Reference Model for Hybrid Broadcast Web 3D TV

Igor G. Olaizola,* Josu Pérez, Mikel Zorrilla, Ángel Martín and Maider Laka
Vicomtech

Figure 1: *Vicomtech Digital Television Lab.*

Abstract

3DTV can be considered as the biggest technical revolution in TV content creation since the black and white to color transition. However, the big commercial success of current TV market has been produced around the Smart TV concept. Smart TVs connect the TV set to the web and introduce the main home multimedia display in the *app* world, social networks and content related interactive services. Now, this digital convergence can become the driver for boosting the success of 3DTV industry. In fact, the integration of stereoscopic TV production and Web3D seems to be the next natural step of the hybrid broadband-broadcast services.

We propose in this paper a general reference model to allow the convergence of 3DTV and 3D Web by defining a general architecture and some extensions of current Smart TV concepts as well as the related standards.

CR Categories: I.3.3 [Computer Graphics]: Three-Dimensional Graphics and Realism—Display Algorithms I.3.8 [Computer Graphics]: Applications—;

Keywords: 3D Web, Smart TVs, 3DTV Broadcast, Hybrid Broadband-broadcast

1 Introduction

The digital convergence has fostered the interoperability and seamless integration of previously isolated technologies and markets such as photography, music, telecommunications etc. However, the TV and Cinema industry have digitalized their processes and technologies in a kind of parallel way where professional solutions did not take into account the mass market oriented technological solutions. As an example of this, interactive TV technologies such as MHP/OCAP[Andreadis et al. 2007] were basically incompatible with the state of the art of web technologies allowing access to the return channel but with extremely high limitations for web interaction.

MPEG-4[MPEG 1999] has been the main standard for professional video transmission since MPEG-2. MPEG-4 includes specific parts for 3D encoding such as BIFS (Binary Format for Scene, based on VRML[W3C 1995]), AFX (Animation Framework eXtension)[Morán 2008], 3D Graphics Compression Model, 3D Graphics Conformance, 3D Mesh Progressive Streaming (3DMC) or MVC (Multiview Video Coding, an amendment to H.264/AVC). However, the success of MPEG-4 for A/V encoding has not been followed by a generalized industrial adoption of its 3D specifications. Sometimes considered as a too broad standard, and probably due to the fact that the 3D production pipeline is more suitable for web technologies than for multimedia compression and transmission standards, there is a lack of implementations and resources for MPEG-4 based 3D solutions[1].

In the same way, 3DTV commercial solutions are still far away of being compatible with the Web3D reality and basically act as displays that create stereoscopic images from left/right frames.

On the other hand, the huge commercial success of Smart TVs is creating *apps* platforms and web technologies that enable the full integration of TV sets into the web media ecosystem together with tablets and mobile phones. According to IHS Screen Digest during 2012 over a quarter of all sets sold (66 million) were smart TVs, and they predict that by 2015 it will be half of all sets sold, or around 141 million[2].

As a response to this great expected growth of the Smart TV market, the technology offer is plenty of commercial solutions where we can distinguish the following:

- Propietary solutions offered by HW manufacturers:
  - TV manufacturers: Samsung, LG, Sony, Panasonic, etc.

*e-mail:{iolaizola, jperez, mzorrilla, amartin, mlaka}@vicomtech.org

[1]http://general3d.com/newsletter/newsflash/3_31_11.html

[2]http://www.worldtvpc.com/blog/smart-tv-sales-surging-as-3d-gets-the-brush-off/

Web3D 2013, June 20 – 22, 2013, San Sebastian, Spain.

ACM 978-1-4503-2133-4/13/06 $15.00

- Game console manufacturers: PlayStation, Xbox, etc.
- Set-top boxes: Roku, Apple TV, etc.

- Web based transversal approaches:Yahoo! Connected TV
- Middleware like solutions: Android, Smart TV Alliance.
- Open Standards: HbbTV[HBBTV+ETSI]

Within this extremely fragmented technology offer, there are very few approaches that barely address how to deal with 3D or either stereoscopic content. LG[3] 3DTV sets, have some specific applications that exploit the stereoscopic capabilities of the TV set (Figure 2) while Samsung offers a *Explore 3D* stereoscopic service for 3D VoD.

Figure 2: *General overview of a hybrid broadband-broadcast stereoscopy content pipeline. 3D content is rendered by the device's local resources.*

In this context and until a real convergence of the aforementioned burgeon technologies will become into reality, we propose a reference model for hybrid broadband-broadcast 3DTV that takes the advantage of current existing 3D Web standard technologies and the most successful interactive TV specfication.

