In Proc. of Visual Communic. and Image Proc., VCIP '97, San Jose, CA, Feb. 1997

Stereo Image Compression based on Disparity Field Segmentation

Woontack Woo and Antonio Ortega

Integrated Media Systems Center
Department of Electrical Engineering-Systems
University of Southern California,
Los Angeles, California 90089-2564
Email: wwoo@sipi.usc.edu and ortega@sipi.usc.edu

ABSTRACT

The increasing demand for 3D imaging and recent developments of autostereoscopic displays will accelerate the usage of 3D systems in various areas. However, limited channel bandwidth is, as for monocular images, the main bottleneck for realizing 3D systems. As a result, an efficient compression algorithm will be essential to reduce the bandwidth requirement while maintaining the perceptual visual quality at the decoder. In this paper, we will focus on compression of stereo images. When it comes to stereo image coding, we can take advantage of binocular redundancy by using disparity compensation. The most popular disparity compensation method approaches so far have been block based methods, due mostly to their simplicity. Block based methods, however, may suffer from blocking artifacts at low bit rates due to the uniform disparity assumption within a fixed block. Meanwhile, if we reduce the block size, the disparity estimation may suffer from various noise effects which result in increases of bit rates for the disparity. Considering these observations, we estimate disparity based on a small block or a pixel with the energy equation derived from the MRF model. In order to prevent oversmoothing across boundaries, we use the combined intensity edges of two images as an initial disparity boundary. Then, we segment the resulting smooth disparity field. Finally, the disparity and the starting position are encoded using DPCM and the corresponding boundary is encoded using Run Length Chain coding. At the end of this paper, we present experimental results.

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

1 INTRODUCTION

Recently, 2-D visual communication technologies have matured so fast that various commercial systems are available for real-time visual communication based on standards such as JPEG, MPEG I and II, or H.263. Nevertheless, these techologies may not be sufficient for the increasing new demands for realistic or more natural representations of scenes. New 3D technologies are a proper solution for these requirements. Also, recent developments of autostereoscopic displays may open a new way for adding realism in the 2-D display and will accelerate the usage of 3D systems in various areas.

In this research, we consider the problem of efficient encoding of stereo images, which are frequently used to provide 3-D realism. In general, depth perception is obtained by simultaneously viewing a scene from different positions. If we can find correspondences in stereo images, we can then construct 3-D structures of the scene by a triangulation with the camera geometry. The stereo (or multi-view) images are bound to have a number of applications in near future; for example, in the fields of 3-D visualization (CAD), 3-D telemedicine, 3-D telerobotics, 3-D visual communications, 3-D HDTV and cinema, and Virtual Reality. However, limited channel bandwidth is, as for monocular images, the main bottleneck for making these systems possible since stereo images require to transmit or store enormous amount of data. As a result, an efficient compression algorithm will be essential to reduce the bandwidth requirements while maintaining the perceptual visual quality at the decoder.

This paper deals with stereo image coding based on disparity field segmentation. In stereo image encoding, we can take advantage of binocular redundancy by adopting disparity compensated coding. Fixed block based displacement estimation has been widely used in motion or disparity compensated coding.[1] Though they are simple and effective to implement, block based methods suffer from several main drawbacks; as the block size becomes smaller the overhead required to handle the disparity becomes too large. Block based methods may also fail to provide accurate matching results because the correspondences are locally ambiguous due to noise, occlusion, lack of texture, and repetitive texture. Conversely, larger block size will result in increases in the corresponding estimation error and thus it requires higher rates to transmit the difference images. In addition, if the block includes object boundaries, block based methods may suffer from visual artifacts such as blockiness at low bit rates. This is because the displacements are assumed to be the same within a block and in general different objects have different disparity. Since the human visual system (HVS) is sensitive to object boundaries, which are usually related to abrupt intensity changes, these blocking errors can be very annoying.

One way to avoid the problems associated with block based methods is to a variable block size by segmenting blocks with higher estimation error into smaller blocks, in order to reduce estimation error while maintaining the encoding efficiency.[2] Another possibility is to use an arbitrary shape segmentation-based approach.[3] Chu *et al.* showed that edge information can help segment range images effectively.[4] Marques *et al.* have adopted object based segmentation approaches in motion compensated coding.[5] In general, a correct displacement estimation based on arbitrary shapes improves the coding efficiency of the residue image by reducing the estimation error at the object boundary. As a result, segmentation or object based approaches fewer visible artifacts. In addition, segmentation simultaneously produces useful intermediate information for various applications such as scene analysis, synthesis, or generation for VR. However in order to achieve better overall coding rates than block-based methods, segmentation-based algorithms have to represent the segmentation information and the corresponding disparity map in an efficient manner. Therefore our goal should be to simplify the disparity/segmentation information so as to minimize the encoding cost while providing a good prediction. This involves finding a smooth disparity map and along with simple segmentation contours.

