Active Capture: Automatic Direction for Automatic Movies (Video Description)

Bibliographic Reference:

Marc Davis. "Active Capture: Automatic Direction for Automatic Movies (Video Description)." In: *Proceedings of 11th Annual ACM International Conference on Multimedia in Berkeley, California*, ACM Press, 602-603, 2003.

Active Capture: Automatic Direction for Automatic Movies

Marc Davis
University of California at Berkeley
School of Information Management and Systems
Garage Cinema Research
http://garage.sims.berkeley.edu
marc@sims.berkeley.edu

ABSTRACT

Current consumer media production is laborious, tedious, and produces unsatisfying results. To address this problem, Active Capture leverages media production knowledge, computer vision and audition algorithms, and user interaction techniques to automate direction and cinematography and thus enables the automatic production of annotated, high quality, reusable media assets. Active Capture is part of a new computational media production paradigm that transforms media production from a manual mechanical process into an automated computational one that can produce mass customized and personalized media integrating video of non-actors. The implemented system automates the process of capturing a non-actor performing two simple reusable actions ("screaming" and "turning her head to look at the camera") and automatically integrates those shots into various commercials and movie trailers.

Categories and Subject Descriptors: D.2.2 [Software Engineering]: Design Tools and Techniques; H.1.2 [Models and Principles]: User/Machine Systems; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Audio input/output, Video; H.5.2 [Information Interfaces and Presentation]: User Interfaces; I.4.9 [Image Processing and Computer Vision]: Applications; J.5 [Computer Applications]: Arts and Humanities.

General Terms: Algorithms, Design, Human Factors

Keywords: Metadata, video capture, human-in-the-loop, recognition, automated direction, automatic direction, automatic movies, active capture.

1. INTRODUCTION

The current media production process for consumers is labor-intensive and inefficient yielding usually unsatisfying results. It is a commonplace that more home video is shot than is edited, but it is also true that more home video is shot than is watched even a first time. While lacking the same time, resources, training, and motivation as professional media producers, consumers must struggle to produce videos in much the same way and with many of the same tools as professionals and as a result are left with piles of unedited, unwatched tapes, guilt, and frustration. Consumers are also confronted on a daily basis with the radical asymmetry between the production values of the videos they can produce themselves and the highly produced content they see on television and in the movies. To change this situation requires not just the computerization of existing consumer or professional production methods, but the reinventing of media production from a manual mechanical process to an automated computational one [1]. In support of that reinvention, Active Capture [1,2,3] focuses on the beginning of the media production process—the point of capture—and works to ensure the production of high quality reusable annotated media assets by encapsulating some of the expertise of a human director in software.

MM'03, November 2-8, 2003, Berkeley, California, USA.
ACM 1-58113-722-2/03/0011.

2. CONSUMER VIDEO PRODUCTION

As we shall see, the lack of media production solutions that can create metadata and that can use that metadata to automate media production and reuse is the principle source of the difficulty and frustration of consumer (and even professional) video production. As briefly demonstrated in our video, consumers face a host of worrisome challenges in the current mode of video production. Because of the lack of metadata, consumers produce video that often resides on unlabelled videotapes and/or in computer files with only a filename. The process of adding metadata is currently so cumbersome (both in the physical realm of videotape labels and on the computer), that it rarely occurs, and if it does, the descriptions produced are often overly terse and even cryptic. Most consumers are not and do not want to be professional motion picture archivists even of their own content. Consumers are also not trained in how to shoot video (cinematography), how to interactively guide who and what they are shooting (direction), or how to select and edit what they have shot (editing). Moreover, most consumers do not have the time or inclination to acquire the skills of a director, cinematographer, or editor (nor should they have to). As a result, consumer videos routinely suffer from poor production values.

However, consumers are surrounded by and relate to popular content with high production values. It is illustrative to think of the differences between consumer (personal) video and professional (popular) video content. As Tim Oren has observed, consumer videos have high salience—we usually care about and are interested in seeing our friends and family—but low production values, while popular media has high production values, but no guarantee of salience. A $100 million movie can be gone from the box office in a week because no one cared about the characters and their story. In Garage Cinema Research (http://garage.sims.berkeley.edu), we are working to enable daily media consumers to become daily media producers in a way that combines the salience of personal media production with the production values of professional media production thereby also opening up the possibility of greater personal participation in popular media.

