Optical Engineering 44(7), 077004 (July 2005)

Zerotree-based stereoscopic video CODEC

S. Thanapirom
Loughborough University
Department of Computer Science
Leicestershire, LE11 3TU
United Kingdom

W. A. C. Fernando
Brunel University
School of Engineering and Design
Electronic and Computer Engineering
Uxbridge, Middlesex UB83PH
United Kingdom

E. A. Edirisinghe
Loughborough University
Department of Computer Science
Leicestershire, LE11 3TU
United Kingdom

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 must 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. We propose a timely and novel framework to transmit stereoscopic data efficiently. We propose a timely and novel framework to transmit stereoscopic data efficiently. We present a new technique for coding stereo video sequences based on discrete wavelet transform (DWT) technology. The proposed technique particularly exploits zerotree entropy (ZTE) coding that makes use of the wavelet block concept to achieve low bit rate stereo video coding. One of the two image streams, namely, the main stream, is independently coded by a zerotree video CODEC, while the second stream, namely, the auxiliary stream, is predicted based on disparity compensation. A zerotree video CODEC subsequently codes the residual stream. We compare the performance of the proposed CODEC with a discrete cosine transform (DCT)-based, modified MPEG-2 stereo video CODEC. We show that the proposed CODEC outperforms the benchmark CODEC in coding both main and auxiliary streams. [DOI: 10.1117/1.1951768]

Subject terms: stereoscopic imaging; imaging systems; display; discrete wavelet transform; zerotree entropy coding.

Paper 040561R received Aug. 18, 2004; revised manuscript received Feb. 2, 2005; accepted for publication Feb. 3, 2005; published online Jul. 15, 2005.

1 Introduction

During the past decade, 3-D visual communication technology has received considerable interest as it intends to provide reality of vision. Various types of 3-D displays have been developed to produce the depth sensation. However, the accomplishment of 3-D visual communication technology requires several other supporting technologies such as 3-D image representation, handling, and compression for ultimate commercial exploitation. Many innovative studies on 3-D 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., wavelet-based stereoscopic video coding.

Stereo vision is used to stimulate 3-D perception capability of humans 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 3-D perception. Thus, every 3-D image can be represented by two 2-D 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.

Many techniques have been proposed in literature for coding stereo image pairs. However, coding of stereoscopic video (i.e., image sequences) has not been exploited to a similar extent. The most straightforward approach to stereoscopic video coding is independent coding (simulcast approach) of the left and right monoscopic sequences. However, this has the inherent disadvantage that binocular redundancy is not exploited. An early implementation of a basic stereoscopic video CODEC was presented in Ref. 1, where a MPEG-2-like monoscopic video coding architecture was extended to code stereoscopic video. In this scheme, frames from the predicted (right) sequence were coded with respect to the reference frame based only on disparity compensated prediction; i.e., no efforts were taken to exploit the binocular redundancy jointly with temporal redundancy. Within the standardization activities of MPEG-2 and later within MPEG-4, various stereoscopic video coding systems based on discrete cosine transform

0091-3286/2005/$22.00

(DCT) technology have been considered. The most common of these uses a field-based approach to coding stereoscopic video. Within this scheme, the odd-numbered lines (say) of a typical video frame are made out of data from the left stereoscopic view and the even numbered lines are made out of data from the right view. Frames thus formed are coded using a relevant standard monoscopic video coding scheme and later separated into left and right views at the decoder. However, this approach is capable of only producing half resolution left and right frames at the decoder and does not fully exploit binocular redundancy. In Ref. 2 an object-based coding scheme for stereoscopic video, which uses 3-D models, was considered. This scheme is more suitable for advanced applications such as generation and transmission of intermediate views, but is computationally costly for general purpose applications such as coded transmission of stereoscopic video. The authors of Ref. 3 provide a comprehensive review of other DCT-based stereoscopic video coding techniques.

With the introduction of discrete wavelet transform (DWT)-based technology for coding still images (e.g., EZW, SPIHT, EBCOT, JPEG2000) and its use within I-frame coding of MPEG-4, application of DWT technology to stereoscopic image[4–6] and sequence[7] coding has recently attracted some research interest. In Ref. 7 authors use a nonembedded wavelet-based coding scheme for encoding stereo image pairs, i.e., it uses a band-by-band approach in coding coefficients of wavelet subbands. In addition, it uses a hierarchical, block-based, joint disparity-motion compensation and prediction scheme. Unfortunately the use of the nonembedded coding scheme limits this CODECs application in current technology domains and results in suboptimal rate-distortion performance.

The proposed CODEC aims to further extend the use of DWT technology to efficient stereoscopic video coding. It adopts the most established technique for stereo video coding, block-based stereoscopic coding, which is similar to block-based motion estimation/compensation, defined in MPEG-2 multi-view profile.[8] In the proposed approach, the left view is coded independently as a reference sequence using a modified zerotree video coder[9] and each frame in the right view is predicted from the decoded left view using disparity compensated prediction or from the previous and following frame of the right view, using motion compensated prediction.[10] Subsequently the prediction error (residual) frame of the right stream (view) is also coded by the zerotree video coder. The use of the zerotree-based coding scheme in coding the reference and residual stream enables embedded coding that is capable of improving results as compared to that obtained by the nonembedded scheme of Ref. 7.

For clarity of presentation, the rest of the paper is organized as follows. Section 2 summarizes the principles of zerotree entropy (ZTE) coding. The proposed method for coding stereo sequences is presented in Sec. 3. Section 4 presents simulation results and a detailed analysis. Finally, Sec. 5 concludes with an insight to future directions of research.

