Stereo video coding based on quad-tree decomposition of B–P frames by motion and disparity interpolation

J.N. Ellinas and M.S. Sangriotis

Abstract: A new optimised technique for coding stereoscopic image sequences is presented and compared with already known methods. The proposed technique, called enhanced interpolated motion and disparity estimation (EIMDE), is based on the joint method, which encodes the frames of the right image sequence by exploiting both the temporal redundancy of the same sequence and the disparity redundancy with the left image sequence. In the proposed method, a variable block size scheme has been employed for motion and disparity estimation. The block size is controlled by quad-tree decomposition of the processed frame based on a rate-distortion splitting criterion. For the prediction of a macroblock, optimised motion and disparity vectors are jointly estimated and the participating proportion of each similarity is suitably searched. In this way, the energy of the resulted residual frame is minimised and the whole framework is optimised. Finally, the residual frame is decomposed by a discrete wavelet transform and is further compressed by morphological encoding the resulting coefficients. The proposed coder has been experimentally evaluated on real image sequences, where it produced good performance over other known methods.

1 Introduction

Stereoscopic vision is based on the projection of an object on two slightly displaced image planes and has an extensive range of applications, such as 3-D television, 3-D video applications, robot vision, virtual machines, medical surgery and so on. Two pictures of the same scene taken from two nearby points form a stereo pair and contain sufficient information for rendering the captured scene depth. The above demanding application areas require the development of more efficient compression techniques of a stereo image pair or a stereo image sequence. In a monoscopic video system the compression is based on the intra-frame and inter-frame redundancy. Typically the transmission or the storage of a stereo image sequence requires twice as much data volume as a monoscopic video system. Nevertheless, in a stereoscopic system a more efficient coding scheme may be developed if the inter-sequence redundancy is also exploited. A typical compression scenario includes the effective prediction of the right sequence frames based on both motion and disparity estimation.

In general, there are two methods of implementing either motion or disparity estimation in monoscopic or stereoscopic compression applications. The first method, based on intensity processing, handles this estimation by the block-matching algorithm (BMA) [1]. The target frame is divided into blocks of fixed size (FBS) and is matched under a matching criterion, which may be the mean square error (MSE) or the mean absolute difference (MAD), which minimises a cost function. The result of employing this technique is to predict the target frame and subsequently code the difference between the initial and the predicted target frame. This difference is called residual frame. The positions of the best matching blocks are denoted by a set of vectors that are also coded. Several compression algorithms have been developed that use block matching or alternative implementations, including hierarchical disparity estimation [2], multiresolution block matching [3], block matching with geometric transform [4], DWT with morphological coefficient-to-coefficient stereo coding [5] and so on.

The second method for either motion or disparity estimation, based on object segmentation, firstly defines or derives the features of the participating objects in the processed frame and then estimates the temporal or disparity field between corresponding frames [6]. Apart from the above-mentioned methods, several others have been proposed that try to improve the performance or combine their characteristics, such as the segmentation-based coding [7], stereo image projection [8], post-compensation residual coding [9], overlapped block disparity compensation [10], a hybrid scheme between block and object based technique [11] and so on.

Of the two methods described above, the former, which estimates either motion or disparity with blocks of fixed size, is the most commonly used because of its simplicity. However, it fails to code homogeneous regions efficiently. Segmenting the frame into blocks of variable sizes that incorporate homogeneous motion and/or disparity characteristics may improve the coding efficiency by allocating fewer bits to larger regions. The segmentation of a region is performed and proves efficient only if there is a reduction of a total rate-distortion cost.

Several schemes have been proposed for the compression of stereoscopic image sequences. In [6] the proposed

IEE Proceedings online no. 20045033

doi: 10.1049/ip-vis:20045033

Paper first received 25th May 2004 and in revised form 7th March 2005. Originally published online 5th July 2005

The authors are with the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Panepistimiopolis, Ilissia, 157 84 Athens, Greece

E-mail: iellinas@di.uoa.gr

3D motion estimation methods are integrated in a stereoscopic inter-frame coding scheme. In [12] an object-based coding algorithm is proposed relying on modelling of objects with 3D wire-frames. A study similar to the present paper proposes the successive exploitation of motion and disparity redundancy based on a disparity segmentation scheme [7,13]. These methods employ rectangular segmentation of the disparity field in a multiresolution framework, in order to improve coding efficiency of the right image. The suggested framework is wavelet decomposition, in which disparity is estimated among the subbands in a hierarchical sense. Another scheme proposes the MPEG-like encoding of the left stream and the joint motion and disparity estimation of the right frames with blocks of fixed size [14].

The benefits and preferences of 3DTV in many applications, where there is a viewing enhancement because of depth impression, have imposed a shift of technology from monocular to binocular vision. Thus, a lot of effort has gone into developing efficient compression algorithms for stereo images and image sequences. The embedding of the multiview profile (MVP) into the MPEG-2 coding standard is towards that direction [15]. Also, the recent emergence of the highly efficient advanced video codec H.264/AVC [16] has provoked the expectation of an equally efficient stereo codec. In parallel, JPEG-2000 and MPEG-4 show a trend to replace DCT-based techniques with DWT-based schemes.

