International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Wireless Technology for Improving E-Learning Classroom with Gesture Analysis

Hari S

Assistant Professor
Department of Electronics and Communication
Mount Zion College of engineering, Kadamanitta

Abstract: ***Video conferencing plays an important role in many corporate and educational fields. E-learning uses the concept of video conferencing for interaction between students' and tutors' in different locations. The tutors' actual presence is in real classroom and the students can view their tutor through a video in a virtual classroom. Wireless microphones and video sensors are used, to facilitate an interaction between the students and tutors but sometimes it may not be as efficient when we use multiple speakers. In that case, it would be helpful if we identify the student who asks a question first, either in virtual or real class-room by using audio and video sensors. To make an E-learning classroom as close to a real classroom, we propose a system that will utilize the professor's gestures; this will decide who can ask questions. This is particularly useful when we use several speakers in a E-learning classroom. The student who is asking a question the first time will be located using audio and video sensors either from virtual or real classroom. The raised hand along with his voice is used for localization. This method helps both the professor and the student get the experience of being in a real classroom.***

Keywords: Gestures, Multi-speaker, vocal vibration, A-view, E-learning

1. Introduction

E-learning has created a revolution in modern teaching methods. By this technique it is possible to provide world class training to students in remote areas with low cost. It also helps corporate companies to train their employees across different location. In order to provide an effective learning experience, the teacher in the remote classroom should be able to interact efficiently with students situated in another location. Audio and video sensors are used to make the E-learning classroom more efficient. Audio sensors such as microphone are used to receive audio input and video-sensors such as cameras are used to get video signals. To get accurate results, there is a need for improvement when we use audio and video sensors.

Gestures are used as a form of non-verbal communication. In a regular classroom gestures will play an important role in communication. Students, in regular classroom will raise their hand if they have questions. The professor will respond to the students by just looking at the particular student. Incorporating gestures of students and tutors in E-learning classroom is a challenge but it will make the classroom as close to a real classroom.

Multi-speakers might cause problems in E-classroom as tutors will not be able to effectively point out who is asking questions. It is necessary to build a system that will work effectively if multiple speakers are used. Amrita had developed a system (A-VIEW) for E-learning purposes. A-VIEW is constantly getting upgraded for efficiency, accuracy and .functionality

2. Related Work

Balaji Hariharan et al [1] say about how audio video localization is used for identifying a student in an E-learning classroom. Real time video tracking is done with the help of Open CV which is a free open source library for real time computer vision. Motion detection algorithms are used to find out the raising hand and the particular face is identified by comparing the frame with the trained sequence. Viola Jones algorithm is the fastest and the most robust algorithm for face detection. Motion detection algorithm basically depends on frame difference calculation. The algorithm runs until there is a motion detected with respect to the reference image obtained. The audio is also detected and localized by using triangulation.

Berman et al [2] tells us how gesture recognition system works. Body language is one of the most powerful ways through which humans can communicate nonverbally. Body language can take on many forms such as arm and hand gestures, face movement, facial expressions, and eye gaze. A GRS consisted of five basic systems or stages: 1) the presentation of an object of interest (the gesture); 2) a gesture-capture device (sensor); 3) a tracking algorithm (for motion capture); 4) feature extraction; and 5) a classification algorithm. It tells us about different types of sensors and it uses. The four main type of sensor stimuli discussed are electrical, mechanical, acoustic and optical. Optical sensors use camera, acoustic uses microphones and mechanical use touch screen as sensors. It tells us how we will get the stimuli using all these sensors.

David Lo et al [3] tells an innovative method for performing joint audio-video talker localization that takes the reliability of the individual localization parameters such as audio, motion detection, and skin-colour detection. The audio and video data are taken and are digitized and processed, and then localized separately. The space outside is divided into sectors for making localization easier. The active sector tells the place where speaker is present. Audio is localized by microphone using delay and sum technique and video is localized using motion detection and skin colour detection.

Chien-San Lin et al [4] talks about a speech radar system is presented for extracting speech information from the vocal vibration signal of a humans. Hua Gu Guangda Su

Cheng Du et al [5] This paper proposes a method to extract the feature points from faces automatically.

