Joint Source and Channel Coding for 3D Video with Depth Image - Based Rendering

B. Kamolrat, W.A.C. Fernando, *Senior Member*, IEEE, M. Mrak and A. Kondoz, *Member*, IEEE

Abstract — *Following recent commercial availability of autostereoscopic 3D displays that allow 3D visual data to be viewed without the use of special headgear or glasses, it is anticipated that the applications of 3D video will increase rapidly in near future. In this paper we propose a joint source channel coding scheme for depth image-based rendering based 3D video coding. We considered different source and channel coding rates to find the optimum coding performance under a given channel bit rate for a WiMAX based communication channel. When the optimum bit allocation combination for color and depth image sequences are found, different protection levels have been considered for coding both image sequences. Finally, an optimum protection levels are proposed for the best video quality.*

Index Terms — 3D video, video coding, image coding.

I. INTRODUCTION

The research on stereoscopic video has received high interest over the past decade in order to provide viewers with more realistic vision than traditional 2D video. Instead of using left and right views to represent 3D video, a new technique, known as depth image-based rendering (DIBR) [1], represent 3D video based on a monoscopic video and associated per-pixel depth information (simply called color and depth maps). While the color consists of three components - Y, U and V as in the traditional video applications, the depth maps video uses only one component to store the depth information of objects within the scene related to the camera position. The advantage of such a scheme is, it can capture the stereoscopic sequences more easily compared to the traditional left and right view techniques. Even though, 3D video can provide more impressive video than conventional 2D video, in the past, there were many factors such as huge bandwidth requirement and the discomfort due to special devices needed to observe 3D effect preventing 3D video from success in commercial services.

With recent advances in both digital video compression and digital transmission technologies like H.264/AVC and WiMAX, the real time 3D video transmission application such as 3D television (3D TV) and 3D-conference can be realized. H.264/AVC is a video compression standard, also known as MPEG-4 Part 10 - Advanced Video Coding (AVC) [2]. It has many new features to achieve significant improvement in coding efficiency comparing to previous standards. Besides, it

provides more flexibility for application to a wide variety of network environments by applying a concept of Network Adaptation Layer (NAL). The high data rate and Quality of Service (QoS) provided by WiMAX technology make it attractive to multimedia applications, such as video telephony, video gaming, and video broadcasting. IEEE WiMAX 802.16e standard, [3], [4], also referred to as Wireless-MAN, is capable to support data rate up to 70 Mbps. The WiMAX standard features Low-Density-Parity-Check (LDPC) codes as an optional channel coding scheme. LDPC codes are linear block codes first introduced by Gallager in 1963 [5]. The advantage of LDPC codes is that they can approach the channel capacity over a very large code length. For example, for the code length of one million bits, the performance of LDPC is only 0.0045 dB lower from the maximum theoretical channel capacity [6].

Highly compressed H.264/AVC video bit-streams are very sensitive to channel errors [7]. To further improve the performance of 3D video transmission over wireless channel which is considered as bandwidth limited and error prone, joint source channel coding (JSCC) is an effective method to overcome such challenges [8]-[12]. The main concept of JSCC is that both the source coding and channel coding are adapted according to channel conditions in order to minimize the distortion. Distortion in video communication can be separated into two major types. The first type is the quantization distortion introduced by a lossy source encoding and the second type is caused by channel noise. These distortions are simply called "source distortion" and "channel distortion" denoted by D_S and D_C, respectively. The overall distortion, D_{Total}, is equal to $D_S + D_C$. A popular measure the distortion is a mean square error (MSE). Source distortion refers to MSE between the decoded frame from uncorrupted bit-stream at the transmitter and the original one. Channel distortion refers to the MSE between the decoded video frame at the receiver and the transmitted frame. The overall picture distortion at the receiver end can be defined as the MSE between the received video frame and the original one. In this paper we are trying to minimize the effect of these two distortions using a JSCC approach.

The rest of the paper is organized as follows. Some related work is briefly discussed in section 2 together with the basic concept of DIBR. The concept of bit stream organization in H.264/AVC is summarized in section 3. The 3D video transmission framework and experimental results for JSCC are presented in section 4. Finally, the conclusions are summarized in section 5.

Contributed Paper
Manuscript received April 15, 2008

0098 3063/08/$20.00 © 2008 IEEE

II. RELATED WORK

Since 3D TV is believed by many to be a new generation of TV broadcasting, the Moving Picture Expert Group (MPEG) developed a compression technology for stereoscopic video sequences as part of the MPEG-2 standard [13]. This approach is based on the multiview profile (MVP) which can be regarded as an extension of the temporal scalability tool. It encodes the left-eye view as a based layer and right-eye view as an enhancement layer. Even though this scheme provides backwards-compatibility to conventional 2D digital TV services, it has not been applied to any commercial services so far.

The latest 3D TV broadcasting approach has been proposed by the European Information Society Technologies (IST) project 'Advanced Three-Dimensional Television System Technologies' (ATTEST), [14]. The 3D representation used in this approach, based on monoscopic video and associated per-pixel depth information, is called DIBR. The concept of the ATTEST 3D TV can be separated into five sections: 3D content creation, 3D coding, transmission, virtual view synthesis and 3D display. At the coding part, monoscopic video is encoded by MPEG-2 and the depth information is encoded with more efficient codec such and MPEG-4. Subsequently, the encoded data streams are transmitted over digital video broadcasting (DVB) network. At the receiver, the received data streams are synthesized and displayed by a 3D display. The conventional digital TV can display 2D video while depth information is ignored.

