Comparison of Stereo Video Coding Support in MPEG-4 MAC, H.264/AVC and H.264/SVC

C.T.E.R. Hewage, H.A. Karim, S. Worrall, S. Dogan, A.M. Kondoz

Centre for Communication Systems Research
University of Surrey
Guildford, Surrey, GU2 7XH, U.K.
{e.thushara, h.karim, s.worrall, s.dogan, a.kondoz} @surrey.ac.uk

Keywords: Stereoscopic video, Stereoscopic video coding, Depth-Image-Based-Rendering, Scalable video coding

Abstract

In the near future, certain 2D (2 Dimensional) video application scenarios are likely be replaced by 3D (3 Dimensional) video, in order to achieve a more involving and immersive representation of visual information, and to provide more natural methods of communication. Stereoscopic video is the simplest form of 3D video, which renders two slightly different views of a scene to perceive the depth information. Currently much research is being carried out in this area including stereo video capture, coding, transmission, and display technologies. This paper presents a review of some of the most prominent methods for coding stereoscopic colour and depth video, and investigates their respective coding efficiencies. This paper also proposes a new configuration for encoding colour and depth stereoscopic video, based on the H.264/SVC video coding standard. The rate-distortion performance comparison of the colour and depth video obtained using the depth-range cameras vs. left and right video is analyzed at low bitrates using the H.264/SVC configuration. The performances are compared in terms of coding efficiency and implementation factors with the MPEG-4 MAC and H.264/AVC configurations. It was found out that the configuration proposed based on H.264/SVC performs similar to H.264/AVC and outperforms the MPEG-4 MAC based configuration in terms of coding efficiency. In terms of implementation factors the proposed configuration provides wider flexibility compared to all other configurations.

1 Introduction

According to the classification of MPEG-3DAV (Motion Picture Expert Group-3D Audio Visual), three scene representations of 3D video have been identified, namely Omni-directional (panoramic) video, interactive multiple view video (free viewpoint video) and interactive stereo video [1]. Stereoscopic video is the simplest form of 3D video and can be easily adapted in communication applications with the support of existing video technologies. Stereoscopic video renders two views for each eye, which facilitates depth perception of the 3D scene. This paper investigates the stereoscopic video coding support possible with current video codecs in order to enhance and scale existing video applications into stereoscopic video applications. Furthermore, it elaborates on other significant features (e.g. scalability, backward compatibility, etc.) available in the respective video codecs, which can be used to support stereoscopic application scenarios.

The rest of the paper consists of the following sections: Section 2 presents some background to stereoscopic video, including capture of stereo video; Section 3 elaborates on stereoscopic video coding support in MPEG4-MAC (Multiple Auxiliary Component) and H.264/AVC (Advance Video Coding); The proposed configuration for stereoscopic video based on H.264/SVC (Scalable Video Coding) video coding standards are also discussed in section 3; The experimental setup, results obtained and discussion of results are given in Section 4; Section 5 concludes the paper.

2 Stereoscopic video capture

(a) (b)

Figure 1: Interview sequence (a). Colour image (b). Per-pixel depth image

There are several techniques to generate stereoscopic content including dual camera configuration, 3D/depth-range cameras and 2D-to-3D conversion algorithms [10]. Stereoscopic capture using a stereo camera pair is the simplest and most cost effective way to obtain stereo video, compared to other technologies available in the literature. The latest depth-range camera generates a colour image and a per-pixel depth image of a scene as shown in Figure 1. This depth image with its corresponding colour image can be used to generate two virtual views for the left and right eye using the Depth-Image-

Based Rendering (DIBR) method described in [6]. Equation (1) is used to generate the two virtual views.

$$P_{pix} = -x_B \frac{N_{pix}}{D}\left[\frac{m}{255}(k_{near} + k_{far}) - k_{far}\right] \qquad (1)$$

Where, N_{pix} and x_B are the number of horizontal pixels of the display and eye separation respectively. The depth value of the image is represented by the N-bit value m. k_{near} and k_{far} specify the range of the depth information respectively behind and in front of the picture, relative to the screen width N_{pix}.

