Stereo Image Coder Based on the MRF Model for Disparity Compensation

J. N. Ellinas and M. S. Sangriotis

Department of Informatics and Telecommunications, National & Kapodistrian University of Athens, Panepistimiopolis, Ilissia, 15784 Athens, Greece

This paper presents a stereoscopic image coder based on the MRF model and MAP estimation of the disparity field. The MRF model minimizes the noise of disparity compensation, because it takes into account the residual energy, smoothness constraints on the disparity field, and the occlusion field. Disparity compensation is formulated as an MAP-MRF problem in the spatial domain, where the MRF field consists of the disparity vector and occlusion fields. The occlusion field is partitioned into three regions by an initial double-threshold setting. The MAP search is conducted in a block-based sense on one or two of the three regions, providing faster execution. The reference and residual images are decomposed by a discrete wavelet transform and the transform coefficients are encoded by employing the morphological representation of wavelet coefficients algorithm. As a result of the morphological encoding, the reference and residual images together with the disparity vector field are transmitted in partitions, lowering total entropy. The experimental evaluation of the proposed scheme on synthetic and real images shows beneficial performance over other stereoscopic coders in the literature.

1. INTRODUCTION

The perception of a scene with 3D realism may be accomplished by a stereo image pair which consists of two images of the same scene recorded from two slightly different perspectives. The two images are distinguished as Left and Right images that present binocular redundancy, and for that reason can be encoded more efficiently as a pair than independently. The stereoscopic vision has a very wide field of applications in robot vision, virtual machines, medical surgery, and so forth. Typically, the transmission or the storage of a stereo image requires twice the bandwidth or the capacity of a single image. The objective on a bandwidth-limited transmission system is to develop an efficient coding scheme that will exploit the redundancies of the two images, that is, intraimage and cross-image correlation or similarities.

A typical compression scenario is the encoding of one image, which is called reference and the disparity compensation of the other, which is called target. In this work, the Left image is assigned as reference and the Right image as target. Transform coding is a method used to remove intraspatial redundancy both from reference and target images. The cross-image redundant information is evaluated by considering the disparity between the two images. Disparity compensation procedure estimates the best prediction of the target image from the reference and results in an error image, which is called residual, together with a disparity vector field. The encoded reference and residual images together with the disparity vectors are entropy coded and transmitted. Therefore, the effectiveness of the encoding algorithm, the energy of the residual image, and the smoothness of the disparity vector field affect the overall performance of the stereo coder.

Several methods have been developed for disparity compensation. The area-based methods, including either pixel or line or area matching, are simple approaches for disparity estimation [1, 2]. The block-based matching method, either fixed or variable size (FSBM or VSBM), finds the distance between two blocks that have similar intensities within a predefined search window [3]. The block-matching algorithm (BMA) may also be applied on the objects that appear in a stereo pair after an object contour extraction in the two images [4] or on the subbands of a wavelet decomposed stereo image pair in a hierarchical way [5]. Nevertheless, the area-matching methods, either pixel or block, often

fail to estimate disparity satisfactorily because the disparity field inherits a nonsmooth variation due to noise and the existence of occlusions. This may be improved by estimating disparity field with the Markov random field (MRF) model, which provides smoothness constraints and takes into account the occlusions [6]. Some other methods code the residual part of the predicted target image using efficient coders for "still" images, as EZW, or mixed coding [7–10]. Another method predicts the blocks transform of one image from the matching blocks transform of the other [11]. The subspace projection technique is another method that combines disparity compensation and residual coding by applying a transform to each block of the target image [12].

MRF model takes into account the contextual constraints by considering that the disparity field is smooth except near object boundaries. Hence, the value of a random variable, which may be a block of pixels, is influenced by the local neighbourhood system. The probabilistic aspect of the MRF analysis is converted to energy distribution through its equivalence to Gibbs distribution (GRF) with Hammersley-Clifford theorem. The usual statistical criterion for optimality is the maximum a posteriori probability (MAP) that provides the MAP-MRF framework. Since Gemans' classical work [13], many methods have been presented in motion estimation of a monoscopic video which is very similar to the disparity estimation case. Some works use either global or local methods for the MAP estimation problem [6, 14]. The global methods, like simulated annealing (SA), converge to a global minimum with high computational cost, whereas the local methods, like iterated conditional mode (ICM), converge quickly but they are trapped to local minima. Some other methods, based on the mean field theory (MFT), provide a compromise between efficiency and computational cost [15, 16].

A robust "still" image encoder and a disparity compensation process, which is based on the MRF/GRF model [17], are the novelties of the proposed coder. According to this model, occlusion field is initially separated into three regions by setting two threshold levels. The blocks of the intermediate region, which is called uncertain, are finally characterized as occluded or nonoccluded. This reduces the number of regions needed for the MAP search procedure, which is normally implemented in the entire occlusion field, making the algorithm simpler and faster. Also, mean absolute error (MAE) is selected instead of mean square error (MSE), in order to render our algorithm less sensitive to noise. The reference image and the resulting disparity compensated difference (DCD) or residual are decomposed by a discrete wavelet transform (DWT) and encoded by employing the morphological representation of wavelet data (MRWD) encoding algorithm [18]. The disparity vectors are DPCM entropy encoded and are embedded in the formed partitions of the morphological algorithm. The outstanding features of the proposed stereoscopic coder are the inherent advantages of the wavelet transform, the efficiency and simplicity of the employed morphological compression algorithm, and the effectiveness of the disparity compensation process.

This paper is organized as follows. In Section 2, there are overviews of the disparity compensation process, the MRF model, and the employed morphological encoder. In Section 3, the proposed algorithm is discussed and in Section 4, the experimental results are presented. Finally, conclusions are summarized in Section 5.

2. OVERVIEW

2.1. Disparity in stereoscopic vision

The problem of finding the points of a stereo pair that correspond to the same 3D object point is called correspondence. The correspondence problem is simplified into one-dimensional problem if the cameras are coplanar. The distance between two points of the stereo pair images that correspond to the same scene point is called disparity. The estimation of this distance (disparity vector or **DV**) is very important in stereo image compression because the target image (Right) can be predicted from the reference (Left) along with the disparity vectors. Then, the difference of the prediction from the original image (disparity compensated difference or DCD) is evaluated so that redundant information is not encoded and transmitted [19, 20]. Disparity compensation usually employs BMA for the estimation of a residual or DCD block:

$$\text{DCD}\,(b_{i,j}) = \sum_{(x,y) \in b_{i,j}} | \, b_{i,j}^{R}(x, y) - \tilde{b}_{i,j}^{L}(x + dv_x, y + dv_y) \, |,$$

(1)

where $b_{i,j}^{R}$, $\tilde{b}_{i,j}^{L}$ are the corresponding blocks of the Right and the reconstructed Left images, respectively and dv_x, dv_y are the disparity vector components for the best match, which is defined as

$$\mathbf{DV}\,(b_{i,j}) = \underset{(dv_x, dv_y) \in A}{\text{argmin}} \ \text{DCD}\,(b_{i,j}),$$

