MUSCADE – Multimedia Scalable 3D for Europe
Guillaume Berenger¹, Greet Verelst², Ariane Isoard³
¹,²,³EADS Astrium, Toulouse, France
E-mail: ¹guillaume.berenger, ²greet.verelst, ³ariane.isoard@astrium.eads.net

Abstract: MUSCADE (Multimedia Scalable 3D for Europe) is a European project, funded under the European Commission ICT 7th Framework Programme. MUSCADE aims at generating major innovations in the fields of 3DTV production equipment and tools, data representation, compression, transmission and rendering on various kinds of 3D displays. This paper provides an overview of the MUSCADE system architecture and the first technical choices made by the consortium. The final objective of MUSCADE is to demonstrate a complete multiview 3DTV live chain over wireline, wireless and satellite networks.

Keywords: MUSCADE, 3D TV, multiview video, scalability, quality of experience

1 INTRODUCTION

The last decade has seen a revolution in the distribution of motion content: from analogue to digital and then from SDTV to HDTV. Both the delivery and the content creation industries exploited these advances, for the final benefit of consumers. Today, it is widely accepted that the next step is the evolution from HDTV to 3DTV, as indicated by the new industrial fora that appeared over the last years.

Combining strengths of twelve European partners, the objectives of the MUSCADE project cover the whole 3DTV chain. The project will define, develop, validate and evaluate the technological innovations in 3DTV capturing, data representation, compression, transmission and rendering required for a technically efficient and commercially successful 3DTV broadcast system. The MUSCADE reference system architecture is shown in Figure 1.

2 VIDEO REPRESENTATION FORMAT AND SCALABLE CODING

First standardization on 3D video beyond standard stereo has been started in 2003, including video-plus-depth coding for 3D-TV in MPEG-C Part III. Multi-view coding (MPEG-MVC) was started in Jan 2006 and has been finalized in July 2008 [1]. A new initiative has been launched in July 2008 through the MPEG-3DV Adhoc group [2]. None of the proposed solutions today are scalable in terms of network performances and display capabilities.

The system developed by the MUSCADE project will target a large range of displays: 2D, stereoscopic, multiview autostereoscopic and advanced lightfield displays. In the typical use case, one display is driven at a time. However, this doesn’t mean that a specific video representation format will be generated according to the targeted display. Instead of that, the concept of view scalability will be part of the representation format of the 3D video.

MUSCADE has selected a MVD (Multiview Video plus Depth) based format: MVD4. It includes 4 HD views and 4 disparity maps, providing per-pixel depth information, as illustrated in Figure 2.

![Figure 2: MVD4 video representation format](image)

This format ensures efficient support of multiview autostereoscopic displays, as well as backward compatibility with stereoscopic and 2D displays. Indeed, a stereoscopic pair or a single HD view can easily be extracted by low complexity devices. Moreover, this format is compatible with the scalable transmission of 3D-TV, i.e. the adaptation of 3D content to propagation conditions, targeted by the MUSCADE project.

3 VIDEO CAPTURE AND POST PRODUCTION TECHNOLOGIES

To avoid undesired perception conflicts and well-known discomfort like eyestrain or visual fatigue, conventional

Corresponding author: Guillaume Berenger, EADS Astrium, Toulouse, France, guillaume.berenger@astrium.eads.net
stereo production as well as any other 3D production requires both, a careful set-up of the multiple camera views during shooting and an appropriate post-production process.

MUSCADE is developing an audio-visual acquisition system being able to capture two or more HD views (in 1080p25 in the first phase of the project) in conjunction with 3D sound. In contrast to the existing solutions, it will concentrate on the special requirements of 3D broadcast productions with its restricted post-production capabilities and its special needs for live broadcast. It includes an intelligent on-site assistance system for multiview and 3D sound acquisition providing all metadata needed for post-production and live broadcast in real-time.

The capture system of MUSCADE is based on a 4-camera set-up as illustrated in figure 3.

This camera set-up consists of an inner stereo pair that is compatible with conventional stereo production and enables backwards compatibility of the whole MUSCADE chain to standard stereo viewing. An outer wide-baseline stereo system provides all additional information for robust depth extraction and depth-image based rendering, allowing the support of a wide range of 3D displays. A special geometrical property of this 4-camera system is that it is calibrated such that all focal points are at one common straight line and that it follows a parallel stereo geometry. The related calibration, rigging and geometrical adjustment will be ensured by an image-based camera assistance system. The 4-camera system will be combined with a set of microphone arrays and individual microphones for recording separated sound sources and ambient acoustics. 3D video and audio acquisition will be properly synchronized and jointly calibrated. Calibration data will be sent as metadata in conjunction with rigging data and 3D stereo parameters from the camera assistance system.

