Behaviour & Information Technology, 2014
Vol. 33, No. 8, 859–869, http://dx.doi.org/10.1080/0144929X.2014.914975

Taylor & Francis
Taylor & Francis Group

New interaction modes for rich panoramic live video experiences

Louise Barkhuus[a]*, Goranka Zoric[b], Arvid Engström[a], Javier Ruiz-Hidalgo[c] and Nico Verzijp[d]

[a]*Department of Systems and Computer Sciences, Stockholm University, Forum 100, 164 40 Kista, Sweden;* [b]*Interactive Institute, Isafjordsgatan 22 164 26 Kista, Sweden;* [c]*Signal Theory and Communications Department, Universitat Politecnica de Catalunya, Jordi Girona, 1-3, 08034 Barcelona, Spain;* [d]*Alcatel-Lucent Bell Laboratories, Antwerp, Belgium*

(*Received 23 December 2012; accepted 6 April 2014*)

The possibilities of panoramic video are based on the capabilities of high-resolution digital video streams and higher bandwidth's opportunities to broadcast, stream and transfer large content across platforms. With these opportunities also come challenges such as how to focus on sub-parts of the video stream and interact with the content shown on a large screen. In this paper, we present studies of two different interaction modes with a large-scale panoramic video for live experiences; we focus on interactional challenges and explore if it is (1) possible to develop new interactional methods/ways of approaching this type of high-resolution content and (2) feasible for users to interact with the content in these new ways. We developed prototypes for two different interaction modes: an individual system on a mobile device, either a tablet or a mobile phone, for interacting with the content on the same and a non-touch gesture-based system for the home or small group interaction. We present pilot studies where we explore the possibilities and challenges with these two interaction modes for panoramic content.

Keywords: panoramic video; interaction modes; interactive television; pilot studies

1. Introduction

The notion of interactive television has been part of both research and new entertainment visions for decades. Encompassing almost any element of interactivity that can make the traditionally passive television watching experience more active, the notion is broad and includes both features such as direct manipulation of television content and indirect interactivity through for example text-message voting (e.g. in reality shows and song contests). Also elements such as computerised platforms where television can be selectively chosen can be considered as part of the new interactive television sphere. In parallel with these new technical developments and increasing use of interactive elements in main-stream television, high-resolution video is becoming feasible to stream and broadcast due to higher bandwidth availability and the transfer to digital broadcast in most western countries. In many cases, the high resolution available is of better quality than the receiving device resulting in loss of data and, conceptually, loss of opportunities. It is, for example, possible to record panoramic images, yet no reasonably affordable private consumer screens are able to show these without a significant loss of screen estate. Large custom screens pose great potential for such broadcasting/showing, especially for large group viewings, such as in arenas, and for special occasions, such as sports events and concerts. However, when taking large panoramic high-resolution images into the home, the experience will diminish and much screen estate be wasted. One challenge in this context is, therefore, to utilise interactive elements for controlling the picture, for example by letting the viewer zoom in and out. By letting the viewer interact directly with the large-scale content, the viewer becomes immersed in the content, justifying the use of high-resolution panoramic video.

In this paper, we explore different possibilities for in-home interaction with large-scale panoramic video, especially focusing on live experiences such as sports events and concerts;[1] such experiences are more likely to be of interest to interact with and watch in the panoramic format. In order to select subsections of a video picture, zoom and pan, we can imagine several methods of interaction such as using a remote control, using hand gestures or even a second screen for controlling the video. Where the traditional remote control or keyboard interaction might seem obvious (and readily available), it might not be the best way of interacting with this new presentation of content. Other methods of interaction can potentially offer more natural interaction in the situation and warrant further exploration. We here propose two different methods of interaction and explore their use in realistic situations with real panoramic video and audio. One interaction method makes use of large arm and hand gestures and the other utilises a touch pad as a second screen to interact with television content. We conducted two studies (one included a pilot study) and found

*Corresponding author. Email: barkhuus@mobilelifecentre.org

that although both interaction methods were useful for navigating the panoramic video, each mode afforded different types of activities, leading to an emphasis on a variety of choices when designing for interaction with rich-content television experiences.

Before describing our systems and potential scenarios, we describe related research within the area of interactive television.

2. Related literature

Interactive television has been a research focus for decades. Although many specific definitions exist, these generally include any form of interaction with television content extending beyond channel changing or volume control. Interactivity can be provided to users as digital video recorder, video-on-demand, or as extra content or manipulating content. Recent advances in camera development, image processing as well as ubiquity of interactivity using for example mobile devices, have enabled changes in both production and viewing practices. In our work, we concentrate on a new type of TV content that makes it possible for viewers to watch a live event using multiple views. It means that the broadcast view is not selected by a camera operator based on assumptions about the actions taking place, and the screen size as in traditional TV production, but instead the event is being captured with more details than the viewers can use, enabling them to choose what they would view. Yet, it also means that current interaction techniques need to be adjusted for viewing such content.

