A Virtual Camera Team for Lecture Recording

Fleming Lampi
Department of Computer Science IV
University of Mannheim
Mannheim, Germany

lampi@informatik.uni-mannheim.de

Manuel Benz
Department of Computer Science IV
University of Mannheim
Mannheim, Germany

benz@pi4.informatik.uni-mannheim.de

Stephan Kopf
Department of Computer Science IV
University of Mannheim
Mannheim, Germany

kopf@informatik.uni-mannheim.de

Wolfgang Effelsberg
Department of Computer Science IV
University of Mannheim
Mannheim, Germany

effelsberg@informatik.uni-mannheim.de

Abstract

We present the design of a virtual camera team for lecture recording based on the teamwork of a real camera team. A major problem with traditional lecture recordings is that they tend to be boring for the students, especially if only the slides and the audio of the lecturer are presented. In a first step, we determine the different roles in a camera team, their tasks and how they collaborate to apply cinematographic rules. We then adapt these results to a distributed computer system and show how they can be implemented. We present early evaluation results, and we conclude that lecture recordings can be much more lively and interesting using our approach.

1. Introduction

Lecture recordings have become very widely accepted, because students can participate without time constraints, repeating parts that are difficult to understand. But in many cases they tend to be boring, independent of how fascinating the original session was, especially if only the slides and the lecturers' speech are recorded.

Television has pushed our expectations by the quality we watch every day. Although students preparing for their exams are highly motivated, it would be really helpful to support their learning from recorded lectures by applying basic cinematographic rules during the recording.

But especially in times when universities have to save money it is far too expensive to hire a real camera team for lecture recording. In some cases it is possible to use university staff to replace a camera team, but even then it is unlikely to get the quality an experienced camera team would produce.

1

Thus we focus on the design and implementation of an automatic system allowing the recording and broadcasting of lectures in real-time. Furthermore, our system can cooperate with interactive learning tools used in the lectures [9, 10].

Close to our approach is the use of pan and tilt operations and image processing for framing and following the lecturer. A sample application is "AutoAuditorium" [1], which shows a basic level of automatic presentation recording but without any cinematographic rules. More advanced is the system developed by Microsoft Research [8], improved in [13]; it uses multiple cameras and implements video production rules. A video director module based on a finite state machine (FSM) is available which can be configured by a scripting language to implement basic cinematography rules.

Nevertheless, these earlier approaches differ significantly from our approach; other systems solely use image processing to determine the image framing and to track the lecturer while we also use an indoor positioning system. We are able to identify the positions of all tracked persons in the room. Thereby, we are able to implement more sophisticated cinematographic rules, e.g., two tracked persons may be framed in such a way that they face each other while the system switches between their shots. Our implementation of the cinematographic rules also differs. Microsoft uses a scripting language in which the rules are rewritten in a note form; this implies fixed durations of the shots leading to more predetermined transitions than with our model. Similar basic rules have been proposed by [3] for the recording of real-time applications.

2. A Human Camera Team

In contrast to the large staff in TV production, for lecture recordings we can focus on the camera team itself; for example, we don't need make-up artists or set constructors. At first, there is a cameraman, for each camera: one for a long shot (complete lecture hall), one for the lecturer, with the ability to follow him and his gestures, one for the slides, and one for the audience, when questions are asked. In addition, a director is needed to coordinate these cameramen and decide which stream to record. In order to capture the audio of the lecturer, of simulations, of videos and as well of the questions of students, we need a sound engineer. Lighting technicians complete the team.

The technical work of a cameraman performed during each shot consists of moving, panning and tilting the camera, and adjusting the exposure, the focus and the zoom. Besides these technical aspects, aesthetic work is an important part of a cameraman's job. To fulfil the viewer's expectations, teamwork of the entire camera team is necessary, and it starts long before the recording. In an initial meeting the director goes over the storyboard of the event and comments it. The cameraman gets the relevant information in three steps: first, out of the storyboard, second, during the meeting where he can amend the information given by the director, and third, during the recording session using the intercom.

