Video Quality Assessment: From 2D to 3D - Challenges and Future Trends

Quan Huynh-Thu, Patrick Le Callet, Marcus Barkowsky

VIDEO QUALITY ASSESSMENT: FROM 2D TO 3D – CHALLENGES AND FUTURE TRENDS

Quan Huynh-Thu

Technicolor
1, av. de Belle Fontaine
CS 17616 – 35576 Cesson-Sévigné, France

Patrick Le Callet, Marcus Barkowsky

Polytech' Nantes
IRCCyN UMR 6597 CNRS
Rue Christian Pauc, 44306 Nantes, France

ABSTRACT

Three-dimensional (3D) video is gaining a strong momentum both in the cinema and broadcasting industries as it is seen as a technology that will extensively enhance the user's visual experience. One of the major concerns for the wide adoption of such technology is the ability to provide sufficient visual quality, especially if 3D video is to be transmitted over a limited bandwidth for home viewing (i.e. 3DTV). Means to measure perceptual video quality in an accurate and practical way is therefore of highest importance for content providers, service providers, and display manufacturers. This paper discusses recent advances in video quality assessment and the challenges foreseen for 3D video. Both subjective and objective aspects are examined. An outline of on-going efforts in standards-related bodies is also provided.

Index Terms— video quality, 3D, objective metrics, subjective assessment, standards

1. INTRODUCTION

The availability of multimedia services has greatly expanded in the recent years thanks to advances in video coding, convergence of networks and increase of transmission bandwidth. One fundamental aspect of the overall quality of experience (QoE) of a multimedia service is the video quality.

Whilst the access to high-definition video content is still to reach everyone in a home environment, three-dimensional (3D) video is gaining a strong momentum both in the cinema and broadcasting industries as it is expected to enhance extensively the viewer's visual experience through a higher level of immersion in the media content. Several 3D video formats and 3D video coding strategies currently co-exist and practical tools (i.e. objective quality metrics) to compare them in terms of QoE would be very useful for researchers and for the industry.

Video quality can be measured using either subjective assessment or objective measurement. Subjective testing requires human observers to view videos and provide their opinion of quality. Quality measurement using objective models (computational algorithms) can provide a more practical solution. However these objective models are only useful if their measurement closely correlates with subjective quality. The development of reliable objective models depends ultimately on the validation of those models using reliable subjective assessment.

From a discussion on recent advances in the field of video quality assessment and an outline of recent standardization efforts focusing mainly on VQEG and ITU, this paper intends to raise the challenges foreseen for 3D video and gives an overview of the ongoing effort towards that goal.

2. 3D VIDEO TRANSMISSION CHAIN

Since 3D video is an extension of 2D, it can be affected by the same types of visual distortions than those encountered in 2D video. However, there are many additional aspects that can influence the 3D visual experience. Figure 1 shows the block diagram of a typical transmission chain. Several blocks are common to a 2D and 3D transmission chain but, in most cases, additional processing steps are required for 3D, and the delivery of 3D signals causes new types of artifacts.

2.1. Acquisition of 3D signals and format conversion

It can be considered that the most complete 3D representation is achieved by computer-generated imagery (CGI). The underlying 3D model can be stored and the scenario can be rendered from any position with as many (virtual) cameras as necessary. However, this might require an excessive processing power.

On the other hand, from a 2D still image only a very small number of cues are available about the 3D structure. Algorithms have been proposed to extract 3D information from a 2D image using monocular cues such as the focus information, the texture or shades. The result is a depth map, which measures the distance to the nearest visible object. This resulting representation is often referred to as 2D plus depth.

In the extraction of depth from video sequences, the motion parallax can be used to achieve a higher accuracy of the depth map. A disadvantage of the 2D plus depth representation is that only information about the first object in the line of sight for a monocular view is available. This causes problems when a second view is rendered and parts of objects that are further away must be disoccluded. Disocclusion designates the recovery of hidden parts of an object, usually by using inpainting algorithms. This can be avoided by adding another layer of texture and depth information which is often referred to as occlusion layer. The occlusion layer can be estimated in videos from camera or object movement.