2 ZTE Coding

ZTE (Ref. 11) is an efficient method for coding wavelet transform coefficients of motion-compensated video residuals or of video frames. Originally proposed by Martucci and Sodagar,[11] it is based on the idea that, in the hierarchical decomposition of the wavelet transform, every coefficient at a given scale, excluding the highest frequency subband, can be associated to a set of coefficients of same spatial location and of same orientation at the next finer scale. The coefficient at the coarse scale is called the parent, and the four coefficients representing the same spatial location and of similar orientation at the next finer scale are called children. According to this relationship, we can build a data structure called a wavelet tree.

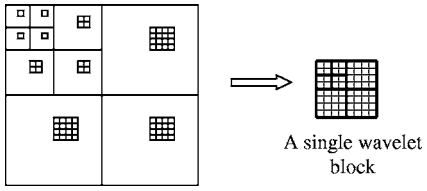

Fig. 1 Formation of a single wavelet block from a three-level DWT decomposed image.

In ZTE coding, the coefficients of each wavelet tree are rearranged to form a wavelet block (see Fig. 1). Each wavelet block contains all coefficients in the same wavelet tree in the frame. The wavelet blocks are located at the same corresponding spatial location as their wavelet tree. To use ZTE coding, a symbol is assigned to each node in a wavelet tree describing the wavelet coefficients corresponding to that node.

The quantization of wavelet coefficients in ZTE coding can be performed either before employing ZTE coding or during the ZTE process. Any quantization scheme can be used. By combining quantization with the construction and coding of zerotrees, the quantization and bit allocation can be done adaptively according to spatial location and frequency band information.

The wavelet trees are coded by scanning each tree from the root in the lowest frequency band through the children, and assigning one of four symbols to each node encountered: zerotree root, valued zerotree root, or value. A zerotree root denotes a coefficient that is the root of a zerotree. A zerotree exists at any wavelet tree node where the coefficient is zero and all the node's children are themselves zerotrees. A valued zerotree root is a node where the coefficient has nonzero amplitude and all four children are zerotree roots. A value symbol identifies a coefficient with amplitude either zero or nonzero, but also with some nonzero descendant. The scanning process of each tree can stop at zerotree root or valued zerotree root symbols. The list of nonzero quantized coefficients that correspond one-to-one with the valued zerotree root symbols are encoded using an alphabet that does not include zero. The remaining coefficients, which correspond one-to-one to the value symbols, are encoded using an alphabet that includes zero. For any node reached in a scan without a node's children, zerotree root and valued zerotree cannot apply. Therefore, bits are saved by not encoding any symbol for this node and encoding the coefficients along with those corresponding to the value symbol.

In the following section we present the proposed stereoscopic video CODEC.

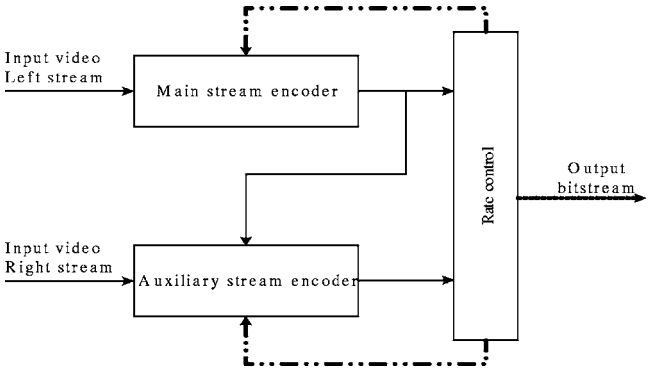

Fig. 2 Proposed stereo image encoder.

3 Proposed Scheme

Figure 2 shows a simplified block diagram of the proposed stereoscopic video encoder. It consists of a rate-control module and two subencoders, namely, a mainstream encoder and an auxiliary-stream encoder. Each subencoder operates similar to a conventional video coder and employs the MPEG frame structure of *I*, *P*, and *B* pictures. The reference, left stream (i.e., the mainstream) is input to the mainstream encoder. The "to be predicted" right stream (i.e., the auxiliary stream) is input to the auxiliary-stream encoder. As the right stream is predicted with reference to the left stream, the decoded left stream acts as a second input to the auxiliary-stream encoder. The rate control module regulates the bit rate according to the target bit allocation for each frame in both streams. If the bit rate exceeds the target bit rate for that frame, the feedback forces the quantizer step size to be increased. Similarly, if the actual bit rate is less, then the quantizer step size is reduced to increase the bit rate and improve the image quality.

A detailed block diagram of the main-stream encoder is illustrated in Fig. 3. Here, the mainstream frames are encoded either as intra- or interframes. The intraframe pixels are transformed into DWT coefficients and are subsequently quantized and coded by ZTE and adaptive arithmetic coding. The succeeding intercoded frames are predicted from their past or future reference frames. To reduce blocking artifacts, overlapped block motion compensation (OBMC) and prediction is used. Subsequently, the residuals are transformed into DWT coefficients, quantized, coded by ZTE and adaptive arithmetic coding as in the case of intraframe coding. Motion vectors from all interframes are coded by Huffman coding. The quantizer step size is manipulated by the number of bits available for the mainstream frame. The rate control does not separately control the bit rate in the two streams, but does so relative to the total bit budget of both streams.

The coding process within the auxiliary-stream encoder is similar to that of the mainstream encoder. However, as the auxiliary-stream frame can be coded with respect to the left coded frame or its reference right coded frames, the auxiliary stream frame is encoded by first finding the best predicted frames from disparity or motion compensated prediction.[10] We denote the *I*, *P*, and *B* pictures of the main stream using I_M, P_M, and B_M. The corresponding pictures in the auxiliary stream are represented by I_A, P_A, and B_A. The following strategies are used in coding these three types of frames in the auxiliary stream:

1. An I_A frame is coded by using disparity compensation prediction with respect to the corresponding I_M frame. The total bit count is greatly reduced compared to independent coding.
2. A P_A frame is coded by selecting the better prediction between the previous reference frame and the corresponding P_M frame.
3. A B_A frame is coded by selecting the best prediction amongst the previous reference frame, the following reference frame and the corresponding B_M frame.

