Virtual Videography

Rachel Heck, Michael Wallick, and Michael Gleicher
Computer Sciences Department
University of Wisconsin, Madison
1210 West Dayton Street
Madison, WI 53706
{heckr, michaelw, gleicher}@cs.wisc.edu

ABSTRACT

Well-produced videos provide a convenient and effective way to archive lectures. In this demonstration, we present a new way to create lecture videos that possess many of the advantages of well-composed recordings without the cost and intrusion of a video production crew. The videos are produced by an automated system called *Virtual Videography* that employs the art of videography to mimic videographer-produced videos, while being unobtrusive when recording the lectures. The system uses the data recorded by unattended video cameras and microphones to produce a new edited video as an offline post-process. By producing videos offline, our system can use future information when planning shot sequences and synthesizing new shots. Using syntactic cues gathered from the original video and a novel shot planning algorithm, the system makes cinematic decisions without any semantic understanding of the lecture.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Video

General Terms

Algorithms, Human Factors

Keywords

Automated Video Editing

1. INTRODUCTION

A "traditional" lecture, where an instructor presents in front of a chalk- or marker-board, is a common method for presenting information. Archiving these events on video can be valuable for those who did not attend or for someone who wants to review its contents. Unfortunately, simply recording a lecture with a fixed, unattended camera does not produce a video with the production value that viewers expect. Effective videos exhibit cinematographic qualities that, in the past, have required the skills of an expensive, intrusive videographer or video production crew. Our goal is to provide inexpensive, yet effective, archiving of traditional lectures without disrupting the lecture.

In this demonstration, we present an approach to lecture video production that uses unattended, stationary video cameras and microphones to record the event and a novel, automatic editing system that post-processes this data to produce a new video that exhibits many of the properties of a videographer-produced one. We call our approach *Virtual Videography* as it simulates many aspects of a production crew. Virtual Videography uses computer vision and signal processing methods to gather information about the lecture, a planning algorithm to choose appropriate shots, and image synthesis methods to generate new images from the original source footage. With little to no human intervention, the system produces videos containing a variety of different shots and visual effects that serve to guide the viewer's attention and maintain visual interest, without violating the rules of cinematography.

A videographer relies on knowledge of the art of filmmaking and understanding of what is happening in a scene to inform their decisions. In our efforts to create a virtual videographer with these same characteristics, four key insights emerged.

1. ***Syntactic cues*, such as marks on the board and motions of the instructor, can inform editing decisions.** Since current techniques cannot reliably extract the semantics of a lecture, the ability to make editing decisions based on easily identifiable cues makes Virtual Videography viable.

2. **By processing the video *offline*, we not only gain more time to perform analysis and synthesis but, more importantly, the ability to use more information to make decisions.** We can process large windows of time to provide robustness to simple image analysis algorithms, look into the future to determine how to best plan shots, and use future images to synthesize visual effects.

3. **We pose the editing problem as a *constrained optimization over space and time*.** A good video not only frames individual shots in a pleasing manner but also combines them effectively. As such, the quality of a video cannot be determined by evaluating each shot in isolation. We therefore encode composition and editing goals in an optimization objective and solve for a whole *sequence* of shots.

4. **Rather than relying on controlling a camera to acquire desired viewpoints, *image synthesis* can be used to create the views deemed necessary by Virtual Videography.** View synthesis allows us to avoid using special purpose movable cameras and to postpone editing decisions until after the lecture is captured.

These four insights allow Virtual Videography to produce lecture videos that exhibit the production value viewers expect. At the same time, the process is unobtrusive and does not require the instructor to alter his or her lecture style for the camera; setup requires no special equipment and works without calibration in a wide variety of classrooms; and editing requires little to no human

MM'06, October 23–27, 2006, Santa Barbara, California, USA.
ACM 1-59593-447-2/06/0010.

input. Furthermore, the system can produce multiple videos of the same lecture, each tailored to a specific need or viewer preference. In particular, Virtual Videography can produce videos of varying aspect ratios from the same source footage, making better use of the available screen space on different devices.

Our emphasis on automatically creating inexpensive videos of traditional lectures places a number of constraints on our system: we capture a limited amount of source footage, the system has no true understanding of the lecture content, and we place few requirements on the structure of the lecture. Yet, despite all of these constraints, Virtual Videography is able to create videos that exhibit desirable properties of good video. The system may not compete with videographers with years of experience, human understanding, and special equipment, but it can create effective lecture videos with much lower expense and intrusion.

