Gesture Controlled Interactive Rendering in a Panoramic Scene

Axel Kochale
Technicolor R&I Hannover
Karl-Wiechert-Allee 74
D-30625 Hannover
++49 511 418-2397
axel.kochale@technicolor.com

J. Ruiz Hidalgo
Universitat Politècnica de Catalunya
Jordi Girona 1
E-08034, Barcelona
++34 934015765
j.ruiz@upc.edu

Malte Borsum
Technicolor R&I Hannover
Karl-Wiechert-Allee 74
D-30625 Hannover
++49 511 418-2482
malte.borsum@technicolor.com

ABSTRACT
The demonstration described hereafter covers technical work carried out in the FascinatE project [1], related to the interactive retrieval and rendering of high-resolution panoramic scenes. The scenes have been captured by a special panoramic camera (the OMNICAM) [2] with is capturing high resolution video featuring a wide angle (180 degrees) field of view. Users can access the content by interacting based on a novel device-less and marker-less gesture-based system that allows them to interact as naturally as possible, permitting the user to control the rendering of the scene by zooming, panning or framing through the panoramic scene.

Keywords
Panoramic content rendering, region of interest, free view point, sub-picture, gesture control, head and hand tracking, depth estimation.

1. INTRODUCTION
The variety of available end terminals require nowadays a format-agnostic production to prepare the content best suited to all. Format agnostic production means that the capture of audio-visual content will not be performed in the same format as it will be displayed and rendered on the receiving device. FascinatE terminals and services will supply interactive, personalized visual perspectives to enrich the user experience. Content navigation like pan, tilt and zoom (PTZ) allows the user to enjoy a real immersive experience beyond simple channel switching. Hence, special technical solutions are required in order to facilitate the retrieval of images on consumer type displays having less resolution and also not providing an aspect ratio and geometry which matches the panoramic scene format [2].

Furthermore, the availability of these new interactions with the content, such as gesture-based pan, tilt, and zoom navigation, which are enabled by format-agnostic rendering, open up new challenges in user interaction. Performing these interactions naturally might not be very easy and intuitive by using general devices like remote controls. Therefore, new devices or mechanisms need to be studied in order to facilitate the interaction with the format-agnostic representation.

This demonstrator showcases a device-less and marker-less gesture recognition mechanism based on a Kinect sensor. The gestures allow the user to perform interactions such as selecting menus presented on the screen or navigating through the high-resolution panoramic views of the scene.

WSICC workshop at EuroITV'13, June 24, 2013, Como, Italy.

2. INTERACTIVE RENDERING
The technical demonstration will show how panoramic content can be rendered on various types of consumer displays. Basically, the consumer will be able to interactively choose her or his preferred region of interest (ROI) as a free view point sub-picture out of the panoramic scene. Sub-picture in this respect implies that the user display shows a specific portion out of the panoramic scene. The portion out of the scene for the ROI-window (zoom-in and zoom-out) as well as the position of the ROI-window (pan and tilt) can be freely chosen and modified. The viewer will not only be in a position to interactively move around in a panoramic scene, moreover she or he will be able to link the ROI-window to a specific event or a dedicated subject of the scene (e.g. player or ball in a football match), forcing the ROI-window to automatically follow the object of his preference.

The technical implementation of this content rendering feature has been studied and developed by Technicolor Research & Innovation, Hannover. The demonstration will show the real-time rendering capability which will allow for a free view-point navigation in a panoramic scene. The captured scenes have been shot in the course of the FascinatE project and cover several genres ranging from sport events to orchestral and choreography performances.