(2)

where A is the window searching area and the matching criterion is MAE. In this work, the general case is considered where the disparity vector has horizontal and vertical components. The above-described disparity compensation process is called closed-loop, because the prediction of the target image is performed with the reconstructed reference image. This is quite reasonable because the reconstruction of the target image will be performed with the assistance of the reconstructed reference image at the decoder's side [8]. Alternatively, disparity compensation may be performed with the reference image and is called open-loop. The open-loop systems, although they are simpler since there is no need for inverse quantization and wavelet transform at the encoder's side, are less effective. The disparity compensation process exploits the spatial cross-image dependency in order to remove redundant information. However, some blocks that have no correspondence may be encountered and are called occluded blocks. The sides of the stereo pair that cannot be seen directly by both eyes as well as the areas from object overlapping are occluded regions. The occluded regions are usually tracked and excluded during the disparity estimation
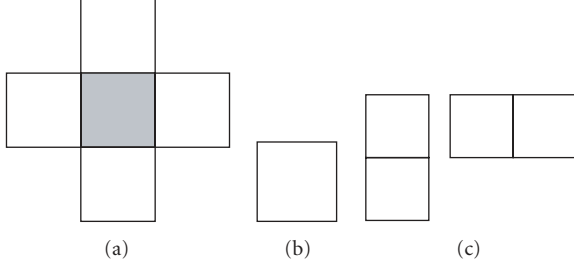
FIGURE 1: (a) First-order neighbourhood system; (b) single-site clique; (c) double-site cliques.

process, since they contribute to high distortion in the residual image. MRF model penalizes the existence of an occluded block and encourages the connectivity of neighbouring occluded blocks, as they usually appear at the boundaries of objects where large intensity gradients prevail.

2.2. The MRF/GRF model

In this section, the basic concepts of the MRF model are reviewed [21, 22]. Let

$$S = \{(i, j) \mid 1 \le i, \ j \le N\} \quad (3)$$

be a rectangular lattice of size $N \times N$, which in this case is the disparity compensated image and

$$D = \{D_{i,j}, \ (i, j) \in S\} \quad (4)$$

a family of random variables defined on S representing the random disparity field. Obviously, each disparity compensated image may be viewed as a discrete sample realization of D, with a configuration d, which is a set of each random variable. Each disparity compensated block of pixels may be viewed as a random variable in the spatial domain:

$$d = \{d_{i,j}, \ (i, j) \in S\}. \quad (5)$$

The MRF model considers a neighbourhood system N on S, which is defined as

$$N = \{N_{i,j}, \ (i, j) \in S\}, \quad (6)$$

where $N_{i,j}$ is the set of sites on the neighbourhood of the (i, j) block. The definition of the neighbourhood is as follows:

$$N_{i,j} = \{(i', j') \mid (i', j') \in S, \ (i', j') \ne (i, j), \\ (i - i')^2 + (j - j')^2 \le k\}, \quad (7)$$

where k is a positive integer defining the order of a neighbourhood system. The first-order neighbourhood $(k = 1)$, which is used in the present work, is a four-connected structuring element as shown in Figure 1.

The cliques are a subset of sites in S, where each site is a neighbour of the other sites in the defined neighbourhood

system. A family of random variables D is an MRF model on S with respect to N if the following properties are satisfied:

$$P(D_{i,j} = d_{i,j}) > 0, \quad \forall d_{i,j} \in D, \quad (8)$$

$$P(D_{i,j} = d_{i,j} \mid D_{m,n} = d_{m,n}, \ (m, n) \in S, \ (m, n) \ne (i, j)) = \\ P(D_{i,j} = d_{i,j} \mid D_{m,n} = d_{m,n}, \ (m, n) \in N_{i,j}). \quad (9)$$

Equation (9) is called Markovianity and indicates that disparity field on site (i, j) has local characteristics, that is, it depends only on neighbouring sites $N_{i,j}$. According to Hammersley-Clifford theorem [23], D is an MRF on S with respect to N if $P(D = d)$ for all configurations d is a Gibbs distribution with respect to N. Gibbs distribution has the following form:

$$P(d) = Z^{-1} \times e^{-(1/T)U(d)}, \quad (10)$$

where

$$U(d) = \sum_{c \in C} V_c(d) = \sum_{(i,j) \in C_1} V_1(d_{i,j}) + \sum_{(i,j) \in C_2} V_2(d_{i,j}) \quad (11)$$

is the energy function for d. $V_c(d)$ represents the clique potential of all possible first-order clique sets, which are single-site C_1 and double-site C_2. Normalization factor Z is called partition function and has the following form:

$$Z = \sum_d e^{-(1/T)U(d)}. \quad (12)$$

The practical value of the above is that the probability of a configuration d may be specified in terms of prior potentials $V_c(d)$ for all the cliques. Let us assume that the observation model r, the configuration d, the a priori probability $P(d)$, and the likelihood density $p(r \mid d)$ are known. Normally, the best value of d is given by an MAP estimation, which can be expressed with Bayes formula as

$$P(d \mid r) = \frac{p(r \mid d)P(d)}{p(r)}, \quad (13)$$

where $p(r)$ is the probability density function of the observation model, which does not affect the solution of (13). Therefore, an MAP solution is given by

$$\hat{d} = \underset{d \in S}{\operatorname{argmax}} P(d \mid r) = \underset{d \in S}{\operatorname{argmax}} [p(r \mid d)P(d)]. \quad (14)$$

According to Gibbs distribution equation (10), the MAP solution may be converted as follows:

$$\hat{d} = \underset{d \in S}{\operatorname{argmax}} [p(r \mid d)P(d)] \\ = \underset{d \in S}{\operatorname{argmax}} [Z_r^{-1} \times e^{-U(r|d)}][Z^{-1} \times e^{-U(d)}] \quad (15)$$

or

$$\hat{d} = \underset{d \in S}{\operatorname{argmin}} [U(d) + U(r \mid d)], \quad (16)$$
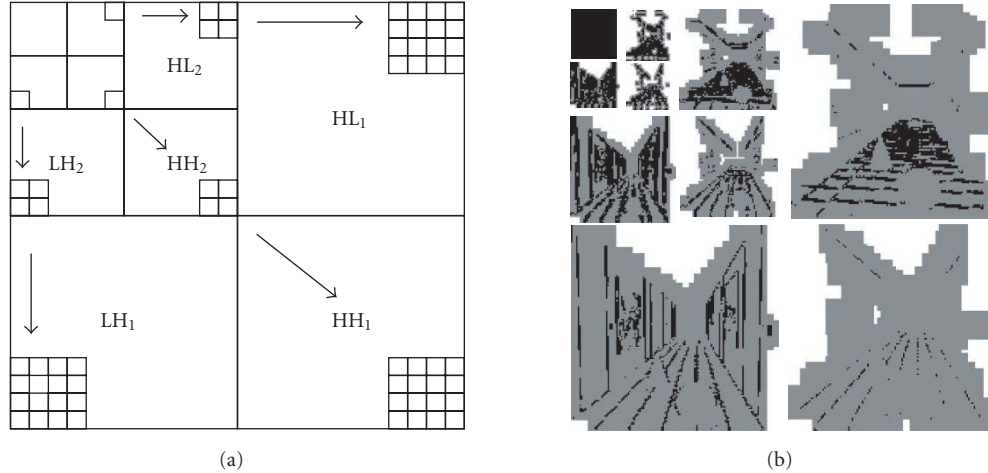
(a)

(b)

FIGURE 2: (a) Spatial dependency of significant coefficients among the subbands of a three-level wavelet decomposition; (b) partitions of significance and insignificance.

where $U(d)$ is the prior and $U(r \mid d)$ the likelihood energies. Finally, configuration d may be estimated by the minimization of energy equation (16) knowing the prior and likelihood energies for a given neighbourhood system.

