

Development of an Automatic Camera Control System for Videoing a Normal Classroom to Realize a Distant Lecture

Suganuma, Akira
Department of Intelligent Systems, Kyushu University

<https://hdl.handle.net/2324/5358>

出版情報 : pp. -. ., 2002-11
バージョン :
権利関係 :

Development of an Automatic Camera Control System for Videoing a Normal Classroom to Realize a Distant Lecture

Akira Suganuma

Abstract— The growth of a communication network technology enables people to take part in a distant lecture. When lecture scenes for a distant lecture are videoed, a camera-person usually controls a camera to take suitable shots; alternatively, the camera is static and captures the same location all the time. Both of them, however, have some defects. It is necessary to control a camera automatically. We are studying a camera control system for the distant lecture for videoing a normal classroom with computer vision techniques. We call it ACE: Automatic Camera control system for Education. ACE is not only a system that controls a camera but also a system that enables students to browse a thing that a teacher has written before. ACE can also inform a teacher of the state of his students in a distant room.

Index Terms— Application for a Distant Lecture, Communication Aids, Image Processing, On-Line Teaching and Learning

1 Introduction

The growth of a communication network technology enables people to take part in a distant lecture. There are mainly two kinds of method to held such a lecture. One is an web-page based method, the other is a method of sending visual and audio of lecture scenes. We are studying some supporting systems for a distant lecture. In the web-page-based method, we have designed and developed two supporting system: a Computer Aided Cooperative Classroom Environment (CACCE)[1] and are developing an Automatic Exercise Generator based on the Intelligence of Students (AEGIS)[2],[3]. On the other hand, we are developing a camera control system for the distant lecture held with the method sending a lecture scene[4],[5]. We call it “ACE”: Automatic Camera control system for Education[6].

A traditional style lecture held in a normal classroom is the target videoed by ACE. Nowadays, a teacher teaches his students with an OHP and/or other visual facilities. Indeed many lectures such as in the information technology or the program-

ming are frequently held by using visual facilities or computers in many universities but there are still many traditional style lectures, in which a teacher explains something with a blackboard. It seems that such a lecture will not completely disappear in the future although they will hold by combining the blackboard and a visual facility such as an OHP or Power Point software. We are, consequently, developing ACE for the distant lecture with videoing the traditional lecture.

When a lecture scene for a distant lecture is videoed, a camera-person usually controls a camera to take suitable shots; alternatively, the camera is static and captures the same location all the time. It is not easy, however, to employ a camera-person for every occasion, and the scenes captured by a steady camera hardly give us a feeling of the live lecture. It is necessary, consequently, to control a camera automatically. ACE enables people to do it for taking suitable shots for a distant lecture. ACE analyses a scene sent from a camera and recognizes the complexion on the lecture. ACE judges what is important in the scene and controls the camera to focus on it.

ACE is not only a system that controls a camera but also a system that enables students to browse a thing that a teacher has written before. The early version of ACE, which only controls a camera, detects a teacher in a scene and controls a camera. This denotes that ACE focuses on a target from teacher’s point of view. However, some students seem to want to watch another scene. We have designed that ACE can generate and store an image from a shot captured by ACE. Students can, then, watch a scene they want with an Web-browser.

ACE can also inform a teacher of the state of his students in a distant room. In a distant lecture, a teacher cannot watch his students in the distant room or he can watch them by only the medium of a monitor. In this situation, he cannot judge their state as well as he judge the state of students in front of him. We designed a function to inform a teacher of th state of his students with cooperation between the function of supplying a scene captured before and CACCE, which we had implemented before.

In this paper, section 2 presents our assumption in the design of ACE, our strategy of designing ACE. Section 3 describes the algorithm that detects a thing focused on. Then, section 4 describes how to inform a teacher of the state of students in the distant room. A cooperation between ACE and CACCE is represented in this section. Finally, concluding remarks are given in section 5.

This research was partly supported by a Grant-in-Aid for Scientific Research (Category C) from Japan Society for the Promotion of Science(JSPS), No.12680215, 2000–2002 and No.14580224, 2002–2004.

A. Suganuma is with the Department of Intelligent Systems, Kyushu University, Kasuga-koen, Kasuga, Fukuoka 816–8580 Japan (e-mail: suga@limu.is.kyushu-u.ac.jp).