Following these trends, we propose a DWT-based implementation that aims to combine a robust morphological coder with a new disparity compensation scheme based on motion-disparity interpolation. This new stereo image sequence compression scheme is called enhanced interpolated motion and disparity estimation (EIMDE) and belongs to intensity processing methods. The left image sequence is MPEG-like encoded, whereas P and B frames of the right image sequence are predicted by a joint motion-disparity interpolative scheme and segmented into variable size macroblocks. The processed frame is initially segmented into blocks of homogeneous intensity by its quad-tree decomposition with an intensity difference threshold criterion. These blocks may probably belong to the same object or the background of an image and may present homogeneous motion or disparity characteristics. Then, quad-tree decomposition follows with a simplified rate-distortion criterion that permits splitting if there is a rate-distortion benefit. During the segmentation of a processed right B-frame, each macroblock is predicted from a weighted prediction of bidirectional (forward and backward) motion predicted macroblocks of I–P frames and from the disparity predicted macroblock of the corresponding left frame. In the same way, the macroblocks of the right P-frames are predicted by interpolating the motion predicted macroblocks from I or P frames and the corresponding disparity predicted macroblocks. The performance of the proposed interpolative scheme is further enhanced by using a suitable search method for the estimation of the best joint motion and disparity vectors and by optimising the weighting factors of the participating frames. Finally, the left and the residual right frames are decomposed using a DWT. The transform coefficients, after their morphological representation and partitioning, are encoded using arithmetic coding that practically achieves the theoretical entropy bound. The motion and disparity vectors, which follow the same partitioning as the transform coefficients, are losslessly transmitted using DPCM and arithmetic encoding.

Typical stereoscopic applications, such as 3DTV or 3D video entertainment, require excessive bandwidth. The aim of the proposed coder is to tackle these increased bandwidth requirements and to keep a fair trade-off between quality and bandwidth. The use of a robust and high efficient wavelet-based morphological coder serves the need of bandwidth reduction in conjunction with more effective motion-disparity compensation. This coder presents excellent compression efficiency, low complexity, fast execution and embedded bit-streams. The proposed motion-disparity compensation exploits the high degree of correlation between the same scene content of the sequences, achieving: frame manipulation with blocks of variable size, allocation of fewer bits to larger homogeneous regions and less annoying artefacts. Also, subband coding methods exploit the non-uniform distribution of energy across the different frequency bands. Since the entire frame is filtered and subsampled to obtain the subbands, these methods do not suffer from blocking artefacts that are common in block-based transform coding methods. The inherent advantages of the wavelet transform are the creation of almost decorrelated coefficients, energy compaction and variable resolution.

2 Overview

2.1 Disparity in stereoscopic vision

The distance between two points of a superimposed stereo pair that correspond to the same scene point is called disparity [17]. Disparity compensation is the process that estimates this distance (disparity vector or ***DV***), predicts the right image from the left one and produces their difference or residual image (disparity compensated difference or *DCD*). The equation that describes disparity compensation, employing the block-matching algorithm (BMA), is:

$$DCD(b_{i,j}) = \sum_{(x,y)\in b_{i,j}} \left[b^R_{i,j}(x,y) - \tilde{b}^L_{i,j}(x+dv_x, y+dv_y) \right] \quad (1)$$

where $b^R_{i,j}$, $\tilde{b}^L_{i,j}$ are the corresponding blocks of the right and the reconstructed left frames, respectively, and dv_x, dv_y are the disparity vector components for the best match that are defined as follows:

$$\boldsymbol{DV}(b_{i,j}) = [dv_x, dv_y]^T = \arg\min_{(dv_x,dv_y)\in S} |DCD(b_{i,j})| \quad (2)$$

where S is the window search area. The selected matching criterion is MAD, because it is less computational expensive and less sensitive to noise than MSE. Estimation of motion is similar to disparity estimation and concerns displacement estimation between two image frames that are offset temporally [18]. Thus, motion compensation excludes temporal redundancies and produces the displaced frame difference (*DFD*) with motion vectors (***MV***), which are correspondingly described by (1) and (2).

2.2 Stereoscopic image sequence compression

The typical methods of stereoscopic video coding are shown in Fig. 1. The simulcast method is based on transmission and reproduction of independently coded channels. Independent coding reduces complexity but requires twice the bandwidth of a single transmission channel. The compatible method utilises MPEG-like encoding for the left sequence and exploits the spatial correlation between corresponding frames of the two sequences. The joint method employs MPEG-like encoding for the left sequence and exploits both the temporal and spatial redundancy of the right frames. MPEG-like encoding means that GOP structure and motion

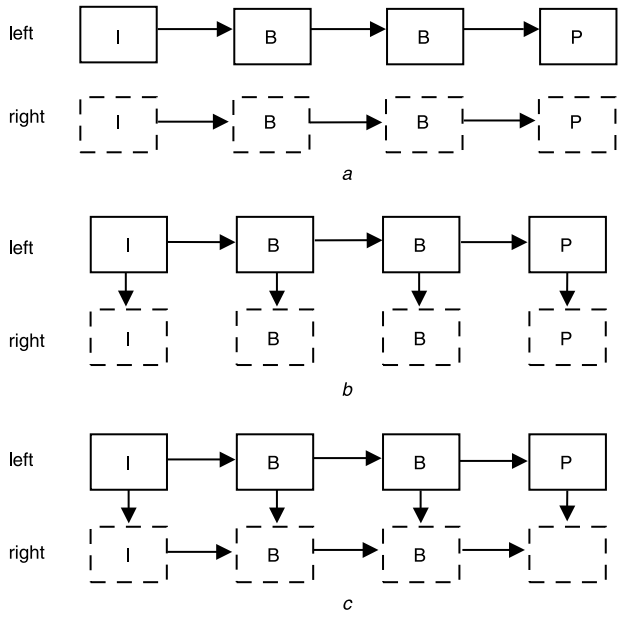


Fig. 1 *Typical methods of stereoscopic video coding*
a Simulcast
b Compatible
c Joint

estimation are treated in a way that is compatible with MPEG1 or MPEG2-MP coding standards.

2.3 Proposed architecture

The proposed method is an optimised version of joint that predicts right P and B frames by interpolating the motion and disparity fields of the participating frames. The characteristics of EIMDE can be summarised as:

- The left sequence is MPEG-like encoded.
- The right I-frames are encoded using the disparity predicted frames.
- The right P- and B-frames are encoded by interpolating the motion predicted frames from the right sequence and the disparity predicted frame from the left sequence. For the first B-frame of a GOP, the initial setting of the motion weighting factors is 35 and 15% for forward and backward prediction, respectively. The motion weighting factors for the second B-frame are in the reverse order. The disparity weighting factor is set to 50%.

