IMMERSIVE PRODUCTION AND DELIVERY OF INTERACTIVE 3D CONTENT

DELIVERABLE 7.3.3

THIRD REPORT ON CURRENT STANDARDS AND CONTRIBUTIONS TO STANDARDISATION BODIES

Contractual Date of Delivery:	November 30, 2014
Actual Date of Delivery :	November 30, 2014
Work Package:	WP7
Dissemination Level:	Public
Nature of Deliverable:	Report

Lead Contractor:	BBC
Coordinating Person:	Graham Thomas
Project Leader:	BBC
Work Package Leader:	BBC
Editor:	Wolfgang Paier (HHI)
Authors:	Wolfgang Paier (HHI), Graham Thomas (BBC), Peng Huang (Surrey)

Classification and Approval

Classification: Public

This document has the status 'Public' and may be published as a whole including this page. Publication of parts of the document must be in unanimous agreement within the RE@CT Steering Board and subsequent EC approval/agreement.

Disclaimer

Neither the RE@CT Consortium nor any of its officers, employees or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing neither the RE@CT Consortium nor any of its officers, employees or agents shall be liable for any direct or indirect or consequential loss or damage, personal injury or death, caused by or arising from any information, advice or inaccuracy or omission herein.

Acknowledgements

All partners of the RE@CT project contributed to this report.

RE@CT Consortium Overview

Participant no. *	Participant organisation name	Short name	Country
1 (Coordinator)	British Broadcasting Corporation	BBC	UK
2	Fraunhofer HHI	HHI	Germany
3	INRIA	INRIA	France
4	University of Surrey	Surrey	UK
5	Artefacto	ART	France
6	OMG	OMG	UK

Abstract

This document is the final of three deliverables reporting on the evolution of relevant standards and possible contributions to the respective standardisation bodies. It is an updated version of D7.3.2. Existing as well as upcoming industry standards play an important role in RE@CT. Wherever RE@CT relies on established technologies, partners carefully evaluate which standards exist and how novel tools can be built on top of these standards to ensure interoperability and connectivity with existing tools and workflows. Building on existing standards will clarify where these standards have to be amended and extended in order to incorporate the workflow and technology developed in RE@CT.

In this document, standards for several key technologies within the RE@CT project are identified, summarized and discussed with respect to the project. Standards in video coding are discussed, with a special focus on upcoming standards in 3D video representation which are developed for the increasingly popular 3D displays, as well as standards for representation and compression of 3D geometry, and standards regarding virtual characters. Finally, relevant standards in production environments are described.

Table of Contents

1 Introduction

In this document, the discussion of the relationship between technology developed in RE@CT and existing and future industry standards is updated. RE@CT is proposing a radically advanced workflow for content production which is systematically explored and evaluated for the very first time. Therefore it was not predictable whether additions to existing industry standards or proposals for new standards will be made during the project. It is the aim of RE@CT to deliver a proof of principle and first demonstrators for the novel production workflow. Standardization of the new techniques should be pursued after the workflow has been thoroughly tested and the tools developed in RE@CT have been refined.

Nonetheless, existing as well as upcoming industry standards play an important role in RE@CT. Wherever RE@CT relies on established technologies, partners carefully evaluate which standards exist and how novel tools can be built on top of these standards. This helps to ensure interoperability and connectivity with existing tools and workflows, and it facilitates the exchange of data within the project as well as with other projects and productions. At the same time, building on existing standards will clarify where and how these standards have to be amended and extended in order to incorporate the workflow and technology developed in RE@CT in the future.

In this third document, the standards identified to take possible influence on the project in D7.3.2 are revisited to include the most recent developments with respect to the project. In section 2, standards in video coding are discussed in the light of the large amount of data generated in the RE@CT content production workflow. Special focus is put on upcoming standards in 3D video representation which are developed for the increasingly popular 3D displays. Section 3 looks at standards for representation and compression of 3D geometry, as well as standards regarding virtual characters. Finally, section 4 summarizes relevant standards in production environments.

2 Standards in Video Coding

Of particular relevance to RE@CT are industry standards in the domain of video coding: RE@CT capture is based on multi-view video with a large number of viewpoints, each captured at broadcast-level quality and image resolution to allow high-quality free-viewpoint re-rendering and re-animation. This results in a large amount of video data. Also, the delivery of RE@CT content over networks is addressed in the project. In the following we briefly outline existing standards in video coding, and we address current standardization activities with respect to 3D and multiview video. Finally, we discuss these developments in the light of RE@CT's requirements.