The main advantage of DIBR technique compared to traditional representation of 3D video with the left-right views is that it provides high quality 3D video with smaller bandwidth required for transmission. This is because the depth map can be coded more efficiently than the two streams of monoscopic views if correlations and properties of the depth map are identified properly. Examples of color and depth map frames of the "Orbi" video test sequence are shown in Figure 1. While the color is stored in the same way as normal 2D video, the depth map is stored using only one component. When 8-bit component depth is used, 256 different depth values are associated to the pixels of the depth map.
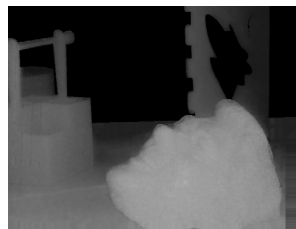
(a) color   (b) depth

Figure 1 Frame from the "Orbi" test sequence.

Conceptually, DIBR utilizes a depth frame to generate 2 virtual views from the same reference (original) view, one for the left eye and the other one for the right eye [15]. This process can be described by a following 2-step procedure.

Firstly, the original image points are re-projected into the 3D domain, utilizing the respective depth values. Thereafter, these intermediate space points are projected into the image plane of a virtual camera located at the required viewing position. This 2-step procedure is usually referred to as "3D image warping" in the computer graphics literature. The virtual views generation process is shown in Figure 2. In this process the original image points at locations (x, y) are transferred to new locations (x_L, y) and (x_R, y) for left and right view, respectively. This process is defined with:

$$x_L = x + \frac{\alpha_x \cdot t_c}{2}\left(\frac{1}{Z} - \frac{1}{Z_c}\right) \qquad (1)$$

$$x_R = x - \frac{\alpha_x \cdot t_c}{2}\left(\frac{1}{Z} - \frac{1}{Z_c}\right) \qquad (2)$$

where α_x is the focal length of the reference camera expressed in multiples of the pixel width and t_C is the distance between the left and right virtual cameras. Z_C is the convergence distance located at the zero parallax setting (ZPS) plane and Z denotes the depth value of each pixel in the reference view. Notes that the y component is constant since the virtual cameras used to capture the virtual views (left-right) are assumed to be located at the same horizontal plane.

The quality of virtual views depends on the quality of received color and depth map. In a video transmission system the impairment of synthesized frames depends on both the compression and transmission processes.
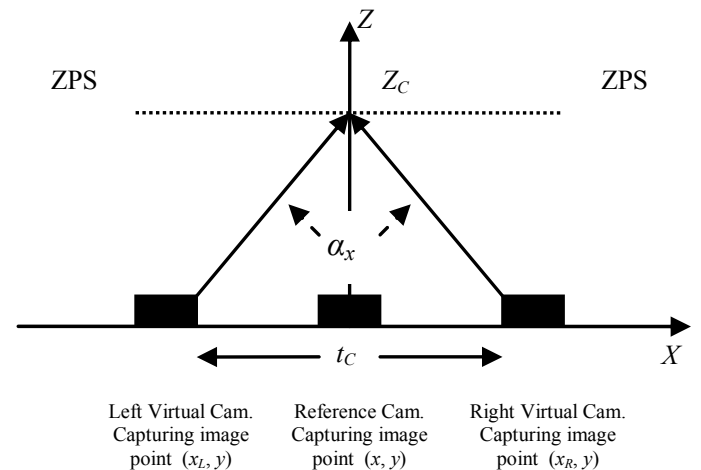

Figure 2 Virtual view generation in DIBR process.

III. VIDEO TRANSMISSION

Even though new communication technologies provide sufficient bit rate to support video communication applications, the bit rate will always be a scarce resource in wireless transmission environments due to physical bandwidth

and power limitations. To reduce the amount of transmitted data, the efficient video compression techniques are necessary. H.264/AVC is a new video coding standard expected to be used by many applications in the near future since it provides better performance than prior video coding standards like H.263 and MPEG-4. However, the highly compressed data is very vulnerable to channel errors. Most efficient video coding techniques are usually based on variable length coded (VLC) where short code words are assigned to the highly probable values and long code words to the less probable ones. Only one error bit can lead to the whole frame to be dropped due to desynchronization between the encoder and the decoder at the receiver. This phenomenon is known as error propagation. Therefore, to prevent error propagation through frame, the concept of slice is adopted where a frame is subdivided into several slices. If errors occur within a slice, the current slice is dropped and the decoder will search for the starting point of next slice and continue the decoding process. Therefore, in the noisy environment it is desirable to use frame segmentation into slices. On the other hand, each slice has a large overhead to indicate the starting point of each slice. The tradeoff between robustness to errors and overhead bits is needed to be carefully considered when designing the video transmission system. Generally, the decoded frame is used as a reference frame for the following frames; therefore the impairment in the reconstructed frame can propagate to successive frames. To reduce the channel errors, error control mechanisms are normally used and the popular techniques used in data transmission system are Automatic Repeat request (ARQ) and Forward Error Correction (FEC). Due to the long delay of ARQ, FEC has been commonly suggested for real-time application. However, FEC incurs constant transmission overhead even when the channel is loss free. To achieve the efficient bandwidth usage, especially over the limited bandwidth of the wireless link, the JSCC technique based on FEC has been introduced.