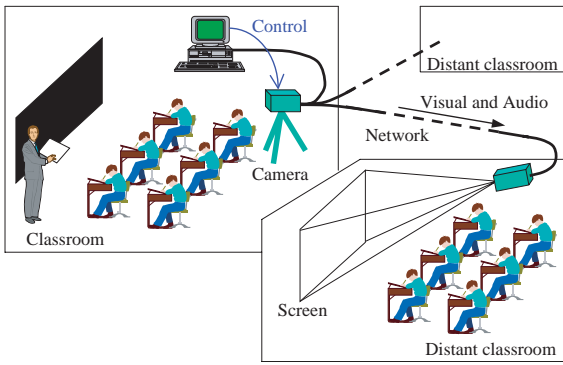


Figure 1: A form of the distant lecture by videoing the normal classroom

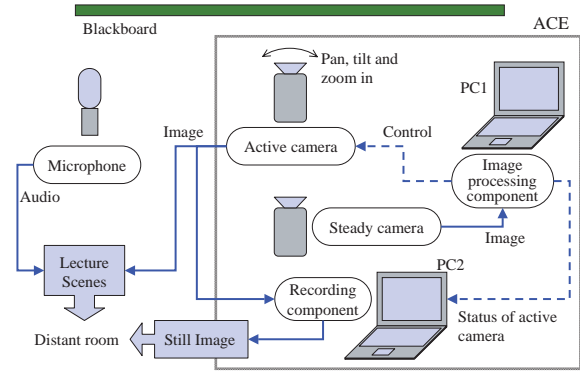


Figure 2: An Overview of ACE

2 ACE System

2.1 Distant Lecture We Envisage

We envisage that a scene of a lecture held in a normal classroom are videoed by a video camera and students in a distant classroom take part in the lecture by watching the scenes projected on a screen. Figure 1 illustrates a form of the distant lecture by videoing a traditional classroom. A teacher teaches his students in a normal classroom. There is a blackboard in the room. He writes and explains something on it. Watching it and listening to his talk, students take part in the lecture. Some cameras are setting in the room and captures a lecture scene in order that the captured scene is sent to distant classrooms. On the other hand, students in the distant rooms take part in the lecture by watching a scene reflected on a screen.

2.2 Assumptions in the Design of ACE

We have designed and implemented ACE, which is an application based on Computer Vision Technique. When we designed it, we assumed the following:

- A teacher teaches his students by using only a blackboard.
- Students aren't reflected in the scenes captured by the camera.
- A teacher isn't required to give the system a special cue.
- Each student in the distant room is assigned a PC connected to the Internet.

The first assumption means that the lecture captured by ACE is a traditional one. The teacher writes something on the blackboard, and explains them. Indeed he teaches his students using an OHP and/or other visual facilities in the recent years, but many traditional ones are held in many schools. The second assumption is made to decrease processing costs.

If students are reflected in the scenes, ACE always has to distinguish a teacher and students. This processing is complex and takes much time. It is easy to satisfy this assumption if a camera sets up on the ceiling.

The third assumption is very important for a teacher. If a teacher gives ACE his special cue such as to press a button of a remote controller, ACE may control a camera more easily. ACE has only to keep waiting his cue in that situation. If the teacher, furthermore, put on a special cloth, on which some color markers are attached, it is easier to detect his position and/or action. The special cue and the special cloth, however, increase the load on the teacher. He may omit to give ACE his cue. He ought to concentrate his attention on his explaining. Consequently, we didn't require him to give ACE his cue.

The last assumption may not satisfy in some classrooms. However, an interface operated by a student and a monitor displaying individually the scene he wants are required in order that he selects a scene. We adopted an Web browser as the interface. We got our system to create Web pages automatically from the scenes captured by ACE. So a student can watch a requested scene with an Web browser running on PC assigned to each student.