The three main forms in which such rich content can be shown are free-viewpoint videos, multi-view videos, or high-resolution panoramas. In a free-viewpoint video, viewers can interactively change their viewpoint in the scene; i.e. viewers are able to freely navigate within real-world visual scenes, as known from virtual worlds in 3D computer graphics (Smolic et al. 2006). For example, work by Hilton et al. (2011) proposes the approach in which conventional monocular broadcast cameras are used for 3D reconstruction of the event and subsequent stereo rendering, applicable for live sports TV production. In a multi-view video, on the other hand, several different views on the event are offered to the viewers with more or less ability to influence what is being broadcast. An example is the LIVE Project the aim of which is to give viewers their personal and interactive broadcast by producing a parallel multi-stream coverage of live events. Similarly, work within the My-eDirector Project provides an interactive broadcasting service enabling end-users to direct their own coverage of large athletic events by taking the role of a virtual director and adapting the broadcast to their own viewing preferences (Patrikakis et al. 2011a). In our project, we explore high-resolution panoramic video and interaction with live event content for this format.

2.1. High-resolution panoramic video

An example of the extended view of the event on which we concentrate is a high-resolution panoramic video. However, in this work we focus on a more standard display like television or computer and therefore disregard approaches where a head-mounted display is used for displaying the panorama view. High-resolution panoramic videos are usually obtained by stitching video streams from several high-quality cameras (HegoOB1system; Fehn et al. 2006). Often, ranges of standard broadcast cameras are used to complement the panoramic picture, i.e. to enable a more comprehensive view of key regions of interest (Schreer et al. 2011). In a panoramic video, a given viewpoint remains the same during the entire event. However, users can be given the ability to influence what they are viewing by panning, tilting, or zooming and in that way be able to choose the viewing direction and view an arbitrary region-of-interests (RoIs) interactively. Our system offers this functionality but this in itself is not new.

Some systems, although providing a panoramic view of the event, are intended for a specific use, like on large screens in a cinema, as in the work by Fehn et al. (2006) and do not offer any interaction for the users. The imLIVE Demo offers live streaming 360 ° video with interactivity for audiences. The panoramic camera has 'only' 2400 × 1200 pixels, which makes it problematic to use for deep zooming. The Camargus System, a production-oriented system, offers panoramic video of a sports event created by combining an array of high-definition video feeds into one feed. The operator is given the possibility of controlling a virtual camera, and the recorded content is mainly used for replays.

2.2. Interacting with extra content

As the above brief overview showed, there are many systems which all, in one way or another, aim to give TV viewers the ability to choose their own view of a live (sport) event. While there is much literature describing technical approaches, challenges, and possible scenarios, there are significantly fewer user studies that aim to show how viewers would interact with the content offering various views on the event.

The work by Olsen, Partridge, and Lynn (2010) describes a user study of an interactive TV sports event over an Internet prototype that allows viewers both to choose the view and to control replays; experiments showed that sports viewers could easily learn the interactive controls and that they would use the interaction such as changing between cameras and 'moving in time', rather than passively watch the broadcast. On the other hand, Patrikakis et al. (2011b) illustrated how users might become annoyed when getting continuous annotations and recommendations of channel switching. Neng and Chambel

(2010) explored 360° hypervideos focusing on navigation and visualisation mechanisms in a panoramic view video, but the first actual user study of the panoramic video was presented by Bluemers et al. (2012). The authors present a study with the focus on omnidirectional video (ODV), a video that enables users to look around in 360°. They were focusing on finding what characteristics make a TV-programme suited for enhancement with ODV according to adult digital TV viewers. Interaction with panoramic video consisted of changing of the viewing angle and zooming in and out. Their findings show that ODV has the potential, yet challenges remain on technological, content, and user levels.

2.3. Interaction techniques

With the extensive development of interactive television content, it becomes necessary to examine how existing interaction techniques could cope with these novel possibilities and the wide range of interactivity that interactive television promotes. Until recent years, interaction with television content has been done mainly through specific devices, such as remote controls. Recently, however, wider acceptance of tablets in the consumer market has changed the basis of the user experience when interacting with multimedia content through the possibility of combining simple movements with finger configurations (i.e. pinch to zoom). The spread of tablet computers and their established interaction paradigm of pinching to zoom and moving the picture to pan has also made it easier to deploy intuitively functioning interaction with new media. Yet most of the solutions available so far have been based on advanced remote controls (including coloured or functional buttons) and hierarchical menus (cursor or numerical navigation), which are often very complex and slow. Several interaction techniques illustrating new research directions are relevant to our work.

An interesting novel interaction method is the speech remote control used by Nakatoh et al. (2007) to control a digital TV, specifically to change between channels or to perform category search. However, for simple actions like volume changing, the button input is still used. In the work by Vatavu and Pentiuc, an interactive coffee table is used to control the TV set using shared wide-area interface via simple hand movements across the video-sensitive surface of the table, which may be performed by any of the viewers at any time. In doing so, the need for negotiation is avoided, and the interface is immediately available for all the participants (Vatavu and Pentiuc 2008). A user interface for personalised live sports viewing on mobile devices by Wang et al. (2009) consists of viewing and navigational parts. Accelerometer sensors, which are incorporated in current mobile devices, are used to switch between viewing screen and navigational menu – it is only necessary to shake such a mobile device.