The advantages and disadvantages associated with depth-range cameras in comparison to stereo camera pairs are given in [7]. Currently ISO/IEC 23002-3 (MPEG-C part 3) are working on standardization of video plus depth image solutions in order to provide: interoperability of the content, flexibility regarding transport and compression techniques, display independence and ease of integration [3]. The ATTEST (Advanced Three-Dimensional Television System Technologies) project consortium is working on 3D-TV broadcast technologies using colour-depth sequences as the main source of 3D video [9]. The experiments presented in this paper use the test sequences ('Orbi' and 'Interview') obtained using the depth-range camera. The sequences can be efficiently compressed using MPEG-4 and H.264/AVC [5]. The rest of the paper will focus on coding of colour-depth sequences.

3 Stereoscopic video coding

The adaptability of MPEG-4 MAC, H.264/AVC and H.264/SVC video coding standards for stereo video coding is analyzed in this section. Furthermore it elaborates on other features which can be exploited to facilitate stereoscopic video applications. Stereoscopic video support in the MPEG family of standards was discussed in [2].

3.1 MPEG-4 MAC

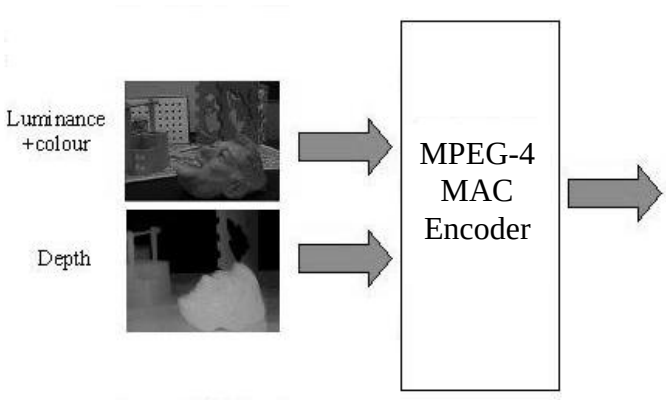


Figure 2: MPEG-4 MAC architecture

3DAV, an ad-hoc group of MPEG has identified MPEG-4 MAC (Multiple Auxiliary Component) for encoding colour and depth stereoscopic video content. MPEG-4 MAC allows the encoding of auxiliary components in addition to the Y, U and V components present in 2D video. The depth/disparity map can be used as one of its auxiliary component [8]. This MPEG-4 MAC configuration for stereoscopic video coding is as shown in Figure 2.

MAC produces a one-stream output. This one-stream approach facilitates end-to-end video communication applications without a system level modification (avoid multiplexing and de-multiplexing stages for different streams) and compliance to ongoing standardization work of ISO/IEC 23002-3 (MPEG-C part 3). In this paper, the rate distortion of MPEG-4 MAC coded 2D and depth stereoscopic video was compared against results obtained with the H.264/AVC and H.264/SVC video coding standards.

3.2 MPEG-4 Part 10/H.264-AVC

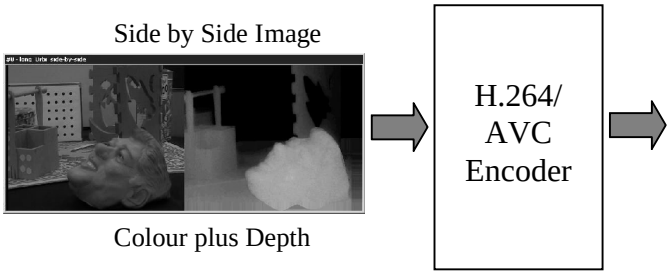


Figure 3: H.264/AVC architecture

The H.264/AVC codec is a video coding standard which provides high compression efficiency, network friendliness and a wide range of error resilient features [12]. Much of the previous research into using H.264 for stereoscopic video uses left-right view disparity estimation, rather than colour-depth coding. For example, an extended H.264 video codec for stereoscopic video, exploiting spatial, temporal, disparity and world-line correlation is described in [4].