2.1 Existing MPEG standards related to RE@CT

Existing standards on 3D video coding for the transmission over networks are mainly driven by the Motion Picture Expert Group (MPEG) of ISO/IEC and the Video Coding Expert Group (VCEG) of ITU-T. In 2008, MPEG has released a standard on Multi-View-Coding (MVC) [8]. It is an extension of the H.264/AVC standard [5] and combines both temporal and inter-view prediction to enable efficient coding of multi-view video as captured in the RE@CT immersive studio setup. Overall, the coding efficiency could further be improved, although there is still the crucial limitation that the data rate is almost linearly dependent on the number of views. Therefore, research shifts more and more towards multiple video-plus-depth (MVD) representations [1]. In MPEG, related standardisation activities have recently been started under the notation 3D Video Coding (3DV) [3] which allows the encoding of a multi-view video sequence with additional depth maps for each video sequence. It can handle occluded areas if they are visible by at least one camera, and it is backward compatible to simple stereoscopic video. All these standards have been developed mainly for sequential video with well-defined unidirectional temporal prediction structures.

Figure 1 Video frame plus depth map

2.2 Current developments for 3D video representations within MPEG

Currently, there are several activities by the Joint Collaborative Team on 3D Video Extension Development (JCT-3V), a joint working group of MPEG and ITU-T VCEG, in order to standardize coding and representation of 3D video. The goal is to extend existing standards for regular as well as multi-view video by depth information to enable auto-stereoscopic and 3D viewing. These 3D extensions are based either on the already finalized H.264 / AVC standard or exploit the new High Efficiency Video Coding (HEVC) initiative which is targeted as a future successor of H.264. In the following, a brief summary of the extensions are provided together with expected timelines.

2.2.1 3D video with H264 / AVC-based coding technology

2.2.1.1 MVC compatible extension including depth (MVC+D)

The main target of this work item is to enable 3D enhancements while maintaining MVC stereo compatibility. The method is invoking an independent second stream for the representation of depth maps that are treated as monochrome video data, as well as high-level syntax signalling of the necessary information to express the interpretation of the depth data and its association with the video data. Macroblock-level changes to the AVC or MVC syntax, semantics and decoding processes are not considered. The specification of this extension (which is 14496-10:2012/Amd.2 as ISO/IEC standard), reached DAM status in July 2012 and was finalized in April 2013 (FDAM).

2.2.1.2 AVC compatible video-plus-depth extension (3D-AVC)

Stereo and multi-view video systems are challenging and require a higher data rate, even more when depth data is included in the representation. Therefore, another possible extension of the AVC standard is currently being investigated, which takes advantage of compressing texture and depth data jointly. This imposes that compatibility with MVC can no longer be retained, however applications exist which only require monoscopic AVC compatibility. Current algorithms investigated in this context outperform the MVC-compatible approach by approximately 20% bit rate reduction for the entire (video-plus-depth) information. The combined coding can however only be achieved when the syntax and decoding process for additional (non-base) texture views and depth information changes at the block level, compared to AVC/MVC. The 3D-AVC extension reached DAM status in 2013 and was finalized in January 2014 (FDAM).

2.2.2 3D video with HEVC-based coding technology

High-efficiency video coding (HEVC) is a new video coding standard which aims to provide the same subjective quality for monoscopic video at half the bit rate of AVC High Profile. A primary usage of HEVC is seen in the area of high and ultra-high definition (UHD) video. Many HD displays already provide the capability for stereo rendering, and it can be expected that the increased resolution and display size of UHD displays makes them even more suitable for such purposes. Beyond that, the improved compression capability of HEVC makes it attractive for the introduction of stereo; for example, it can be expected that by using mechanisms that exploit the redundancy between views, HEVC would be able to encode full resolution stereo at significantly lower rates than AVC would be capable of for just one (monoscopic) view at the same quality and resolution.

