Interactive 360° Video and Storytelling Tool

Louay Bassbouss, Stephan Steglich
Future Applications and Media
Fraunhofer FOKUS
Berlin, Germany
{firstname}.lastname@fokus.fraunhofer.de

Igor Fritzsch
BitTubes GmbH
Berlin, Germany
igor.fritzsch@bittubes.com

Abstract— Existing 360° video players on the market playback only one type of media in their timeline. This paper introduces a new tool that allows creating interactive 360° videos with storytelling elements that can switch between flat and 360° videos in one timeline. The tool also integrates other media types like images, audio, and web resources that can be interlinked with each other to create one or multiple branched stories via non-linear video technology. The benefits of this approach are far-reaching, i.e., it allows broadcasters to create genres of programming that combine both traditional footage and immersive 360° video segments in one seamless experience. It enables a highly effective and attractive lean-back and lean-in experience for the audience. Moments of immersion can then be coupled with traditional storytelling, which is an attractive option for content creators who can enhance existing material with immersive moments in a way that is cost effective and allows them to fall back on their pre-existing skills as story and program creators. It will pave the way for much broader adoption and production of spherical content among content providers of the first wave, as it lowers the barrier of entry. The tool will enable content producers and editors to create stories in 360° videos and in combination with non-360° media content by providing an easy-to-use web-based editor that targets devices with various characteristics and input capabilities such as TVs, tablets and head-mounted displays (HMDs).

Keywords—360° Video; VR; Storytelling; HMD; Interactive Video; Web Tools

I. INTRODUCTION

The purpose of telling a story is to create an emotion to influence action, especially in the marketing world. The Forrester Research study says that one minute of video is worth 1.8 million words [1], when combined with the immersive features special to virtual reality (VR) in which viewer actively interacts with the content as the narrative is revealed, makes VR an ultimate storytelling platform. VR is revolutionizing how information is conveyed and processed. Marketers, advertisers, filmmakers, and news publishers are also embracing this transformation. In April 2017, during the National Association of Broadcasters Show held in Las Vegas [2], it was acknowledged that 360 is rapidly being adopted as the next platform for compelling storytelling. As the microprocessing technology is continuously advancing and growing, so does the number of subscribers, as predicted by GSMA, there are 5 billion subscribers worldwide [3]. The future of VR is also evolving and shifting towards mobile VR, i.e., mobile experience compared to conventional tethered experience. The tech giants like Samsung and Google are leading the way when it comes to selling the VR equipment, e.g., Samsung alone shipped 6.3 million VR headsets in 2016 [4], Google sold 8.4 billion Cardboards and 0.26 million Daydream, the percentage of mobile VR headsets was 98% [5]. The art of constructing compelling storytelling compromises some techniques; it is not like applying traditional cinematic principals to a mounted headset. The understanding of audience, environment, and the story itself is required to construct engaging 360°/VR experiences. Below are some key aspects to consider in storytelling:

A. Presence

Let the viewer feel presence in the space of the experience and sharing that environment with characters. This sensation makes the viewer active and to have a desire to play a role. As explained by Katy Newton & Karin Soukup, there is no such thing as a neutral observer [6].

B. Point of View (POV)

When it comes to 360° storytelling, another critical aspect is deciding which POV to use in the story. It is the basis of creating presence. The POV can be based on an objective (third-person POV) or through the eyes of a character (first-person POV). As Maria Alexander explains how POV affects the narrative in VR, according to her, the biggest challenge in movie scripts to novel adaptations is addressing point of view (POV) [7]. Intrinsically VR asserts first-person POV that is contrary to theatrical narration experience. Therefore, whatever POV you choose the aim should be to maintain presence of the audience.

C. Effective use of Sound

As applicable in real life, sound also plays a vital role in capturing the audience’s attention to critical sights and creating a complementary environment in an immersive experience.

II. STATE OF THE ART

This section discusses some of the latest developments, technologies, and studies for immersive storytelling.

A. Spotlight Stories for YouTube

Google has introduced immersive 360° animated videos with "Spotlight Stories" for YouTube. Movie creators can take advantage of smartphone’s sensors, e.g., gyroscope and
accelerator, to create an immersive and interactive experience for 3D, 2D animations, and 360° videos [8]. To make every view of the movie unique, the story can be traced in different ways, i.e., a person can follow an action based on a chosen angle, and by moving around the phone, other sub-stories within the movie are revealed.