2.3. The morphological encoder

The conventional wavelet image coders decompose a "still" image into multiresolution bands providing better compression quality than the so far existing DCT transform [24]. The statistical properties of the wavelet coefficients led to the development of some very efficient encoding algorithms such as the embedded zero-tree wavelet coder (EZW) [25], the coder based on set partitioning in hierarchical trees (SPIHT) [26], the coder based on the morphological representation of wavelet data (MRWD) [18], and its enhanced version called significance-linked connected component analysis for wavelet image coding (SLCCA) [27].

MRWD algorithm, which is used in the present work, exploits the intraband clustering and interband directional spatial dependency of the wavelet coefficients. This spatial dependency is shown in Figure 2(a) for a three-level wavelet transform. Hence, a prediction of the significant coefficients in a hierarchical manner is feasible starting from the coarsest scale. This may be accomplished using the morphological dilation operation with a structuring element. A dead-zone uniform step-size quantizer quantizes all the subbands and the coefficients of the coarsest detail subbands constitute either the map of significance or insignificance, that is, a binary image with two partitions in every subband. The intraband dependency of wavelet coefficients or the tendency to form clusters suggests that significant neighbours may be captured applying a morphological dilation operator. The finer-scale significant coefficients, in the children subbands, may be predicted from the significant ones of the coarser scale, parent subbands, by applying the same morphological operator to an enlarged neighbourhood, because children

subbands have double size than their parents. However, the significant partitions comprise insignificant coefficients that were captured as significant, and correspondingly the insignificant partitions comprise significant coefficients that were isolated. So, each of these two partitions may be further partitioned into two groups, so that the elements of each group have the same properties.

Figure 2(b) shows binary images of the detail subbands with the formed clusters after the aforementioned morphological operation. The black areas denote significant coefficients, the white areas denote insignificant ones, while the grey areas illustrate insignificant coefficients that are captured as significant by dilation operation with a 3×3 structuring element. The approximation subband, which contains the low-frequency components, is not subjected to this operation and all of its coefficients are considered as significant.

Consequently, the coefficients of the wavelet transform are partitioned into groups with the same characteristics and total composite entropy is lowered. The transmitted sequence of these partitions has a certain order of transmission including side information, which consists of the headers that define each partition, needed at the decoder's stage. The performance of this algorithm for "still" images is quite good with respect to other state-of-the-art compression techniques. It provides PSNR values of about 1 dB better than EZW and has about the same performance as SPIHT [18]. The morphological encoder has also the capability, by assigning a set of embedded quantizers, to produce an embedded coding which insures resolution scalability at the decoder's side.

3.   THE PROPOSED ALGORITHM

The disparity field of a stereo image pair is an MRF/GRF model consisting of disparity D and occlusion O fields. The problem is to determine disparity and occlusion fields from the observations which are the pair of images. The

configurations d and o, for disparity and occlusion fields, respectively, may be estimated by (16):

$$(\hat{d}, \hat{o}) = \underset{(d,o) \in S}{\operatorname{argmin}} \left[U(S^R \mid S^L, d, o) + U(d \mid o) + U(o) \right], \tag{17}$$

where S^L, S^R are the observations and represent the Left and Right images, respectively. The first term represents the likelihood energy, the second term represents the prior disparity field when occlusion field is given, and the third term represents the prior occlusion constraint.

3.1. The likelihood energy

The likelihood energy, which is also called similarity constraint, indicates how similar two corresponding images are when disparity and occlusion fields are known. Typically, this may be expressed as

$$U(S^R \mid S^L, d, o) = \sum_{(i,j) \in S} (1 - o_{i,j}) \sum_{(k,l) \in b_{i,j}} \left(c_{(k,l)}^R - \tilde{c}_{(k,l) \oplus d_{i,j}}^L \right)^2, \tag{18}$$

where $o_{i,j}$ is a binary indication for the presence of an occluded block, $c_{(k,l)}^R$ are the pixels of the processed block $b_{i,j}$, and $\tilde{c}_{(k,l) \oplus d_{i,j}}^L$ are the predicted pixels of the reconstructed Left block that are translated by $d_{i,j}$ in order to have a best match to the corresponding ones of the processed block. The best matching between two corresponding blocks is decided by the minimum value of their MAE.

3.2. The smoothness constraint

The prior disparity field, when occlusion field is given, is also called smoothness constraint. Minimization of the respective term in the general equation (17) provides a smooth disparity field except on the occluded points. This is expressed as follows:

$$U(d \mid o) = \sum_{N_{i,j}} (1 - o_{N_{i,j}}) (d_{i,j} - d_{N_{i,j}})^2, \tag{19}$$

where $d_{N_{i,j}}$ is the disparity field of the first-order neighbourhood system. As it is clear from the above equation, the occluded neighbours are not taken into account, since they represent local discontinuities. The effect of this procedure is to result in a more uniform disparity field, which provides better encoding. In this work, MAE is selected instead of MSE as a measure of the energy terms in (18) and (19), because it is simpler and less sensitive to outliers.

3.3. The occlusion constraint

The prior occlusion field, which is called occlusion constraint, is a binary field that defines local discontinuities. The occluded blocks are not compensated and their disparity vector is set to zero. The energy equation of the occlusion field has the following form:

$$\begin{aligned} U(o) &= \sum_{c \in C} V_c(o_{i,j}, o_{N_{i,j}}) \\ &= \sum_{(i,j) \in C_1} o_{i,j} V_c(o_{i,j}, o_{N_{i,j}}) + \sum_{(i,j) \in C_2} V_c(o_{i,j}, o_{N_{i,j}}), \end{aligned} \tag{20}$$

where $o_{N_{i,j}}$ are the occluded neighbours of the processed $o_{i,j}$, C_1 and C_2 are the single and double clique sites, respectively. First term provides the energy cost if a block becomes occluded and second term encourages occlusion connectivity.

3.4. The final equation for disparity estimation

The general equation (17) of the MRF/GRF model, taking into account (18), (19), and (20), may be expressed as

$$(\hat{d}, \hat{o})$$

$$= \underset{(d,o) \in S}{\operatorname{argmin}} \left\{ (1 - \lambda_d) \sum_{(i,j) \in S} (1 - o_{i,j}) \sum_{(k,l) \in b_{i,j}} \left| c_{(k,l)}^R - \tilde{c}_{(k,l) \oplus d_{i,j}}^L \right| \right.$$
$$+ \lambda_d \sum_{N_{i,j}} (1 - o_{N_{i,j}}) \| d_{i,j} - d_{N_{i,j}} \|$$
$$+ \sum_{(i,j) \in C_1} o_{i,j} V_c(o_{i,j}, o_{N_{i,j}})$$
$$\left. + \lambda_o \sum_{(i,j) \in C_2} V_c(o_{i,j}, o_{N_{i,j}}) \right\}, \tag{21}$$

where λ_d and λ_o are weighting constants that control each of the participating fields. Each term of the above equation depicts the energy cost of likelihood, smoothness constraint, and occlusion functions, respectively.

3.5. The proposed disparity compensation

The disparity field, which is estimated by (1) and (2), consists of the residual image and the vector field. The initial occlusion field is formed by employing a double-threshold procedure as in [16]:

$$\begin{aligned} &\text{nonoccluded block at } (i,j) \in S \\ &\quad \text{if } \left| C_{i,j}^R - C_{(i,j) \oplus d_{i,j}}^L \right| < T_1, \\ &\text{occluded block at } (i,j) \in S \\ &\quad \text{if } \left| C_{i,j}^R - C_{(i,j) \oplus d_{i,j}}^L \right| \geq T_2, \\ &\text{uncertain block at } (i,j) \in S \\ &\quad \text{if } T_1 \leq \left| C_{i,j}^R - C_{(i,j) \oplus d_{i,j}}^L \right| < T_2. \end{aligned} \tag{22}$$

Hence, the occlusion field is separated into three regions:

(i) the nonoccluded region, where the blocks are always predictable;

(ii) the occluded region, where the blocks are always occluded and excluded from the MAP search;

(iii) the uncertain region, where the blocks are subjected into an MAP search, in order to enrol them as occluded or nonoccluded.