The concept of JSCC is that source and channel coding are jointly adjusted. This concept enables minimization of the distortions since the distortion in video transmission can be separated into two types: source distortion and channel distortion. The source distortion does not depend only on a given source-coding bit rate, but also on characteristics of the input videos and the data representation scheme which is employed by the coding algorithm. In case of H.264/AVC, the bit rate and video quality are adjusted by selection of suitable quantization parameter (QP). QP regulates strength of quantization and its value is selected in the rate-distortion optimization process. In order to meet specific bit rate requirements, QP also has to be adjusted depending on the underlying video content [16]. The channel distortion is due to transmitted data corrupted by noise over channel.

One of the design goals of H.264/AVC is to allow coded video to be integrated to all current protocol and multiplex architectures. The H.264/AVC consists of two conceptually different layers: Video Coding Layer (VCL) and Network Adaptation Layer (NAL) [17]. The VCL is designed to represent the content of the video data whereas the NAL is responsible for packaging and conveying data in a manner appropriate to transmission channels. Encoded video data and parameter sets are sent from the VCL to the NAL and encapsulated into units called NAL units. The format of NAL unit is shown in Figure 3.
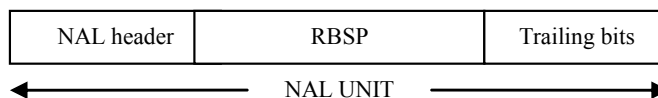
NAL header	RBSP	Trailing bits

← NAL UNIT →

Figure 3 NAL unit format.

A NAL unit consists of a 1-byte NAL header, a variable byte length Raw Byte Sequence Payload (RBSP) and a payload trailing bits. The NAL header is used to indicate the type of the NAL unit. Compressed video data and parameter sets are stored in the RBSP. The payload trailing bits are used to adjust the payload to become a multiple of bytes.

The interface between VCL and NAL is a slice layer. A slice is a group of macroblocks (MB) that does not need any information from other slices to be encoded or decoded. In video transmission, the order in which the NAL units have to be sent is constant. The first NAL unit to be sent is the Sequence Parameter Set (SPS) followed by the Picture Parameter Set (PPS). Both SPS and PPS include parameters which set in the encoder configuration for all pictures in the video sequence, for example: entropy coding mode flag, number of reference index, weighted prediction flag, picture width in MB, picture height in MB and number of reference frames. The next NAL unit is the Instantaneous Decoder Refresh (IDR). After receiving this NAL unit type all the buffers are deleted. Following the IDR frame is the NAL unit type slice. Figure 4 shows NAL units order in the case when slice mode 0 is selected and no data partitioning is used. As SPS and PPS are used for all pictures to reconstruct, they are the most sensitive to errors following with IDR. The NAL unit type slice is the least sensitive to errors.
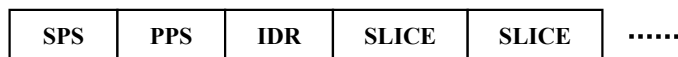
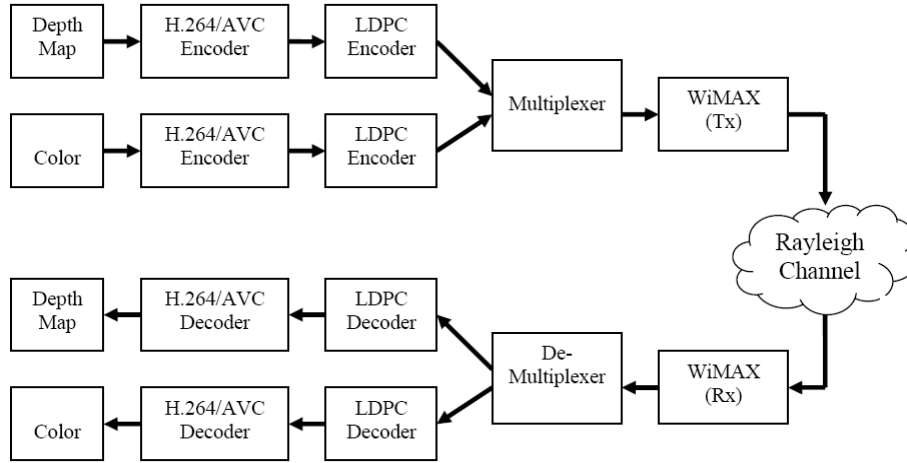| SPS	PPS	IDR	SLICE	SLICE	······

Figure 4 Order of NAL unit.

Figure 5 Simulation model.

IV. SIMULATION MODEL AND EXPERIMENTAL RESULTS

In this section, we explain the simulation model that we use to encode DIBR based 3D video data and the considered JSCC scheme for the video data. The overall system model is illustrated in Figure 5.

At the transmission side (Tx), color and depth map are separately compressed by H.264/AVC encoders and then protected by LDPC codes. The output bit streams are rearranged to get single output at the multiplexer. Subsequently, the output from multiplexer is transmitted by WiMAX over a Rayleigh fading channel. At the receiver (Rx), received data stream is separated back to 2 data streams before decoded by LDPC and H.264/AVC decoders, respectively. At the end of the process, color and depth map are reconstructed.