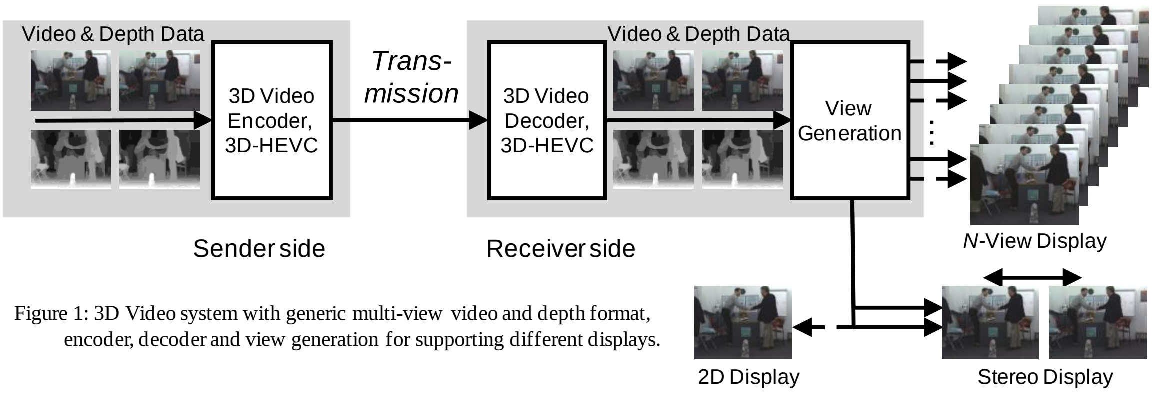

Figure 2 3D-HEVC encoding and transmission overview

HHI has largely contributed to the standardisation of HEVC which was officially approved in April 2013 [17]. The codec uses a number of tools for reducing the correlation with video sequences, including spatial prediction within a picture, temporal motion-compensated prediction between pictures at different time instances; transform coding of the prediction residual, and entropy coding.

In HEVC, these tools are highly optimized, such that it achieves the same subjective video quality at only about 50% of the bit rate on average [18], in comparison to its predecessor (H.264/ MPEG 4 AVC High Profile).

2.2.2.1 Multiview extension of HEVC (MV-HEVC)

To achieve higher compression efficiency than frame compatible stereo and provide extension for applications with more than 2 views, HEVC-based multi-view coding is being investigated by JCT-3V, where the redundancy between different views captured from the same scene can be exploited in a similar fashion as by AVC's multi-view approach (MVC), and is therefore called Multiview HEVC (MV-HEVC). Here, the base view is fully compatible with monoscopic HEVC and can thus be interpreted by legacy HEVC decoders. This is facilitated by extending the high-level syntax appropriately, and by rearrangement of decoded picture buffers to store the reference pictures as needed, without any changes to the core of the coding layer below the level of coded tree blocks (CTB).

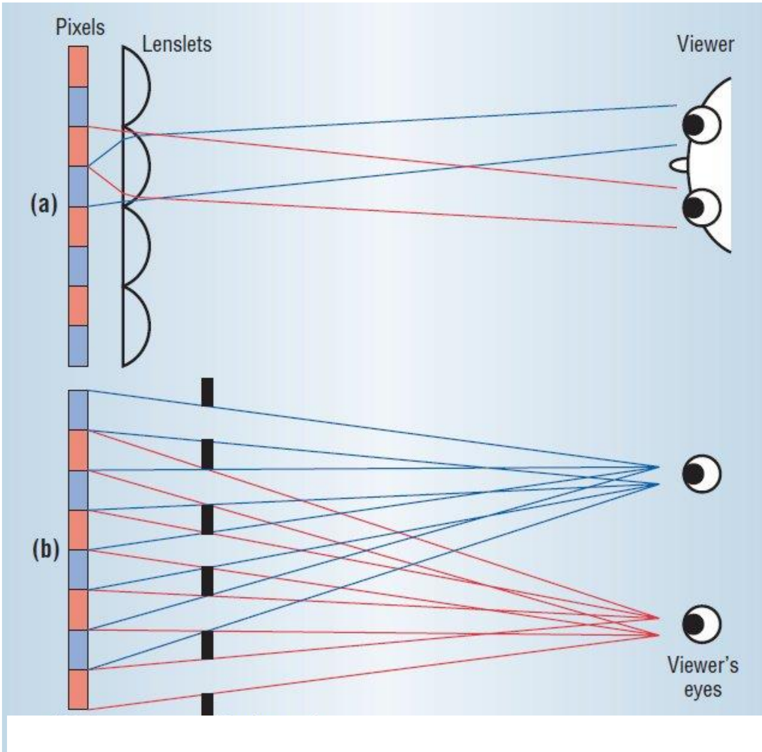

Figure 3 Principles of autostereoscopic displays

In a stereo coding application, the number of pixels to be processed and the decoding complexity would not be higher than for frame-compatible coding with the same resolution of left and right images, whereas the data rate could be reduced through exploitation of inter-view redundancies. Similar to AVC's MVC depth extension (i.e. carrying depth maps as an independent monochrome channel without change of low-level coding tools), MV-HEVC provides a representation for optional depth information. The standardization work for MV-HEVC was finished in July 2014.