Using the intercom, information about "who is on air", "who will be on air next" and "which detail or framing each cameraman should show" are given during the shooting. A cameraman also informs the director about his status, his inability to fulfil a requested shot for technical reasons, or about an extraordinary detail he wants to show. So, throughout the event, there is continuous communication among the team members to improve the aesthetic aspects of the recording. This communication is necessary to apply cinematographic rules. Typical rules are: Mind the line of action. Choose the duration of a shot so that all necessary details may be perceived and that the shot does not get boring. Define a beginning and an end for a pan. Show an overview or neutral shot after two or three close-up shots. Show the important details as close-ups to make them clear after showing the entire scene as a long shot. Do not

show the same series of shots one after another so as to not get predictable. Professional cameramen intuitively apply these rules. Many more cinematographic rules are known to professionals; see [11, 12] for good examples.

3. The Virtual Camera Team

In our approach we mapped the roles of each member of the team to a corresponding virtual pendant.

The virtual director is based on an extended finite state machine (FSM). The states correspond to the different types of shots. The transitions describe the possibilities to go from one shot to another. Each transition is initialized with a given probabilistic value which is increased or decreased by inputs from the sensors. Based on recent history, a transition leading to a camera shown recently gets decreased reducing its probability. Using automatic motion detection algorithms, well known from the multimedia community, transitions leading to shots with more activity get an increase of their probabilistic values. If a question is asked by a student and recognized by external sensors the transition to a shot showing the student is increased considerably. When time has come to make a transition, the transition with the highest value is selected. The behaviour of the director is always similar but seldom identical and thus less predictable. The finite state machine with all its details is loaded at runtime from an XML file. This enables easy adaptation of the FSM to different recording scenarios. More details on the implementation of the virtual director can be found in [7]. Figure 1 shows an FSM example for the director of our system.
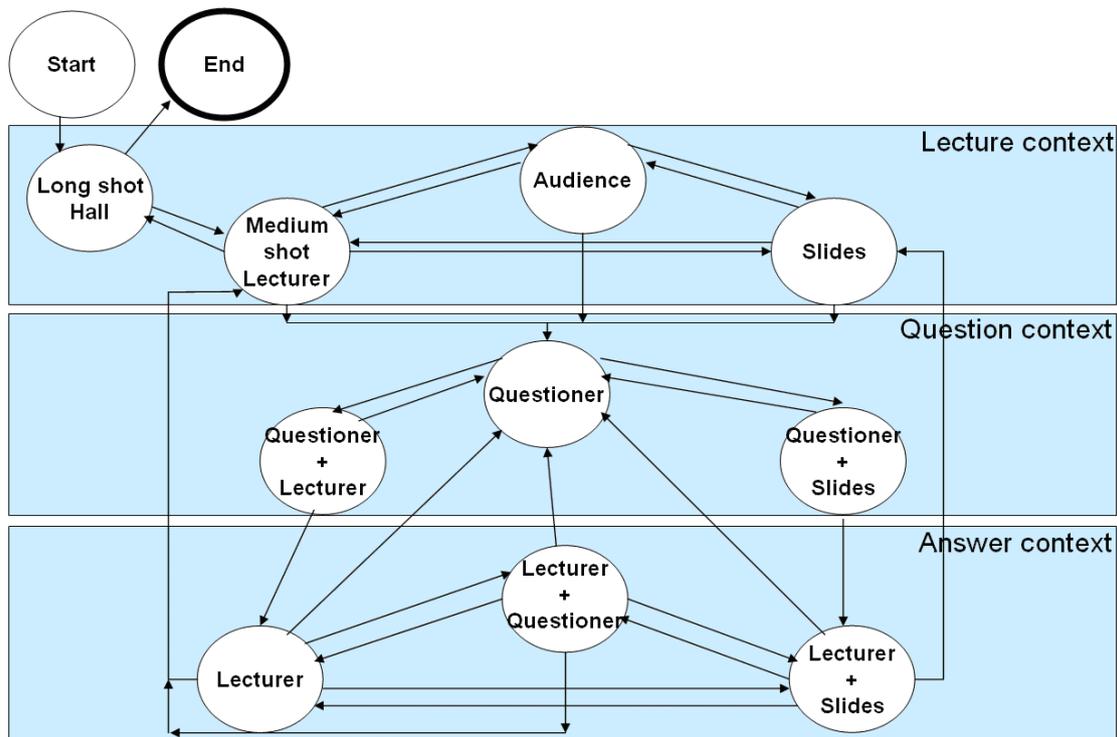
Figure 1: Example of the FSM of the director