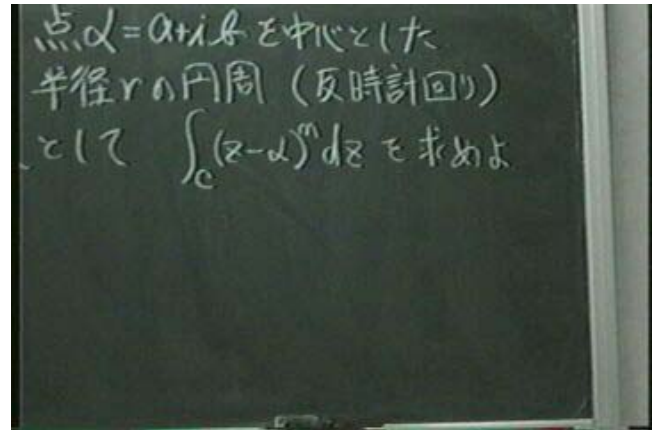
2.3 An Overview of ACE

An overview of ACE is shown in Figure 2. ACE requires two cameras. One is a steady camera and the other is a active one. The steady camera captures a whole blackboard at a constant angle for image processing. The captured image is sent to an image processing component of ACE running on PC1 over an IEEE-1394 protocol. The component analyzes the image and decide how to control the active camera according to a camera control strategy shown in section 2.4. The control signals are sent to the active camera over an RS-232C protocol. The active camera, hereby, takes a suitable shot.

ACE has a recording component which generates effective still image files from the scene captured by the active camera. It runs on PC2 because the procedure of the image processing component requires high CPU power. It receives a status



(a) An ordinary shot



(b) A key shot

Figure 3: Sample shots of a lecture scene captured by ACE

of the active camera from the image processing component, analyzes the scene sent from the active camera and makes a decision whether to generate a still image file or not.

A visual captured by the active camera and an audio picked up by a microphone are sent to the distant room. Students in the distant classroom watch and listen to them. They can also watch a requested scene as a still image. In our study, we are interested in how to video a lecture held in normal classroom. We are using a known method or product as a way sending the video via the network.

2.4 A Strategy of Camera Control

What does ACE capture? It is a very important thing for the system such as ACE. One solution is to take the scenes that students want to watch, but in this case many scenes are probably requested by many students at the same time. Although this solution needs the consensus of all students, it is very difficult to make it. We decide, therefore, that ACE captures the most important things from a point of view of a teacher. We guessed the objects (such as words, sentences, formulas and so on) that teacher was explaining were the most important things for all students. We adopted, therefore, a strategy that ACE captured something explained by a teacher. However to extract directly what a teacher explains is very difficult. Consequently, we made an assumption that the latest object written on the blackboard closely parallel something explained by a teacher because he frequently explains the latest object. ACE captures the latest object written on the blackboard.

When a lecturing scene is videoed, both a constantly changing shot and an over-rendering shot are not suitable because they tire students. A change-less shot is, if anything, more appropriate than those shots. It is very important that students can easily read contents on the blackboard. Sample shots captured by ACE is shown in Figure 3. ACE usually takes a shot containing the latest object and a region near it in a discernible size. The blackboard often consist of four or six

small boards like a picture in Figure 4. A teacher frequently writes relational objects within one small board in the situation. ACE takes a shot by the small board like figure 3-(a) in the ordinary way. On the other hand, when once ACE detects an object on the blackboard as the latest object, ACE takes a shot zoomed in on it like figure 3-(b) after the teacher has written it. After a-few-second zooming, ACE takes an ordinary shot again.

If a steady camera takes the lecture scene, the shot may be like a shot in Figure 4. The steady camera must capture the whole blackboard in this situation, because a teacher writes something anywhere. The characters in this shot are too small for students to read. Then, the shot of ACE is superior to that of the steady camera.

2.5 A Strategy of Recording a Scene

Indeed ACE captures the objects explained by a teacher, but some students probably want to watch another scene. ACE has a function recording the scene captured by the active camera before. Objects on the blackboard doesn't change when once a teacher writes or erases them. If a object is updated, it is detect as the latest object again. This is reason why a still image is good enough for sending information.

The key shot captured by ACE is not suitable for recording because ACE cannot always detect the latest object on the blackboard. It is not to be denied that some objects are not timely detected as the latest object and that a meaningful object is divided into some parts of object. We decided that the ordinary shot, which consists a small board containing the latest object in a discernible size, was recorded. Because a teacher often wrote a meaningful object on one small board.

