

# FRAME CONCEALMENT ALGORITHM FOR STEREOSCOPIC VIDEO USING MOTION VECTOR SHARING

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

I-LAB, CCSR, University of Surrey, Guildford, Surrey, GU2 7XH, UK.

## ABSTRACT

Stereoscopic video is one of the simplest forms of multi view video, which can be easily adapted for communication applications. Much current research is based on colour and depth map stereoscopic video, due to its reduced bandwidth requirements and backward compatibility. Existing immersive media research is more focused on application processing than aspects related to transfer of immersive content over communication channels. As video over packet networks is affected by missing frames, caused by packet loss, this paper proposes a frame concealment method for colour and depth map based stereoscopic video. The proposed method exploits the motion correlation of colour and depth map image sequences. The colour motion information is reused for prediction during depth map coding. The redundant motion information is then used to conceal transmission errors at the decoder. The experimental results show that the proposed frame concealment scheme performs better than applying error concealment for colour and depth map video separately in a range of packet error conditions.

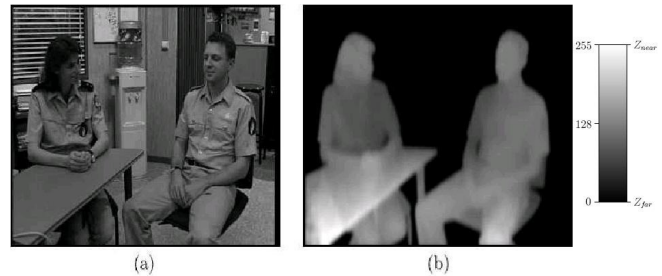
*Index Terms*—Depth-image-based rendering, Frame concealment, Motion vector sharing, Stereoscopic video

## 1. INTRODUCTION

Stereoscopic video (SSV) renders two slightly different views of a scene, one for each eye, to enable the perception of depth. Recently, colour and depth map based SSV has been widely used in research and standardization activities [1,2]. Monoscopic video together with its associated per-pixel depth map can be used to scale existing video communication applications into SSV with low overheads [3]. This format is more flexible and adaptable compared to storage and transmission of the left and right views at low bitrates [4].

Depth map sequences typically have the same spatial resolution as colour video, and only contain a single plane of luminance data (see Fig. 1). The per-pixel depth value determines the position of the associated colour texture in the 3-D space. A specialized image warping technique known as the Depth-Image-Based Rendering (DIBR) is used

to synthesize the desired binocular viewpoint image [5]. As the motion of colour images and corresponding depth information is highly correlated, the colour motion information can thus be used as candidate motion information for the depth video. Hence, the bandwidth for low bitrate SSV applications can be reduced as motion vectors (MVs) are transmitted only once. Furthermore, this reduces the encoding time and complexity of the SSV encoder. The analysis of motion correlation of ‘colour and depth’ sequences and a MV sharing scheme based on MPEG-2 and H.264/AVC is described in [6] and [7], respectively. These developments are mainly focused on reducing the bandwidth, complexity and encoding time of the SSV encoder. This paper exploits the correlation of motion to conceal the frame losses that occur during transmission of the SSV.



**Fig. 1.** ‘Interview’ sequence: (a) Colour image, (b) Per-pixel depth image. The depth-images are normalized to a near clipping plane  $Z_{near}$  and a far clipping plane  $Z_{far}$

The transmission of high resolution 3-dimensional (3-D) video is studied for broadcast (DVB) and Internet Protocol (IP) based streaming video applications in [8]. Streaming 3D video over IP networks may lead to degraded quality due to frame losses. The effect of frame losses on the perceived quality of reconstructed 3-D video may be worse compared to that of conventional video (i.e., 2-D video) applications. Hence, efficient error concealment algorithms are required to improve the reconstructed quality of streamed 3-D video. Error concealment has been performed for left and right view based SSV using the additional data from the corresponding image sequence [9,10]. This paper uses the shared MVs of the corresponding view (Colour or depth map video) to conceal the lost frames. The proposed scheme

reuses the colour MVs during depth map encoding, and this information is transmitted in both colour and depth bit-streams as redundant data in order to conceal transmission errors. Lost frames are then concealed using the correctly received MVs of the corresponding colour or depth map frames accordingly.

