VIC – An Interactive Video System for Dynamic Visualization in Web and Mobile Platforms

Benjamim Fonseca[1,2], Hugo Paredes[1,2], Paulo Martins[1,2] , André Alberto[1], José Rego[1], Leonel Morgado[2,3], and Arnaldo Santos[4]

[1] Universidade de Trás-os-Montes e Alto Douro (UTAD), Vila Real, Portugal
{benjaf,hparedes,pmartins}@utad.pt,
{andrealbertto,zerego.utad}@gmail.com
[2] INESC TEC - INESC Technology and Science, Porto, Portugal
[3] Universidade Aberta, Lisbon, Portugal
leonel.morgado@uab.pt
[4] Portugal Telecom Inovação, Aveiro, Portugal
arnaldo@ptinovacao.pt

Abstract. This paper presents an interactive video system that enables users to change the flow of video playback by interacting with hotspots that were predefined throughout the video streams. These hotspots are synchronized with the underlying video streams and the interactions result in smooth transitions between the preloaded targets. This approach allows the dynamic visualization of content by interacting with the hotspots and producing the consequent changes in the flow of the story. The system includes web-based and mobile video players specifically developed to deal with the interactive features, as well as a configuration tool that allows content managers to choose which pre-produced interaction possibilities will be used for a specific target audience. The interactive video solution presented herein has potential to be used as a powerful communication tool, in commercial, e-learning, accessibility and entertainment contexts.

1 Introduction

Video is often a key element in the communication strategy of many organizations, but it is frequently convenient to have distinct content directed to different target audiences, which in turn tend to consume it differently, depending on several factors such as visualization equipment, time constraints or even the mood.

The way users access and consume information has changed significantly over the last decade, namely with the emergence of social networks. Indeed, Facebook and YouTube provided users with great freedom in choosing which content they want to visualize and when they want to do it. This growing demand for personalized content challenges efficiency in producing content that is able to suit as much people as possible. One possibility to cope with this issue is to put in the same monolithic media file all the information required to satisfy the majority of people, but this will probably lead to cognitive overload or can make it too boring for many users. Alternatively, content

C. Stephanidis and M. Antona (Eds.): UAHCI/HCII 2014, Part II, LNCS 8514, pp. 38–49, 2014.

providers can produce a lot of smaller and distinct media files, but this can make it harder for users to choose the information they want. These two approaches seem to be inadequate to be used isolated, but its combination can provide a suitable solution. Indeed, the combination of all the information in one master stream, from which the user can interactively select the appropriate sub-streams corresponding to the desired content, can allow a great degree of personalization, reaching a large audience, providing multilevel information (e.g., allowing incremental learning). This solution poses challenges related with the technologies required to provide the interactivity features, but also in the narratives, which can be quite different from the traditional ones.

Some technologies have emerged in the last decade, which facilitate the development of interactive applications and the combination of media in increasingly sophisticated multimedia content. However, there are not yet known solutions for the production of interactive video as envisaged above.

In this paper we present an interactive video solution, which main principle is the interaction with hotspots marked in a video, resulting in a change to another video or other content that presents additional information about the underlying subject. This approach, named VIC (Portuguese-language acronym for Interactive Video for Communication), required the development of solutions for coping with the interaction, the media synchronization and the smooth transitions between sub-streams, as well as the players required to visualize this new kind of media in desktop computers, laptops and mobile devices. It was also necessary to specify some video pre-production procedures and to develop a management tool that allows media managers to configure interaction, selecting the appropriate hotspots and actions from the ones specified in the pre-production stage.

This paper is organized as follows: in section 2 we present some previous work related with the solution and technologies used for our approach; section 3 describes the solutions adopted in the VIC project, presenting its requirements, model of interaction and the video production stages; section 4 shows the VIC prototypes, presenting the technologies used for the presentation and interaction with the video streams, platforms and formats supported, screenshots of the web and mobile players, as well as a tool for configuring interaction options for each video and the results obtained with these prototypes; finally, section 5 presents some concluding remarks.

2 Related Work

In the last decades a series of technologies emerged that allow developing and manipulating multimedia products. Some of these technologies, such as MPEG-21 [1], have been researched partially, but there are no commercial products available that explore the whole potential. There are a few players available, such as QuickTime [2] and AXMEDIS [3] that are able to play MPEG-21 files, but access to content and its production is still difficult.

