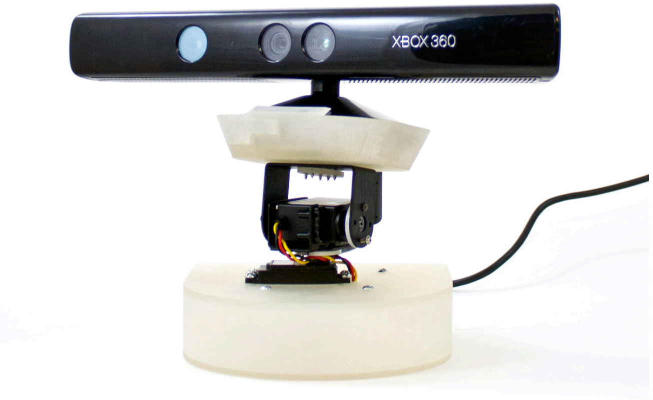

Body-Tracking Camera Control for Demonstration Videos

Derrick Cheng
University of California, Berkeley
354 Hearst Memorial Mining
Building
Berkeley, CA 94720 USA
derrick@cs.berkeley.edu

Pei-Yu (Peggy) Chi
University of California, Berkeley
354 Hearst Memorial Mining
Building
Berkeley, CA 94720 USA
peggychi@cs.berkeley.edu

Taeil Kwak
University of California, Berkeley
102 South Hall
Berkeley, CA 94720 USA
taeil2@ischool.berkeley.edu

Björn Hartmann
University of California, Berkeley
533 Soda Hall
Berkeley, CA 94720 USA
bjoern@cs.berkeley.edu

Paul Wright
University of California, Berkeley
5133 Etcheverry Hall
Berkeley, CA 94720 USA
pwright@me.berkeley.edu

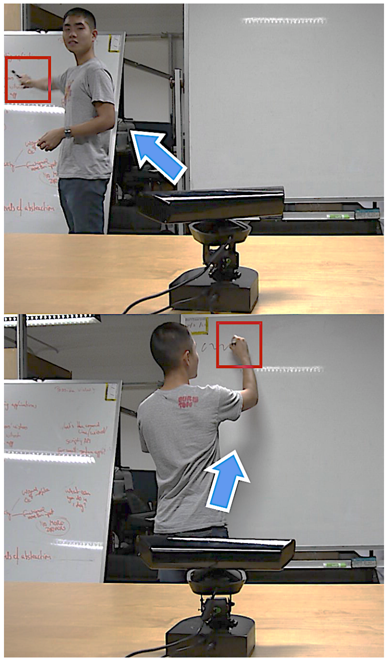

Figure 1. Kinectograph includes a Kinect camera to track user movement and a motorized dock to pan and tilt the camera so that the user (or their hand) remains centered in the recorded video. Here the device follows the user's hand while he is illustrating.

CHI 2013 Extended Abstracts, April 27–May 2, 2013, Paris, France.
ACM 978-1-4503-1952-2/13/04.

Abstract

A large community of users creates and shares how-to videos online. Many of these videos show demonstrations of physical tasks, such as fixing a machine, assembling furniture, or demonstrating dance steps. It is often difficult for the authors of these videos to control camera focus, view, and position while performing their tasks. To help authors produce videos, we introduce Kinectograph, a recording device that automatically pans and tilts to follow specific body parts, e.g., hands, of a user in a video. It utilizes a Kinect depth sensor to track skeletal data and adjusts the camera angle via a 2D pan-tilt gimbal mount. Users control and configure Kinectograph through a tablet application with real-time video preview. An informal user study suggests that users prefer to record and share videos with Kinectograph, as it enables authors to focus on performing their demonstration tasks.

Author Keywords

Capturing tools; camera; videos; motion tracking; Kinect; How-To; DIY

ACM Classification

H.5.m [Information Interfaces and Presentation (e.g., HCI)]: Miscellaneous.

