A H.264 COMPLIANT STEREOSCOPIC VIDEO CODEC

A.B.B. Adikari W.A.C. Fernando H. Kodikara Arachchi K.K. Loo
School of Engineering and Design, Brunel University, Uxbridge UB8 3PH, UK
Adikari.Adikari@brunel.ac.uk Anil.fernando@brunel.ac.uk hemantha@ieee.org Jonatnan.loo@brunel.ac.uk

ABSTRACT

Due to the provision of a more natural representation of a scene in the form of left and right eye views, a stereoscopic imaging system provides a more effective method for image/video display. Unfortunately the vast amount of information that needs to be transmitted/stored to represent a stereo image pair / video sequence, has so far hindered its use in commercial applications. However, by properly exploiting the spatial, temporal and binocular redundancy, a stereo image pair or a sequence could be compressed and transmitted through a single monocular channel's bandwidth without unduly sacrificing the perceived stereoscopic image quality. In this paper, we present a new technique for coding stereo video sequences based on H.264 [1] video codec. The proposed codec exploits disparity and worldline correlation in addition to the advance compression techniques inherited by the H.264 standard to achieve a higher video quality especially in the low bit rates. We compare the performance of the proposed CODEC with a DCT-based, modified MPEG-2 stereo video CODEC [2] and ZTE based stereo video CODEC [3]. We show that the proposed CODEC outperforms the benchmark CODECs in coding both main and auxiliary streams by up to 9.0 dB PSNR gain.

1. INTRODUCTION

During the past decade, 3D visual communication technology has received considerable interest as it intends to provide reality of vision. Various types of 3D displays have been developed in order to produce the depth sensation. However, the accomplishment of 3D visual communication technology requires several other supporting technologies such as 3D image representation, handling, and compression for ultimate commercial exploitation. Many innovative studies on 3D visual communication technology are focused on the development of efficient image compression technology. Within the research context of this paper we intend to continue this effort to a further paradigm, *i.e.* H.264 [1] based stereoscopic video coding.

Stereo vision is used to stimulate 3D perception capability of human psychovisual system by acquiring two pictures of the same scene from two horizontally separated positions and then presenting the left frame to the left eye and the right frame to the right eye. The human brain can process the difference between these two images to yield 3D perception. Thus, every 3D image can be represented with two 2D image frames. These frames are said to form a stereo image pair. If a stereo pair is to be stored or transmitted without exploiting the inherent redundancy, twice as many bits will be required to represent it compared to a monocular image representing the same scene. Fortunately as the two images are projections of the same scene from two nearby points of view, they are bound to have a considerable amount of redundancy between them. By properly exploiting this redundancy, the two image streams can be compressed and transmitted through bandwidth-limited channels without excessive degradation of the perceived stereoscopic image quality.

The proposed codec aims to extend the use of H.264 technology, which represents the state-of-the-art in monoscopic image compression, to stereoscopic coding. In the proposed approach, the left view is coded as a reference sequence using a modified H.264 video coder and each macroblock in the right view is predicted from the decoded left view using disparity [8] or worldline correlation [9] compensated prediction or from the previous/following frames of the right view, using motion compensated prediction. Subsequently the prediction error (residual) frame of the right stream (view) is also coded with the conventional H.264 video coder.

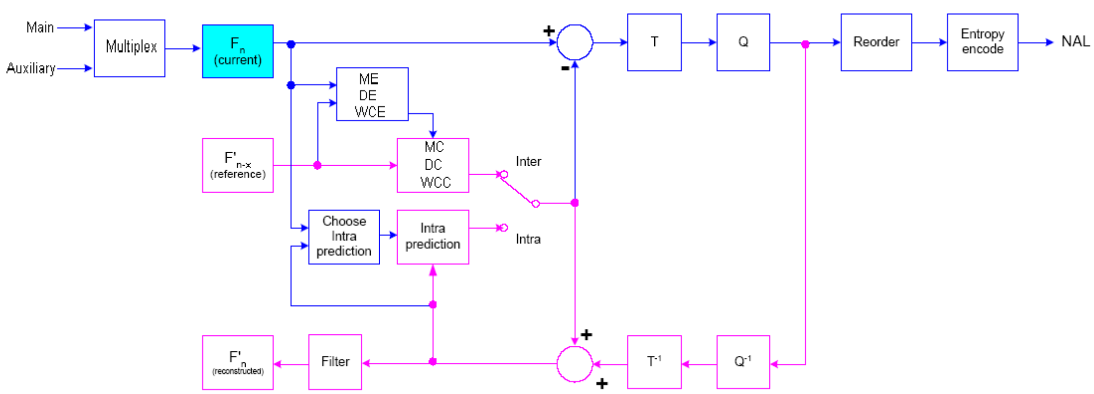
For clarity of presentation, the rest of the paper is organized as follows. Section 2 presents some related work on stereoscopic video coding briefly. In section 3, we summarize the proposed algorithm. Section 4 presents simulation results and a detailed analysis. Finally section 5 concludes with an insight to future directions of research.

