Multi-Camera Live Video Imaging

Yoshinari Kameda

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: Making a lecture video which conveys essence of the lecture is difficult because they should know not only locations of teacher and students but also who is in focus as the lecture goes on. This is especially difficult when a *live* video broadcasting is required in interactive distance learning courses which are held between two remote classrooms in only one of which a teacher is there. Our aim is to establish a live lecture video imaging system which can work like an automatic broadcasting studio.

Motivation: Although E-style distance learning systems are gradually being widespread these days, there are some sorts of strong requests that students in a classroom want to share a teacher in a remote classroom with students there. We call this type of lectures interactive distance learning. One video duplex transmission is usually allowed to use to support this type of lectures whereas a person in focus changes from time to time and sometimes walks around. Therefore, two functions are required to make one video stream from one classroom to the other. One is to observe the scene and track moving persons with active shooting cameras, and the other is to select one of the shooting cameras so as to send its image to the remote classroom. In broadcasting studios, there are cameramen and a director who are responsible for these functions and their mutual work is a good example of collaboration. Our goal is to realize these two functions with computers through computer vision.

Previous Work: Since it is assumed that there are several persons to be shot in the scene and resultant video is viewed by students as a lecture, not only a method of tracking one object with one active camera but also switching target object in the video should be considered. Many researches are focused on tracking problems whereas our concern is beyond a simple tracking problem. It is important to design the functional system which could conduct many cameras at the same time.

Approach: Our approach is divided into two steps. One is a shooter module which is responsible for controlling several shooting cameras by observing a part of the scene. The other is a selector module which selects one of the shooting cameras that belong to the shooters.

A shooter module has one camera for observing the scene and at least one active camera for shooting a target person. Each active camera can be panned, tilted, and zoomed. Each observation camera is set where trajectories of moving person doesn't overlap themselves. Once the moving person is detected at $I(x, y)$ on the image plane of the observation camera, predefined value $C_i(p, t, z)$ of pan, tilt, zoom for the location $I(x, y)$ is sent to active camera i. This predefined $C_i(p, t, z)$ implies a composition of a video frame for the person in the scene. For example, if a person in front of a lecture table is found in the classroom at $I(x_0, y_0)$, and associated z value to $I(x_0, y_0)$ is big, it indicates he/she should be framed in close-up shot. It is no need to set $C(p, t, z)$ so as to put a target in the center of shooting image because location estimation process doesn't use the shooting image.

A selector module accepts statuses of all the shooting cameras from the shooter modules and select one of them to transmit it to a remote site. Motion status of each shooting camera is defined as one of *out-of-service*, *changing-object*, and *tracking-object*, and associated object information is also sent to the selector. Each video can be split by these two information into segments each of which contains one successive action of one person or one object. The selector checks the duration time of video segments on transmission according to the following two rules. First; a video segment should continue on transmission for least short seconds to let viewers understand what they see. Second; a video segment should not be transmitted for a long time if there are other video segments available so as not to bore the viewers. It selects the video segments to be sent under these two rules.

Impact: As this approach enables us to make live video in a classroom automatically while lectures are being held, any class could welcome students in a remote classroom without preparing extra faculty staffs for video

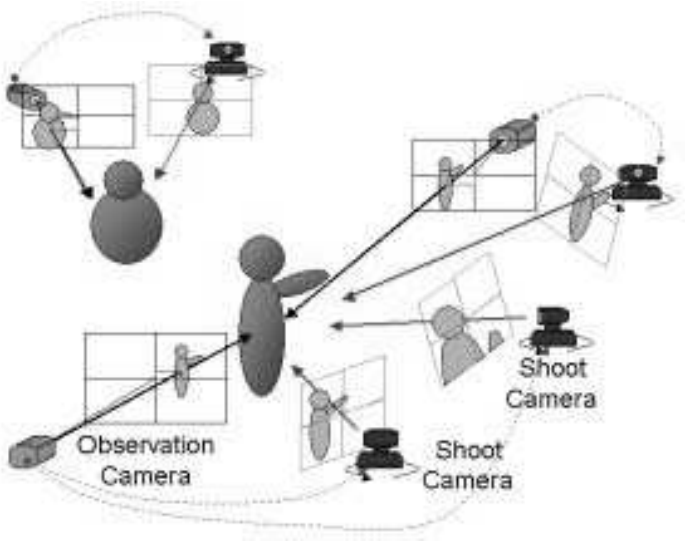

Figure 1: Shooters with Cameras

transmission.

Future Work: An interactive distance learning project has been held in the past two years between Kyoto university in Japan and UCLA and we have been improving our prototype system through its courses.

Sometimes the images from different shooters are very similar to each other because the shooters work independently based on their own image processing result and don't communicate each other. This could be prevented by adding a method of sending some orders from the selector to the shooters so as not to shoot the same object in the similar composition.

We are also investigating video selection algorithm that can deal with the context of lectures in order to achieve more sophisticated video switching. In addition, audio and other types of sensors such as on-line handwriting recognition on a white-board should be taken into acount for that purpose.

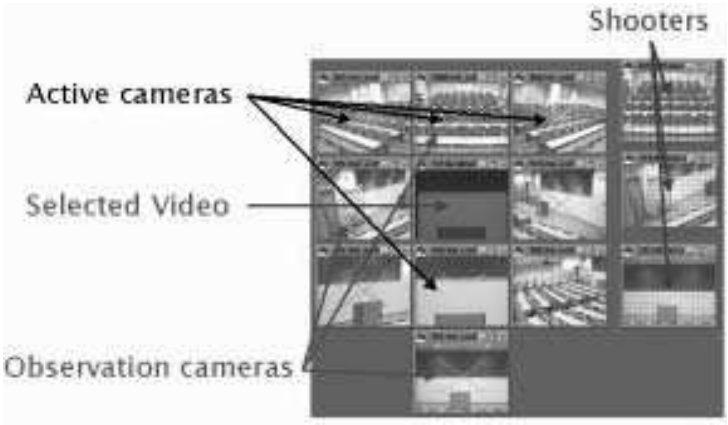

Figure 2: An Example Setup with 9 Cameras