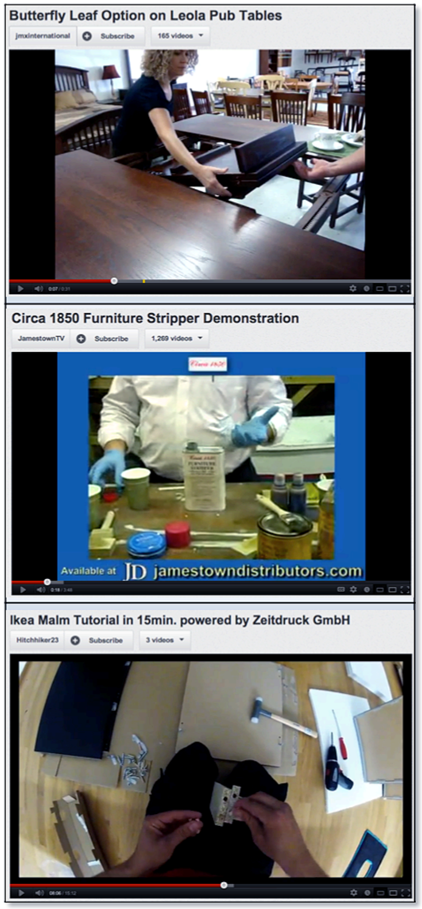
Butterfly Leaf Option on Leola Pub Tables
jmxinternational
Subscribe
165 videos
Circa 1850 Furniture Stripper Demonstration
JamestownTV
Subscribe
1,269 videos
Available at JD jamestowndistributors.com
Ikea Malm Tutorial in 15min. powered by Zeitdruck GmbH
Hitchhiker23
Subscribe
3 videos

Figure 2. Common video recording views of online How-To tutorials: (From top to bottom) Video taken by a cameraman standing aside; Self-Recording by focusing on the tabletop; Using a head mounted camera to capture the workspace.

Introduction

Popular online video-sharing websites such as YouTube have enabled the growth of a large community of users who share their knowledge and expertise in video tutorials. How-To or DIY (Do-It-Yourself) videos demonstrate specific skills and procedures for tasks as varied as cooking, building a treehouse, or fixing a machine [7]. By presenting a skilled demonstration visually, online tutorials help learners observe the manipulations and then put them into practice [8]. However, in recording these videos, the instructors often find it challenging to control the camera while performing their tasks. Based on our reviews of popular DIY tutorial sites, there are mainly three ways to record a How-To procedure (Figure 2):

- **Working with a cameraman** who controls the device and viewpoints. This method ensures that the video captures the movements that the audience would want to see, but it requires having a second person to direct the recording and work together with instructors.
- **Self-recording** with a static viewpoint, by setting a camera on a tripod. This is the most common and easy way to record a video; however, authors are unsure whether their actions are properly in frame at recording time. They may have to record multiple takes, or stop the recording during adjustments.
- **Wearing a head mounted camera** to capture what the instructors see. This may record unwanted and distractive head movements, making it difficult for the audience to watch. Additional camera views of the overall workspace and video stabilization might be needed to assist learners with understanding the context of demonstrated actions [1].

Seeing these filming challenges, we would like to enable users to gain the flexibility of real-time camera control, without the requirements of a dedicated camera-person. In this paper, we propose a new device that provides automatic and user-controlled camera orientation for end users at home. This device tracks user and moves continuously to track their activities (Figure 1). Users can configure the camera and preview video streaming through a tablet device.

Related Work

Existing commercial products, such as video conferencing cameras and surveillance tools, have considered human tracking in order to provide full or partial automatic viewpoint control. Polycom designs video conferencing cameras that feature face recognition and voice detection to enable a group of users to talk in an office room setting [4]. This approach assumes people's faces should be in the frame, which may not be true for demonstration videos that focus on actions rather than "talking heads." Swivl provides an automatic motion tracking iPhone dock that always keeps the user in view by tracking an infrared emitter that the user must wear [6]. In contrast, our system offers both options of real-time off-camera tablet control and automatic tracking without requiring the user to wear sensors.