As shown before, the work of cameramen consists of two parts, the technical work and the aesthetic part. We regard the technical work as a control-loop, which starts even before the recording, e.g., by selecting whether to use a certain grey filter, so called ND-filter. We use well-known image content analysis algorithms to find people in the image, to determine a correct exposure setting, even in backlight situations, etc. For example, in lecture recordings the background of an image is of no interest. But the people in front have to be shown in an appropriate way. We use algorithms like skin colour detection and face recognition to determine the areas of an image showing a person. Then, we adjust the iris to optimize the exposure for this person. It does not matter if the background gets too bright or too dark. The flowchart of the control-loop process is shown in Figure 2.
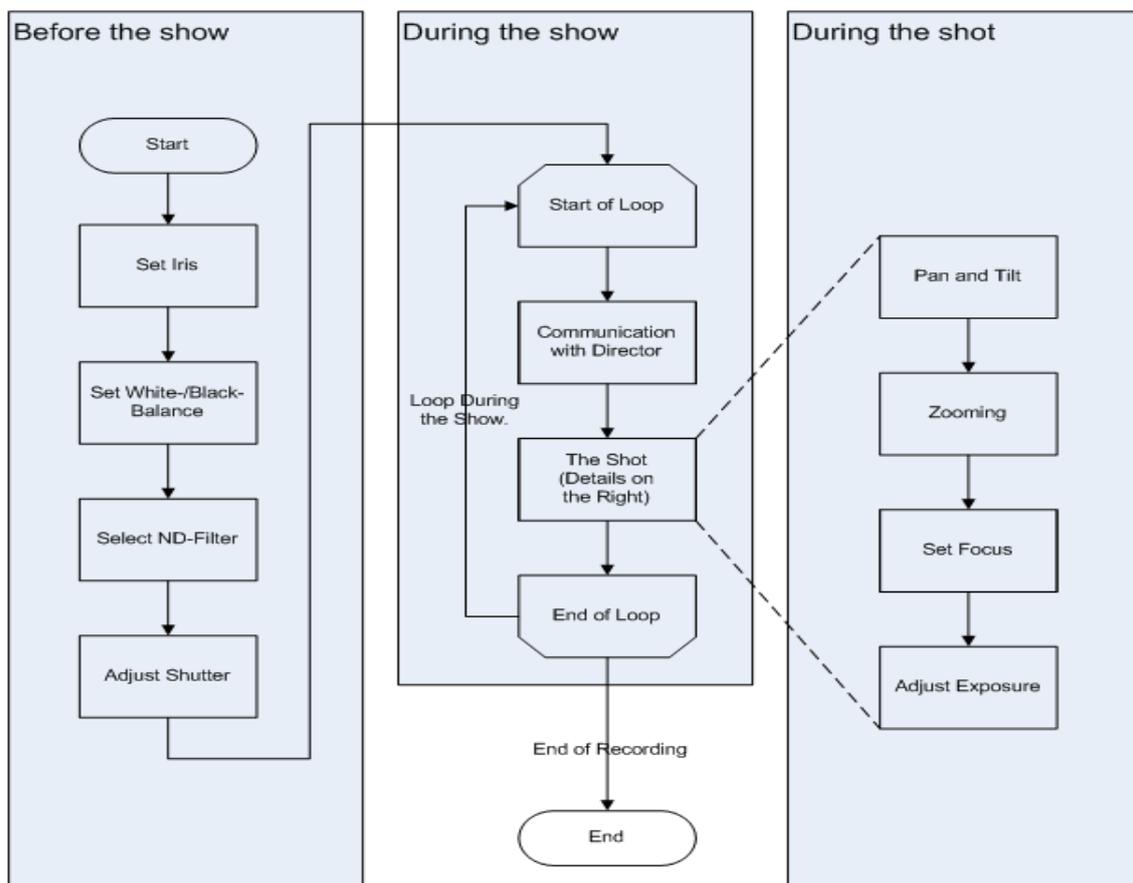
Figure 1: The camerawork as a flowchart