In this system the quality of received video is evaluated for left and right views that are synthesized from color and depth map. The average peak signal to noise ration (PSNR) of reconstructed left and right views is measured comparing to the original left and right views, Figure 6.

In order to maximize the system performance, the bit rate ratio between color and depth map has to be optimized. In the following section this process is explained.
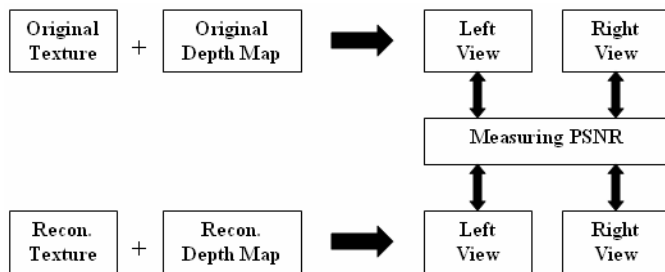

Figure 6 Average PSNR for a video sequence is obtained from the left and right views.

### A. Evaluation of bit rate allocation between color and depth map

To maximize the quality of 3D video when the total bit budget is fixed, optimized bit allocation is needed for both color and depth maps. In conventional approaches, where 8 bits per pixel are used for the depth map, a small change in the pixel value for the depth map does not have much impact on the overall 3D video quality.

A suitable bit rate ratio between color and depth map, when the bit budget is fixed, has been found experimentally. Color and depth map are encoded by H.264/AVC. Several 3D video sequences are considered. However, the "Orbi" and "Interview" video sequences are presented in this paper with 25 fps, CIF resolution (352×288), 8-bits per each colour and depth component and there are 30 P-frames between I-frame. The bit rate of depth map is varied between 10 - 90 % of the total bit rate and the remaining bit rate is assigned to the color. The total bit rate is varied from 200 kbps to 1 Mbps. Then the color and depth map are used to generate left and right videos. Finally, the average PSNR of left and right video output is measured by referencing to the left and right video generated from original color and depth map. The results are summarized in Figure 7. The average PSNR of left and right views are illustrated in Figure 7 a)-d), while the graphs in Figure 7 e) and f) show average result for both views.

The results suggests that if the total bit rate (color and depth map) is fixed, the highest decoding quality is obtained when the percentage of depth map bit rate to the total bit rate is about 20 %. This observation is used in the proposed system.

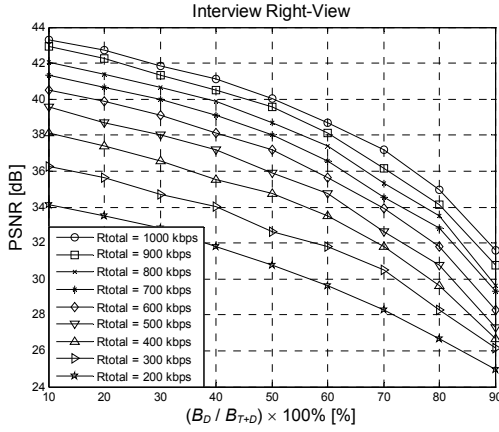### B. Simulation Results and Discussion

In this subsection the effect of channel coding on transmission of DIBR based 3D video is investigated.
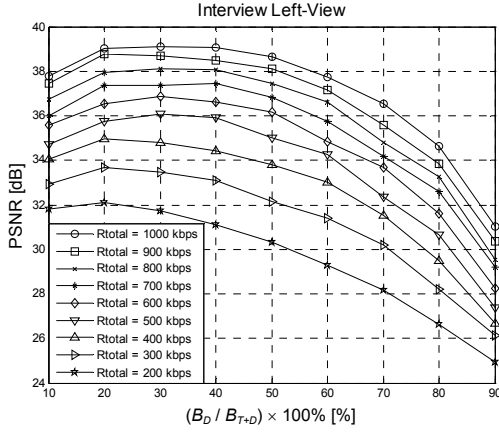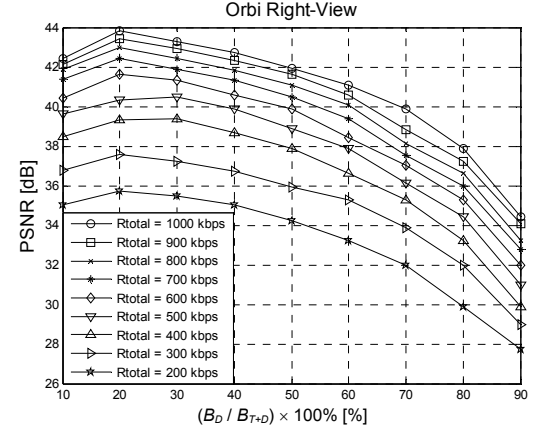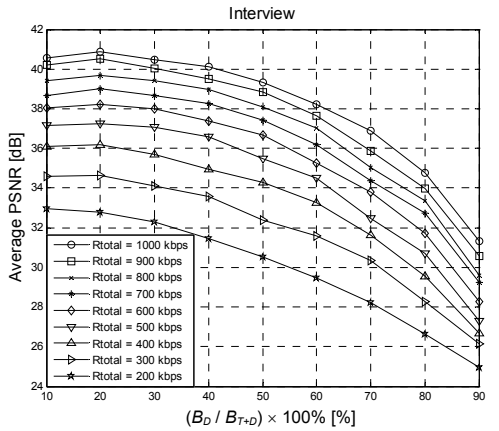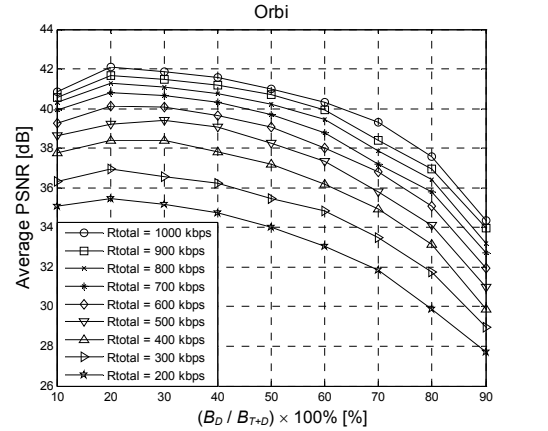
(a) Interview, left-view
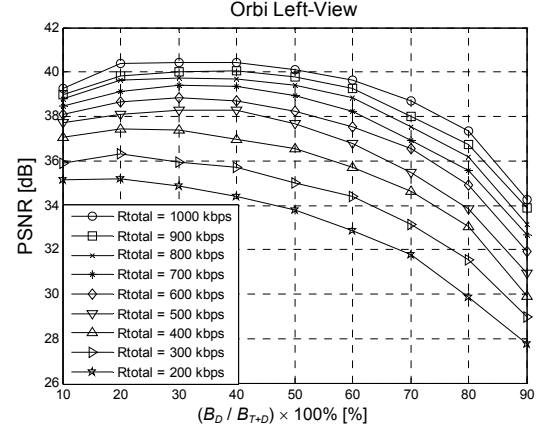
(b) Orbi, left-view

