Stereo Image Compression with Disparity Compensation Using the MRF Model

Woontack Woo and Antonio Ortega

Signal and Image Processing Institute, Department of Electrical Engineering-Systems
University of Southern California, Los Angeles, California 90089-2564

ABSTRACT

In coming years there will be an increasing demand for realistic 3-D display of scenes using such popular approaches as stereo or multi-view images. As the amount of information displayed increases so does the need for digital compression to ensure efficient storage and transmission of the sequences. In this paper, we introduce a new approach to stereo image compression based on the MRF model and MAP estimation. The basic strategy will be to encode the right image as a reference, then estimate the disparity between blocks in the right and left images and transmit the disparity and the error between the disparity compensated left image and the original. This approach has been used in the literature and is akin to the block matching technique used for motion compensation in video coders. The main drawback in this approach is that as the block size becomes smaller the overhead required to transmit the disparity map becomes too large. Also, simple block matching algorithms frequently fail to provide good matching results because the correspondences are locally ambiguous due to noise, occlusion, and repetition or lack of texture. The novelty in our work is that to compute the disparity map we introduce an MRF model with its corresponding energy equation. This allows us to incorporate smoothness constraints, to take into account occlusion, and to minimize the effect of noise in the disparity map estimation. Obtaining a smooth disparity is beneficial as it reduces the overhead required to transmit the disparity map. It is also useful for video coding since the robustness against noise ensures that disparity maps in successive frames will be very similar. We will describe this new formulation in detail and will provide compression results.

Keywords : stereo image coding, image compression, disparity estimation, Markov random field(MRF)

1 INTRODUCTION

A popular approach for 3-D display of scenes is to use stereo or multi-view images. In the stereo case, as for monocular images, compression will be needed to efficiently transmit or store the digitized images. In general, the stereo pair of images can be encoded more efficiently than two independent images because of the binocular redundancy between stereo images.

So far, stereo images have been studied in the field of computer vision and used in applications such as aerial mapping, robot vision, autonomous navigation, military surveillance, and remote machine control.[1] Stereo has also been proposed for entertainment in cinema and television. Only a few works, however, have dealt with digital stereo image compression since Lukacs first introduced disparity compensated prediction.[2] Dinstein *et al.*

compared 3-D DCT and the disparity compensation approach for stereo image compression.[3] The disparities were also compensated in the DCT domain[4] and subband domain.[5] Aydinoglu *et al.* used a lossy compression method to find smooth disparity maps and applied the subspace projection method as a postprocessing technique after disparity compensation.[6] A hierarchical block matching(BM) method has also been proposed to increase the efficiency of the disparity estimation.[7]

In stereo image processing and analysis, a major goal is to find the exact correspondence between the two images to reconstruct the depth information. This task is complicated by local ambiguities in the correspondence due to noise, occlusion, and repetition or lack of texture. Since blockwise correspondence is also used in stereo image compression to take advantage of the redundancy in the stereo pair, the same problems arise here as well. Typically, for each block in the left image we find the best match in the right image and then encode the disparity and the difference between these two blocks. Simple disparity estimation algorithms based on block matching provide results that are optimal in the sense of minimizing the mean squared error (MSE) between the block and its disparity compensated prediction. However, this advantage may not be so clear once the overall system is considered. First, the resulting disparity map will not be smooth in general, and thus the overhead required to transmit it may be high, especially when small block sizes are used. Second, at low bit rates, where few bits can be used to encode the residue images, non-smooth disparity will result in blocking artifacts in the decoded left image. Our goal is to overcome the various noise problems and achieve higher compression ratios and better image quality than what is achievable with simple block matching.

To do so we introduce several useful constraints that make the estimated disparity map smoother without increasing excessively the energy of the residual image. We thus introduce a new approach to stereo image compression based on the Markov random field(MRF) model and maximum a posterior probability(MAP) estimation. A growing number of applications of the MRF model in the field of image processing and computer vision have been proposed since Geman's introduction.[8] Recently, Konrad *et al.*[9] proposed the MRF based approach and used stochastic relaxation to estimate motion vectors from time-varying images. Zhang *et al*[10] showed that the mean field theory(MFT) can be applied to MRF based motion estimation instead of stochastic relaxation. The main advantage of the MRF model is that one can systematically develop algorithms. It provides a general and natural model for the interaction between spatially related random variables and it tends to be local so that the MRF based model can be implemented in parallel.[11] It is also easy to integrate different information such as stereo and motion.[12] We model the disparity map as an MRF with appropriate a priori assumptions on its smoothness, as well as the noise and occlusion effects. According to the theory of equivalence of MRF and Gibbs random fields,[13] we can then find the disparity map by minimizing an energy equation which takes into account smoothness, noise and occlusion.

This paper is organized as follows. In section 2, we review the basic structure of stereo imaging,[14] describe the theoretical background of the MRF model, and derive an energy equation for disparity estimation. In section 3, we explain a disparity estimation algorithm based on the energy equation derived in section 2 and the encoding algorithm for the stereo image encoding based on disparity compensation and JPEG encoding.[15] The experimental results for synthetic and real images are presented in section 4. Finally, conclusions and future research directions are given in section 5.

2 STEREO MODELING, MRF, AND MAP ESTIMATION

2.1 Stereo Modeling

Figure 1 shows the basic structure for stereo image formation.[14] Two pictures of the same scene are acquired from slightly different perspectives and are called a stereo pair or a stereo image. The center of the lens is called the camera focal center and the ray extending from the focal center is referred to as the camera focal ray. The line connecting the focal centers is called baseline, b. The plane passing through an object point and the focal centers is the epipolar plane. The intersection of two image planes with an epipolar plane is the epipolar line. The disparity, d, is defined by the difference vector, (x_l-x_r,y_l-y_r), between two points corresponding to the same object in the stereo images.