2.4. Gesture interaction with TV content

Over the last years, deviceless and touchless interaction, inspired by the touch interaction of tablet devices, has seen tremendous growth, particularly due to two factors. First, the emergence of new sensors that facilitate the recognition of human pose, as well as hand and finger gestures, has led to more opportunities for gesture-based interaction; second, the latest advances in image processing algorithms have opened up the possibility of implementing real-time systems that recognise these gestures with high fidelity. In regards to human pose and gesture recognition, the release of, for example, Kinect by Microsoft has raised depth cameras from marginal research sensors to be an actual alternative to strategies in the field of human motion capture. These new depth cameras, which work in a range between 0.5 and 6 m, provide a very fast and handy way of obtaining 3D information from human body parts. In this connection, Vatavu (2013) developed freehand gestures through participatory design studies to explore and compare which gestures are useful for interacting with television content in a home setting.

Finally, recent works and studies on human, hand, and finger pose using the aforementioned depth cameras have emerged allowing the possibility of navigating through the new multimedia formats. In this direction, Soutschek et al. (2008) propose a user interface for the navigation through 3D data sets using a depth camera. With a similar objective, Van den Berg et al. (2009) combine colour and depth information to recognise user gestures. In another related example, Jota, Pereira, and Jorge (2009) looked at three different metaphors for using gestures to interact with large screen content.

These factors have opened up possibilities of recent consumer examples of real deviceless interaction between users and televisions. For instance, Samsung offers Samsung Smart Interaction that allows users to control the graphical user interface of the TV with their hands. Unfortunately, the hand control is limited to moving a mouse pointer through the screen. Recently, Panasonic presented its 'Gesture Control' that allows the TV to be controlled through some basic gestures (Panasonic 2010). Users can change channels and access specific contents waving their hand. However, the system only works if the hand is very close (10 cm) to the screen. Although Vatavu (2012) compares handheld with freehand gestures, he focuses on the development of appropriate gestures, not the actual interaction activities when watching the content. The existing studies also mostly look at interaction with 'classic'-type television and fairly obvious interaction such as channel changing. In this project, we look at direct manipulation with high resolution, panoramic content, for experiencing live event shows. We investigate two types of interaction with high-resolution panoramic video content: large gestures for potential group interaction with a large wall screen and small touch screen gestures on a secondary screen (a tablet).

3. User scenarios

The scenarios we envision the system being used in are based on home environment scenarios of one or more users. The gesture-based system can facilitate seamless interaction with the live experience on a large living room screen where the viewer can choose between different regions of interest and turn up and down the volume by the use of hand gestures instead of reaching the remote. This increases the viewer's sense of immersion. The viewer is also able to choose between the regular broadcast view and a more specialised view, for example following a favourite football player. In the tablet user scenario, the control of the screen only takes place on the actual table, manipulating content on the tablet. The larger screen shows continuously the full panoramic view of the live event. This enables a personal view for one person at the same time as the rest of the family views the full panoramic view of the event.

These scenarios illustrate most of the functionality that we imagine our system will be able to support. Although much of the basis for this functionality has been developed and is being tested as part of the broader scale project, in this paper we limit our scope to exploring the two modes of interaction, hand gestures to navigate a large screen and a small-scale touch screen interaction. Both modes share many features in terms of, for example, zooming and panning but they also have their own allowances, enabling them to work well in diverse situations. The user does not need to be close to or in front of the TV. The provided interaction includes navigating through the panorama (panning, tilting, and zooming) and changing channels using only their hands with no extra device. Users are also able to interact with the associated sound channel by raising, lowering, or muting the volume. We continue by describing each system in detail, what features have been implemented and how to interact with them in detail.

4. Hand gestures interaction system

The hand gestures system that we designed allows users to perform simple interactions, such as changing channels on their TV, and more innovative interactions, such as selecting menus presented on the screen, navigating through high-resolution panoramic views of the scene, control the audio by changing the volume, muting, or selecting the speaker.

The current implemented gestures include:

Swipe (moving hand right to left) is used to select channels (Figure 1(a)).

Pointing and dragging an item to the centre of the screen allows users to select the menu item on the screen (Figure 1(b)).

Navigation inside the panoramic scene is provided using one closed hand for panning (not shown) and two pointing fingers to zoom in or out of the scene (Figure 1(c)).

A *Tee* gesture is used for taking and releasing control of the system (Figure 1(d)).

A *ParallelHands* gesture is used to pause or resume the reproduction or streaming of video content (Figure 1(e)).

The volume is controlled with three gestures, *Cross* to mute/activate the audio (Figure 1(f)), *FingerOnLips* to lower the volume (Figure 1(g)), and *HandOnEar* to raise the volume (Figure 1(h)).

The selected gestures were designed as a compromise between providing a natural and intuitive user experience and, at the same time, a feasible solution from a technical perspective. Gestures were selected to ensure that the gesture recognition system always provided a responsive, convenient, and intuitive experience for the user, even in crowded and/or low-lighted environments. They were selected following two different design decisions:

- Navigation in the panoramic video is performed with gestures (*Swipe*, *Point*, and *Navigation* gestures shown in Figure 1(a)–(c)) designed to simulate a virtual tablet in front of the user. In this manner, common gestures, such as *Drag to Move* or *Pinch to Zoom*, were translated to a virtual paradigm where no tablet is needed.
- Other functionalities (take control, increase/decrease volume, etc.) are activated with static gestures (*Tee*, *ParallelHands*, *Cross*, *Sush,* and *HandOnEar* shown in Figure 1(d)–(h)). These gestures provide a fast way to simple tasks acting as shortcut access to common menus thus increasing the usability of the system.