B. Storytelling to Storyliving

Google has conducted a study [10], which comes out with a concept of experiencing stories and brand messages through VR, termed as "Storyliving". The study concludes that VR can convey brand messages and advertising that is more effective than traditional text-based storytelling, still images, and videos. The study also finds that storytelling experience should make a strong emotional impact through VR.

III. INTERACTIVE 360° AND STORYTELLING

Considering the studies and products as mentioned above, they all have one thing in common: Their focus is on mobile or head mounted devices with their main input methods movement (gyroscope) and touch. However, what about consuming video content aside from small screens? What about TV screens and their particular input method remote control? This device category, among providing an easy to use and extensible editing tool, is in the focus of our research. Of course, the commonly targeted device categories are covered as well to provide an open platform reaching as many consumers as possible. Before going into detail of the Interactive 360° and Storytelling platform, this section introduces the key elements that can be used in any combination to create stories among a set of flat and immersive media content as listed below:

A. Interactive Overlays

Sometimes it is desirable to obtain more detailed information from a mix of images, texts, or comparable content, but due to their spatial dimensions, they cannot be placed unobtrusively near the point of interest. This requires a superposition of the field of view either entirely or only partially utilizing a "clickable" overlay. Once defined and configured, the visibility of such an interactive overlay can be triggered by events fired by any other element after a user interaction or implicitly, e.g., via a video playback event.

B. Sticky Marker

A Sticky Marker is a component that consists of a single element or composition of elements, such as images or texts, that serves primarily as a reference to a specific spot or coordinate within the 360° space. Sticky Markers are superimposed on the video and, in contrast to traditional overlays, always remain in the field of view regardless of the viewing direction. However, they are only visible if the corresponding spot or coordinate is out of certain bounds or violates previously defined rules. For example, Sticky Markers might notify at the margins of the video that certain hotspots exist outside the field of view.

C. Autofollow

The Autofollow feature is an action that tracks an object in the 360° video. During playback, the camera will be aligned according to the position of the marker, so that the object is always focused at the center of the field of view (FOV) at all times.

D. Scene Detection with default FOV

This feature analyses and detects scenes in a 360° video automatically, which allows the viewer to switch very quickly between them. It can be combined with other actions so that in addition to the change of scene/position, e.g., the change of the FOV, zooming into a specific object or pausing of the video can be made as well. The scenes are provided as sections in the video (see next point).

E. Captioning and Sectioning in 360° Videos

Captions like subtitles are relevant for 360° videos as well as for flat videos. The same applies for sections as mentioned in the previous point. A section can be a scene in the video or manually defined with a custom start and end position. In 360° videos, subtitles and sections should consider additional aspects that are not necessarily relevant for flat videos. For subtitles, a dynamic positioning mechanism has been introduced that only displays a person's subtitle when it is in the current FOV. Moreover, regarding sections in 360° videos, a set of relevant FOVs were introduced for each section. When the viewer selects a section, the video player seeks to the corresponding position in the video timeline, and the camera moves and zooms to one of the defined FOVs related to the selected section. The first defined FOV is considered as the default.

F. Picture in Picture

Picture in Picture is a common way of displaying another downsized traditional footage or 360° video over the actual video. Using web technologies, this can be done either in HTML or WebGL only, with WebGL providing considerable value to HTML, fitting seamlessly into the space of a 360° video to give the viewer an even more realistic experience.

G. Video Branching

Video Branching exchanges only the underlying video. The superimposed interactive levels and thus, the interactive elements embedded in them are retained during this change. An application would be, for example, a stationary 360° recording, in which it would be possible to switch between day and night views.

H. Project Branching

Project Branching is the switch between two interactive videos in which both the video and the interactive layers superimposed on the video will be exchanged. It can be switched between two videos of the same format or between traditional footage and 360° videos. It enables not only new forms of storytelling but also increases the Quality of Experience (QoE) at the same time, leading to a potential increase in engagement.
IV. INTERACTIVE 360° PLATFORM

The platform provides the tools and technologies that are needed to distribute an interactive 360° video experience. Fig. 1 depicts the interplay between front- and backend, customers, and providers with their appropriate tools, technologies, and systems.

Figure 1: Architecture - 360° Interactive Storytelling Platform

The following sections provide an overview of the editing tool, and the various players running on the customer's target devices.