Disparity and occlusion fields are iteratively updated according to the nonoptimal deterministic method proposed in [28], in order to reduce complexity.

(i) Given the best initial estimate of the occlusion field, update disparity field by minimizing the first two terms of the final equation (21). This phase refers to blocks that belong to the nonoccluded and uncertain regions, because occluded blocks are not compensated.

(ii) Given the best estimate of the disparity field, update occlusion field by minimizing the last two terms of the final equation. This phase is applied on blocks that belong to the uncertain region, in order to enrol them in one of the other two regions. First term penalizes the conversion of an uncertain block to an occluded or nonoccluded block and second term favours the connectivity of the processed block.

(iii) The whole process is repeated until no further energy minimization takes place. The proposed MRF method converges in three or four iterations.

The last two terms of (21) represent the potential costs for the occlusion phase and are defined as follows:

$$U(o_{i,j}) = o_{i,j} \big[C_0 - \lambda_p \, \text{mean} \, (|b_{i,j}^R - \tilde{b}_{i,j}^L|) \big] + \lambda_o \sum_{(i',j') \in N_{i,j}} h(o_{i,j}, o_{i',j'}), \tag{23}$$

where C_0, λ_p, λ_o are weighting constants. The first term of the above equation is the energy cost if an uncertain block is assigned as occluded and is expressed in terms of the mean residual block. The second term penalizes the connectivity of an uncertain block to its neighbours. The function $h(\cdot)$ is defined as

$$h(o_{i,j}, o_{i',j'})$$
$$= \begin{cases} |o_{i,j} - o_{i',j'}| & \text{if } (i', j') \notin \text{uncertain}, \\ [1 - \text{sign}(||d_{i,j} - d_{i',j'}|| - \lambda_q)] \\ \qquad \times [1 - 2\delta(o_{i,j} - o_{i',j'})] & \text{if } (i', j') \in \text{uncertain}, \end{cases} \tag{24}$$

where $\delta(\cdot)$ is the Kronecker delta function and sign is the signum function.

If the neighbours of an uncertain block are occluded or nonoccluded blocks, the cost increases with the number of neighbours that are of different kind. This term favours the connectivity of an uncertain block to its neighbourhood. If a neighbour of an uncertain block is also uncertain, the cost depends on their disparity vectors difference. If this is greater than a threshold λ_q, there is no energy cost. If the difference is less than the prespecified threshold, the energy cost increases if the two uncertain blocks are of different kind. The threshold λ_q becomes smaller over the iterations, as the disparity vector field becomes more uniform and in this work is defined as $\lambda_q = \max(2e^{-i/8}, 1)$.

3.6. The computational complexity of the proposed algorithm

It is well known that the computational complexity of a disparity estimation algorithm is defined by the search algorithm, the cost function, and the search range. Assume the macroblock size of 8×8 pixels and that the search range parameter is p. Let us also assume that a disparity estimation algorithm employs the MSE cost function and that the image size is $M \times N$ pixels. The computational complexity of BMA is given by

$$O_{\text{BMA}} = \left(\frac{MN}{64} - n_{\text{occ}} \right)(2p+1)^2 O_{\text{MSE}} + O_{\text{OCC}}, \tag{25}$$

where O_{MSE} is the complexity of the cost function requiring 259 operations. The exhaustive search technique requires $(2p+1)^2$ searches per macroblock. If $p = 16$, as proposed in this paper, the number of searches is 1089. The disparity field, unlike motion field, depends on the distance from the camera and thus is less uniform, as different parts of the background show different disparity. The occlusion field is defined by comparing the magnitude of the MAE of a matching block with a preselected threshold and is assumed consisting of n_{occ} occluded macroblocks. The disparity compensation procedure is performed for macroblocks that are not occluded. Thus, the complexity of defining the occlusion field is given by the term O_{OCC} that is about the complexity of the MAE cost function.

The computational complexity given by (25) is considered as the initial step of a typical MRF algorithm, [6]. To this complexity, the required operations for updating the disparity field and the operations for updating the occlusion field have to be added, as mentioned in the previous subsection. The update of the disparity field is performed by the first two terms of (21), whereas the update of the occlusion field is performed by the last two terms of the same equation or (23) and (24). This may be expressed as

$$O_{\text{MRF}} = O_{\text{BMA}} + (O_{\text{DCD}} + O_O)k, \tag{26}$$

where O_{DCD} and O_O represent the computational complexity for updating the disparity and occlusion fields, respectively, and k is the number of required iterations. The update of the disparity field concerns only the nonoccluded macroblocks, whereas the update of the occlusion field concerns all the macroblocks.

In our proposed algorithm, the update of the occlusion field is performed only on the uncertain region as indicated by (22), which is a fraction of the total image size. This reduces the computational complexity of a typical MRF method, which is expressed by (26), and renders the execution time faster. Moreover, MAE has been chosen as the cost function because of its simplicity compared to MSE, its direct hardware implementation, and its robustness to outliers. It has been estimated that the time consumed by our proposed algorithm is about three times that of BMA and about 30% less than that of a typical MRF algorithm. The complexity of our proposed scheme may be reduced if the search range in the vertical direction is confined to ± 2 pixels. In that case,

TABLE 1: Values assigned to weighting constants.

Parameter	Value	Parameter	Value
T_1	$2 \times \text{mean}(|\text{DCD}|)$	λ_o	10
T_2	$\text{mean}(|\text{DCD}|)$	λ_p	1
C_0	50	λ_q	$\max(2e^{-i/8}, 1)$
λ_d	0.5	—	—

the number of searches is reduced to 165. This is reasonable because the natural images used for experimental evaluation have been captured by fixed and aligned cameras. The complexity may be further improved if a fast searching algorithm is employed for disparity estimation, as for motion estimation. For example, the three-step search algorithm presents a complexity $O(\log p)$ compared to $O(p^2)$ that the exhaustive search presents. Also, the complexity of the hierarchical block-matching algorithm is 50 times lower compared to the exhaustive search.

4. EXPERIMENTAL RESULTS

In this section, the experimental evaluation of the proposed coder is reported. Three grey-scale stereo image pairs were employed for the experimental evaluation, from which two images are synthetic: "Room" (256×256) and "SYN.256" (256×256); and two images are real: "Fruit" (256×256) and "Aqua" (360×288) [29–31]. The proposed stereoscopic coder employs four-level DWT with symmetric extension, based on the 9/7 biorthogonal Daubechies filters [32]. The parameter values are obtained by trial and error and are listed in Table 1.

(i) T_1, T_2 are thresholds that define an initial occlusion field. They are defined in terms of the average value of the initial disparity compensated field. The initial DCD or the initial residual image is attained after disparity estimation for all the macroblocks employing BMA.

(ii) λ_d controls the smoothness of the disparity vector field. Large values of this parameter may lead to blurring across object boundaries.

(iii) C_0, λ_p control the energy cost of an uncertain block to be assigned as occluded. In the final energy equation, they represent the single-site cliques.

(iv) λ_o controls the double-site cliques and enforces the connectivity of neighbours.

(v) λ_q is a variable threshold value in each iteration that penalizes the disparity vector difference between uncertain neighbouring blocks.

