ImmersiaTV: an end-to-end toolset to enable customizable and immersive multi-screen TV experiences

Juan A. Núñez, Mario Montagud, Isaac Fraile, David Gómez, Sergi Fernández

i2CAT Foundation

C/ Gran Capità 2-4, Edifici Nexus I, Barcelona, Spain

{juan.antonio.nunez, mario.montagud, isaac.fraile, david.gomez, sergi.fernandez}@i2cat.net

ABSTRACT

ImmersiaTV is a H2020 European project that targets the creation of novel forms of TV content production, delivery and consumption to enable customizable and immersive multi-screen TV experiences. The goal is not only to provide an efficient support for multi-screen scenarios, but also to achieve a seamless integration between the traditional TV content formats and consumption devices with the emerging omnidirectional ones (e.g. 360º video and spatial audio), thus opening the door to new fascinating scenarios. This paper presents the end-to-end toolset that is being developed to overcome existing challenges, with a special focus on the building blocks that require a re-definition to meet the targeted goals. The contents that have been created and the considered pilot scenarios to demonstrate the feasibility and benefits of the project are briefly described as well. Overall, the contributions of ImmersiaTV allow providing new forms of customizable, interactive and immersive experiences across multiple devices, including main TVs, tablets, smartphones and Head Mounted Displays (HMDs).

Author Keywords

360º Video, Immersive TV, Media Synchronization, Omnidirectional Media, Storytelling, QoE, Virtual Reality.

ACM Classification Keywords

H.5.1 Information interfaces and presentation (e.g., HCI). Multimedia Information Systems.

INTRODUCTION

The combined usage of (connected) TVs and companion devices for the consumption of TV-related contents is attracting the interest of the research community, industry and, especially, of consumers (e.g. [1-2]). One step further, the integration of Virtual Reality (VR) and omnidirectional contents (e.g. 360º video and spatial audio) and devices within the TV landscape can open the door to new fascinating scenarios.

ImmersiaTV is a H2020 European project (http://www.immersiatv.eu/) targeted at overcoming the existing challenges to enable customizable and immersive multi-screen TV experiences. By considering the current heterogeneity in terms of contents and consumption devices, with a special focus on omnidirectional media, ImmersiaTV proposes backward-compatible and standard-compliant re-definitions to the end-to-end chain to make these new experiences a reality. In particular, novel forms of content capturing, production, storytelling, encoding, distribution and consumption are being created. These contributions allow providing different versions and formats of media contents, which can be interactively selected and presented on multiple devices in multi-screen scenarios (see Figure 1) in a coherent, intuitive and personalized manner.

This paper initially provides an overview of the key building blocks of the end-to-end toolset developed in ImmersiaTV. Then, the created contents and considered pilot scenarios to demonstrate the feasibility and to evaluate the benefits of the project are briefly described.

IMMERSIATV PLATFORM

Figure 2 provides a high-level overview of the key building blocks of the end-to-end ImmersiaTV toolset / platform. Next, these building blocks are briefly described.

Content Capture and Stitching

A novel sub-system for the capture, replay and processing of both traditional and omnidirectional contents has been developed. The system adopts a distributed architecture and allows a seamless integration between multiple 360º video camera rigs and conventional ones, while leveraging current practices in camera and computing technology systems.