The paper is organized as follows. Section 2 elaborates on the proposed frame concealment method based on scalable video coding (SVC) architecture. The experimental setup, results obtained and discussion of those results are given in Section 3. Section 4 concludes the paper.

## 2. FRAME CONCEALMENT USING SHARED MOTION VECTORS

The proposed MV sharing scheme for frame concealment is implemented using the SVC layered encoding architecture, as shown in Fig. 2. The colour and depth information is encoded at the base and enhancement layers, respectively. The scalable extension of H.264/AVC is modified to implement the MV sharing scheme [11]. This SVC architecture is backward compatible to colour video as H.264/AVC compatible base layer can be decoded using a standard H.264/AVC decoder. A performance analysis of a backward compatible SVC encoding configuration for SSV is discussed in [12].

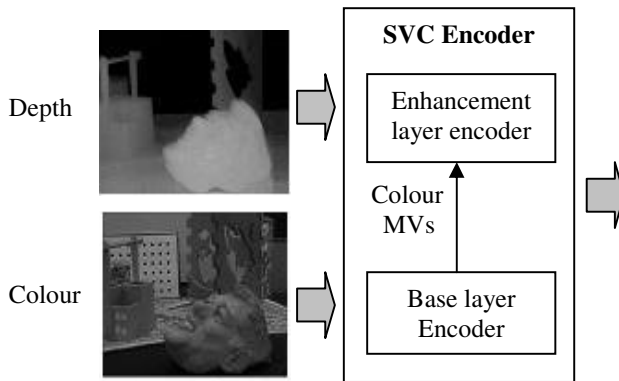


Fig. 2. The MV sharing architecture based on SVC

For the colour image sequence, one MV is generated for each macroblock (MB) using the motion estimation process. Two encoding modes are used to perform rate-distortion (R-D) optimization during depth map coding, namely ‘MB skip’ and ‘Motion compensation’. As large numbers of MBs in the depth image sequence have a uniform texture without high frequency components, the ‘MB skip’ mode is used to increase the compression efficiency. The ‘Motion compensation’ mode does not perform motion estimation. Instead, it reuses the MVs of the corresponding colour image MBs to predict the current depth image at the enhancement layer. Then, the difference between the predicted and original frames is transform coded. The ‘Motion compensation’ mode is performed faster than

conventional motion compensation modes as motion estimation does not take place during depth map coding. Finally, an R-D optimized encoding mode is selected to encode the MBs of the depth image. Hence, the global SVC bit-stream consists of both base and enhancement layer units, which include headers, shared MVs and coded residual texture data.

The overhead added due to the depth information needs to be kept at a smaller percentage of its corresponding colour video. Due to the areas where the motion of colour and depth map is not highly correlated, the proposed MV sharing method imposes a bit budget penalty during depth map coding. However, high quality depth maps can be obtained by adjusting the depth information percentage within an acceptable percentage of colour bitrate. Thus, the average depth bitrate is kept below 25% of its corresponding colour video bitrate.

