Computer Technologies for 3D Video Delivery for Home Entertainment

Atanas Gotchev

Abstract: *The paper overviews the state-of-the-art in the area of 3D video content creation, coding, transmission, and display. Computer technologies for 3D video capture, format conversion and compression, as well as network technologies for 3D video delivery are briefly described. Standardization activities in the area are reviewed as well. As of 3D displays, special attention is driven to the most mature technology in that field – the so-called auto-stereoscopic displays delivering 3D visual sensations to single or multiple users without any special glasses. Conclusions about the open questions before the 3D technology reaches the users homes are drawn.*

Key words: *3D video, multi-view, auto-stereoscopic displays, MPEG, DVB.*

INTRODUCTION

3D is expected to be the next big thing in video delivery technologies. In a historical retrospective, visual content representation has been continuously evolving to offer greater realism to the users. First, colour was added then high-definition was introduced. While these technology movements added a more realistic gamut of light and colour and increased the spatial resolution, 3rd dimension has been so far missing for consumer use like home entertainment. In cinemas, there have been 3D technologies, such as IMAX and RealD, relying on special glasses and polarized light projection to create stereo illusion but the true 3D content delivery to user at homes is still to come. However, there are optimistic signs that this move is rather close than faraway. Critical technology components, such as multi-camera and multi-sensor systems, stereoscopic and auto-stereoscopic displays, broadband channels and high-capacity storage devices are available and in good level of maturity. Relevant standards are supported by big industry players, such as Philips, Samsung, Nokia and others. High-profile research institutions, such as Fraunhofer HHI in Germany, Korean Electronics and Telecommunications Research Institute (ETRI), Finnish Agency for Technology and Innovations (TEKES) have steered their research priorities toward 3D. At European level, the European Commission has identified 3D Media as a top research priority as well.

This paper presents an overview of the current situation in the area of 3D video delivery technologies. We first address the general requirement for such technologies and the barriers for their mass market introduction. Then we review the techniques for 3D capture, representation, coding and display and emphasize the research challenges from a signal processing point of view. At the end, we make some conclusions.

3D MEDIA NOW

3D is supposed to enhance the visual user's experience substantially. More appealing, attracting, intelligent and informative visual scenes can be delivered by 3D. The rapid developments in computer technologies have also increased the general enthusiasm regarding the 3D. There are numerous experts who predict that home entertainment will be completely transformed by stereoscopy [1]. In order these optimistic forecasts to happen, there are three factors needed as emphasized by experts:

1. Ability to show 3D on high-quality TV systems.
2. Increased source material available to the consumer.
3. Increased consumer interests through exposure to 3D visual content.

The first factor is related with the technology; the second is related with the content availability and the third is related with the quality to be achieved.

Technology developments have been facing the following requirements: utilize the existing infrastructure and media; develop techniques with minimal change to device components such as hardware enders/decoders; ensure backward compatibility; and support wide range of display devices. Utilization of the existing infrastructure, media and devices is important as these are expensive and have been built for many years. Technologies based on existing infrastructure are also attractive for rapid commercialization. To this end, backward compatibility is something always required for media technologies. It is the same way as mono audio is supported by stereo devices or black and white movies are watched on colour TV sets. As there is no dominating display technology for 3D, the video delivery techniques have to support if not all, then at least major of them. A related important problem is the availability of suitable data delivery formats. At the moment there is no single such format. A unique format would make the subsequent developments easier and more focused. Since several competing formats are available, the standardization efforts have been concentrated on avoiding format wars. Systematic standardization efforts in this area have been initiated [2].

C*ontent creation* developments have been influenced by the technology ones. A television station in Japan started broadcasting ordinary television contents in high-definition stereoscopic form in December 2007. The NBA All-Star game in 2007 was captured, transmitted and displayed in 3D [3]. On March 8, 2008, a joint venture between the BBC and The3DFirm, organized stereo-video delivery via satellite of A Rugby Six Nations match. 3D cinema is gaining momentum in Hollywood with plenty of new movies created in 3D.

However, the decisive factor for the acceptance of the new technology is the *quality* of the respective 3D content. In many cases, stereoscopic presentations have been uncomfortable or even impossible to view. The issue of quality has to be addressed systematically and thoroughly in order to ensure that 3D video content will reach end users at their homes.

