

# Social Editing of Video Recordings of Lectures

**Margarita Esponda-Argüero**  
esponda@inf.fu-berlin.de

**Benjamin Jankovic**  
jankovic@inf.fu-berlin.de

Institut für Informatik  
Freie Universität Berlin  
Takustr. 9  
14195 Berlin

January 2008

## Abstract

In this paper we show how video recordings of lectures can be enhanced through “social editing”, a special sort of “human computation”. We record lectures using six cameras arranged to produce a panoramic view (360 degrees) of the classroom. After recording, viewers can replay the lectures on a computer and can navigate through them tilting and zooming a virtual camera. All user interactions are recorded by the system, and are later used to decide which views should be stored and their zoom level. The result is an automatically edited video as a result of the social navigation of the first few initial viewers of the lecture.

## Keywords

Automatic video editing, social navigation, virtual videography, human computation.

## ACM Classification Keywords

H.3.3 Information Search and Retrieval: Information filtering, Relevance Feedback. H.3.5 Online Information Services: Web-based services.

## 1 Introduction

Video recording of lectures has become increasingly inexpensive and easy to implement in the classroom. The price of the video cameras and the associated equipment has fallen steadily during the last years. Unfortunately, the main cost factor is still the cameraman, and/or the editing process necessary for transforming a raw recording into a high-quality movie. Unfortunately, there are many examples of recordings of lectures on the Web which are almost unusable because the slides cannot be read, or because the handwritten symbols on the blackboard are too small, because the lecturer disappears from the frame, etc.

We have been experimenting during the last years with a multimedia classroom in which the lecturer writes on a contact sensitive screen. The handwritten input and pasted pictures can be rendered in high-quality during replay of the lecture, and are also clearly visible in the printed version of the chalkboard contents. The E-Chalk system [1] provides a video inset of the lecturer, together with a window with a high-resolution view of the chalkboard.

Unfortunately, a multimedia lecture room is not available everywhere at the university, and some lecturers do not really want to interact with a computer. In some cases, the lecturer just reads from prepared notes (for example, in philosophy courses), in others the interaction with the chalkboard is negligible. Many lecturers prefer to use PowerPoint slides, or similar presentation systems, because it makes course preparation and material reuse easier.

For this kind of lecturers we wanted to provide an unobtrusive system, one which can be started at the push of a button, and which does not require a cameraman or video post-processing. Instead of having several cameras and use some kind of computer vision method for locating and tracking the lecturer, we decided to record *everything* happening in the classroom. We record the lecturer, but we also record the audience and all sides of the lecture room using a panoramic camera. When the lecture is replayed in the computer, the viewer has a limited view –she can only observe the picture taken by one or 1.5 cameras. However, using the mouse it is possible to “turn around”, that is, it is possible to navigate in the panoramic view of the classroom. For example, one can observe the lecturer, but if a question comes from the back of the classroom it is possible to turn around the virtual camera and look at the person asking the question, just as we would do if we were physically present.

This kind of video recordings are very attractive for students because they enhance the feeling of physical presence. The downside of this technique, however, is the large size of the stored video files. Since we are recording from six multi-megapixel cameras, the video files are very large.

In our project we reduce the size of the video files by editing the contents. This is done by the viewers themselves, who navigate in the lecture (in the original large file), and thus provide us important clues about the frames which should be preserved in the final version. “Social navigation”, used already in other domains, can be used here to compress the data files and produce a more appealing recording, one in which even the questions from the audience can be seen in the edited version, and in which anything pointed to by the lecturer is also clearly visible in the video.

## **2 Virtual Videography**

There are two main approaches for automatically editing video recordings: on-line automatic editing works with several cameras and possibly several microphones, or with a movable camera. In this case a computer tries to identify the best possible shots and selects the camera to be used and the level of zooming. In the off-line approach, the raw material coming from several cameras is selected automatically by the computer and is arranged in a final edited video.

The off-line approach has been championed across the years by Heck, Wallick, and Gleicher, who call their technique “virtual videography” [2]. In this framework, the complete raw material is available to a computer (from different cameras). The computer uses past, present, and future frames, for example, to find the background. Given the background it is easy to segment the lecturer and track him, or it is even possible to render his image semitransparent (“ghosting”). Gestures from the lecturer can be interpreted and can be used to enhance blocks of symbols on the blackboard, or zoom into them.

The on-line approach to automatic editing has been followed by many different groups [3]. In some cases, tracking the lecturer already provides some clues as to the best camera shot. More complex systems try to analyze the gestures of the lecturer and its movements, in order to provide an optimal view from cameras at different angles [4].

Both off-line and on-line automatic editing is still unsatisfactory. Questions from the audience remain unrecorded, and when the lecturer calls the

attention to some part of the presentation (“look at formula 3 on the blackboard”), it is impossible to extract the adequate frame without semantic analysis.

For these reasons we decided to follow an alternative approach: *Let the user decide where the information is.*

### 3 Recording Hardware

For our project we acquired a panoramic camera from Point Grey Research, the Ladybug. This composite optical system consists of six cameras arranged in such a way as to provide a composite image of around 70% of a sphere centered around the system (see Fig. 1).