The design of static gestures was performed taking into account both the usability and intuitiveness of the system and the technical limitations of the system. An initial brainstorming with end-users and implementers of the system was conducted to propose gestures associated with the functionalities of the system. Functionalities such as decrease or increase the volume were mapped to fairly intuitive gestures such as the *Hush* or *HandOnEar* gestures. Other functionalities, even though some intuitive gestures could be mapped to them, were limited by the technical robustness of the system. This was the case for instance of the *Cross* gesture and finally was selected as the system could recognise it robustly over time. Other cases, such as the taking control functionality, did not have a clear intuitive gesture behind. Initially, a gesture where the user with the control of the system 'touched' the head of the user to pass the control was implemented. However, initial prototypes and user studies showed that user did not like the gesture and found it intrusive and difficult to perform. Therefore, the *Tee* gesture

Figure 1. The different hand gestures for the gestures interaction: (a) swipe, (b) pointing and dragging, (c) zooming in, (d) the Tee gesture, (e) the pause gesture, (f) mute/activate audio, (g) lower volume and (h) increase volume.

(a gesture easy to recognise by the system as it is used often) was proposed to take control of the system.

The current implementation of the gesture system is written in C++ and can run on a single laptop. The system has been split into multiple connected components where each is responsible for a single task (head location, hand tracker, gesture classification, etc.). Each component is based on SmartFlow, software developed by the National Institute of Standards and Technology (NIST) (http://www.nist.gov/smartspace/) to facilitate the communication of software modules.

The set-up of the system allows the user to be standing or seated in a chair or sofa. The gesture system is multi-user in the sense that several users can ask for the control of the system and interact with it while the others might still be present in the scene. The system uses the colour and depth video provided by a single Kinect camera as the principal sensor. No other data provided by the Kinect are employed. The Kinect camera should be placed at heights between 120 and 250 cm and approximate at the middle of the TV screen. A free space in front of the sensor of 2–4 m is also recommended.

5. Tablet interaction

The tablet interface is simpler than the gesture-based one since it is developed as a second screen. At the time of testing of the tablet interaction we had not implemented the RoI mode and the sound was not interactive. The tablet interaction follows conventional tablet finger movements: pinch and spread for zooming and selecting views with a simple touch on a virtual button. The system runs on tablets with a Wi-Fi connection for direct streaming of content from the server.

6. Studies

Both types of interaction, gesture interaction and tablet interaction, were studied in laboratory-based tests with 20 (half of them in pairs) and 16 participants, respectively. None of the participants took part in both studies. Due to the early stage of the overall project with panoramic live experiences we had only one type of content to use for the test of the tablet: a football match between Chelsea and Tottenham played in 2010. For the gestures test, we also had material from a dance performance. We had a loop of 10 minutes of the content that was running in both tests.

6.1. Hand gesture interaction pilot study

6.1.1. Method

The hand gesture laboratory study was based on an earlier pilot test performed with 10 people, between the ages of 20 and 32. The participants were instructed to interact with the system, using the described gestures to zoom, pan, and turn the volume up, down, and mute. We video recorded their interactions from two directions, front and back, where it was possible to see the screen and the resulting interactions. The results led to the changes of a few of the gestures as well as fine-tuning of some of the gesture recognition parameters; for example, the panning function was reversed from panning *with* the hand to *opposite* the hand movement, reflecting smart phone map navigation, which was more intuitive for these users.

The study was conducted with 20 people, between the ages of 20 and 39. Most of them had viewed panoramic videos before, either on big cinema screens or on tablets or mobile devices. Three of them had some experience using gesture controls in this context while seven were new to this type of interaction. Where 14 of the participants interacted as pairs, six of them interacted alone.

The participants were instructed to interact with the system, using the described gestures to zoom, pan, tilt, and turn the volume up, down, and mute. We video recorded their interactions from two directions, front and back, where it was possible to see both the screen and the gestures simultaneously. The participants interacted with the system for about 10 minutes for each of two available contents, though in some cases even longer. Afterwards, the researcher asked them a set of questions about their immediate experiences and impressions, and their responses were audio recorded and transcribed.

6.1.2. Observed interaction with gesture interaction system

The ability to control the video with hand gestures was strikingly diverse among our testers. We noticed that although 14 of the 20 people found it simple and were able to select the different modes of watching (region of interest versus panning and zooming) easily but three people struggled with it. Of the individual gestures, the selection of different channels/ROIs was by far the easiest to perform and get 'right' by the users. We found that several of the participants would do the gestures differently, for example, the pause gesture was explained as two parallel hands, but two flat hands in front of the body also worked and were used by one participant seamlessly; this was probably due to different interpretation of the instructions.

We found that several people tried to pan with an open hand despite the interface 'help pictures' clearly showing that panning around should be done with fingers. Zooming, however, appeared complicated and this was the most difficult of the tasks due to the system easily recognising a partly closed hand as fingers and thereby continue zooming in when the user tried to take his/her hands away to zoom more out. This resulted in 'swimming motions' being adopted by four users, in order to 'fool' the system into not reading the hands when moving them towards each other again.

