Stereo Image Coding Using Hierarchical MRF Model and Selective Overlapped Block Disparity Compensation

Woontack Woo[†], Antonio Ortega[‡] and Yuichi Iwadate[†]

Media Integration & Communications Labs[†]
Advanced Telecommunications Research
Kyoto 619-0288, Japan
E-mail:{wwoo,yiwadate}@mic.atr.co.jp

Integrated Media Systems Center, EE-Systems[‡]
University of Southern California
Los Angeles, CA 90089, USA
ortega@sipi.usc.edu

Abstract

In this paper, we propose a novel hierarchical disparity estimation/compensation (DE/DC) algorithm for stereo image coding. One way to limit the well-known drawbacks of block matching (e.g., inaccurate disparity and blocking artifacts in the decoded image) is to resort to variable size block matching (VSBM). However, VSMB may result in an inconsistent disparity estimation (i.e., such that the estimated disparity field does not correspond to the true disparity) especially as the subblock becomes small, thus leading to high frequency energy at block boundaries in the residue image. To address these problems, in this paper we propose a hybrid quadtree-based DE/DC scheme, where a Markov Random Field (MRF) model is used in combination with VSBM and selective overlapped disparity compensation to improve the disparity field consistency and lead to higher coding efficiency. Our experimental results demonstrate that the proposed block segmentation scheme achieves a higher PSNR and a more consistent disparity field, as compared to conventional VSBM schemes.

1 Introduction

As for motion compensation, a well-known drawback of fixed block-size disparity compensation schemes is that they assume that disparity can be well represented by using a single vector per block, and that the block size is the same for every vector. A more accurate disparity estimation can be achieved instead by using variable size block matching (VSBM), which can use larger blocks in homogeneous areas (background or inside of an object) and smaller blocks in object boundary areas.

However, VSBM may also have some disadvantages. In most previous approaches, block segmentation is determined by various intuitive criteria such as thresholds, features, local variance of the disparity compensated difference (DCD) and/or local motion activities [1–3]. If the block size is determined by simple thresholds it may not be possible to control efficiently the number of bits used for segmentation. In addition, even in rate-distortion (RD) based approaches, the consistency of the disparity field cannot be guaranteed because the criteria are based only on *local* measures (i.e., mean square error (MSE), mean absolute error (MAE) or Lagrangian RD costs based on these metrics [1].) Therefore, while the segmented subblocks may represent the local minimum for energy in the residue image, they may not correspond to a consistent disparity field. This affects performance in three ways. First, it means that the rate required to transmit the disparity field may be higher. Second, the resulting DCD frame may contain high frequency components along subblock boundaries, and this will be particularly harmful if small blocks are used (e.g., smaller than the block size used for DCT coding). Finally, if generation of intermediate images from the stereo pair is required, the lack of consistent disparity may result in poor quality intermediate images.

In this paper, we propose a novel hierarchical VSBM (H-VSBM) scheme that addresses the above mentioned weaknesses of conventional VSBM-based schemes. The efficiency of the proposed H-VSBM is provided mainly by three added components, which allow more consistent DE/DC, efficient representation of the segmented DV field and less blocking artifacts along the object boundaries. First, *DE with a hierarchical MRF model* allows estimating a relatively smoother DV field by considering blocks in the upper levels of the tree, as well as neighboring blocks, while maintaining encoding efficiency of DV field. Note that obtaining a smooth DV field is beneficial as it reduces the rate for the DV field itself. Then, *quadtree pruning based on the simplified RD cost* improves the encoding efficiency. Finally, *selective overlapped block disparity compensation (OBDC) for the segmented subblocks* improves the encoding efficiency by reducing blocking artifacts within segmented blocks, using overlapped windows with different sizes and shapes for the segmented subblocks [4]. For selective OBDC, no side information is required because a quadtree already contains block segmentation information.

The paper is organized as follows. In Section 2, we briefly explain each step of the proposed hierarchical VSBM scheme. The entire procedure of the proposed scheme is described in Section 3. In Section 4, we provide some experimental results to compare the effectiveness of the proposed scheme over conventional VSBM.

2 Hierarchical VSBM

2.1 Disparity Estimation

To estimate a smooth DV field, we consider the DV field, V, as an MRF model and then formulate the DE problem in a manner similar to the one we introduced in [4–6]. The main difference is that the neighborhood we define here is based on a hierarchical quadtree block decomposition. Note that this is an iterative procedure where MRF-based DE and RD-based segmentation are alternatively used. In the first iteration the block segmentation is considered according to MSE rather than RD cost.