Figure 1. Envisioned multi-screen scenarios in ImmersiaTV.
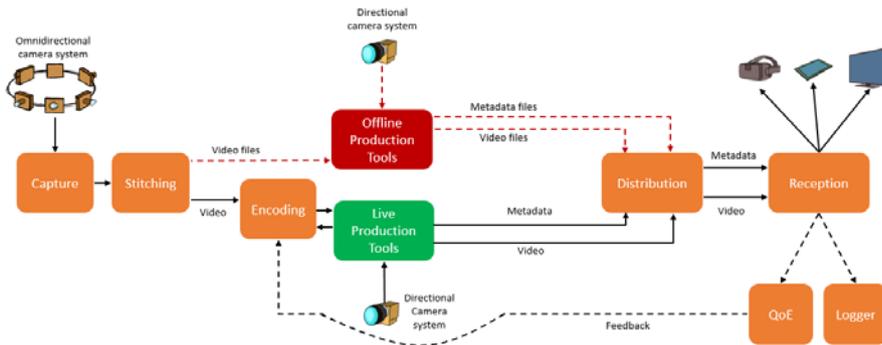
Figure 2. High-Level Overview of the end-to-end ImmersiaTV platform.
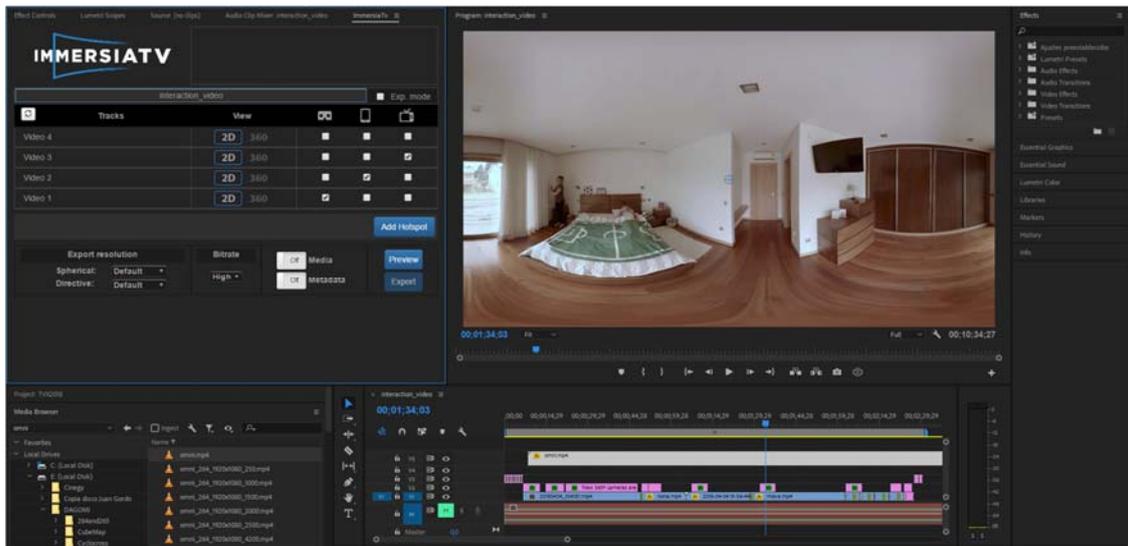
**Figure 3. Screen Capture of the Video Production / Edition Tools.**

Content Production and Edition

A custom production workflow supporting the live and offline edition of traditional and omnidirectional contents has been developed in ImmersiaTV. It is based on both new and adapted tools, and includes a new plugin for Adobe Premiere Pro (Figure 3). Among other functionalities, the plugin allows the definition of: i) inserts to be placed within the omnidirectional videos (e.g. overlaid contents and notifications, icons…); ii) proper metadata for and regarding the available contents (e.g. their location, media types, dependences between them…); iii) interactivity features, transitions or effects to be included for the available contents; and iv) the targeted devices on which the available contents can be presented, by properly preparing them.

Content Encoding, Packaging, Signaling and Delivery

On the one hand, ImmersiaTV is developing new standard-compliant solutions for the adaptive encoding of the traditional and immersive contents, taking into account the heterogeneity of the targeted consumption devices. Regarding 360º videos, the use of tiling strategies and of both Equirectangular and Cubemap projection formats is supported, taking into account Regions of Interest (RoI). In particular, strategic tiling solutions when using Cubemap projection have been proposed, which consist of creating two tiles, and assigning higher quality to one of the tiles covering the 270º horizontal panorama or just a single face of the cube. These strategies allow for a significant bandwidth reduction compared to when using traditional Equirectangular- and Cubemap-based solutions, without having a negative impact on the perceived Quality of Experience (QoE). Moreover, they are lightweight enough to be successfully integrated in current web browsers, enabling a synchronized playout and a seamless exploration of the full 360º area.

On the other hand, MPEG-DASH standard [3] has been adopted for the adaptive delivery of the media contents. In addition, appropriate metadata for media discovery, association, orchestration, playout and interaction are provided, by specifying standard-compliant extensions to the MPEG-DASH signaling and manifest file formats.