This prediction process is illustrated in Fig. 4.

A detailed block diagram of the auxiliary stream encoder is illustrated in Fig. 5. After the encoder finds the best prediction, the rest of the auxiliary-stream encoder operates similar to the mainstream encoder, i.e., performing DWT, quantization, ZTE coding, and arithmetic coding, respectively.

The rate control module controls the bit rate of the auxiliary-stream, proportionate to the total bid budget allo-

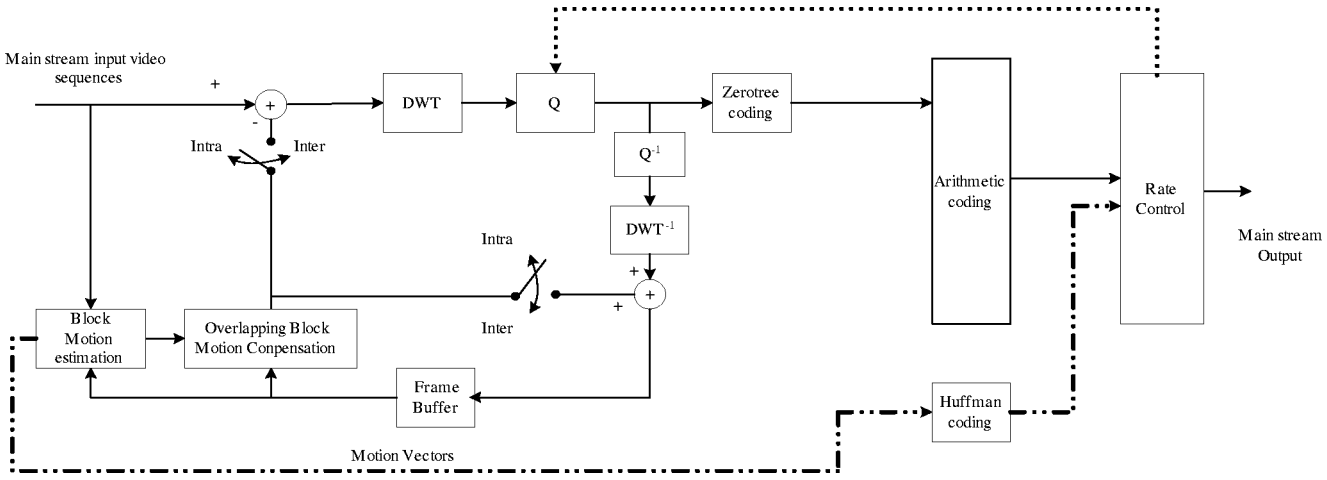

Fig. 3 Mainstream encoder.

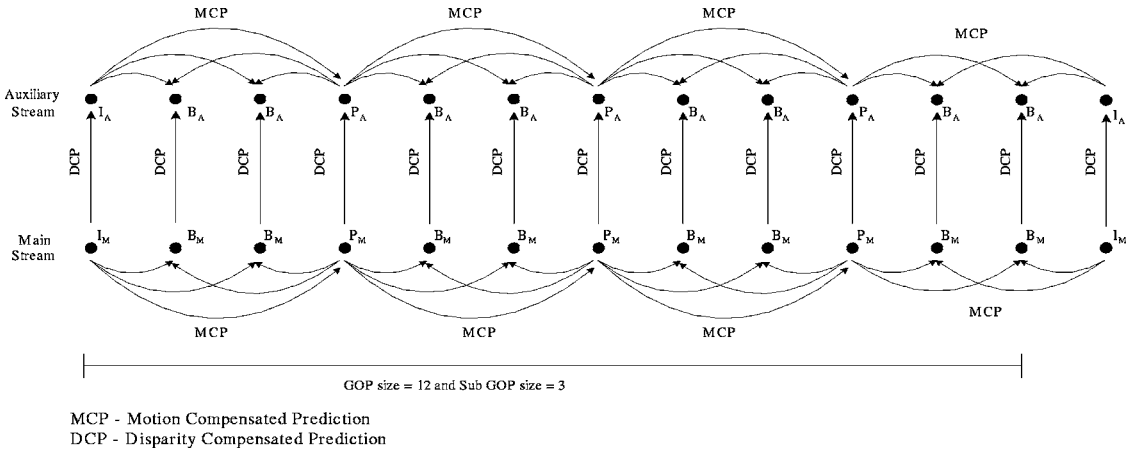

Fig. 4 Disparity and motion-compensated prediction in stereo video sequences.

cated to both streams. Nevertheless, it should be noted that both stereo streams are not necessarily coded at the same level of quality. In most cases, the quality of one of the views can be significantly lower (3 dB) when compared to the other view without sustaining perceivable visual distortions. This observation is supported by the results of many experiments presented in Ref. 12.

The following subsections present operational details of the main modules of the encoder.

3.1 *Motion Estimation and Compensation*

3.1.1 *Block motion estimation*

For motion estimation the auxiliary-stream frame (anchor frame) is divided into nonoverlapping 16×16 blocks. For each block in the anchor frame, a search range of ± 16 pixels is applied in both directions on the reference frame. The predicted block is found by minimizing the sum of absolute differences (SAD). To favor zero motion vectors MVs, the SAD of zero displacement is reduced by a value of 100. The half-pel refinement is also exploited in all four directions using bilinear interpolation of the eight surrounding pixels to diminish the mismatch error. The motion estimation is performed only in the luminance component. MVs thus obtained are divided by two and used in conjunction with quarter-pel interpolation to obtain the prediction of the chroma blocks. The MV is differentially coded by predicting from a spatial neighborhood of three MVs that are transmitted. The prediction in differential encoding of the MVs is same as that of the differential pulse code modulation encoding of dc band coefficients in DCT-based technologies.