In this paper, we first derive an energy equation based on the MRF/GRF model to find a smooth disparity map with simple boundaries. In order to achieve these goals, we impose some useful and realistic constraints on the disparity estimates, such as similarity between intensity fields of corresponding stereo images, smoothness of the disparity field, a connected occlusion field, and smooth contours. We also combine two intensity edge fields and then use them as an initial disparity boundary which prevents disparity from being oversmoothed across boundaries. Then, given a smooth disparity field we make the disparity contour as simple as possible to reduce the bit rate of the target image of the stereo pairs using run length chain coding. The proposed compression algorithm can be used not only to reduce bandwidth but also to achieve efficient rendering of scenes. The potential for a parallel implementation is another desirable feature of our algorithm. This paper is organized as follows. In section 2, we briefly introduce the MRF method that will be used for image segmentation based on intensity and for disparity field estimation and segmentation. In section 3 we describe how these two MRF techniques can be used to estimate disparity, using the intensity field segmentation to help obtain a more efficient segmentation of the disparity field. We also describe our method for encoding the resulting disparity field based on the disparity field segmentation. Experimental results are presented in section 4. Finally, conclusions and future research directions are given in section 5.

2 IMAGE SEGMENTATION AND MRF

2.1 Segmentation and Region Representation

The main goal of segmentation-based coding is an efficient representation of images, *i.e.* one based on the classical principles of achieving the best visual quality for a given rate. An image can be segmented into homogeneous regions, and the contours (or shapes of segmented regions), textures (intensities) and segmentation errors have then to be encoded. The same ideas as will be seen can be used to segment a disparity field or a motion vector field. Segmented regions can be represented by global object descriptors such as the surface area, the perimeter, the center of mass, the length and width, or the convexity. These type of parameters are useful for object recognition and classification but not for coding since such low precision descriptions do not allow a good reconstruction. Other popular ways to represent segmented regions are region description methods and boundariy description methods.

An example of a region-based method is the quadtree representation, which assumes the image is unit square and the root node represents the entire image. The image is successively subdivided into quadrants until no further division is necessary. It guarantees a compact representation due to its hierarchical data structure. It, however, may not be good for image sequence compression since the descriptions are shift variant.

Boundary description methods are also widely used since they provide compact representation and are easy to implement.[6] The contour, the boundary representation of segmented region, can be found by edge detection or region growing and then can be encoded using chain coding. The total rate, B, for the contour coding of an $M \times N$ intensity image can be represented (not including the cost of sending the intensities within each segment) as follows

$$B = \sum_{i=0}^{c} \{|l_i| \times log_2(M \times N)\} \tag{1}$$

where c represents the number of contours and $|l|$ denotes contour length. It can be encoded using a chain code which encodes the sequence of contour directions. The total number of bits for the chain coding of an $M \times N$ intensity image can be represented as follows

$$B = \sum_{i=0}^{c} \{log_2(M \times N) + |l_i| \times log_2 k\} \tag{2}$$

where k denotes the neighbor connectivity used in the chain coding, say 4 or 8. The first term is needed to represent the starting point of a contour and the second term is required to describe the contour. Roughly speaking, for a 256×256 image we can achieve about 8 to 1 compression ratio when compared to a plain contour coding. One simple extension of the contour coding exploiting the redundancy, smoothness in contours, is differential or run length chain coding. However, chain coding methods are noise sensitive and also require more bits for detailed contour description. Therefore, we can tradeoff between given bit budget and desired precision by allowing approximate reconstruction using straight lines, curves, or polygons. Methods to achieve approximate curve reconstruction using rate-distortion criteria have recently been proposed.[7]

Once all the pixels in the image have been labeled by the segmentation procedure we need to encode the contours. We start by finding labeled contours for each segmented area using a 4-connected neighborhood, *i.e.*, if one of the 4 neighbors has a different value, then we decide it as a boundary. Then, we scan the image and detect a starting point. We keep the starting point and label, (d, x, y), and then start tracking the contour using 8-connected neighborhood. The tracking is finished when the starting point is reached. This process is repeated until whole contours in the image are tracked. We can save more bits by applying DPCM to the starting points and labels. We also can use differential or run length chain coding to further reduce the bit rate of the contour coding.

2.2 MRF/GRF Model and MAP Estimation

The main advantage of the MRF model based approach is that it provides a rigorous mathematical framework and a general model for the interaction among spatially related random variables. Another advantage of the MRF model is its ability to combine discontinuity and occlusion indicators into the energy equation. It reduces the disparity estimation error resulting from an occlusion effect by treating those blocks independently.[8] The resulting algorithm also can be implemented in parallel due to its inherent localization property.