The camera system compresses the individual images as JPEG files. A library and SDK provided with the system allows the user to replay videos. The video player shows a view on the screen of the captured hemisphere, as if the viewer was positioned inside the camera. With this player it is then possible to rotate and zoom the camera view. The rotation is used to “turn” the head around, and zooming to see details better.



**Figure 1.** The Ladybug camera used for recording lectures in the classroom. The system consists of five cameras arranged radially, and sixth camera pointing up. The six cameras cover 70% of a spherical view centered at the system.



**Figure 2:** The Ladybug records a panoramic view of its surroundings. In this example we can see the complete classroom: the lecturer stands to the right, all students are sitting from the middle of the picture towards the left.

The Ladybug does not come with a microphone, and frame synchronization is not part of the inherent features of the system (the player plays frames as fast as it can). Recording video implies therefore, that the programming library had to be augmented with frame synchronization to a parallel audio channel.



The upper row of Fig. 4 shows a lecture in the panoramic view (which would not be what the user usually selects). The two middle images show two frames selected by panning the virtual camera: one can see the lecturer pressing a key on the laptop. A student asks a question (right middle image). The viewer can turn the virtual camera towards the student, then back toward the lecturer, who is answering the question, and again towards the student. The user himself is the camera man and selects the best virtual camera position in order to better understand the lecture.



**Figure 4:** The upper image shows a panoramic view of the classroom. The sequence of smaller images show some views selected by a user replaying the lecture who turns the virtual camera back and forth to a student who is asking questions.

As can be seen from this few examples, the result is very natural. Having a complete panoramic view of the lecture allows us replay and view everything that happened in the classroom.

## **5 Compressing and editing the lecture**

We have experimented producing compressed lectures where the panoramic images have been replaced by the sequence of views selected by “expert viewers”. These are teaching assistants or the lecturer himself, who know exactly what kind of information should be visible on the screen. The traces of the views are stored as the final version of the edited lecture. This reduces the space needed for the files by a considerable factor.

In future experiments we will collect the traces of videos viewed by naïve users, that is, students who learn the material for the first time. There will be some conflicts in the views selected by such viewers and we will have to adopt an automatic mechanism for selecting between a range of frames. In most cases we do not expect too much variation (as we observed with the expert users). In case of conflicts we will have to adopt some averaging or median selection mechanism.

## **6 Summary and Future Work**

Our social video editing system is being developed. The first lectures have been recorded and we are extending the SDK for the replay of panoramic lectures with more functionality.

Our approach is an example of a novel application of “human computation” [6]. In this paradigm, computer users collaborate to solve a problem, usually by interacting in a computer game (for labeling images, for example). In our approach, the viewer has a direct interest in understanding the lecture and we do not have to embed the application into a computer game. There might many other problems solvable by “human computation” on the web which go beyond computer gaming.

One problem we have with the current hardware is that the resolution of the individual cameras is sometimes too low to capture small details on the blackboard. Also, contiguous views from two cameras are not stitched together, they are interpolated. At the boundary between cameras the image resolution suffers.

Therefore we are developing our own multicamera system. For recording the lecturer we will use three or more cameras pointed directly to the blackboard, in order to enhance the resolution. For the audience we will only reserve one camera. The geometric arrangement will be different and

will have to be mapped to the video player, but zooming into details will be greatly enhanced. Experiments with large camera arrays show that even custom hardware can produce high-quality results [7].

We are also developing a miniature camera system, with an arrangement similar to the Ladybug, but much smaller. We would like to reduce the size of the system to the volume of an aluminum can. Such small cameras could be transported to classrooms, and all processing would be done internally. Storage of the videos would be done by synchronizing the device with a laptop or a desktop.

### **Acknowledgements**

We thank the “Ernest-Reuter-Gesellschaft der Freunde, Förderer und Ehemaligen der FU Berlin” for the acquisition of the hardware necessary for our project. The lectures produced with our system will become available through the “Network University” just started by FU Berlin with offices in seven countries.

### **References**

- [1] Rojas R., Knipping L., Raffel U. and Friedland G. Elektronische Kreide: Eine Java-Multimedia-Tafel für den Präsenz und Fernunterricht. *Informatik: Forschung und Entwicklung* 16, 3 (2001), 159–168.
- [2] Heck, R., Wallick, M., and Gleicher, M. Virtual Videography. *ACM Transactions on Multimedia Computing, Communications and Applications* 3, 1 (2007), 1-27.
- [3] Lampi, F., Kopf, S., Benz, M., and Effelsberg, W. An Automatic Cameraman in a Lecture Recording System. *ACM Educational Multimedia and Multimedia Education* (2007), 11-18.
- [4] Ozeki, M., Nakamura, Y., Ohta, Y. Video Editing Based on Behaviors-for-Attention. *IEEE International Conference on Multimedia and Expo* (2004), 2215-2218.
- [5] Shamma, D. A., Shaw, R., Shafton, P., Liu, Y. Watch what I watch – Using Community Activity to Understand Content. *ACM Multimedia Information Retrieval* (2007), 275-283.
- [6] Ahn, L. Games with a Purpose, *Computer* 39, 6 (2006), 92- 94.
- [7] Wilburn, B. et al. High-Performance Imaging Using Large Camera Arrays, *ACM Transactions on Graphics* 24, 3 (2005), 765-776.