CHALLENGES AND OPPORTUNITIES IN VIDEO CODING FOR 3D TV

Hari Kalva, Lakis Christodoulou, Liam Mayron, Oge Marques, and Borko Furht

Dept. of Computer Science and Engineering, Florida Atlantic University, Boca Raton, FL 33431

ABSTRACT

This paper explores the challenges opportunities in developing and deploying 3D TV services. The 3D TV services can be seen as a general case of the multi-view video that has been receiving significant attention lately. The keys to a successful 3D TV experience are the availability of content, the ease of use, the quality of experience, and the cost of deployment. Recent technological advances have made possible experimental systems that can be used to evaluate the 3D TV services. We have developed a 3D TV prototype and have currently conducting our first user study to evaluate the quality and experience. These experiences have allowed us to identify challenges and opportunities in developing 3D TV services.

1. INTRODUCTION

The recent interest in 3D and multi-viewpoint (MV) TV can be attributed, in part, to the success of the MPEG-4 AVC/H.264 video coding standard. The coding gains made possible by H.264 can be applied to provide enhanced services such as multi-viewpoint TV and 3D television. Another reason for the increasing interest in 3D TV is the recent advances in the display technologies that have lowered the cost of projectors and 3D displays. While these technological advances have renewed interest in 3D/multiview coding, the successful deployment of 3D services still faces key challenges. Given the current state of the technology and the maturity of the marketplace, this maybe the right time to overcome the barriers to the 3D and MV TV services.

Ever since the invention of the Wheatstone stereoscope in 1938, there has been a constant interest in stereo imaging and this extended to the 3D motion pictures and television. There have been a large number of 3D movies and the technology has been steadily improving. One of the recent successes is "Chicken Little" shown in 3D using special digital projectors and polarized glasses to separate the left

and right eye view. Alternative technologies use synchronized/active eyeglasses to present the proper images to the left and the right eye. The successes of the 3D movies, however, have not translated into successes for 3D TVs. The primary obstacles have been the quality of the displays and the cost of deployment. The TV industry has been experimenting with 3D TV by offering some special programming in 3D. NBC recently (Nov 2005) showed portions of an episode of "Medium" in 3D. The glasses to experience the 3D program were distributed in that week's issue of TV Guide magazine. The world will see another 3D TV experiment during the 2006 World cup Soccer in Germany, in summer 2006. The soccer matches are expected to be simulcast in 3D for users equipped with autostereoscopic displays with eight view points. Autostereoscopic displays provide a 3D experience without any need for glasses and the advances in autostereoscopic display technology is expected to be one of the key drivers for the successful deployment of 3D TV.

One of the reasons for the lack of success of the 3D TV so far is the ease of use and viewing comfort. Most of the displays today use standard TV with anaglyph video and a pair of glasses to generate 3D perception. Watching such TV for long periods causes eye strain. Even the current generation autostereoscopic displays have limited viewing angle and are not suitable for viewing for longer periods. The application where 3D TV has had reasonable success are the applications where viewing comfort is secondary to the objective; applications such as security, medicine, design automation, and scientific visualization.

The digital video revolution launched by the MPEG-1 and MPEG-2 video coding standards also resulted in an active 3D and multi-view video coding research [1, 2]. The MPEG-2 multiview profile is a form of temporal scalability that encodes left view of the stereo pair as a base layer and the right view is coded as a temporal enhancement. The existing studies on the quality of 3D video are based on MPEG-2 view coding and not directly applicable to the H.264 based coding that is expected to be used in 3D TV services [3]. The studies also did not use autostereoscopic displays that are expected to be the dominant display types for 3D TV. The MPEG-2 based coding is inefficient compared to the H.264 based view coding; furthermore, the coding artifacts in MPEG-2 and H.264 are different and are likely to have different effects on the 3D perception. One of the goals of our work in 3D coding area is to understand the

impact of the coding artifacts on asymmetric view coding. Since each compression algorithm causes different artifacts, especially in the view coded at a lower bitrate, the effect of the compression algorithm on the quality of the 3D video would be different. The quality of 3D video experiences is also influenced by the type of displays used. A good summary of the perceptual quality requirements and evaluations for 3D video is presented in [4]. Our current focus is on developing efficient coding and representation algorithms for 3D and multi-view video. We are using H.264 as the basis for view coding and autostereoscopic displays for rendering the 3D video.

In this paper we discuss the key challenges to the success of 3DTV and discuss ways to overcome some of these challenges. We describe our ongoing research in the area of 3D and multiview video coding with emphasis on video coding for 3D TV that exploits the potential of asymmetric view coding. We report the results of a user study to evaluate the quality of 3D video on autostereoscopic displays with H.264 coded stereo views. The study provides an understanding of the bounds of the asymmetric coding and additional bandwidth necessary to provide 3D TV services.