There are two main approaches to coding colour-depth video using H.264. The first is to encode the respective colour and depth video sequences using two parallel H.264 codec implementations. However, a single encoder output is advantageous, compared to using two bit-streams, as it will not affect the end-to-end communication chain (i.e. no extra signalling is needed to accommodate the additional bitstream). Hence, one possible approach is to encode stereoscopic video after some source processing. At the source the colour and depth images are combined into a single source to serve as the input for the H.264/AVC codec. For example, side by side colour-depth images (see Figure 3), or interlaced colour-depth images. But this approach lacks backward compatibility and flexibility for stereoscopic video communication applications. A study of stereoscopic video coding based on source processing is given in [11]. This paper analyzes the coding efficiency of colour and depth image sequences using H.264/AVC, where colour and depth sequences were combined to form a side by side image sequence. The H.264/SVC single layer configuration was used to obtain H.264/AVC results as the H.264/SVC base layer is compatible with H.264/AVC, and to avoid differences

arising from the different software implementations of H.264/AVC and H.264/SVC. The proposed configuration based on H.264/AVC is shown in Figure 3.

3.3 H.264-SVC

H.264/SVC developed by JVT is a proposed video coding standard, which supports spatial, temporal and quality scalability for video [9]. This paper proposes a configuration to code stereoscopic video sequences based on the layered architecture proposed in H.264/SVC. The colour and depth/disparity image sequences are coded as the base and enhancement layers respectively as shown in Figure 4. As the base layer is compatible for H.264/AVC decoding, users with a H.264/AVC decoder will be able to decode the colour image, whereas users with an SVC decoder will be able to decode the depth/disparity image and will experience stereoscopic video. This backward compatibility feature of H.264/SVC can be used to enhance or scale existing video applications into stereoscopic video applications. Furthermore it can be used to exploit asymmetric coding of the left and right images as H.264/SVC supports a range of temporal, scalable and quality scalable layers. The Inter layer prediction mode can be selected based on the correlation between the colour and depth images. The results presented here make use of the adaptive inter layer prediction option in the JSVM (Joint Scalable Video Model) software. This configuration complies to the ongoing standardization work of ISO/IEC 23002-3 (MPEG-C part 3). Section 4 discusses the coding efficiency of the proposed configuration of H.264/SVC compared to MPEG-4 MAC and H.264/AVC.

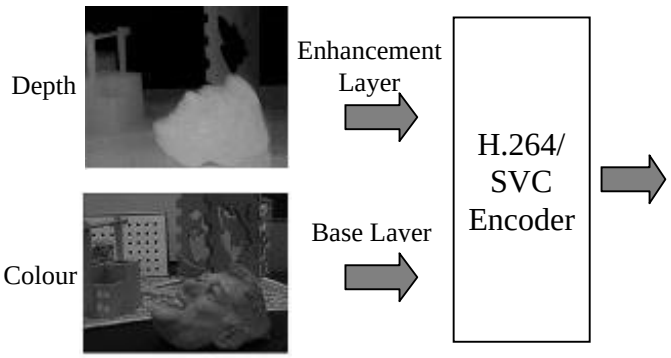


Figure 4: H.264/SVC architecture

4 Results and discussion

Two 2D (colour) and depth image sequences are used to obtain the rate distortion results, namely 'Orbi' and 'Interview'. 'Orbi' is a very complex sequence with camera movement and multiple objects, whereas 'Interview' is a sequence captured with a static camera and featuring a stationary background. The tests are carried out using CIF format (352x288) video. The image sequences are encoded using the three proposed configurations based on existing video coding standards, MPEG-4 MAC, MPEG-4 Part 10/H.264-AVC (using the H.264/SVC single layer coding) and H.264-SVC. The basic encoding parameters are: 300 frames, IPPP… sequence format, 30 frames/s original frame rate, a single reference frame, variable length coding (VLC), 16 pixel search range and no error resilience. The Quantisation parameter (QP) in the configuration file is varied to obtain the bitrate range shown in the rate-distortion curves. For the H.264/SVC double layer configuration the same QP was used at both layers. The rate distortion curves show the image quality measured in PSNR (Peak-Signal-to-Noise Ratio) against the resulting bitrate.