There are also research projects aiming to provide automatic or interactive filming experiences. Okumura *et al.* designed an optical gaze control camera to automatically focus on fast-moving objects such as a Ping-Pong ball using rotational mirrors [3]. Their system only tracks predetermined targets and cannot be used for general demonstrations. TeleAdvisor assists a helper to remotely observe a physical task via video

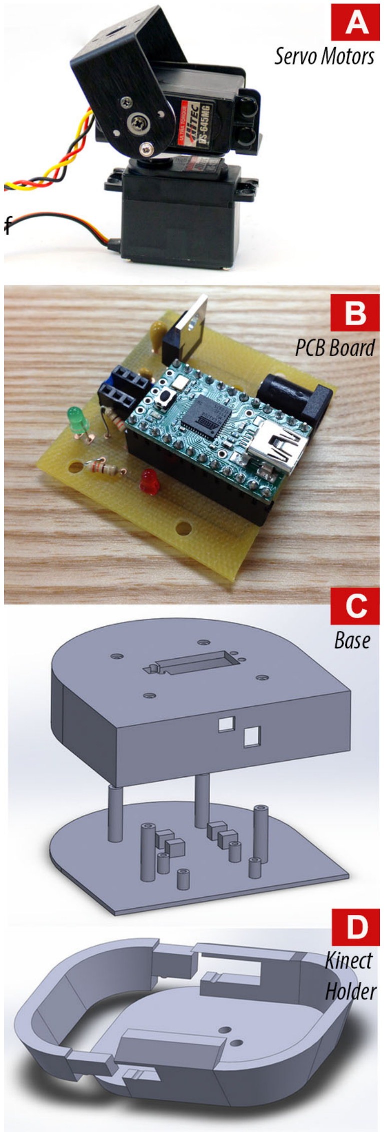

Figure 3. Kinectograph components

streaming in real-time and provide instructions [2]. However, the system is limited by the static camera view without automatic tracking or highlighting. Beamatron enhances a dynamic environment with graphical projection by tracking user movements using Kinect [9]. This work shares similar components with our system in that it uses both a Kinect and automatically controlled motors but is for a different purpose - augmented reality projection, rather than video recording for user demonstrations.

Introducing Kinectograph

Kinectograph serves as both the camera and the cameraman. It provides a motorized dock for a Kinect sensor and a tablet-based user interface (separate from the camera) to control the camera orientation (Figure 1). By using the Kinect to track the user's movement, Kinectograph automatically pans and tilts the camera in real time. The portable size of Kinectograph makes it easy to be placed on a tabletop surface or a TV stand to capture the room where the user will perform.

The user controls Kinectograph with a tablet that runs our web-based user interface (Figure 4). The Kinectograph UI shows a real-time video feed from the Kinect camera. On this video feed, Kinectograph overlays detected body features: the user's hands and head. By tapping on one of these features, a user can instruct Kinectograph to continuously track this body part. The user is free to change these settings in real-time during the recording process. If the user decides to record only an object in some parts of the demonstrations, she can also switch from automatic tracking into manual control mode. In this mode, swipe gestures on the video preview are translated in to pan and tilt rotation commands for the motorized dock.

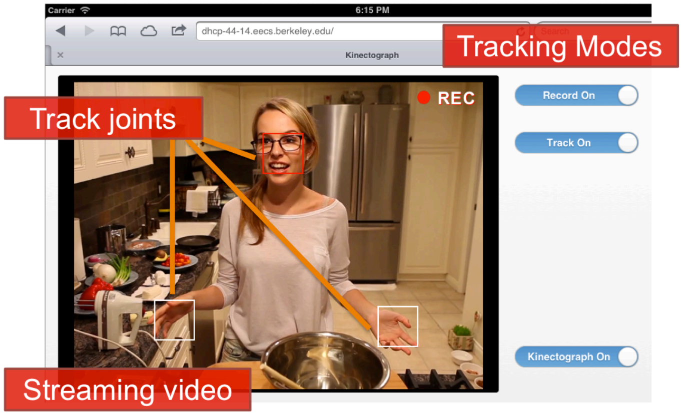

Figure 4. Kinectograph UI on a tablet device

System Design

Kinectograph streams the video view captured from a Kinect camera to a PC. This PC also acts as a web server, publishing the control interface to tablet or phone clients. When a user enables automatic tracking, the PC analyzes user movements using the skeletal data from the Kinect SDK. Based on the user position, it controls the two-dimensional servos by sending appropriate commands to the motorized dock via USB to move the camera dynamically.

