Gesture Interaction with Rich TV Content in the Social Setting

Goranka Zoric

Mobile Life @ Interactive Institute
Box 1197, SE-164 26 Kista,
Sweden
goga@mobilelifecentre.org

Arvid Engström

Mobile Life @ Interactive Institute
Box 1197, SE-164 26 Kista,
Sweden
arvide@tii.se

Louise Barkhuus

Mobile Life @ Interactive Institute
Box 1197, SE-164 26 Kista,
Sweden
barkhuus@mobilelifecentre.org

Javier Ruiz Hidalgo

Signal Theory and
Communications Department,
Technical University of Catalonia
Jordi Girona 1-3, edifici D5
Campus Nord UPC
08034 Barcelona, Spain
j.ruiz@upc.edu

Axel Kochale

Deutsche Thomson OHG
Karl-Wiechert-Allee 74,
30625 Hannover, Germany
axel.kochale@technicolor.com

Abstract

The appearance of new immersive TV content has increased the interactive possibilities presented to the viewers. Increased interactivity is seen as a valuable feature in viewing richer television content, but new functionalities are limited by what can be done naturally and intuitively using available devices like remote controls. Therefore, new interaction techniques, such as visual gestures control systems, have appeared aiming to enhance the viewers' viewing experience. In this work we begin uncovering the potential and challenges of gesture interaction with ultra high definition video for people watching TV together. As a first step we have done a study with a group of people interacting with such content using a gesture-based system in the home environment.

Author Keywords

Gesture interaction, panorama, video, interactive TV, user studies, live, broadcasting.

ACM Classification Keywords

H.4 [Information Systems Applications]: Miscellaneous; H.5 [Information interfaces and presentation (HCI)]: User interfaces

General Terms

Design, human factors

Introduction

Emerging TV systems offer unprecedented possibilities for interacting with rich audiovisual media content. The combination of new ultra high definition capture technology and novel interaction techniques seem to suggest a more active and lean-forward form of viewing than the lean-back traditional TV consumption. Advanced TV content that allows for more extensive interaction often appears in the form of multiple feeds available within the same program, free viewpoint selection, user controlled shot selection or automatic scripts to follow interesting regions of the screen. High-speed networks as well as powerful viewing devices offer more content and more interactivity for the end user. As a result we have rich and possibly overwhelming viewing options available.

In this work we are interested in how interactivity can enhance the experience of people watching TV together. We begin to uncover the potential and challenges of gesture interaction with ultra high definition video in real life settings in the home. As a first step we have done a user study with a group of people interacting with such content using a gesture-based system. Here we specifically concentrate on new interactive high definition panoramic TV – a wide video image representing a view of the scene from a given viewpoint which has a resolution high enough to enable interactive viewing by panning, tilting or zooming, and is obtained by stitching streams from several standard high definition cameras.

Such content is the focus of the FascinatE (Format-Agnostic SCript-based INterAcTive Experience) project [6]. The FascinatE system enables broadcasting of live events, and allows end-users to interactively view and navigate around an ultra-high resolution video panorama showing a live event no matter what device they are using to view the broadcast. Next, we will mention other systems offering panoramic viewing, as well as motivate gesture interaction in this context.

Background

Currently, there are a number of panoramic video based systems available. They all differ in picture quality, and in the interactivity they offer. For example, the Immersive Media's Telemmersion System [4] enables creating live interactive 360° videos which includes panning and tilting using standard computer input, but with the resolution of "only" 2400x1200 pixels, which makes it problematic to use for deep zooming. The Camargus [5], a production oriented system, offers a panoramic video of the sport event created by stitching 16 high-definition video feeds into one feed. The operator is given the possibility to control a virtual high definition camera, i.e. to do virtual panning and zooming.

How we interact with such rich content plays an important role in how we will experience it. While traditional remote control based interaction is readily available, it might not be the best option for navigating within large volumes of content. Advanced interactive commands are not easy to perform on a remote control *"(…) since the number of keys is limited to the arrows, channel numbers, the OK button, and the color buttons*." [1]. Instead more natural interaction methods, like speech and gesture modalities, are becoming popular. Research efforts within iTV in the last decade have resulted in numerous innovative approaches for interacting with TV content ranging

Figure 1: Details from navigation in panoramic video of a dance show

from advanced remote controls and handheld mobile devices to specifically designed digital artifacts.