Back in the 1980s, an early approach to interactive video used videotape recorders, audio track tones to control it and light flashes as a feedback mechanism, introducing

a basic interactive experience for learning activities [4]. In 1992, Tani *et al.* [5] presented an interactive video technique that allowed interacting with a graphics layer that interfaces physical objects being displayed in live video, allowing users to remotely control the movement of those objects.

The "Featured Media" prototype [6] implemented the ability to interact with objects in metadata enriched video streams being transmitted through IPTV. The user interacts with the system using the TV remote control and interactive objects are highlighted under pause of the video playing.

Wang *et al.* presented an interactive system that allows extracting foreground objects from a video sequence and its use with another different video [7]. A model for linking and describing interactive TV programs was presented in a work by Goularte *et al.* [8], in which MPEG-7 was used to describe media and to segment objects in a video frame. In [9] the authors present an object tracking system for videos streams, using a sprite layer for implementing interactivity.

3 The VIC Solution

VIC's aim is to enable interaction with video objects without the cumbersome pauses caused by pre-buffering alternative streams, which results in the provision of several playing flows in the same video stream, allowing for content personalization. To achieve this, a technological solution was designed and implemented, as well as an interaction model and application scenarios. The technological solution and the interaction model will be presented further in this section, but regarding the scenarios, VIC was initially targeted to support the following corporate business activities:

- Commercial presentation, presenting functional and commercial features of a product, to be used in corporate websites but also by the product manager at the client premises, as well as by the client to support buying decisions.
- Technical presentation, describing the product's technical characteristics (product datasheet), to be used by the product manager to clarify technical questions at the client premises, and by the client as a basic technical support tool.
- Step-by-step wizard, working as an installation or user manual.

However, VIC's characteristics make it potentially attractive for other scenarios, such as e-learning, advertising, and entertainment.

3.1 Requirements

The VIC solution requires the use of technologies that enable the creation of multimedia content in a format that allows the inclusion of interaction objects (hotspots) and several presentation flows. The interaction with these interactive objects shall result in a change in the flow of visualization, either by showing a different video stream (sub-flow), or by showing additional information, such as text, photos, graphics or web pages. At the end of a sub-flow, playback must return automatically to the point of interruption in the parent stream. It is also convenient to have the possibility

to interrupt sub-flow payback earlier and return to the point of interruption in the parent stream. The interactive video players shall also enable VCR-like commands (play, pause, fast forward, rewind), as well as zooming a region of the video or repositioning additional content windows, using drag-and-drop capabilities provided by mice or touchscreens.

The interactive video content is intended to be visualized in a web page, but also in mobile devices such as smartphones and tablets.

From this set of main characteristics, a number of functional (Table 1), nonfunctional (Table 2) and interface requirements (Table 3) were specified.

Table 1. Functional requirements

ID	Requirement
F01	Change video presentation flow (reproduction of sub-flows) by interacting with hotspots
F02	Control video playback with VCR-like commands
F03	Allow return to normal flow interruption point, after completion of sub-flow
F04	Allow digital zoom in a video region
F05	Show additional information by interacting with hotspots
F06	Follow hyperlink by interacting with hotspots
F07	Allow high level interaction editing. trhough a configuration tool, allowing product manager to select which interactive functionalities to enable
F08	Reposition video and additional information windows
F09	Explicitly change playing video, without interaction with hotspots
F10	Allow the automatic change of the organization of a mobile interface, through orientation change (using device sensors)

Table 2. Nonfunctional requirements

ID	Requirement
N01	To have a data model supporting the functionalities related with the additional information and the editing tool
N02	To use visualization formats suitable for wen and mobile distribution
N03	To use mechanisms of synchronize between visualization flows and sub-flows that allow smooth transitions
N04	To use common editing tools in the early stages of video editing (prior the configuration tool)

Table 3. Interface requirements

ID	Requirement
I01	To have easily identifiable and nonintrusive video interaction regions (hotspots)
I02	To have additional information viewing areas (text, photo, graphics, hyperlinks, auxiliary videos)
I03	To provide VCR-like controls
I04	Configuration tool must provide activation and deactivation buttons for the interactive features
I05	Provide web and mobile interfaces with equivalent interaction features

3.2 Interaction Model

To cope with the scenarios envisaged for VIC, it is desirable to have solutions for diverse computational environments, from desktop and laptop computers, to smartphones and tablets. This diversity requires mainly two types of interaction devices: the mouse, for desktops and laptops, and the touchscreen, for smartphones and tablets. Based on these interaction devices and on the functional requirements identified in Table 1, the interactive functionalities were specified, which can be seen in Table , with its relationship with the corresponding functional requirements shown in Table 5.