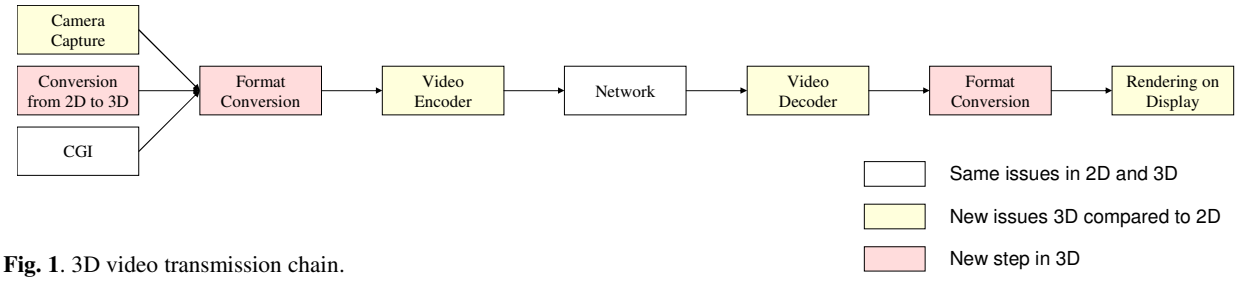

Fig. 1. 3D video transmission chain.

In the acquisition of stereo content, often two separate cameras are used, thus resulting directly in a stereo pair. Depending on the precision of the alignment, several artifacts can be introduced such as vertical misalignment, color misalignment, different focus points or zoom levels, temporal offsets or geometry distortions resulting for example from toed-in configurations. The captured stereo pair is often stored and transmitted in a combined manner. Mostly, top/bottom, left/right or line interleaved formats are used.

A higher flexibility for the rendering is achieved by recording more than two views. The additional information can be used to adjust the depth range or to support free viewpoint navigation or to enable motion-parallax rendering for example by autostereoscopic multiview displays. The captured data is typically converted into Layered Depth Video (LDV), Multiple Video plus Depth (MVD) or Depth Enhanced Stereo (DES) format.

The wide variety of possible representations for 3D content often requires a format conversion. This conversion can only be considered as lossless if the output is a subset of the input, e.g. when converting from MVD to just a single view plus depth. In most cases, the artifacts introduced by the format conversion lead to noticeable degradations. For example, when estimating the depth from stereoscopic video, the resulting depth map contains many errors due to the ambiguity of features found in the two images. Sometimes background objects are mistakenly arranged in the foreground. When re-rendering the original stereoscopic view, the depth impression is severely distorted and in addition, visible artifacts around the borders of those objects occur.

2.2. Transmission of 3D signals

Numerous approaches exist to encode, transmit and decode 3D video signals, the easiest being a simulcast transmission of the different views or depth maps using standard 2D video codecs such as H.264. An extension to H.264, called Multi View Coding (MVC), was developed to allow the compression, transmission and storage of 3D video. MVC was adopted as a standard format on the Blu-Ray Disc. The independent or combined transmission of 3D video signals leads to new artifacts which most often lead to binocular rivalry. Moreover, each compression algorithm requires a specific input representation, thus conversions between formats frequently occur, leading to information loss.

2.3. Display of 3D signals

At the display side, another format conversion may occur when either 2D plus depth representation was used for transmission or a different viewpoint needs to be rendered. Depth Image Based Rendering (DIBR) approaches are frequently used. These render the stereo pair before the display, producing a dedicated image for the left and the right eye. Because at least one viewpoint differs slightly from the transmitted view, inpainting algorithms need to fill the previously occluded image regions. The inaccuracy of the inpainting often produces artifacts around the edges.

3. VIDEO QUALITY ASSESSMENT

3.1. Subjective quality assessment

Subjective assessment of (2D) video quality can be considered to be a mature field. The International Telecommunication Union (ITU) has recommended several methodologies for standard-definition [1], high-definition [2] and low-resolution video [3][4].

On the other hand, this is not the case for the subjective quality assessment of 3D or stereoscopic video. A first international recommendation was published in 2000 [5]. However, it mostly discusses the way to measure the stereo acuity of subjects. It also mentions the vergence-accommodation conflict that occurs on most of today's displays due to the fact that flat screens are used. While the accommodation of the Human Visual System (HVS) focuses on the screen because the objects appear to be most sharp on the display plane, the disparity of the objects between the left and the right eye leads to a convergence of the eyes towards a point in front or behind the display plane. This is an unnatural condition.