2. 3DTV CHALLENGES

While public fascination with 3D technology remains, the success of the 3D TV will depend on the user adoption and the willingness of service providers to deploy new services. The key factors that influence the success of 3D TV are: 1) effortless viewing 2) content availability 3) compatibility with the existing infrastructure and 4) role of multi-view video coding

2.1 Effortless Viewing

Viewer comfort is the most important factor that can make or break the 3DTV. Depth perception, new displays, and experiences all affect the viewing experiences. TV watching is different from watching movies and so is 3D TV experience from 3D movie experiences. Users tend to watch TV for longer periods of time with breaks in between while movies are on average less than 1 ½ hours long. Most of the users also tend to do other activities while watching TV. These factors rule out the use of glasses to experience 3DTV. Active eyewear or other types of glasses would be cumbersome and "unnatural" for users to wear continuously. Users will expect to watch some videos in 2D and the display technologies have to support this seamless mode switching.

There are a variety of 3D display technologies that still in their early stages of development [5]. The complexity and cost considerations make most of these technologies unsuitable for 3DTV applications at this time. Autostereoscopic displays are good candidates to provide

realistic 3D experiences in the home environment. Autostereoscopic displays based on lenticular imaging have been relatively successful. Display manufacturers such as Sharp and Philips have developed autostereoscopic LCD displays based on lenticular techniques at a reasonable cost. These first generation displays have restricted viewing angles and hence restrict the movement of users. This is expected to change with more viewing freedom as the technologies mature.

The effortless viewing experiences also depend on the 3D content and how the usage patterns evolve. There have been no long term studies on the impacts of 3DTV on user experiences, usage patterns, and most importantly, vision. One of the common complaints about 3D experiences is eye strain. Eye strain is attributed to a number of factors including the type of displays, content, and the constant vergence/accommodation adjustments in the human eye. There has been work on developing algorithms that attempt to reduce the strain on the eye when watching 3D videos [6]. The long term studies are necessary to understand the effects of 3D viewing on the eye strain and human vision. Users can be expected to adjust their viewing patterns based on their personal experiences.

2.2 Content Availability

Content availability is another factor that will influence the success of 3DTV. Content production is a difficult task and it is even more difficult to create 3D effects in the content that are pleasing to the eye. Since most of the available TV content today is available as 2D video, ability to convert 2D content to 3D will make large amounts of 3D content available. Dynamic digital depth (DDD) is a pioneer in this area and has been successful in developing 2D to 3D conversion [7]. While 2D to 3D conversion makes available large amounts of content in 3D quickly, such converted content was not originally created for 3D viewing and effectiveness of such content is unclear.

2.3 Infrastructure Compatibility

The 3DTV deployment should be seen as an evolution similar to the black and white to color TV migration. The compatibility with the existing systems cannot be compromised. The TV industry is currently beginning a slow migration from analog to digital and to HDTV. Introduction of 3DTV will perhaps begin after migration to digital/HDTV, which will put 3DTV deployment at least 5 years away.

Technology compatibility is another factor that has to be considered. The two main approaches to delivering 3D video are 1) stereo coding where the left and right views are encoded and 2) depth image based rendering (DIBR) where a single view and an associated depth map are transmitted to the receiver [8]. DIBR systems generate the left and right

views at the receiver based on the single view and the depth information. These two approaches have their advantages and disadvantages. However, from a production and compatibility point of view the stereo coding methods are more suitable as studios do not need new depth sensing cameras. Furthermore, the free viewpoint TV (FTV) based on multi-view video coding (MVC) is gaining momentum and this makes DBIR approaches unnecessary as the MVC would enable view synthesis to generate the left and right views necessary for the 3DTV.

2.4 The Deployment Costs and Benefits

The success of the 3D TV depends to a large extent on cost of deployment and additional revenues that will justify the investments. Hardware manufacturers will have an incentive as they can sell more hardware. However, the TV networks and content providers have a harder challenge. Advertisements in 3D are expected to be the initial drivers as it is generally believed that a 3D ad will make a better impression on the consumers. Broad studies supporting this general belief are yet to be conducted. Philips recently (April 2006) announced a 42″ flat panel 3D TV with a price tag of $18,000 intended for advertisements and demos by commercial establishments. Another recent development is a 3D advertising network launched in Thailand that is expected to be deployed in other countries. These developments are a pointing in the direction of advertiser driven 3D TV.