3.1.2 *OBMC*

A significant drawback of DWT-based coding when compared to DCT based coding is that one has to sacrifice many bits to code the artificial high frequency information at the block boundaries of motion-compensated *P* or *B* frames. Therefore, we propose to use OBMC to offer substantial reductions in the prediction error. OBMC predicts the current frame by repositioning overlapping blocks of pixels from the previous frame, each weighted by a smoothing window. In this method, each 8×8 block in a macroblock is overlapped with three major adjacent blocks that do not belong to the same macroblock. In our OBMC algorithm, we used a raised cosine window, as it performed better when compared to a bilinear window frequently used in conventional motion compensation.[13]

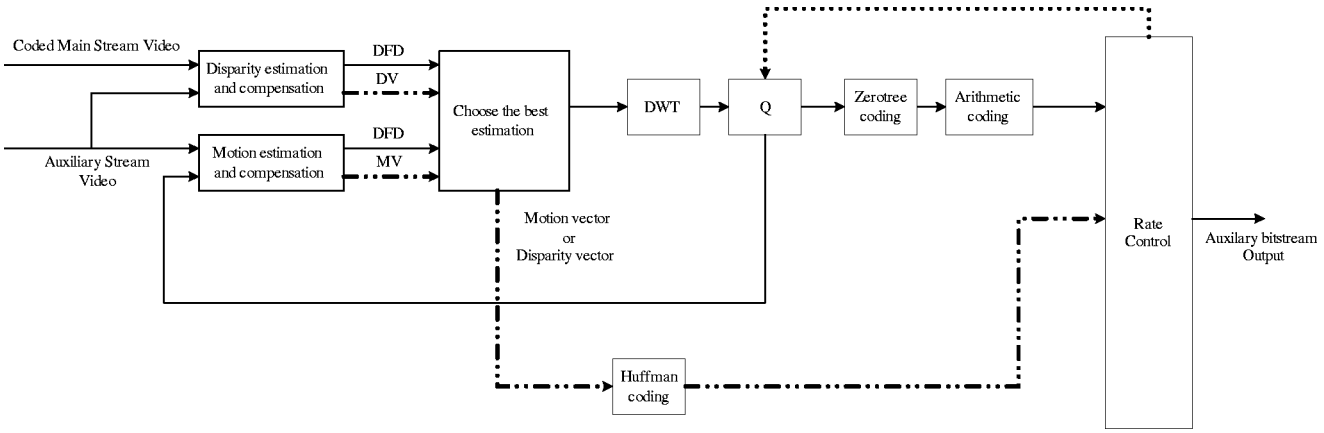

Fig. 5 Auxiliary stream encoder.