In the subjective evaluation of 3D images and video sequences, the video quality is closer to the concept of a quality of experience and should be considered to be multi-dimensional: visual quality, depth quality/perception, and comfort. The first dimension may be considered to be the visual quality in the 2D sense because observers usually view a 3D video for the first time in a subjective test, whilst they have a lot of experience with 2D television quality. The added value of depth was often proposed as a second criterion, and the term "naturalness" was proposed to express the combination of the perceived depth and the overall quality [6]. Comfort is crucial as it has also been reported that some observers experience visual fatigue with symptoms like eye strain, headache or nausea. This effect is often measured using questionnaires [7]. A recent excellent summary of the causes can be found in [8].

The 3D display itself has a large impact on the stability and reproducibility of the subjective experiment. As the 3D display technology is still advancing, different technologies exist and none can be recommended as a reference. The viewing angle, the field of view, the amount of crosstalk and the brightness are often limiting factors. The International Committee for Display Metrology (ICDM) will soon release a Display Measurement Standard (DMS) to unify the measurement of display properties [9].

As was mentioned earlier, the 3D content has to be prepared specifically to fit the 3D display, e.g. the depth range has to be adapted. This adjustment depends on the display characteristics and on the viewing distance [10].

Special attention is required on the way the display itself processes the 3D content. Often, crosstalk reduction is applied by the playout program or a format conversion takes place, e.g. from 2D plus depth to nine distinct views, and the rendering artifacts may easily outweight the added value of depth [11].

ITU-R WP6C is working towards the identification of requirements for the broadcasting and subjective testing of 3DTV [12], whilst ITU-T Study Group 9 added 3D video quality in its scope in 2009 [13]. However, all the issues mentioned previously constitute major challenges in finding a standardized way to characterize and measure the perceived quality of 3D video.

3.2. Objective quality metrics: from 2D to 3D

Whilst quality assessment of video impaired by coding distortions has been widely covered in the literature, research on 2D video models that handle distortions due to transmission errors (e.g. packet loss) has only flourished recently. Since 1997, the Video Quality Experts Group (VQEG) is a good witness of this effort. This group has been investing a lot of effort to examine, test and validate objective models using subjective data collected around the world. Based on VQEG's results, several international recommendations for objective video quality metrics have been established. ITU-T Rec. J.144 and its counterpart ITU-R Rec. BT.1683 were published in 2004. Both provide four full- reference objective video quality assessment models for standard-definition television signals impaired by coding distortions only. ITU-T Rec. J.247 was published in 2008 and recommends four full-reference models for low-resolution video impaired by both coding and transmission errors, whilst reduced-reference models are included in ITU-T Rec. J.246. The results of the on-going validation phase for HDTV [14] are expected to produce new recommendations for objective models in 2010/2011. Pursuing their effort to align with the trends in the field, VQEG has recently extended their work to 3D video quality and is first investigating the issues related to subjective testing protocols.

Compared to 2D, the objective assessment of video quality in 3D is more complex:

- there are additional steps in the transmission chain that need to be addressed, as depicted in Fig. 1;
- the observer's opinion may be considered as multidimensional, including factors like visual fatigue and depth perception;
- more aspects of the HVS need to be addressed, e.g. binocular rivalry, binocular suppression.

In the 3D transmission chain, visible artifacts occur at several locations. The camera capture or the conversion and rendering steps may introduce geometric degradations [15] or distortions in 3D size leading to the puppet theater effect. For a recent definition of these artifacts, see [16].

As the 3D artifacts partially stem from the newly introduced steps in the transmission chain, it may be advantageous to perform the objective measurement with signals extracted at different stages as well. For example, a no-reference measure for geometric distortions may be used to indicate artifacts resulting from the capture process. Later, a full-reference approach may be applied to evaluate the quality of the conversion algorithms from a stereo pair to 2D plus depth representation by using a reference implementation of the inverse step and comparing the input stereo pair to the re-rendered views [17]. The classical 2D quality assessment algorithms measure the lost information between the video encoder and decoder and may thus be used as basis for this part of the 3D transmission chain. Enhancements are necessary for binocular artifacts, e.g. occlusions, perspective distortions, depth distortions or the detection of binocular rivalry. In the optimal case, the screen itself should be considered as perfect by the objective measurement but it might also be worthwhile to model the influence of display artifacts such as crosstalk.