2.2.2.2 Multiview plus depth coding with HEVC-based technology (3D-HEVC)

With the advancement of UHD display technology and the capability to present stereoscopic views with higher resolution, the demand for higher compression capability is again expected to arise together with advanced display features supported by depth maps. This could again be achieved by joint compression of video texture and depth maps, similar as in the 3D-AVC concept discussed above, where it is likely to be necessary to include more sophisticated video and depth coding tools and/or tools exploiting inter-component dependencies with possible new definitions at the sub-CTB level of the core codec. It is anticipated that the market adoption of such an approach would require providing about 25% of total bit-rate reduction in comparison to MV-HEVC. HHI proposed a set of novel tools to exploit several types of redundancies in multiview data for data-rate reduction in [19] and [20]. The 3D-HEVC extension reached DAM status in January 2014. Since the specification work is still in progress, the final version of the extension is expected for June 2015.

2.3 Relevance to RE@CT

We have identified three aspects of the RE@CT project to which video coding standards are of particular relevance:

1. Compression of multi-view video.
2. Delivery of RE@CT content to a user device over the network.
3. Display of RE@CT content on a 3D display.

2.3.1 Compression of multi-view video

RE@CT capture is based on multi-view video with a large number of viewpoints, each captured at broadcast-level quality and image resolution for high-quality re-rendering, resulting in a vast amount of data. In a production environment, maximal video quality must be retained, and no compression should be applied to the raw material. However, for delivery of a finalized RE@CT-based production (e.g. a video game or an interactive application), it is advisable to investigate the use of 3D video coding standards for compressing RE@CT multiview video textures.

The current standardization activities discussed above are driven by the growing market for 3D display technology. display technology. Therefore, they aim primarily at the efficient encoding of a limited number of views with small inter-views with small inter-view baselines, as required by current 3D displays (see

Figure 3): Simple displays provide a different view for each eye of a single viewer (see example in Figure 4.), which requires a baseline related the average distance of human eyes. Alternatively, multi-viewpoint displays offer a range of different views, such that adjacent pairs of views provide a stereo image for one of several viewpoints in front of the display. Again, the disparities between adjacent views are related to human eye distance, and the total disparity is a multiple of that.

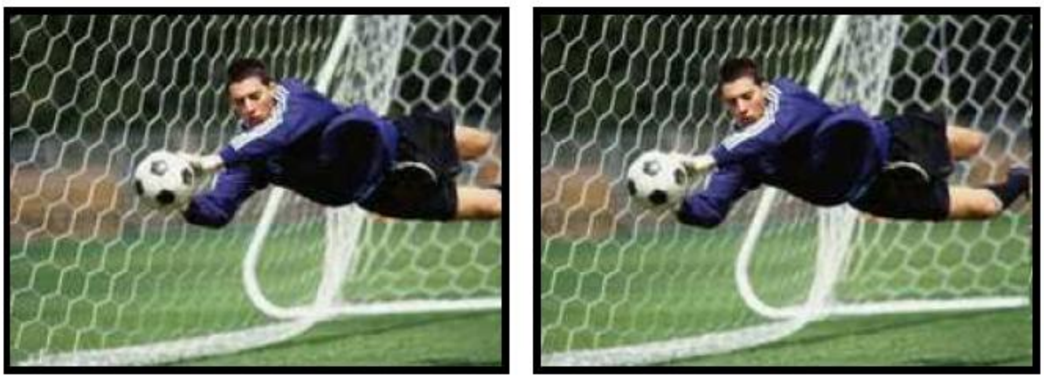

Figure 4 Example of viewpoints required by a 3D display

RE@CT's multi-view video textures, however, differ significantly from this 3D display content, covering an entire scene over 360 degrees in order to provide the maximal amount of freedom in re-rendering and re-animation. Figure 5 shows an example from the first RE@CT test production at BBC. The baselines between the cameras are in the range of 0.5 to 1.5 meters, and images from adjacent viewpoints are less similar than adjacent views for 3D displays.