Let F_1 and F_2 represent the reference image and the target (or dependent) image in a pair of stereo images, respectively. First, the target image, F_2, is segmented into square blocks, *i.e.* $F_2 = \{b_{2,i}, 0 \leq i < N\}$, where N denotes the number of blocks. Then, we define an a set of segmentation indicators $\Phi = \{\phi_i\}$. If the i-th block has a higher estimation error, then we set $\phi_i = 1$ and the block is segmented into subblocks to improve the overall encoding efficiency. Let b_i^l be the block i at level l and $b_i^{l+1,k}$ be its children, where $k \in \{0,1,2,3\}$ for a quadtree-based segmentation. Then we slightly change the equation in [4] to estimate the disparity field, V, for segmented subblocks. The changed equation is defined as follows,

$$U(V|F_1,F_2) = \sum_{i \in N} \{(1-\phi_i) \cdot u_i + \phi_i \cdot u_i^*\}, \qquad (1)$$

where

$$\begin{aligned} u_i &= (1-\alpha)\|b_{2,i} - b_{1,i\oplus v_i}\|^2 \\ &+ \alpha \sum_{\eta} (1-\phi_i^{\eta})(v_i - v_i^{\eta})^2\}, \\ u_i^* &= \sum_{l+1,k} \{(1-\alpha) \cdot \|b_{2,i}^{l+1,k} - b^{l+1,k}_{1,i\oplus v_i^{l+1,k}}\|^2 \\ &+ \alpha\{\sum_{\eta^{l+1,k}} (1-\phi_i^{l+1,\eta^k})(v_i^{l+1,k} - v_i^{l+1,\eta^k})^2 \\ &\quad +(v_i^{l+1,k} - v_i^l)^2\}\} \end{aligned}$$

where v_i^{lk} and ϕ_i^{lk} denote the disparity vector and the segmentation status of b_i^{lk}. The superscript η represents predefined neighborhoods in the DV field. The degree of smoothness can be controlled using a weighting constant $(0 \leq \alpha \leq 1)$.

In (1), as explained, either u_i or u_i^* is used for the DE. For example, the MRF-based DE is performed for the block i using u_i, if the MSE of the block i is smaller than the threshold. Otherwise, the hierarchical MRF-based DE is performed for the segmented subblocks using u_i^* and then the block segmentation is determined by comparing encoding costs between the block b_i^l and its subblocks b_i^{l+1}.

2.2 Block Segmentation

We extend quadtree-based approaches mainly adopted in video coding [2,7], to stereo image coding. In case of a quadtree, the original block is segmented into four square subblocks and then the segmentation is repeatedly applied to the subblocks until a stop-condition is satisfied, *e.g.* the subblock size reaches 1×1. In this method only one bit is required to represent whether the block is segmented or not. Then, the segmentation of a block can be represented using a series of bits that represent termination with a "0" (a leaf node) and continuation with an "1" (child nodes). Similarly, the entire segmentation of an image can be represented by a set of quadtrees.

Let c_i^l and c_i^{l+1} be encoding costs of b_i^l and $\sum_k b_i^{l+1,k}$, respectively. Then the Lagrangian cost c_i^l and c_i^{l+1} can be represented as follows,

$$\begin{aligned} c_i^l &= d_i^l + \lambda r_i^l \\ c_i^{l+1} &= \sum_k \{d_i^{l+1,k} + \lambda r_i^{l+1,k}\} \end{aligned} \qquad (2)$$

where d_i^l and r_i^l denote the distortion and the rate of the block, b_i^l, respectively. The Lagrange multiplier $(\lambda \geq 0)$ controls the weight between d and r. The block segmentation is performed, only if the segmentation reduces the Lagrangian cost in (2), *i.e.*, $c_i^l > c_i^{l+1}$.

The decision of block segmentation can be made by comparing the encoding costs of the block and its consecutive subblocks. To achieve an efficient hierarchical segmentation, we simplify the blockwise segmentation cost as follows. The rate r^l in (2) is defined as a sum of bit rates, $r(v^l) + r(dcd^l)$, where v and dcd denote disparity vectors and the resulting DCD of the block. To support a simple decision, we assume that fixed length coding is used, *e.g.* $r(v^l) = 4 \times r(v^{l+1})$, and the rate $r(dcd^l)$ can be ignored based on the observation that the role of $r(dcd)$ can be replaced by the distortion $d(dcd)$. Note also that the energy of the DCD, rather than the MSE in the quantized DCD, is used to measure the distortion, $d(dcd)$.