Table 4. Interactive functionalities

ID	Requirement
X01	Mouse click
X02	Touch interaction
X03	Mouse double click
X04	Multitouch interaction
X05	Drag-and-drop
X06	Touch with drag-and-drop
X07	Device rotation

3.3 Production Stages

Traditional video production is usually a complex work, requiring careful planning (pre-production), capturing the several video sequences and editing them trough an often long and skilled process of cutting, dubbing, subtitling, adding special effects or exporting to distribution formats (post-production).

Beyond the traditional production chain, interactive video requires the addition of the interactive functionalities, which depend much on a careful definition of all the interactions that can take place. This requires the development of a storyboard that takes into consideration the interaction and the possibility of multiple stories, in a combination of traditional screenplay with interaction design skills.

Table 5. Relation between functional requirements and interactive functionalities

ID	Requirement	Functionality ID
F01	Change video presentation flow (reproduction of sub-flows) by interacting with hotspots	X01, X02
F02	Control video playback with VCR-like commands	X01, X02
F03	Allow return to normal flow interruption point, after completion of sub-flow	X01, X02
F04	Allow digital zoom in a video region	X03, X04
F05	Show additional information by interacting with hotspots	X01, X02
F06	Follow hyperlink by interacting with hotspots	X01, X02
F07	Allow high level interaction editing. trhough a configuration tool, allowing product manager to select which interactive functionalities to enable	X01, X02
F08	Reposition video and additional information windows	X05, X06
F09	Explicitly change playing video, without interaction with hotspots	X01, X02
F10	Allow the automatic change of the organization of a mobile interface, through orientation change (using device sensors)	X07

Taking these new issues into consideration, after the prior definition of the interactive storyboard, the interactive video production chain can be divided in three stages:

- Stage 1: basic editing of digital content – this corresponds to the traditional video editing, resulting in the video sequences that can be manipulated and combined to form the interactive video.
- Stage 2: interaction editing – definition of all the possible hotspots, without making it correspond to any specific action. The result is an XML file that contains a list of editing decisions that identify potential hotspots, as well as synchronization information. The definition of the hotspots requires that the corresponding region can be delimited spatially but also in time. This is achieved using a tracking mechanism of well-known commercial editing software.
- Stage 3: interaction configuration – from the XML file produced in stage 2, the user can enable or disable any of the hotspots, define its shape and specify its targets.

Stages 1 and 2 are carried out by video editors while stage 3 is performed typically by product managers (at least in the corporate scenarios considered).

4 VIC Prototypes

The project which originated VIC has produced 3 completely functional prototypes:

- A web player – web application that allows visualization and interaction with the interactive video, requiring only a common web browser.

- A mobile player – mobile application developed for the Windows 8 platform, that allows visualization and interaction with the interactive video.
- An interaction configuration tool – web application that allows configuring the interactive video from the possibilities created during stage 2 of the production chain.

Interaction in the VIC players is achieved through the use of interaction layers that are overlapped to the video stream and synchronized with it. This allows to define the hotspots in this overlay layer and trigger the corresponding actions. The hotspot is thus a mask that defines the interactive area while keeping the video visible to the user.

All the content to be presented is formatted with HTML5 and CSS3 is used to implement the interaction layer, both technologies widely supported by current browsers. The choice of the CSS3 technology was also due to the ability to animate the interaction objects, which enabled us to have dynamic hotspots that track video objects.

The prototype of the web player is compatible with all common browsers, operating systems and video formats. A similar situation occurs with the mobile player but the prototype was built only for the Windows 8 platform, due to organizational constraints and for the sake of interface homogeneity.

In the case of the web player, the video content is delivered via streaming, using double buffering to achieve the N03 nonfunctional requirement (quick transition between flows, upon hotspot interaction). However, for the mobile player video can be delivered with the application installation, through app update or in the first time the video is requested, keeping it in the device's memory and thus achieving N03 and keeping synchronization of video with other content.

4.1 Web Player

When a user selects a video from the list the player starts its visualization. Fig. 1 shows a message that is shown at this time, alerting for the interactive possibilities, while Fig. 2 shows the reproduction of an interactive video that has an active hotspot (double dot on the upper left corner) and Fig. 3 shows the result of interactions with hotpots.