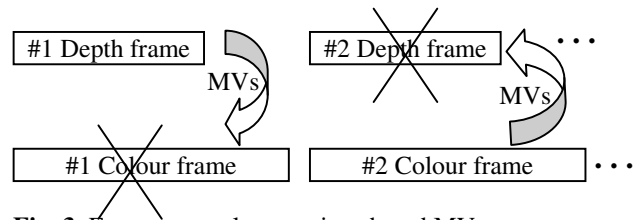


Fig. 3. Frame concealment using shared MVs

The proposed frame concealment scheme using shared MVs is shown in Fig. 3. If a colour video frame is missing due to packet losses, the MVs of the lost frame can be recovered using the correctly received corresponding depth video frame. Similarly, when a depth frame is lost, then the MVs can be recovered from the uncorrupted corresponding colour video frame. The recovered MVs from the corresponding view are used to predict the current frame. The quality of the reconstructed images are lower compared to that of error free decoding as residual information is not used during reconstruction. However, the motion of the reconstructed sequences will be intact and error propagation will be minimized by this proposed method. If both corresponding video frames are lost, then a conventional frame concealment algorithm needs to be used to recover the lost frames. The proposed scheme improves the image quality of both colour and depth map video, and thus ultimately improves the perception of depth in SSV.

## 3. EXPERIMENTS AND RESULTS

Two colour and depth map sequences are used for the experiments, namely ‘Orbi’ and ‘Interview’. ‘Orbi’ is a very complex sequence with camera motion and multiple objects, whereas ‘Interview’ is a sequence captured with a static camera and featuring a stationary background. The tests are carried out using standard definition (SD) TV resolution (720x576 pixels) videos. The test sequences are encoded using the modified JSVM codec version 9.4. The basic

encoding parameters are: 100 frames, IPPP... sequence format, 25 frames/s original frame rate, a single reference frame and Content Adaptive Binary Arithmetic Coding (CABAC). In order to accommodate SD resolution colour and depth map video, the average SVC encoder bitrate is set to 2Mbps. The quantization parameter (QP) of both layers is independently varied to obtain the target bitrate. The depth bitrate is kept below 25% of its corresponding colour video bitrate. For each test sequence two SVC bit-streams are generated using separate MV estimation at both layers and the proposed MV sharing method. The image qualities of the generated SVC bit-streams using separate MVs and the proposed MV sharing scheme are shown in Table 1. Due to the areas of uncorrelated motion, the image quality with the proposed MV sharing is comparably lower compared to the image quality with separate MVs at the given bitrate.

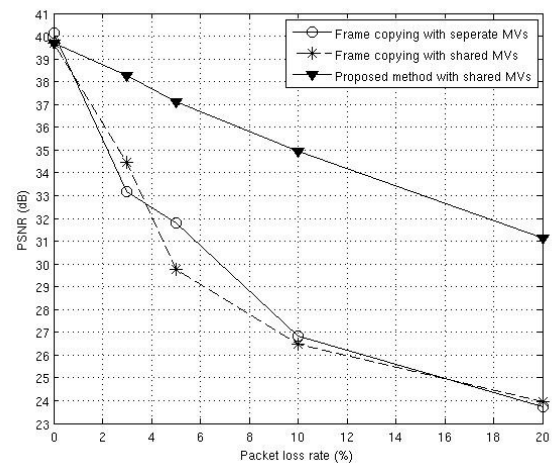
Table 1: Image quality of 'Orbi' and 'Interview' at 2Mbps

	Orbi		Interview	
	Colour PSNR/dB	Depth PSNR/dB	Colour PSNR/dB	Depth PSNR/dB
Separate MVs	40.13	38.10	41.05	42.34
Proposed MV sharing scheme	39.69	37.25	41.05	40.27