(c) Interview, right-view

(d) Orbi, right-view

(e) Interview, average results for left and right views

(f) Orbi, average results for left and right views

Figure 7 Decoding quality of reconstructed sequences for different coding conditions.

The total bit budget B_{total} for transmission is fixed to 2 Mbps. The coding rates of color and depth map are denoted as R_T and R_D, respectively. The relation between the total bit budget and the color bit rate B_T and the depth map bit rate B_D is:

$$\frac{B_T}{R_T} + \frac{B_D}{R_D} \approx B_{total} \qquad (3)$$

When the ratio of the color and depth map bit rates is set to the optimized values of the total source coding rate B_{T+D} the

total bit rate is:

$$\frac{0.8 \cdot B_{T+D}}{R_T} + \frac{0.2 \cdot B_{T+D}}{R_D} \approx B_{total} \qquad (4)$$

In this experiment, each frame of "Orbi" and "Interview" test sequences is subdivided into 18 slices. Color and depth map data streams are protected by the LDPC code with coding rates of 1/2, 2/3, 3/4 and 5/6. With all combinations, the 16 protection schemes are available. The WiMAX-LDPC is used

to convey color and depth map data streams over wireless channel which has the characteristic of Rayleigh fading. WiMAX frame size is set to 1056 bits. The data stream is modulated with 64-QAM. At the receiver, the data stream is demodulated by log-MAP algorithm and decoded by sum-product decoding algorithm with maximum iteration set to 50. Moreover, it is assumed that all errors can be perfectly detected and if an error is detected within slice after channel decoder, the whole slice is simply dropped. Finally, the error

concealment mode 1 of the reference software JM version 10 [18] is applied by simply copying the pixel area from the reference frame that is at the same location as the macroblock in the current image.