Hardware Modules

To control the camera view, we designed a specialized dock that holds the Kinect sensor and can be rotated using a *pan-and-tilt servo kit* in two axes (Figure 3a). A bottom servo moves the Kinect left and right, while the top servo moves it up and down. The mount allows full 180-degree rotation for both horizontal and vertical axis. To control the servo motors of the pan-tilt mechanism, an embedded microcontroller (8-bit

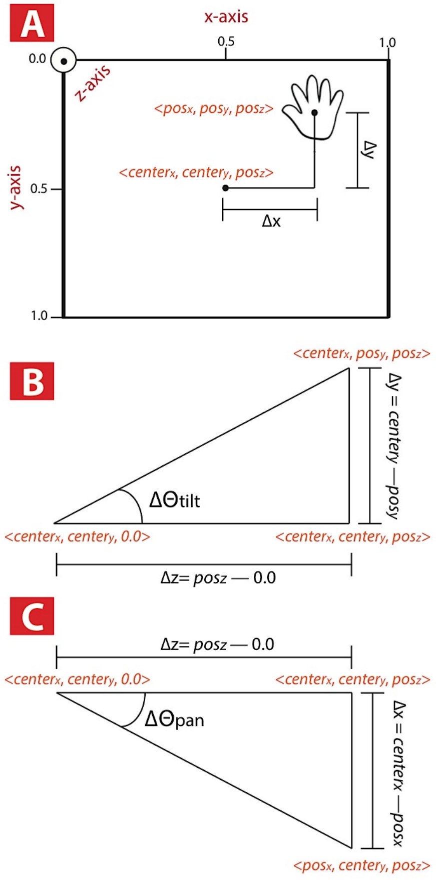

Figure 5. (A) describes the positions of the center and the position of the tracked hand where the X-Y plane is the field of view of the Kinectograph. (B) and (C) depict the values needed to calculate the tilt and the pan degrees.

ATmega32U4[1]) generates PWM (Pulse-Width Modulation) signals.

A base (Figure 3c), fabricated on a Projet HD 3000 printer[2], stabilizes the system and houses the electronics, such as a custom PCB board (Figure 3b). The bottom servo of the pan-tilt system fastens to the middle of the base, while the Kinect holder attaches the Kinect to the pan-tilt system (Figure 3d).

Motion Tracking and Servo Adjustment

To track the user position and determine the camera angle, our system analyzes the skeletal tracking data and depth information of user's body parts received from the Kinect sensor in real-time. Using Kinect enables the user to freely move or turn around, with support of self-occlusion [5]. Currently Kinectograph only tracks the person of the nearest distance to the camera and filters the background crowd.

The goal of Kinectograph is to move the camera angle to position the target, such as a user's hand, to the center. When a user is found in the view, at first we receive the position of the tracked joint located at *<*pos_x, pos_y, pos_z*>*. Second, we determine the rotation angles in order to align the joint to the position of the center of the field of vision located at *<*$center_x$, $center_y$, pos_z*>* on the same X-Y plane as the joint position. We compute the angles $\Delta\Theta_{tilt}$ and $\Delta\Theta_{pan}$ in degrees to tilt or pan the camera using the following formulas:

Tilt angle of y-axis: $\Delta\Theta_{tilt} = \frac{\Delta y}{\Delta z} = \arctan\left(\frac{center_y - pos_y}{pos_z}\right)$

Pan angle of x-axis: $\Delta\Theta_{pan} = \frac{\Delta x}{\Delta z} = \arctan\left(\frac{center_x - pos_x}{pos_z}\right)$

Figure 5 depicts the geometric relations from the Kinectograph view. Every 10 milliseconds the top motor will be turned by $\Delta\Theta_{tilt}$ and the bottom motor will be turned by $\Delta\Theta_{pan}$. In order to avoid extraneous small camera movements, we set a bounding box around the center of the field of view, such that the Kinectograph only moves once the tracked object leaves this bounding box.