A. Editing Tool

To enrich a 360° video with overlays and interactive elements, a tool with a graphical user interface is a convenient way to allow editors and other parties to access this technology for their content. A graphical interface has been developed that will enable editors to place items easily to specific coordinates via drag and drop. A variety of object types (called "models") support the process to create a rich interactive 360° experience in a short time. Fig. 2 shows the user interface for an editor to choose from a range of interactive models (bottom left box). The middle box on the left contains all available settings for the selected model. On the right, there is a preview of the 360° video where the editor can seek to desired scenes and rotate the camera to favored viewports.

Figure 2: Screenshot of the editing tool to place interactive elements on top of a video

The interactive objects can be placed manually just by dragging them to a spot via mouse control. In the settings dialog, the behavior of the model can be configured. It allows defining actions that can be assigned to specific events such as user clicks, video ends, and others. One of the most remarkable features concerns the chaining of actions. It allows every conceivable combination of actions to be executed based on a single event.

- branch video, branch project
- zoom, focus, auto follow object or coordinate
- play, seek, pause the video
- show, hide an object
- sleep

List 1: Excerpt from possible actions that can be applied to an event of an interactive object ("model")

The editor was planned to be flexible, adaptive, and scalable to support the usage across various device types. Flexibility and adaptivity are crucial when planning an editing tool since an editor should be able to realize a variety of different interaction scenarios using an extensible range of interaction objects. Several requirements had to be considered as the following:

- Interactions between objects or viewer and objects
- Movement of objects
- Support of both worlds HTML and WebGL
- Support of a high number of concurrent interactive elements (scales without decreasing performance)
- Simple but usable primitives and building blocks: model-based, event-driven approach with interactive objects as so-called "models" with their events and actions

To meet these requirements, the player consists of an additional layer on top of the video to hold the interactive objects. "Clicking" onto the video means clicking the interactive overlay, which tracks time and position, and reacts accordingly.

Interactions generally refer to the interrelation between two actors or systems. In this context, this is to be understood between the viewer and the objects within the video. However, the editor's architecture also takes into account cases in which objects interact with each other. These implicit actions that are not directly triggered by the viewer are, e.g. cases in which objects move out of the user's field of view or events are triggered at specific times.

Specific times or the duration of a video also play an essential role in another way. While the positions of objects in pictures do not change at any time, this may well apply to videos. As an example, suppose 22 field players of a football game. If you wanted to make each of these players interactively "clickable," the 22 interactive elements would have to follow the movement profile of the respective player.
Therefore, moving objects are also considered by the editing tool.

The basic principle for creating interactivity in 360° videos is to position one or more additional layers above the video, eventually embedding the interactive elements. Every "click" of the viewer into the video is registered in an exact position by the superimposed layers and evaluated to see whether there is a corresponding interactive element at this point in the video or not. If so, the action defined with the interactive element will be executed.

Taking into account, the fundamental principle described here and all the requirements defined, a highly flexible, adaptable and scalable system for the interactive design of 360° video can be realized using a model-based event-driven approach in which the models are represented by the interactive objects and the events basically by the user interactions. Besides, this approach is convincing by the structure and use of simple but usable primitives (small, useable building blocks). The structure follows this according to the three basic components model, event, and action.

The model is the actual interactive element or object which most probably can be seen by the viewer. It is based on web technologies HTML, CSS, and JavaScript. In the simplest case, a model consists of an HTML element such as a text, image, audio, or video. The renaming of an interactive object into a model becomes clearer once looking at the "more advanced" models. These do not consist solely of a single HTML element, but rather consist of a kind of microapp or widget made up of a composition of HTML elements, WebGL objects, a layout description, and an inner logic. It can be a chat, weather, or basic compass app for orientation in 360° videos.

The editor is a web-based UI built using HTML5, CSS3, JavaScript, and besides Desktops, it works fine on tablets as well. It can handle adaptive video formats like HLS or DASH alongside traditional MP4, WebM, and OGG. The tool uses W3C WebGL API to perform 3D, and 360° rendering in browser and XHR calls perform the operations between the frontend and the backend.