A sample of a interface to students is shown in Figure 5. This Web page has two frames. In the left one, some thumbnail images of a small board are placed in order of generating, and in the right one, ones are placed in its position. When clicking each image, students can watch its enlarged image.



Figure 4: A sample shot captured by a steady camera

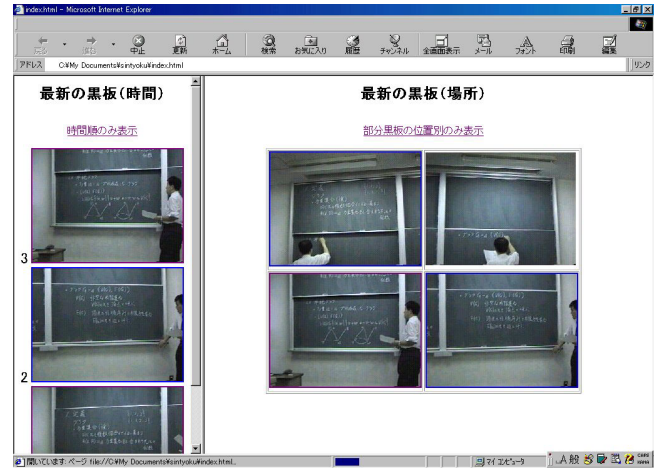


Figure 5: A sample of displaying small boards on which a teacher wrote some objects

2.6 A Strategy of Spying Students

In a distant lecture, a teacher cannot comprehend the state of his students in a distant classroom as well as in a local classroom. It is a useful function to inform the teacher of the state of his student.

ACE supplies students the feature that they can refer a scene captured before with a browser whenever they don't understand its detail or whenever they want to make sure it. We guessed that their operation may denote the state of students. If the browser used by the student can inform a program running on the teacher's PC of a URL of the page which it displays, the information probably helps progress of the lecture.

Students who take part in the distant lecture with ACE can watch a scene, which the active camera captured before, on their own pace. The recording component of ACE selects and stores an informative scene and generates a web page automatically. Students browse around the generated pages by their browser. On the other hand, CACCE, which is a system implemented by us, consists of a teacher's browser and student's browsers[1]. We adopted the student's browser of CACCE as the browser that runs on the PC assigned to each student.

3 How Does ACE Take a Suitable Scene?

3.1 Extracting the Latest Object

Background Subtraction: We use a background subtraction technique to detect objects on the blackboard. This technique is a method to detach the foreground image from an image. In the method, a background image is prepared in a pre-process. In the case of ACE, the background image is captured by the steady camera before opening the lecture.

The image contains only the blackboard on which no object is written. It does not contain a teacher. When once the background image is made, the foreground image is detached from the image, which is captured by the same camera as captured the background image, by calculating the difference between the image and the background image. ACE can detect something written on the blackboard and something in front of blackboard by subtracting the background image from the image captured by the steady camera during the lecture.

We adopted a background model[7] in our system. This model is robust against a noise. A platform and a blackboard are lightened by fluorescent lamps in a normal classroom. Since the fluorescent lamps cause a flickered noise, there are usually many noises in the shot captured by the steady camera. ACE needs a robust method against noises for this reason.

We specialized the background model for the lecture scene. The foreground objects segmented by the technique are something written on the blackboard, something erased on it, a teacher and so on. We need only the written object and want to distinguish them. The pixels of the object written on the blackboard appear only above the upper-bound of the background model because the object is brighter than the blackboard. Our method detects, therefore, pixels whose brightness is more than the upper-bound. ACE segments the object by using the following inequality:

$$I(p) - \text{Max}(p) \leq D(p)$$

where $I(p)$ stands for an intensity values of pixel p , $\text{Max}(p)$ is a maximum intensity values of pixel p observed during capturing the background image. $D(p)$ is a maximum intensity in the subtract image between two pieces of successive frames observed during capturing the background image. The foreground objects are extracted by thresholding and noise clearing. The objects represent highlighted pixels in the background subtraction image.

Separating an Object from the Foreground Image: The foreground image almost always includes a teacher because he is not included in the background image. We would like to detect only the written object. If a region of the teacher can be masked, the written object is extracted correctly. We have to, therefore, extract a teacher's region.