Except for the thresholds T_1 and T_2, which define the three regions of the occlusion field, minor alterations to the other parameters will not change considerably the experimental results. It is very difficult to estimate automatically their values or to correlate their estimation with the source stereoscopic image pair. For this reason, the parameters listed in Table 1 were kept constant throughout the experiments and for all the tested images.

The experimental evaluation of the proposed method is performed with the following criteria.

(i) The subjective quality measure, which is the optical quality of the reproduced target image. The smoothness of the residual image and the disparity vector field are indicative of the final target image quality. The abnormalities that appear in the residual image due to occlusions make the bit cost larger. Also, the detection of the occlusion field by using thresholds is simple but contributes to a larger bit cost.

(ii) The objective quality measure of the reproduced images, which is expressed by the PSNR value in terms of the total bit rate:

$$\text{PSNR} = 10 \log_{10} \frac{255^2}{(\text{MSE}_L + \text{MSE}_R)/2}, \tag{27}$$

where MSE_L and MSE_R are the mean square errors of Left and Right images, respectively. The total bit rate is the entropy of the DWT subband coefficients of reference and residual images, after their morphological representation and partitioning by the morphological encoder and the disparity vectors, which are DPCM encoded, since their transmission must be lossless.

(i) The entropy of the disparity vector field, which is defined as

$$H_{\text{DV}} = -\sum_{dv_x} P(dv_x) \log_2 P(dv_x) - \sum_{dv_y} P(dv_y) \log_2 P(dv_y), \tag{28}$$

where $P(dv_x)$ and $P(dv_y)$ denote the probability of the horizontal and vertical disparity vector components. This measure indicates the randomness of the disparity field and it is intended to be as low as possible. This is normal in most images which consist of smooth intensity objects, except around object boundaries. The MRF method, in contradiction to the classical BMA, takes care of that vector smoothness.

(ii) The normalized average energy or MSE of the residual image, which is defined as

$$E_{\text{DCD}} = \frac{\sum_{(i,j) \in S} [\text{DCD}(i,j)]^2}{N \times N}, \tag{29}$$

where S is the image lattice of $N \times N$ dimensions. A lower residual energy means that fewer bits are needed for encoding, so it is indicative of the matching algorithm effectiveness.

The experimental evaluation involves the comparison of the proposed disparity compensation process, which is based on the MRF model, with respect to the classical BMA method and the performance of the proposed stereo coder with respect to other state-of-the-art coders. In this coder, the disparity compensation process is implemented with blocks of 8×8 pixels in a searching area of 16 pixels. This size of blocks is found to be the best choice in terms of the produced noise and coding efficiency.

(a)   (b)

FIGURE 3: (a) The initial occlusion field as it has been formed after a two-threshold-level classification. The grey colour indicates the uncertain blocks, whereas the black colour indicates the occluded blocks. (b) The final occlusion field after the occlusion phase of the energy minimization process. The occluded region has been augmented because the employed algorithm favours occlusion connectivity.

TABLE 2: Comparative results between BMA and MRF.

Image	Method	Total bit rate (bpp)	E_{DCD}	H_{DV} (bpp)
Room	BMA	0.20 (bpp)	0.0308	0.1275
	MRF		0.0198	0.0975
SYN.256	BMA	0.21 (bpp)	0.0186	0.1381
	MRF		0.0186	0.1284

Table 2 shows the normalized average energy of the residual image and the entropy of the disparity vector field for BMA and MRF processes at a specific total bit rate. As expected, the MRF residual images present lower energy and the disparity vector field is smoother than that of BMA processing. This lower energy and the smoothness of the vector field insure lower total entropy values. The occluded regions are usually tracked and excluded from the disparity compensation process, since they contribute to distortions increasing excessively the bit rate. The occlusion indicators are transmitted because their residual coding results in a total bit-rate benefit. Also, their main role is to avoid mismatching blocks containing object boundaries and preventing disparity oversmoothing across discontinuities. The MRF model penalizes the existence of an occluded block and encourages the connectivity of neighbouring occluded blocks, which usually appear at objects boundaries where large intensity gradients prevail.

Figure 3(a) and 3(b) show the initial and final occlusion fields for the "Room" stereo pair, respectively. In (a), grey regions represent the uncertain field, black areas represent the occluded field, and white areas represent the nonoccluded field. The isolated occluded blocks are initially assigned as uncertain blocks, because it is desirable to exclude them as they increase entropy cost. It is also apparent in (b) that occlusion connectivity is favoured, as black areas have been enlarged.

Figures 4(a)–4(d) show the residual image and the disparity vector of a BMA- and an MRF-based disparity compensation process for the "Room" stereo pair, at a bit rate of 0.20 bpp. Figures 5(a)–5(d) show the residual image and the disparity compensation process for the "SYN.256" stereo pair, at a bit rate of 0.21 bpp. In both stereo pairs, the performance of the MRF disparity compensation process is better than the corresponding BMA. Apparently, the MRF model residual images present lower energy and their corresponding disparity vector fields are smoother than their BMA counterparts validating the results of Table 2.

Figures 6(a) and 6(b) show the reconstructed target image of stereo pair "Room," for BMA and MRF, respectively. The objective quality of BMA and MRF processes is 26.02 dB and 28.24 dB, respectively, for a bit rate of 0.2 bpp. Figures 7(a) and 7(b) show the reconstructed target image of stereo pair "SYN.256," for BMA and MRF, respectively. The performance of BMA and MRF processes is 29.08 dB and 29.92 dB, respectively, for a bit rate of 0.21 bpp.

Table 3 demonstrates the performance of the proposed coder for all the tested stereo pairs at discrete bit rates.

Figure 8 illustrates the quality performance of various stereoscopic coders for the "Room" stereo image pair, over the examined bit-rate range from 0.25 to 1 bpp. The proposed MRF stereo coder outperforms Frajka and Zeger coder by about 1 dB [9], Boulgouris and Strintzis coder by 2 dB [8], disparity-compensated JPEG2000 coder by about 2.5 dB [33], and optimal blockwise-dependent quantization by about 3 dB [34]. The optimal blockwise-dependent quantization stereo coder by Woo et al. employs a JPEG-like coder for both the reference and residual images, whereas Boulgouris and Strintzis use DWT and EZW followed by arithmetic encoding. Frajka and Zeger employ JPEG for the reference image and a mixed transform coder followed by arithmetic encoding for the residual image. The disparity compensated JPEG2000 stereo coder is based on a JPEG2000 coder for the reference and residual images with a disparity compensation
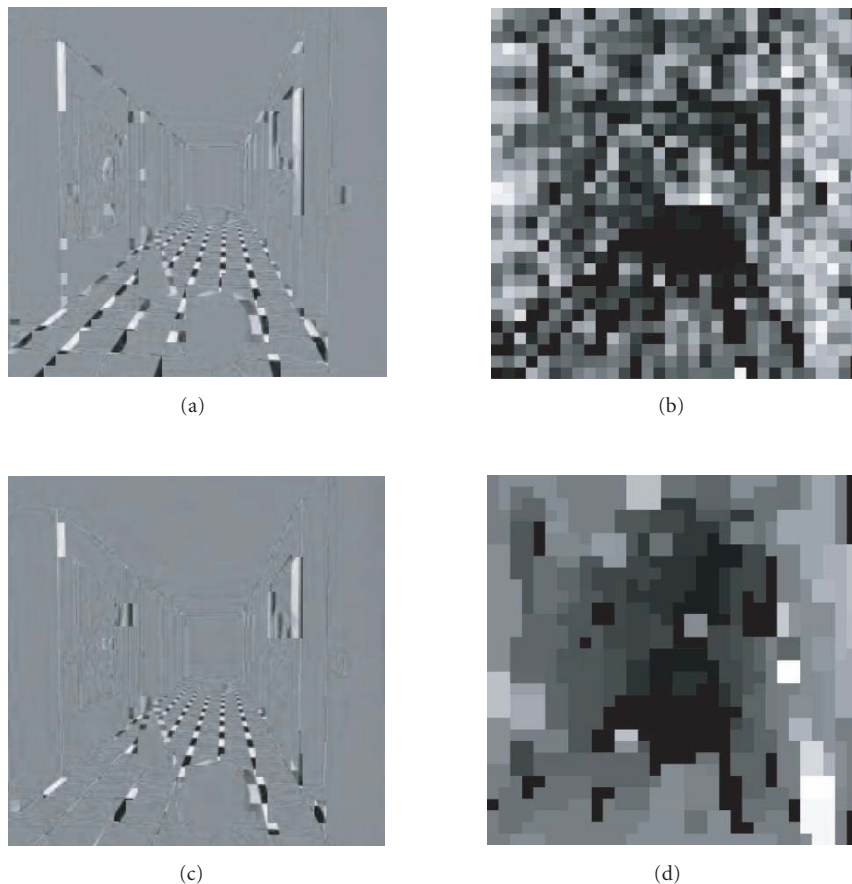
(a)