3. ACTIVE CAPTURE

Our work in Garage Cinema Research focuses on three synergistic areas:

1) *Media Streams* provides a framework for creating metadata throughout the media production cycle to make media assets searchable and reusable

2) *Active Capture* automates direction and cinematography using real-time audio-video analysis in an interactive control loop to create reusable annotated media assets

3) *Adaptive Media* uses adaptive media templates and automatic editing functions to mass customize and personalize media and thereby eliminate the need for editing on the part of end users

The challenges of creating and using metadata to automate media production and reuse can best be met by automating metadata creation *at the point of capture*. At the point of capture, the world, the user, and the capture device are available to enter into an interactive dialogue to ensure the creation of *high quality media assets with detailed metadata*. To accomplish this goal, Active Capture integrates capture, interaction, and processing. Active Capture is a new paradigm of media capture and algorithm design that connects people, machines, and the world into a cybernetic system utilizing active engagement and communication among the capture device, the human agent(s), and the environment. Active Capture re-envisions capture as a control process with feedback in which the Active Capture device is able to issue the types of instructions and corrections a human director offers an actor or a portrait photographer offers a subject Furthermore, Active Capture overcomes some of the limitations of the standard approach in computer vision and audition—that of striving for complete automation that avoids interaction with human beings—by employing "human-in-the-loop" algorithms that interactively simplify the capture scenario through dialogue with a human agent More detailed descriptions of the ideas behind Active Capture and the process can be found in [1,2,3].

The Active Capture system depicted in the video automates the functions of human director and cinematographer in software (See Figure 1).

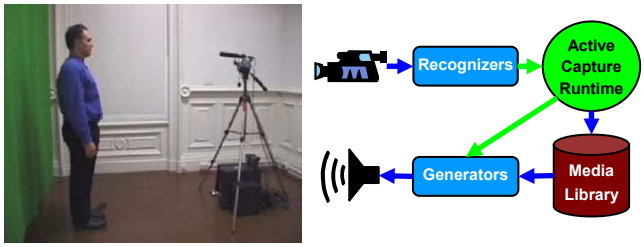

Figure 1: Active Capture System View and Diagram

The Active Capture system talks to the human capture subject and analyzes and responds to their motions and sounds in order to capture two highly reusable shots: a "Scream" shot (See Figure 2) and a "Head Turn" shot (See Figure 3). The Active Capture system uses simple video and audio recognizers (eye detection, motion detection, and audio pause detection) together with human-computer interaction scripts represented in a Finite State Machine in the Active Capture Runtime to effectively function as a more robust human-in-the-loop recognizer. By interacting with the human via generators which deliver the next instruction in the state machine using the appropriate file from the Media Library, the Active Capture system can ensure high quality of both the media captured and the metadata assigned to it.

Figure 2: Storyboard of the Scream Shot

Figure 3: Storyboard of the Head Turn Shot

Active Capture leverages what humans and computers are respectively good at by coupling effective (ideally game-like) interaction design with computer vision and audition recognizers. To understand how this approach to algorithm design and media capture differs from the standard approach of complete automation, imagine designing the following recognizers using *no processing at all*: a motion "recognizer" that uses the "Simon Says" interaction paradigm to capture and annotate various user motions (e.g., jumping, clapping, etc.); an object "recognizer" that uses the "Treasure Hunt" interaction paradigm to capture and annotate shots of various objects (e.g., cars, dogs, etc.). With the addition of simple recognizers and a control process with feedback, these game-like interactions become robust human-in-the-loop recognizers. Our Active Capture scream shot and head turn shot employ this algorithm design paradigm.

Once the Active Capture system has captured, annotated and parsed out its shots they can be used as inputs to adaptive media templates that can automatically integrate these shots into high quality customized and personalized media. In the video, we show an excerpt from the Terminator 2 movie trailer adaptive media template which has automatically incorporated the captured "Head Turn" shot of Prof. Marc Davis.

4. ACKNOWLEDGEMENTS

I want to thank the members of the Garage Cinema Research Active Capture Applications Team who worked on this video: Rachel Strickland, Ana Ramirez, Rita Chu, My Huynh, Erick Herrarte, Jeff Heer, and Ted Hong. And special thanks to Brian Williams for his prior work on this technology at Amova.

5. REFERENCES

[1] M. Davis. Editing Out Video Editing. *IEEE MultiMedia*, 10(2), April-June 2003, 54-64.

[2] M. Davis. Active Capture: Integrating Human-Computer Interaction and Computer Vision/Audition to Automate Media Capture. In: *Proceedings of ICME 2003 in Baltimore, Maryland*, IEEE Computer Society Press, Vol. II, 185-188, 2003.

[3] M. Davis, J. Heer, and A. Ramirez. Active Capture: Automatic Direction for Automatic Movies (Demonstration Description). In *Proceedings of ACM Multimedia,* Forthcoming 2003.