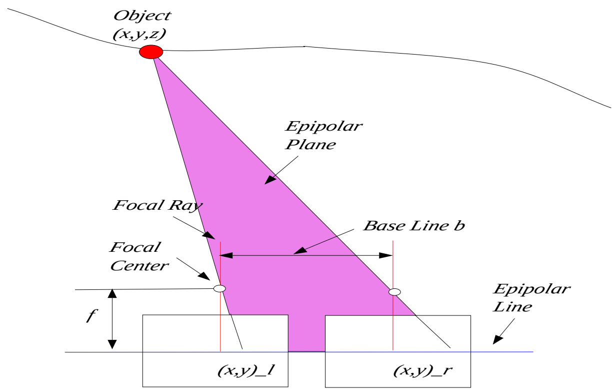


Figure 1: Simple camera geometry for stereo photography

In conventional stereo matching, we assume that the camera is comprised of a thin lens and that the parallel axis constraints are satisfied; that is, the focal rays of the two cameras are parallel and perpendicular to the stereo baseline. The two image planes are coplanar and the scan line, x, is parallel to the baseline or epipolar line. We also assume that the pixels corresponding to each point in the scene have the same brightness. If these constraints are satisfied, the search area for correspondence is restricted to a 1-D line and the matching process is accelerated significantly. Finally, 3-D information can be computed by binocular disparity and triangulation as follows.

$$x = \frac{b(x_l + x_r)}{2(x_l - x_r)},\ y = \frac{b(y_l + y_r)}{2(x_l - x_r)},\ z = \frac{bf}{2(x_l - x_r)} \tag{1}$$

where b represents the baseline and f denotes the camera focal length. As can be seen in (1), the disparity represents the relative depth, *i.e.*, the distance is inversely proportional to disparity. By increasing the baseline the accuracy of the distance measure is increased though the common area in the two images is decreased.

Disparity estimation (DE) is similar to motion estimation (ME) in the sense that they both are used to exploit the similarity between two images in order to reduce the bit rate. Additionally, in both cases block based estimation is preferred for practical reasons, even though the true disparity/motion fields are obviously not blockwise constant. There are two main differences between DE and ME. First, if the parallel axis constraint is met, as is usually assumed, we will only need to find the disparity along the x axis, thus simplifying the search. Second, in typical video scenes there are just a few moving objects and the background is either static (zero motion) or shows uniform motion due to panning. However disparity depends on the distance from the camera and thus different parts of the background will show different disparity. Thus, from a compression perspective we can expect the disparity field to be less uniform than typical motion fields and thus to require a larger number of bits to be encoded (for same block size).

2.2 MRF Model and MAP Estimation

To estimate the disparity, we consider a coupled MRF model consisting of a disparity process, D, and an occlusion process, Φ. Geman *et al.* considered images as realizations of a stochastic process that consists of an observable noise process and a hidden edge process.[8] We extend this stochastic model to modeling of the stereo disparity space. Figure 2 shows some neighborhood systems commonly used in image processing. These can be used similarly for disparity and occlusion.

5	4	3	4	5
4	2	1	2	4
3	1	1	1	3
4	2	1	2	4
5	4	3	4	5

(a)

(b)

Figure 2: Neighborhood systems and cliques: (a) Geometry of neighborhoods; the number denotes i^{th} order neighborhood system. (b) First order neighborhood η^1 and cliques used for occlusion; we can quantify the effect of each clique according to the characteristics of the random fields

We model the spatial interactions among neighboring disparities based on the discrete MRF model and Gibbs distribution. Consider a random field of disparity $D = \{d_{ij}, (i,j) \in \Omega\}$ defined on a discrete, finite, rectangular lattice $\Omega = \{(i,j) | 0 \leq i \leq N_x, 0 \leq j \leq Ny\}$ where N_x and N_y are the vertical and horizontal size of the disparity map, respectively. Note that here we consider a blockwise disparity field, that is, d_{ij} represents the disparity for a intensity block, F_{ij}, of $B \times B$ pixels. Obviously the more traditional problem, where correspondence is computed for each individual pixel, corresponds to $B = 1$ in our framework. Assume D is a MRF with respect to a neighborhood system $\eta = \{\eta_{ij}, (i,j) \in \Omega\}$, where η_{ij} is the neighborhood of d_{ij} such that $(i,j) \notin \eta_{ij}$ and $(k,l) \in \eta_{ij}$, *i.e.*,

$$P\{D = d_{ij} | d_{kl}, (k,l) \in \Omega\} = P\{d_{ij} | d_{kl}, (i,j) \neq (k,l), (k,l) \in \eta_{ij}\} \tag{2}$$

For the disparity problem, the goal is to estimate the most likely solution for the disparity, D, and occlusion, Φ, from the observations, *i.e.*, left, F^l, and right, F^r, images. The posterior probability for the disparity estimation is defined by the Bayesian rule as follows

$$\begin{aligned} P(D, \Phi | F^r, F^l) &= \frac{P(D, \Phi, F^r, F^l)}{P(F^r, F^l)} \\ &= \frac{P(F^l | F^r, D, \Phi) P(D | \Phi, F^r) P(\Phi, F^r)}{P(F^r, F^l)} \\ &= \frac{P(F^l | F^r, D, \Phi) P(D | \Phi) P(\Phi)}{P(F^l | F^r)} \end{aligned} \tag{3}$$

where we assume that the disparity, D, depends on the occlusion, Φ, and the occlusion is independent of the right image, F^r (because the right image itself does not affect directly the decision of disparity and occlusion in the MAP estimation.) The denominator of the above equation can also be ignored in the MAP estimation because it is not a function of the disparity, D, or the occlusion, Φ. Therefore, the posterior probability is proportional to the numerator in (3) as follows

$$P(D, \Phi | F^r, F^l) \propto P(F^l | F^r, D, \Phi) P(D | \Phi) P(\Phi). \tag{4}$$

This MAP estimation process is difficult because of the ill-posedness of the problem, *i.e.*, the solution to the problem may not exist, may not be unique, or may be discontinuous with respect to the data. According to the Clifford-Hammersley theorem,[13] if a measure can be modeled by a MRF, then the probability mass of the measure can be represented as a Gibbs distribution. It can be represented in the Gibbs distribution form as follows

$$P(D, \Phi | F^r, F^l) = \frac{1}{Z} exp\{-U(D, \Phi | F^r, F^l)\} \tag{5}$$

where Z is a normalization constant and $U(D, \Phi | F^r, F^l)$ is an energy function. The energy function $U(D, \Phi | F^r, F^l)$ is defined as follows

$$U(D, \Phi | F^r, F^l) = \sum_{(i,j) \in \Omega} U(d_{ij}, \phi_{ij} | F^r_{ij}, F^l_{ij}) \tag{6}$$

where Ω represents a discrete and finite rectangular lattice with the same size as the disparity field. For the block based disparity estimation d_{ij} and ϕ_{ij} denote the disparity and occlusion of the block of pixels, F_{ij}.