Geman and Geman considered images as realizations of a stochastic process that consists of an observable noise process and a hidden edge process.[9] We apply this stochastic model to modeling of the stereo disparity field as well as intensity image. We model the spatial interactions among neighboring intensity pixels (disparities) based on the discrete MRF model and Gibbs distribution. Consider a random field of intensity, $F = \{f_{ij}, (i,j) \in \Omega\}$ (or 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, respectively, the vertical and horizontal size of the image (or the disparity field). Assume the intensity image, F (or D), is a MRF with respect to a neighborhood system $\eta = \{\eta_{ij}, (i,j) \in \Omega\}$, where η_{ij} is the neighborhood of f_{ij} (or d_{ij}) such that $(i,j) \notin \eta_{ij}$ and $(k,l) \in \eta_{ij}$, *i.e.*,

$$\begin{aligned} P\{F = f_{ij} | f_{kl}, (k,l) \in \Omega\} &= P\{f_{ij} | f_{kl}, (i,j) \neq (k,l), (k,l) \in \eta_{ij}\} \\ P\{D = d_{ij} | d_{kl}, (k,l) \in \Omega\} &= P\{d_{ij} | d_{kl}, (i,j) \neq (k,l), (k,l) \in \eta_{ij}\} \end{aligned} \tag{3}$$

The spatial interaction among the motion in the image sequences also can be defined similarly.

Fig. 1 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 1: Neighborhood systems and cliques: (a) Geometry of neighborhoods; the number denotes ith order neighborhood system. (b) First order neighborhood η^1 and cliques used for intensity, the disparity and the occlusion; we can quantify the effect of each clique according to the characteristics of the random fields.

Fig. 2 shows two different neighborhood systems for edge processes, horizontal and vertical edges, respectively.[10] Considering the model constraints, an isolated edge is inhibited and a connected edge is encouraged even if the intensity (or disparity) changes slightly.

2.3 Image Segmentation using MRF

Our first goal is to find simple and connected contours for the objects/regions in the image. The segmented regions will then be used to help segment the disparity field using the MRF model. The resulting edge is a good initial guess of the disparity edge, though the intensity discontinuities may not correspond to physical boundaries.

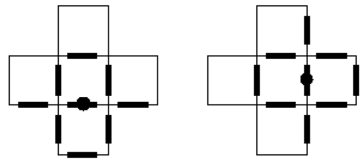

Figure 2: Neighborhood System for Edge Process for: (a) Vertical Edge (b) Horizontal Edge

We can formulate an intensity image segmentation as follows. For a given intensity image, G, we want to find a smooth intensity image, F, and intensity edge, L^I, such that solutions maximize the *a posteriori* probability (MAP), $P(F, L^D|G)$. We can decompose the posterior probability using Bayes theorem as follows

$$P(L^I, F|G) = \frac{P(G|L^I, F)P(F|L^I)P(L^I)}{P(G)} \propto P(G|F)P(F|L^I)P(L^I) \tag{4}$$

where F is smoothed image of given intensity image G. In the above equation, the first term is called a noise process or a degradation model, the second term denotes a smoothness model and the third term represents connected edge process. The denominator $P(G)$ can be ignored since it is independent on the smooth image or intensity edge.

According to the Clifford-Hammersley theorem,[11] if a measure can be modeled by a MRF, then the probability mass of the measure can be represented in the Gibbs distribution form. The first term of the right side in (4) is called as the observation process. It can be represented in the Gibbs distribution form as follows

$$P(G|F) = \frac{1}{Z} exp\{-\frac{\alpha}{T} U(G|F)\} \tag{5}$$

where Z is normalization constant α and T is a temperature. The energy functions designate the constraints of the strong similarity between the noise image and the original image. The energy function for a pixel (i, j), $U(g_{ij}|f_{ij})$, can be represented as follows

$$U(g_{ij}|f_{ij}) = (g_{ij} - f_{ij})^2 \tag{6}$$

where the second g denotes given intensity image and f represents reconstructed intensity image.

The second term in (4) represents an *a priori* assumption on the smoothness of the intensity field, F, given an intensity edge, L^I. The *a priori* distribution for F with L^I also can be represented as Gibbs distribution form

$$P(F|L^I) = \frac{1}{Z} exp\{-\frac{\beta}{T} U(F|L^I)\} \tag{7}$$

where Z is normalization constant β and T is a temperature. The energy function $U(f_{ij}|l_{ij})$ can be represented as follows

$$U(f_{ij}|l_{ij}) = \sum_{\eta_{ij}} (1 - l_{\eta_{ij}})(f_{ij} - f_{\eta_{ij}})^2 \tag{8}$$

where f_η represents neighboring pixels and l_η represents corresponding discontinuities.