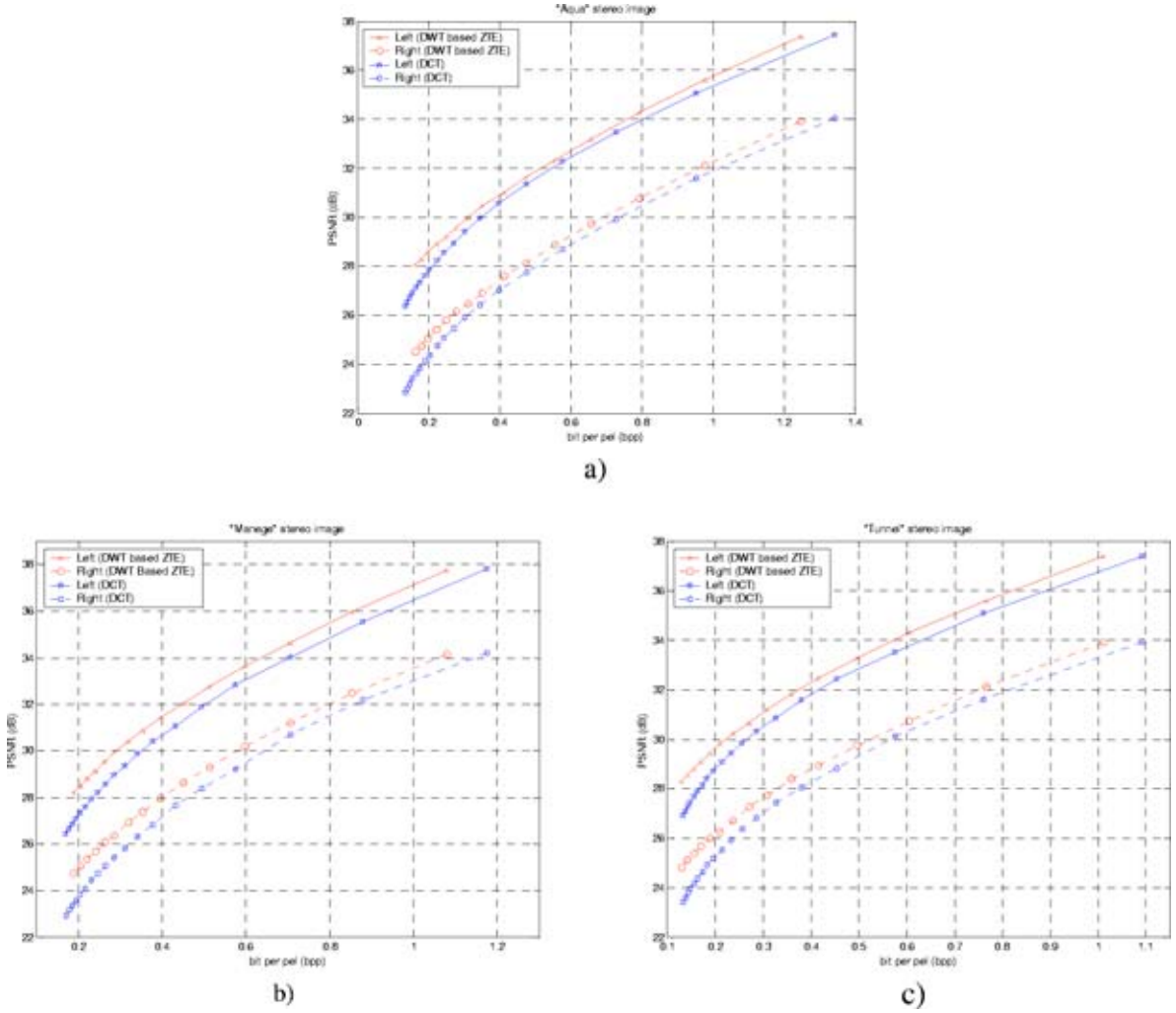

Fig. 6 Rate-distortion performance of I frame encoding/decoding process of proposed stereo video CODEC for (a) "Aqua," (b) "Manege," and (c) "Tunnel."

3.2 *Disparity Estimation and Compensation*

As disparity refers to the displacements between the 2-D points in the left and the right images that correspond to the same 3-D point in a real scene, the disparity estimation is similar to the motion estimation in that both require establishment of the correspondence between pixels in two images. Hence, in the disparity estimation and compensation, we use the same estimation and compensation scheme, i.e., block motion estimation followed by overlapped block motion compensation, as presented in Sec. 3.1. However, assuming that the stereoscopic video sequences have been captured with a parallel axis camera configuration, we could limit the disparity estimation in the vertical direction. Thus, the search window size can be modified to 0 to 16 in the horizontal direction and ±4 in the vertical direction. The reason for having a small search range in the vertical direction comes from the fact that the camera axes may not always be strictly parallel due to calibration difficulties and errors.

3.3 *DWT*

The wavelet transform performs decomposition of video frames, motion-compensated residual frames or disparity-compensated residual frames. The DWT can be made flexible by allowing explicit specification filters to be used at each decomposition level. In the proposed CODEC, DWT is implemented up to four decomposition levels in luminance components and up to three decomposition levels in chrominance components.

However, a method for image boundary treatment[14] is required such that the image size does not expand and no artifacts occur at the image borders. We use the symmetric periodic extension method to avoid these problems. Another important factor is the choice of the wavelet filters in performing DWT. We prefer linear phase filters associated with biorthogonal wavelet bases as they do not introduce high phase distortions around edges and the symmetry of filters enables solving the border extension problem. The use of longer filters provides a good frequency localization but can cause ringing artifacts along the edges of objects. In

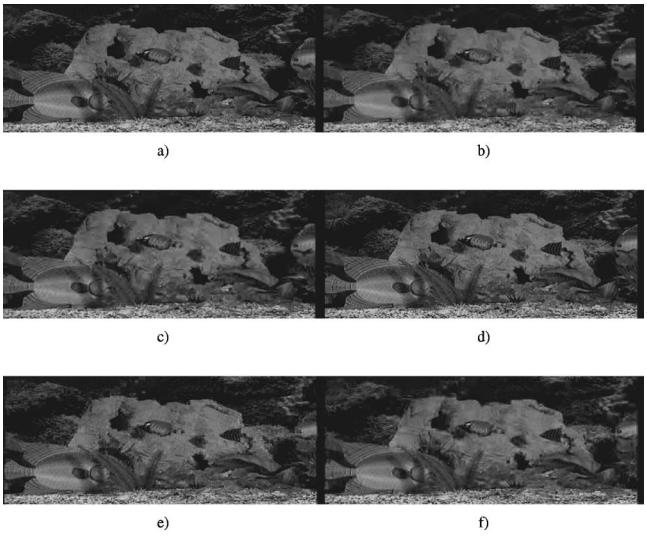

Fig. 7 Subjective results of "Aqua" stereo image pair at 0.163 bpp. (a) Original left image (b) reconstructed left image (DCT), (c) reconstructed left image (proposed), (d) original right image, (e) reconstructed right image (DCT), and (f) reconstructed right image (proposed).

contrast, the use of shorter filter banks, such as Haar, results in less ringing artifacts but more blockiness is caused in the reconstructed frame. Therefore, a combination of long filters for the initial levels and shorter filters for later levels provides a good tradeoff between ringing and blocking artifacts. In our encoder, in the DWT of the luminance components, we use the biorthogonal 9/7 filter for the first two levels, biorthogonal 5/3 filter for the third level, and Haar for the last level. In the chrominance components, DWT is performed by using biorthogonal 9/7 filter in all three decomposition levels.

3.4 *Quantization and ZTE Coding*

To preserve the reconstructed image quality over the threshold of visibility, an optimal quantization approach would be designed in such a way that the actual bit rate is under the constraint target bit rate. In the proposed work, we use a midriser uniform scalar quantizer with quantization scale q and a perceptual weighting matrix depending on the frequency content of a given subband. We use a quantization matrix that is based on the effect of the frequency band to noise sensitivity.[15]

To absorb the wavelet coefficients, we have implemented a dead zone, a large zero bin $[-\tau, \tau]$. The ratio $\eta = (2\tau/q)$ of zero bin size to step size had been chosen $\eta_{\text{intra}} = 1.5$ for intraframe coding and $\eta_{\text{inter}} = 2$ for interframe coding. The preceding choice of η_{intra} and η_{inter} was experimentally proven to be a good approximation of the optimum.[15]

3.5 *ZTE Coding*

After quantization, the quantized values are coded by scanning each tree depth first, from the values in the lowest frequency band through to the highest frequency band and assigning one of four ZTE symbols, discussed in Sec. 2, to each quantized coefficient encountered. The wavelet coefficients of dc bands are encoded independently from the other bands.