3. Problem Statement

Automatic speaker identification is not possible in the current E-learning classroom because the audio video sensors are not programmed to detect students who are asking doubts. Using multiple speakers in E-classroom does not help since it is difficult to spot a student when there are multiple students who want to ask questions. So we need to develop an effective system that will incorporate gestures of tutors and students either from real or virtual classrooms

4. Proposed System

An E-class consists of students located in local classroom as well as distant classroom. The professor will be teaching a class in the local classroom. It has cameras as well as audio sensors to detect the events happening in the class. Students who have questions will either raise their hand or talk. These audio video sensors will collaboratively work together and detect the first event either in virtual or local classroom. The PTZ camera will be zoomed to a particular location and the focus will be on a specific student. The Professor can decide whose question he or she wants to answer first. Professor's gestures play an important role here to make the decision. The students hand gestures and voice will be used to track and localize the events. So the E-Learning class will work more close to a normal classroom.

The speaker in the remote class room and normal classroom is identified by using microphone array and PTZ camera. The speaker who first talks are identified either from remote or real classroom using audio/video signals. The tutor who is in the remote classroom can make the decision whether to allow the students in local or remote classroom to talk by using gesture analysis of the professor's face. The PTZ camera and the audio sensors are used to track the students who are asking doubts on a priority basis. The students who are asking doubts first will be focused and professor's facial expression and the direction he faces is incorporated when allowing a student to ask questions Here an algorithm is developed to detect tutor's gestures. Fig[1-4] shows the various phases that are addressed in the project. The virtual classroom is a place where the students need a screen to view the professor. We need three cameras one for taking pictures of professor and the other two are used for taking images of the students in the real and virtual classroom. These classes are connected using internet and we need to consider the delay factor to get results in real time.

The tutor is taking class. His video will be displayed on the remote classroom and the remote students video will be displayed in real class

Figure 1

A student in remote classroom raises his hand for doubt. His face is focused in the real classroom as he produces first interrupt

Figure 2

A student in real classroom also raises hand.

Figure 3

Professor takes decision and give chance to remote student.

Figure 4

5. Block Diagram

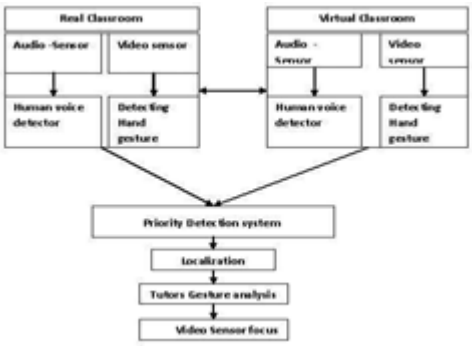

Figure 5: Block Diagram

The Audio sensors will sense the students who are asking doubts and the video sensors will sense the images of the students either from real or virtual classroom. The audio sensor will be fed to human voice detecting system for detecting human voice and the video sensor will be used

to detect hand raise of the students. Then we need to use priority detecting system to detect which event happens first. The priority is based on outputs from audio/video sensors from real and virtual classroom. After it's prioritized, the camera will focus the particular student who asks doubts first. The tutors' gesture will be analyzed in order to allow the students to ask questions just like normal classroom

Audio Localization

For Audio Localization we are using the concept of estimating time delay between pair of microphone for calculating the angle from the source. Cross correlation between audio signals is used for getting the time delay. Refer Figure 6

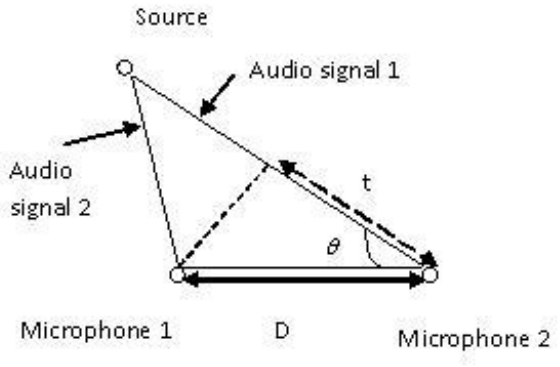

Figure 6: Audio Localization

Algorithm for audio localization

- Obtain audio signals
- Convert to frames
- Calculate average energy
- If it is above a threshold it is speech
- Cross correlate to find the time delay
- Calculate the angle

Thus by taking average of angles from different pairs of microphone, we will be able to detect the direction of students speech.

