A LIVE VIDEO IMAGING METHOD FOR CAPTURING PRESENTATION INFORMATION IN DISTANCE LEARNING

Yoshinari KAMEDA†, *Kentaro ISHIZUKA*‡, *and Michihiko MINOH*†

†Center for Information and Multimedia Studies, Kyoto University
‡Graduate School of Informatics,, Kyoto University
Yoshidahonmachi, Sakyoku, Kyoto, Japan. 606-8501
kameda@media.kyoto-u.ac.jp, kentaro@kuis.kyoto-u.ac.jp, minoh@media.kyoto-u.ac.jp

ABSTRACT

We propose a new approach of video imaging method suitable to support distance learning. Automatic active camera control and video selection system visualizes presentation information of a lecturer and active students. We consider dynamic situation of the lecture to determine the appropriate target object to be shot which gives remote students live lecture information. We have implemented our method and used it in several distance learning courses held between Kyoto university and UCLA.

1. INTRODUCTION

We consider that *learning material* consists of two types of elements from a viewpoint of serving lecture with multimedia equipment. One is *teaching material* which is prepared before lecture starts and the other is *presentation information* which is generated or added to the teaching materials when the lecture is being held. Both teaching materials and related presentation information are given by a lecturer, but in different time (before lecture / during lecture). Ordinary multimedia learning materials which are provided by CDROM or WWW[2] are produced by involving presentation information into teaching materials. Order of chapters in CDROM or hyper-links in WWW are embedded presentation information and are prepared by the lecturers on producing them.

On supporting distance learning, remote students should receive both kinds of elements simultaneously. We should recognize what kind of presentation information arises in real-time in a classroom and send appropriate elements to the remote students because it is not only impractical for the lecturers to prepare all the presentation information in advance but also difficult to predict reaction of the students in the lecture. This point is different from producing CDROM or WWW for education provided by some universities.

One of the most important presentation information is considered to be a gesture of the lecturer and behavior of the students. Therefore, we concentrate on visual aspect of presentation information in this research.

We assume two kinds of objects to be visualized in a local classroom in our distance learning framework. Those are the lecturer and the students. They generate visual presentation information in the lecture. We explain how to shoot these objects and visualize them. We use multiple active cameras for shooting in real-time. Description of dynamic situation is introduced to determine appropriate camera control and video selection.

The framework for supporting distance learning is described in Section 2 and we propose an imaging method with dynamic situation in Section 3. We show experimental results in Section 4 and conclude our paper in Section 5.

2. DISTANCE LEARNING SYSTEM

Distance learning system is not only the system for audio/video transmission as is advertised widely but also includes lecture, video imaging, hand-writings and teaching materials. The framework of distance learning influences the design of our imaging method because the objects that the active cameras have to shoot would be limited according to the framework.

In this paper, we consider a kind of distance learning in which there are two remote classrooms, one with a lecturer and students, and the other with only students. The lecturer uses WWW pages to present his/her teaching materials, and uses an electric whiteboard to draw subsidiary figures and diagrams during the lecture.

There is a layer structure in the distance learning system as is shown in Figure 1. The lowest layer is network layer. The network layer has to assure data transmission, particularly streaming audio/video data transmission.

Above the network layer, we have system layer in which video imaging system, audio recording system, hand-writing system, and teaching material handling system exist. The purpose of these four systems is to detect and capture pre-

0-7803-6536-4/00/$10.00 (C) 2000 IEEE

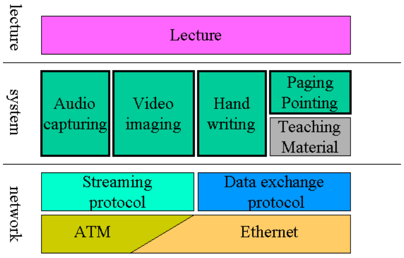

Figure 1: Layer structure of distance learning

sentation information. There are not only persons but also text/figures in a classroom to be imaged from a viewpoint of visualization. As we prepare the electric whiteboard which can record everything the lecturer writes on it during his/her lecture and WWW pages which can be displayed directly to the screen in the remote classroom, these objects are out of our focus on imaging the lecture in this research.

Our framework of distance learning system is shown in Figure 2. The lecturer in the left classroom uses the electric whiteboard and WWW synchronization software on PC. Transmitted data of the electric whiteboard and WWW pages are displayed on one screen-T in the remote classroom. There is the other screen V which is used to display video of the lecturer and the students.

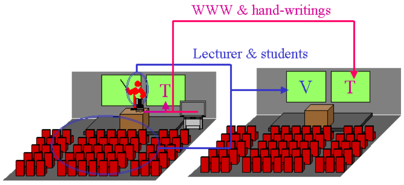

Figure 2: Distance learning framework

3. VIDEO IMAGING IN DISTANCE LEARNING

3.1. Imaging Philosophy in Distance Learning

We consider that there are three styles of imaging the lectures for distance learning. These are characterized by the way of viewing presentation information.