3.6 *Adaptive Arithmetic Coding*

ZTE symbols and the quantized coefficient values generated by the zerotree scanning are all encoded using an adaptive arithmetic coder. The arithmetic coder is run over symbols and quantization values so we require two separate adaptive models for each alphabet set. Nevertheless, we employ the different statistical models for different decomposition levels, since the range of wavelet coefficients is enlarged as the decomposition level increases. The output bit stream is separately generated for each luminance and color component.

3.7 *Rate Control*

The rate control is implemented in the picture level, i.e., we use the same quantization scale for all wavelet trees. The number of bits to be spent for each frame is determined based on the specific bit allocation for that type of frame and the number of bits available. The specific bit allocation for each type of frame in both main and auxiliary stream is calculated by the following method.

Table 1 The average PSNRs for $\{Y, U, V\}$ at 0.240 bpp of the "Booksale" stereo sequence.

		Main Stream PSNR (dB)			Auxiliary Stream PSNR (dB)			
		Y	*U*	*V*	*Y*	*U*	*V*	Total Bits (Main+Aux.)
	DCT	31.11	38.65	38.84	28.03	37.42	37.93	123,853
I	ZTE	32.38	42.36	42.17	31.63	41.65	41.95	126,956
	DCT	29.16	37.91	38.20	28.08	36.50	36.95	67,270
P	ZTE	31.12	40.95	40.95	28.83	39.12	39.35	60,838
	DCT	30.80	38.51	38.79	29.42	37.49	38.04	40,379
B	ZTE	32.00	41.90	41.96	30.54	40.39	40.77	42,589

Table 2 The average PSNRs for $\{Y, U, V\}$ at 0.235 bpp for the "Crowd" stereo sequence.

		Main Stream PSNR (dB)			Auxiliary Stream PSNR (dB)			
		Y	*U*	*V*	*Y*	*U*	*V*	Total Bits (Main+Aux.)
	DCT	31.48	39.84	40.59	27.96	38.81	39.62	122,762
I	ZTE	32.41	42.93	43.79	29.76	40.87	41.64	105,199
	DCT	30.59	38.68	39.26	27.94	38.20	38.82	62,522
P	ZTE	31.94	42.07	42.93	29.84	40.53	41.29	60,856
	DCT	31.46	38.83	39.52	28.89	38.58	39.29	41,235
B	ZTE	32.59	42.16	43.04	30.77	41.06	41.88	42,578

Due to the following assumptions on rate-distortions characteristics[16]:

1. The bit rate (r) used in a frame is inversely proportional to the distortion (d).
2. The distortion increases linearly with quantization scales (q). The rate of increase for different frame types is different.

We can define the bit allocation (or complexity) for each type of frame as

$$X = rq$$

where r is the bit rate and q is the quantization scale for each frame type in both streams.

The target bit budget for each frame in the stereo stream is calculated based on this complexity. After coding each frame the complexity value is updated according to the new bit rate and quantization scale.

4 Simulation Results

First, to investigate the performance of the proposed CODEC in coding the *I* frames of both streams, we performed experiments on three stereoscopic image pairs made available by CCETT (DISTIMA). The left image is assumed to be the reference image and the right image is predictive coded. As for comfortable stereoscopic viewing one image (say left) can be coded with higher quality than the other (i.e., right), we have allowed a quality difference of 3 to 3.5 dB between the reconstructed images. This assumption increases the total compressibility of a given stereoscopic image pair. Figure 6 compares the performance of the *I* frame encoding/decoding process of the proposed stereo video CODEC with that of an equivalent DCT-based stereo image coding benchmark, for stereo image pairs, "Aqua," "Manege," and "Tunnel." The only difference between the two techniques is the difference in the base technology used, i.e., DWT versus DCT. The rate-distortion curves of all stereo image pairs show a peak SNR (PSNR) enhancement of 0.5 to 1 dB (for both left and right images) as compared to the DCT-based benchmark. Further, the proposed DWT-based technique performs comparatively better at bit rates lower than 0.4 bpp. It is also shown that the predictive error image could be compressed between 40 to 60% the size of the reference (left) image, still maintaining the quality of the reconstructed right image at an

Table 3 The average PSNRs for $\{Y, U, V\}$ at 0.333 bpp for the "Booksale" stereo sequence.

		Main Stream PSNR (dB)			Main Stream PSNR (dB)			
		Y	*U*	*V*	*Y*	*U*	*V*	Total Bits (Main+Aux.)
	DCT	33.90	41.10	41.16	29.59	39.00	39.31	172,184
I	ZTE	34.60	44.12	44.10	31.83	41.95	42.26	150,509
	DCT	32.92	39.63	39.81	28.98	37.47	37.84	91,968
P	ZTE	33.02	42.36	42.49	30.51	40.39	40.69	86,549
	DCT	33.33	40.31	40.51	30.20	38.61	39.10	56,823
B	ZTE	34.30	43.26	43.47	32.29	41.69	42.13	60,994

Table 4 The average PSNRs for $\{Y, U, V\}$ at 0.333 bpp for the "Crowd" stereo sequence.

		Main Stream PSNR (dB)			Main Stream PSNR (dB)			
		Y	*U*	*V*	*Y*	*U*	*V*	Total Bits (Main+Aux.)
	DCT	33.76	41.45	42.19	29.35	39.46	40.21	171,948
I	ZTE	34.57	44.39	45.30	31.23	41.70	42.55	149,833
	DCT	32.79	40.06	40.68	29.08	38.70	39.22	89,216
P	ZTE	33.42	43.17	44.05	31.17	41.40	42.23	86,669
	DCT	33.66	40.20	40.88	29.83	39.17	39.82	57,469
B	ZTE	34.01	43.17	44.03	32.08	41.85	42.74	61,048

equivalent level to that of the left image, at bit rates 0.4 to 1.0 bpp. This is a significant achievement in terms of compressibility of a stereo image pair.