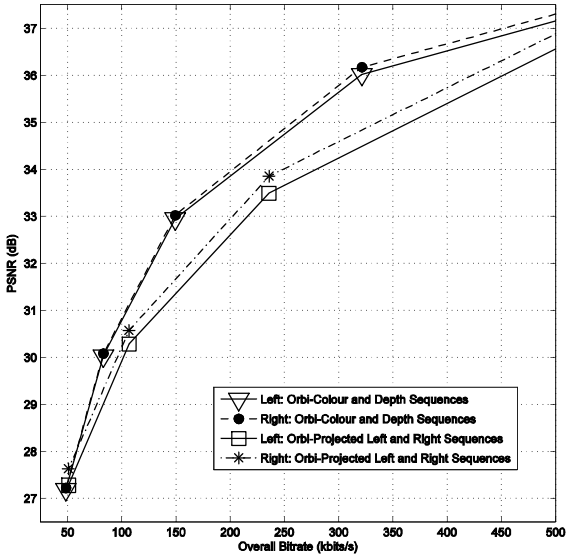


Figure 5: Rate-Distortion curves for 'Orbi' (using colour and depth sequences and projected left and right sequences)

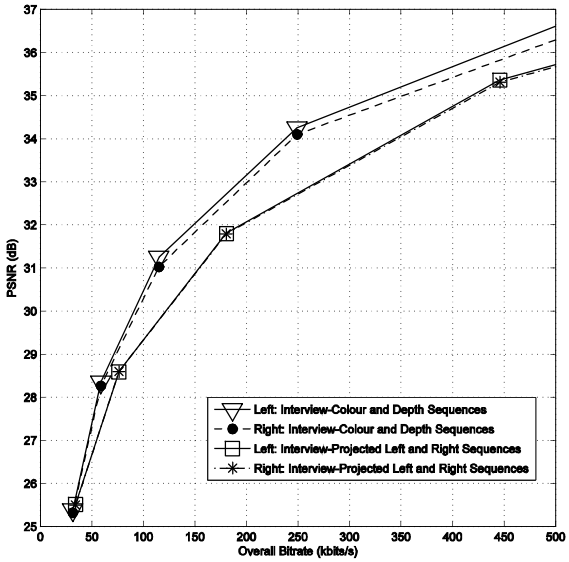


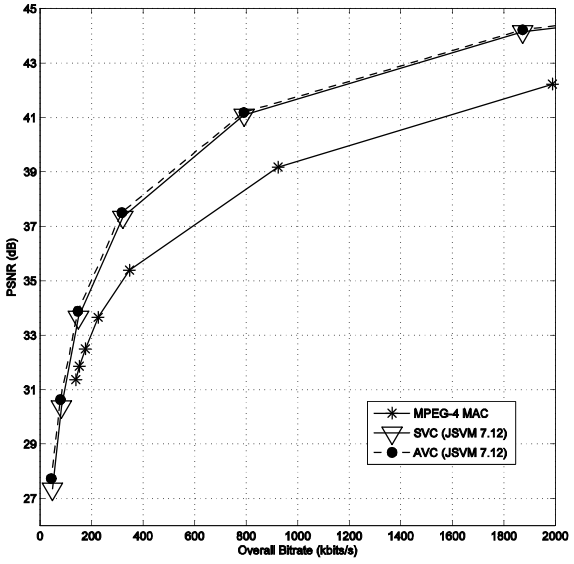
Figure 6: Rate-Distortion curves for 'Interview' (using colour and depth sequences and projected left and right sequences)

The initial test was carried-out to compare the Rate-Distortion performance of colour and depth image sequences vs. left and right image sequences using the H.264/SVC configuration. In

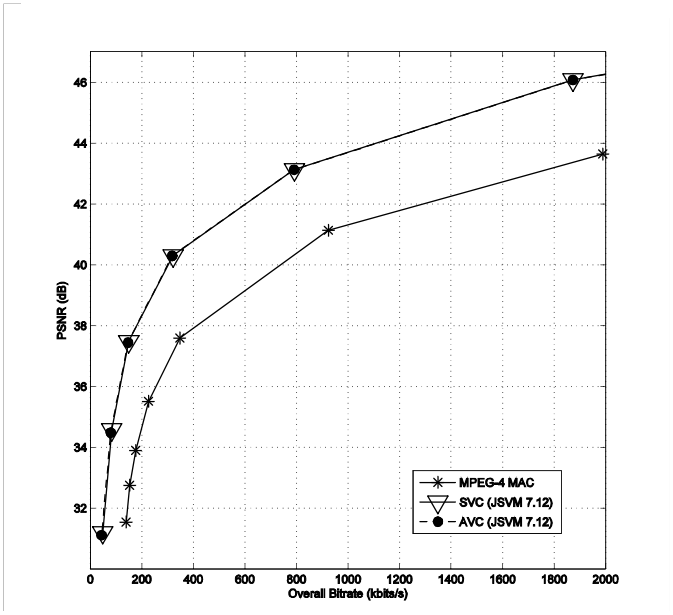
order to produce left and right image sequences the 'Orbi' and 'Interview' sequences were projected in to virtual left and right image sequences using Equation (1) and were coded as the base and enhancement layers. The produced left and right sequences representing video generated using a stereo camera pair. The coded colour and depth image sequences at the base and enhancement layers were converted to virtual left and right video to be compared with the coded left and right video. Figures 5 and 6 show the R-D performance for the 'Orbi' and 'Interview' sequences respectively at low bitrates up to an overall bitrate of 500kbits/s.

