

Active Capture: Automatic Direction for Automatic Movies

Marc Davis

University of California, Berkeley
School of Information Management
and Systems
Garage Cinema Research
marc@sims.berkeley.edu

Jeffrey Heer

University of California, Berkeley
Computer Science Division
Group for User Interface Research
jheer@acm.org

Ana Ramírez

University of California, Berkeley
Computer Science Division
Garage Cinema Research
anar@cs.berkeley.edu

ABSTRACT

The Active Capture demonstration 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. Active Capture leverages media production knowledge, computer vision and audition, and user interaction design to automate direction and cinematography and thus enables the automatic production of annotated, high quality, reusable media assets. 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.

General Terms

Algorithms, Design, Reliability, Human Factors

Keywords

Metadata, video capture, human-in-the-loop, recognition, automated direction, active capture.

1. INTRODUCTION

Media production today requires more time, money, skill, and equipment than most people have. As a result, the production of high quality media is limited to select professional settings or a few dedicated hobbyists while the consumption of media is frequent and widespread. One of the research goals of Garage Cinema Research is to *enable daily media consumers to become daily media producers* (<http://garage.sims.berkeley.edu>). In order for this goal to become a reality, the current media production process must be transformed from a manual mechanical process into an automated computational one [2]. Active Capture [1,2,3] focuses on the beginning of the media production cycle—the point of capture—and automates the directorial and cinematographic processes. Active Capture solves several significant problems affecting consumer media production: it enables the automated capture of high quality, reusable media

assets; as a byproduct of this process, it produces content descriptions (metadata) of the captured media assets; and the metadata created during Active Capture enable the captured media assets to be automatically integrated into customized and personalized media productions. By reconceptualizing metadata creation as a human-in-the-loop algorithmic process, we leverage what computers and humans are respectively good at in order to create a combined human-machine system more capable than either alone.

2. AUTOMATING DIRECTION

In traditional motion picture direction, as well as in professional (and even consumer) portrait photography, a director/photographer directs an actor/subject in order to capture a given shot she has in mind. She might use a variety of techniques such as describing or demonstrating the desired action (e.g., “Say Cheese” or “Smile like this”), or making the actor react spontaneously to a stimulus. The director/photographer reviews each capture (a “take”) and iterates until she “gets the shot.” This entire process is automated in an Active Capture system.

In motion picture direction, professional portrait photography, and consumer “Say Cheese” interactions, humans both prompt and evaluate the responses of the captured subjects. In Active Capture both prompting and response evaluation are achieved by the media capture device itself. The Active Capture system’s direction is made as realistic and effective as possible by using a variety of mediation techniques from the design space described in [3]. Mediation is necessary when the captured media does not meet the predefined capture parameters. The mediation techniques are the system’s way of gracefully clarifying to the user what the system is trying to get the user to do and how specifically the user can improve their performance. These mediation techniques are common in directing or even portrait photography settings.

3. DEMONSTRATION DESCRIPTION

The Active Capture system that is currently implemented creates recombinable, annotated media assets of a guest screaming (See Figure 1a) and turning her head to look at the camera (See Figure 1b). These two shots are highly reusable media components that can be used to produce a variety of short movies. We have implemented several computer programs, which we call “adaptive media templates” [1], that use these shots to create personalized versions of an MCI commercial, a 7Up commercial, a Godzilla movie scene, a Blair Witch movie trailer, and a Terminator 2 movie trailer (See Figure 2 for the T2 movie trailer example).

The Active Capture system uses audio and/or video cues to prompt the user to scream or turn her head to look at the camera. Through real-time audio and video analysis the Active Capture

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM MM '03, November 2-8, 2003, Berkeley, CA.

Copyright 2003 ACM 1-58113-000-0/00/0000...\$5.00.

system determines the fitness of the user's response in relation to predetermined capture parameters. If the capture satisfies these parameters, the capture process is complete. If not, the Active Capture system prompts the user again (using a progressive assistance technique) until a suitable response is achieved or the process has timed out (graceful failure) [3]. Currently, an Active Capture application is specified in terms of a finite state machine (FSM), which is then implemented in C++.