(b)

(c)

(d)

FIGURE 4: Residual image and disparity vector field: (a), (b) BMA method; (c), (d) MRF method.

procedure that is performed with fixed-size block BMA. Finally, our proposed coder presents inferior quality compared to Woo et al. hierarchical MRF stereo coder at medium bit rates [35].

At the lower bound, the two algorithms converge, whereas at bit rates greater than 0.5 bpp, our proposed scheme outperforms Woo et al.'s coder. The hierarchical MRF stereo coder incorporates the typical MRF model and a variable-size block-matching scheme for disparity estimation. Consequently, we believe that a variable-size block disparity estimation scheme, adapted to our MRF model, would improve the performance of our coder.

Figure 9 shows the experimental evaluation of various stereoscopic coders for the "Fruit" stereo image pair. The disparity compensated EZW coder is based on EZW encoding for both the reference and residual images employing fixed-size block BMA for disparity estimation. The proposed MRF stereo coder presents beneficial PSNR values in comparison with the other coders. This proves that our coder behaves equally well not only with synthetic stereo images but with camera-acquired images, which present a more difficult disparity field as this field depends on cameras distances and their alignment. It should be observed that the quality difference mitigates at lower bit rates, which may be ascribed to

the fixed-size block matching. BMA disparity compensation with fixed-size blocks does not exploit the constant disparity areas that exist in a scene and assigns more bits than actually required. Figure 10 illustrates the performance of the proposed coder in comparison with other state-of-the-art coders for the "Aqua" stereo image pair. Again, our proposed coder outperforms in the middle and high bit rates of at least 0.8 dB the other stereo coders and its performance converges to the others at lower bit rates. It is worth to note that hierarchical MRF stereo coder presents inferior quality to the specific natural image, whereas our proposed scheme has a stable performance both in synthetic and natural images.

Apart from the MRF model, which treats disparity compensation very effectively, the wavelet-based morphological encoder contributes to the good performance of our proposed scheme because it is more efficient than other coders. It presents, for "still" images, about 1 dB better performance over EZW and also outperforms DCT because of its wavelet nature. Apart from its simple implementation, fast execution, and efficiency, MRWD encoder may provide embedded bit streams and spatial scalability that are prerequisites of a modern coding scheme.

The proposed algorithm may be applied to stereoscopic video coding with advantageous results as the smoother
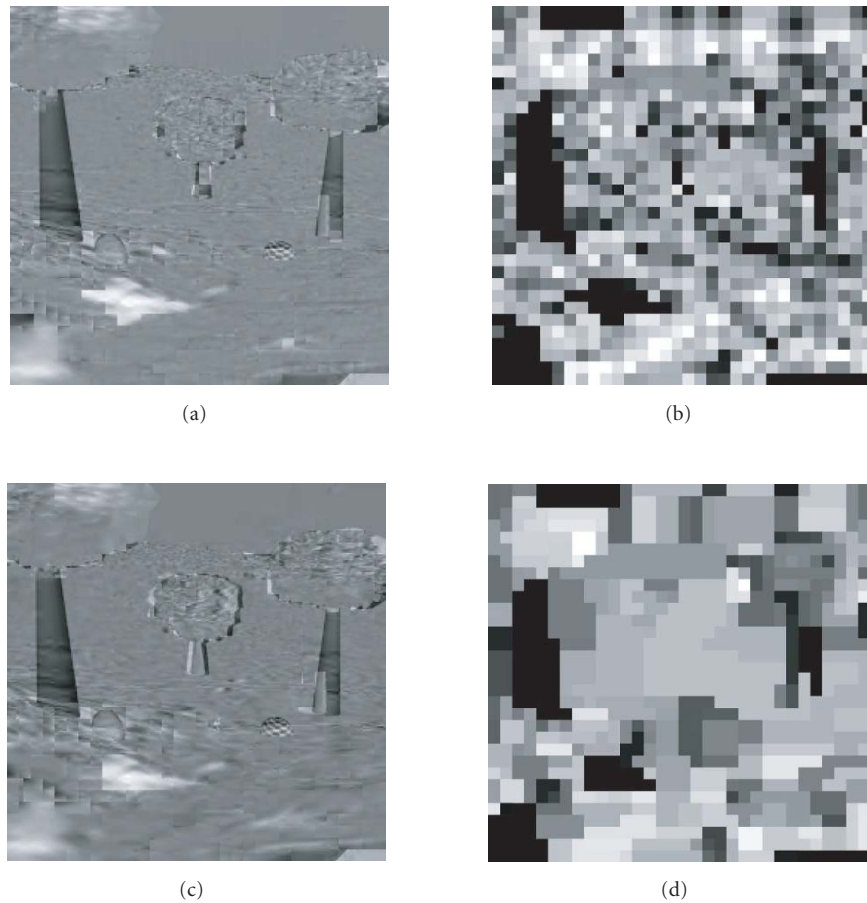
(a)

(b)

(c)

(d)

FIGURE 5: Residual image and disparity vector field: (a), (b) BMA method; (c), (d) MRF method.
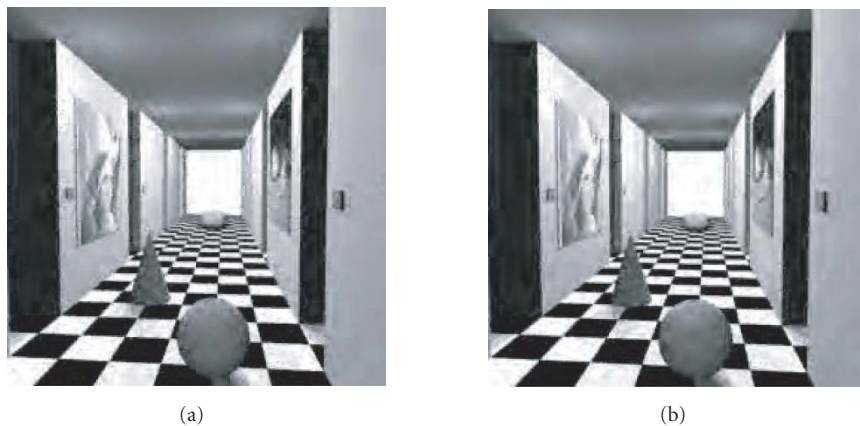