The main advantage of representing each probability in the Gibbs distribution form is that it can be formulated with an energy function and the multiplication of the probabilities can be replaced by the sum of the energy equations. The MAP estimation problem can be replaced by the problem of the finding a solution minimizing the energy equation. Therefore, the solutions, D and Φ, minimizing the energy function $U(D, \Phi | F^r, F^l)$, are the solutions maximizing the posterior probability $P(D, \Phi | F^r, F^l)$. Finally, we can formulate the disparity estimation problem as follows

$$\begin{aligned}(\hat{D}, \hat{\Phi}) &= \arg\max_{(D,\Phi)} P(D, \Phi | F^r, F^l) = \arg\min_{(D,\Phi)} U(D, \Phi | F^r, F^l) \\ &= \arg\min_{(D,\Phi)} \{U(F^l | F^r, D, \Phi) + U(D|\Phi) + U(\Phi)\} \end{aligned} \tag{7}$$

where each term represents useful constraints for the disparity estimation, *i.e.*, similarity, smoothness, and discontinuity constraints, respectively. We now describe appropriate definitions for these terms in our application.

2.2.1 Observation Model- Similarity Constraint

The first term of the right side in (7) represents the constraints of the similarity between stereo image for a given disparity and occlusion. In general, the block based disparity estimation assumes only that the image intensities in the stereo pair, F^r and F^l, are similar along the disparity and tries to find the matching block with an optimal cost value between F^r and F^l. We can represent the block matching method for the disparity estimation as follows

$$\hat{d}_{ij} = \arg\min_{d_{ij}} \; C(F^l_{ij}, F^r_{ij \oplus d_{ij}}) \tag{8}$$

where F_{ij} denotes a block of pixels and d_{ij} represents the disparity vector of the block F_{ij}. we can use a minimum error or a maximum correlation as a cost function, C. Since the intensity levels may not be same even though the images are taken at the same time and at the same place, we can express the difference as follows

$$F^l_{ij} - F^r_{ij \oplus d_{ij}} = W_{ij} \tag{9}$$

where W_{ij} denotes a matrix of the difference, ϵ, between corresponding blocks. The measurement error ϵ is assumed white Gaussian noise with zero mean and variance σ^2. Unfortunately, even assuming ideal conditions, *e.g.* no camera noise, the error will not be zero at object boundaries due to the occlusion effect, that is, some region of the left image may not appear in the right image. When, as is the case here, a block based approach is used, this results in erroneous matching for those blocks which contain object boundaries. To include the occlusion effect into (9) we adopt an occlusion indicator and express the matching cost as follows

$$(1 - \phi_{ij}) \cdot C(F^l_{ij} - F^r_{ij \oplus d_{ij}}) = \begin{cases} c_{ij}, & \phi_{ij} = 0 \\ 0, & \phi_{ij} = 1 \end{cases} \tag{10}$$

where ϕ_{ij} is a binary process which acts as the occlusion indicator for block (i, j) and c_{ij} denotes the corresponding cost value for the nonoccluded block.

For a given disparity and occlusion the constraint of similarity between blocks in the stereo image can be represented in the Gibbs distribution form. For block matching, the energy function of the similarity constraint,

$U(F^l|F^r, D, \Phi)$, can thus be represented as follows

$$U(F^l_{ij}|F^r_{ij}, d_{ij}, \phi_{ij}) = (1-\phi_{ij}) \sum_{(m,n)\in F_{i,j}} (f^l_{mn} - f^r_{mn\oplus d_{ij}})^2 \tag{11}$$

where F_{ij} represents a block of pixels and f_{mn} represents pixels within the block. In (11), we select a MSE as the cost C.

2.2.2 Disparity Process- Smoothness Constraint

The second term of (7) represents an *a priori* assumption on the smoothness of the disparity field, D, given the occlusion, Φ, which will be used to trade-off smoothness and estimation error. The BM methods are sensitive to the intensity variations and various noise effects and then result in blocking artifacts in the low bit rate encoding because the adjacent blocks may have slightly different disparity vectors. It thus may differ from the actual disparity field. We assume that the real disparity field is smooth except for the object boundaries that are related to the depth discontinuities. As before, the disparity minimizing the energy function $U(D|\Phi)$ is the solution maximizing the posterior probability $P(D|\Phi)$. The corresponding energy function $U(D|\Phi)$ can be represented as follows

$$U(d_{ij}|\phi_{ij}) = \sum_{\eta_{ij}} (d_{ij} - d_{\eta_{ij}})^2 (1-\phi_{\eta_{ij}}) \tag{12}$$

where η_{ij} represents a neighborhood of the disparity, d_{ij}. The larger η, the greater the influence of neighboring disparities, $d_{\eta_{ij}}$. The occlusion indicator ϕ_{ij} controls the discontinuity between the disparity d_{ij} and its neighborhood. If the block is detected as an occlusion block, the smoothness constraint should not be applied across the occluded blocks. Note that generating a smooth disparity field not only mitigates the effects of noise, it can also increase the encoding efficiency for the disparity (similar disparities in adjacent blocks results in lower entropy).