Figure 2 shows the architecture employed for coder and decoder. The left frames are coded and reconstructed within the coder, so that the joint motion-disparity compensation is performed in a closed-loop mode. This is a common technique used in video coding as it provides less reconstruction distortion at the decoder's side. At the coder, both sequences are initially subjected to motion compensation. The resulting motion vectors of the right frames together with the disparity vectors, which follow from the disparity estimator, are combined in the interpolation unit and provide the residual frame. The interpolation unit comprises the proposed disparity compensation that is based on frame segmentation. The residual frame, in turn, is decomposed by DWT, quantised, morphologically encoded and finally guided to an entropy coder in order to form the transmission bit-stream. At the decoder, the incoming sequences follow the reverse order for the reconstruction of the initial frames.

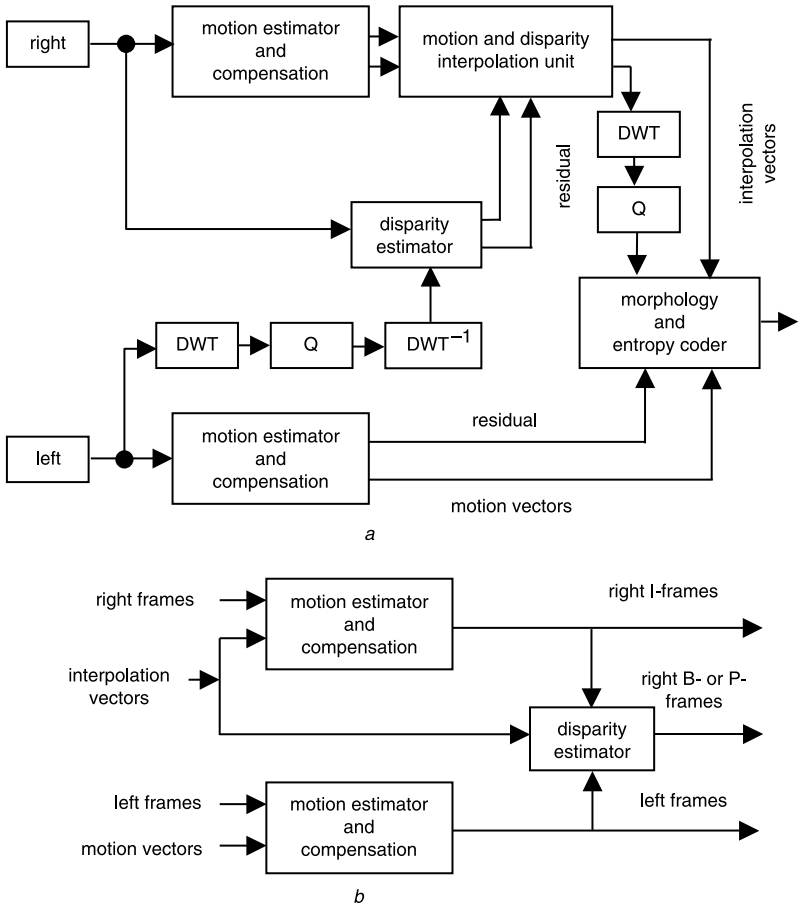


Fig. 2 *EIMDE architecture of stereoscopic video coding*
a Coder
b Decoder

2.4 Coding based on morphological representation of DWT coefficients

The conventional wavelet image coders decompose a 'still' image into multiresolution bands [19], providing better compression quality than the existing DCT transform. An alternative adaptive wavelet packet scheme can enhance the benefits of this transform [20]. This type of coder is subjected to the fact that they include all the coefficients in the transmitted sequence, even those that are zero or nearly zero, and their absence would have little effect on the reconstructed image quality. The statistical properties of the wavelet coefficients have led to the development of some very efficient algorithms including the embedded zero tree wavelet coder (EZW) [21], the coder based on set partitioning in hierarchical trees (SPIHT) [22], the coder based on the morphological representation of wavelet data (MRWD) [23], and the embedded block coding with optimised truncation of the embedded bit streams (EBCOT) [24].

The MRWD algorithm, which is used in the present work, exploits the intra-band clustering and inter-band directional spatial dependency of the wavelet coefficients in order to capture and separate them into significant or non-significant partitions. Their prediction is conducted in a hierarchical manner from the coarsest to finer scales by a morphological dilation operation that uses a 3×3 structuring element. This partitioning of wavelet coefficients reduces the overall entropy and the bit-rate becomes smaller than the non-partitioned case.

This coder is selected in the present work because it presents excellent compression efficiency, low complexity, fast execution and embedded bit-streams. The specific coder has, for 'still' images, a better performance of about 1 dB over the popular EZW [21]. The EZW compression algorithm outperforms significantly DCT at low bit-rates and provides reconstructed images free of blocking artefacts. A 2.4 dB better performance at 0.39 bpp has been reported [21]. In lossy compression, the wavelet-based JPEG 2000 is 10–20% better than DCT-based JPEG for high-quality imaging applications (at 0.5–1.0 bpp) [25]. The wavelet-based coders turn out to be particularly well-suited to capturing both the transient high-frequency phenomena, such as image edges, and long spatial duration low-frequency phenomena such as image backgrounds.

The proposed algorithm employs four-level wavelet decomposition with symmetric extension, based on the 9/7 biorthogonal Daubechies filters, for both left and right frames of the stereo pair. The linear phase of these filters together with the symmetric extension ensures minimum distortion across the image's boundaries [26].