Preliminary Evaluation

To understand how users would interact with Kinectograph with what tasks, we evaluated our system through two means:

Demo at an Expo

We demonstrated Kinectograph at a public exhibition to approximately 60 people. Each participant was allowed to enter our capturing space and experience the device. Based on our observation and conversations collected, we found that people were convinced by the idea as soon as they walked into the scene when Kinectograph started to move along. These questions were often asked: "*How fast was Kinectograph able to follow me?*" and "*Can I switch to track other parts (like my hand)?*". With the tablet device control, participants soon were successful in controlling the camera. They often walked, ran, and danced to test the tracking. We also learned that people expected the device to provide fast response in various conditions such as turning, rapid change of directions, or partial occlusions (when people were hidden by furniture or large objects).

[1] http://www.atmel.com/devices/atmega32u4.aspx

[2] http://printin3d.com/3d-printers

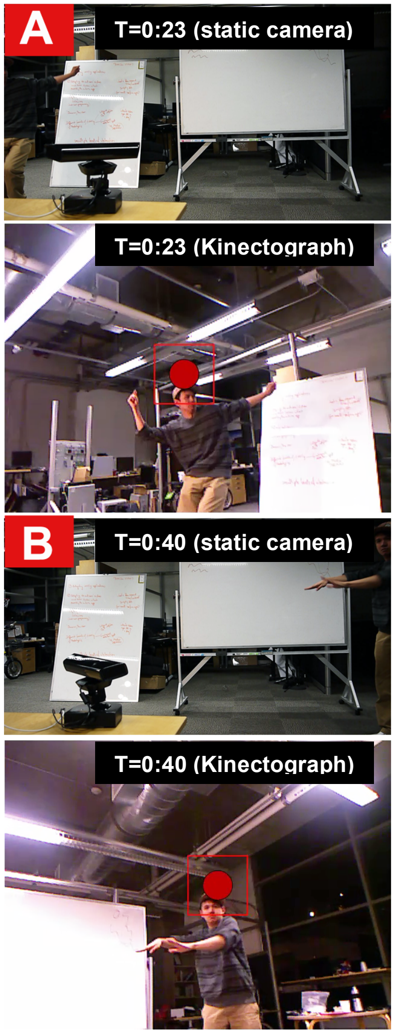

Figure 6. Examples of camera views captured by a static camera and Kinectograph at two specific moments in time

Informal User Study

To understand how Kinectograph can support users in demonstrations, we invited four participants (3 male and 1 female, aged 22-29) who did not join the exhibition to our user study in a home environment. We aimed to explore two hypotheses:

H1. Users prefer to watch the video captured by Kinectograph over video recorded with a static camera.

H2. Kinectograph can capture complete demonstrations that a static camera cannot achieve.

We first introduced Kinectograph by having participants walk around while the device tracked. We encouraged participants to brainstorm some activities they wanted to record. Once the task was decided, they were asked to set up both a static camera and the Kinectograph with our tablet device and start the recording. There was no time constraint during the study. A short post interview was then conducted, in which we showed the recorded videos from both cameras on a PC.

To answer **H1**, we designed a questionnaire with 5-point Likert-scale questions and gathered feedback. To answer **H2**, we analyzed the videos captured by both cameras. We recorded quantitative data of the number of times and the total length that each user moves out of the camera view.

Results and Discussion

Table 1 shows details of the four tasks and analysis of the recorded videos. We categorized physical activities into three movement types: *Continuous* (user continuously moves around), *Periodic* (user moves, stays, and moves again periodically), and *Occasional* (no clear motion pattern was observed). There were two Continuous and two Periodic tasks that participants designed. The moving range was about 15 feet in a home environment, and participants set the static camera about 8 feet away from the center of their workspace. Participants chose this distance to avoid out-of-frame problems with the static camera: "*The distance was chosen so that all of the activity could be captured*" (P4). Kinectograph was placed 6 feet away on a tabletop by the experimenters to capture the participant's whole body. Participants were allowed to adjust the camera angle via our tablet UI before recording the demonstration.