We can also decide edge process initially by the intensity difference of the noisy image. The initial discontinuity process is defined as

$$l_{ij} = \begin{cases} 1, & |f_{ij} - f_{\eta_{ij}}| \geq T_d \\ 0, & o.w. \end{cases} \tag{9}$$

where T_d is threshold for edge decision. If the difference between the intensity and its neighborhood exceed a threshold T_d, then there is discontinuity. In this case the smoothness constraints should not be performed across this discontinuity.

Finally, we have

$$P(L^I, F|G) \propto exp\{-\frac{\alpha}{T}U(G|F)\}exp\{-\frac{\beta}{T}U(F|L^I)\}exp\{-\frac{\gamma}{T}U(L^I)\} \tag{10}$$

where each term represents a noise process, a smooth intensity field, and a intensity edge process, respectively. The overall energy function for intensity image segmentation can be represented as equation (4) and the energy for a pixel (i,j) ca be calculated as follows

$$U(l_{ij}^I, f_{ij}|g_{ij}) = \alpha(f_{ij} - g_{ij})^2 + (1-\alpha)\sum_{\eta_{ij}}(1 - l_{\eta_{ij}})(f_{ij} - f_{\eta_{ij}})^2 + \gamma V_c(l_{ij}, l_{\eta_{ij}})\} \tag{11}$$

where α and γ are weighting constants. We set β to be equal $(1-\alpha)$ since they are related to the noise level of the given image.

2.4 MRF Model for Disparity Estimation

We estimate disparity using smoothed images and intensity edges of the previous step. To estimate the disparity, we consider a coupled MRF model consisting of a disparity process, a disparity edge, and an occlusion process. The estimation is also 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.

We can similarly formulate a disparity estimation (DE) problem as we did in our previous work.[8] The main difference is the disparity boundary process which is used to compactly represent the segmented area as well as prevent the disparity field to be oversmoothed across the boundary.

For given stereo pairs, F^L and F^R, we want to find disparity, D, occlusion, Φ, and disparity edge, L^D, such that solutions maximize the *a posteriori* probability, $P(D, \Phi, L^D|F^L, F^R)$. We can decompose the posterior probability using Bayes theorem as follows

$$P(D, \Phi, L^D|F^R, F^L) = \frac{P(F^L|F^R, D, \Phi, L^D)P(D|\Phi, L^D)P(L^D)P(\Phi)}{P(F^L|F^R)} \tag{12}$$

where we assume that the disparity, D, depends upon the occlusion, Φ, and disparity edge, L^D. The disparity edge and the occlusion are independent of the right image, F^R, because the right image itself does not directly effect on the decision of the disparity edge and the occlusion in the MAP estimation. The denominator of the above equation also can be ignored in the MAP estimation because it is not a function of the disparity, D, the disparity edge, L^D, or the occlusion, Φ. Therefore, the posterior probability is proportional to the numerator in (12) as follows

$$P(D, L^D, \Phi|F^R, F^L) \propto P(F^L|F^R, D, L^D, \Phi)P(D|L^D, \Phi)P(L^D)P(\Phi) \tag{13}$$

Using again the Clifford-Hammersley theorem,[11] we have

$$P(D, \Phi, L^D|F^R, F^L) = \frac{1}{Z}exp\{-U(D, \Phi, L^D|F^R, F^L)\} \tag{14}$$

where Z is a normalization constant and $U(D, \Phi, L^D|F^L, F^R)$ is an energy function. The MAP estimation problem can be replaced by the energy minimization problem, *i.e.*, the solutions, D, Φ, and L^D minimizing the energy function $U(D, \Phi, L^D|F^L, F^R)$, are the solutions maximizing the posterior probability $P(D, \Phi, L^D|F^L, F^R)$. We can formulate the disparity estimation problem as follows

$$\begin{aligned}
(\hat{D}, \hat{\Phi}, \hat{L}^D) &= \arg\max_{(D,\Phi)} P(D, \Phi, L^D|F^R, F^L) \\
&= \arg\min_{(D,\Phi)} U(D, \Phi, L^D|F^R, F^L) \\
&= \arg\min_{(D,\Phi)} \{U(F^L|F^R, D, \Phi) + U(D|\Phi, L^D) + U(L^D) + U(\Phi)\}
\end{aligned} \tag{15}$$

where each term represents useful constraints for the disparity estimation, *i.e.*, similarity, smoothness, discontinuity constraints, and occlusion, respectively.

The first term of the right side in (15) 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 image pair, F^L and F^R, are similar after disparity compensation and tries to find the matching block with an optimal cost value between F^L and F^R. Therefore, block based approaches result in erroneous matching for those blocks which contain object boundaries. In order to avoid the occlusion effects we adopt an occlusion indicator and then we segment the block selected as an occlusion candidate into smaller blocks. If the small block (or pixel) has still lager DE error than theshold, it is classified as an occlusion.

The second term in (15) represents an *a priori* assumption on the smoothness of the disparity field, D, given the disparity edge, L^D, and the occlusion, Φ, which will be used to trade-off smoothness and estimation error. We assume that the real disparity field is smooth except for the object boundaries that are related to the depth discontinuities. 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).