2 Related Work

Stereoscopic TV broadcasting has not required new standard specifications. While, the production chain still faces big scientific and technological problems, stereoscopic content delivery has been addressed in a quite straight forward manner. The same DVB or ATSC standards allow the stereo content flow by using MPEG-2 or MPEG-4 standards[Chiariglione 2012], typically organizing left and right frames sequentially, side by side or in a top and bottom format. For more advanced and disruptive features like depth information, *multiview video coding* or *free-viewpoint TV*[Tanimoto et al. 2012] new requirements that difficult the backwards compatibility must be introduced.

TV interactive middleware solutions (OpenTV, MediaHighway, MHEG, MHP, OCAP, HbbTV) do not include specific resources for 3D rendering and interaction and furthermore, the low computational power of TV sets and set-top boxes that include such technologies don't allow real-time 3D applications[Ugarte et al. 2007].

In order to overcome these limitations, *ad-hoc* solution that avoid broadcast side regulatory constraints have to be carried out[Ugarte et al.]. Such approaches have demonstrated the potential of TV as the main home device for 3D content consumption.

[3]http://www.lg.com/uk/press-release/lg-announces-global-launch-of-3d-world-next-generation-premium-3d-content-service

For the Japanese/Brasilian equivalent of DVB-HbbTV (ISDB-NCL), there is a proposal to integrate 3D content in the specification[Azevedo and Soares 2012]. However, the combination of broadcasting signal and IP services is not addressed in this work.

3 Reference Model

In order to define a realistic technological and market approach, our proposed reference model is based on extensions upon current existing standard technologies. We propose two scenarios where TV set requirements are very similar to those offered by existing commercial Smart TVs.

The first scenario, is designed for TV sets which include 3D rendering resources and where the *apps* that are executed on the device have access both to web and broadcast content. The second scenario is less restrictive and only needs of capability to display the remotely rendered 3D content overlaid on top of the broadcasting signal. Both scenarios are deployed on top of a set of open standards bridging the convergence of broadcasting and web technologies while preserving the market open for the different agents involved in this new framework (Broadcasters, web content providers, network operators, content producers, advertisers, etc.) where the classic lineal value chain evolves to a totally new and complex "ecosystem". These standards allow the technological convergence around HTML-5:

CSS-3+JavaScript: foundations for multi-device interactive service provision.

SMIL[W3C 2008]: standard markup language for multimedia content description

X3D + WebGL: resources for 3D content description and rendering

DVB-HbbTV: Signaling, content provision and interactivity specifications from the broadcast channel. The trend of HbbTV towards HTML-5 will be one of the key factors to enable a real hybrid broadcast broadband 3DTV

WebSockets[Fette and Melnikov 2011] full-duplex communication protocol for low-latency interaction

3.1 Architecture for local rendering scenario

Within this scenario we establish the following requirements for 3DTV TV sets or set-top boxes (STBs):

- HTML-5 web browser compatible X3D and WebGL and WebSockets
- DVB tuner and MPEG-4 parser/decoder
- HbbTV interpreter
- SMIL Multimedia player
- A runtime environment with simultaneous access to the previous three components

The proposed system architecture for this scenario includes two main parts as it can be observed in Figure 3.

The infrastructure has three main actors. The first one, at the left, deals with the broadcasting production including a stereo camera system (camera rig) and a typical digital work-flow where the content is properly edited, encoded and delivered (the delivery process will depend on the specific broadcast channel, terrestrial, cable, satellite or even IP). The second one, at the right, is the main

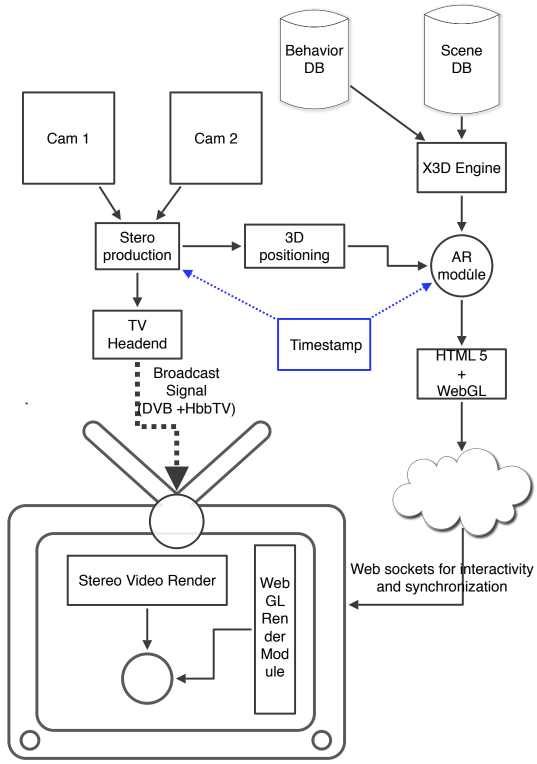

Figure 3: *General overview of a hybrid broadband-broadcast stereoscopy content pipeline. 3D content is rendered by the device's local resources.*