We assumed that all the moving object is a teacher. A method of subtracting between successive frames is usually used when people wants to detect a moving object. In a lecture scene, a teacher almost always moves in front of a blackboard and hence is partly detected by this method. It sometimes detects only a part of teacher's body or sometimes misses a teacher. ACE makes a preliminary teacher's region by joining the rectangle of the moving objects, which is circumscribed the moving objects detected by this method, with the teacher's region in the previous frame. ACE extracts highlighted pixels in the preliminary region of the foreground image, and regards a rectangle circumscribed them as the teacher's region.

After all pixels in teacher's rectangle in the foreground image are changed to dark ones, the remnant highlight pixels are the written object if teacher's region is segmented correctly enough. ACE makes a rectangle circumscribed the highlight pixels, and deals with it after this processing.

Remake the Background Model: We have to distinguish the latest object and others. If once the object has been detected as the written object, it doesn't have to be detected more than twice. After detecting the latest object, ACE recalculates the values of the background model for each pixel in the region of the object. ACE always detects, therefore, only the latest one.

3.2 Finding the Timing of Zooming in

We cannot control a camera even if we get the region of the latest object. We have to find the timing of zooming in. If ACE zooms in on the written object while a teacher continues writing, ACE must take a scene in which the teacher occlude the object. After guessing whether the teacher finished writing, consequently, ACE zooms in on the object.

The rectangle circumscribed the latest object usually change frame by frame. This main reason is the following:

- The rectangle increases or decreases because the teacher wrote something new or erased something.
- The masked region changes because the teacher moved to write something in front of the object. Then the rectangle of the latest object increases or decreases.

Shortly, the rectangle changes when the teacher is writing something.

On the other hand, he usually clears the object to make his students watch it after he has written. ACE takes advantage of this feature to guess whether he finished. The rectangle does not change when he cleared the object because neither

above two reasons occurs. ACE counts the number of frames in which the rectangle does not change. If the number is over a threshold, then ACE judges the teacher finished writing and controls a camera to zoom in on the written object.

4 How Does ACE Inform a Teacher of His Students' State?

ACE integrates two functions in order to inform a teacher of his students' state. One is a function of recording a scene captured by the active camera, the other is a function of communicating between a teacher's PC and his students' PC.

4.1 Recording the Scenes Captured Before

Detecting the State of the Active Camera: We adopt the ordinary shot as a still image stored by ACE as we discussed in section 2.5. ACE can easily generate an image from the ordinary shot according to the status because the recording component receives the status of the active camera from the image processing component. The ordinary shot, however, is not always suitable for the stored image. ACE has to check whether the objects on the small board is the same as that in the image which has been stored already. The image does not have to be stored if it has been already stored. The component checks it whenever the status of the active camera changes from a key shot into an ordinary shot.

Estimating Whether a Teacher Occludes a Small Board: Contents on the blackboard in the stored image have to be clear. A teacher sometimes occludes the contents in the ordinary shots. Then, the recording component also has to check whether he does not occlude the contents on the blackboard. The image processing component detects a teacher's region. The position of the teacher calculated by the image processing component is not good enough for the recording component to check the occlusion because the lines of vision of the steady camera and the active camera are differ.

The recording component detects a teacher in the image captured by the active camera. The ordinary shot contains the whole small board, so four edges of the board appears a complete rectangle in the image. If the teacher is in front of the board, the bottom edge is divided into two pieces. When the edge is completely connected, the component guesses that a teacher does not occlude the board, and stores the shot from the active camera.

Making an Web page: The recording component automatically re-makes an HTML file of a top page containing some thumbnail images of the still image such as fig. 5. The HTML file is sent to a Web server and is able to be referred by students. It makes a thumbnail file from the image stored by ACE, sends both the original file and the thumbnail file to the Web server for students to refer them.