The small help illustrations of gestures were interpreted widely and had an interesting effect: when people were not getting feedback right away they often attempted to make the gesture differently, possibly interpreting the system as not reading their gestures. This is intuitive for users but not necessarily correct in a system sense. The system could be malfunctioning or simply not recognise the person at all, or just be slow. Yet, as users of technology on an everyday basis, these participants are familiar with many different interpretations from technology and the tinkering that they often have to do in order to get the technology to do a desired action.

Several of the participants found it difficult to determine which mode (ROI or zoom/pan mode) the video was in and were using gestures irrelevant to the mode. They needed some instructions to be able to go back to the desired mode, which indicates a more distinct type of feedback. This could indicate a different mental model from the user, which might be alleviated through frequent use, but which also questions the feasibility of two main modes as foundation for the interface.

Similarly to what Vatavu describes from his studies of gestures, where participants preferred one-hand gestures (Gesture Control), our participants most often also just used one particular hand, especially the right hand (we did not ask

if the participants were right or left handed but it is assumed that most of them were using their dominant hand). This often resulted in almost awkward arm movements where they would put one arm in front of themselves to reach the desired target. It was not specified that only one hand or arm should be used for pointing and selecting, in fact it was possible to seamlessly shift from one hand to another according to our own testing of the system prior to the actual user testing. Yet this use of one hand was very prominent. Similarly, many of the gestures were exaggerated in comparison to what type of gestures would minimally be recognised by the system. Often participants would rather make a gesture bigger if it was not immediately recognised than repeating it, for example, slower. This highlights the human interpretation of how a system 'reads' behaviour. While there is no logic in a system recognising a larger gesture (unless it has been programmed to only recognise larger than X gestures), the interpretation from the participant's perspective is that the system did not quite *see* the gesture, hence, a bigger on is necessary.

We now describe the method and results from the tablet interface study before discussing our findings in broader perspective.

6.2. *Tablet interaction study*

6.2.1. *Method*

For the tablet interaction, we tested 16 people, between the ages of 22 and 46 with a median age of 28 years. Nine of them were not very interested in football and the rest mainly indicated a moderate interest with one being an avid football fan. Almost half of the participants had watched panoramic video before in different contexts, including large-scale movie experiences. We studied the interaction on the tablet as a second screen, running the full size panorama video on a large computer display while running the zoomable content on the table (Figure 2). After a short introduction to the system, the participants in this study were instructed to simply interact with the content, trying to follow the ball and enjoy the game. We recorded the interaction with two cameras similarly to the hand gestures interaction study, from the front and back. After the participant had been interacting with the system for about five minutes the researcher asked them a set of questions about their immediate experiences and impressions. Their responses were video recorded and transcribed.

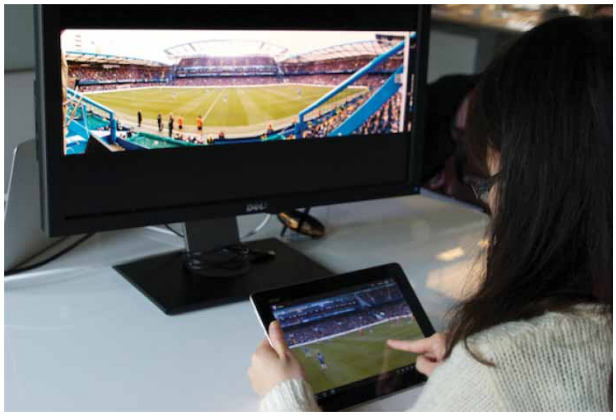

Figure 2. Tablet interaction set-up.

6.2.2. *Observed interaction with tablet system*

The tablet interaction seemed at first glance to be easier for participants compared to the hand gesturing, as it was evident that the participants were familiar with interacting with touch interfaces and were able to draw on this familiarity when performing the basic navigation features available in the tablet prototype. Their instructions for navigating in the image were broadly formulated (i.e. follow the action, look into details of your choice) and followed by a brief demonstration of the touch interaction techniques (swiping for panning and tilting, pinching and using the invisible scroll bar for zooming in and out). The participants were able to replicate these interaction techniques without any major issues. However, although all the participants were familiar with touch screen navigation, it was not straightforward to, for example, follow the ball in the football match. This stemmed from our calibration of the zooming, which was a bit too fast and the participants then had to use an 'invisible zoom bar' on the side of the screen. When this was pointed out to the participants, they were able to zoom more fine-grained, but still, it was difficult for most of them to follow the ball. We looked at how often the users looked up at the main screen to either find the ball or get an overview of the screen, and interestingly we found that only five of the 16 participants looked up more than once during the five-minute test and seven did not look up at all. So although 10 of the participants found it difficult to follow the ball and pan to the area they wanted, they still rarely used the larger screen for an overview.

In this study the participants pointed out specifically how a panoramic view lacks different angles of the camera; in a normal broadcast of a football match there would be several cameras at different positions in the stadium but with the panoramic view the camera angle is fixed. The participants highlighted that this contributed to the difficulty of following the ball on such a small screen as a tablet-sized screen.