When the influence of each step in the transmission chain is known, an open question remains on how to combine the different degradations, as each artifact at a given point in the transmission may be emphasized or diminished by the following steps in the transmission. In this sense, a holistic video quality estimation algorithm that takes inputs from several stages in the 3D transmission chain might be advantageous. This model would also benefit from the data generated during the transmission, e.g. a depth map.

In 2D video quality assessment, the best performing algorithms use a full-reference approach comparing the displayed video with the captured sequence pixel by pixel. This is not universally applicable to 3D as the displayed video might be rendered from a different perspective than the one that was captured.

It should be mentioned that the video content itself has a larger impact on the perceived visual quality in 3D than in 2D. Typical issues are objects that are clipped by the frustum, a moving camera perspective destabilizing the human sense of orientation and fast moving objects in the foreground which result in visual discomfort [18]. While 2D proposals have been presented for the automated characterization of content, little work has been done for 3D.

Studies have indicated that viewers tend to focus their attention on specific areas of interest in the image and models of visual attention have been proposed [19]. There is an increasing interest in using visual attention models (saliency maps) inside video quality assessment models in order to improve their accuracy [20]. Visual attention is without any doubt also a crucial factor in the perception of 3D video.

3.3. Objective quality metrics: status on 3D

Besides the lack of reliable subjective QoE assessment methodologies for 3D, several approaches to objective 3D video quality assessment have been proposed so far. This might affect the validation value of the metrics themselves. We might suspect that those metrics have been able to capture one dimension of the 3D QoE, which represents nonetheless valuable first contributions towards an ideal QoE metric.

An objective assessment algorithm which uses the depth map as well as the stereoscopic views is proposed in [17]. It includes parts of the Structural Similarity Index (SSIM), the detection of edge and color degradations.

The binocular suppression indicates that one view of the stereo pair might be transmitted at a worse visual quality than the other. This is investigated in [21] using a rate-distortion model based on the estimation of visual quality with an adapted Peak Signal to Noise (PSNR) and a jerkiness metric. The influence of a reduction in spatial, temporal and quality dimension is analyzed. A model based on the idea that the inferior view should not add high frequency components is proposed and analyzed in [22].

The applicability of PSNR and 2D video models (SSIM and VQM) to 3D was investigated on a small dataset both for the case of stereoscopic video and monoscopic video with depth information [23]. Results show that 3D video quality might be estimated from separate assessment of each stereo-view, whilst models for 2D could also be used to estimate the quality of depth perception. A similar analysis is performed in [24]. Both subjective experiments used a Philips 42-inch autostereoscopic display for presenting the 2D plus depth content. However, a study has shown that subjects prefer to switch off the 3D effect on this display [11].

4. CONCLUSION AND FUTURE DIRECTIONS

Subjective quality assessment of 2D video is a very mature field. However, subjective assessment of 3D video quality is still facing many problems to solve before the performance of 3D video models can be properly evaluated in order to capture the essential QoE involved by such media. Standardized protocols for measuring display characteristics and for characterizing the different dimensions of perceived video quality in 3D are still needed.

The research literature has started to investigate the applicability of 2D objective video quality assessment models to 3D video, as quality issues in 2D and 3D video are related and present some similarities. However, the lack of reliable ground truth (subjective dataset) reflecting the essence of 3D QoE limits the value of this initial effort. Moreover, 3D video presents in addition significantly different quality issues that are not encountered or don't have their equivalent in 2D. As opposed to 2D video where a direct analysis of the transmitted signal can produce a quality measure that correlates with subjective opinion, in 3D it is not the signal itself but rather the rendered version that needs to be analyzed. For these reasons, it should be difficult to measure 3D video quality simply by using existing or straightforward extensions of 2D video quality heritage. Objective quality assessment of 2D video is somewhat a mature field where researchers are considering very complex degradations beyond compression artifacts, such as those caused by transmission errors. Although 3D video will face the same problems, the objective assessment of 3D video quality is today in its infancy, with many problems to solve already in the characterization of the concept of 3D video quality

5. REFERENCES

[1] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," Rec. BT.500, 2002.