2.3 Overlapped Block Compensation

In video coding, overlapped block matching (OBM) has been introduced as a promising way to reduce blocking artifacts by using multiple vectors for a block, while maintaining the bit rate of the motion vector fields. In [4], we proposed a modified OBDC scheme for FSBM-based stereo image coding and showed that the scheme can reduce blocking artifacts as well as computational complexity as compared to conventional OBM schemes proposed in video coding. However, FSBM-based OBDC scheme has limitation in reducing estimation/compensation error along the object boundaries. Therefore, in this paper, we extend the FSBM-based OBDC scheme in [4] to VSBM-based OBDC and only apply the OBM scheme to the segmented subblocks to reduce the blocking artifact along segmented block.

As explained in [4], the selective OBDC works because OBDC is efficient only when the energy level of the DCD block is significantly different from that of its neighboring blocks or when high frequency components exist in the DCD block. Note that, for blocks with $\phi_i = 0$, we assume that the block can be compensated effectively with a flat window, *i.e.* without OBDC. Then, for blocks with $\phi_i = 1$, we partially apply OBM windows to compensate the subblocks, according to the reliability of the corresponding vector. Otherwise, we encode the block in intra mode, instead of the DCD block. Note also that the adaptively changing window shape prevents oversmoothing.

3 Encoding Procedure

First, the F_1 is encoded in intraframe mode using transform methods such as DCT or wavelet transform. Then the formal procedure of the proposed scheme for the target image is as follows.

- **Step 0** (Initialization) F_2 is segmented into initial blocks with size of 32×32 (or 16×16). Set the segmentation level l to be 0.
- **Step 1** (Initial Disparity Estimation) For each block in F_2, the best matching block is estimated in F_1 within the search window using (1). Go to next block, if the resulting MSE of the DCD is smaller than a threshold, T_ϕ. Otherwise, set the segmentation indicator $\phi_i = 1$ and go to step 2 for further segmentation.
- **Step 2** (Disparity Estimation in Subblocks) At $l = l + 1$, for each subblock, a DV is estimated. In this process, a smooth DV field is estimated by removing outliers in the DV fields using the MRF-based cost function in (1). If the DV's are the same as those generated in the upper level, we do not need to consider further segmentation. Otherwise, go to step 3.
- **Step 3** (Block Segmentation) Using 2 the encoding costs are compared to determine whether the block should be segmented. The block is segmented only if the coding cost of the subblocks is lower than that of the block, *i.e.* $c^{l+1} < c^l$.
- **Step 4** Segmentation and DE are repeatedly applied until the coding cost of the block is smaller than that of segmented subblocks or the block size is reached the preselected size, *e.g.* 1×1.
- **Step 5** (Selective OBDC) For the resulting DV field, based on the quadtree, OBDC is selectively performed for the blocks with $\phi = 1$. If there is no gain from OBDC, the original block is encoded instead of the DCD.

The resulting quadtree, the DV field and the DCD frame are encoded for subsequent transmission or storage. The DV vectors of subblocks are ordered in a Z shape. The DVs are then ordered in row by row and later encoded using DPCM. After selective OBDC, the smoothed DCD frame is encoded using JPEG and then stored or transmitted to improve the overall quality of the decoded image. Notice that no side information is needed for the selective OBDC since the information is already contained in the quadtree. The decoding is the inverse of the encoding process. At the decoder, the reference image is first decoded. Afterwards, the target image is reconstructed using the reference image and the side information such as disparity and compensated error.

4 Experimental Results

To show the effectiveness of the proposed scheme we test our algorithm on a synthesized image pair, *Room*, and a natural image pair, *Aqua* as shown in Figures 1 and 2, respectively. The search window for DE is $(\pm 2, \pm 16)$. Then, the results are compared with those of the MSE-based VSBM (in this case simple MSE thresholding is used for segmentation and neither OBM, MRF-based estimation or RD segmentation are used.) The performance is measured in terms of the bit rate and the peak signal to noise ratio (PSNR) of the target image.

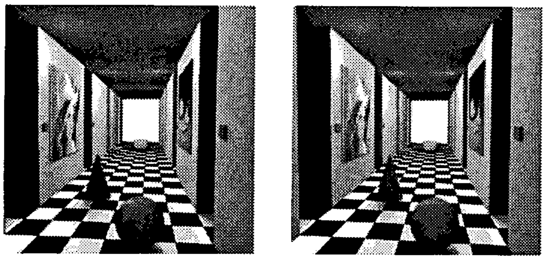

(a) (b)

Figure 1: Test stereo images *Room.* (a) left image (b) right image.