1. Lecture understanding (Broadcast)
2. Remote student's interest
3. Common Attention (Surveillance)

The first style is the most suitable style for distance learning because it aims to let the remote students understand the lecture. Although several cameras are used in this style, only one video stream is generated to carry visual information due to the communication cost of the network. We call this style a broadcast imaging style which is usually programmed in broadcast studio where a director controls cameramen by considering his/her audience.

The second style is acceptable when the interests of the remote students take precedence over the context of the lecture. We have already proposed the imaging method for this style[1].

The third style is for an event which happens in the classroom but is not related to the lecture. As no one can predict such kind of unexpected event, there is no systematic imaging way against it. Therefore, we should model common reaction when people encounter the unexpected events.

We focus on the first style in this paper. Since gestures of the lecturer and the behaviors of the students are very important presentation information in the lecture, the following sections describe our method for shooting them based on the situation of the lecture.

3.2. Dynamic Situation

On watching the lecture, people usually pay attention to only one object at the same time. Therefore, in order to send the video stream to the remote students, the system should know: "1. Which object is to be shot now ? And how ?"

Because this is based on a *dynamic situation* in the classroom at that time, then it should know: "2. What kind of dynamic situation is now ?"

The first question is linked to the second question through the dynamic situation[1]. See Figure 3.

A representation gap of the dynamic situation is an obstacle on interpreting dynamic situation between the two questions. As the first question should be answered from the viewpoint of capturing presentation information, it is preferable to describe the imaging rules symbolically. “Watch the lecturer closely when he/she is talking to the students” is an example where the main clause of the imaging rule is called *camera-work* and the subordinate clause which is called *A-component* is a symbolic representation which describes a part of the dynamic situation. The dynamic situation sometimes consists of several A-components during the lecture. A pair of the camera-work and the A-component represents one imaging rule, and several imaging rules which has the same A-component but different camera-works can be specified simultaneously.

The second question should be answered by computer vision technique. The system estimates certain values numerically such as the location of the objects in the classroom by the technique. We call these numerically estimated values *situation features*.

We introduce a mapping function which maps the situation features onto the A-components as shown in Figure 3. The system extracts the situation features in the classroom and detects several A-components which represents the current dynamic situation according to the mapping function, then controls an active camera to shoot the specified object.

The camera-work which is realized at the active camera consists of four elements as (object, object-location, camera-location, priority). It indicates that the *object* in the *object-location* should be shot by the shoot camera at the *camera-location*.

In the broadcast imaging style, the camera-works for answering the first question are given to the system in advance.

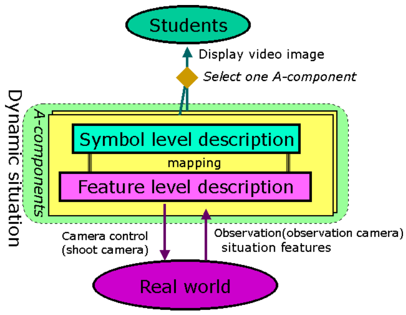

Figure 3: Two aspects of dynamic situation in a classroom

3.3. Situation Feature

We use two situation features which are sufficient to formulate the mapping function.

- Location of the lecturer
- Location of the active students

These locations are estimated by observing the classroom. We call the cameras used for this purpose *observation cameras*.

The region of the lecturer on the image plane of the observation camera is detected by checking motion region in the successive two video images. Then, the location of the lecturer is calculated with single camera by introducing the constraint that the lecturer always locates on a certain horizontal plane in the 3D world and by giving the camera parameters of location, direction, focal length, image aspect, and so on in advance. The location of the active students is also estimated by almost the same technique.

We use multiple observation cameras to cover almost all the area of the classroom.

3.4. Automatic Camera Control

We prepare the cameras some of which can pan, tilt, and zoom in/out and some are static as the shoot cameras. Although only one shoot camera is needed at a time, we use multiple shoot cameras in the classroom.

One of the reasons is that the lecturer may walk around in the classroom and the students in any seats may become active. In this case, the object to be shot may easily go out of the visible area of the static shoot camera. The other reason is that the lecturer may walk faster than the maximum speed of pan/tilt/zoom of the active camera. Therefore, the system should control multiple shoot cameras simultaneously.

When one A-component is detected and the corresponding camera-work uses an active camera, the camera will follow the object as long as the A-component is being detected.

3.5. Automatic Video Selection

As our automatic camera control method can let the shoot cameras image the object, the remaining and the most important issue is to select the most appropriate video stream for the remote classroom.

Our video selection method consists of two rules. The first rule is processed based on the dynamic situation in the classroom. The second rule is based on human intrinsic feature.