Video + depth data formats are designed to enable the interpolation of many in-between with small baseline variations views for multiple-viewer type displays. We do not expect these techniques to be applicable to RE@CT video data without significant research effort, which is beyond the scope of the project.

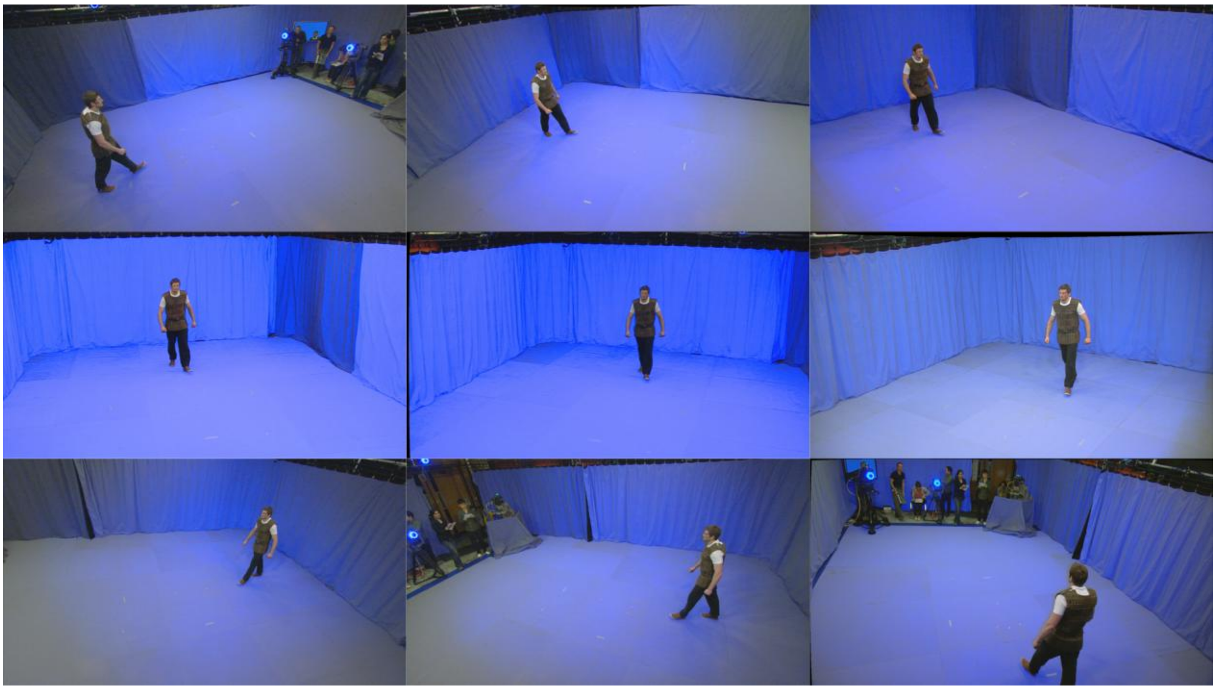

Figure 5 Example of viewpoints in RE@CT multiview footage

Multi-view data formats such as MV-HEVC, where the redundancy between different views captured from the same scene is exploited for compression without an explicit depth map layer, are a more promising candidate for RE@CT multiview data. It remains to be investigated which compression rates can be achieved given the lower inter-view similarity of the RE@CT footage in comparison to stereo display content. Also, RE@CT scenes do not rely on the entire image content, but only the segmented actor(s) and props. This can be exploited for compression but also violates established scene assumptions which compression standards are based on.

2.3.2 Delivery of RE@CT content to a user device over the network

In the project, techniques for delivery of RE@CT content to an end-user device over the network are to be discussed. Due to the high data load of RE@CT multi-view textures, transmitting them live over the network to large number of clients is not a realistic scenario in the foreseeable future. It was therefore proposed to exploit server-side rendering of the scene, and to deliver only the final image over the network as explained in D5.4. This can be achieved on the basis of existing and future standards of video compression, especially H.264/MPEG-4 AVC and the upcoming HEVC. Transmission of server-rendered RE@CT content is not different from standard video transmission over a network. Extensions like Scalable Video Coding (SVC) are applicable to RE@CT content as to any other video content delivered over the network.