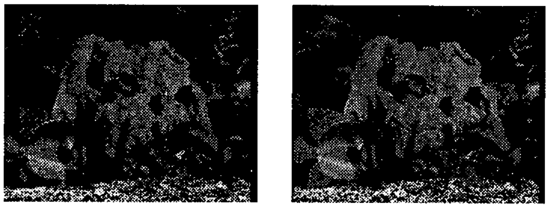
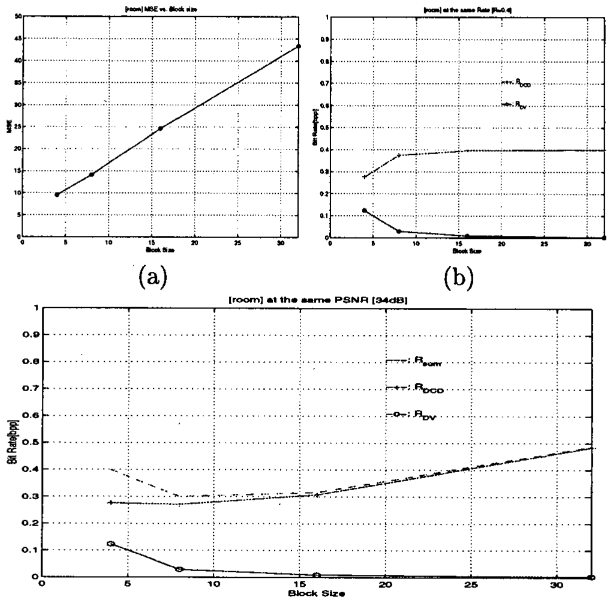

(a) (b)

Figure 2: Test stereo images *Aqua.* (a) left image (b) right image.

We first investigate the effects of block size in terms of MSE and bit rate. As shown in Figure 3, the MSE between the disparity estimated block and the original block is linearly decreased, as the block size is reduced for DE. However, as shown in (b), the bit rate required to transmit the DV field increases. Another noteworthy observation is that, as shown in (c), as the block size of DE becomes smaller than that of DCT, decreasing block size does not always increase the encoding efficiency, due to block boundary effects of smaller blocks, even though the smaller block decreases the MSE of the DCD frame. In our experiments DCT is performed on the 8 × 8 block and thus the DCD resulting from FSBM with 4 × 4 block may not be efficient in the RD sense, which is one of the main motivations of introducing OBDC for the segmented subblocks.



(a) (b)

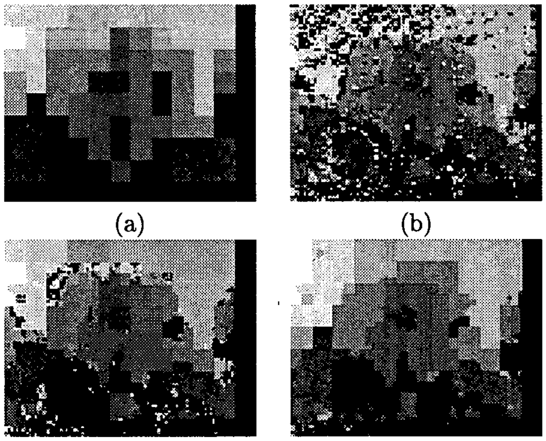
(c)

Figure 3: Effects of block size (Room). (a) MSE vs. block size. (b) Bit rate vs. block size rate (at the same rate, 0.4 [bpp]). (c) Bit rate vs. block size (at the same PSNR, 34dB).

Figure 4 compares the disparity estimation results of the proposed H-VSBM scheme versus those of MSE-based VSBM. Figure 4 (a) and (b) show DV fields based on the FSBM. The resulting DV field of FSBM has a blocky appearance. The errors occur, as expected, along the object boundaries for the FSBM, where a different estimate for object and background is needed. The disparity field in Figure 4 (c) and (d) show DV fields based on the VSBM. As shown in (c), simple MSE-based VSBM shows inconsistencies in the DV field, as the subblock size becomes small. Meanwhile, as shown in (d), MRF-based DE in H-VSBM allows a consistent DV field estimation, which reduces the bit rate for the disparity field itself, while reducing the disparity compensated error.