3D CAPTURE AND REPRESENTATION

Mainly three approaches for 3D visual content creation can be underlined. First, such content can be captured by two or more synchronized cameras in a multi-camera setting. Second, such content can be created from 2D video applying video processing methods. Third, video output can be augmented by depth information captured by another sensor.

Stereo cameras of high definition already exist in the market. Usually such cameras are customised versions of high-end sensors from major manufacturers such as Sony, Panasonic or Canon, placed in a specially designed rig, so to form the left and right views. For example, PACE (Burbank, CA) is delivering high-end systems for 3D and Digital Cinema Systems based on the Sony HDC-950 camcorder; 21st Century is delivering 3D video kit based on two Panasonic AG-CVX100 cameras, with customized electronic synchronization of the shutters. Such cameras can be used for direct capture of stereoscopic imagery with varying parameters. In addition various multi-camera settings exist [4]. Their purpose is to capture 3D scene from different directions in a view-redundant representation so to allow for subsequent depth estimation and inter-view interpolation.

The extraction of 3D structure from 2D video is another content creation approach particularly interesting as it can provide an easy way to convert already existing 2D material to 3D [4]. Shape-from-motion (or multi-view) has been favoured to be the most promising approach. Pioneering approaches for reconstructing 3D scenes from broadcast video are known [5].

Dense depth map representation [6] is one of the most widely used 3D representation formats of a scene. The scene is represented by two channels – a 'normal' 2D video channel accompanied with a gray-scale channel carrying out the quantized depth values of each video pixel. Such representation can be used to render desired views. Depth map estimation utilizes stereo or multiple views. An excellent taxonomy for depth from stereo algorithms is proposed in [7]. Occlusion problems, characteristic for depth from stereo can be handled by adding more views.

A promising method for acquiring depth information from natural 3D scenes is the so-called *multisensory* capture. In this approach, depth measurement data of various sensors are registered, and fused with the visual channel. Various types of sensors are used – PMD cameras, laser and acoustic scanners, movement sensors. The depth measurements acquired by such sensors are accurate but usually with low spatial resolution.

An alternative approach measures depth using structured light. Such system consists of one or more cameras, aided by active light sources. The sources are used to project known pattern sequence onto the scene, which enables labelling of each pixel. Though measuring pixel correspondences between images, depth maps can be obtained with high precision. Other advantages of this approach are the relatively low costs of the equipment involved, and that no re-calibration is needed during operation. Disadvantages of the method are weak performance when dealing with shadows, specular objects, and partially occluded areas. Also, depth acquisition by using structural light is particularly prone to aliasing errors, coming from the limited spatial and dynamic resolution of the light capturing system [8].

In order to produce a depth map with sufficient quality for artistic 3D content, various signal-processing techniques need to be applied in a post-processing stage. These include super-resolution techniques for increasing the resolution of the depth map, suppression of noise and other acquisition artefacts, and proper alignment between the contours in the video stream and these in the depth map. Recent state-of-the-art image- and video-filtering algorithms are based on sophisticated, though quite efficient, spatially-adaptive transform-domain operations. These methods typically utilize both the sparsity and the statistical properties of multiresolution representations as well as the inherent correlations between frames in temporal and spatial dimensions.

3D COMPRESSION AND BROADCAST

Efforts to find an efficient way to compress 3D moving imagery have reached a standardization stage. Compression standards for video are investigated by the Moving Picture Experts Group (MPEG), a working group of ISO/IES and the corresponding standards are issued with ISO/IES designations [9]. There are mainly three standardization attempts in the area of 3D known as MPEG-C, part 3; multi-view coding (MVC) and 3D video coding initiative (3DVC), respectively.

ISO/IEC 23002-3 Auxiliary Video Data Representations (MPEG-C part 3) is meant for applications where additional data needs to be efficiently attached to the individual pixels of a regular video. The specification is directly applicable to 3D video. Such video is represented in the format of single view + associated depth, where the single channel video is augmented by the per-pixel depth attached as auxiliary data. As such, it is susceptible to efficient compression. Rendering of virtual views is left for the receiver side [10], [11],[12].

Multiview Video Coding (MVC, ISO/IEC 14496-10:2008 Amendment 1) is an extension of the Advanced Video Coding (AVC) standard. It targets coding of video captured by multiple cameras. The video representation format is based on N views, i.e. N temporally synchronized video streams are input to the encoder which outputs one bitstream to be decoded and spit into N video signals at the decoder [13], [14].

