3D Video: new techniques and challenges

Emilie Bosc, Muriel Pressigout, Luce Morin

First Sino-French Workshop on Education and Research collaborations in Information and Communication Technologies
SIFWICT 2011 - May 19 - 20, 2011, Nantes, France

3D Video: new techniques and challenges

Emilie Bosc, Muriel Pressigout, Luce Morin
IETR UMR CNRS 6164
INSA of Rennes
35708 RENNES CEDEX 7, France
firstname.name@insa-rennes.fr

***Abstract*—Three-dimensional television (3DTV) is meant to enhance conventional 2D television by the added feeling of depth. The introduction of 3DTV will be successful if the perceived image quality and the viewing comfort are at least comparable to conventional television. Indeed, 3DTV raises a number of challenges concerning various fields (technical and technological, evaluation protocols, etc.). The success of 3DTV at home is strongly related to the resolution of such questions. This paper discusses the 3DTV technology and the challenges to overcome.**

I. Introduction

3D video has gained a growing interest for the last few years. Industrials have turned their attention to this technology and aim at providing users a brand new quality of experience that should offer more immersion. 3D video refers to tow main applications, namely 3D television (3DTV) and Free Viewpoint Video (FVV). 3DTV provides a depth feeling, and Free Viewpoint Video allows navigation inside the scene.

3DTV and FVV are thought as the logical evolution of conventional 2D television. Besides, 2D conventional techniques and protocols were naturally thought as appropriate candidates in view to an extension to 3D. However, it turns out not to be that simple: indeed, along the whole processing chain, from the acquisition to the display, choices should be studied in order to improve users' quality of experience.

This document identifies the new challenges brought out by 3DTV technology into different categories. Section II addresses the problem of data format and compression. Section III addresses the view synthesis question. Section IV addresses the evaluation question. Finally, Section V concludes the paper.

II. Compression methods challenge

Among the numerous data formats that were proposed as for 3D video [1], Multiview Video plus Depth (MVD) data is one of the most studied lately, but not exclusive (Layer depth Image (LDI) is also a candidate, derived from MVD). Indeed, most suited input data format is still questioned, regarding the targeted application.

MVD consists of a set of multiple sequences capturing the same scene at different viewpoints (namely texture data), with their associated per-pixel depth value (namely depth data) (Figure 1). Depth data provide information on scene geometry and help in virtual intermediate view generation.

Then, MVD data refer to a huge amount of data to be processed. Considering texture and depth information are required for view synthesis, efficient coding tools are needed. Numerous methods have been proposed and an important part consists in improving the efficient H.264/AVC codec [2], [3] : H.264/MVC [4] can be used as a compression method for both texture and depth videos of MVD. MVC stands for Multiview Video Coding [5].

However, most of the proposed methods are inspired from 2D codecs that are optimized for 2D perception. Yet, artifacts that may not be perceived in a 2D color image, may have drastic effects on depth data, as pointed out in [6]. Indeed, the importance of depth data and texture data in view reconstruction is still discussed: because of the bandwidth constraints, it may be judicious to efficiently set this bit ratio between depth adn texture. Considering its capital role in the view synthesis processing, compression artifacts of depth data may lead to fatal synthesis errors when generating virtual views. For this reason, bit allocation between depth and texture data is still addressed [7].

III. View synthesis algorithms

From depth data and texture data, new intermediate viewpoints can be synthesized with the help of depth-image-based rendering techniques. The generated views can then be rendered on a conventional display, or a stereoscopic or an autosterescopic display.

Generating a "virtual" view consists in synthesizing a novel view of the scene, from a viewpoint which differs from those captured by the cameras, relying on the available texture and depth data. The texture, that is the conventional 2D color sequences, gives the color information. The depth data are gray-scales images and are considered as a monochromatic signal. Each pixel of a depth map indicates the distance of the corresponding 3D-point from the camera. Based on projective geometry [8], the 3D representation of a scene can be retrieved

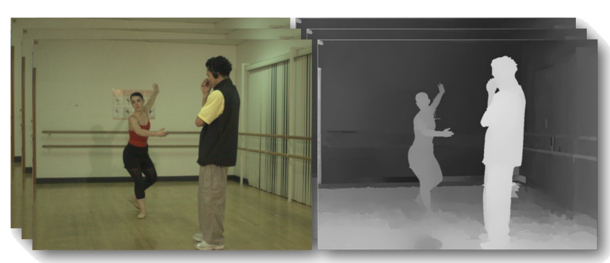

Fig. 1. MVD data consist of texture (left) and depth information (right).

Fig. 2. Rendering objects edges is a tough task for view synthesis algorithms.

from a depth map.

Virtual view synthesis is essential because captured video sequences are generally not adapted to all 3D screen displays: to ensure depth feeling, synthesizing new viewpoint stereoscopic pairs is usually required. Depth Image Based Rendering algorithms (DIBR) algorithms are used but they are prone to projection errors, especially around depth discontinuities (Figure 2). Considerable effort ([9]) has been made to overcome the challenge related to rendering. However, the proposed methods need to be validate through objective or subjective protocols that may differ from the 2D ones [10].