Video Localization

The students hand raise gesture as well as professors gestures needs to be find out for taking decision in E-class. The Gesture analysis Algorithm works on basis of comparison between the references frames with the frame to be checked. For creating reference image, we need to train the gestures of different category and save in a database. The captured image is compared with each of the reference frame. Those who get the maximum correlation will be detected as the match. Students hand raise is used for locating a student who is asking doubts. Here we are using correlation algorithm to detect gesture of the professor as well as hand raise. It consists of two phase training phase and Recognition phase.

Algorithm for Video localization

Training Phase

- Take snapshot each of two positions (left and right)
- Convert colour to gray
- Save in a matrix

Recognition Phase

- Take snapshot of image to be recognized
- Convert RGB to gray scale and save in a matrix
- Find correlation between each stored image with snapshot.
- Choose the position which has got maximum correlation coefficient.

6. Results

The code for cross correlation was done in MATLAB and the system was able to recognize the direction of facial change of the professor as well as the hand raise of the student. The audio localization was simulated in Labview and the time delay was found out. This can be used to calculate the direction of audio signal and hence will be able to localize the students. Refer Fig 7-9.

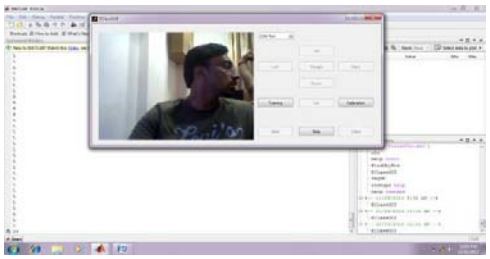

Figure 7: Video Frame Right

Figure 8: Video Frame Left

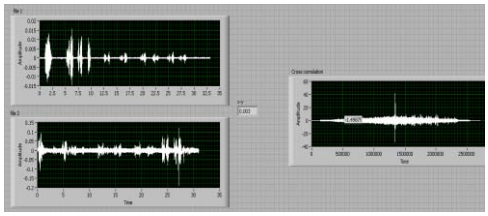

Figure 9: Time delay estimation

7. Conclusion

The main purpose of the project is to make the E-Learning classroom more natural by effectively using gesture analysis of tutor in case of taking decision in case of multi speaker scenario. The PTZ camera and the audio sensors are used to track the students who are asking doubts. The system will then focus on the student who talks first.

Professor's facial direction is observed when letting a student ask a question. Here an algorithm is developed to detect tutors gestures as well as students hand raising gesture and audio localization. It also uses internet for connecting virtual and real classroom for efficient way of communication.

Acknowledgement

We would like to express our immense gratitude to our management and staffs of Mount Zion College of Engineering, Kadammanitta for providing a very good motivation and inspiration for my work.

Reference

[1] Amrita Vishwa Vidyapeetham, Remote Student Localization using Audio and Video Processing for Synchronous Interactive E-Learning Balaji Hariharan, Aparna Vadakkepatt, Sangeeth Kumar, Amrita Centre for Wireless Networks and Applications
[2] Berman, Member, IEEE, and Helman Stern, Member, IEEE, Sensors for Gesture Recognition Systems-IEEESigal
[3] David Lo, Rafik A. Goubran, Member, IEEE, Richard M. Dansereau, Member, IEEE, Graham Thompson, and Dieter Schulz, Robust Joint Audio-Video Localization in Video Conferencing Using Reliability Information
[4] Chien-San Lin, Sheng-Fuh Chang, Senior Member, IEEE, Chia-Chan, Microwave Human Vocal Vibration Signal Detection Based on Doppler Radar Technology
[5] Hua Gu Guangda Su Cheng Du, Feature Points Extraction from Faces
[6] J. Wilder, P. J. Phillips, C. Jiang, and S. Wiener, "Comparison of visible and infrared imagery for face recognition," in Proc. 2nd Int. Conf. Autom. Face Gesture Recog. Oct. 1996, pp. 182–187