2. RELATED WORK

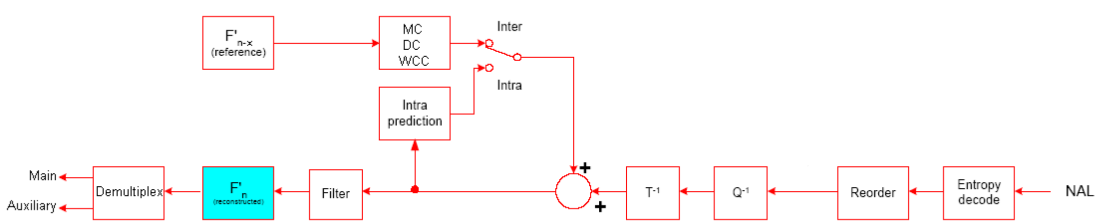
Over the last decade there have been a number of studies on stereo image compression based on intensity or feature. The first technique is to code sum and difference of the two images in a stereo pair [4]**.** After that three-dimensional discrete cosine transform (DCT) coding of stereo image was presented [5]. It is equivalent to the sum-difference coding in transform domain. However, the performance of these two techniques decreases with increased disparity values. Global translation and correlation enhanced techniques assume that objects in the scene have same disparity values. Nevertheless, these methods are not principally efficient since objects in the scene have normally different disparity values. Lukacs [6] constituted the concept of disparity compensation, which established correspondence between similar areas in a stereo pair using binocular disparity information. This method is used to predict the rest of the views from an independently coded view. Other disparity compensation based methods have followed. Perkins [4] formalized disparity compensation based coding as a conditional coding approach that is optimal for lossless coding and suboptimal for lossy coding. Puri *et al.* [7] presented results of MPEG-2 compatible coding. In the standard, one view is coded as base layer and another view is coded within enhancement layer of the temporal scalability model of the MPEG-2 standard. Thanapirom *et al.* has proposed a stereoscopic CODEC based on ZTE and has shown that this CODEC can outperform MPEG-2 based stereoscopic CODEC [3].

3. PROPOSED CODEC

The proposed codec is based on the latest H.264 standard and it is built so that both main and auxiliary streams can be fed. Figure 1 shows a simplified block diagram of the proposed stereoscopic video codec. In the encoder, the two streams 'main' and 'auxiliary' are first fed to the multiplexer to make one stream of images and then the output is sent to the modified H.264 based encoder, where the correlation of the auxiliary stream is exploited mainly using motion, disparity and worldline correlation. In order to exploit the worldline correlation we use 8 multiple references in the main stream. The proposed codec uses all the advanced features of the H.264 such as multiple reference frames and strong motion isolation (¼-pel resolution) with added capabilities to predict based on disparity and worldline correlation to maximize the performance. The encoded bit stream is fed to the modified H.264 based stereoscopic decoder and then the decoded bit stream is sent to the demultiplexer to separate the stream again to main and auxiliary streams.



(a) Encoder



(b) Decoder

Where,

F_n	-	Current frame			
F'_n	-	Reconstructed frames	F'_{n-x}	-	Reconstructed previous frames (X = 1 to 8)
ME	-	Motion Estimation	MC	-	Motion Compensation
DE	-	Disparity Estimation	DC	-	Disparity Compensation
WCE	-	Worldline Correlation Estimation	WCC	-	Worldline Correlation Compensation