IV. 3D Perception challenge

The compression question already consists of a serious challenge and in addition to this, is the problem of evaluating the proposed coding framework, or any new device. Those methods or devices need to be validated through robust measurements of particular features. The demand for high-quality visual content makes the need for a reliable assessment protocol as essential. And indeed, the difficulty comes from the fact that 3D vision involves physiological mechanisms that are still not fully understood. In point of fact, thanks to recent studies that brought this complexity into focus, it is now more understandable. [11] reported that not only image quality should be taken into account for an assessment framework, but visual comfort and depth feeling as well. Concerning image quality, new distortion types have appeared, as discussed in [12], [13]. Some of them are related to synthesis process, others are related to compression methods or display technologies. A few of them are:
- *keystone* effect : the image look like a trapezoid.
- *ghosting* effect: this is a shadow-like artifact.
- *cardboard* effect: depth is perceived as unnatural, as discrete incoherent planes.
- *puppet-theater* effect: characters in the scene appear miniaturized.
- *staircase* effect: discontinuities appear between adjacent blocks.

3D video brings new types of artifacts that usual metrics may not be able to efficiently assess [14]. Especially, pixel-based metric PSNR (Peak Signal to Noise Ratio) revealed not to be sufficient for assessing synthesized views in which object translation may occur without affecting the perceived image quality. Furthermore, the problem of assessing 3D video refers first to the evaluated target. Assessment studies in 3D are not mature and there is no standardized assessment framework for 3D video. [10] proposed new requirements for subjective video quality assessment. In particular, the authors point out the fact that the minimum of 15 observers, as recommended in ITU-BT.500, may not be sufficient because of the instability of viewers' assessment in 3D. As well, they suggest reconsidering the statistical analysis of viewers' opinion for appropriate rejection of incoherent responses. For 3DTV applications, the authors explained, above all, how test material (3D displays and content disparity) affects depth rendering. This invites to carefully analyze future experiment results.

V. Conclusion

There is a serious need to study and develop methods, tools, and requirements, in 3D video, that allow great 3D quality of experience. Task forces are required first to understand human visual system. That should be the key for future technical choices. This knowledge will help in defining more perceptual-based metrics and new assessments protocols. Then improvements are expected in the fields of compression, that should focus on 3D perception criteria, and view synthesis.

Acknowledgment

This work is supported by ANR-PERSEE project n^o ANR-09-BLAN-0170, ANR-CAIMAN project n^o ANR-08-VERS-002.

References

[1] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G. B Akar, G. Triantafyllidis, and A. Koz, "Coding algorithms for 3DTVa survey," *IEEE transactions on circuits and systems for video technology*, vol. 17, no. 11, pp. 16061620, 2007.

[2] ITU-T Recommendation H.264, "Advanced video coding for generic Audio-Visual services," 2009.

[3] T. Wiegand, G. J Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h. 264/AVC video coding standard," *IEEE Transactions on circuits and systems for video technology*, vol. 13, no. 7, pp. 560576, 2003.

[4] ISO/IEC JTC1/SC29/WG11,, "Text of ISO/IEC 14496-10:200X/FDAM 1 multiview video coding," 2008.

[5] Y. Chen, Y. K Wang, K. Ugur, M. M Hannuksela, J. Lainema, and M. Gabbouj, "The emerging MVC standard for 3D video services," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, 2009.

[6] P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Muller, and T. Wiegand, "The effects of multiview depth video compression on multiview rendering," *Singal Processing: Image Communication*, vol. 24, no. 1-2, pp. 73–88, 2009.

[7] E. Bosc, V. Jantet, M.l Pressigout, L. Morin, and C. Guillemot, "Bit-rate allocation for multi-view video plus depth," in *to appear in 3DTV Conference 2011*, Turkey, 2011.

[8] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge Univ Pr, 2003.

[9] M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, and Y. Mori, "Reference softwares for depth estimation and view synthesis," Apr. 2008.

[10] W. Chen, J. Fournier, M. Barkowsky, and P. Le Callet, "New requirements of subjective video quality assessment methodologies for 3DTV," 2010.

[11] P. Seuntiens, *Visual Experience of 3D TV*, Ph.D. thesis, 2006.

[12] M. Meesters, W. Ijsselsteijn, and P. Seuntiens, "A survey of perceptual evaluations and requirements of three dimensional TV," *IEEE Transactions on Circuits And Systems for Video Technology*, vol. 14, no. 3, pp. 381–391, Mar. 2004.

[13] M. Cancellaro, V. Palma, and A. Neri, "Stereo video artifacts introduced by a distributed coding approach," 2010.

[14] E. Bosc, M. Koppel, R. Pepion, M. Pressigout, L. Morin, P. Ndjiki-Nya, and P. Le Callet, "Can 3D synthesized views be reliably assessed through usual subjective and objective evaluation protocols?," in *to appear in ICIP 2011*, Bruxels, 2011.