Recent work on gesture use and advanced depth sensors has enabled appealing device-less gesture based interaction with this type of content. New techniques for gesture recognition have been employed; for instance, the Kinect (based on depth cameras) by Microsoft has become widely chosen alternative in the field of human motion capture. For example, Panasonic presented recently their gesture control [2] that allows controlling the TV through some basic gestures – users can change channels and access specific contents waving their hand. The main benefits provided by these gesture-based systems are that they bring natural and intuitive mechanisms to control and navigate through rich TV content, but also personalization of the gestures. For the user, the control is always available, as they only need their hands, not any extra device, to control and interact with the system.

Even though a device-less and marker-less gesture based techniques present many advantages to traditional means of interaction, such gesture techniques come with challenges. One of the most prominent ones is control of interactive television while viewing socially. Juhlin and Önnevall [3] emphasize the role body movements have in the social TV environment: "*(…) movements can have many different purposes and are directed simultaneously at the screen and co-viewers.*" Thus, gesture control takes place in the social space to a larger extent than traditional remote control interaction does. The following study begins to explore what that implies for the social viewing experience.

User study

The gesture system used is based on a single Kinect camera and allows the user to interact with panoramic video displayed on a high definition TV screen or projected on the wall without using only their hands. User control includes among other navigating through a high-resolution panoramic views of the scene by panning, tilting and zooming, and control of the audio. The system is multi-user, i.e. a group of users can interact with the system, but only one person at a time can have control. Another user can ask for control of the system, and once the person has it, she/he can start interacting with the system.

The study was done in the autumn of 2012 with two types of panoramic content – video of a football match and of a dance show (Figure 1). Seven pairs took part of the study. They were between 20 and 39 years old, and where almost all had experiences with panoramic videos before, only 5 of them had some experience using gesture controls in this context. We video recorded their interactions. First, the participants interacted with the system for about 10 minutes for each of two available contents, and secondly, a researcher asked them a set of questions about their immediate impressions.

Gestures in social TV viewing

In our study it was clear that interacting with the system and getting used to the gesture control was taking a lot of the participants' attention. However, pairs were managing to communicate in between themselves. It was apparent that the gestures themselves were a conversation topic and something that added to an enjoyable social viewing experience. Users actively commented on the gestures both while

doing them and while observing the other. This communication seemed to serve the dual purpose of social interaction and learning the gesture techniques by putting words to them, both for oneself and to others. This may be partly due to the novelty of the system in the testing situation, but it suggests that social interaction and gesture control of the system do not necessarily need to be treated as to separate and possibly conflicting activities. Rather, the two modes both contributed to the viewing experience.

Although the participants were trying out the system's gesture interaction and navigation in the video image for the first time, they were not active at all times throughout the test sessions. Users shifted between active and passive modes, and between interacting with the system and each other. Video recordings show that the participants were looking mostly on the screen, just sometimes to each other, though there was no need to look for a control device.

In this system, a specific gesture is assigned for taking the control. Although the system gestures were individual, participants enacted them socially, e.g. mimicking the movement of taking the remote control, leaning towards the person with the control (Figure 3). The range of movement in gestures, like panning in the panorama varied between participants. Bigger gestures were sometimes experienced as tiring, but participants also stated that they were taking up the space of the person sitting next to them (Figure 4). As the study in [3] shows "*ordinary gestures (…) already occur in conjunction with TV viewing",* and that includes both gesturing towards the screen and nonverbal communication with co-viewers.

These observations suggest that the design of multi-user gesture systems should allow for unintrusive socially adapted gestures for controlling and navigating in rich video content, i.e. social needs should to be combined with gesture control in a natural way. In our future work we will further explore how we can make the use of natural gesture interaction in the context of emerging high definition video.

Acknowledgements

References

[1]   Cesar, P., and Chorianopoulos, K. The evolution of TV systems, content and users toward interactivity. NOW Foundation and Trends in HCI, vol.2: n°4. 373-95, 2009.

[2]   Panasonic, D-Imager, Panhttp://pewa.panasonic.com/news/press-releases/omek-interactive-panasonic-3d-gesture-recognition/ (Accessed 08.01.2013.).

[3]   Oskar Juhlin and Elin Önnevall, On the relation of ordinary gestures to TV screens: General lessons for the design of collaborative interactive techniques; Accepted for publication, 31st international conference on Human factors in computing systems, CHI'13, 2013.

[4]   Immersive Media, http://immersivemedia.com/demos/ (Accessed 08.01.2013.).

[5]   Camargus system, www.camargus.com/ (Accessed 08.01.2013.).

[6]   FascinatE project, http://www.fascinate-project.eu/ (Accessed 08.01.2013.)

Figure 3: Taking control gesture

Figure 4: Taking physical space of a co-viewer