7. Overall findings

Both studies revealed insights into the challenges and benefits of interacting with the panoramic content, as well as issues regarding each method of interaction. Where the tablet study mostly focused on detailed zooming and panning, we were able to explore social interaction around live event watching in the study of gesture interaction, where we observed seven couples. We now discuss two themes around

our findings before presenting more specific design-related considerations.

7.1. Controlling of the view

The main advantage of navigating (e.g. zoom and pan) within a large image is of course to be able to control the actual view. We confirmed that one of the reasons why participants wanted more control was that it would give them a personal view of the live show/event:

> It is always the wrong pictures they are showing, showing something completely different, and then you see something like 'that was really weird' or if you think the referee made some mistake and you want to watch it and form your own opinion,

said one participant from the tablet study after talking approvingly about the navigation option on the smaller screen. It was evident that most sports viewers knew where to look on the screen, further supporting our observation that participants knew where they wanted to pan the zoomed-in section of the image. They acknowledged that it can be useful to control their own view, as one of the gesture study participants expressed: 'For this it makes sense if you have a lot of things going on and you want to follow.' However, there was clearly a limit to the desire to control and constantly having to actively choose the view: 'I think it's nice to be able to zoom, but I was thinking about constantly doing that. I'd love that sometimes somebody does it for me', said another participant from the same study. It also reflected the observation from the gestures study that participants got tired of their large gestures, gestures that were in fact not always necessary for the functionality.

In terms of using the larger screen where the full panoramic content was running during the tablet test, participants found it useful to get an overview; one participant explains:

> If I cannot see the ball when I move the picture [on the tablet] then I look up [at the big panoramic image] to see where the ball is. Even if the ball is here and I want to see the whole image, so it quite depends. I think it is quite important to have this overview. You can see the players, where they stand. And the two teams. (…) I guess the position of the players is quite important.

This illustrates well how part of controlling the view was also about being able to view the show 'from above'. The panorama was thereby giving viewers the visual experience of seeing the game from the grandstand, '(…) I like it, it is really cool to see the whole field, it is like I'm sitting there', said one of the tablet study participants.

Controlling the view by zooming in and out was highly appreciated. It was used by all participants in the interaction user studies and it became a preferred functionality to zoom in and out, although it was not trivial to master technically. However, some participants were worried about content being lost. 'There is a danger with all this zooming in because you lose the overview and then it just takes a few seconds and "oh where are they now".' The issue here is how to get the best information, and the zooming is seen as a delicate instrument for getting more details suiting personal needs. Constantly zooming in and out creates a distraction in the interaction with the panoramic content. 'It would be good to be able to move between regions of interest without having to zoom out to the panorama again', one participant commented, suggesting that a feature similar to a cut between cameras would be useful.

7.2. Expecting technical advancement

Our studies also highlighted the harsh requirements that users have on technical workings of these types of systems today, and although this system is capable of ultra high-definition capture there are technical limits to the resolution that can be provided in a detailed subsection of the larger image. This most critically became evident when participants zoomed deep into details of the image. 'The resolution is the biggest killer of the application, if I zoom I want a proper resolution, otherwise it doesn't make sense. If I could get detailed view I would zoom in more', one of them commented. Several participants reported that they sometimes zoomed in past what they perceived as acceptable image quality. 'I wanted to look at the people but then it's out of focus, to see what they are doing', one viewer stated, giving a concrete example of a situation where they steered their view into a very narrow section of the scene without getting more detail in the image. The sensitivity and speed of the zoom were other factors that limited the interaction. 'I was able to follow the ball, except for the long shot, when the ball would disappear. But it is also the lag of the system, if zoom was faster, it would help.' Hence, the allowances of the zoom control – responsiveness, speed and zoom level – are important both individually and in combination, in the experience of interactive navigation in panoramic images.

Another issue was highlighted by participants who pointed out specifically how a panoramic view is lacking different angles of the camera, like 'It is quite a freedom, but I'm missing different angles like in real broadcast.' Because of the camera viewpoint being fixed, it was problematic in some situations to get a good view despite the possibility to zoom, compared to a normal broadcast of a football match, where there would be several cameras at different locations in the stadium. This was an issue in both studies.

8. Design-related considerations

We set out to explore, broadly, two interaction modes with panoramic video content, not just in order to compare these two but to find characteristics of successful and complicated use for each. We inevitably have to compare the two modes in some ways, in order to find which one is more relevant to use for different functionality. In this way, it was clear

that zooming was easier for the participants on the tablet, using two fingers to pinch and spread the picture. Yet, due to the calibration of the zoom, it was not unproblematic on the tablet either. We envision that a smoother implementation will alleviate this problem on the tablet. Four specific themes in relation to the interaction design are worth highlighting: transfer of interaction mode from one platform to another; complications with second screen interaction; different functionality for different interaction paradigms; and wide system interpretation of gestures.