(a)

(b)

FIGURE 6: Reconstructed target image at a bit rate of 0.2 bpp: (a) BMA; (b) MRF.

disparity field will imply better temporal prediction. Of course, this will imply a more complicated framework because motion and disparity fields along with their unpredictable fields must be integrated. The motion-disparity estimation procedure for compensating the auxiliary channel with techniques like the joint motion-disparity estimation, vector regularization, as well as the GOP structure of the two channels should be considered for an effective coding scheme with low complexity. Also, the fixed-size block framework employed in this paper assumes that all the pixels of a block have the same disparity, which is not the case. This assumption does not take advantage of the constant disparity areas.
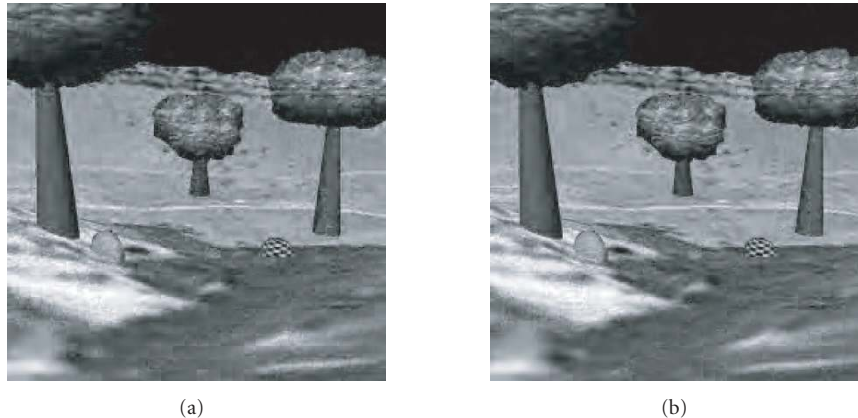
(a)   (b)

FIGURE 7: Reconstructed target image at a bit rate of 0.2 bpp: (a) BMA; (b) MRF.

TABLE 3: PSNR versus bit rate of the proposed coder.

Image	PSNR (dB)			
	0.25 (bpp)	0.5 (bpp)	0.75 (bpp)	1 (bpp)
Room	30.67	36.98	41.85	45.06
SYN.256	30.76	35.60	38.63	41.20
Fruit	—	37.00	38.65	39.86
Aqua	26.35	28.87	30.81	32.57

Therefore, as a plan for future work, the application of the proposed scheme in a variable-size block framework should be considered.

5. CONCLUSIONS

In this work, an algorithm employing the MRF model is proposed for the disparity estimation of a stereo image pair. The MRF model is a popular method in video community for a consistent evaluation of motion fields. It provides the means to accomplish smooth disparity field without increasing the residual energy, and thus to devote fewer bits to encode them. The proposed coder consists of a disparity compensation unit and an encoding unit. The disparity compensation unit constructs initially the disparity and occlusion fields using BMA. The occlusion field is separated into three regions by employing a two-level threshold and the MAP search is performed on the uncertain region, which consists of blocks that have to enrol in the occlusion or nonocclusion regions. This approach permits faster execution times, as the MAP search is conducted in a fraction of the whole image. In addition, the choice of MAE provides more reliable disparity estimation than MSE because it is more robust and simpler. The encoding unit decomposes the reference and residual images with DWT and employs the morphological algorithm MRWD for compression. This algorithm partitions the coefficients of the wavelet transform and lowers their entropy. The obtained results show that the proposed method improves the quality
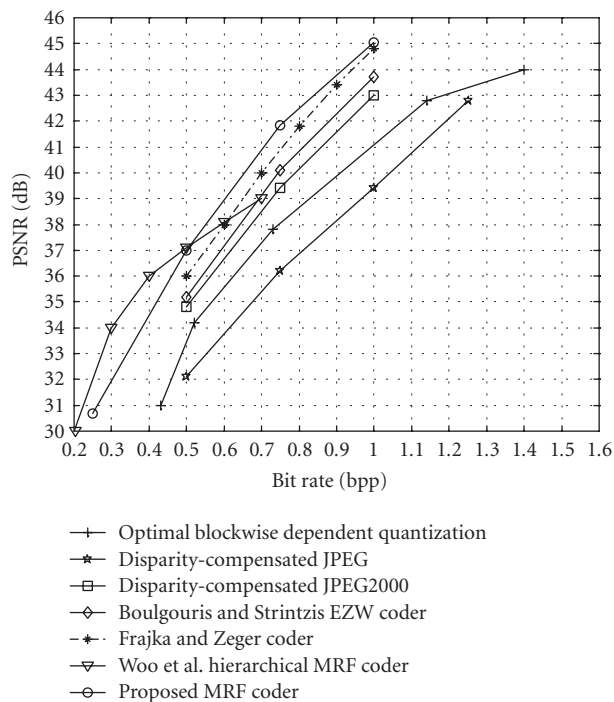

FIGURE 8: Quality performance evaluation of various stereoscopic coders for "Room" stereo pair.

of the reconstructed target images compared with the results which a plain BMA method may provide. Also, the proposed stereo coder outperforms some known state-of-the-art coders.

To further investigate the contribution of the MRF model to the efficient handling of disparity estimation, its application on the subband domain may be tested. This may be done using BMA or coefficient matching, which may be embedded into the morphological encoder by using the same structuring element as that of the first-order neighbourhood system.
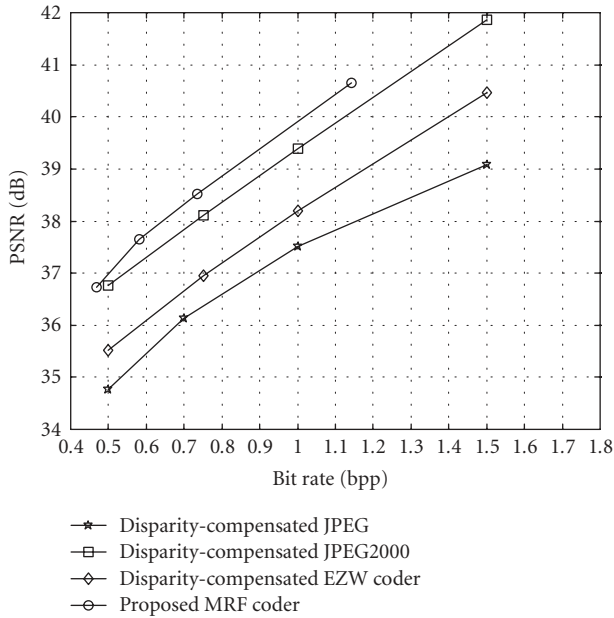
FIGURE 9: Quality performance evaluation of various stereoscopic coders for "Fruit" stereo pair.
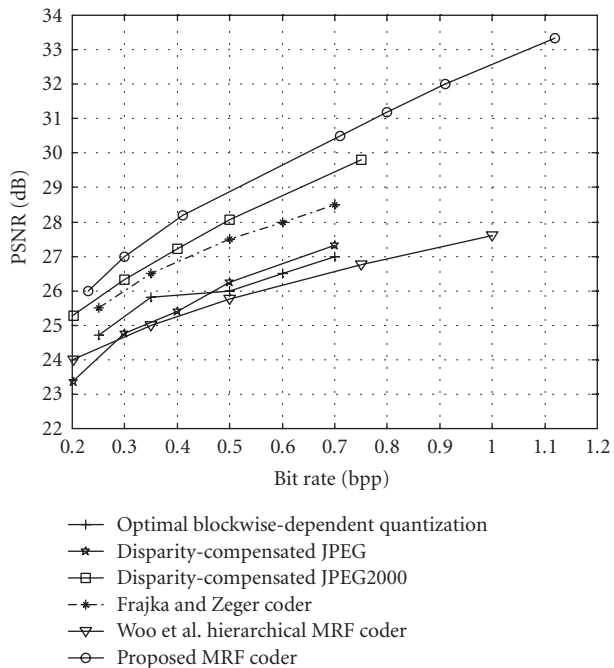

FIGURE 10: Quality performance evaluation of various stereoscopic coders for "Aqua" stereo pair.