2. FRAMEWORK

The main challenge to automatically producing effective lecture videos is to recreate the skills of a videographer. Rather than recreate the senses and thought processes of human videographers, we cast the decision-making process in the computational framework described in this section.

A goal in videography is to direct the viewer's attention towards what is important [Katz 1991]. Therefore, a central portion of our system is a representation for potentially important objects in the scene. We use a concept called a *region object*, which is the smallest rectangular region of the board that contains a semantically-linked group of writing [Wallick et al. 2005]. For example, a region object could contain a sentence, a diagram, a bullet list, or a single thought. We have found that these region objects not only provide a useful implementation strategy but also provide a way to cast filmmaking knowledge in a computational framework. These objects, as well as the region surrounding the instructor, describe the parts of the original video that could contain important information. As such, each phase of the Virtual Videography system uses these region objects.

There are four phases in the Virtual Videography system.

Media Analysis: Provides the system with simple events and properties of the original video data, such as a list of region objects, an audio analysis, and gesture recognition results.

Attention Modeling: Using the set of region objects supplied by media analysis, builds an evolving probability for each region specifying the likelihood of that region being a focus of the audience's attention at any instant in time.

Computational Cinematography: Does the job of the cinematographers, directors, camera operators, and editors in a conventional video production crew. First, virtual cameras analyze the available information to frame effective shots throughout the video. These shots are then assembled into a *shot sequence graph* that encodes all of the possible shot sequences. Next, an optimization strategy is used to find the shot sequence that minimizes an objective function designed using the art of film and knowledge of the lecture domain. Finally, transitions are added between appropriate shots to smooth the virtual camera movements.

Image Synthesis: Constructs the final video by synthesizing the views and visual effects specified by the computational cinematography component.

See Figure 1 for a diagram of the Virtual Videography system.

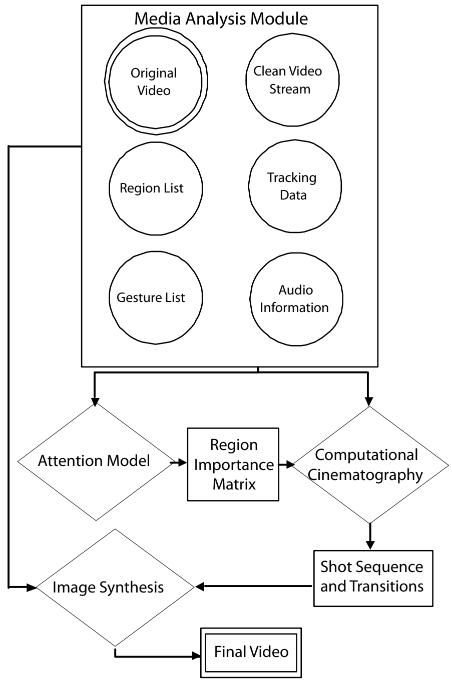

Figure 1: A diagram describing the structure of Virtual Videography.

3. CONCLUSION

The examples presented in this demonstration show that our system can assemble videos that exhibit multiple shot types, visual effects, and transitions while following many of the rules of cinematography. Details of our approach and implementation can be found in [Heck et al. 2007].

Since our system postpones the video planning process until after the lecture, it is able to produce several different videos of the same event tailored to specific needs. In fact, unlike any other automated editing system, Virtual Videography can even create novel videos from archived lectures. We have shown that our system is capable of producing videos of the same lecture at different aspect ratios. In particular, we have been able to produce 4:3 video from a 16:9 HD video. Our ability to deal with this type of data will become more useful in the future since HD video is becoming increasingly popular because of its high resolution.

Our system meets our goal of producing videos in our chosen editing style from limited inputs. One particulary interesting area of future research is determining what makes lecture videos educationally effective. Our system is extensible, making it possible to incorporate better rules as they are discovered.

4. REFERENCES

HECK, R. M., WALLICK, M. N., AND GLEICHER, M. L. 2007. Virtual videography. *To appear in ACM Transactions on Multimedia Computing, Communications, and Applications*.

KATZ, S. 1991. *Film Directing Shot by Shot: Visualizing from Concept to Screen*. Michael Wiese Productions.

WALLICK, M. N., HECK, R. M., AND GLEICHER, M. L. 2005. Chalkboard and marker regions. In *Proceedings of Mirage - Computer Vision/Computer Graphics Collaboration Techniques and Applications*.