This set of data (multiple synchronized audio and video streams plus metadata) forms the input for the MUSCADE 3D post-production system. Metadata from the camera assistance system are first used for initial colour matching, geometric correction and rectification. Subsequent operator-assisted control processes supervise these initial settings and guide further manual corrections if necessary. Then the corrected images are used as input for the extraction of depth information and the semi-automatic generation of reliable and depth-consistent disparity maps for each image. As a result, this process will provide the MVD4 (Multiple Video-Plus-Depth with 4 Views) format consisting of the four post-produced video streams and the four generated depth streams. After generating the MVD sequences, the 3D video sequences are edited together with 3D sound. Apart from 3D sound editing and dubbing, this includes cutting, fading and blending of 3D sequences, mixing of different 3D sources like MVD4 and CGI data, graphical overlay and subtitle editing. The post-production process will be completed with final finishing and packaging of all data in a format that is suitable for coding and delivery.

4 ROBUST AND SCALABLE TRANSMISSION OF 3DTV

After postproduction, the processed audio-visual 3D data is compressed into a suitable transmission format and encapsulated into appropriate transport streams. The packet streams are then fed into the transmission channels investigated in the project i.e. wireline (ADSL, ADSL2+, VDSL), wireless (DVB-T(2), WLAN, WiMAX) and satellite (DVB-S2) using both emulated and actual links.

The audio encoding algorithm is based on MPEG Surround and also supports Spatial Audio Object Coding (SAOC) by down mixing 8 audio objects (or object groups) maximum and producing related side information. AAC core codec is to be used for compressing downmixed signals.

The video encoding algorithm selected by MUSCADE is based on MVC, which is a natural solution for the coding of multiview videos. A first approach is to encode the views and the depth using two independent MVC encoders.

The carriage of MVC has been specified in an amendment of ITU-T Rec. H.222.0 | ISO/IEC 13818-1 [3]. This
amendment enables the splitting of the MVC bit streams into multiple elementary streams (ES) with unique PIDs, each of them corresponding to individual views, and thus allowing for de-multiplexing on Transport Stream (TS) level, which is the pre-requisite for selective access to a certain set of views and unequal error protection mechanisms. Figure 5 illustrates the separation of MVC data into elementary streams.

Since 3D-TV demand more system resources (e.g. transmission power, bandwidth) than conventional 2D video, optimum resource allocation schemes have to be developed in order to match the available resources and provide good quality 3D perception. The combined use of MVC coding and MVD4 is an efficient way to implement the scalable transmission of 3D-TV.

Scalability refers to the capability of dropping data fragments of a video bit-stream, which results when applied correctly in a new video bit-stream decode-able at lower quality (in terms of number of views in MUSCADE) at reduced data rate and reduced processing demands on the decoding platform. A multi-view scalable bit-stream consists of a base layer (base view) and one or more enhancement layers (enhancement views), which enhances the decode-able 3D quality of the bit-stream.

The MUSCADE consortium defined 3 scalability levels, as shown in table 1.

| Table 1: MUSCADE scalability scheme |
|-----------------------------|------------------|
| Stereoscopic pair | Preliminary bit rate estimation (1080p30, 8-bit depth encoding) |
|                        | 13 Mbps |
| ∆MVD2                  | 7 Mbps  |
| ∆MVD4                  | 17 Mbps |

The first layer consists in a stereoscopic pair. The second layer consists in two depth maps, providing depth information for the layer 1 stereoscopic pair. The third layer includes the two outer views and the associated depth maps. This choice has been lead by the nature of the 3D displays targeted by MUSCADE and the capabilities of the networks addressed in MUSCADE.

A typical use case of the scalable transmission of a multi-layered 3D video over a satellite link using DVB-S2 is shown in Figure 6. The different quality layers are broadcasted in VCM (Variable Coding and Modulation) mode with different MODCOD (Modulation and Coding) protection, enabling to make the 3D video displayable on every display with the best possible quality and depending on transmission conditions (clear sky or rainy conditions).

The mapping between the different layers and MODCOD is performed according to the quality layers to be transmitted: the base layer (stereo) is transmitted with a robust MODCOD and the additional quality layers (∆MVD2, ∆MVD4) are transmitted with a more spectral efficient MODCOD. Such technique allows reducing the required satellite frequency resources at the expense of reduced link availability for high video quality layers (which can be counteracted by installing a larger antenna at the receiver side).