According to Figures 5 and 6 at, low bitrates both 'Orbi' and 'Interview' sequences demonstrate better performance for colour and depth image coding using H.264/SVC than coding projected left and right view video. With high QP values H.264/AVC coded depth images are of high image quality compared with the colour images, which results in high quality left and right image sequences. The amount of disparity between the projected left and right images does not allow the use of adaptive inter layer prediction more effectively between the base and enhancement layers.

Figure 7 shows the rate-distortion curves for 'Orbi' colour and depth sequences based on MPEG-4-MAC, H.264/AVC and H.264/SVC configurations. The rate-distortion curves of 'Interview' colour and depth sequences are shown in Figure 8. All of the results are plotted against the overall bitrate (output bitrate of the codec), which includes all overhead bits, texture, colour, motion vectors and depth. In order to highlight R-D performance at a range of bitrates, the final bitrate is shown from 0kbits/s to 2Mbits/s. H.264/AVC coded stereoscopic video sequences were separated into colour and depth sequences in order to calculate the PSNR with respect to their original colour and depth sequences.



(a)



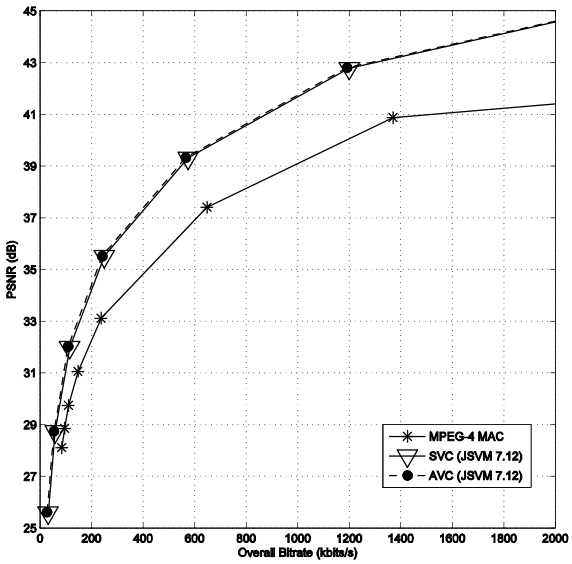
(b)

Figure 7: Rate-Distortion curves for 'Orbi' sequence (a) Colour image sequence (b) Depth image sequence

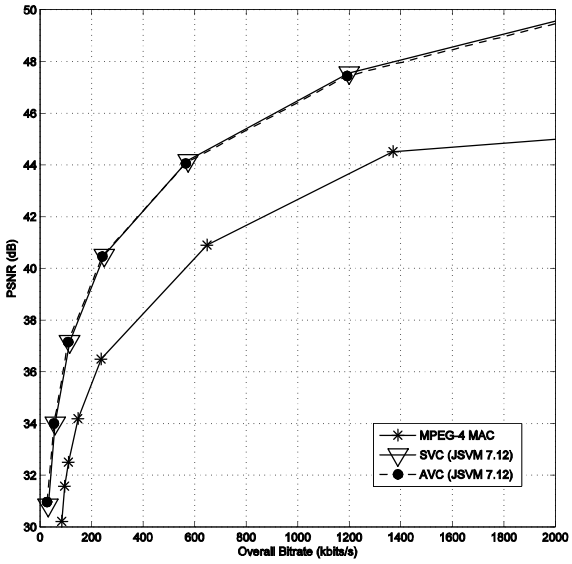
Rate-Distortion curves for both 'Orbi' and 'Interview' sequences show that the proposed configuration with H.264/SVC performed similarly to the H.264/AVC configuration, and outperformed the MPEG-4 MAC based configuration at all bitrates. The H.264/SVC configuration was unable to achieve high compression efficiency compared to the H.264/AVC configuration due to the negligible usage of inter layer prediction between the base and enhancement layers. The correlation between the colour and depth sequences is small. Hence the H.264/SVC configuration could not gain any advantage from inter layer prediction and performed similarly to the H.264/AVC configuration. However the flexibility of the H.264/SVC configuration for stereoscopic video coding, such as asymmetric coding support (temporal, spatial and quality scalability for depth image sequences) and backward compatibility, facilitates end-to-end stereoscopic video communication chain to a greater extent. The flexible macroblock (MB) sizes and skipped MB features available in H.264/AVC standard have helped to provide high image quality for the H.264/AVC and H.264/SVC based configurations compared to the MPEG-4 MAC based configuration at all bitrates. Furthermore, it can be observed that the configurations based on the H.264/AVC and H.264/SVC provide reasonable image quality at very low overall bitrates compared to the MPEG-4 MAC configuration.