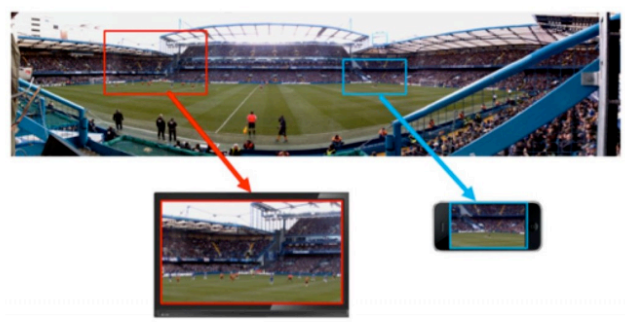

Figure 1. Examples of different Regions of Interest (ROI) within a panoramic video stream.

3. GESTURE CONTROL
It became obvious that this form of novel interactive content consumption will ask for new and innovative means of audiovisual screen interactions. Because one major scope of the FascinatE project was also to push for the paradigm shift from lean-back passive TV consumption towards lean-forward interactivity, special investigations and user studies have been carried out on the question on how to control such interactive content consumption. One proposed solution is an approach of gesture controlled user interaction.

The Universitat Politècnica de Catalunya (UPC) developed for this purpose a fast and robust head and hand tracking algorithm using depth information from a range sensor, allowing interactive and immersive applications. This functionality is used to control the real-time rendering platform developed by Technicolor with interaction commands like pan/tilt/zoom navigation and audio control commands.

In particular, the gesture demonstrator supports the following functionality to the user:

- Selection of items by pointing at an overlaid menu on the screen.
- Navigation through the panorama scene by panning, tilting and zooming in the content.
- Change between the available ROIs in the current scene.
- Pausing or resuming the content replay.
- Increasing or decreasing the audio volume and muting it completely.
- Passing the control of the system between users.

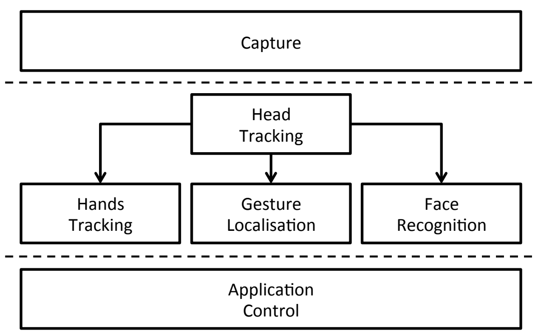

Figure 2. Architecture of the gesture recognition system.

The gesture control is divided into three main layers (Figure 2). In the first layer, the capture component is responsible of communicating with a single Kinect camera and feeding the images into the system. The second layer is the core of the interface where the processing, detection and recognition algorithms take place [3], [4]. The second layer is composed of a head tracking, where the depth image (obtained in the capture module) is analyzed to detect heads, a face recognition component where users of the system are identified, a hand tracking component, where dynamic gestures (gestures performed with moving hands) are recognized, and a gesture localization component, where static gestures (gestures performed with still hands) are detected. Finally, the third layer of the architecture contains the application control. This layer is responsible of acquiring all the detections and tracking information obtained by the components in the middle block (head, hands, user recognized and classified gestures) and mapping them into the functionality previously listed.

In the demonstrator, the gesture control system runs in real-time using basic hardware, an 8 CPU laptop and a single Kinect sensor as the input for all algorithms.

4. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 248138.

5. REFERENCES

[1] FascinatE Project web site: (http://www.fascinate-project.eu/)

[2] Niamut O., Kaiser R., Kienast G., Kochale A., Spille J., Schreer O., Ruiz Hidalgo J., Macq JF., Shirley B., Towards A Format-agnostic Approach for Production, Delivery and Rendering of Immersive Media, *Proceedings of ACM conference on Multimedia Systems* (Oslo, Norway, Feb 27-Mar 01, 2013), to be published

[3] López-Méndez, A. and Casas, J., Can our TV robustly understand human gestures? Real-Time Gesture Localization in Range Data, in *Conference on Visual Media Production*, London, UK, 2012.

[4] Suau, X., Ruiz-Hidalgo, J., and Casas, J., Real-time head and hand tracking based on 2.5D data, *IEEE Transactions on Multimedia,* vol. 14, no. 3, pp. 575-585, 2012.