For the aesthetic part of a cameraman's job the cinematographic rules have to be implemented. We divide the cinematographic rules into two categories: One group can be realized directly by one cameraman alone, the second group requires the collaboration of the team or at least of the cameraman with the director. A typical example of the first category is the reaction to a person starting to gesticulate: The

cameraman zooms out until the person and his movements are completely visible in the picture. This type of rules is implemented again by using image content analysing algorithms, here motion detection. Typical for the second category is the shot/counter shot arrangement of a dialog. One person is shown looking from the left edge of the frame to the right; the next shot shows the other person looking into the opposite direction. The director gives the order to the cameramen and is then able to switch between these two cameras. Cameramen and director have to communicate a lot. We have implemented that communication in our virtual system with an XML-based protocol over TCP. Cameraman and director exchange all necessary information like commands, acknowledgements, alerts and status reports. More details concerning the virtual cameraman can be found in [6].

Unlike a real camera team our virtual team is additionally based on sensors. For example, we use an indoor positioning system based on 802.11 access points to identify the places of the students. We use the interactive devices already used in our lectures and implemented a client/server based question manager to cope with students asking questions and their determined locations. Thus we are able to adjust the audience camera accordingly. As there are many difficulties using 802.11 indoor location systems we have taken the circumstances in our lecture hall into account, as it is described in [4, 5].

For the sound engineer we plan to use the work of Gerald Friedland as described in [2]. The automatic lighting technician is foreseen for a later time, because the lighting conditions in lecture halls are usually sufficient.

4. First Evaluation Results

In autumn semester 2007 we started to test our system in the lecture hall. Step by step one module after another is brought into the test system. The director already performs well and communication with the cameramen is stable. The cameramen itself is basically working well, but still need some fine tuning to not overreact. As expected the indoor position system has to be perfectly adjusted to the lecture hall to minimize the position error and the question manager needs a good interface. Besides improving the system, the main work will go into a virtual video switcher/mixer and the implementation of the sound engineer. Figure 3 gives an overview of the entire system in the lecture hall. The areas highlighted in red mark the cameras for the long shot, the lecturer and the audience and the hardware to record the slides.

Figure 3: The virtual camera team in action

5. Conclusion

A real-world camera team recording or broadcasting a lecture can be described as one that artfully reacts to events and to changes of contexts as the recording goes on. Cinematographic rules are guidelines how to best record specific types of scenes, and how to react to changes as a team. The experience of a director, of each cameraman and of the entire team determines how and to which extent these rules are applied. We have implemented our virtual camera team, applying the same rules. Our distributed approach, with well-defined tasks for each module, has two significant advantages: First, the workload is distributed, e.g., the cameraman modules and not the director module produce the images. Second, it is easier to implement even complex cinematographic rules using the well-defined roles of the virtual team members and the communication between them. Using this approach, the behaviour of the virtual camera team comes closer to the behaviour of a human camera team and thus leads to more lively recordings.

One major difference compared to a human camera team is that some tasks analysing a picture are deferred from the virtual director to the virtual cameramen to better distribute the workload. The virtual camera team is also limited to the set of implemented cinematographic rules. Therefore, it will always be an imitation of the human original. Our long term goals are the implementation of further modules for lecture recordings, the improvement of the implementation of cinematographic rules and a more complete evaluation of the recorded courses.