As a standard, 3DVC targets serving various multi-view displays. N output display views are assumed. However, a lower number of K views is assumed to be coded and transmitted. To enhance the rendering capabilities, the transmitted views are equipped with additional depth information. The number of views and depth channels is an open research question. This representation generalizes the possibilities of MPEG-C, Part 3 and MVC. 3DVC is an ongoing MPEG activity, and a standard is expected in 2010/2011 [15],

[16], [17].

The work of digital TV broadcasting standards is carried out within the DVB Project. A brief summary of the standards is given on the DVB fact sheets page [18]. They usually base the broadcast on the MPEG-2 transport stream (TS), so 3D video streaming relates on MPEG-2 TS as well. Within a FP6-funded project, called ATTEST, backward-compatible delivery of 3DTV was demonstrated on DVB-T system [19]. The 3D video content was represented in the video-plus-depth (2D+Z) format where MPEG-2 coded 2D video was augmented by an MPEG-4 coded depth map sequence and then encapsulated in MEPG2 TS and broadcast over DVB-T.

3D DISPLAYS AND RELATED ARTEFACTS

Modern 3D display systems address the binocular parallax through providing different visual information to the left and right eyes. This could be done by using stereoscopic displays equipped with glasses or by using autostereoscopic displays where no glasses are needed.

Autostereoscopic displays are typically fabricated using an optical filter on top of a conventional LCD or OLED display. Two types of optical filters are used - lenticular sheet [20] and parallax barrier [21].

Autostereoscopic displays present several artefacts such as picket-fence, image flipping, or cross-talk can cause eye strain to the viewer [22]. Techniques based on viewer-tracking have been suggested as an effective way to adaptively soften the visibility of the above-mentioned artefacts. Such approach is beneficial for a single user and not for multiple users in front of a multi-view display. When multiple observers are targeted, *slanted* optical filter is used in order to mitigate picket-fence and image flipping artefacts [23]. However, the image in each view appears on a non-rectangular grid, which creates the need of specially designed anti-aliasing filters [24], [25]. Cross-talk is another common artefact in auto-stereoscopic displays. Its presence can hinder the ability of an observer to fuse the views of a 3D scene [26]. For stereoscopic displays with two views, signal processing techniques employing image pre-compensation enables significant reduction of the perceived crosstalk [27], [28].

CONCLUSIONS

There is a clear trend toward developing 3D video delivery technologies for home users. Displays, networks and communication channels, and video processing algorithms and standards are on the way. So far, 3D content representation formats have been studied mainly for their compressibility and susceptibility to transmission errors. However, much more studies are needed about how different formats affect user perception. In general, these should deliver quality and performance comparable and competitive if not superior to the best available monoscopic techniques. The question of the quality of new 3D content is very much related with the way how it is measured. The quality measurement and assessment should be done in combination of measurement and assessment of the 2D image quality plus the associated depth plus the naturalness of the scene.

As of 3D visual content visualization, it is not clear yet which technology: either glasses-based or glasses-free will prevail. Fewer artefacts versus free viewing? Would the auto-stereoscopic display technology achieve such quality so to dismiss the use of obtrusive glasses? Apparently, the research challenge here is to develop efficient, quality-optimized, signal processing techniques and visualization tools.

REFERENCES

[1] Insights Media, “2007 3D report”, available from Insight Media, http://www.insightmedia.info

[2] ISO/IEC JTC1/SC29/WG11 (MPEG) Call for Contributions on FTV Test Material, (http://www.chiariglione.org/mpeg/hot_news.htm.

[3] National Basketball Association press release, “NBA Goes 3D HD for 2007 NBA All-Star”, available online at http://www.nba.com/allstar2007/news/3d_hd_070212.html.

[4] E. Stoykova, A. A. Alatan, P. Benzie, N. Grammalidis, S. Malassiotis, J. Ostermann, S. Piekh, V. Sainov, C. Theobalt, T. Thevar, and X. Zabulis ,3D Time Varying Scene Capture Technologies – A Survey, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, No. 11, pp. 1568-1586, 2007.

[5] E. İmre, S. Knorr, B. Özkalaycı, U. Topay, A. Alatan, T. Sikora, Towards 3-D Scene Reconstruction from Broadcast Video, Signal Processing: Image Communication 22, pp 108-126, 2007.

[6] A. A. Alatan, Y. Yemez, et.al., Scene Representation Technologies for 3DTV - A Survey. IEEE Trans. Circuits Syst. Video Techn. 17(11): 1587-1605 (2007.