2.3.3 Display of RE@CT content on a 3D display

For displaying finalized RE@CT content on a 3D display, the video + depth standards discussed above are particularly well suited: The RE@CT workflow encompasses a high quality 3D reconstruction of the actors and props involved in the scene. The 3D data is used for the free viewpoint rendering process. Therefore, the depth maps required for display on 3D screens can easily be obtained by reading out the depth buffer which is generated in the rendering process anyway. This is an advantage over traditional stereo capture techniques, where depth maps have to be generated per image by stereo reconstruction. RE@CT's highly sophisticated and specialized 3D reconstruction can be expected to provide better depth maps of scenes than generic stereo methods.

3 Standards in Geometry Representation, Geometry Coding and Virtual Characters

RE@CT is introducing a completely new technique and pipeline for 3D Character Animation production, and therefore there are no existing industry standards that natively support the Virtual Character Representation required in RE@CT. The most relevant existing standards are created by the Web3D Consortium which coordinates with ISO and IEC. Those standards are summarized below.

3.1 Relevant existing standards for 3D Geometry

Extensible 3D (X3D) is an ISO-ratified standard that provides a system for storage, retrieval and playback of real-time graphics content embedded in applications, all within an open architecture to support a wide array of domains and user scenarios. X3D defines a software system that integrates network-enabled 3D graphics and multimedia. Conceptually, each X3D application is a 3D time-based space that contains graphic and aural objects that can be dynamically modified through a variety of mechanisms. **X3D part 1: Architecture and base components** [14] defines the architecture and base components of X3D. It describes an abstract functional behaviour of time-based, interactive 3D, multimedia information. **X3D Part 1: Extensible Markup Language (XML) encoding** [15] provides a Web-compatible format that maximizes interoperatibilty with other Web languages. An XML representation has several advantages:

- XML supports structuring data,
- it is similar to HTML,
- it is readable by systems as well as humans,
- it is verbose for clarity,
- it represents a modular family of technologies,
- it is fundamentally interoperable with Web technologies.

However, XML-based formats are expensive to transmit and parse. Therefore **X3D Part 3: Compressed binary encoding** [16] provides a compact transmission format that minimizes delivery size and maximizes parsing speed.

Besides X3D, MPEG-4 standards [2] exist that describe 3D geometry encoding, addressing not only representation but also compression for efficient delivery. ISO MPEG has recognized the importance of efficient 3D mesh compression and included a tool called 3D Mesh Compression (3DMC), for static meshes exploiting spatial dependencies of adjacent polygons [10]. Later, research was extended to compression of animated meshes and mesh sequences with dynamic vertex positions. Temporal dependencies can be exploited by key meshes and vertex prediction schemes [7]. This approach called Interpolator Compression (AFX-IC) was adopted as an extension of the computer graphics part of MPEG-4 called Animation Framework eXtension [4] for dynamic mesh coding. Recently, the MPEG community has introduced a new model for 3D graphics compression, by means of Part 25 of the MPEG-4 standard [2]. The model makes it possible to use MPEG-4 compression on top of a third party XML description, facilitating its deployment. This includes the following tools for compression: 3D Mesh Compression 3DMC, Wavelet Subdivision Surface, different Interpolators, Bone-based Animation, and Frame-based Animated Mesh Compression [6], an amendment to AFX for compressing time-consistent 3D meshes. These tools already define skeletal motion representations as well as animated surface meshes but are, similarly to the video coding standards, defined for unidirectional prediction.

X3D standards provide the base point of interoperability with the MPEG-4 interactive profile. X3D standards are widely adopted by governments such as the EU, US, and AU, agencies such as NASA, US Navy, NSF, and major industry players such as Apple, IBM, Intel, Microsoft and Walt Disney.

3.2 Relevant existing standards for Animation

Humanoid Animation (H-Anim) [13], also within X3D standards, is an ISO/IEC standard for humanoid modelling and animation. H-Anim specifies a systematic method for representing humanoids in a network-enabled 3D graphics and multimedia environment. Conceptually, each humanoid is an articulated character that can be embedded in different representation systems and animated using the facilities provided by the representation system. H-Anim specifies the semantics of humanoid animation as an abstract functional behaviour of time-based, interactive 3D, multimedia articulated characters. H-Anim does not define physical shapes for such characters but does specify how such characters can be structured for animation. The intension is for a wide variety of presentation systems and application, and provides wide latitude in interpretation and implementation of the functionality.