2.2.3 Occlusion Process- Discontinuity Constraint

It is natural to discard the estimates for the occluded blocks because the occluded blocks tend to increase the entropy of the error image. Therefore, by adopting an occlusion indicator we can reduce the bit rate for the disparity compensated difference image. We model the occlusion process using a binary MRF model and assume that it tends to be connected along object boundaries. We thus impose an *a priori* assumption on the occlusion field so as to force it to be connected and define the corresponding energy function, $U(\Phi)$, as follows

$$U(\phi_{ij}) = \sum_{c\in C} V_c(\phi_{ij}, \phi_{\eta_{ij}}) \tag{13}$$

where C represents a pre-specified set of cliques for the occlusion process and V_c is a potential function for the cliques. Spatial connections of occluded blocks are determined according to the energy equation, *i.e.*, the isolated occlusion is inhibited and the connected one is encouraged. The initial occlusion is decided by comparing the magnitude of the mean absolute error(MAE) of the block matching with a pre-selected threshold, i.e.,

$$\phi_{ij} = \begin{cases} 1, & |F^l_{ij} - F^r_{ij\oplus d_{ij}}| \geq T_\phi \\ 0, & o.w. \end{cases} \tag{14}$$

where T_ϕ is threshold value for the initial occlusion detection and d_{ij} represents the disparity obtained by the BM algorithm. In stereo encoding, we only transmit or store the initial occlusion indicators because the connected occlusions tend to increase total bit rates for the disparity compensated error image. The main role of the occlusion indicators is to avoid mismatching blocks containing object boundaries and preventing disparity from being oversmoothed across the occluded blocks.

2.2.4 Total Energy

Given the above model, the overall energy function $U(D, \Phi|F^r, F^l)$ for the DE can now be written as

$$\begin{aligned} U(D, \Phi|F^r, F^l) &= \sum_{(i,j)\in\Omega} U(d_{ij}, \phi_{ij}|F^r_{ij}, F^l_{ij}) \\ &= \sum_{(i,j)\in\Omega} \{(1-\alpha)(1-\phi_{ij}) \sum_{(m,n)\in F_{i,j}} (f^l_{mn} - f^r_{mn\oplus d_{ij}})^2 \\ &\quad + \alpha \sum_{\eta_{ij}} (d_{ij} - d_{\eta_{ij}})^2 (1-\phi_{\eta_{ij}}) + \gamma \sum_{c\in C} V_c(\phi_{ij}, \phi_{\eta_{ij}})\} \end{aligned} \qquad (15)$$

where the F_{ij} denotes the intensity image block and f_{mn} represents the pixels within the block F_{ij}. C represents a pre-specified set of cliques for the occlusion and V_c is a potential function for the cliques. The parameters α and γ are weighting constants and η_{ij} denotes a neighborhood. The constant α is determined according to the noise level of the two images. For example, if we set α to be zero for the noise-free images the equation will be similar to the simple BM algorithm. We increase it according to the noise level to reduce the noise effects.

3 DE AND STEREO IMAGE ENCODING

3.1 Disparity Estimation

Due to the nonlinearities of the energy equation $U(D, \Phi|F^r, F^l)$ in (15), optimal disparity solutions are difficult to obtain. There are several ways to solve the problem. One popular approach is to use a stochastic relaxation algorithm, such as simulated annealing[8,9] or MFT,[10] which require a high degree of computation to find an accurate disparity solution. To reduce the computational burden we can use non-optimal, deterministic relaxation which converges to a local minimum. The problem can also be solved using neural networks; a parallel implementation is possible due to the localization properties of the problem.[11] When we implement a parallel algorithm, we have to partition the disparity field into disjoint regions and update the disjoint regions simultaneously at each iteration using the previous iteration result.

In this paper instead of using the relaxation algorithm, for simplicity, we estimate the disparity with a minimum energy by full search within a search window. For disparity estimation, first we divide the corresponding image into blocks and then for each block the best matching block is estimated within the search window according to the energy equation. The disparity map is formed by arranging the individual disparity of each block. Next, given the disparity, we label blocks with high energy value in (15) as occlusion blocks. We estimate disparity and occlusion iteratively until good solutions are found.

As can be seen in Figure 3, if we reduce the block size used for disparity estimation, the MAE between the disparity estimated blocks and the original blocks is decreased but obviously the bit rate required to transmit the disparity field increases. Figure 4(a) and (c) show the resulting disparity map of the SYN.256 image obtained from BM for two different block sizes, 4×4 and 2×2. Note that as the block size is reduced the resulting disparity map appears to be noisy. Compare those results with those of Figure 4(b) and (d). As expected, using the MRF method a smoother disparity map can be achieved. In this paper, we will choose a block size of 8×8 as a compromise between overhead and error image energy. Figures 5 and 6 show the results of the disparity estimation for the SYN.256 image and the LAB.480 image, respectively. Figure 5 (a) shows the intensity disparity maps from disparity estimation based on BM and (c) shows the disparity map from the MRF model. Figure 5 (b) presents the initial occlusion indicator decided by the simple threshoding after disparity estimation based on block matching and (d) shows the connected occlusion indicator by the MRF model. Figure 6 compares the results for the LAB image.

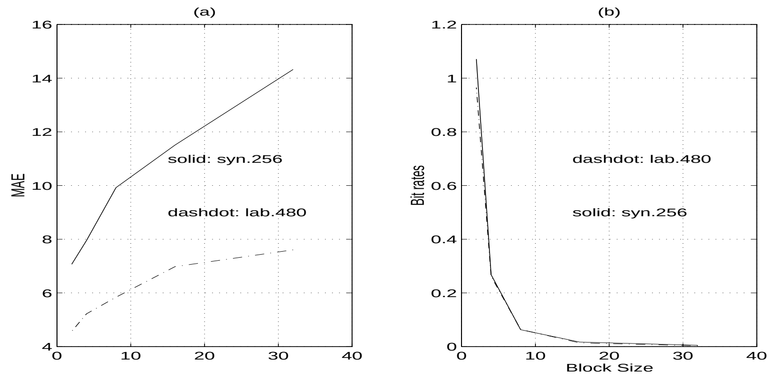


Figure 3: Effects of block size. (a) MAE vs. block size (b) Bit rates (bits per block) for Disparity vs. block size