Interactive and Personalized Content Consumption

The immersive media contents are adaptively presented on the involved consumption devices, based on their resources and capabilities, on the narrative(s) specified during production and on the specific preferences, interactions and behaviors of the consumers. Different presentation modes are enabled, such as adding overlays ("*portals*") to the

omnidirectional videos with additional media assets, as well as adding intuitive and user-friendly notifications, transitions and effects. Exploration around the 360º area is provided via the built-in sensors within the mobile devices (e.g. gyroscope or tactile screens), the mouse or keyboard in traditional desktops and laptops, and head movements or joysticks in HMDs.

Two media players have been developed. One of the players is completely web-based, and has been developed by using well-known components, such as *DASH.js* and *three.js*. The other media player has been developed by using the GStreamer framework, together with a custom rendering module that connects to the Unity3D engine, called Gstreamer Unity Bridge (GUB). Both players provide cross-platform support.

In addition, modules to monitor and log Quality of Service (QoS) and QoE related metrics during media consumption have been also developed (see Figure 2). The collected statistics can be useful to analyze the performance of the system and the users' activity during the session, but can also be additionally used for dynamically adapting the media encoding and delivery processes.

Multi-Screen Scenarios
Key requirements to enable comfortable and coherent multi-device scenarios consist of providing proper: i) service and device discovery mechanisms; ii) association mechanisms between the involved devices and processes; iii) app launching mechanisms; iv) inter-device synchronization solutions; and v) interaction features. ImmersiaTV leverages the features and guidelines provided by the HbbTV standard [4] with respect these features. In addition, these features are augmented with ad-hoc and lightweight extensions (e.g., protocols, algorithms, techniques…) to provide full-fledged and efficient solutions for the mentioned requirements.

PILOTS & AVAILABLE CONTENTS
Apart from designing and developing the necessary technological components, ImmersiaTV believes in the convenience of creating appropriate contents and developing diverse pilots. This is essential to evaluate and demonstrate the proper performance of the overall platform but, most importantly, to validate the benefits provided to the consumers. Next, the pilots considered in the project are briefly explained.

Pilot 1: Documentary
The first pilot has been based on the production of a novel form of documentary, combining omnidirectional and traditional TV contents. The documentary has been titled "*DRAGON FORCE – The Making of Future Heroes*", and relates the daily life of a young boy who pursues the dream of becoming a successful football player. The making off of this pilot can be watched at: http://goo.gl/qrmqay A demo of the ImmersiaTV platform for this pilot was presented in [5], and a demo video can be watched at: http://goo.gl/HLyPL2.

Pilot 2: Live Sports Events
The second pilot has been focused on a live cyclocross sports event, by placing traditional and omnidirectional cameras at the most spectacular points of the track. This live pilot in a crowded and highly dynamic event was selected to gain insight in how and when to insert omnidirectional video and sound in these events, and how to combine and present the media contents with the proper layout, layering and interaction features. The complexity of the live event also becomes a challenge to test the real-time production tools and to explore the most appropriate methodologies to provide the proper storytelling, interaction features and maximize the QoE perceived by the audience. A demo video for this pilot can be found at: http://goo.gl/Bp9Rp9.

Pilot 3: Evolved Documentary + Live Event
The third pilot is being prepared at the time of writing this text. It has been defined from the lessons learned in the first two pilots and aims at testing the most recent contributions of the project. In particular, the pilot consists of refined iterations for both an offline documentary and a live event, integrating improvements along the end-to-end chain, in terms of: encoding and stitching methods; media formats; delivery infrastructure; interaction and presentation features (e.g. introduction of multiple simultaneous viewpoints); and advanced synchronization solutions.

On the one hand, the documentary adopts novel storytelling and exploration modes to allow consumers to customize their experience. While watching it, users will be able to freely explore and navigate between contents and scenes (with specific order, timing, and layout properties), as well as between devices, thus being provided with a highly personalized and immersive story, determined by their own actions. The documentary is based on a fiction "whodunit" scenario, in which users have to look for clues to solve a crime. Herein, each device plays it specific role: the main storyline is developed on TV, the exploration mode in HMD, and examination of facts and characters on the tablet.

On the other hand, the live event in pilot 3 introduces a focus shift from the one in pilot 2. While director-control scenes were provided to the consumers in pilot 2, a free exploration of multiple views of the event will be possible in pilot 3.