role of the proposed system, the 3D service back-end, addresses the 3D context to provide the 3D objects and the features to display the final view. Last but not least, at the bottom, the TV device receives all the signals, composes the result and generates the interaction events. Due to the multi-path nature of the signals involved, synchronization rise as a major concern. To get a frame level synchronization accuracy, timestamps are generated in this audiovisual production infrastructure and sent to the 3D service back-end. It includes a pose estimation module for real-time 3D positioning of 3D objects in the real scene. This information will be employed to estimate the corresponding homographies and locate virtual objects within the real scene. The virtual elements are described in X3D and sent to the TV set with the corresponding timestamps.

In order to deal with this envisaged scenario where complex models are rendered on local devices, it is needed that TV sets or STBs include GPUs and Web browsers with WebGL compatibility (e.g.: through X3DOM[Behr et al. 2010]).

Once 3D data and broadcast audiovisual content are received by the TV set, the runtime environment has to able to access both data sources, synchronize the flows and present them together in the same image. HbbTV extensions will be necessary for that purpose where the compatibility with SMIL W3C's standard [W3C 2008] would provide a standard way for audiovisual description.

Even if the proposed solution only implies the integration of currently available technologies and standards, it is not realistic to foresee commercial mass market devices with the described features. Therefore, we propose a second scenario with lower end device requirements widen the potential audience.

3.2 Architecture for remote rendering scenario

This second scenario shares the underlying idea keeping the general schedule, but it is modified by moving the 3D rendering process from the end device to the cloud pushing the 3D rendering responsibility to the 3D service back-end. As it is shown in Figure 4, the broadcasting stuff is maintained as it has been described in the previous section. However, the 3D information is not sent to the TV set but rendered and sent via video streaming from the server side. The video content sent via streaming includes an *alpha* channel that will be used by the HTML-5 browser to remove the parts of the rendered video where the broadcast content should appear.

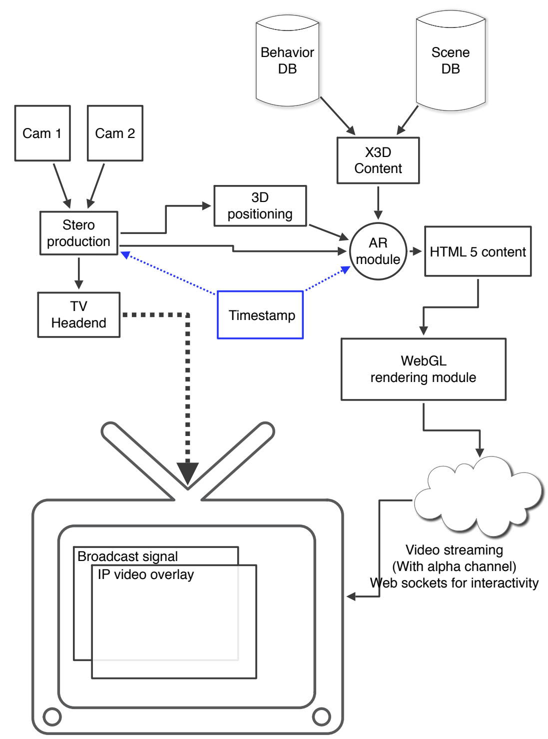

Figure 4: *General overview of a hybrid broadband-broadcast stereoscopy content pipeline. 3D content is rendered remotely and delivered through streaming.*

Since there is not a standard specification for alpha channel in video for HTML-5, the most suitable way is to send an extra video that will be used as a mask. Figures 5 and 6 show an example of alpha channel on videos embedded in HTML-5.

Once the broadcast content and streaming videos are received by the TV set, both are mixed by applying the mask. Timestamps ensure the synchronization of both data sources. However, for network conditions where the video streaming may have low QoS parameters, special policies will be needed (e.g.: remove virtual objects if QoS is under a certain threshold).

For user interaction, we propose the use of WebSockets that will allow the synchronization of events from the client to the server side.