2.5 The Role of Multi-view Video Coding

Video coding technologies have matured sufficiently over the last few years to make possible new generation of video applications. Multi-view video coding has been receiving significant renewed attention among researchers and the industry [9]. Multi-view video coding (MVC) is also being standardized by the MPEG committee [10]. The goal of MVC is to allow coding of multiple camera views such that the user has the freedom of choosing the view point within a small field of view. The biggest challenge here is in developing compression technologies that can exploit the redundancies among the multiple views to achieve high a compression ratio.

The 3D video can be seen as a subset of multi-view video coding where the stereo pair required for the 3DTV can be synthesized at the receiver. Multi-view video systems require a multi-view decoder at the receiver and can use existing TV monitors. Since the multi-view video encoding is expected to be based on H.264 video coding, the incremental cost of a multi-view receiver is small. These factors perhaps will result in faster deployment of multi-view video compared to the 3D. With 3D being a special case of multi-view with two coded views, migration from multi-view to 3D will be an incremental step. Efficient

compression of the multiple views or stereo views continues to be important. For the special case of stereo views intended for 3D TV, the human visual may be exploited to improve the compression efficiency further.

3. EFFICIENT 3D CODING AND QUALITY EVALUATION

The human visual system has the property of binocular mixture where the left and right eye views are compared and combined to generate a single 3D percept. The binocular mixture makes possible encoding of left and right eye views at different bitrates. This asymmetric view coding has been exploited before to improve the compression efficiency [3,12]. The H.264 video coding used in our system is much more efficient than MPEG-2 and also has support for de-blocking that improves the perceptual quality of video. The effects of these improved compression algorithms on video quality cannot be understood from the past MPEG-2 based studies.

Figure 1 shows the general architecture of a 3D video system used to evaluate the efficiency of the asymmetric view coding. The stereo views are encoded at the sender by exploiting the large amount of redundancies among the views. We use H.264 as the core compression engine with inter-view prediction to increase compression efficiency [11]. The coded views are communicated to the receiver where the decoded views are rendered on an appropriate display. The 3D displays use a pair of coded views to display 3D video with depth perception.
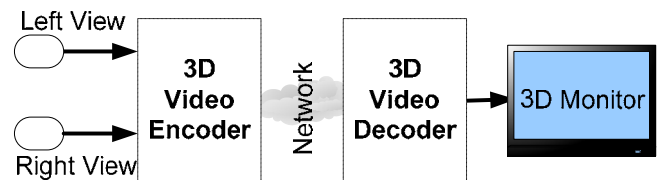

Figure 1. 3D/Multiview Video System

We used the Sharp LL-151-3D autostereoscopic display that uses lenticular imaging techniques to test the rendering of stereoscopic videos. The display is 15-inches, XGA resolution (1024 by 768 pixels). The perception of depth is achieved by a parallax barrier that diverts different patterns of light to the left and right eye. It should be noted that our player architecture accommodates a variety of formats for 3D playback and can be extended to include others.

We are currently conducting a large user study with to evaluate the impact of asymmetrically coded 3D views on the quality of the 3D video rendered on the Sharp autostereoscopic display. The goal of this study is to understand the bounds of asymmetric coding, relationship between the eye-dominance and 3D quality of asymmetrically coded video, and to understand the effects

of the H.264 coding artifacts. The results are reported based on the evaluations from 14 users that have evaluated the subjective quality.
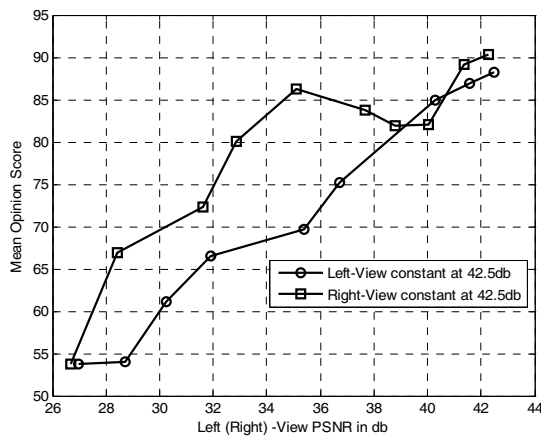

Figure 2. Mean Opinion Scores for Asymmetric View Coding

The subjective video quality evaluation was done using two stereo video sequences (Akko & Kayo and Ballroom). The left and right eye views are encoded at variety of qualities. The test sequence pairs were created with one view at a high quality and the other at lower qualities. Each 3D sequence is 10 seconds long with a 5 second mid-gray displayed between the sequences. The sequences at different qualities are shown in a random order on the 15-inch Sharp autostereoscopic 3D displays. The users evaluated the subjective quality on a scale of 1 (unacceptable) to 100 (excellent). Figure 2 shows the mean opinion scores (MOS) for the Akko and Kayo sequence with one view kept at a constant quality while the quality of the other view is varied. The figure shows a plot of MOS vs. the PSNR of the left (right) view while keeping the right (left) view PSNR constant at 42.5 dB.