Figure 5 (a) and (b) show the corresponding RD plots of both target images in the pairs of stereo images. In our experiments, we fix the reference image and only measure the RD performance for the target image in the stereo pairs. The DV field and quadtree are encoded using DPCM and the resulting DCD frame is encoded using JPEG.

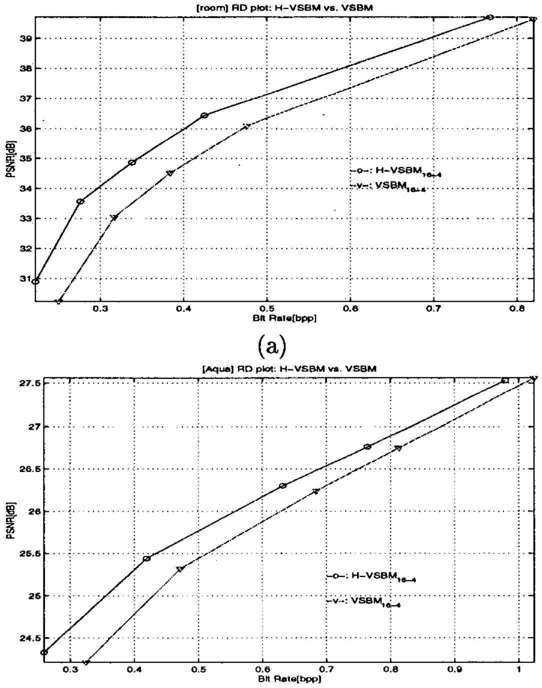
According to our experimental results, the proposed H-VSBM method improved overall encoding performance for the target image in a stereo pair. The



(a) (b)

(c) (d)

Figure 4: Results of the disparity estimation for the natural image, *Aqua*. (a) DV field with the FSBM (32 × 32) (b) DV field with the FSBM (4 × 4) (c) DV field with the MSE-based VSBM (32 × 32 to 4 × 4) (d) DV field with the proposed H-VSBM (32 × 32 to 4 × 4)



(a)

(b)

Figure 5: R-D Plot of reconstructed target images in the stereo pairs. (a) *Room* (b) *Aqua*.

DE with smoothness constraint estimates relatively accurate and more consistent DV fields, which results in less bit rate for the DV. The encoding efficiency has been further improved by selectively applying OBDC for blocks with higher energy levels of estimation errors, which significantly reduces the compensation error along the object boundaries. In terms of PSNR, the proposed H-VSBM scheme achieved 0.5-1.5 *dB* higher PSNR, as compared to a simple MSE-based VSBM. In addition, the reconstructed image based on the proposed scheme resulted in fewer annoying blocking artifacts, by reducing the blocking errors along the subblocks using OBDC.

This research will be extended to the intermediate view generation (or synthesis) to provide the look-around-capability at the decoder. The RD-based contour representation of the segmented region remains a crucial future work because the RD-based shape coding will be an important technique for object-based coding, providing various functionalities such as interactive manipulation of objects.

References

[1] G. J. Sullivan and R. L. Baker, "Efficient quadtree of images and video," *IEEE Trans. on IP*, vol. 3, no. 3, pp. 327–331, May 1994.

[2] J. Lee, "Optimal quadtree for variable block size motion estimation," in *Proc. IEEE ICIP*, Oct. 1995, vol. 3, pp. 480–483.

[3] H. Bi and W. Chan, "Rate-Distortion optimization of hierarchical displacement fields," *IEEE Trans. on CSVT*, vol. 8, no. 1, pp. 18–24, Feb. 1998.

[4] W. Woo and A. Ortega, "Modified overlapped block matching for stereo image coding," in *Proc. SPIE EI-VCIP*, Jan. 1999, vol. 3653.

[5] W. Woo and A. Ortega, "Stereo image compression based on the disparity compensation using the MRF model," in *Proc. SPIE VCIP*, Mar. 1996, vol. 2727, pp. 28–41.

[6] W. Woo and A. Ortega, "Stereo image compression based on the disparity field segmentation," in *Proc. SPIE EI-VCIP*, Feb. 1997, vol. 3024, pp. 391–402.

[7] M.R. Banham, J.C. Brailean, C.L. Chan, and A.K. Katsaggelos, "Low bit rate video coding using robust motion vector regeneration in the decoder," *IEEE Trans. on IP*, vol. 3, no. 5, pp. 652–665, Sept. 1994.