The results are expressed by the PSNR of reconstructed videos for different channel signal-to-noise ratio SNRs. The protection levels are indicated by coding rate of color and depth map where "X and Y" means R_T equals X and R_D equals Y.
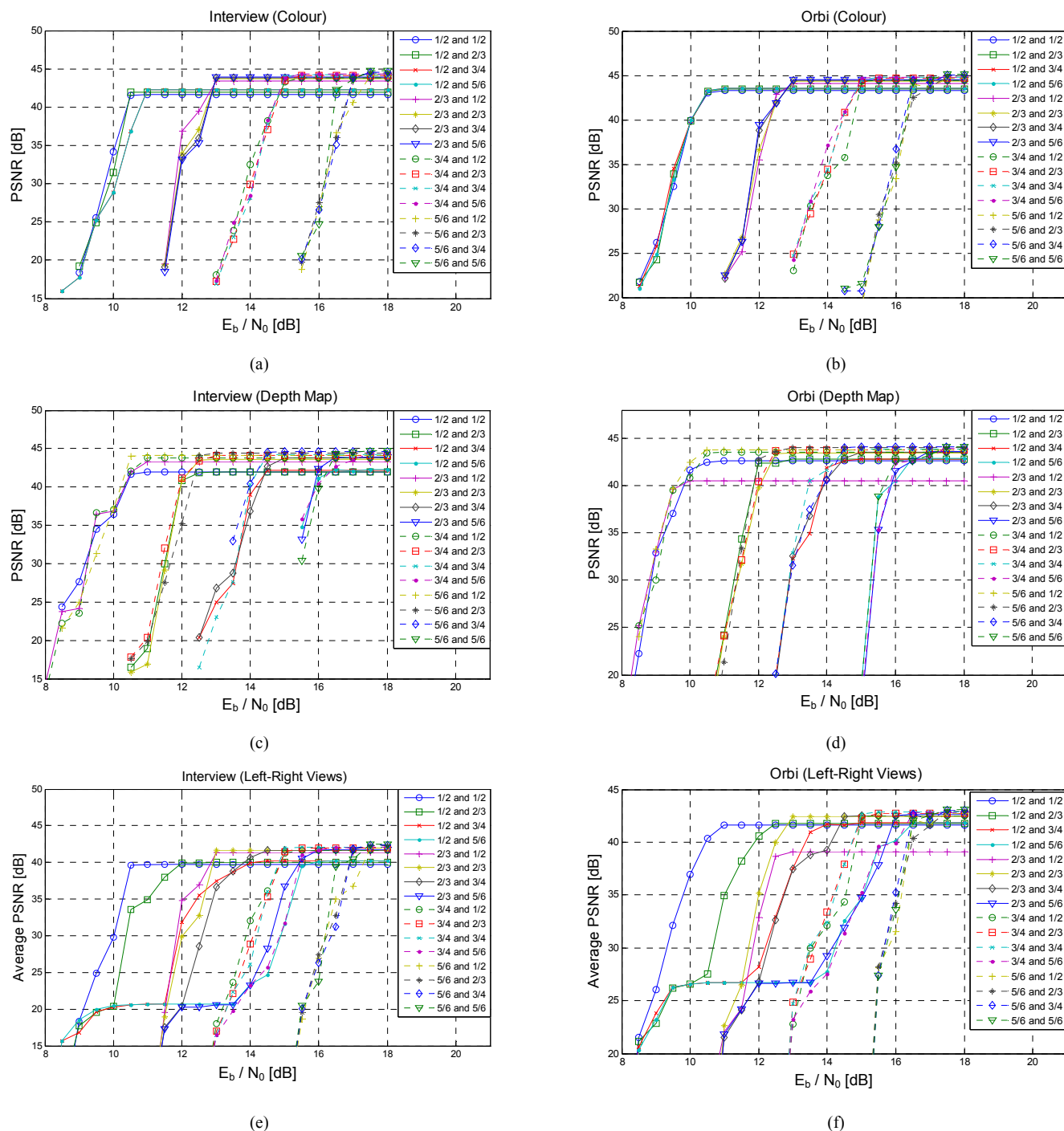

(a)

(b)

(c)

(d)

(e)

(f)

Figure 8 The quality of reconstructed 3D video bit streams when the levels of protection are varied.

Figure 8 presents the PSNR of color, depth and the average left and right views for different channel SNRs. The results suggest that the PSNR of reconstructed left and right views is dominated by the quality of color images. For example, in the case of "1/2 and 5/6" even though the depth map is not available until the SNR of 14 dB, the average PSNR of left and right views is about 26 dB (in SNR range of 10 dB to 13 dB). On the other hand, in case of "5/6 and 1/2" the average PSNR of reconstructed left and right views is very low or not available at SNR below 16 dB since no or very low quality of color images is available. The effect of this phenomenon on visual quality is presented in Figure 9 for the 15th frame of color, depth map, left-view and right-view In case of "1/2 and 5/6" at SNR 11 dB even though no depth map is available, the visual quality of synthesized left and right views is acceptable. However, when displaying them on 3D display, viewers cannot perceive the sense of depth due to the absence of depth information. In the case of "5/6 and 12" at SNR 15 dB, where the quality of color is low but the quality of depth map is excellent, the quality of reconstructed left and right views is very low. When displaying them on 3D display, even though the video quality is not good, viewer can perceive the sense of depth.

When the protection level is varied, the system performance can be summarized as follow. The "1/2 and 1/2" protection scheme has the best performance in low SNR region since it provides the strongest protection level. However, at high SNR region the "5/6 and 5/6" which provides the lowest protection, can obtain the highest PSNR due to low redundant bits. Therefore, to maximize the system performance both the channel condition and the source coding scheme have to be considered. Such JSCC approach maximizes the decoding quality since it enables selection of the appropriate protection level according to channel conditions.

In the final experiment the impact of the number of slices per frame to the overall system performance is investigated. A single frame of "1/2 and 1/2" coding pair is separated into 36, 18, 12 and 6 slices per frame. The experiment results are shown in Figure 10.