[2] ITU-R, "Subjective assessment methods for image quality in high-definition television," Rec. BT.710, 1998.

[3] ITU-R, "Methodology for the subjective assessment of video quality in multimedia applications," Rec. BT.1788, 2007.

[4] ITU-T, "Subjective video quality assessment methods for multimedia applications," Rec. P.910, 2008.

[5] ITU-R, "Subjective assessment of stereoscopic television pictures," Rec. BT.1438, 2000.

[6] R.G. Kaptein, A. Kuijsters, M.T.M. Lambooij, W.A. IJsselsteijn, and I. Heynderickx, "Performance evaluation of 3D-TV systems," in *Proc. SPIE Image Quality and System Performance V*, vol. 6808, Jan. 2008.

[7] M. Pölönen, T. Jarvenpaa, J. Hakksinen, "Comparison of near-to-eye displays: subjective experience and comfort," *J. Display Technol.*, vol. 6, no. 1, pp. 27–35, Jan. 2010.

[8] M.T.M Lambooij, W.A. IJsselsteijn, and I. Heynderickx, "Visual discomfort in stereoscopic displays: a review" in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XIV*, vol. 6490, Jan. 2007.

[9] Society for Information Display, "Display measurements standard", http://icdm-sid.org/Public/DMS/ICDM-DMS.html

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

[11] M. Barkowsky, R. Cousseau, and P. Le Callet, "Influence of depth rendering on the quality of experience for an autostereoscopic display," in *Proc. Int. Workshop QoMEX*, pp. 192–197, July 2009.

[12] ITU-R, "Digital three-dimensional (3D) TV broadcasting," Question ITU-R 128/6, 2008.

[13] ITU-T, "Objective and subjective methods for evaluating perceptual audiovisual quality in multimedia services within the terms of Study Group 9," Question 12/9, 2009.

[14] Video Quality Experts Group, "VQEG HDTV test plan," available at http://www.vqeg.org, June 2009.

[15] A. Woods, T. Docherty, and R. Koch, "Image distortions in stereoscopic video systems," in *Proc. SPIE Stereoscopic Displays and Applications IV*, vol. 1915, pp. 36–48, Feb. 1993.

[16] L.M.J. Meesters, W.A. IJsselsteijn, P.J.H. Seuntiëns, "A survey of perceptual evaluations and requirements of three-dimensional TV," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 3, pp. 381–391, Mar. 2004.

[17] H. Shao, X. Cao, and G. Er, "Objective quality of depth image based rendering in 3DTV system," *in Proc. 3DTV Conference*, pp. 1–4, May 2009.

[18] F. Speranza, W.J. Tam, R. Renaud, and N. Hur, "Effect of disparity and motion on visual comfort of stereoscopic images," in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XIII*, vol. 6055, pp. 94–103, 2006.

[19] O. Le Meur and P. Le Callet, " What we see is most likely to be what matters: visual attention and applications", in *Proc. IEEE ICIP*, pp. 3085–3088, Nov. 2009

[20] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Does where you gaze on an image affect your perception of quali*ty? Applying vi*sual attention to image quality metric," in *Proc. IEEE ICIP*, vol.2, pp. 169–172, Sep 2007.

[21] N. Ozbek and A.M. Tekalp, "Unequal inter-view rate allocation using scalable stereo video coding and an objective stereo video quality measure," in *Proc. IEEE Int. Conf. Multimedia and Expo*, pp. 1113–1116, 2008.

[22] F. Lu, H. Wang, X. Ji, and G. Er, "Quality assessment of 3D asymmetric view coding using spatial frequency dominance model," *in Proc. 3DTV Conference*, pp. 1–4, May 2009.

[23] S.L.P. Yasakethu, C.T.E.R. Hewage, W.A.C. Fernando, and A.M. Kondoz, "Quality analysis for 3D video using 2D video quality *models," IEEE Trans. Consum. Electron.,* vol. 54, no. 4, pp. 1969–1976, Nov. 2008.

[24] C.T.E.R. Hewage, S.T. Worrall, S. Dogan, S. Villette, and A.M. Kondoz, "Quality evaluation of color plus depth map-based stereoscopic video," *IEEE. J. Sel. Topics Signal Process., vol.* 3, no. 2, pp 304–318, 2009.