3 Proposed interpolative scheme for stereoscopic video coding

3.1 Interpolative procedure in stereoscopic video

In a stereoscopic video, the predicted right B-frames are estimated by interpolating the motion predicted frames from the reconstructed right I- or P-frames and the disparity predicted frame from the corresponding reconstructed left B-frame. As Fig. 3 shows, the interpolation must be applied to every macroblock of the frame and this is described as follows:

$$\hat{b}(\boldsymbol{v}_f, \boldsymbol{v}_b, \boldsymbol{v}_d) = w_f \tilde{b}_f(\boldsymbol{v}_f) + w_b \tilde{b}_b(\boldsymbol{v}_b) + w_d \tilde{b}_d(\boldsymbol{v}_d) \quad (3)$$

where b, $\boldsymbol{v}$, w stand for macroblock, vector, weighting factor and indexes f, b, d denote forward, backward, disparity,

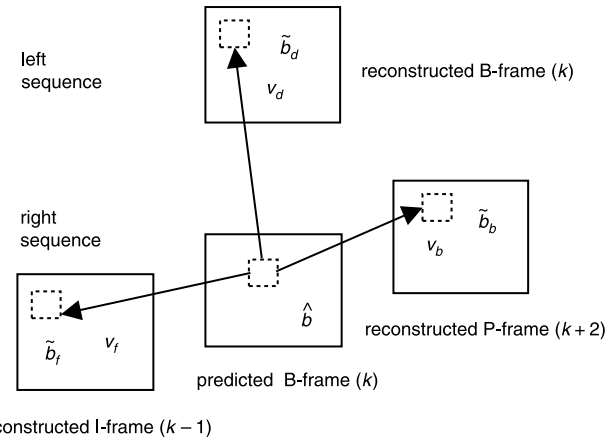


Fig. 3 *Interpolative scheme for stereoscopic video coding*
Predicted macroblock of B-frame is estimated through interpolation of motion and disparity-related macroblocks

respectively. The $(\hat{})$ sign denotes the predicted macroblock. The $(\sim)$ sign indicates that the macroblocks are the reconstructed ones.

The aim is to find the vectors that provide the best prediction of the right frame:

$$\left(\boldsymbol{v}_f^{opt}, \boldsymbol{v}_b^{opt}, \boldsymbol{v}_d^{opt}\right) = \arg\min_S |b - \hat{b}(\boldsymbol{v}_f, \boldsymbol{v}_b, \boldsymbol{v}_d)| \quad (4)$$

where S is the search area for block matching. Usually, a good approximation for the solution of (4), which reduces the number of the performed matches, comes from the best independent matching of the macroblocks that participate in the interpolation, that is:

$$\begin{aligned} \boldsymbol{v}_b^{opt} &= \arg\min_S |b - \tilde{b}_b(\boldsymbol{v}_b)| \\ \boldsymbol{v}_f^{opt} &= \arg\min_S |b - \tilde{b}_f(\boldsymbol{v}_f)| \\ \boldsymbol{v}_d^{opt} &= \arg\min_S |b - \tilde{b}_d(\boldsymbol{v}_d)| \end{aligned} \quad (5)$$

However, this solution is sometimes suboptimal. A better solution may be obtained by the joint full search that has been developed for monoscopic video [27]. The extension of the joint search method to stereoscopic video is as follows:

- The best independent vectors and the minimum difference according to (5) are estimated, with the weighting factors set to a predefined value.
- Successively, two vectors are kept constant and the third vector is updated for a lower minimum of the difference.
- The above procedure is iterated until the minimum value no longer becomes lower.

3.2 Estimation of weighting factors in interpolative scheme

The normalised energy of a macroblock b_{ij} in a residual B-frame is:

$$E_{ij} = \frac{1}{mby \times mbx} \sum_{k=1}^{mby} \sum_{l=1}^{mbx} \{b_{ij}(k,l) - \hat{b}_{ij}(k,l; \boldsymbol{v}_f, \boldsymbol{v}_b, \boldsymbol{v}_d)\}^2 \quad (6)$$

where, $\hat{b}_{ij}$ is the predicted macroblock from (3) after the previously described vector optimisation procedure. The total normalised energy of the residual frame is:

$$E_{tot} = \sum_{i=1}^{I} \sum_{j=1}^{J} E_{ij} \quad (7)$$

where $I = M/mby$ and $J = N/mbx$, for an image of size $M \times N$ and variable size macroblocks of $mby \times mbx$ pixels.

It is apparent that the energy of the residual B-frame depends on the proper selection of the weighting factors between motion and disparity. For a given bit-rate, the reduction of this energy provides higher PSNR for the reproduced frame, as the resulting distortion is decreased.

In a monoscopic video, the weighting factors for the frames that participate in an interpolative procedure are usually $w_f = w_b = 0.5$. However, it is reasonable that their values are inversely proportional to the time interval of the processed frame from the interpolated frames [28]. In this work, the motion weighting factors, which are forward and backward, are kept in a constant relation, 7:3 for the first and 3:7 for the second B-frame of every GOP. Correspondingly, the motion and disparity weighting factors are equal to 0.5 for all the P-frames. The relation between motion and disparity weighting factors is adjusted so that the energy of the residual frame is minimised, but their sum must be unity, i.e. $w_f + w_b + w_d = 1$.

Basically, the relation between motion and disparity weighting factors should be adjusted for every macroblock and their choice must minimise the energy of each residual macroblock through the previously described vector optimisation. Instead of estimating the weighting factors for every macroblock, which is time consuming and bit-rate expensive, a suboptimal scheme is proposed. The motion weighting factors are initially considered to play a role equal to that of the disparity weighting factor. Therefore, for a B-frame, $w_f + w_b = 0.5$ and $w_d = 0.5$. Considering the aforementioned relationship between forward and backward prediction, the motion weighting factors are set to $w_f = 0.7 \times 0.5 = 0.35$ and $w_b = 0.3 \times 0.5 = 0.15$ for the first B-frame and $w_f = 0.15$, $w_b = 0.35$ for the second B-frame. These values are applied to every macroblock and for all B-frames in a GOP. Also, for all the P-frames in a GOP, the motion and disparity weighting factors are set to 0.5. The proposed algorithm, which involves quad-tree decomposition of a P or B frame with a rate-distortion splitting criterion, employs the previously described interpolative scheme with the initial values of the weighting factors. These values are adjusted so that the total energy of the resulting residual frame is minimised according to (7). For example, in some B-frames it is found that the energy of the residual frame is minimised when motion and disparity contribute to the interpolative scheme by 60 and 40%, respectively. Thus, $w_f = 0.7 \times 0.6 = 0.42$ and $w_b = 0.3 \times 0.6 = 0.18$ for the first B-frame and $w_f = 0.18$, $w_b = 0.42$ for the second B-frame.