ACKNOWLEDGMENT

REFERENCES

[1] H. Yamaguchi, Y. Tatehira, K. Akiyama, and Y. Kobayashi, "Stereoscopic images disparity for predictive coding," in *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP '89)*, vol. 3, pp. 1976–1979, Glasgow, Scotland, UK, May 1989.

[2] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*, pp. 674–679, Vancouver, BC, Canada, August 1981.

[3] W. Woo and A. Ortega, "Overlapped block disparity compensation with adaptive windows for stereo image coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 2, pp. 194–200, 2000.

[4] J. Jiang and E. A. Edirisinghe, "A hybrid scheme for low bit-rate coding of stereo images," *IEEE Transactions on Image Processing*, vol. 11, no. 2, pp. 123–134, 2002.

[5] S. Sethuraman, A. Jordan, and M. Siegel, "Multiresolution based hierarchical disparity estimation for stereo image pair compression," in *Proceedings of Symposium on Applications of Subbands and Wavelets*, Newark, NJ, USA, March 1994.

[6] W. Woo and A. Ortega, "Stereo image compression with disparity compensation using the MRF model," in *Visual Communications and Image Processing (VCIP '96)*, vol. 2727 of *Proceedings of SPIE*, pp. 28–41, Orlando, Fla, USA, March 1996.

[7] H. Yamaguchi, Y. Tatehira, K. Akiyama, and Y. Kobayashi, "Data compression and depth shape reproduction of stereoscopic images," *Systems and Computers in Japan*, vol. 22, no. 12, pp. 53–64, 1991.

[8] N. V. Boulgouris and M. G. Strintzis, "A family of wavelet-based stereo image coders," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 10, pp. 898–903, 2002.

[9] T. Frajka and K. Zeger, "Residual image coding for stereo image compression," *Optical Engineering*, vol. 42, no. 1, pp. 182–189, 2003.

[10] J. N. Ellinas and M. S. Sangriotis, "Stereo image compression using wavelet coefficients morphology," *Image and Vision Computing*, vol. 22, no. 4, pp. 281–290, 2004.

[11] M. G. Perkins, "Data compression of stereopairs," *IEEE Transactions on Communications*, vol. 40, no. 4, pp. 684–696, 1992.

[12] H. Aydinoglu and M. H. Hayes, "Stereo image coding: a projection approach," *IEEE Transactions on Image Processing*, vol. 7, no. 4, pp. 506–516, 1998.

[13] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[14] J. Konrad and E. Dubois, "Bayesian estimation of motion vector fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 9, pp. 910–927, 1992.

[15] J. Zhang and G. G. Hanauer, "The application of mean field theory to image motion estimation," *IEEE Transactions on Image Processing*, vol. 4, no. 1, pp. 19–33, 1995.

[16] J. Wei and Z.-N. Li, "An efficient two-pass MAP-MRF algorithm for motion estimation based on mean field theory," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 6, pp. 960–972, 1999.

[17] J. N. Ellinas and M. S. Sangriotis, "Stereo image coder based on MRF analysis for disparity estimation and morphological

encoding," in *Proceedings of 2nd IEEE International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT '04)*, pp. 852–859, Thessaloniki, Greece, September 2004.

[18] S. D. Servetto, K. Ramchandran, and M. T. Orchard, "Image coding based on a morphological representation of wavelet data," *IEEE Transactions on Image Processing*, vol. 8, no. 9, pp. 1161–1174, 1999.

[19] M. Lukacs, "Predictive coding of multi-view point image sets," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '86)*, vol. 11, pp. 521–524, Tokyo, Japan, April 1986.

[20] J.-L. Starck, F. D. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis: The Multiscale Approach*, Cambridge University Press, Cambridge, UK, 1998.

[21] S. Z. Li, *Markov Random Field Modeling in Computer Vision*, Springer, Tokyo, Japan, 1995.

[22] S. Z. Li, "Markov random field models in computer vision," in *Proceedings of 3rd European Conference on Computer Vision (ECCV '94)*, vol. 2, pp. 361–370, Stockholm, Sweden, May 1994.

[23] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society B*, vol. 36, pp. 192–225, 1974.

[24] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 205–220, 1992.

[25] J. M. Shapiro, "Embedded image coding using zero trees of wavelet coefficients," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, 1993.

[26] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, 1996.

[27] B.-B. Chai, J. Vass, and X. Zhuang, "Significance-linked connected component analysis for wavelet image coding," *IEEE Transactions on Image Processing*, vol. 8, no. 6, pp. 774–784, 1999.

[28] E. Dubois and J. Konrad, "Estimation of 2-D motion fields from image sequences with application to motion-compensated processing," in *Motion Analysis and Image Sequence Processing*, M. I. Sezan and R. L. Lagendijk, Eds., pp. 53–87, Kluwer Academic, Boston, Mass, USA, 1993.

[29] Stereo images from Bonn University, available from: http://www-dbv.cs.uni-bonn.de.

[30] Stereo images from Carnegie Mellon University, Pittsburgh, Pa, USA, available from: http://www-2.cs.cmu.edu/afs/cs.cmu.edu/project/sensor-9/ftp/.

[31] http://www.code.ucsd.edu/~frajka/images/stereo/stereo_images.html.

[32] B. E. Usevitch, "A tutorial on modern lossy wavelet image compression: foundations of JPEG 2000," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 22–35, 2001.

[33] M. D. Adams, H. Man, F. Kossentini, and T. Ebrahimi, "JPEG 2000: The next generation still image compression standard," ISO/IEC JTC 1/SC 29/WG 1 N 1734, 2000.

[34] W. Woo and A. Ortega, "Optimal blockwise dependent quantization for stereo image coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 6, pp. 861–867, 1999.

[35] W. Woo, A. Ortega, and Y. Iwadate, "Stereo image coding using hierarchical MRF model and selective overlapped block disparity compensation," in *Proceedings of IEEE International Conference on Image Processing (ICIP '99)*, vol. 2, pp. 467–471, Kobe, Japan, October 1999.

J. N. Ellinas received his B.S. degree in electrical and electronic engineering from the University of Sheffield, England, in 1977, and his M.S. degree in telecommunications from Universities of Sheffield and Leeds, in 1978. He received his Ph.D. degree from the Department of Informatics and Telecommunications at the National & Kapodistrian University of Athens in June 2005. Since 1983, he has been with Technological Educational Institute of Piraeus, Department of Computer Engineering, Greece, where he is currently an Associate Professor. His research interests include image processing, image and video compression.

M. S. Sangriotis received his B.S. and Ph.D. degrees from Athens University in Greece. In 1981, he was with the Department of Physics in Athens University. Since 1990, he has been with the Department of Informatics and Telecommunications in National & Kapodistrian University of Athens, Greece, where he is currently an Associate Professor. His research interests include image analysis and image coding.