The third term in (15) denotes a disparity contour process. We use the intensity edge obtained at the image segmentation step as an initial disparity edge process. The disparity edge controls the discontinuity between the disparity and its neighborhood. The main role of the disparity edge is to prevent disparity from being oversmoothed across object boundaries. Thus, smoothness constraints should not be performed across this discontinuity. The smooth contours are also used to compactly represent segmented disparity fields.

The last term in (15) denotes an occlusion process. We impose an *a priori* assumption on the occlusion field so as to force it to be connected and then the isolated occlusion is inhibited. The initial occlusion candidate is decided by comparing the magnitude of the mean absolute error (MAE) of the block with a pre-selected threshold. It is natural to discard the estimates for the occluded block (pixel) because the occluded block tend to increase the entropy of the error image. Even though, the connected occlusions increase the bits for DE errors, they reduce the bits for the representations of occlusion areas. From the view point of perceptual image quality, the advantages of occlusion indicators outweigh the disadvantages, *i.e.*, the increases in bits for DE.

Given the above model, we can derive the overall energy function, $U(D, \Phi, L^D|F^R, F^L)$, for the segmentation-based disparity estimation as follows

$$\begin{aligned} U(D,\Phi|F^R,F^L) &= \sum_{(i,j)\in\Omega} U(d_{ij},\phi_{ij},l_{ij}|F^r_{ij},F^l_{ij}) \\ &= \sum_{(i,j)\in\Omega} \{(1-\alpha)(1-\phi_{ij}) \sum_{(m,n)\in F_{ij}} (f^l_{mn} - f^r_{mn\oplus d_{ij}})^2 \\ &\quad +\alpha \sum_{\eta_{ij}} (d_{ij} - d_{\eta_{ij}})^2 (1-l_{ij})(1-\phi_{\eta_{ij}}) + \beta \sum_{c\in C} V_c(l_{ij}, l_{\eta_{ij}}) + \gamma \sum_{c\in C} V_c(\phi_{ij},\phi_{\eta_{ij}})\} \end{aligned} \tag{16}$$

where F_{ij} denotes a block of pixels and f_{mn} represents pixels within the block, *i.e.*, the block size is 1×1, F and f is the same. In the above equation, d, ϕ, and l represent a disparity, an occlusion, and a disparity contour, respectively. The disparity edge l controls the discontinuity between the disparity, d, and its neighborhood, d_η. An C represents a pre-specified set of cliques, c, for the occlusion and V_c is a potential function for the cliques. The larger neighborhood, η, the greater the influence of neighboring disparities. The parameters α, β, and γ are weighting constants. 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 STEREO IMAGE CODING BASED ON SEGMENTATION

3.1 Disparity Estimation and Segmentation

Using (16), we estimate a smooth disparity field. The occluded area can be decided by comparing an estimation error and a clique potential of an occlusion indicator. Similarly, the disparity contour can be decided. Finally, we segment the resulting smooth disparity map in terms of disparity and occlusion.

For the disparity estimation we only use the first and the second term in (16) since the other terms are not a function of disparity. We estimate disparity by tradeoffs between similarity of intensities and smoothness of the disparity field. To find occlusion candidates, the first and the last terms are used. In this case, the role of the first term is a kind of dynamic thresholding for the decision of an occlusion according to the state of the neighboring occlusions. For example, a pixel with lower MAE can be decided as an occlusion, if its neighboring pixels are in an occlusion area. Also, even though a pixel has high MAE, it can be labelled as a nonocclusion, if its neighboring pixels are in nonocclusion area. Similarly, the disparity edge is decided using the first and the third term. The same statement with occlusion can be done for the disparity boundary.

Due to the nonlinearities of the energy equation $U(D, \Phi, L^D | F^R, F^L)$ in (16), 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[9,12] or MFT,[13] 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.[10] 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.

First, we assume the given stereo images are smooth enough and then segment intensity images using a simple threshold method. Using the MRF model we try to find smoothly connected intensity edges. The resulting combined intensity edges are used as an initial disparity boundary. It is a good initial guess, though all intensity edges do not correspond to disparity boundaries. For disparity estimation, we divide the target image into blocks and estimate initial disparity based on the fixed block matching.

Then, for each block the best matching block is estimated within the search window. Though the disparity minimizing MSE is an effective solution in terms of bit rate for Disparity compensated error (DCE), it may not be optimal when we consider the DCE and disparity together. Thus, we use in the energy equation containing smoothness constrained. Then, we segment a block with higher Mean Absolute Error (MAE) than given threshold into small blocks and for each subblock we estimate disparity again. We repeat this process until the block with higher MAE reaches a pixel or the MAE of the block is smaller than the given threshold. When the block or pixel has larger MAE than the threshold, we consider it an initial occlusion. The resulting disparity map is formed by arranging the individual disparity of each block and used as an initial disparity.