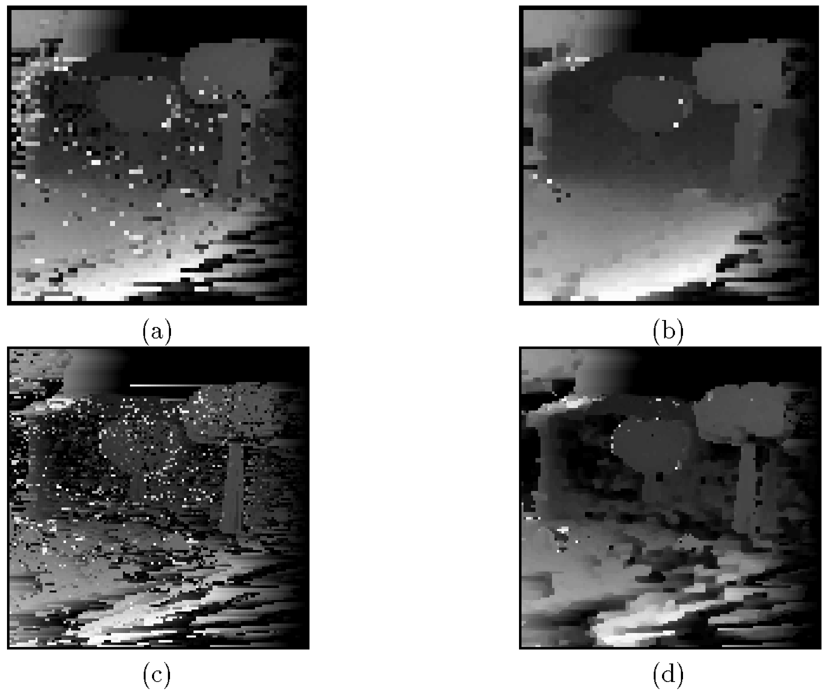
3.2 Stereo Image Encoding

Stereo pairs of images can be encoded more efficiently than two independent images because of the binocular redundancy between two images.[2] Figure 7 shows the basic structure for stereo image encoding and decoding structure. In this experiment, the right image is selected as reference image to make disparity always positive. The basic procedure of stereo image compression is as follows. In the encoding, we encode the right image independently and then estimate the disparity and the occlusion based on the energy equation derived from the MRF model. The resulting difference between two stereo images after disparity compensation is encoded. There exist two types of error image blocks, namely occluded and non-occluded blocks. For the occluded blocks we encode the original left image instead of the error image to increase the encoding efficiency. We use the JPEG algorithm to encode the reference image and the error image. We use smaller quantization steps to allocate more bits for the occluded blocks because the performance of the reconstructed image highly depends on the occluded blocks. For each group of blocks, we use two different Huffman tables. The image quality of the reconstructed image can be adjusted by changing the quality factor, which controls the JPEG quantization table, for the error image. The disparity and the initial occlusion maps are sent or stored as side information using a simple DPCM. At the decoder, we first decode the reference image and then decompress the left image using the reference image, error image, disparity and occlusion map.

4 EXPERIMENTAL RESULTS AND DISCUSSION

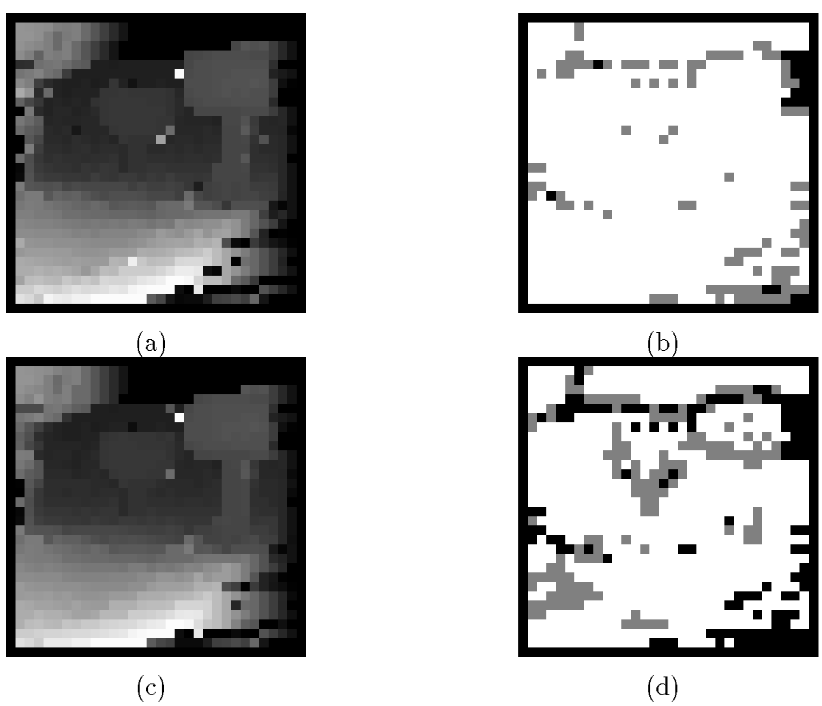
In order to test the effectiveness of the proposed algorithm using the MRF model we have simulated for the two test stereo images; *a synthesized image*, *SYN.256*, and *an in-door scene*, *LAB.480*, satisfying the parallel axis constraint. The image sizes used are 256×256 and 480×480, respectively. The selected block size for DE and DCT is 8×8. The performance is measured in terms of the bit rates of the encoded image and peak signal to noise ratio(PSNR). The bit rates for the disparity and occlusion maps are estimated using the entropy (in the case of disparity we use the entropy of the difference between the disparity of adjacent blocks).

In Figure 9, (a) and (b) show the original stereo pair $SYN.256$. Figure 9(c) shows the reconstructed left image from a simple block matching based method and (d) shows the result from the MRF based method. As expected, disparity compensated encoding achieves better performance than the single image encoding using JPEG. For this image we had relatively less gain because the resulting disparity map from the BM is smooth. As seen, for the reconstructed images corresponding to a PSNR of about 28 dB the block matching method results in the annoying blocking artifacts along the object boundary which are related to the occlusion effects. A comparison



(a) (b) (c) (d)

Figure 4: Disparity Estimation vs. Block size for the synthesized image(256x256). (a) 4×4 BM (b) 4×4 MRF (c) 2×2 BM (d) 2×2 MRF. Note that for the BM method the images appear to be noisier as the block size decreases. This noise disappears using the MRF model. The shapes of the objects can then be clearly seen.