5 DISPLAY AGNOSTIC VIDEO RENDERING

Various kinds of 3D ready displays for stereo are now available. LCDs with micro-polarisers with passive glasses and time sequential techniques with active glasses use two video streams for the two eyes.

Currently, several 3D “frame compatible” formats are proposed. However, these formats are not compatible with multi-view and light-field displays that are emerging. The display and technology dependent video formats make it very hard to transfer live 3D content from one 3D display to another one without losing serious amounts of information. Therefore, MUSCADE selected a video plus depth based format that is convenient for large range of displays. As explained in previous sections, the advantage of this type of format is the possibility to render the content from stereo (with glasses) display, to autostereoscopic and light field displays (without glasses). MUSCADE will implement a broadcast chain to
demonstrate the display agnostic property of this format in real-time.
For that purpose, MUSCADE is developing innovative rendering algorithms able of adapting the incoming 3D contents to any 3D display. A challenge consists in adapting 3D content to multiview displays and light-field displays, improving the received content by in-between view interpolation.
Another innovation of MUSCADE consists in providing to the user the possibility of adjusting the baseline in order to tune the depth perception and to reduce the eventual eye-strain.

Figures 7 and 8 present the overall diagram of the A/V stream player and the interactivity modules (redirecting the user to a 3D video interactive application platform, also developed in the frame of MUSCADE) connected to the various display technologies that are going to be addressed.

The rendering module will be a PC based player ensuring the following functions:
- Network adaptation
- A/V decoding
- A/V synchronization
- A/V rendering
- A/V formatting for the various displays and loudspeakers

During the first phase of MUSCADE, real time rendering will only be performed for the 2D and 3D stereo cases. Taking into account the complexity of the light field display, a specific interface based on a Gigabit Ethernet connection will be set up to connect the player to the Lightfield adaptation system. The other displays will be connected either through DVI or HDMI interfaces.
The audio rendering will be synchronized with video using the scene description in the A/V player and will drive a multichannel audio interface equipment connected to a set of loudspeaker rigs that could be Binaural, Stereo, Multichannel surround or Wave Field Synthesis (WFS).
6 QUALITY OF EXPERIENCE EVALUATION

QoE (Quality of Experience) is inherently a subjective concept, and as such, the ultimate way of measuring QoE is using human subjects in controlled experiments to obtain subjective measurements of the different dimensions of QoE. Several methodologies for subjective evaluation of video quality, sound quality and multimedia quality for particular applications have been standardised within the ITU [4] [5] [6].

Previous work has only considered audio and video independently and there is no QoE metric for 2D audiovisual content. Moreover, there is no metric available for 3D content quality.

QoE for audiovisual contents is composed of a number of basic quality components. These are identified as the QoE dimensions. The spatial audio experience, 3D visual experience and comfort are the QoE dimensions that are considered in the MUSCADE project. These three QoE dimensions are studied and modelled independently. Once the models for QoE dimension are available, they are combined to obtain the overall QoE model for 3D audiovisual contents.

KPIs (Key Performance Indicators) are used for modelling spatial audio and 3D visual experience dimensions. For modelling spatial audio experience QoE dimension, basic audio quality, front spatial fidelity and surround spatial fidelity are used as the KPIs. Image quality and depth perception are considered as KPIs for modelling 3D visual experience QoE dimension. The correlation between the sound and picture is also considered for modelling the overall QoE model. This attribute includes the spatial synchronicity of audio and video, which is the correlation of source positions including azimuth, elevation and depth derived from aural and visual cues. It is known that, the user demographics, audiovisual displays available to the user and different production types may affect the user expectations, hence the QoE. These will be considered, albeit to a limited extent, defined by the availability of different 3D audio and video displays, realistic test durations and the variance in demographics of subjects that pass pre-screening. The comfort dimension will be modelled by itself from the direct subjective results acquired through a questioner.

The relationship between the QoE dimensions and KPIs is depicted in figure 9.

7 CONCLUSION

This paper has given an overview of the work being carried out by the MUSCADE consortium in the fields of 3DTV data representation, capture and production, encoding, transmission and rendering. The system architecture and the first technological choices have been explained.

MUSCADE is a 3-year project running until the end of 2012.

References


Acknowledgment

The author would like to thank all MUSCADE consortium members for their contribution to this work.