The results show that asymmetric coding can be exploited in encoding 3D video sequences. The MOS is significantly higher when the right eye view is coded at a higher quality. This may be due the eye-dominance as most people are right-eye dominant and hence perceive a better quality when the right eye view is at a higher quality. The role of eye-dominance in visual perception is not well understood. A recent study found that eye dominance improves the performance of visual search tasks by perhaps aiding visual perception in binocular vision [13]. Further evaluation with additional sequences and careful user screening is necessary to fully validate our claim.

4. CONCLUSION

3DTV has the potential to become the next big evolution in television. The realization of this potential is influenced by the technological advances as well as the human factors. The key factors that influence the success of 3D TV are: 1) effortless viewing 2) content availability 3) compatibility with the existing infrastructure and 4) role of multi-view video coding. Autostereoscopic displays are likely to become the dominant mode of displaying 3D video. The autostereoscopic displays based on lenticular imaging techniques have advanced sufficiently to be able to large scale user tests for 3DTV services. Larger scale studies are necessary to gain proper understanding of the 3DTV experiences. The paper also presents a short overview of the 3D/multiview video coding system we are currently developing. Preliminary results of the asymmetric view coding experiments show that the H.264 can exploit the asymmetric coding well resulting in smaller incremental bandwidth for 3DTV services.

5. REFERENCES

[1] B.L. Tseng and D. Anastassiou, "Multi-Viewpoint Video Coding with MPEG-2 Compatibility," IEEE Trans. Circuits and Systems for Video Tech., Vol. 6, No. 4, Aug. 1996, pp. 414-419.

[2] A. Puri, R. V. Kollarits and B. G. Haskell, "Stereoscopic Video Compression Using Temporal Scalability," Proc. SPIE Visual Communications and Image Processing'95, Taiwan, May 1995.

[3] Lew B. Stelmach, W. James Tam, "Stereoscopic image coding: Effect of disparate image-quality in left- and right-eye views", Signal Processing: Image Communication, Vol. 14, pp.111-117, 1998.

[4] L.M.J. Meesters, W.A.Jsselsteijn, and P.J.H. Seuntiens, "A Survey for Perceptual Evaluations and Requirements of Three-Dimensional TV," IEEE Trans. Circuits Syst.Video Technol., Vol. 14, No. 3, March 2004, pp. 381-391.

[5] N.A. Dodgson, "Autostereoscopic 3D Displays," Computer, Volume 38, Issue 8, Aug. 2005 pp. 31 – 36.

[6] J.S. McVeigh, M.W. Siegel, and A.G. Jordan, "Algorithm for automated eye strain reduction in real stereoscopic images and sequences," Proc. SPIE Vol. 2657: Human Vision and Electronic Imaging pp. 307 - 316, 1996.

[7] P. Harman, J. Flack, S. Fox, and M. Dowley, "Rapid 2D to 3D conversion," Proc. SPIE, vol. 4660, pp. 78–86, 2002.

[8] C. Fehn, "A 3D-TV approach using depth-image-based rendering (DIBR)", Proc. of VIIP 03, Sept. 2003.

[9] ISO/IEC JTC1/SC29/WG11, "Survey of Algorithms used for Multi-view Video Coding (MVC)," MPEG Document MPEG2005/N6909, January 2005.

[10] ISO/IEC JTC1/SC29/WG11, "Call for Proposals on Multi-view Video Coding (MVC)," MPEG Document MPEG2005/N7327, July 2005.

[11] H. Kalva and B. Furht, "Hypercube based inter-view prediction for multi-view video coding," Proceedings of the 2nd Workshop on Immersive Communication and Broadcast Systems (ICOB), October 2005.

[12] Daniel V. Meegan, Lew B. Stelmach, and W. James Tam, "Unequal Weighting of Monocular Inputs in Binocular Combination: Implications for the Compression of Stereoscopic Imagery", Journal of Experimental Psychology: Applied, Vol 7(2) 143-153, Jun 2001.

[13] E. Shneor and S. Hochstein, "Effects of eye dominance in visual perception," International Congress Series, Volume 1282, Vision 2005, September 2005, pp. 719-723.