User	*Recording Task*	*Location*	*Movement Type*	*User Moving Range*	*Static Camera Distance*	*Video Length*	*Out of view from static camera*		*Out of view from Kinectograph*	
							Counts	*Length*	*Counts*	*Length*
P1	Hip-hop dance	Meeting room	Continuous	16ft	7ft	1'45"	3	15"	0	0"
P2	Workouts	Computer room	Periodic	15ft	8ft	1'30"	0	0"	0	0"
P3	PiggyBack ride tutorial	Living room	Continuous	15ft	8ft	2'00"	9	15"	0	0"
P4	Fight scene	Living room	Periodic	15ft	8ft	0'55"	2	5"	0	0"

Table 1. Task information and results collected in the preliminary user study.

All the participants chose to track their heads, but note that their activities involved frequent turning where pure face recognition might fail. Participants did not change this setting during the performance, although they were allowed to. P2 changed to the manual mode for testing, switched back, and then continued the activity. The average video length is one and half minutes long.

All the participants agreed or strongly agreed that Kinectograph captured what they intended to show, while only half of them agreed that the static camera captured as expected (**H1**). The main reason was the limited static camera angle; in three tasks, participants moved out of the static camera view more than once (**H2**). Figure 6 shows two examples where our system captured what the static camera missed. It was worth noting that although P3 had set and confirmed the viewpoint before recording, he was not aware that he shortly but frequently (9 times) went over the boundaries when he was demonstrating. He explained that he preferred using Kinectograph because it "*kept us in the center of view no matter how we moved around*." This shows that Kinectograph successfully ensured the activities would be captured and therefore enabled users to focus on their tasks.

Participants provided suggestions for further improvement. P2 suggested including different body parts (e.g. feet) or a combination of joints and objects, for scenarios such as playing a basketball. P1 and P3 would like to see how Kinectograph could be used in a multiple user scenario where different people can take control on the fly and add more flexibility. We plan on incorporating in the feedback we received from the preliminary user study. We will make motor movement smoother, add more features to address more scenarios, and add controls and editing effects such as zoom capabilities. A formal user study will be designed and conducted to record more complicated and longer demonstrations that would meet wider user needs.

Acknowledgement

We thank all the participants of the class CS294-84 "Interactive Device Design" at UC Berkeley for their insights and comments in Fall 2012.

References

[1] Fussell, S.R., Setlock, L.D., and Kraut, R.E. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. CHI'03 (2003).

[2] Gurevich, P., Lanir, J., Cohen, B., and Stone, R. TeleAdvisor: a versatile augmented reality tool for remote assistance. In CHI'12, ACM Press (2012).

[3] Okumura, K., Oku, H., and Ishikawa, M. High-speed gaze controller for millisecond-order pan/tilt camera. ICRA 2011: IEEE International Conference on Robotics and Automation (2011).

[4] Polycom. UC solutions for telepresence, video conferencing and voice. Available at: http://www.polycom.com/ (2012)

[5] Shotton, J., Fitzgibbon, A., Cook, M., et al. Real-time human pose recognition in parts from single depth images. In CVPR (2011).

[6] Swivl. Personal cameraman for hands free video. Available at: http://www.swivl.com/ (2012)

[7] Torrey, C., McDonald, D.W., Schilit, B.N., and Bly, S. How-To pages: Informal systems of expertise sharing. In ECSCW 2007 (2007), 391–410.

[8] Torrey, C., Churchill, E.F., and McDonald, D.W. Learning How: The Search for Craft Knowledge on the Internet. In CHI'09, AMC Press (2009), 1371–1380. 2.

[9] Wilson, A., Benko, H., Izadi, S., and Hilliges, O. Steerable augmented reality with the beamatron. In UIST'12, ACM Press (2012), 413–422.