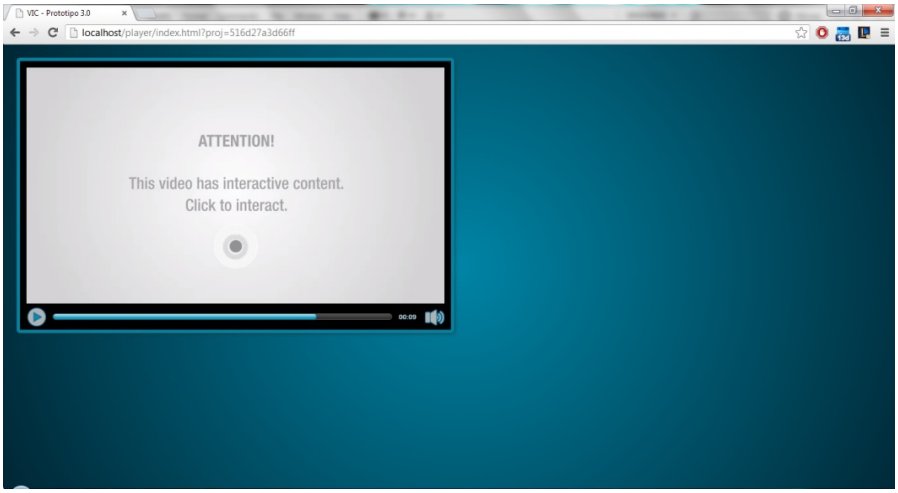

Fig. 1. Web player – starting visualization of an interactive video on a web page

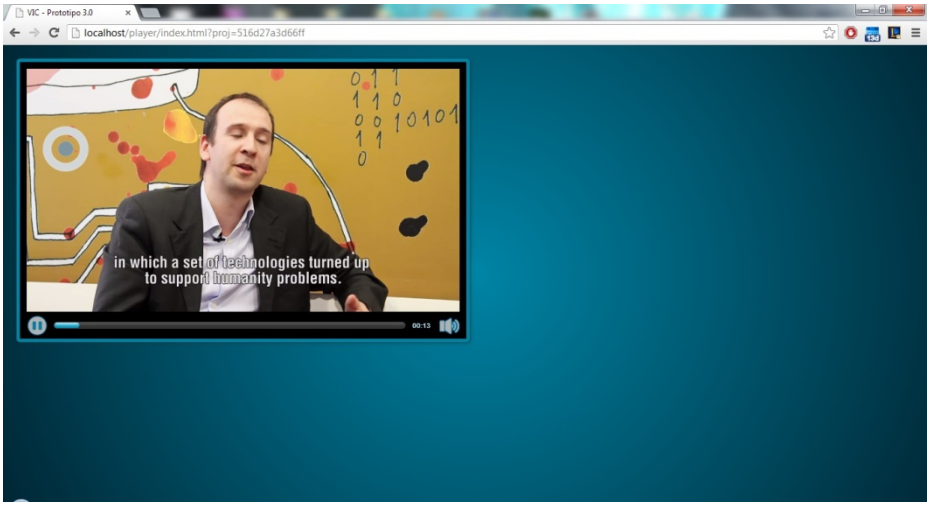

Fig. 2. Web player –an active hotspot (concentric circles on the left) visible during video play