Extra Pilot: "Roméo et Juliette" Opera
In a joint initiative with the H2020 European ImAc (Immersive Accessibility, http://www.imac-project.eu) project, ImmersiaTV has participated in an extra pilot: an immersive recording of "Roméo et Juliette" at the Gran Teatre del Liceu Opera House, in Barcelona. The goal is to create a ground-breaking product that will allow consumers enjoying an opera performance in a highly interactive, immersive and personalized manner.

In particular, the following video capturing equipment was used in the recording, being placed at strategic points, both on and off the stage:

- 4 360º cameras.

- 2 320º cameras.
- 4 170º cameras.
- 5 classical directive cameras.

This multi-camera and multi-view recording allows users to enjoy the opera performance in the classical (directive) format on a main TV, but they will be additionally able to experience the event using the preferred content format, from the many different viewpoints, by dynamically selecting the contents recorded by the available directive and (nearly-)omnidirectional cameras, and by also using the preferred companion devices (e.g. tablets, smartphones and HMDs), being able to dynamically switch between them. Moreover, as opera is of the most demanding scenarios when it comes to audio quality, more than 80 distributed audio sources (including 3D –Ambisonics– and binaural formats) were also used in the recording. This allows providing detailed, position- and view-dependent sound landscapes (aka soundscapes), depending on the selected camera(s), the current point(s) of view and the active consumption devices, thus increasing the feeling of immersion and of realism.

An example of this multi-screen scenario for the opera use case can be seen in Figure 4.

Finally, a customizable, adaptive and assistive presentation of accessibility contents (e.g. text and audio subtitles, audio description, and sign language interpreting) related to the 360º videos is being provided thanks to the contributions of the ImAc project. Regarding subtitles, different colors for each speaker can be added, and different font sizes and languages can be dynamically selected. Subtitles are, by default, presented at the bottom region, as in typical video players. However, it is also possible to present them at the top region of the player (Figure 5). This presentation option is known as super-titles or surtitles, and is typically used and preferred in musical and theatre performances. In order to properly identify the active speaker(s) and main actions while freely exploring the 360º area, different guiding mechanisms are provided, such as adding arrows, a compass, or sided text. Similar guiding mechanisms are also provided for sign language and audio description.

The playback of all the selected contents will be accurately synchronized, regardless on the number and types of consumption devices being used. A demo video showing all the mentioned features for the opera pilot can be watched at: https://goo.gl/xpqVPF

CONCLUSIONS

This paper has provided an overview of the ImmersiaTV project, which targets the creation of novel forms of TV content production, delivery and consumption to enable customizable and immersive multi-screen TV experiences. The key building blocks and features of the end-to-end toolset being developed have been briefly presented. Then, the created contents and considered pilot scenarios to evaluate the contributions of the projects and demonstrate its benefits have been described.

Figure 4. Multi-screen consumption experience for an opera use case, including multi-location omnidirectional videos and spatial audio.

Figure 5. Personalized and Assistive Presentation of Subtitles in 360º Environments.

REFERENCES

1. V. Vinayagamoorthy, R. Ramdhany, M. Hammond, "Enabling Frame-Accurate Synchronised Companion Screen Experiences". ACM TVX '16, June 2016.

2. F. Boronat, D. Marfil, M. Montagud, J. Pastor, "HbbTV-Compliant Platform for Hybrid Media Delivery and Synchronization on Single- and Multi-Device Scenarios", IEEE Transactions on Broadcasting, vol. PP, no. 99, pp. 1-26, 2017

3. ISO/IEC 23009-1: 2012. Information Technology. Dynamic Adaptive Streaming over HTTP (DASH). Part 1: Media Presentation Description and Segment Formats. April 2012.

4. Hybrid Broadcast Broadband TV (HbbTV) 2.0.1 Specification, HbbTV Association Resource Library, https://www.hbbtv.org/resource-library, July 2016

5. J. A. Núñez, S. Malewski, S. Fernández, J. Llobera, "Production and Delivery of Video for Multi-device Synchronized Playout", ACM TVX '17, June 2017.

6. D. Gómez, J. A. Núñez, M. Montagud, S. Fernández., "ImmersiaTV: Enabling Customizable and Immersive Multi-Screen TV Experiences", ACM MMSYS '18, June 2018.