Real time tracking using an active pan-tilt-zoom network camera," *IEEE International Conference on Intelligent Robots and Systems*, Oct 2009.

[26] U. Park, H. Choi, A. K. Jain, S. Lee, “Face tracking and recognition at a distance: A coaxial and concentric PTZ camera system,” *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 10, 2013.