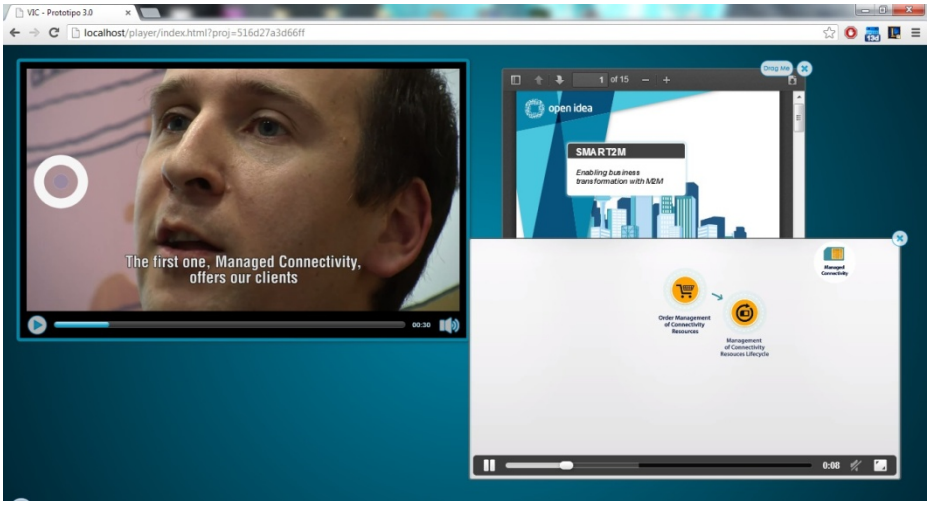

Fig. 3. Web player – showing additional content after interaction with hotspots

4.2 Mobile Player

The mobile player has the same capabilities as the web player, being the main difference the interaction with the touchscreen – for the web player the mouse is used for interacting with the hotspots). Fig. 4 shows the interactive video selection panel that is shown in the mobile player prototype, in this case running in a Windows 8 tablet. Fig. 5 shows an example of text information shown upon interaction with a hotspot.

Fig. 4. Mobile player – interactive videos selection panel

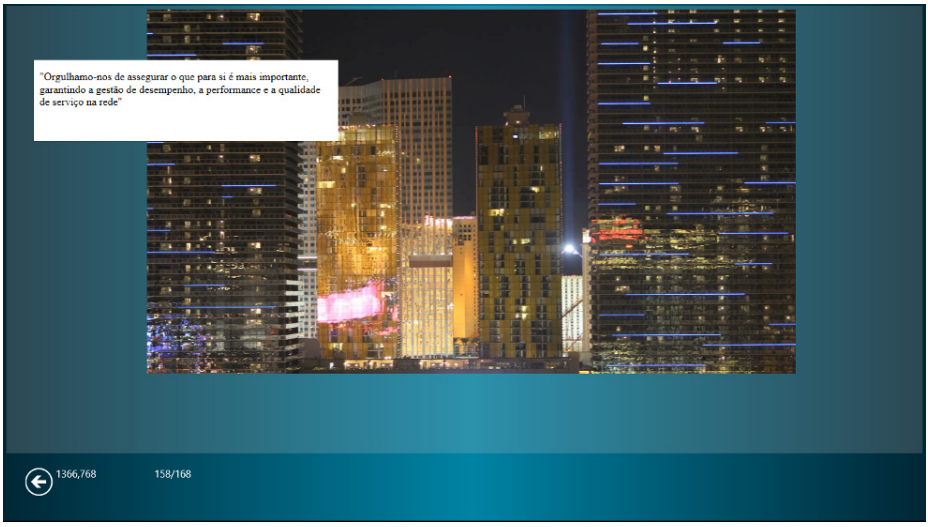

Fig. 5. Mobile player – text box shown upon interaction with hotspot

4.3 Interaction Configuration Tool

The configuration tool is a web application that allows product managers to build the interactive video to be shown to the audience, choosing the hotspots and resulting actions from a list of possible ones, built during stage 2 of the production chain, and according to a storyboard where interaction is crucial. Fig. 6 shows an example of the page that allows configuring several properties of the hotspot, such as shape, size, position, type of content that is shown upon interaction or the “Draggable” property.

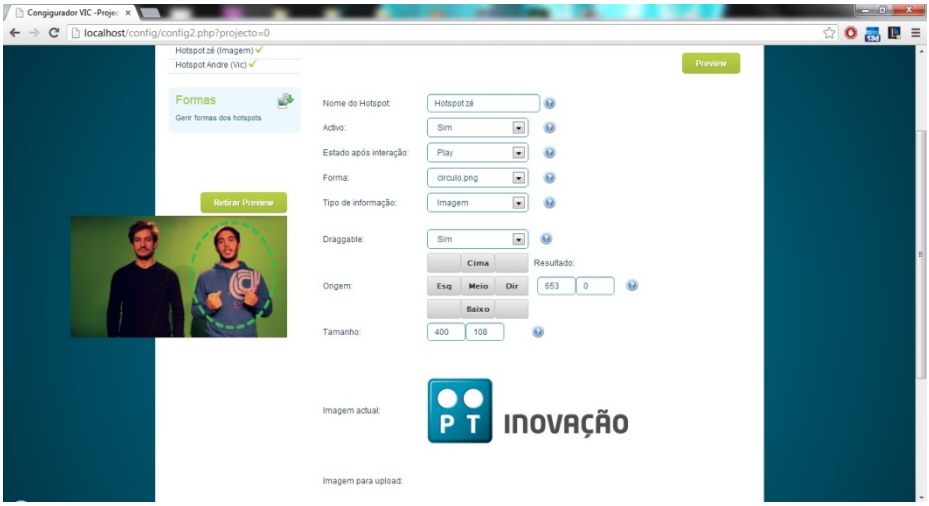

Fig. 6. Configuring the hotspot

4.4 Results

The implementation of the prototypes fulfilled all the requirements defined for the VIC project. However, a few problems were found and considered for further developments. For the web player, the following limitations were identified:

- Synchronization problems occur when the visualization flow is changed manually (for example, clicking in the time bar to return to a specific moment).
- Browser processing capabilities affect synchronization and response to events.
- Bandwidth, browser processing capabilities and target video duration affect transition between videos.
- Necessity to have the videos available in 3 formats (Ogg, MPEG4 and Webm) because browsers do not support all the same formats.

The mobile player has an additional set of problems that are essentially due to the option for the platform and also to usual performance limitations of mobile devices:

- There's no possibility to play 2 videos simultaneously.
- Transitions between videos are slower than in the web version and sometimes the delay is sensible.
- The visualization of a PDF file requires an external application which implies leaving the player application.

5 Concluding Remarks

The development of VIC has shown the viability of the interactive video concept and allowed envisaging its potential as a powerful communication tool, in commercial

contexts (product information, datasheets, demonstrations), e-learning (adaptation of learning flows according to learners' profiles and needs, including accessibility options), and entertainment (personalization of narratives). Future evolutions of the VIC concept can incorporate other technologies such as augmented reality and motion detection to achieve higher levels of interaction and immersion.

Developing a product for different platforms requires overcoming challenges such as the need for guarantying similar user experiences, despite technological diversity. The functionalities, requirements and technologies adopted in the VIC project were all successfully implemented and a few limitations were identified, most of them predictably overcome by the evolution of the current and future mobile and web technologies.

The production model for the interactive video is based in a combination of well-known software and techniques with a simple configuration tool developed for the specific purpose of the VIC project.

Despite the technological challenges that were identified, the biggest challenge in adopting interactive video as a valuable communication tool is the care required in the definition of adequate narratives, incorporating interactive characteristics that are not familiar to storyboarding professionals. For interactive video production, video capture and basic editing have to be carefully planned and executed in order to achieve a coherent set of stories and not just a unordered sequence of information, which could avoid users from keeping a notion of the sequence or main narrative. It is also important that interaction points are envisaged and several aspects like its duration planned from the very beginning, because it can impact video production and direction.

Acknowledgements. This work is funded by Portugal Telecom Inovação under the Innovation Plan of the PT Group, and by FEDER funds through the Programa Operacional Fatores de Competitividade – COMPETE and National Funds through FCT - Foundation for Science and Technology under the project «FCOMP - 01-0124-FEDER-022701».

References

1. The Moving Picture Experts Group, http://mpeg.chiariglione.org/standards/mpeg-21 (accessed in February 20, 2014)
2. Apple Inc., http://www.apple.com/quicktime/what-is/ (accessed in February 20, 2014)
3. AXMEDIS, http://www.axmedis.org/com/ (accessed in February 20, 2014)
4. Emmens, C.A.: slj/Video Watch. SLJ School Journal, 40–41 (February 1981)
5. Tani, M., Yamaashi, K., Tanikoshi, K., Futukawa, M., Tanifuji, S.: Object-oriented video: interaction with real-world objects through live video. In: CHI 1992 Proceedings, pp. 593–598. ACM (1992)
6. Rack, C., Seeliger, R., Arbanowski, S.: Featured Media Nonlinear, clickable multimedia content for an interactive video experience. In: Proceedings of the 2010 Second International Conferences in Advances in Multimedia, pp. 39–43. IEEE Computer Society (2010)

7. Wang, J., Bhat, P., Colburn, R.A., Agrawala, M., Cohen, M.F.: Interactive Video Cutout. In: ACM Transactions on Graphics (TOG) – Proceedings of ACM SIGGRAPH 2005, vol. 24(3), pp. 585–594. ACM (2005)
8. Goularte, R., Moreira, E.S., Pimentel, M.G.: Structuring Interactive TV Documents. In: Proceedings of DocEng 2003, pp. 42–51. ACM (2003)
9. Smith, J.C., Stotts, D.: An extensible Object Tracking Architecture for Hyperlinking in Real-time and Stored Video Streams. In: Technical Report TR02-017. Department of Computer Science, University of North Carolina at Chapel Hill (2002)