The subjective results of "Aqua" stereo image pair at 0.163 bpp are shown in Fig. 7, which shows that the reconstructed right image obtained using the proposed techniques [Fig. 7(f)] is of a marginally better quality as compared to its DCT-based counterpart [Fig. 7(e)].

Second, we compared the performance of the proposed CODEC with that of a DCT-based benchmark stereoscopic CODEC implemented by us. The preceding benchmark CODEC is identical to an existing MPEG-2-based stereoscopic CODEC, but for fairness of comparison with the proposed CODEC we have replaced run-length coding with arithmetic coding and implemented an identical rate-control scheme, as proposed in Sec. 3.7. The results of the simulations performed on AVDS sequences from Carnegie Mellon University, Pittsburgh, Pennsylvania, are shown in Tables 1–6 and could be summarized as follows.

For the luminance component of the main stream, at low bit rates (Tables 1 and 2), the proposed CODEC achieves marginally better performance as compared to the benchmark. For the luminance component of the main stream, at high bit rates, the proposed CODEC performs better as compared to the benchmark by 0.5 to 2.0 dB. For luminance component in the auxiliary stream, the proposed CODEC shows 1.0 to 2.5-dB enhancement in PSNR when compared to the benchmark. For chrominance components in the main and auxiliary streams, the proposed CODEC shows 2- to 3-dB PSNR improvement as compared to the benchmark.

Further analysis of the results in Tables 1–6 suggests that under the present experimental arrangements it could be concluded that the proposed CODEC outperforms the benchmark in coding *I* (intracoded) and *P* (predictive-coded) frames. However, the results are not that conclusive in pointing out the same for *B* (bidirectionally coded) frames. Given the fact that PSNR represents a function that is harder to converge as compared to the bit rate, further critical experimental analysis indicated that the high PSNR values for *B* frames obtained with the proposed CODEC as compared to using the benchmark CODEC (even though at an expense of extra bits) does mean that it performs marginally better for *B* frames as well.

Figure 8 illustrates the original and reconstructed frames for the crowd image when coding the 13th left frame at 0.233 bpp. It clearly shows the improved perceptual quality that is obtainable at low bit rates by a DWT-based CODEC as compared to a DCT-based CODEC.

Table 5 The average PSNRs for $\{Y, U, V\}$ at 0.466 bpp for the "Booksale" stereo sequence.

		Main Stream PSNR (dB)			Main Stream PSNR (dB)			
		Y	*U*	*V*	*Y*	*U*	*V*	Total Bits (Main+Aux.)
	DCT	36.39	43.47	43.60	31.16	40.19	40.53	239,258
I	ZTE	37.16	45.86	46.05	33.65	43.16	43.64	210,620
	DCT	35.59	41.22	41.49	30.18	38.26	38.67	127,465
P	ZTE	34.69	43.73	43.92	32.11	41.55	41.93	121,339
	DCT	36.03	41.82	42.11	31.45	39.53	40.05	79,623
B	ZTE	35.93	44.50	44.75	33.86	42.84	43.32	85,352

Table 6 The average PSNRs for $\{Y, U, V\}$ at 0.466 bpp for the "Crowd" stereo sequence.

		Main Stream PSNR (dB)			Auxiliary Stream PSNR (dB)			
		Y	*U*	*V*	*Y*	*U*	*V*	Total Bits (Main+Aux.)
	DCT	36.04	43.50	44.31	30.79	40.07	40.99	243,203
I	ZTE	36.88	45.93	46.86	32.64	42.64	43.46	209,973
	DCT	35.36	41.35	42.10	30.52	39.15	39.85	123,024
P	ZTE	34.82	44.30	45.12	32.40	42.28	43.08	121,245
	DCT	35.44	41.42	42.23	31.27	39.70	40.52	80,698
B	ZTE	35.32	44.14	44.98	33.26	42.66	43.52	85,462

5 Conclusion and Further Work

We proposed a state-of-the-art stereoscopic video coding technique that makes use of a modified version of the ZTE technology originally proposed for monoscopic video coding. The core modules of the CODEC consist of an OBMC module, DWT module with variable length filters, a ZTE module and an adaptive arithmetic coding module for coding quantized wavelet coefficients. We provided experimental results to prove that the proposed CODEC performs better than an equivalent CODEC using DCT as the basis technology.