1. Select one of the detected A-components in the dynamic situation
2. Change camera under time constraint

First, if multiple A-components are detected at the same time, the A-component of which the object is a group of students takes precedence over the other A-components. This is because we conclude that Japanese students rarely become active during the lecture and so it is a noteworthy activity rather than anything else. If the behavior of the students varies, this rule should be changed accordingly.

Then, the shoot cameras that are assigned to the camera-works of the selected A-component become candidates to be used. Since there may be multiple camera-works (it means multiple cameras) for one A-component, we have to choose one of them under the time constraint we introduce here. There are two criteria for it; one is that a man/woman needs certain time to understand what he/she watches. Based on the criterion, we introduce “glance time” which limits the shortest time of keeping the video stream from a single camera. The other is that a person become bored if he/she watches the same object from the same camera location for a long time. To avoid this, we introduce “release time” which limits the longest time of keeping the video stream from the same camera.

Therefore, if there is a camera-work that lasts less than the glance time, it is always adopted. On the other hand, if there is no change of dynamic situation even after the release time is over and if there are the other camera-works corresponding to the same A-component, one camera-work among them is newly selected.

4. EXPERIMENT

We installed a distance learning system based on our method with 4 shoot cameras and 4 observation cameras.

We conducted an experiment on actual distance learning lectures which were held between Kyoto university and UCLA. The prepared A-components are as follows:

A	The lecturer is pointing on the screen-T.
B_{1-3}	The lecturer is talking to the students.
C_{1-12}	Some of the students are active.

B_{1-3} are the same A-component but are found on three different places. We divide audience area into 12 blocks so we need 12 A-components for C_i. As A, B_1, B_2 and B_3 represent the lecturer and $C_1, \cdots, C_{12}$ represent the students, they may occur at the same time.

The set of imaging rules given to the system in advance is shown in Table 1, and the mapping function is shown in Figure 4. If a moving object is found in each area, the corresponding A-component is detected. We use three static shoot cameras in the imaging rules.

Table 1: Imaging rules

A-component	camera-work	
	object	camera
A	lecturer	7
A	screen-T	5[fixed]
B_1	lecturer	7
B_1	lecturer	4[fixed]
B_2	lecturer	7
B_3	lecturer	7
C_{1-12}	student(1-12)	6[fixed]

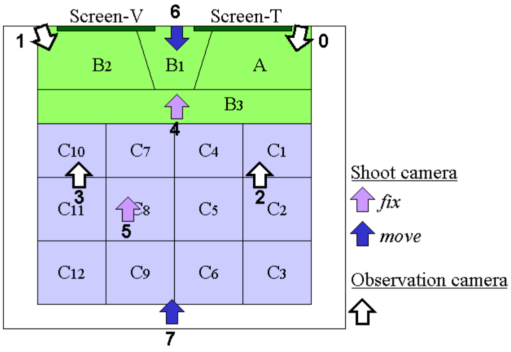

Figure 4: Mapping function of situation features

Figure 5 shows a sequence of generated lecture video transmitted to the remote classroom. Two kinds of arrows indicate video selection rule in the sequence.

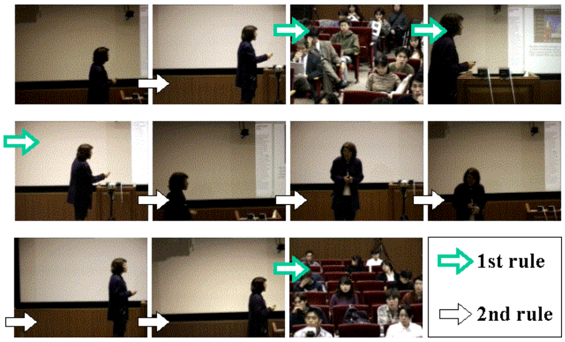

Figure 5: A generated video sequence

5. CONCLUSION

We have proposed a new approach of live video imaging method that is suitable to support distance learning. In this method, an automatic active camera control and video selection is realized by considering dynamic situation of the lecture.

Some students in the remote classroom reported us that the video imaging result contained the presentation information such as behaviors of the lecturer and the students, and was suitable to the lecture. It is considered that our experiment was successful on capturing the visual presentation information through the questionnaire survey we conducted against the remote students. We are planning to add hyperlinks onto the generated video from the other information such as hand-writings recorded on the electric whiteboard in the lecture.

Acknowledgment

This work has been partly supported by "Cooperative Distributed Vision for Dynamic Three Dimensional Scene Understanding (CDV)" project (JSPS-RFTF96P00501, Research for the Future Program, the Japan Society for the Promotion of Science).

References

[1] Yoshinari Kameda, Hideaki Miyazaki, and Michihiko Minoh, "A Live Video Imaging for Multiple Users," Proc. ICMCS'99, Vol.2, pp.897-902, 1999.

[2] Gregory D. Abowd, "Classroom 2000: An Experiment with the Instrumentation of a Living Educational Environment," IBM Systems Journal, Special Issue on Pervasive Computing, Vol. 38, No. 4, pp. 508-530, 1999.