B. Players

As shown in Fig 1, a web-based player framework has been developed, providing support for a high number of end devices, such as Desktop PCs, tablets, mobiles and head-mounted displays (HMD) but especially for connected TVs and set-top boxes. In addition to analysis and media research capabilities, the framework offers as an essential feature of the presentation and playback of interactive non-linear 360° content. The ability to switch between 360° media and traditional (non-360°) content is essential leveraging the idea of having video proactively experience, to increase engagement of the viewer and monetization at the producer's side. Interactive 360° non-linear content allows the viewer to display supplemental information through user interactions, such as clicking on an object within the video or to actively influence the plot of the content. The means of creating such immersive interactive experiences are very far-reaching, while some of them are described in Sec III.

To support this kind of experience, the player must be designed to superimpose the actual video with a transparent layer.

Apart from the standard user end devices, the use of the web-based player has already been successfully tested on TV platforms like HbbTV, Chromecast, Amazon Fire TV, and Android TV concerning playback and multi-screen capabilities. Since none of these devices is capable of rendering a 360° video locally due to the limited graphical and computational resources, the cloud-based 360° video playout provided by Fraunhofer FOKUS [11] has been integrated. The player is recently used in production during FIFA world cup 2018 in Russia with more than 170 360° clips for all games [12].

1) Mobile

On a mobile device, the primary input methods are either touch inputs, where a user can navigate in the 360° video by swiping the finger over the display. By tapping on interactive elements the corresponding action can be triggered. Another essential way to maneuver around is by using the gyroscope. The user can tilt or rotate the device to change the viewport where the change of position in space should reflect the change of the viewport directly.

2) TV

Controls on a TV (or devices attached to it like streaming boxes) differ widely from the aforementioned mobile devices. A user cannot tilt or rotate the TV, and most of them do not support touch input. Instead, the user has to rely on the standard remote control. To deliver the best possible user experience, the user navigates via one set of arrow keys and the "Enter"/"OK" button. However, the user has the opportunity to switch between two basic control modes:

1. The user wants to change the perspective in the video by pressing left/right/up/down keys.
2. The user wants to highlight and select the interactive elements which can be placed on top of the video, either sticky or floating like a head-up display.

3) HMD

Head-mounted displays can be seen as extensions of mobile devices. They share the input "method" gyroscope in terms that the user can move his head to change the viewport accordingly and physically correct to achieve an immersive effect. Controlling the interactive elements differs in the way that there is no touch input available since the display is within the head mount and could only be accessed by removing it from the head. Currently, the most common method is "staring" at the desired element for a few seconds to activate it. This method is assisted by a small dot that overlays the screen in the middle of the viewport to indicated the action spot. Some HMDs support an external hardware controller, comparable with a small joystick, which communicates wirelessly with the HMD. It can be used to select certain interactive elements.
C. Platform’s Implementation

The entire platform is deployed in the Amazon Web Services infrastructure, comprising the frontend's application hosting, persistence, media content storage and delivery, user identity management. The platform itself and all its tools are hosted on S3 while DynamoDB serves as a persistence layer, e.g., to store metadata of the content and interactive objects created on top of the videos. The AWS SDK client library is used to perform authenticated operations. Amazon S3 service stores the media content and works seamlessly with Amazon CloudFront as a global content delivery network (CDN) for securely transferring data at high speed. Amazon Cognito Identity handles the user identity and access management; e.g., it creates unique identities for users and authenticates them for secure access to AWS resources like Amazon S3 or DynamoDB.

V. Evaluation

Using the presented tool, an interactive 360° video was created in order to carry out performance tests in comparison to other competing platforms and their web-based and publicly accessible interactive 360° experiences. The continuous tests during the implementation phase were conducted using Google Chrome’s DevTools Performance panel [13]. The approach presented here was up to 30% faster than the other 360° projects. This performance could already be noticed by subjective tests in which in full screen mode at a resolution of 2560x1080 no smooth display could be identified for most 360° projects, whereby the interactive core component of the tool presented here showed proper performance, which has a significant impact on the viewer's QoE.

VI. Conclusion and Future Work

In this paper, a vertical end-to-end solution has been introduced that facilitates the creation of immersive interactive 360° media experiences focusing on storytelling. The integrated editing tool allows content creators to easily create stories from a set of media assets, including flat and 360° videos that are distributed to a broad audience. The current tool provides the essential elements required to create immersive interactive video experiences as well as other relevant elements. To further improve the storytelling and user experience, other narrative stylistic devices as spatial audio or personalized paths in 360° videos based on the viewing behavior of other users are considered in future releases.

REFERENCES