8.1. *Transfer of interaction mode from one platform to another*

In terms of familiarity with touch screens versus gestures, there are a few possible design implication: in cases where the interaction techniques in themselves are not novel, their affordances will benefit from replicating those of familiar touch interfaces on other hardware/systems, i.e. the swiping motion is broadly used for navigating in images in a wide range of applications on mobile phones, tablets, and fixed installation touch screens. The established norms for how this manual interaction is typically translated into changes in the display of the image by the computer are similar across application areas. For instance, the scrolling speed following a swipe movement in a digital map or a collection of photos, and the way this scrolling movement decelerates as the finger leaves the screen, are familiar and recognisable by users of those applications. These familiar behaviours in digital media can be used as a resource when designing for panoramic video navigation. The specifics of large-scale imagery, such as panoramic television, may call for adjustment of these established interaction behaviours but any such adjustments should be carefully considered so that they do not introduce unnecessary complexity. Rather than introducing new behaviours that may seem to correspond to the specific features of the panoramic image, it is important to try to adapt established techniques to the extent possible.

8.2. *Complications with second screen interaction*

In terms of second screen use and how users use the large screen for reference, our results indicate that although users have access to both an overview and a second screen device, they may not intuitively switch between the two and make use of the different affordances – detail/interaction and overview – of the two displays. Instead, the interactive features in the tablet prototype seemed to take focus from the overview in the main screen. One potential added value for an interactive second screen device in panoramic live television would be that the viewer could go back and forth between overview and detailed shots of the action, similarly to how a broadcast producer mixes views of the action in a live broadcast (Perry et al. 2009). But for viewers to be able to take advantage of these two views in an intuitive way, it seems that merely providing two displays is not enough. A more intuitive use of both displays may build up over time, but our observations would suggest that it would be helpful to guide the user in how to use the two screens in parallel, by design. As an example, events in the live action that involve the risk of losing track of important actions may trigger an alert to direct the viewer's attention to the overview image.

8.3. *Different functionality for different interaction paradigms*

When considering the overall experience of a set of features available in one interaction mode or the other (here, gestures or touch), the design of features for each mode could be made to emphasise the positive affordances of that mode of interaction, while still providing the basics for interacting with the panoramic TV content. For instance, features for navigating through swiping movements could be more detailed on a tablet application, while the boundaries for the corresponding interaction in a gesture application could be set tighter, in order to avoid unexpected behaviours in the image. This would imply a slightly more limited responsiveness in the gesture interaction case, in favour of a more coherent and predictable user experience across the two modes. Conversely, the gesture application may take advantage of a larger set of easily accessible control features, such as audio and volume controls that may be assigned individual gestures, than the tablet interface where the same features may need to be embedded in menus that require averting the attention from viewing the live action.

8.4. *Wide system interpretation of gestures*

The wide interpretation of gestures that we observed leads to one distinct conclusion: it is important to design for a wide interpretation, leaving the smaller details of the gestures open within the system. The users all had the same instruction and watched the same icons of example gestures, yet they did not interpret these in the same way. When designing gesture interaction, identical gestures from person to person are impossible, but their different interpretations by users result in an even bigger spread in terms of gesture performance. Thus, different sets of gestures should ideally be designed so that overlaps between sets are avoided to the greatest extent possible. This was a key consideration in the development of the prototype in this test, yet sets of gestures that were designed to be visually separate were mixed up by the system's image recognition, due to individual users' interpretations of those gestures. This points to an urgent need to take a range of interpretations into account and to try to predict and avoid any intermittent poses that could potentially confuse the image recognition. The amplitude of gestures was another aspect that was observed to be distinctly more prominent in the free-hand gesture tests than in the tablet tests. This amplitude can be managed, as seen for instance in Microsoft Kinect games, where the system recognises the shape of a child

and adjusts the predicted amplitude of the child's motions accordingly. Similarly, the system could be trained to calibrate for the amplitude of individual user's gestures and store these values in individual user profiles.

9. Conclusion

In this paper, we presented two new interaction modes for consuming live video experiences in a panoramic view, wide gesture interaction for a large communal screen scenario and small tablet-based touch interaction with a second screen. We found that although both interaction methods were useful for navigating the panoramic video, each mode afforded different contexts, and suggest that when designing for interaction with these rich-content television experiences emphasis is placed on providing a variety of choices in terms of interaction possibilities. These two interaction modes show interesting promise but also pose many challenges. It is not clear that either one of them is seamlessly fitted for interaction with panoramic television in the home but instead each type of interaction could be explored for different types of functionality. We emphasise that it is often more relevant to refine a system's already existing interaction mode rather than radically change it and it is, therefore, important to make sure that future scenarios, like the ones presented in this paper, should be adjusted to include simple choices from the user within the context.

Acknowledgements

We thank all the participants in our studies.

Funding

This work was supported by the European Union's Seventh Framework Programme (FP7/2007-2013) [grant no. 48138].

Note

1. This project is part of a larger frameworkFascinatE, which aims to provide a panoramic experience of live event content. See also http://www.fascinate-project.eu.

References

Bleumers, Lizzy, Wendy Van den Broeck, Bram Lievens, and Jo Pierson. 2012. "Seeing the Bigger Picture: A User Perspective on 360 ° TV." In *Proceedings of the 10th European Conference on Interactive TV and Video (EuroiTV '12), Berlin, Germany, July 4–6*, edited by Stefan Arbanowski, Stephan Steglich, Hendrik Knoche, and Jan Hess, 115–124. New York: ACM.

Fehn, C., C. Weissig, I. Feldmann, M. Muller, P. Eisert, P. Kauff, and H. BloB. 2006. "Creation of High-resolution Video Panoramas of Sport Events." Proceedings of the eighth IEEE international symposium on multimedia, ISM '06, 291–298, Washington, DC.