3.3 Proposed method of stereo video coding

The proposed method of coding a stereoscopic video, EIMDE, is an enhanced method of the IMDE scheme proposed in [14]. The proposed motion and disparity compensation is based on the segmentation of a frame, given the motion and disparity corresponding frames and achieves a coding representation that is commensurate with the local motion and disparity detail. Typical stereoscopic sequences consist of frames that contain areas of almost constant motion or disparity. The motion or disparity estimation schemes based on blocks of fixed size divide these areas into small blocks creating more motion or disparity vectors than those actually needed. To overcome this drawback, a joint motion-disparity estimation based on a quad-tree segmentation of the right frames is proposed. The following relation provides the residual of a processed frame, either B or P:

$$B_{res} = B_R - B_{pr} \quad (8)$$

where, B_R is the right frame, B_{res} is the residual frame and B_{pr} is the prediction of the right frame consisting of the predicted macroblocks $\hat{b}_{ij}$ that are estimated by (3).

In the proposed algorithm the residual frame is estimated by (8) but the predicted macroblocks are of variable size, according to the quad-tree decomposition of the processed frame. The summary of this new method is as follows:

- The B_R frame is quad-tree decomposed using an intensity difference splitting criterion. According to this criterion, a block splits into four children blocks if the maximum value minus the minimum value of the block elements is greater than a threshold. The threshold is defined as a value between 0 and 1 multiplied by 255 for greyscale images. The lowest permissible block size is set to 8×8 pixels, whereas blocks of larger size are formulated. The largest possible block size is half the frame's dimension, because of quad-tree splitting. In this way, the intensity homogeneous regions are located.
- The resulting 8×8 blocks are located at the boundaries of frame objects, where there are larger intensity gradients. Their predictions are estimated by (3) with the proposed interpolative method. Although the smallest block could have any dimension, it has been found that 8×8 blocks provide a good trade-off between accuracy of the estimation and the number of bits necessary to encode the motion or disparity vector for each block.
- The quad-tree analysis is continued for the blocks of larger size but with a different splitting criterion. The splitting criterion for a node is the cost of the residual for this node, defined by the following relations:

$$J_p = D_p + \lambda R_p \quad (9)$$

$$J_{ch} = \sum_{k=1}^{4} \{D_c(k) + \lambda R_c(k)\} \quad (10)$$

where J_p and J_{ch} are the costs of parent and children nodes, respectively. The Lagrange multiplier λ defines the relation between distortion and bit-rate and its value affects the segmentation depth of the processed frame. The distortion D is the MSE for the specific node. The rate R is defined as:

$$R = r_v + r_{res} \quad (11)$$

where $r_v = r_f + r_b + r_d$ and r_{res} are the bit-rates of the vectors (motion and disparity) and the residual, respectively.

The residual of a node is estimated by the proposed interpolative scheme and (8). Therefore, a parent node splits into four child nodes if and only if the cost of the parent is greater than the cost of the children.

The r_v increases when splitting occurs whereas r_{res} and D both decrease. The splitting criterion can be formed as:

$$D_p + \lambda R_p > \sum_{k=1}^{4} D_c(k) + \lambda \sum_{k=1}^{4} R_c(k) \quad (12)$$

$$D_p - \sum_{k=1}^{4} D_c(k) > \lambda \left\{ \sum_{k=1}^{4} [r_v^c + r_{res}^c] - [r_v^p + r_{res}^p] \right\} \quad (13)$$

Equation (13) is reduced to:

$$\Delta D + \lambda \sum_{k=1}^{4} (r_{res}^{p} - r_{res}^{c}) > \lambda \sum_{k=1}^{4} (r_{v}^{c} - r_{v}^{p}) \quad (14)$$

Considering that $r_v^c > r_v^p$ and $r_{res}^p > r_{res}^c$, (14) becomes:

$$\Delta D + \lambda \Delta r_{res} > \lambda \Delta r_v \quad (15)$$

which is satisfied if the following relation is valid:

$$\Delta D > \lambda \Delta r_v \quad (16)$$

This suggests that a parent node splits into four children if the benefit from the distortion is greater than the benefit from the vector's bit-rate.

• After the completion of quad-tree analysis, the right frame consists of variable size macroblocks each of which has smooth motion and disparity characteristics. In this way, the resulting residual frame is less distorted. The initial preset relationship of the weighting factors between motion and disparity is trimmed so that the energy of the entire residual frame is minimised.

4 Experimental results

The proposed coder, which estimates the residual P and B frames by employing a motion-disparity interpolative scheme, was tested on three stereoscopic image sequences, namely 'crowd', 'book-sale' and 'Sergio' [29]. The first two stereo sequences are the only ones available with a sufficient temporal length of 169 and 89 frames, respectively, whereas the third one has a temporal length of ten frames. The size of each frame is 320 × 240 pixels for the first two sequences and 512 × 512 pixels for the third one, the type of sequence is IBBPBBPBB, the largest macroblock size is half the frame's dimension and the smallest macroblock size is 8 × 8 pixels. Both motion and disparity compensation employ the classical block-matching algorithm. The searching area is 12 pixels (six pixels around the macroblock, which is a typical value for MPEG and H.261/H.263) and the matching criterion is MAD. The size of the search space and the selected cost function counterbalance the complexity of the exhaustive search algorithm used in the proposed coder. Additionally, the tested sequences have small camera and object displacements that are fairly supported by this selection.