The H.264/AVC and H.264/SVC configurations outperform the MPEG-4 MAC configuration by a considerable margin for depth image quality at all overall bitrates (see Figure 7.b and 8.b). The smoothness of the depth image and the availability of constant chrominance (U and V) planes help to achieve superior image quality for the H.264/AVC and H.264/SVC configurations. The depth images are highly compressed using flexible MB sizes and skipped MB modes

available in H.264/AVC and H.264/SVC configurations. This is more visible in the 'Interview' sequence, which has less motion and stationary background. The analysis of depth map compression using H.264/AVC against other video coding standards given in [5] concludes H.264/AVC outperforms other MPEG video coding standards.



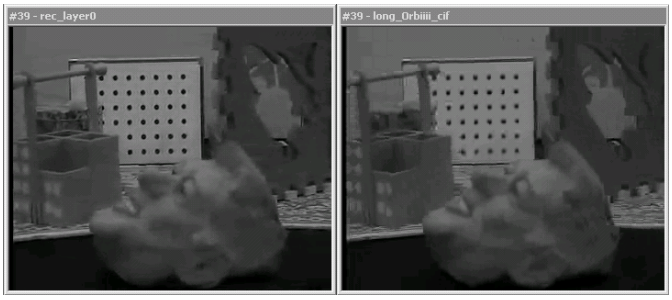
(a)



(b)

Figure 8: Rate-Distortion curves for 'Interview' sequence (a) Colour image sequence (b) Depth image sequence

The subjective image qualities of the 'Orbi' colour and depth image sequences are shown in Figures 9 and 10 respectively. The image sequences were obtained at an overall bitrate of 150kbits/s using the configurations based on H.264/SVC and MPEG-4 MAC. According to Figure 9, subjective quality of the H.264/SVC coded colour image is better compared to the MPEG-4 MAC coded colour image. This is more visible in the depth image sequences as shown in Figure 10. The H.264/SVC coded depth image sequence demonstrates a sharp and superior image quality compared to MPEG-4 MAC coded depth image sequence at the given low bitrate of 150kbits/s



(a) (b)

Figure 9: Subjective image quality of the 'Orbi' colour sequence at an overall bitrate of 150kbits/s (a). H.264/SVC configuration (b). MPEG-4 MAC configuration

(a) (b)

Figure 10: Subjective image quality of the 'Orbi' depth sequence at an overall bitrate of 150kbits/s (a). H.264/SVC configuration (b). MPEG-4 MAC configuration

	'Orbi' PSNR (dB)		'Interview' PSNR (dB)	
	Colour	Depth	Colour	Depth
H.264/SVC	34.74	38.31	34.22	39.29
MPEG-4 MAC	33.05	34.68	32.25	35.52
H.264/AVC	35.01	38.33	34.41	39.41

Table 1: Image quality at overall bitrate of 200 kbits/s

Table 1 shows the image quality of both sequences at an overall bitrate of 200kbits/s. This shows that the proposed stereoscopic video coding configuration based on H.264/SVC provides reasonable quality compared to the MPEG-4 MAC configuration at overall bitrates as low as 200kbits/s. The high performance and flexible features (backward compatibility and temporal, spatial and quality scalability) associated with H.264/SVC can be used to convert low bitrate video

applications into stereoscopic video applications. Furthermore, at an overall bitrate of 200kbits/s the 'Orbi' depth image sequence can be coded at 49% of the 'Orbi' colour image bitrate using the propose configuration based on H.264/SVC. The depth image bitrate can be further reduced by using a high QP value or reduced temporal or spatial scalability for the enhancement layer (depth image) without affecting the perceptual quality of stereoscopic video. In order to avoid occlusion problems associated with the DIBR method several layers of depth images (LDI) can be coded using this proposed H.264/SVC configuration.