(a) (b) (c) (d)

Figure 5: Disparity estimation results for the Synthesized Image. (a) Disparity with BM(0.063[bits/pixel]) (c) Disparity with MRF(0.063[bits/pixel]) and (d) Occlusion with MRF. Note that in this case we have not imposed strong smoothness constraints on the disparity field. The initial occlusion of (b) has been modified to produce a connected occlusion field.

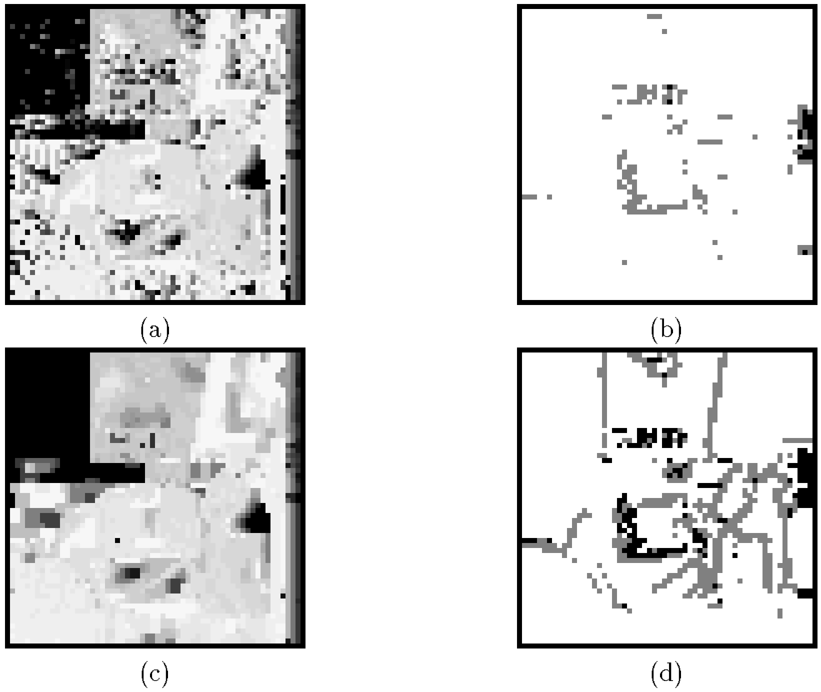


Figure 6: The results of the disparity estimation for the LAB image. (a) Disparity with BM(0.064[bits/pixel]) (b) Initial Occlusion Indicator(0.001[bits/pixel]) (c) Disparity with MRF(0.045[bits/pixel]) and (d) Occlusion with MRF. Note that in this case, through an appropriate choice of the parameter α there is a significant reduction in noise in the disparity field.

of the bit rates and PSNR for the disparity compensated image is given in Table 1. The MRF method converges in 3 or 4 iterations.

Figure 10, (a) and (b) show the original stereo pair LAB.480. Figure 10 (c) shows the reconstructed left image from a simple block matching based method and (d) shows the result from the MRF based method. For the reconstructed left image of the LAB image a comparison is given in Table 2 in terms of the bit rates and PSNR. Table 2 represents analogous results to those of Table 1 but the reconstructed left image has a PSNR of 30.25dB at 0.104 bits per pixel, which is more than 2dB over the PSNR obtained from BM (28.89 dB at 0.124 bits per pixel). For this image, the gains come from both the smooth disparity and the occlusion indicator. According to the results, the MRF based method obtained about 1-2dB higher PSNR than the simple block matching method at the low bit rates ranges, *e.g.*, 0.1-0.2 bits per pixel. Figure 8 (a) and (b) show performance comparison in terms of the rate-distortion for synthesized image, SYN.256, and indoor image, LAB.480, respectively.

Method	Disparity		Occlusion		Error Image		Total	PSNR
	[/block]	[/pixel]	[/block]	[/pixel]	[/pixel]	MAE	[/pixel]	[dB]
JPEG	-	-	-	-	-	-	0.282	26.95
BM	4.029	0.063	-	-	0.167	9.923	0.230	28.18
MRF	4.029	0.063	0.191	0.003	0.152	4.403	0.221	28.33

Table 1: Results of the Synthesized Image(search window=96, α=0.5, γ=100, T_ϕ=15, and iteration=3)

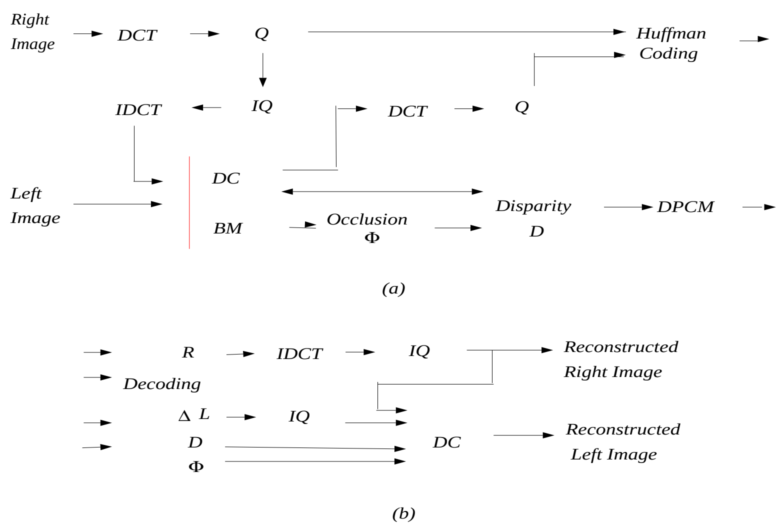


Figure 7: Disparity Compensated Coding for Stereo Image (a) Encoding (b) Decoding