Given an initial disparity edge and an initial disparity we estimate a smooth disparity field to make the entire coding efficient based on the energy equation derived from MRF model and MAP. Next, given the disparity, we label blocks with high energy value in equation (16) as occlusion candidates with higher DCE. Then, we eliminate redundant disparity boundaries and find smooth disparity contour since fewer numbers of region and shorter boundary length requires less bit rate. We repeat these steps until we get a reasonable solution in terms of bit rates and distortion. The resulting disparity field and corresponding contours are encoded using differential chain coding.

3.2 Stereo Image Coding based on Disparity Field Segmentation

In general, we can get perceptually better image quality as well as reduced DCE by encoding the target image based on the disparity segmentation if we assume that disparity contours coincide with object boundaries. Figure 3 shows the block diagram of the stereo image encoding using the MRF model and segmentation.

The reference image is encoded in intraframe mode using transform methods such as DCT or wavelet transform. We do not encode the reference image based on segmentation since the quality of the reconstructed image highly depends on that of the reference image and generally conventional transform methods provide better performance at higher bit rates than segmentation based coding does. First, we find disparity contour using a 4-connected

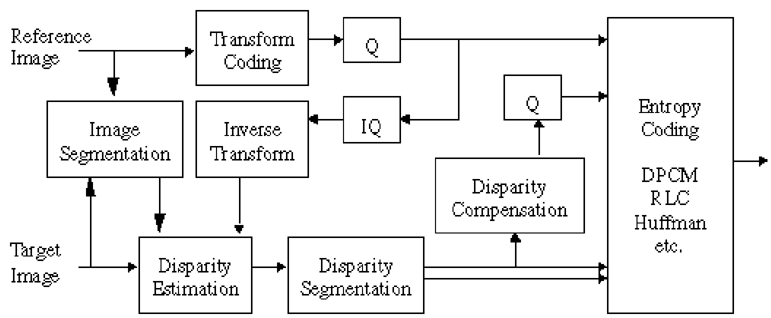

Figure 3: Stereo Image Encoding using Segmentation based Disparity Compensation

neighborhood from the smooth disparity map resulting from MRF model based disparity estimation. Then, we trace the contours using an 8-connected neighborhood. Then we encode contour using the differential chain coding or run length chain coding followed by Huffman coding. Simultaneously, we encode the disparity and the starting point using DPCM and Huffman coding. Finally, we encode the estimation error which includes the intensity error for the occluded error. There is no side information for the occlusion since the maximum disparity represents an occlusion.

The DCE is computed using the smooth disparity map and stereo images. The segmented areas with high energy, *i.e.*, requiring higher bit rate than intra mode coding, are encoded independently by encoding the original image instead of DCE. In conventional approach, DCE is encoded in a similar manner to that of intra frame by applying DCT to every square block. We use scalar quantizer and DPCM followed Huffman encoding for DCE.

The decoding is the inverse process of the encoding process. At the decoder, the reference image is first decoded. Then the target image is reconstructed using the reference image and side information such as disparity with occlusion and compensated error.

4 EXPERIMENTAL RESULTS AND DISCUSSION

In order to show the effectiveness of the proposed algorithm we test the proposed algorithm on simple synthesized images. The performance is measured in terms of the bit rate of the target image and the peak signal to noise ratio (PSNR). The first test image consists of a rectangle, a circle, and a triangle and has random texture as shown in Figure 4 (a). Figure 4 (b) shows the disparity estimation results based on the simple block matching. The resulting DCE has a blocky appearance as we are limited to a single disparity vector per block. As expected the errors occur along the object boundaries where a different estimate for object and background is needed.

In order to solve this kind of problem which comes from the uniform disparity assumption within the block we included line processes in the MRF Model. Figure 4 (c) shows the combined intensity edge fields which are obtained based on the energy equation driven in Section 2.3. The combined edge is used as an initial disparity boundary to prevent the disparity field from beeing oversmoothed accross the boundary of segmented areas. The disparity field can then be found using the energy equation derived in Section 2.4. The resulting smooth disparity field and corresponding disparity boundary are shown in Figure 4(e) and (f), respectively. As expected, the MRF model based method found a smooth and accurate disparity field which can be used to generate intermediate scenes in the virtual environment while reducing the disparity compensated error. The smooth disparity field with connected occlusion areas reduces the bit rate for the disparity field itself. Furthermore, main advantages of separation of occlusion areas are in the 3D visual perception since the occluded areas usually do not have any effect on depth perception. Therefore, even if we reduce bit rate for the occluded areas, with the corresponding loss in overall PSNR, we can perceive 3-D without loss of depth information. Finally, Figure 4(d) shows the reconstructed image using the disparity contour, segmentation based disparity, and DCE. The segmentation based coding results in no blocking artifacts which are obvious in human visual system.