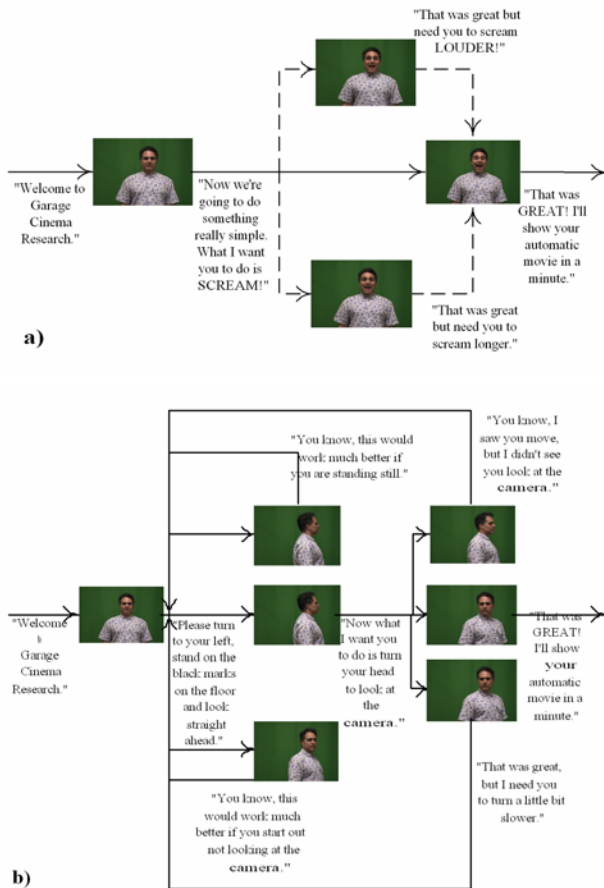


Figure 1. Active Capture process for capturing (a) “Scream” shot and (b) “Head Turn” shot. (Quotes are verbal instructions from the Active Capture system. The continuously horizontal left-to-right arrows represent an error-free path. The other arrows are error correction loops.)

In the “Scream” shot, the user is directed to scream while looking at the camera (See Figure 1a). The system has a predetermined loudness range and minimum length the guest needs to scream in order to capture an asset that can be effectively reused. If the user screams loud enough and long enough based on the predetermined thresholds, the system tells her she did a “great” job and lets her know she will be able to see her automatic movies in a few minutes. If the user does not scream loud enough or long enough, she is asked to scream again, only this time to either scream “LOUDER” or “longer.” The user and the system will go through this feedback cycle until the system is able to get the user to produce the kind of shot it needs or the system gracefully exits and makes the best movie possible with the shots captured.

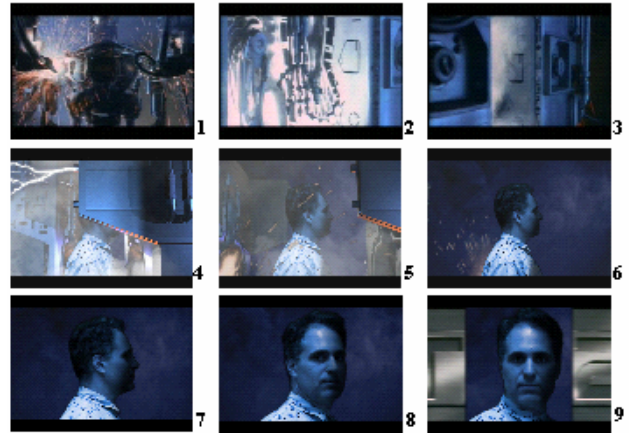


Figure 2. Storyboard of Terminator 2 automatic movie.

In the “Head Turn” shot (See Figure 1b), the possible failure modes and the requisite need for mediation are significantly greater than in the “Scream” shot. The Active Capture application must recognize and then mediate to ensure the proper execution of the following aspects of the user's actions:

- The head must be properly framed in the camera view
- The head must face away from the camera at the beginning of the shot
- The head must be still at the beginning of the shot
- The head must stop turning at the end of the shot
- The head must face the camera at the end of the shot
- The head must turn slowly enough that sufficient video frames can be captured to enable the shot to be sped up or slowed down during later reuse

The Active Capture system interactively directs the user to perform the head turn in such a way that a shot can be captured where all of these things are true. The corresponding interaction script is illustrated in Figure 2b.

4. FUTURE WORK

We are currently designing a language for prototyping Active Capture applications called “ACAL” (Active Capture Automation Language) as well as developing new Active Capture scenarios (such as a “Running in Place” shot) and new adaptive media templates that use these and existing scenarios to automatically produce customized and personalized media content.

5. REFERENCES

[1] M. Davis. Active Capture: Integrating Human-Computer Interaction and Computer Vision/Audition to Automate Media Capture. In *2003 IEEE Conference on Multimedia and Expo Special Session on Moving from Features to Semantics Using Computational Media Aesthetics*, Baltimore, MD, July 6, 2003.

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

[3] J. Heer, N. S. Good, and M. Davis. Mediating Automated Media Capture: Leveraging Interaction Design for Metadata Creation. Submitted to *ACM Multimedia 2003*, Berkeley, CA.