Figure 1: The proposed CODEC

4. SIMULATION RESULTS

In all simulations, we assumed that the left image is to be the reference image and the right image is predictive coded. We compare the performance of the proposed CODEC with that of MPEG-2 based and ZTE [10] based benchmark stereoscopic CODECs [2,10]. The results of the simulations performed on AVDS sequences (*Booksale* and *Crowd*) from Carnegie Mellon University, Pittsburgh, PA, USA are shown in Figure 2 and Figure 3 for the auxiliary streams. The simulation results show that a significant PSNR gain with the proposed scheme compared to the MPEG-2 based codec and ZTE based codec for both sequences. Furthermore it is clear that the proposed codec can achieve up to 7.5dB and 9.0dB gain in PSNR (luminance component – *Booksale* sequence) compared to ZTE based and MPEG-2 based stereoscopic CODECS respectively. Similarly for the chrominance components, the proposed encoder provides up to about 6dB PSNR improvement. Furthermore, simulation results for the main stream also has a significant gain in PSNR compared to the benchmark stereoscopic CODECs. We tested the proposed algorithm with other stereo video sequences such as *"train"* and observed similar PSNR gains. Therefore, it could be concluded that the proposed CODEC outperforms both MPEG-2 based and ZTE based benchmark stereoscopic CODECs.

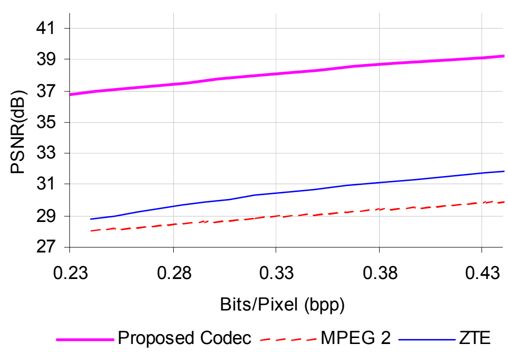


Figure 2.a: Performance comparison (Y signal) of the Proposed Stereoscopic Codec on Booksale sequence

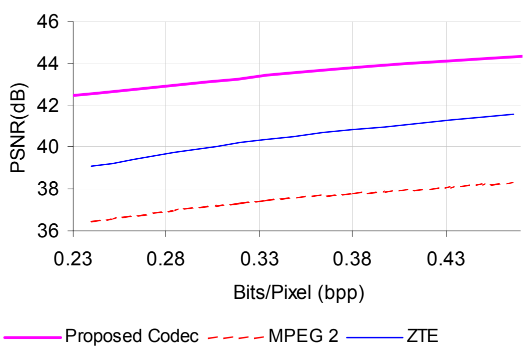


Figure 2.b: Performance comparison (U signal) of the Proposed Stereoscopic Codec on Booksale sequence

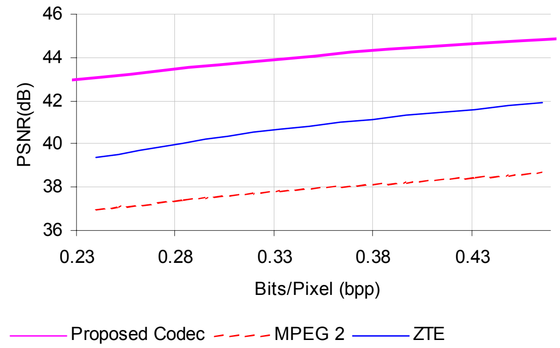


Figure 2.c: Performance comparison (V signal) of the Proposed Stereoscopic Codec on Booksale sequence

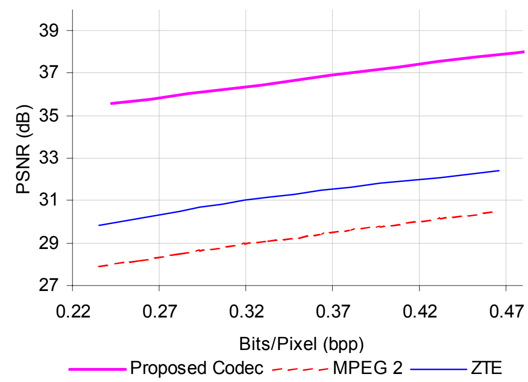


Figure 3.a: Performance comparison (Y signal) of the Proposed Stereoscopic Codec on Crowd sequence

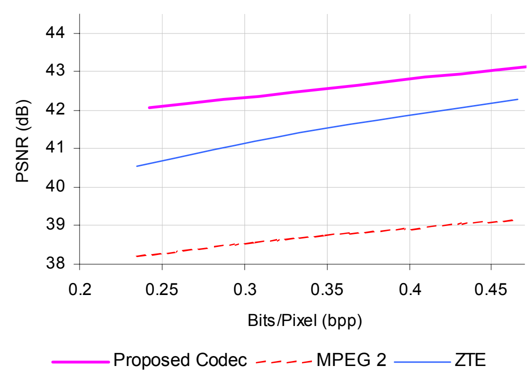


Figure 3.b: Performance comparison (U signal) of the Proposed Stereoscopic Codec on Crowd sequence