Method	Disparity rate		Occlusion rate		Error Image		Total	PSNR
	[/block]	[/pixel]	[/block]	[/pixel]	[bpp]	MAE	[bpp]	[dB]
JPEG	-	-	-	-	-	-	0.116	25.10
BM	4.108	0.064	-	-	0.160	5.840	0.124	28.89
MRF	2.711	0.045	0.068	0.001	0.058	4.403	0.104	30.25

Table 2: Results of the Synthesized Image(search window=32, α=0.95, γ=100, T_ϕ=15, and iteration=3)

It is clear that the MRF model based approach also provides perceptually better image quality than the BM method from Figure 9 and Figure 10. In particular, it gives perceptually more satisfactory results because the human visual system is more sensitive to the visible artifacts at boundaries. This is especially important at low bit rates. Though the wrong disparity obtained from BM reduce bit rates for the difference between the original image and disparity compensated image, they may increase the entropy of the disparity field. By treating the occlusion indicator separately the MRF method increases encoding efficiency because the occluded blocks have relatively higher matching error and result in higher bit rates.

5 CONCLUSION AND FUTURE DIRECTIONS

In this paper, we have developed a stereo image compression algorithm based on the MRF model to improve encoding efficiency by exploiting several useful constraints. We estimate a smooth disparity field using the spatial correlation between blocks in a neighborhood to avoid various error effects due to noise and occlusion. We also introduced an occlusion indicator to reduce the matching error for blocks containing object boundaries. According to the result it is clear the MRF method provides more realistic vector field and it also can be used as an initial disparity to reconstruct a denser disparity map at the decoder.

We showed that this approach is more efficient than a simple block matching method for the stereo image coding. However, there still remain several problems to be tackled. JPEG may not be well suited for encoding the error image because we can consider the difference image as a white random noise. Another problem is that

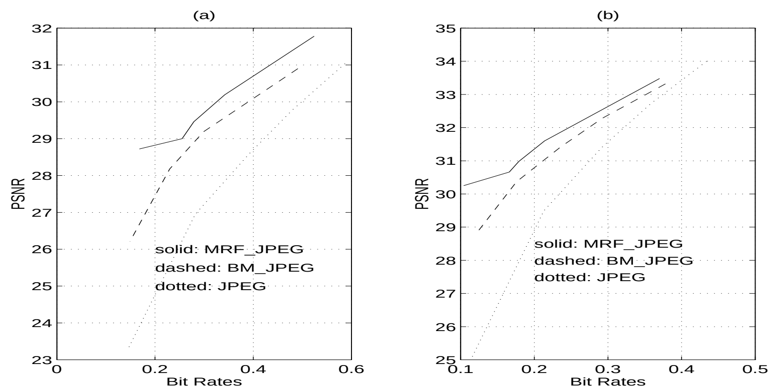


Figure 8: Rate-Distortion (R-D) Results. (a) Synthesized image (b) Lab Image.

for the fixed block based disparity estimation a vector represents the disparity of all pixels within the block and it results in matching error. To take into account these deficiencies and improve the coding efficiency, we plan to test adaptive block size matching and use different techniques to encode the error images. It is worth noting that applying our method to stereo video sequence coding, will be advantageous since the smooth disparity can help reduce the temporal redundancy between two consecutive disparity maps. Thus we need to integrate the motion and disparity information to compress the stereo video sequence more effectively. Finally, we also plan to study and measure how well the depth perception is preserved in the reconstructed image.

Acknowledgement: We wish to thank Dr. Haluk Aydinoglu from Georgia Tech for providing us with a test stereo image and for his friendly answer to our questions.

6 REFERENCES

[1] D. R. Shres, F.F. Holly, and P.G. Harnder. High ratio bandwidth reduction of video imaging for teleoperation. *SPIE Image and video Processing*, 1903:236–245, November 1993.

[2] M. E. Lukacs. Predictive coding of multi-viewpoint image sets. In *Proc. on ICASSP 86*, pages 521–524, October 1986.

[3] I. Dinstein, G. Guy, and J. Rabany. On the compression of stereo images: Preliminary results. *Signal Processing*, 17(4):373–381, August 1989.

[4] M. G. Perkins. Data compression of stereopairs. *IEEE Tran. on Comm*, 40:684–696, April 1992.

[5] T. Ozkan and E. Salari. Coding of stereoscopic images. *SPIE Image and video Processing*, 1903:228–235, 1993.

[6] H. Aydinoglu, F. Kossentini, Q. Jiang, and M. H. Hayes. Region based stereo image coding. In *Proc. IEEE ICIP'95*, pages 57–60, Washington, October 1995.

[7] M. Accame, F. G. De Natal, and D. D. Giusto. Hierarchical block matching for disparity estimation in stereo sequence. In *Proc. IEEE ICIP'95*, pages 374–379, Washington, October 1995.

[8] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Tran. on PAMI*, pages 721–741, November 1984.

[9] J. Konrad and E. Dubois. Bayesian estimation of motion vector field. *IEEE Tr. On PAMI*, 14:910–927, September 1992.

[10] J. Zhang and J. Hanauer. The mean field theory for image motion estimation. In *Proc. on ICASSP 93*, volume 5, pages 197–200, 1993.