In this scenario, TV sets do not have to integrate new standards as in the previous case, the only requirement is a runtime environment where both streams can be synchronized and the capability of create an overlay with videos as it is currently done by common HTML-5 web browsers.

Figure 5: *Video with mask sent via streaming.*

Figure 6: *Final render performed by an HTML-5 browser (Chrome) where the mask is applied and a background is added in real time.*

4 Conclusions

We have presented two possible solutions to allow the convergence of the 3D Web and 3DTV. The high degree of isolation of these two fields makes that current standards do not consider fundamental aspects for a seamless integration of 3D Web technologies and 3DTV. However, instead of creating new technologies we propose the integration of existing ones and extensions of most relevant standards like HbbTV and HTML-5.

The presented two scenarios differ in the point where the 3D rendering takes place. The first scenario (local rendering) has much higher requirements for TV manufactures that should include more capabilities in the browser that their devices embed. Furthermore, hardware requirements would also be higher where the integration of GPUs could have a strong impact in the final price of the devices. This fact does not seem very realistic in a short term unless 3D services become real killer application in the TV mass market. The evolution of TV set or STBs towards advanced 3D capabilities would also enter in the game console arena increasing the number of market uncertainties.

The second scenario relies on remote rendering and keeps the TV set much thinner in terms of hardware requirements and browser APIs. However, content synchronization turns critical due the higher bandwidth needs. User interaction experience could drop by latency brought by remotely processed user requests. Moreover, it is crucial to keep in mind that TV is a really massive service where millions of users access simultaneously to the same content. Thus, this approach could be unaffordable for massive service provision.

From the market point of view, it is worth to take into account that there is still not a clear market dominant technology and that big players are offering vertical solutions that in the end could limit the success of open standards. However, the trend to HTML-5 of all platforms and applications is right now the most likely option.

It is expected that the evolution of HbbTV from its current 1.5 version to 2.0 will include a much higher interoperability with HTML-5. However, 3D rendering capabilities are still out of the scope of HbbTV that has other priorities such as the interoperability with other devices and the market share where the standard competes with a bunch of proprietary technologies controlled by big market players. On the other hand, in the same way that HTML-5 apps are extended to access the local resources of mobiles devices (GPS location, camera, compass, accelerometer, etc.), specific extensions would be needed for the TV environment (specially in order to access broadcast content). The liaison of the different standardization groups (HbbTV,W3C,MPEG, etc.) and the technological convergence will then enable the biggest home display for full 3D experience.

References

ANDREADIS, A., BALDO, D., BENELLI, G., AND DAINO, G. L. 2007. Towards itv applications portability across digital terrestrial television frameworks. In *Software, Telecommunications and Computer Networks, 2007. SoftCOM 2007. 15th International Conference on*, IEEE, 1–4.

AZEVEDO, R. G. D. A., AND SOARES, L. F. G. 2012. Embedding 3d objects into ncl multimedia presentations. In *Proceedings of the 17th International Conference on 3D Web Technology*, ACM, 143–151.

BEHR, J., JUNG, Y., KEIL, J., DREVENSEK, T., ZOELLNER, M., ESCHLER, P., AND FELLNER, D. 2010. A scalable architecture for the html5/x3d integration model x3dom. In *Proceedings of the 15th International Conference on Web 3D Technology*, ACM, New York, NY, USA, Web3D '10, 185–194.

CHIARIGLIONE, L. 2012. Multimedia standards: Interfaces to innovation. 893–904.

FETTE, I., AND MELNIKOV, A. 2011. The websocket protocol.

HBBTV+ETSI. Technical specification 102 796,". *Hybrid Broadcast Broadband TV 1*, 2010–06.

MORÁN, F. 2008. Animation framework extension (afx). Tech. rep., Universidad Politécnica de Madrid.

MPEG, 1999. ISO/IEC 14496 MPEG-4.

TANIMOTO, M., TEHRANI, M. P., FUJII, T., AND YENDO, T. 2012. Ftv for 3-D spatial communication. 905–917.

UGARTE, A., FLÓREZ, J., OLAIZOLA, I. G., AND OYARZUN, D. Discot: A platform to approach information society to everyman users.

UGARTE, A., GARCÍA, I., ORTIZ, A., AND OYARZUN, D. 2007. User interfaces based on 3d avatars for interactive television. *Interactive TV: a Shared Experience*, 107–115.

W3C, 1995. VRML virtual reality modeling language.

W3C. 2008. SMIL. *Synchronized Multimedia Integration Language*.