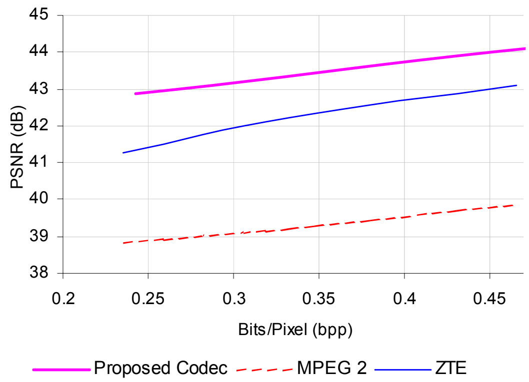


Figure 3.c: Performance comparison (V signal) of the Proposed Stereoscopic Codec on Crowd sequence

5. CONCLUSIONS

We have proposed a state-of-the-art stereoscopic video coding technique that makes use of a modified version of the H.264 technology originally proposed for monoscopic video coding. We have provided experimental results to prove that the proposed CODEC performs better than equivalent CODECs based on MPEG-2 and ZTE up to 9.0dB and 7.5 dB PSNR gain respectively.

Due to the use of state-of-the-art base technology (*i.e.* H.264), the proposed CODEC could be used as a stereoscopic video coding extension to a likely future H.264 based monoscopic video coding standard. At present we are further extending the idea in improving its rate-control strategy and fine-tuning quantization parameters to provide optimum subjective stereoscopic quality.

5. REFERENCES

[1]. T. Wiegand and G. Sullivan, "Final draft international standard (FDIS) of joint video specification (ITU-T rec. H.264 ISO/IEC 14 496-10 AVC)," Joint Video Team (JVT) of ITU-T SG16/Q15 (VCEG) and ISO/IEC JTC1/SC29/WG1, Annex C, Pattaya, Thailand, Doc. JVTG050, Mar. 2003.

[2] Yun-Jeong **"**Improved disparity estimation algorithm with MPEG-2's scalability for stereoscopic sequences," *IEEE Transactions on Consumer Electronics*, Vol. 42, No. 3, Aug. 1996.

[3] S. Thanapirom, W.A.C. Fernando, E.A. Edirisinghe, "A zerotree stereo video encoder,**"** *Proceedings of the 2003 International Symposium on Circuits and Systems*, Vol. 2, 25-28 May 2003.

[4] Perkins, M. G., "Data Compression of Stereo pairs", *IEEE Transaction on Communications*, Vol. 40, No. 4, April, pp. 684-696, 1992.

[5] I. Dinstein, G. Guy, J. Rabany, J. Tzelgov, and A. Henik, "On Stereo Image Coding," *Proceedings of the 9th International Conference on Pattern Recognition*, Rome, Italy, 14-17 Nov, pp. 357-359, 1988.

[6] Lukacs, M.E., "Predictive Coding of Multi-viewpoint Image Sets", *Proceedings of the IEEE-IECEJ-ASJ International Conference on Acoustics*, Speech, and Signal Processing, Tokyo, Japan, pp. 521-524, 1986.

[7] Puri, A., Kollarits, R. V. and Haskell, B. G., "Stereoscopic Video Compression Using Temporal Scalability", *Proceedings of SPIE on Visual Communications and Image Processing*, Taiwan, Vol. 2501, May, pp. 745-756, 1995.

[8] W. Woo, "Rate-Distortion Based Dependent Coding for Stereo Images and Video: Disparity Estimation and Dependent Bit Allocation", Doctoral Dissertation, University of Southern California, December, 1998.
<URL:
http://vr.kjist.ac.kr/~Publications/PhDThesis_Woo.pdf>.

[9] P. D. Gunatilake, M. W. Siegel, A. G. Jordan "Compression of Stereo Video Streams," Department of Electrical & Computer Engineering and School of Computer Science Carnegie Mellon University
<URL: http://www-2.cs.cmu.edu/afs/cs/user/mws/OldFiles/OldFiles/ftp/papers/gunatilake_hdtv93.pdf>

[10] S. A. Martucci, and I. Sodagar, "Zerotree Entropy Coding of Wavelet Coefficients for Very Low Bit rate Video," *Proceedings of the 1996 IEEE International Conference on Image Processing*, Lausanne, Switzerland, 16-19 Sep., Vol. 2, pp. 533-536, 1996.