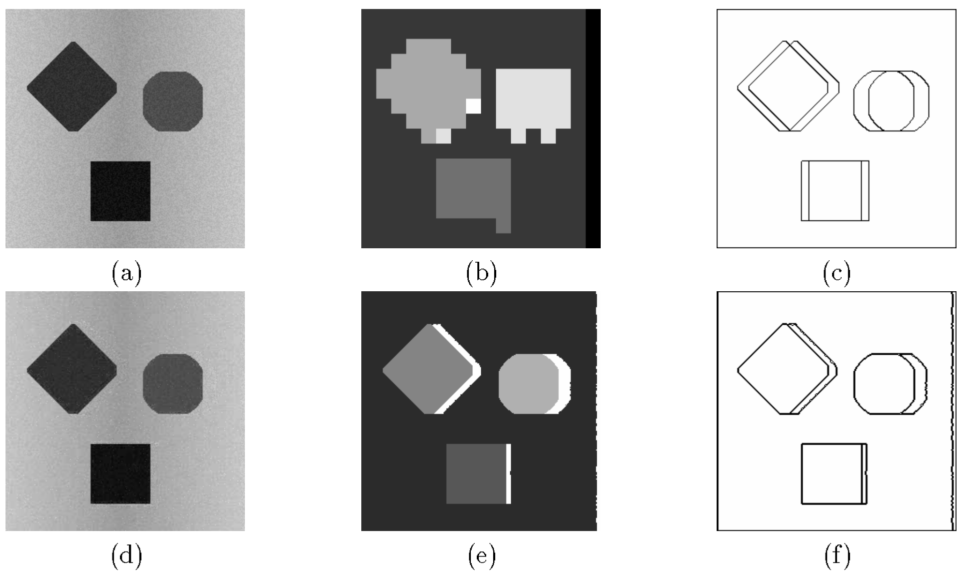

Figure 4: Results of Disparity Estimation for synthesized image. (a) target image (b) initial disparity (c) combined intensity edges (d) reconstructed target image (e) disparity with occlusion (f) disparity contour

Figure 5 shows disparity estimation results and a reconstructed image. The original image, *bust*, was provided by *INRIA-Syntim* in the gif format with size of $512x512$ and we coverted it intensity image with size of $256x256$. Similarly, the proposed algorithm provides a smooth and exact disparity field.

We encoded the disparity contour using an eight-connected chain code and the disparity compensated error using constant quantizer followed by DPCM and RLC. As a result, we reduced the bit rate for the disparity itself as well as for the disparity compensated error. We also can take advantage of the fact that neighboring disparity contour tend to change smoothly by adopting DPCM after chain coding. The proposed algorithm provides better perceptual quality as well as higher PSNR than block based methods do. Figure 6 shows the comparison of the rate-distortion(RD) performance among the results of JPEG, Block based, and MRF with segmentation based coding.

According to the experimental results, the proposed algorithm achieves 0.5-0.7dB better performance. The main gain comes from a smooth disparity field and compact boundary descriptions. At the same low bit rate, though the reconstructed target image based on the fixed block suffers from blocking artifacts, especially at

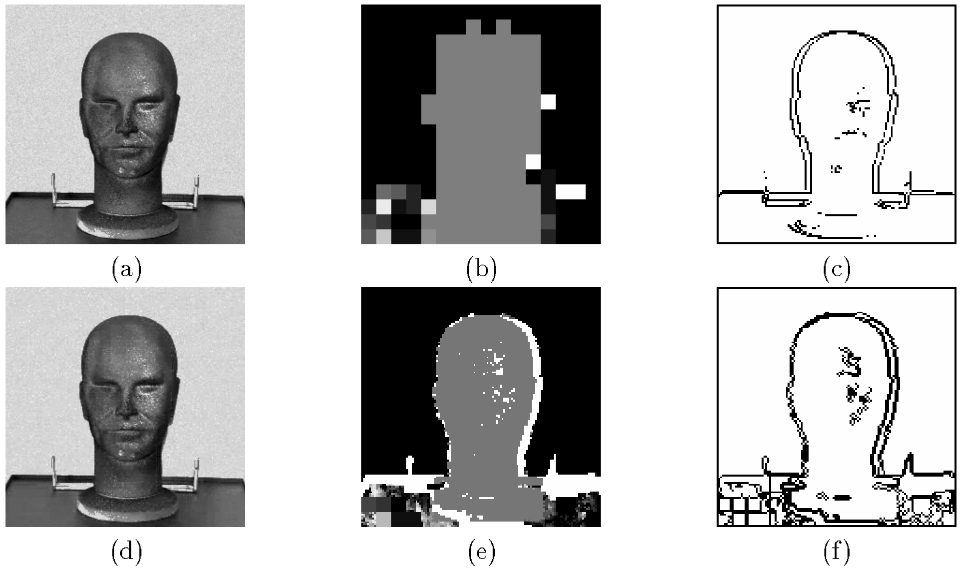

Figure 5: Results of Disparity Estimation for synthesized image. (a) target image (b) initial disparity (c) combined intensity edges (d) reconstructed target image (e) disparity with occlusion (f) disparity contour