[7] D. Sharstein and R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal on Computer Vision, Volume 47, pg 7-42, April 2002.

[8] Thomas P. Koninckx, Pieter Peers, Philip Dutré, Luc J. Van Gool: Scene-Adapted Structured Light. CVPR (2) 2005: 611-618.

[9] Moving Pictures Experts Group, http://www.chiariglione.org/mpeg/

[10] ISO/IEC JTC1/SC29/WG11, “Text of ISO/IEC FDIS 23002-3 Representation of Auxiliary Video and Supplemental Information”, Doc. N8768, Marrakech, Morocco, January 2007.

[11] ISO/IEC JTC1/SC29/WG11, “Text of ISO/IEC 13818-1:2003/FDAM2 Carriage of Auxiliary Data”, Doc. N8799, Marrakech, Morocco, January 2007

[12] ITU T and ISO/IEC JTC 1, “Advanced video coding for generic audiovisual services”, ITU-T Rec. H.264 and ISO/IEC 14496-10 AVC, 2003, most recent Version: 2005.

[13] ISO/IEC JTC1/SC29/WG11, “Study Text of ISO/IEC 14496-10:2008/FPDAM 1 Multiview Video Coding”, Doc. N9760, Archamps, France, May 2008.

[14] ISO/IEC JTC1/SC29/WG11, “Joint Multiview Video Model (JMVM) 8”, Doc. N9762, Archamps, France, May 2008.

[15] ISO/IEC JTC1/SC29/WG11 “Overview of 3D Video Coding”, Doc. N9784, Archamps, France, May 2008.

[16] ISO/IEC JTC1/SC29/WG11 “Description of Exploration Experiments in 3D Video Coding”, Doc. N9783, Archamps, France, May 2008.

[17] ISO/IEC JTC1/SC29/WG11 “Applications and Requirements of FTV”, Doc. N9466, Shenzhen, China, Oct. 2007.

[18] DVB-H Broadcasting to Handhelds, The Global Technology Standard for Mobile Television, available online at: http://www.dvb.org/technology/fact_sheets/

[19] C. Fehn, “Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3DTV”, SPIE, proceedings of, vol. 5291, 2004

[20] Van Berkel, “Lenticular screen adaptor”,US Pat. 6801243, issued Oct. 5, 2004.

[21] Schmidt and A. Grasnick, "Multi-viewpoint autostereoscopic displays from 4D-vision", in Proc. SPIE Photonics West 2002: Electronic Imaging, vol. 4660, pp. 212-221, 2002

[22] L.M.J. Meesters, W.A. IJsselsteijn, and P.J.H. Seuntiëns, "A Survey of Perceptual Evaluations and Requirements of Three-Dimensional TV", in IEEE Trans. Circuits and Systems for Video Tech., vol. 14., no. 3., March 2004.

[23] C. van Berkel, D. Parker and A. Franklin, “Multiview 3D LCD,” in Proc. SPIE Vol. 3012, Stereoscopic Displays and Virtual Reality Systems III, (Fisher, Merritt, Bolas, edts.), p. 32-39, 1996.

[24] J. Konrad and P. Angiel, “Subsampling models and anti-alias filters for 3-D automultiscopic displays”, Image Processing, IEEE Transactions on , vol.15, no.1, pp. 128-140, Jan. 2006.

[25] A. Jain and J. Konrad, "Crosstalk in automultiscopic 3-D displays: Blessing in disguise?" in Proc. SPIE Stereoscopic Displays and Virtual Reality Systems, vol. 6490, Jan. 2007, pp. 12.1-12.12.

[26] Y. Yeh and L. Silverstein, “Limits of Fusion and Depth Judgement in Stereoscopic Colour Displays”, Human Factors, vol 32(1), pp. 45-60, 1990.

[27] J. Konrad and M. Halle, "3-D displays and signal processing: An answer to 3-D ills?," IEEE Signal Process. Mag., vol. 24, pp. 97-111, Nov. 2007.

[28] J. Lipscomb and W. Wooten, "Reducing crosstalk between stereoscopic views," in Proc. SPIE Stereoscopic Displays and Virtual Reality Systems, vol. 2177, Feb. 1994, pp. 92-96.

ABOUT THE AUTHOR

Atanas Gotchev, Department of Signal Processing, Tampere University of Technology, phone: +358 3 3115 4349, E-mail: atanas.gotchev@tut.fi