4 Conclusion

This paper analyzes the rate-distortion performance of stereoscopic video using three configurations based on MPEG-4 MAC, H.264/AVC and H.264/SVC. The proposed H.264/SVC configuration based on the layered architecture performs similarly to the configuration based on H.264/AVC and outperforms the configuration based on MPEG-4 MAC at all bitrates in terms of objective and subjective quality. Furthermore the configuration based on H.264/SVC produces high quality stereoscopic video using colour and depth sequences (obtained from the depth–range camera) compared to the virtual left and right sequences produced from the same colour and depth sequences at low bitrates. This configuration can be used to explore the possibilities of replacing existing video communication applications by stereoscopic video applications. This will be further facilitated by the backwardly compatible nature of H.264/SVC, which supports base layer decoding using H.264/AVC decoders. The scalable layers present in H.264/SVC can be exploited to encode multi-resolution video, where one view is coded differently compared to the other view. For example, the depth image can be coded as several quality layers, in order to eliminate the occlusion problem present in DIBR, and gain optimum image quality output based on user capabilities. At a given bitrate, depth image sequences achieve high image quality in both H.264/AVC and H.264/SVC configurations. Hence it can be stated that depth image sequences can be compressed efficiently using H.264 based configurations.

Acknowledgements

The work presented was developed within VISNET II, a European Network of Excellence (http://www.visnet-noe.org), funded under the European Commission IST FP6 programme.

References

[1] A. Smolic and H. Kimata, "Applications and Requirements for 3DAV", *ISO/IEC JTC1/SC29/WG11 W5877*, (July 2003).

[2] A.Puri, R.V. Kollarits, B.G. Haskell, "Basics of stereoscopic video, new compression results with MPEG-2 and a proposal for MPEG-4", *Signal Processing: Image communication*, **Vol. 10**, pp. 201-234, (1997).

[3] Arnaud Bourge and Christoph Fehn, "White paper on ISO/IEC 23002-3 Auxiliary Video Data Representation", *ISO/IEC JTC1/SC29/WG11 N8039*, (April 2006).

[4] Balasubramaniyam Balamuralii, Edirisinghe Eran and Bez Helmut, "An extended H.264 CODEC for stereoscopic video coding", *Proceedings of SPIE - The International Society for Optical Engineering*, pp. 116-126, (2005).

[5] C. Fehn, K. Hopf and Q. Quante. "Key Technologies for an Advanced 3D-TV System", *In Proceedings of SPIE Three-Dimensional TV, Video and Display III*, pp 66-80, (October 2004).

[6] C. Fehn, "A 3D-TV Approach using Depth-Image-Based Rendering (DIBR)", *In Proceedings of VIIP-2003*, (September 2003).

[7] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression and Trans-mission for a New Approach on 3D-TV", *Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI*, pp.93-104, (January 2004).

[8] H. A. Karim, S. Worrall, A. H. Sadka, A. M. Kondoz, "3-D video compression using MPEG4-multiple auxiliary component (MPEG4-MAC)", *IEE 2nd International Conference on Visual Information Engineering (VIE2005)*, (April 2005).

[9] http://ip.hhi.de/imagecom_G1/savce/index.htm, 2006.

[10] Meesters L.M.J., IJsselsteijn W.A. and Seuntins P.J.H., "Survey of perceptual quality issues in three-dimensional television systems", *Proceedings of the SPIE*, **Vol. 5006**, pp. 313-326, (2003).

[11] Shijun Sun, Shawmin Lei and Toshio Nomura, "Stereo Video Coding Support in H.264", *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6*, (December, 2003).

[12] Thomas Wiegand and Gary J. Sullivan, "Overview of the H.264/AVC Video Coding Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, **Vol. 13**, (July 2003).