4 Standards in Studio Production Environments

Most standards for production technology and video standards are produced by the Society for Motion Picture and Television Engineers (SMPTE) and the International Telecommunications Union (ITU). Some existing relevant standards are given below.

4.1 Relevant existing SMPTE standards

SMPTE 292M [11] defines the serial digital interface used for uncompressed HD video, which is used by the broadcast-standard HD cameras and capture hardware in the RE@CT capture system. Most professional HD camera equipment uses this format, and as long as all cameras, video switchers, cables and capture systems that the project uses are compatible with this standard, they should all inter-operate correctly.

4.2 Relevant existing ITU standards

ITU-R Recommendation BT.709 [12] covers image standards for HDTV, including image size, colour space, and transfer characteristics. The images captured by broadcast-standard HDTV cameras in the RE@CT production system are in accordance with this standard. The details of how images are represented are relevant for those parts of the project that are performing image processing tasks, e.g. it may be necessary to consider the transfer characteristics (gamma) when blending images.

4.3 Standards under development

SMPTE is currently working on a replacement for the timecode standard (SMPTE 12M), which includes methods for synchronising devices over networks rather than using a video reference signal. It should also support a wider range of frame rates, and so for example might be relevant to high-speed motion capture cameras as well as conventional broadcast cameras. BBC has a representative working in the relevant SMPTE group (33TS).

References

[1] P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, A. Smolic, and R. Tanger: Depth Map Creation and Image Based Rendering for Advanced 3DTV Services Providing Interoperability and Scalability, *Signal Processing: Image Communication. Special Issue on 3DTV*, 2007

[2] ISO/IEC 14496-25: MPEG-4 Part 25: 3D Graphics Compression Model, 2008

[3] ISO/IEC JTC1/SC29/WG11, "Vision on 3D Video", Lausanne, Switzerland, 2009

[4] ISO/IEC JTC1/SC29/WG11, "ISO/IEC 14496-16/PDAM1", Doc. N6544, Redmond, WA, USA, 2004

[5] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC 14496-10 Advanced Video Coding", 2003

[6] ISO/IEC JTC1/SC29/WG11 (2007) a.k.a., "MPEG4 Part 16 AMD2: Frame-based Animated Mesh Compression," ISO, 2007

[7] E. Jang, J. Kim, S. Jung, M. Han, S. Woo, S. Lee, Interpolator Data Compression for MPEG-4 Animation, IEEE Trans. on Circuits and Systems for Video Technology, 2004

[8] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC 14496-10:200X/FDAM 1 Multiview Video Coding", Doc. N9978, Hannover, Germany, 2008

[9] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC 14496-10 Version 8 Scalable Video Coding", 2007

[10] ISO/IEC JTC1/SC29/WG11 "Information Technology - Coding of Audio-Visual Objects. Part 2: Visual; 2001 Edition", Doc. N4350, Sydney, Australia, 2001

[11] SMPTE standard ST 0292-1-2012, 1.5 Gb/s Signal / Data Serial Interface.

[12] ITU-R Recommendation BT.709: BT.709 : Parameter values for the HDTV standards for production and international programme exchange. http://www.itu.int/rec/R-REC-BT.709/en

[13] ISO/IEC 19774: "Humanoid Animation (H-Anim)", 2006.

[14] ISO/IEC 19775-1: "Extensible 3D (X3D) Part 1: Architecture and base components", 2008

[15] ISO/IEC 19775-1.2: "Extensible 3D (X3D) Part 1: XML encoding", 2009

[16] ISO/IEC 19776-3: "Extensible 3D (X3D) Part 1: Compressed binary encoding", 2007

[17] G. J. Sullivan et al, "Overview of the High Efficiency Video Coding (HEVC) Standard," IEEE Trans. Circuits and Systems for Video Technology, Dec. 2012

[18] J.-R Ohm et al, "Comparison of the Coding Efficiency of Video Coding Standards – Including High Efficiency Video Coding (HEVC)," IEEE Transactions on Circuits and Systems for Video Technology, Dec. 2012

[19] K. Müller et al: "3D High Efficiency Video Coding for Multi-View Video and Depth Data", IEEE Transactions on Image Processing, Sept. 2013

[20] A. Vetro, K. Müller, "Depth-Based 3D Video Formats and Coding Technology", in "Emerging Technologies for 3D Video: Creation, Coding, Transmission and Rendering", John Wiley & Sons, Chichester, UK, ISBN 978-1-1183-5511-4, 2013