For low SNRs, the application of 36 slices per frame achieves the best result. For example, at SNR of 10 dB with source coding with 36 slices per frame the video quality is approximately 0.9 dB higher than in the case of 18 slices per frame, 3 dB higher than in the case of 12 slices per frame and 7 dB higher than in the case of 6 slices per frame. However, at error free region the PSNR values for source coding schemes with 36, 18, 12 and 6 slices per frame are saturated at 40.28 dB, 41.59 dB, 41.79 dB and 41.93 dB, respectively. It is clear that at high channel SNRs, the 6 slices per frame is the best option since it has the minimum rate of overhead bits. As mentioned before, the immunity to channel error can be enhanced by splitting a single frame into many slices with a tradeoff of redundant bits.
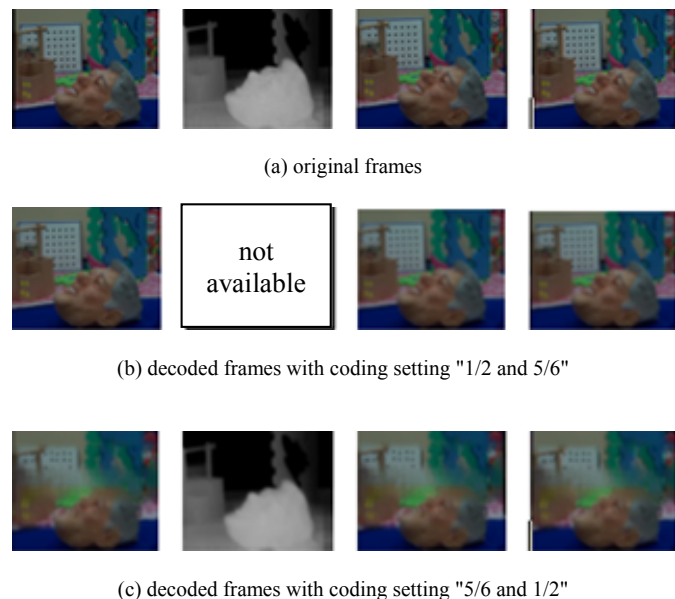

(a) original frames

(b) decoded frames with coding setting "1/2 and 5/6"

(c) decoded frames with coding setting "5/6 and 1/2"

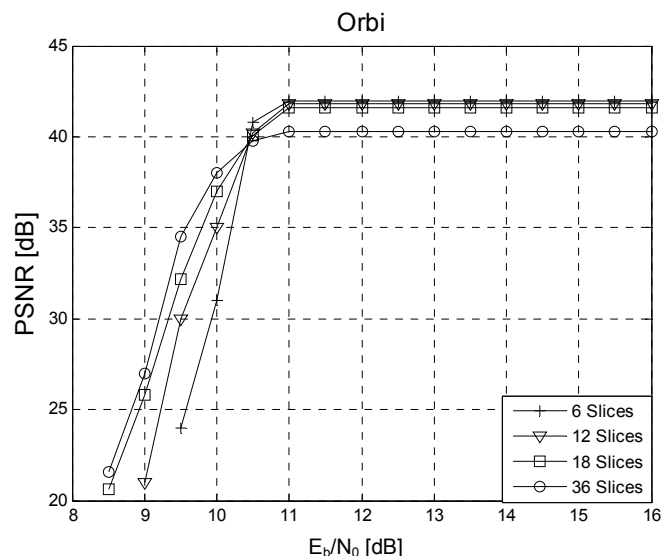**Figure 9 Decoded frames for different channel code rates.**

Figure 10 The performance of "1/2 and 1/2" when frame is spited into 36, 18, 12 and 6 slices.

V. CONCLUSIONS

In this paper, we consider optimized bit allocation and JSCC for DIBR based 3D video to achieve high compression efficiency and robustness to channel errors. Instead of using left and right views to represent 3D video, the monoscopic video and associated per-pixel depth information are used to gain better compression. Experimental results clearly show that the quality of 3D video is dominated by quality of the color. Besides, when they are encoded by H.264/AVC, the ratio of color to depth map bit rate is found to be suitable when depth map bit rate is about 20 % to the total source coding bit rate. This result comes up with the fact that the complexity of depth map is much lower than color leading to significantly redundant information that can be compressed.

When channel coding is used, results show that stronger channel coding at lower bit rates produce higher video quality and at high bit rates, weaker channel codes produces better picture quality. Results also suggest that quality of the depth image does not have a big impact on final picture quality and therefore, a lower channel code can be used to code the depth sequence.

REFERENCES

[1] Fehn, C., "A 3D-TV system based on video plus depth information," *Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on* , vol.2, no., pp. 1529-1533 Vol.2, 9-12 Nov. 2003

[2] "Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14 496-10 AVC," in Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVTG050, 2003.

[3] IEEE 802.16. Worldwide Interoperability for Microwave Access, www.wimaxforum.org.

[4] IEEE 802.16e. Air interface for fixed and mobile broadband wireless access systems. IEEE P802.16e/D12 Draft, Oct. 2005.

[5] R. G. Gallager, "Low-Density Parity-Check Codes," M.I.T. Press, Cambridge, Massachusetts, 1963.

[6] S. Y. Chung, G. D. Forney, T. J. Richardson, and R. Urbanke, "On the Design of Low-Density Parity-Check Codes within 0.0045 dB of the Shannon Limit," *IEEE Comm.*, vol. 5, pp. 58-60, February 2001.

[7] L. Hanzo, P. J. Cherriman, and J. Stereit, *Wireless Video Communications Second to Third Generation Systems and Beyond*, New York: IEEE PRESS, 2001.

[8] A. H. Murad and T. E. Fuja, "Joint Source-Channel Decoding of Variable-Length Encoded Sources," *in Proc. IEEE Information Theory Workshop 1998*, pp. 94-95, June 1998.

[9] M. J. Ruf and J. W. Modestino, "Rate-Distortion Performance for Joint Source and Channel Coding of images " *IEEE Trans. on Image Processing*, pp. 305-320, 1999.

[10] J. Cai and C. W. Chen, "Operational Rate-Distortion Design for Joint Source and Channel Coding over Noisy Channels," *Proc. IEEE Wireless Communications and Networking Conference 1999*, September 1999.