6. Acknowledgement

7. References

[1] Bianchi, M. H. AutoAuditorium: A fully automatic, multicamera system to televise auditorium presentations, *Proc. Joint DARPA/NIST workshop on smart spaces technology*, 1998.

[2] Friedland, G. *Adaptive Audio and Video Processing for Electronic Chalkboard Lectures*, Dissertation, Faculty of Mathematics and Computer Science, Freie Universität Berlin, October 2006.

[3] He, L., Cohen, M. F., Salesin, D. H. The virtual cinematographer: A paradigm for automatic real-time camera control and directing, *Proc. ACM SIGGRAPH,* 1996, pp. 217-224.

[4] King, Th., Haenselmann, Th., Effelsberg, W. Deployment, Calibration, and Measurement Factors for Position Errors in 802.11-based Indoor Positioning systems, *Proc. 3rd International Symposium on Location- and Context-Awareness (LoCA)*, 2007, pp. 17-34.

[5] King, Th., Kopf, S., Effelsberg, W. Position detection of students in lecture halls using the Chi-Square-Adaptation-Test. (In German: Positionserkennung von Studierenden in Hörsälen mit dem Chi-Quadrat-Anpassungstest), *Proc. 3rd GI/ITG KuVS Fachgespräch "Ortsbezogene Anwendungen und Dienste"*, 2006, pp. 44-48.

[6] Lampi, F., Kopf, S., Benz, M., Effelsberg, W. An Automatic Cameraman in a Lecture Recording System, *Proc. ACM Multimedia, EMME Workshop*, 2007, pp. 11-18.

[7] Lampi, F., Scheele, N., Effelsberg, W. Automatic Camera Control for Lecture Recordings, *Proc. ED-MEDIA*, 2006, pp. 854-860.

[8] Rui, Y., Gupta, A., Grudin, J., He, L. *Automating lecture capture and broadcast: Technology and videography*. ACM Multimedia Systems Journal. Vol.10, No.1, pp. 3-15, 2004.

[9] Scheele, N., Mauve, M., Effelsberg, W., Wessels, A., Horz, H., Fries, St. The Interactive Lecture - A new Teaching Paradigm Based on Ubiquitous Computing, *Poster. Proc. CSCL*, 2003, pp. 135-137.

[10] Scheele, N., Seitz, C., Effelsberg, W., Wessels, A. Mobile devices in Interactive Lectures, *Proc. ED-MEDIA*, 2004, pp. 154-161.

[11] Thompson, R. *Grammar of the edit*, Elsevier Focal Press, Oxford, 1993.

[12] Thompson, R. *Grammar of the shot,* Elsevier Focal Press, Oxford, 2nd edition, 2002.

[13] Zhang, C., Crawford, J., Rui, Y., He, L. An Automated End-to-End Lecture Capturing and Broadcasting System, *Proc. ACM Multimedia*, 2005, pp. 808-809.

Fleming Lampi is a PhD student and research assistant at the Department of Computer Science IV, University of Mannheim, Germany. His research interests include video recording, processing and transcoding. He received an MS in computer science and multimedia from the University of Applied Science in Karlsruhe, Germany.

Stephan Kopf received his Diploma in Business Administration and Computer Science in 2000 and his doctoral degree in computer science in 2007 from the University of Mannheim, Germany. He is working as a postdoctoral researcher at the Computer Science IV research team in Mannheim. His research interests are multimedia content analysis and new learning technologies.

Manuel Benz received his diploma in computer science from the University of Mannheim, Germany in 2007. His research interests include video processing and analysis.

Wolfgang Effelsberg is head of the Department of Computer Science IV at the University of Mannheim, Germany. His research interests include computer networks, multimedia systems and e-learning. He received his PhD from the Technical University of Darmstadt, Germany. He is a member of IEEE and ACM and serves as a member of the editoral board of several major multimedia journals as well as of the program committee of the IEEE and ACM multimedia conferences.