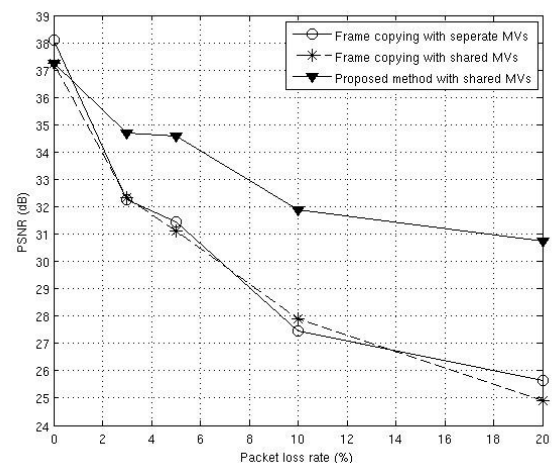
The IP error patterns generated for the Internet video experiments are used to corrupt the generated SVC bit-streams at different packet loss rates [13]. Each coded video frame is encapsulated in a single network abstraction layer (NAL) unit and fragmented into packets of 1400 bytes, which is below the Internet's Maximum Transfer Unit (MTU) size of 1520 bytes. The first fragment contains the important header data of the frame NAL unit. Hence, if the first fragment of a frame is lost, it is assumed that the whole frame is lost, as the header data is unrecoverable without adding redundant information. Furthermore, it is assumed that the sequence headers and the first I frame of both colour and depth image sequences are received uncorrupted. The corrupted bit-streams are later decoded using a JSVM decoder. The bit-streams generated using the separate motion estimation processes are concealed using frame copying. In frame copying, if the current frame is corrupted, the previously decoded frame is copied to the current frame position. Initially, for the bit-streams generated using the MV sharing scheme, the corrupted frames are concealed using frame copying. Then, the lost frames are concealed using the proposed frame concealment technique. The proposed technique uses the duplicated MVs from the corresponding depth or colour frame. If the corresponding MVs are also corrupted, then frame copying is used to conceal the current frame. In order to average the results, the simulations are run ten times starting from different positions in the error traces.

Figures 4 and 5 show the reconstructed image qualities of the 'Orbi' and 'Interview' sequences using different frame

concealment methods. According to Figures 4 and 5, the proposed frame concealment method using shared MVs outperforms both of the frame copying methods, based on separate motion estimation and shared MVs. This is more visible in the case of colour image quality as depth frames are more likely to survive compared to corresponding colour video frames which are larger in size. The gain is significant in highly dynamic 'Orbi' sequence as motion of the 'Orbi' sequence is correctly restored using the duplicated MVs from the corresponding view. Due to low motion of the 'Interview' sequence, even frame copying method achieves comparable results compared to the proposed frame concealment method. The frame copying method with or without shared MVs performs similarly apart from the error free condition. Although the error free quality of the proposed scheme is comparably lower, it improves the quality of reconstructed SSV over packet networks.

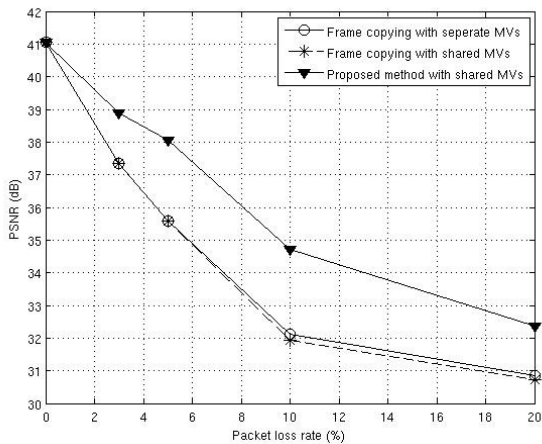


(a)

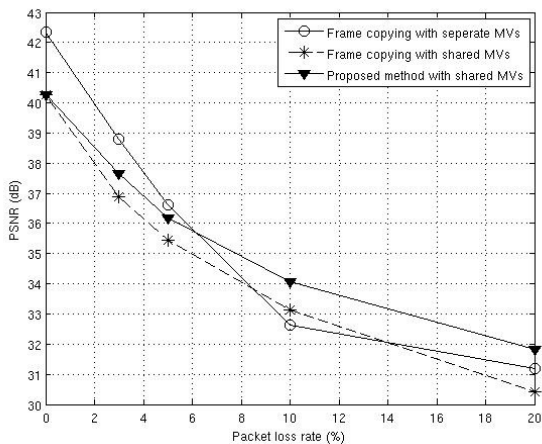


(b)