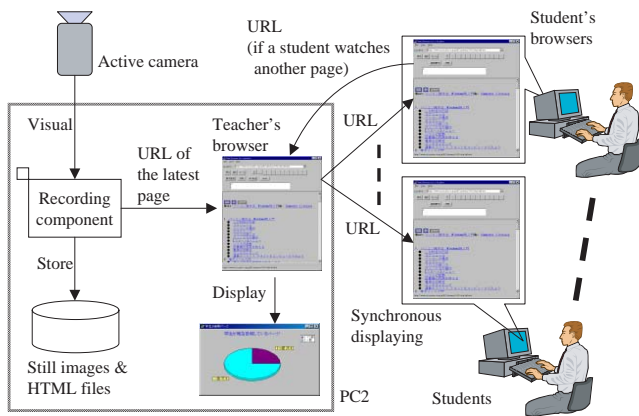


Figure 6: Overview of a flow of data around the recording component

4.2 Communication between a Teacher and Students

The teacher's browser and student's browser of CACCE are connected by a socket connection and communicates each other. CACCE has a function of synchronously displaying among a teacher's browser and student's browsers by sending an URL of the page displayed by the teacher's browser. The student's browser is, therefore, usually kept displaying the latest page stored by ACE with this feature. The student's browser, however, also allows a student to show any page just like an ordinary browser. Hence they can show the scenes captured before on their own pace. Student's browser also reports to teacher's browser about a URL of the page displayed on it whenever it displays another page. Teacher's browser makes a pie chart about the state of student's browsing, and informs consequently a teacher about it.

We have designed cooperation between the recording component of ACE and teacher's browser. Figure 6 illustrates an overview of a flow of data around the recording component. Both the recording component and teacher's browser run on PC 2 in fig. 2. The component selects and stores a informative shot as a still image, makes an web page automatically. After that, it sends the URL of the latest stored page to teacher's browser. Teacher's and student's browsers perform their ordinary role.

5 Conclusion

ACE takes a suitable shot if the teacher explains the object as soon as he writes on the board. ACE cannot take, however, a suitable shot when he explains something in front of it and when he explains something written before. In the former case, the teacher have to change his position because he occludes the objects and his students can not look at them. In the latter case, the teacher usually teaches his students pointing the objects which he wants them to watch. If interpreting a

teacher's action and/or posture, ACE could capture more suitable scene. We will get ACE to interpret them. We assume that a teacher teaches his students with only a blackboard. But he sometimes also uses with a visual facility. We will also get ACE to apply to such a situation.

ACE only informs a teacher of a graph showing which page students browse in. Indeed this graph may be used as one of the guidelines for teaching, but the state of the browser does not always denote student's state. They might watch a screen with keeping their browser displaying a meaningless page. Then, we will devise a method analyzing the information from student's browsers and informing a teacher of more useful information for teaching.

References

- [1] A. Suganuma, R. Fujimoto, and Y. Tsutsumi, "An WWW-based Supporting System Realizing Cooperative Environment for Classroom Teaching," Proc. of World Conference on the WWW and Internet, pp.830-831, 2000.
- [2] T. Mine, A. Suganuma, and T. Shoudai, "The Design and Implementation of Automatic Exercise Generator with Tagged Documents based on the Intelligence of Students: AEGIS," Proc. of International Conference on Computers in Education, pp.651-658, 2000.
- [3] A. Suganuma, T. Mine, and T. Shoudai, "Automatic Generating Appropriate Exercises Based on Dynamic Evaluating both Students' and Questions' Levels," Proc. of World Conference on Educational Multimedia, Hypermedia & Telecommunications, CD-ROM, 2002.
- [4] A. Suganuma, S. Kuranari, N. Tsuruta, and R. Taniguchi, "An Automatic Camera System for Distant Lecturing System," Proc. of Conference on Image Processing and Its Applications, Vol.2, pp.566-570, 1997.
- [5] A. Suganuma, S. Kuranari, N. Tsuruta, and R. Taniguchi, "Examination of an Automatic Camera Control System for Lecturing Scenes with CV Techniques," Proc. of Korea-Japan Joint Workshop on Computer Vision, pp.172-177, 1997.
- [6] A. Suganuma and S. Nishigori, "Automatic Camera Control System for a Distant Lecture with Videoing a Normal Classroom," Proc. of World Conference on Educational Multimedia, Hypermedia & Telecommunications, CD-ROM, 2002.
- [7] I. Haritaoglu, D. Harwood and L. S. Davis, "W⁴: Who?When?Where?What? A Real Time System for Detecting and Tracking People," International Conference on Face and Gesture Recognition, pp.14-16, 1998.