The objective quality measure of the reconstructed right frames is estimated by peak signal-to-noise ratio (PSNR):

$$\mathrm{PSNR} = 10 \log_{10} \frac{255^2}{MSE_R} \quad (17)$$

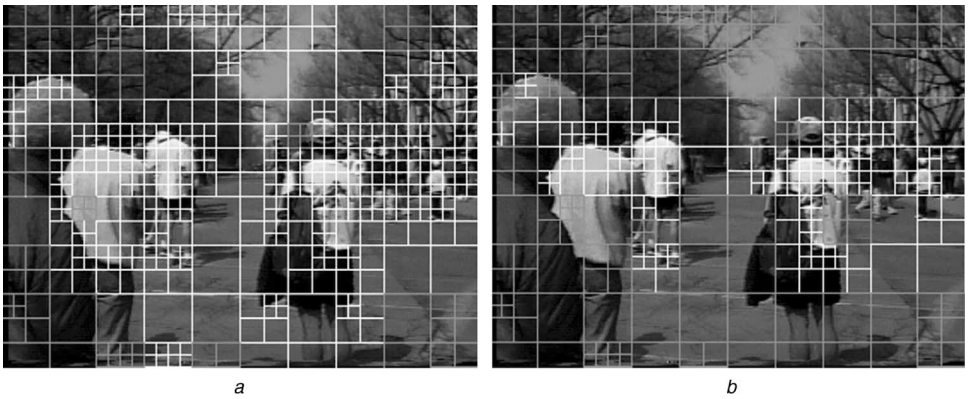
where MSE_R is the mean square error of the right frames.

The DWT subband coefficients of the left and residual frames are quantised, are partitioned by the morphological encoder and are arithmetically encoded. The motion and disparity vectors are partitioned by the morphological encoder and are losslessly transmitted using DPCM and arithmetic encoding. The experimental results refer to P and B frames of the right sequence. The coding of the right I-frames is based only on disparity compensation for all the above mentioned methods. This is because I-frames are access index points of a video stream, as in the MPEG standard of a monoscopic video and are coded in intra-frame mode that does not employ temporal prediction.

Figure 4*a* illustrates the quad-tree decomposition of a B-frame with an intensity difference splitting threshold of 0.6, whereas Fig. 4*b* shows the frame's segmentation with a threshold of 0.8. The Lagrange multiplier is set to a value of $\lambda = 500$. The white-line blocks show decomposition with an intensity criterion, whereas the black line blocks show decomposition according to a rate-distortion splitting criterion that is provided by inequality (16). The selection of the intensity difference threshold defines the number of blocks that will be further processed with the rate-distortion splitting criterion. As this threshold increases, more blocks are subjected to the rate-distortion splitting criterion, giving slightly better performance, but the computational complexity increases as well. In fact, the intensity quad-tree decomposition is used as a first step for the location of homogeneous regions that may or may not be further decomposed. The high-intensity gradient regions are excluded from the rate-distortion decomposition and they are treated by the interpolative scheme as fixed size blocks of 8 × 8 pixels.

The Lagrange factor selection affects the depth of rate-distortion decomposition. As this factor becomes large enough, the segmentation criterion of inequality (16) is no longer valid and a smaller number of blocks are further segmented. The experimental results of this work were attained with an intensity threshold of 0.6 and unity Lagrange factor.

Table 1 provides the average objective performance of the right sequence of compatible, IMDE and EIMDE compression methods at bit rates of 0.70 and 1.15 Mbit/s (0.30 and 0.50 bpp) and frame rate of 30 frame/s. The bit rate for the sequence of 'Sergio' is provided only in bpp because of its restricted temporal length. The motion and



a *b*

Fig. 4 *B-frame segmentation with different thresholds*

a Intensity difference threshold of 0.6
b Intensity difference threshold of 0.8

Table 1: Average performance evaluation of compatible, IMDE and EIMDE methods of stereoscopic coding at bit rates of 0.70 and 1.15 Mbit/s

		Right image sequence					
		B			P		
Sequence	Method	PSNR	Mbit/s	bpp	PSNR	Mbit/s	bpp
Crowd	compatible	30.89	0.74	0.32	30.06	0.83	0.36
	IMDE	31.45	0.76	0.33	30.99	0.69	0.30
	EIMDE	32.35	0.76	0.33	32.15	0.71	0.31
Book-sale	compatible	31.09	0.85	0.37	30.13	0.99	0.43
	IMDE	32.01	0.81	0.35	31.23	0.76	0.33
	EIMDE	32.89	0.78	0.34	32.57	0.78	0.34
Sergio	compatible	41.35	-	0.34	39.80	-	0.31
	IMDE	41.87	-	0.34	41.16	-	0.29
	EIMDE	43.79	-	0.32	43.19	-	0.30
Crowd	compatible	34.53	1.13	0.49	34.10	1.52	0.66
	IMDE	35.04	1.15	0.50	34.79	1.20	0.52
	EIMDE	35.43	1.13	0.49	35.26	1.20	0.52
Book-sale	compatible	35.00	1.24	0.54	34.45	1.77	0.77
	IMDE	35.75	1.15	0.50	35.28	1.31	0.57
	EIMDE	36.42	1.15	0.50	35.81	1.27	0.55
Sergio	compatible	45.17	-	0.43	44.21	-	0.52
	IMDE	45.81	-	0.44	45.31	-	0.45
	EIMDE	47.73	-	0.41	47.33	-	0.44

disparity estimation is performed with a pixel precision. The use of half-pixel accuracy in the proposed coder is estimated to produce on average a gain of 0.3 dB at the expense of computing complexity.