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

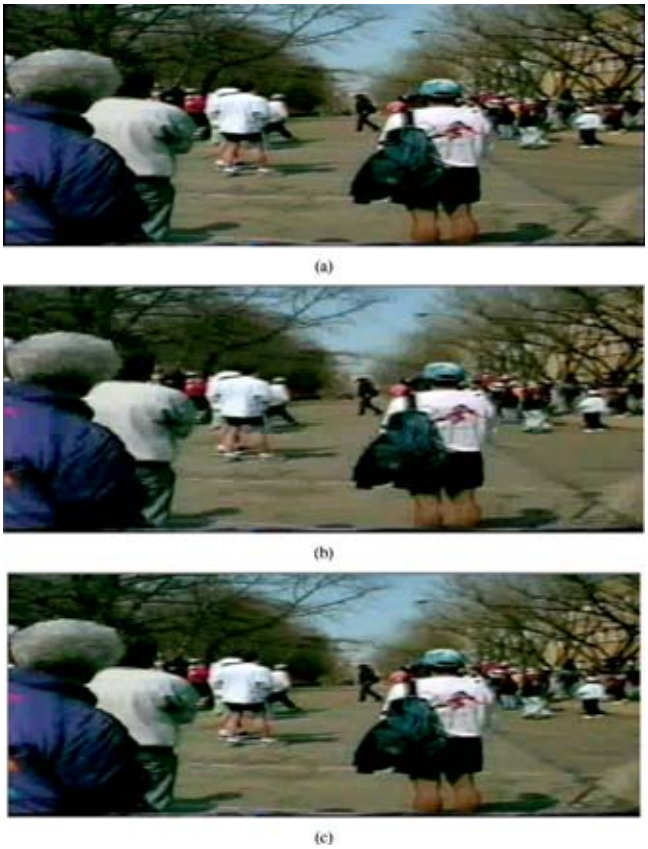

Fig. 8 Subjective results of "Crowd" sequence (frame 13) at 0.233 bpp: (a) original left image, (b) reconstructed left image (benchmark), and (c) reconstructed left image (proposed).

Acknowledgment

The authors would like to thank Dr. Mel Siegel of Robotics Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, for providing invaluable advice.

References

1. M. Siegel, P. Gunatilake, S. Sethuraman, and A. Jordan, "Compression of stereo image pairs and streams," *Stereosc. Displ. Virt. Reality Syst.* **2177**, 258–268 (1994).
2. S. Malassiotis and M. G. Strintzis, "Object based coding of stereo image sequences using three-dimensional models," *IEEE Trans. Circuits Syst. Video Technol.* **7**(6), 892–905 (1997).
3. A. Puri, R. Kollarits, and B. Haskell, "Basics of stereoscopic video, new compression results with MPEG-2 and a proposal for MPEG-4," *Int. J. Signal Process. Image Commun.* **10**, 201–234 (1997).
4. A. Nantheera, W. A. C. Fernando, and E. A. Edirisinghe, "An EBCOT based stereoscopic image codec," in *Proc. Int. Symp. on Communications and Information Technologies*, pp. 695–699, Songkhia, Thailand (2003).
5. G. Rajkumar, M. Y. Nayan, and E. A. Edirisinghe, "RASTER: a JPEG-2000 stereo image CODEC," in *Proc. IASTED Int. Conf. on Visualization, Imaging & Image Processing*, pp. 233–238 (2003).
6. E. A. Edirisinghe, M. Y. Nayan, and H. E. Bez, "A wavelet implementation of the pioneering block-based disparity compensated predictive coding algorithm for stereo image pair compression," *Int. J. Signal Process. Image Commun.* **19**(1), 37–46 (2004).
7. P. Chang and M. Wu, "A wavelet multiresolution compression technique for 3D stereoscopic image sequences based on mixed-resolution psychophysical experiments," *Int. J. Signal Process. Image Commun.* **15**, 705–727 (2000).
8. Yun-Jeong, "Improved disparity estimation algorithm with MPEG-2's scalability for stereoscopic sequences," *IEEE Trans. Consum. Electron.* **42**(3),306–311 (1996).
9. S. A. Martucci, I. Sodagar, T. Chiang, and Y. Zhang, "A zerotree wavelet video coder," *IEEE Trans. Circuits Syst. Video Technol.* **7**(1), 109–118 (1997).
10. S. Sethuraman, M. W. Siegel, and A. G. Jordan, "Segmentation based coding of stereoscopic image sequences," *Proc. SPIE* **2668**, 420–430 (1996).

11. S. A. Martucci and I. Sodagar, "Zerotree entropy coding of wavelet coefficients for very low bit rate video," in *Proc. IEEE Int. Conf. on Image Processing*, pp., Lausanne, Switzerland (1996).
12. B. G. Haskell, A. Puri, and A. N. Netravali, *Digital Video: An Introduction to MPEG-2*, Kluwer Academic (1996).
13. M. T. Orchard and G. J. Sullivan, "Overlapped block motion compensation: an estimation-theoretic approach," *IEEE Trans. Image Process.* **3**(5), 693–699 (1994).
14. C. Chrysafis, "Wavelet image compression rate distortion optimizations and complexity reductions," PhD Dissertation, University of Southern California (Mar. 2000).
15. A. S. Lewis and G. Knowles, "Image compression using the 2-D wavelet transform,," *IEEE Trans. Image Process.* **1**(2), 244–250 (1992).
16. L. J. Lin, "Video bit-rate control with spline approximated rate-distortion characteristics," PhD Dissertation, University of Southern California (May 1997).

S. Thanapirom Biography and photograph not available.

W. A. C. Fernando received his BSc degree in engineering (first class) in electronic and telecommunications engineering from the University of Moratuwa, Sri Lanka, in 1995, his MEng degree (with distinction) in telecommunications from the Asian Institute of Technology (AIT), Bangkok, Thailand, in 1997, and his PhD degree from the Department of Electrical and Electronic Engineering, University of Bristol, United Kingdom, in 2001. He is currently a lecturer in signal processing with Brunel University, United Kingdom. Prior to that, he was an assistant professor with AIT. His current research interests include digital image and video processing, intelligent video encoding, optical frequency-division multiplexing and code division multiple access for wireless channels, channel coding, and modulation schemes for wireless channels. He has published more than 100 international papers in these areas.

E. A. Edirisinghe received his BSc Eng (hons.) degree in electronic and telecommunications engineering from Moratuwa University, Sri Lanka, in 1994 and his MSc degree in digital communication systems and his PhD degree from Loughborough University, United Kingdom, in 1996 and 1999. He joined the Department of Computer Science, Loughborough University, as a lecturer in computing in June 2000 and became a senior lecturer in February 2004. He currently heads the Digital Imaging Research Group of Loughborough University. His research interests include image and video coding and processing, computer graphics, pattern recognition, and signal processing. He is the author of more than 50 conference and journal papers.