[11] M. Bystrom and J. W. Modestino, "Combined Source-Channel Coding Schemes for Video Transmission over an Additive White Gaussian Noise Channel," *IEEE in Comm.*, vol. 18, pp. 880-890, 2000.

[12] J. Xu, Q. Zhang, W. Zhu, X. G. Xia, and Y. Q. Zhang, "Optimal Joint Source-Channel Bit Allocation for MPEG-4 Fine Granularity Scalable Video over OFDM System," in *Proc. of the 2003 International Symposium on Circuits and Systems*, vol. 2, pp. 360-363, May 2003.

[13] H. Imaizumi and A. Luthra, "MPEG-2 Multiview Profile", *Three-Dimensional Television, Video, and Display Technologies*, pp. 169-181, 2002.

[14] W. A. IJsselsteijn, P. J. H. Seutiens, and L. M. J. Meesters, "State-of-the-Art in Human Factors and Quality Issues of Stereoscopic Broadcast Television," Technical Report D1, IST-2001-34396 ATTEST 2002.

[15] C. Fehn, R. Rarre, and S. Pastoor, "Interactive 3-DTV—Concepts and Key Technologies," Proc. IEEE, vol. 94, no.3, pp. 524-538, 2006

[16] D.-K. Kwon, M.-Y. Shen, and C. C. J. Kuo, "Rate Control for H.264 Video with Enhanced Rate and Distortion Models," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, pp. 517-529, 2007.

[17] A. Ksentini, A. Gueroui, and M. Naimi, "Improving H.264 video transmission in 802.11e EDCA,". *in Proc. of 14th International Conference on Computer Communications and Networks, ICCCN 2005*, pp. 381-386, 2005.

[18] H.264/AVC JM Refernce Software, http://iphome.hhi.de/suehring/tml/

B. Kamolrat received the BSc in Electrical Engineering from Royal Thai Naval Academy (RTNA), Thailand in 2000. He had served for the Royal Thai Navy (RTN) from 2000 to 2003 before he got a scholarship from the RTN to further his study at the University of Newcastle upon Tyne, Uk in 2004 and received MSc degree (Distinction) in Communication and Signal Processing in 2005. Currently he is a research student completing his PhD in the Centre for Communication Systems Research (CCSR) in the University of Surrey, UK. His research interests are in 3D video coding and video communications.

W.A.C. Fernando received the B.Sc. Engineering degree (First class) in Electronic and Telecommunications Engineering from the University of Moratuwa, Sri Lanka in 1995 and the MEng degree (Distinction) in Telecommunications from Asian Institute of Technology (AIT), Bangkok, Thailand in 1997. He completed his PhD at the Department of Electrical and Electronic Engineering, University of Bristol, UK in February 2001. Currently, he is a senior lecture in signal processing at the University of Surrey, UK. Prior to that, he was a senior lecturer in Brunel University, UK and an assistant professor in AIT. His current research interests include Distribute Video Coding (DVC), 3D video coding, intelligent video encoding for wireless communications, OFDM and CDMA for wireless channels, channel coding and modulation schemes for wireless channels. He has published more than 150 international papers on these areas. He is a senior member of IEEE and a fellow of the HEA, UK. He is also a member of the EPSRC College.

M. Mrak received the BSc and MSc degrees in electrical engineering from University of Zagreb, Croatia, in 2001 and 2003, respectively, and the PhD degree from Queen Mary University of London, UK, in 2006. From 2001 to 2003 she was a research assistant at the University of Zagreb. In 2002 she was a visitor in the Image Processing Department, Heinrich-Hertz-Institut, Germany. From 2004 to 2007 she was a member of Multimedia and Vision Research Group at Queen Mary University of London where she was a work-package leader at EU FP6 integrated project aceMedia. In 2007 she joined the Centre for Communication System Research, University of Surrey, UK, as a research fellow. Her main research interests are visual data processing, video analysis, video coding and adaptive transforms for image coding. She acts as reviewer for several international conferences and journals, and she ia author of more than 50 publications. Dr Mrak is the recipient of the principal's award for best student work and has received two awards and the bronze medal from the Faculty of Electrical Engineering and Computing in Zagreb for the exemplary success during study. She was awarded a scholarship from DAAD in 2002. She has been a member of the organizing/technical committee of several conferences, including IET VIE 2008 and CBMI 2008, co-chair of workshop on "Scalable coded media beyond compression" and she served as session chair and organiser of conference special sessions.

A. Kondoz studied B.Sc. (Hons.) degree in engineering, and M.Sc. degree in telematics in 1983 and 1984, respectively, before receiving his Ph.D. degree in communication in 1986. He became a Lecturer in 1988, a Reader in 1995, and then in 1996, a Professor in Multimedia Communication Systems and deputy director of Center for Communication Systems Research (CCSR), University of Surrey, Guildford, U.K. He has over 250 publications, including two books on low-bit-rate speech coding and several book chapters, and seven patents. He has graduated more than 50 Ph.D. students in the areas of speech/image and signal processing and wireless multimedia communications, and has been a consultant for major wireless media terminal developers and manufacturers. Prof. Kondoz has been awarded several prizes, the most significant of which are The Royal Television Societies' Communications Innovation Award and The IEE Benefactors Premium Award. Prof. Kondoz is also a director of Mulsys Ltd. a University of Surrey spin-off company marketing worlds first secure GSM communication system through the GSM voice channel.