It is apparent that the proposed method (EIMDE), for the tested stereo sequences, provides better overall performance compared with the compatible method of coding and improved performance compared to the IMDE method. At a bit-rate of 0.76 Mbit/s (0.33 bpp), PSNR of EIMDE for a B-frame is about 1.5 dB better than that of the compatible and 0.9 dB better than that of IMDE for the 'crowd' sequence. At the same bit-rate, PSNR of EIMDE for a P-frame is over 2 dB better than that of the compatible and 1.2 dB better than that of IMDE. The same performance is attained for the 'book-sale' sequence. It may be observed that EIMDE performs even better in the third stereo sequence of 'Sergio'. At a bit rate of 0.32 bpp, EIMDE outperforms compatible by 2.4 dB and IMDE by 1.9 dB on the average for B-frames. The interpretation lies in the fact that the background of the scene is a uniform area of grey colour. Thus, the proposed frame segmentation scheme creates larger blocks with constant motion-disparity characteristics and therefore the algorithm behaves more efficiently. Also, the only moving object of the scene is subjected to small displacements.

Figure 5 shows PSNR against bit-rate of a right B- and P-frame for the 'crowd' stereo sequence. As illustrated, the rate-distortion graphs show that the proposed stereo coding algorithm outperforms the other methods both at medium and low bit-rates. Figure 6 shows the subjective performance of the proposed EIMDE coder with respect to compatible coder. The reconstructed quality of the second right B-frame for the 'crowd' sequence at a bit rate of 0.46 Mbit/s (0.2 bpp) is illustrated. The observed subjective quality degradation near object boundaries is of ringing nature and is an inherent property of the wavelet transformation. The variable size block segmentation of

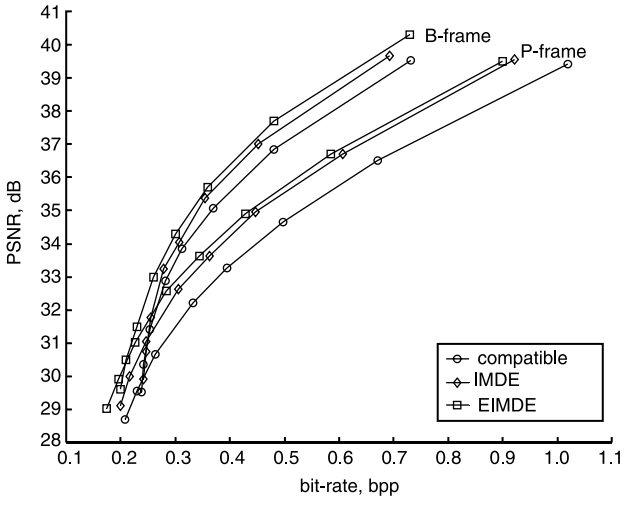


Fig. 5 *Objective quality measure of right B- and P-frame for 'crowd' sequence*

the processed right frames makes this distortion non-periodic and thus less annoying.

Table 2 gives the performance of our proposed coder and MPEG-2 for the 'crowd' and 'book-sale' test sequences. Employing the high-profile double-layer structure, the base layer supports the left stereoscopic sequence while the enhancement layer manages the disparity predicted right sequence. The experimental results show that the proposed coder EIMDE provides comparable or superior performance against MPEG-2 at medium and low bit-rates. It has to be mentioned that the proposed coder employs an MPEG-like encoder that does not include the advanced features of MPEG-2, such as nonlinear quantisation, motion estimation with sub-pixel accuracy etc.

It should also be noted that H.264/AVC provides more than 50% bit-rate savings than MPEG-2 for the same quality but does not support stereoscopic coding. The aim of the

a *b*

Fig. 6 *Subjective quality of second right B-frame at 0.46* Mbit/s *(0.2 bpp) for 'crowd' sequence*
a EIMDE coder
b Compatible coder

Table 2: Performance evaluation of EIMDE and MPEG-2 coders for stereoscopic coding at bit rates of 0.70 and 1.15 Mbit/s

		Right image sequence		
Sequence	Method	PSNR (dB)	Mbit/s	bp p
Crowd	EIMDE	33.38	0.83	0.36
	MPEG-2	34.10	0.83	0.36
Book-sale	EIMDE	32.15	0.71	0.31
	MPEG-2	32.50	0.90	0.39
Crowd	EIMDE	35.06	1.24	0.54
	MPEG-2	35.40	1.18	0.51
Book-sale	EIMDE	35.80	1.20	0.52
	MPEG-2	34.5	1.27	0.55

present work is to propose a new framework of stereo coding and show that it outperforms the already existing algorithms based on simulcast, compatible or joint methods. The adaptation of the proposed scheme to the advanced coding capability of H.264 is an issue for further research.

Finally, the rate-distortion algorithm, which is applied on P and B frames of the right sequence, employs the proposed segmentation procedure in order to create blocks of variable size so as to spend fewer bits on homogeneous regions. The interpolative scheme together with the weighting factors adjustment aim to reduce the distortion of the residual frame. The experimental evaluation proves the effectiveness of the proposed EIMDE method.

5 Conclusions

A stereoscopic image sequence can be encoded in an effective way if redundant information that exists between the frames of the same sequence (temporal redundancy) and the corresponding frames of the two sequences (disparity redundancy) is taken into account. Among the typical methods of coding, the most attractive are the compatible and the joint. The compatible method compresses the right frames by taking into account the spatial redundant information in relation to the left frames. The joint method compresses further the right P- and B-frames by taking into account both temporal and spatial redundancy. While the two above mentioned methods have been formulated in the literature, the optimised framework of the proposed interpolative scheme has not, to the best of the author's knowledge, been addressed.

The proposed scheme, enhanced IMDE, is an optimised version of the joint method and an enhanced version of IMDE method that predicts P- or B-frames of the right sequence by employing joint motion-disparity interpolation. It initially segments the processed P- or B-frame by employing an intensity difference splitting threshold, in order to localise the intensity homogeneous blocks. Then, it splits further the resulting blocks by a simplified rate-distortion relationship. The segmentation is performed through the proposed interpolative scheme that is optimised by a suitable selection of motion and disparity vectors. As a result, the processed frame is segmented into variable size blocks that present homogeneous motion-disparity characteristics and its residual is coded by using DWT transform with the MRWD compression algorithm. Furthermore, the adjustment between motion and disparity weighting factors may minimise the energy of the residual frame. The experimental results show that the proposed stereo coding scheme provides better overall performance in the whole examined range over the typical compatible method of coding and IMDE.