object boundaries, the proposed algorithm provides no artifacts at object boundaties and produces less perceptual artifacts. Currently, the main drawback of this approach is that lots of bits are used for encoding occluded areas to increse the PSNR. One way to overcome this problem is for occlusion areas to adopt arbitrary shape based transform technique which takes advantages of the segmentation. Usually, however, the occlusions do not affect on perceiving 3D structure, i.e, even though we do not send any side information for occlusion areas, we can perceive 3D from stereo images. Therefore, another way to reduce this kind of overhead is to interpolate occluded areas from their neighboring blocks. The remaining factor we have to consider to reduce bit rates is using boundary approximation techniques.

5 CONCLUSION AND FUTURE DIRECTIONS

In this paper, we have proposed an algorithm for stereo image coding based on MRF model and disparity field segmentation. The MRF based method provides a reasonable solution in the Rate-Distortion (R-D) sense since the solution is selected by tradeoffs between compensated error and smooth disparity resulting in a lower bit rate for the disparity. It also encourages smoothly connected disparity contours corresponding to effective representation of segments. According to the results, the proposed algorithm also provides better visual quality as well as PSNR. We will apply the algorithm onto the more complicate real multi-view images and stereo video sequences. We also will consider R-D constrained segmentation and an efficient encoding of the resulting contours and compensated error with given bit budget. The boundary approximation based on the R-D constraints can be another issue to solve.

6 REFERENCES

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

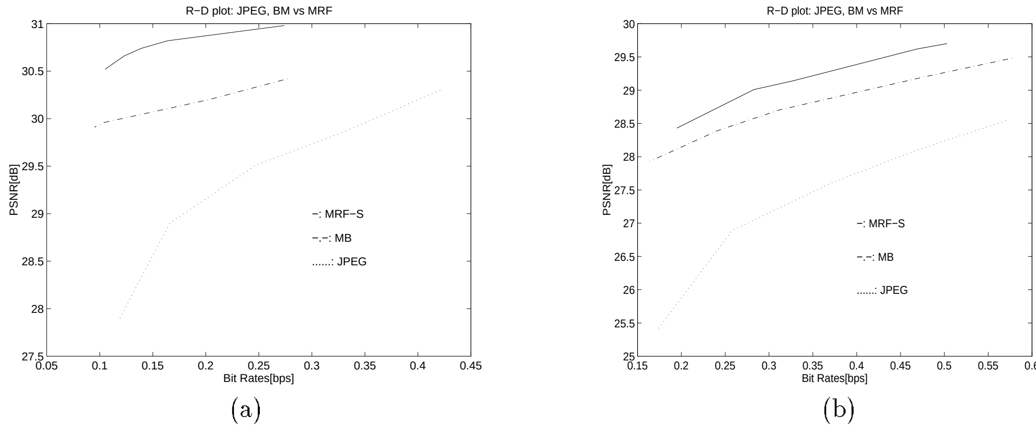

Figure 6: R-D Plot of Reconstructed Target Image in Stereo Image Coding; JPEG, Block based disparity compensation coding, and MRF and segmentation based coding. (a) Synthesized image (b) Bust image

[2] M. T. Orchard. Predictive motion-field segmentation for image sequence coding. *IEEE Tr. on Circuits and Video Technology*, 3(1):54–70, February 1993.

[3] M. Kunt, A. Ikonomopoulos, and M. Kocher. Second-generation image coding techniques. *Proceedings of the IEEE*, 73(4):549–574, April 1985.

[4] C. Chu and K. Aggarawal. The integration of image segmentation maps using region and edge information. *IEEE Tr. on PAMI*, 15(12):1241–1252, December 1993.

[5] F. Marques, M. Pardas, and P. Salembier. *Coding-Oriented Segmentation of Video Sequences*. In Video Coding: The second Generation Approach, 1996.

[6] P.J.L van Beek. *Edge-based Image Representation and Coding*. Thesis Technische University Delft, 1995.

[7] G. M. Schuster and A. K. Katsaggelos. Anefficient boundary encoding scheme which is optimal in the rate distortion sense. In *IEEE Proc. on ICIP 96*, pages 77–80, September 1996.

[8] W. Woo and A. Ortega. Stereo image compression based on the disparity compensation using the mrf model. In *SPIE Proc. on VCIP 96*, volume 2727, pages 28–41, March 1996.

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

[10] 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.

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

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

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