HegoOB1system. Accessed November 17, 2011. www.ob1.hegogroup.com/

Hilton, A., J. Y. Guillemaut, J. Kilner, O. Grau, and G. Thomas. 2011. "3D-TV Production From Conventional Cameras for Sports Broadcast." *IEEE Transactions on Broadcasting* 57 (2): 462–476.

imLIVE demo. Accessed November 18, 2011. www.immersivemedia.com/markets/imLIVE/index.html

Jota, R., J. M. Pereira, and J. A. Jorge. 2009. "A Comparative Study of Interaction Metaphors for Large-scale Displays." Proceedings of CHI 2009, work in progress, April 4–9, 4135–4140. Boston, MA: ACM Press.

LIVE Project. Accessed October 7, 2011. http://www.ist-live.org/

My-eDirector Project. Accessed October 7, 2011. http://www.myedirector2012.eu/

Nakatoh, Y., H. Kuwano, T. Kanamori, and M. Hoshimi. 2007. "Speech Recognition Interface System for Digital TV Control." *Acoustical Science and Technology* 28 (3): 165–171.

Neng, Luís A. R., and Teresa Chambel. 2010. "Get Around 360 ° Hypervideo." *Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek'10), Tampere, Finland, October 6–8*, edited by Heljä Franssila, Olli Sotamaa, Christian Safran, and Timo Aaltonen, 119–122. New York: ACM.

Olsen, D. R., B. Partridge, and S. Lynn. 2010. "Time Warp Sports for Internet Television." *ACM Transactions on Computer-Human Interaction* 17:16:1–16:37.

Panasonic. 2010. "Panasonic Releases New 3D Image Sensor for Gesture Based Applications." Press release, Panasonic, November 29. http://pewa.panasonic.com/news/press-releases/panasonic-releases-d-imager-3d-image-sensor/

Patrikakis, C. Z., N. Papaoulakis, C. Stefanoudaki, A. Voulodimos, and E. Sardis. 2011a. "Handling Multiple Channel Video Data for Personalized Multimedia Services: A Case Study on Soccer Games Viewing." PerCom workshops 11, Seattle, WA, 561–566.

Patrikakis, C. Z., N. Papaoulakis, P. Papageorgiou, A. Pnevmatikakis, P. Chippendale, M. S. Nunes, R. Santos Cruz, S. Poslad, and Z. Wang. 2011b. "Personalized Coverage of Large Athletic Events." *IEEE MultiMedia* 18 (4): 18–29.

Perry, M., O. Juhlin, M. Esbjörnsson, and A. Engström. 2009. "Lean Collaboration through Video Gestures: Co-ordinating the Production of Live Televised Sport." *Proceedings of CHI 2009*, Boston, USA, April 4–9, edited by Ken Hinckley, Meredith Ringel Morris, Scott Hudson, and Saul Greenberg, 2279–2288. New York: ACM.

Samsung Smart Interaction. http://www.samsung.com/us/2012-smart-tv/

Schreer, O., G. Thomas, O. A. Niamut, J.-F. Macq, A. Kochale, J-M. Batke, J. Ruiz Hidalgo, R. Oldfield, B. Shirley, and G. Thallinger. 2011. "Format-agnostic Approach for Production, Delivery and Rendering of Immersive Media." NEM summit, Turin, Italy, September.

Smolic, A., K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand. 2006. "3d Video and Free Viewpoint Video – Technologies, Applications and MPEG Standards." IEEE international conference on multimedia and expo, Toronto, ON, 2161–2164.

Soutschek, S., J. Penne, J. Hornegger, and J. Kornhuber. 2008. "3D Gesture Based Scene Navigation in Medical Imaging Applications using Time-of-Flight Cameras." Computer vision and pattern recognition workshops, Anchorage, AK, June, 1–6.

Van den Bergh, M., E. Koller-Meier, F. Bosche, and L. Van Gool. 2009. "Haarlet-based Hand Gesture Recognition for 3D Interaction." Workshop on applications of computer vision, Snowbird, UT, December, 1–8.

Vatavu, Radu-Daniel. 2012. "User-defined Gestures for Free-hand TV Control." Proceedings of the 10th European conference on interactive TV and video (EuroiTV '12), 45–48. New York: ACM.

Vatavu, R. D. 2013. "A Comparative Study of User-defined Handheld vs. Frehand Gestures for Home Entertainment Environments." *Journal of Ambient Intelligence and Smart Environments* 5 (2): 187–211.

Vatavu, Radu-Daniel, and Stefan-Gheorghe Pentiuc. 2008. "Interactive Coffee Tables: Interfacing TV within an Intuitive, Fun and Shared Experience." In *EuroITV 2008: The 6th European interactive TV conference*, LNCS 5066, edited by M. Tscheligi, M. Obrist, and A. Lugmayr, 183–187. Salzburg. Heidelberg: Springer-Verlag.

Wang, Zhenchen, Stefan Poslad, Charalampos Z. Patrikakis, and Alan Pearmain. 2009. "Personalised Live Sports Event Viewing on Mobile Devices." UBICOMM, 2009 third international conference on mobile ubiquitous computing, systems, services and technologies, Sliema, Malta, 59–64.