Fig. 4. Image quality of the 'Orbi' sequence (a) Colour image sequence (b) Depth image sequence



(a)



(b)

**Fig. 5.** Image quality of the 'Interview' sequence (a) Colour image sequence (b) Depth image sequence

#### 4. CONCLUSION

Video streaming over packet networks is affected by packet losses. Hence, efficient error concealment algorithms are required to improve the perceptual quality. This paper proposes a novel frame concealment algorithm for colour and depth map based SSV. The motion of colour and depth map components is highly correlated. Hence, the proposed method reuses the MVs of colour information during depth map encoding. The SVC based encoding architecture is used to encode the source video with backward compatibility and the rate of the coded depth information was kept below 25% of the colour bitrate. Due to the areas of uncorrelated motion, the coded image quality using shared MVs is slightly lower compared to separate MV estimation method. The duplicated colour MVs are used to conceal the lost frames at the decoder. The experimental results show that the quality of reconstructed colour and depth map video

over packet networks can be vastly improved by the proposed method.

#### 5. ACKNOWLEDGEMENTS

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

#### 6. REFERENCES

- [1] A. Bourge and C. Fehn, "White paper on ISO/IEC 23002-3 Auxiliary Video Data Representation", *ISO/IEC JTC1/SC29/WG11 N8039*, 2006.
- [2] C. Fehn, R. De La Barre and S. Pastoor, "Interactive 3-D TV: Concepts and Key Technologies", *Proceedings of the IEEE*, vol. 94, pp. 524-538, 2006.
- [3] C.T.E.R. Hewage, S. Worrall, S. Dogan, H. Kodikara Arachchi and A.M. Kondoz, "Stereoscopic TV over IP", *In Proc. of the 4th IET European Conf. on Visual Media Production*, 2007.
- [4] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV", *Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI*, pp. 93-104, January 2004.
- [5] C. Fehn, "A 3D-TV Approach using Depth-Image-Based Rendering (DIBR)", *In Proc. of Visualization, Imaging, and Image Processing (VIIP'03)*, pp. 482-487, 2003.
- [6] S. Grewatsch and E. Muller, "Sharing of motion vectors in 3D video coding", *International conference on Image processing (ICIP)*, pp. 3271-3274, 2004.
- [7] Han Oh and Yo-Sung Ho, *H.264-Based Depth Map Sequence Coding Using Motion Information of Corresponding Texture Video*, Springer Berlin/Heidelberg, Advances in Image and Video Technology, vol. 4319, 2006.
- [8] G.B. Akar, A. M. Tekalp, C. Fehn and M.R. Civanlar, "Transport methods in 3DTV-A survey", *IEEE Transactions on Circuits and Systems for Video Technology*, vol.17, pp. 1622-1630, November 2007.
- [9] S. Knorr, C. Clemens, M. Kunter and T. Sikora, "Robust concealment for erroneous block bursts in stereoscopic images", *In Proc. 2<sup>nd</sup> Int. Symp. 3-D Data Process., Visual., Transmiss. (3DPVT)*, pp. 820-827, 2004.
- [10] C. Clemens, M. Kunter, S. Knorr and T. Sikora, "A hybrid approach for error concealment in stereoscopic images", *presented at the 5<sup>th</sup> International workshop Image Anal. Multimedia Interactive Services (WIAMIS)*, 2004.
- [11] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz and M. Wien. "Joint draft 8 of SVC amendment", *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.69 (JVT-U201)*, 2006.
- [12] C.T.E.R. Hewage, H.A. Karim, S. Worrall, S. Dogan and A.M. Kondoz. "Comparison of Stereo Video Coding Support in MPEG-4 MAC, H.264/AVC and H.264/SVC", *In Proc. of IET Visual Information Engineering-VIE07*, 2007.
- [13] W. Stephan. "Error patterns for Internet video experiments", *ITU-T SG16 Document Q15-I-16-R1*, 1999.