6 References

1 Perkins, M.G.: 'Data compression of stereopairs', *IEEE Trans. Commun.*, 1992, **40**, pp. 684–696
2 Sethuraman, S., Jordan, A.G., and Siegel, M.W.: 'Multiresolution based hierarchical disparity estimation for stereo image pair compression'. Proc. Symp. on Application of subbands and wavelets, 1994
3 Tzovaras, D., Strintzis, M.G., and Sahinoglou, H.: 'Evaluation of multiresolution block matching techniques for motion and disparity estimation', *Signal Process. Image Commun.*, 1994, **6**, pp. 59–67
4 Ghanbari, M., *et al.*: 'Motion compensation for very low bit-rate video', *Signal Process. Image Commun.*, 1995, **7**, pp. 567–580
5 Ellinas, J.N., and Sangriotis, M.S.: 'Stereo image compression using wavelet coefficients morphology', *Image Vis. Comput.*, 2004, **22**, (4), pp. 281–290
6 Tzovaras, D., Grammalidis, N., and Strintzis, M.G.: 'Object-based coding of stereo image sequences using 3D motion/disparity compensation', *IEEE Trans. Circuits Syst. Video Technol.*, 1997, **7**, pp. 312–327
7 Sethuraman, S., Siegel, M.W., and Jordan, A.G.: 'Segmentation based coding of stereoscopic image sequences', *Proc. SPIE*, 1996, **2668**, pp. 420–429
8 Aydinglou, H., and Hayes, H.: 'Stereo image coding: a projection approach', *IEEE Trans. Image Process.*, 1998, **7**, pp. 506–516
9 Moellenhoff, M.S., and Maier, M.W.: 'Transform coding of stereo image residuals', *IEEE Trans. Image Process.*, 1998, **7**, pp. 804–812
10 Woo, O., and Ortega, A.: 'Overlapped block disparity compensation with adaptive windows for stereo image coding', *IEEE Trans. Circuits Syst. Video Technol.*, 2000, **10**, pp. 194–200
11 Jiang, J., and Edirisinghe, E.A.: 'A hybrid scheme for low bit-rate coding of stereo images', *IEEE Trans. Image Process.*, 2002, **11**, (2), pp. 123–134
12 Malassiotis, S., and Strintzis, M.G.: 'Coding of video-conference stereo image sequences using 3-D models', *Signal Process. Image Commun.*, 1997, **9**, pp. 125–135

13 Sethuraman, S., Siegel, M.W., and Jordan, A.G.: 'A multiresolutional region based segmentation scheme for stereoscopic image compression', *Proc. IS&T/SPIE*, 1995, **2419**, pp. 265–274
14 Ellinas, J.N., and Sangriotis, M.S.: 'Stereo video coding based on interpolated motion and disparity estimation'. EURASIP 3rd Int. Conf. on Image and Signal Processing and Analysis, 2003, pp. 301–306
15 MPEG-2 Video Subgroup. Proposed Draft Amendment 3 Multi-view Profile. ISO/IEC/JTC1/SC19/WG11 Doc. N1088, 1995
16 Wiegand, T., Sullivan, G.J., Bjontegaard, G., and Luthra, A.: 'Overview of the H.264/AVC video coding standard', *IEEE Trans. Circuits Syst. Video Technol.*, 2003, **13**, (7), pp. 560–576
17 Starck, J.-L., Murtagh, F.D., and Bijaoui, A.: 'Image processing and data analysis: the multiscale approach' (Cambridge University Press, 1998)
18 Tekalp, A.M.: 'Digital video processing' (Prentice Hall, 1995)
19 Antonini, M., Barlaud, M., Mathieu, P., and Daubechies, I.: 'Image coding using wavelet transform', *IEEE Trans. Image Process.*, 1992, **1**, (2), pp. 205–220
20 Ramchandran, K., and Vetterli, M.: 'Best wavelet packet bases in a rate-distortion sense', *IEEE Trans. Image Process.*, 1993, **2**, pp. 160–175
21 Shapiro, J.M.: 'Embedded image coding using zero trees of wavelet coefficients', *IEEE Trans. Signal Process.*, 1993, **41**, (12), pp. 3445–3462
22 Said, A., and Pearlman, W.A.: 'A new, fast and efficient image codec based on set partitioning in hierarchical trees', *IEEE Trans. Circuits Syst. Video Technol.*, 1996, **6**, (3), pp. 243–250
23 Servetto, S.D., Ramchandran, K., and Orchard, M.T.: 'Image coding based on a morphological representation of wavelet data', *IEEE Trans. Image Process.*, 1999, **8**, (9), pp. 1161–1174
24 Taubman, D.: 'High performance scalable image compression with EBCOT', *IEEE Trans. Image Process.*, 2000, **9**, (7), pp. 1158–1170
25 Skodras, A., Christopoulos, C., and Ebrahimi, T.: 'The JPEG 2000 still image compression standard', *IEEE Signal Process. Mag.*, 2001, **18**, pp. 36–58
26 Usevitch, B.E.: 'A tutorial on modern lossy wavelet image compression: foundations of JPEG 2000', *IEEE Signal Process. Mag.*, 2001, **18**, pp. 22–35
27 Wu, S.-W., and Gresho, A.: 'Joint estimation of forward and backward motion vectors for interpolative prediction of video', *IEEE Trans. Image Process.*, 1994, **3**, (5), pp. 684–687
28 Puri, A., *et al.*: 'Video coding with motion compensated interpolation for CD-ROM applications', *Image Commun.*, 1990, **2**, (2), pp. 127–144
29 Stereo video sequences from Carnegie Mellon University, Pittsburgh, PA, USA. Available from URL: http://www-2.cs.cmu.edu/afs/cs.cmu.edu/project/sensor-9/ftp/