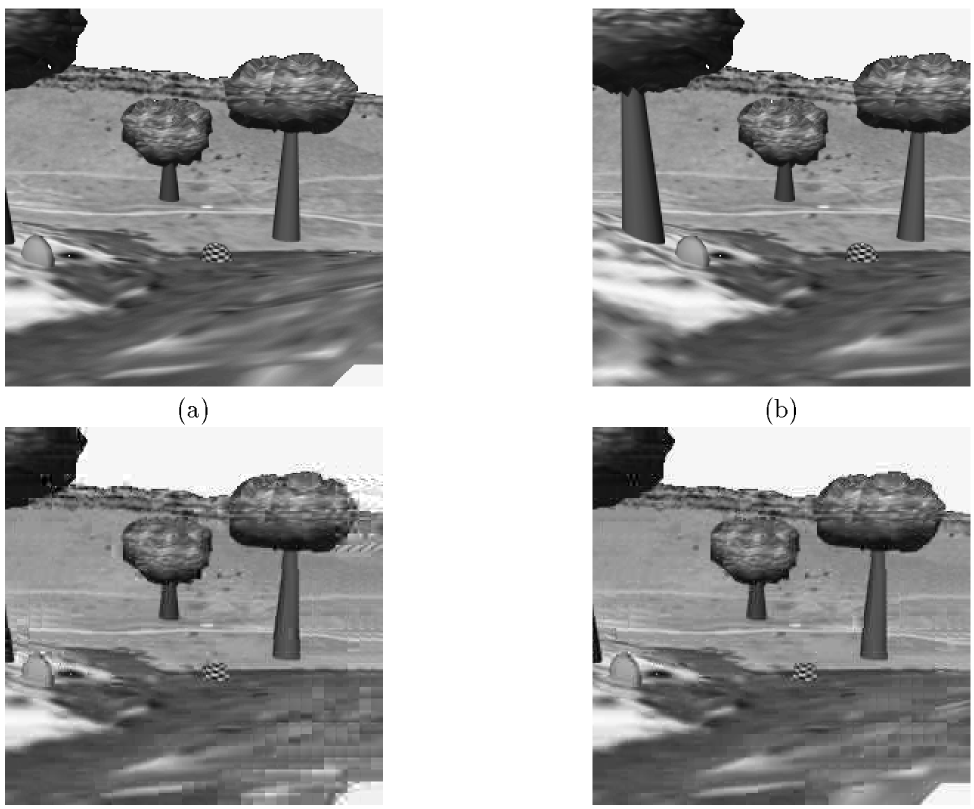
[11] H. Jeong, W. Woo, C. Kim, and J. Kim. A unification theory for early vision. In *Proc. on First Korea-Japan Joint Conf. on the Computer*, pages 298–309, October 1991.

[12] N. M. Nasrabadi, S. P. Clifford, and Y. Liu. Integration of stereo vision and optical flow by using an energy minimizing approach. *Optical Society of America*, 6:900–907, June 1989.

[13] J.E. Besag. Spatial interaction and the statistical analysis of lattice systems. *J. Royal Statistical. Soc.*, B36:192–236, 1974.

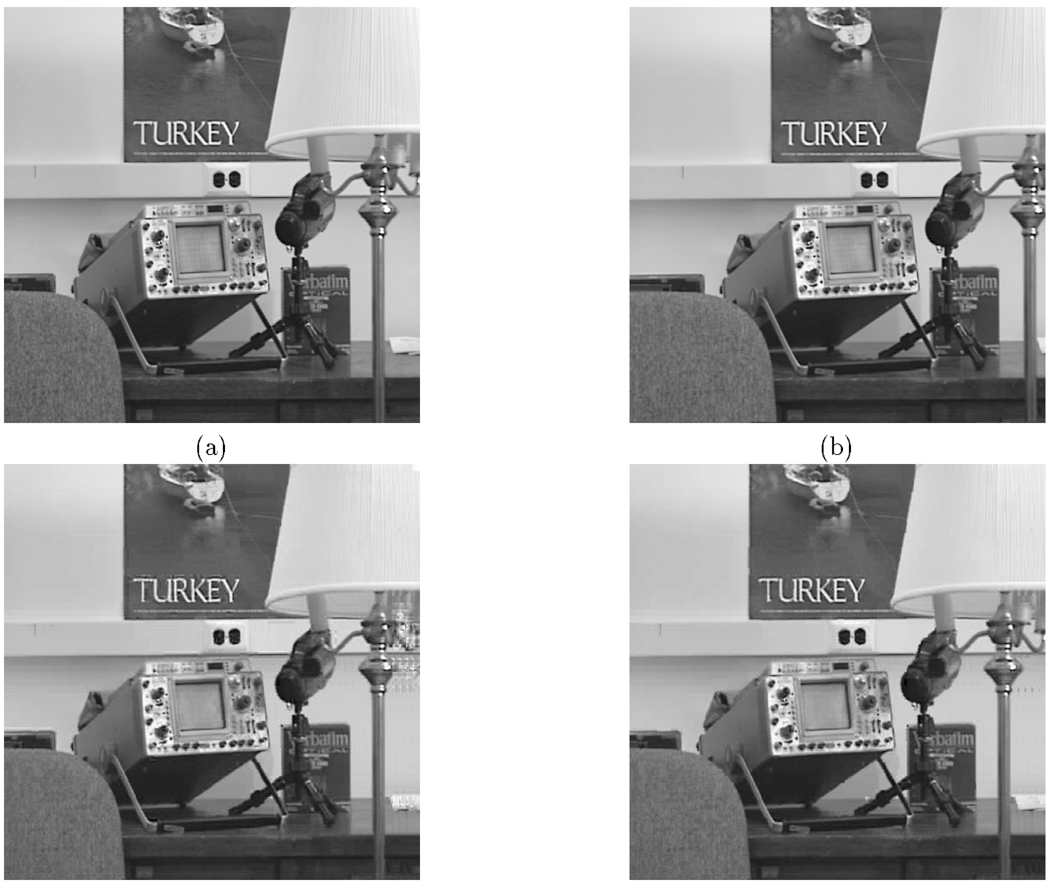
[14] B.K.P. Horn. *Robot Vision.* The MIT Press, 1986.

[15] W. Pennebaker and J. Mitchell. *JPEG Still Image Compression Standard.* Van Nostrand Rheinhold, 1994.



(a) (b) (c) (d)

Figure 9: Coding results, (synthesized image, 256x256 pixels, 8x8 DCT, 8x8 BM). (a) Original left image (b) Original right image (c) Reconstructed image by BM (0.230 bpp, 28.18 dB) (d) Reconstructed image by MRF (0.221 bpp, 28.33 dB)



(a) (b)

(c) (d)

Figure 10: Coding results (LAB image, 480x480 pixels, 8x8 DCT, 8x8 BM). (a) Original left image, (b) Original right image (c) Reconstructed image by BM (0.124 bpp, 28.89 dB) (d) Reconstructed Image by MRF(0.104 bpp, 30.25 dB)