A Robotic Framework for Video Recording and Authoring

Jun Kato Masataka Goto
National Institute of Advanced Industrial Science and Technology (AIST), Japan
{jun.kato, m.goto}@aist.go.jp

ABSTRACT
Authoring videos that demonstrate interactive applications with human-computer and human-robot interactions are not easy. It typically involves many retakes, each of which requires repetitive actions in the authoring tools. To address the issue, we propose a robotic framework that covers not only the camera but also the recording environment. It allows the user to annotate the video when being recorded to eliminate tedious post-processing. This demonstration shows an example recording setup consisted of a robot or tablet device (interactive applications) and the proposed robotic camera system. The attendees can examine the framework through video authoring experience with the system.

Keywords
Robotic camera system, video authoring, toolkit.

1. INTRODUCTION
Recording, editing, and sharing videos that demonstrate human-computer and human-robot interactions (HCI, HRI) have become common not only for the research community but also for the industry to appeal the products and for end-users to share their experience with the interactive applications. However, authoring high-quality videos is still time-consuming and not easy. There have been proposed robotic camera systems that follow aesthetic guidelines for taking photos [1] and recognize the scene to control the camera for framing interesting subjects [2]. Meanwhile, prior work has mostly focused on the recording process, and the overall difficulty remains. It comes from potentially a lot of retakes and the resulting repetitive operations in the video authoring tools to apply appropriate editing effects [3].

To eliminate the tedious post-processing, DemoCut [4] provides a dedicated user interface for annotating the recorded video. The system then automatically edits the video with appropriate effects. Our work shares its goal but focuses on adding annotations *during* (not *after*) the recording process. To achieve this, we extend the robotic camera approach to augment not only the camera itself but also the recording environment. In this one-page abstract, we briefly explain the proposed framework and discuss its potential.

2. PROPOSED ROBOTIC FRAMEWORK
Our robotic framework helps to author videos of interactive applications by adding annotations *during* the recording process. The annotations can be categorized into a) session management, b) camera control, and c) overlaying information. These can be made in two ways after the user starts recording. One way is to show our predefined markers to the camera (e.g. "add a new session" or "discard the ongoing session," Figure 1a). These markers are recognized by the camera server with the ARToolkit.

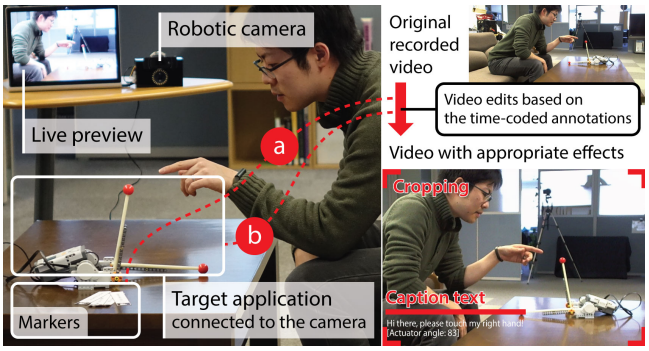

Figure 1. Overview of the proposed robotic camera system.

The other way is to embed a small program developed with our toolkit in the interactive applications (Figure 1b). The toolkit provides APIs that tell the camera to manage sessions, look at a specific direction, and show or hide captions. In the current system, the toolkit supports JavaScript and Java applications and connects to the camera server through the Socket.io library.

Once the user is satisfied and stops recording, the camera server edits the video following the annotations, producing video clips with captions for each session. Even when the user wants a retake, there is no need for tedious post-processing tasks.

3. DISCUSSION AND FUTURE WORK
We proposed a robotic framework for addressing repetitive post-processing operations in the authoring process of demonstration videos. The framework essentially adds time-coded metadata to the recorded videos that are used to edit the video. The metadata can be added instantly by the user, which eliminates the need to keep in mind which session was success/failure. The meta-data can be added programmatically, which eliminates the manual operations to input captions at appropriate timings.

While our current implementation only allows showing static text-based captions on the videos, the future work could allow its animation as shown in TextAlive [5] and allow a robot application to stream its first-person view to be presented as picture-in-picture video. The video and timeline interface help to understand the internal state of physical interactive applications. We foresee that our framework can also be used for debugging robot applications.

4. REFERENCES
[1] Gadde, R., et al. Aesthetic Guideline Driven Photography by Robots. In *Proc. of IJCAI '11*. 2060-2065.

[2] Pixio. 2011. Move 'N See – Your Personal Robot Cameraman. http://www.movensee.com.

[3] Muller, E. 2009. Where Quality Matters: Discourses on the Art of Making a YouTube Video. *The YouTube Reader*. 126-139.

[4] Chi, P., et al. DemoCut. In *Proc. of UIST '13*. 141-150.

[5] Kato, J., et al. TextAlive. In *Proc. of CHI '15*. 3403-3412.