HIERARCHICAL DISPARITY ESTIMATION USING SPATIAL CORRELATION

Chia-Wen Lin
Computer and Communication Research Labs
Industrial Technology Research Institute
Hsinchu, Taiwan 310, ROC

Er-Yin Fei and Yung-Chang Chen
Department of Electrical Engineering
National Tsing Hua University
Hsinchu, Taiwan 300, ROC

ABSTRACT

In this paper, we propose a fast hierarchical algorithm for disparity estimation, which is based on the multiresolution block matching technique with prospective disparity vectors. The disparity vectors of each template block's spatially adjacent blocks are adopted as the initial references for the template block to speed up the estimation process. The experimental results show that the proposed hierarchical estimation algorithm can reduce computation cost drastically while maintaining comparable PSNR quality. The proposed scheme also results in smoother disparity vector field compared to the exhaustive search scheme. A rate-distortion optimized hierarchical disparity estimation algorithm is also proposed to minimize the distortion of the reconstructed image sequence under a target bit rate constraint. The experimental results show that its PSNR quality is better than the full search algorithm while taking more computing power for the optimization process.

1. INTRODUCTION

Disparity between two images is defined as the differences which occur between two simultaneous images shot by a stereoscopic vision system which may be induced by the relative motion of the camera and the scene, by the relative displacement of two cameras, or by the motion of objects in the scene. Disparity estimation may be broadly defined as the determination of the geometric differences between two or more simultaneous images of the same scene. The differences may be the result of binocular parallax, motion parallax, object motion, or some combination of the above. That is, disparity provides a way of determining the spatial relationship between points and surfaces in a scene. The disparity vector field that is derived during the disparity estimation process can be used to predict one image of the stereoscopic image pair from the other. So the disparity vector field and the reconstruction errors have to be transmitted to the decoder in a stereoscopic image sequence coding system.

Disparity compensation is a powerful tool for stereoscopic image/video coding. Disparity estimation is, however, a very time consuming operation. Since most processing time of stereoscopic image sequence coding is spent on the disparity estimation process, it is necessary to simplify the estimation algorithm without seriously affecting the quality of the reconstructed images, such as the hierarchical block matching (HBM) methods for both monocular and stereoscopic image sequence coding proposed in the literatures [1]. These methods are seen to drastically reduce the amount of computation needed for block matching and the quality of the reconstructed images is acceptable. The proposed method of motion and disparity estimation is based on the multiresolution block matching technique [1] and utilizing minimum absolute difference (MAD) as the matching criterion, the MAD of two $N \times N$ blocks X and Y is defined as

$$\text{MAD}(X,Y) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} |X(i,j) - Y(i,j)| \qquad (1)$$

In the following, if $[i, j]$ are the coordinates of the upper left-hand corner of a block, the block is referred to as block $[i, j]$. For expanding multiresolution images, we utilize the Laplacian pyramid [3] to construct the L-level hierarchical image structure, which involves low-pass filtering and downsampling the image with a factor of 2 as shown in Fig. 1

$$block_{\,l+1}[i, j] = \sum_{m=-2}^{2} \sum_{n=-2}^{2} w[m][n]block_{\,l}[2i + m, 2j + n]$$

(2)

Where $w[m][n] = w[m]w[n]$, and $w[0] = a$, $w[1] = w[-1] = 1/4$, $w[2] = w[-2] = 1/4 - a/2$.
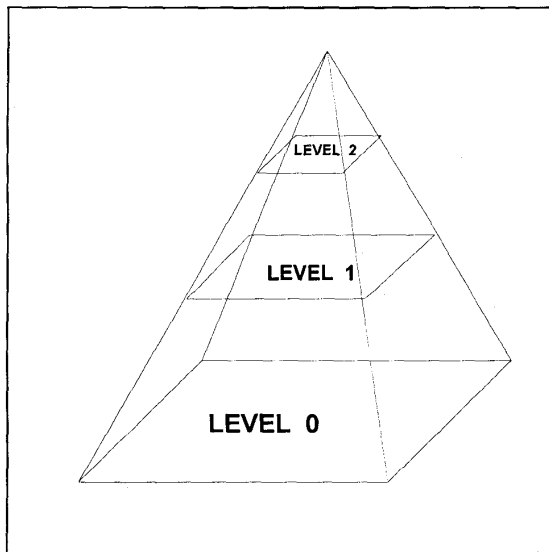

Fig. 1. A 3-level pyramid image data structure

Block-wise structure is widely used for the currently existing coding techniques and for disparity estimation. For the block-based techniques of disparity estimation, it is important to choose a proper block size. Using large-size blocks will lead to inaccurate disparity estimation results. On the other hand, small block size decreases the reliability of disparity estimation. Using hierarchical block matching (HBM) method [1] can solve this problem. There are two ways for hierarchical block matching

method, that is, variable block-size matching and constant block-size matching. The realization of the two HBM methods is as follows.

(i) Variable block-size matching : The size of the block for matching varies with the resolution level : at level N it is $b_x \times b_y$; then at level N-1 the block size becomes $2\,b_x \times 2\,b_y$. If v_N (i, j) is the disparity vector for the block $[i, j]$ at level N of the pyramid, the initial estimation v^i_{N-1} (i, j) of the vector for the same block at level N-1 of the pyramid is :

$$v^i_{N-1} \ (i, j) \ = 2v_N \ (i/2, j/2) \quad (3)$$

(ii) Constant block-size matching : A constant block size of $b_x \times b_y$ is used in all levels of the pyramid. The difference from the variable block-size method is that one block at level N of the pyramid corresponds to 4 blocks at level N-1 of the pyramid. That is,

$$v^i_{N-1} \ (i, j) \ = 2v_N \ (i/2, j/2) \quad (4)$$

$$v^i_{N-1} \ (i + bx, j) \ = 2\,v_N \ (i/2, j/2) \quad (5)$$

$$v^i_{N-1} \ (i, j + by) \ = 2\,v_N \ (i/2, j/2) \quad (6)$$

$$v^i_{N-1} \ (i + bx, j + by) \ = 2\,v_N \ (i/2, j/2) \quad (7)$$

Due to the epipolar constraint, the search range of disparity estimation in horizontal direction is much larger than in vertical direction. For lacking of epipolar constraint, the search range of motion estimation in horizontal direction is equal to vertical one. The search range for the 2-level HBM for disparity estimation is specified in Table.1, where the full pixel precision is used in level 1 and the half-pixel precision is used in level 0. Reconstruction of the total vector field is performed by using:

$$v_{total} = 2^{L-1} v_{L-1} + \sum_{P=0}^{L-2} 2^P dv_P \qquad (8)$$

where L is the total level number of the pyramid and dv_P is the difference value between the estimated vector and the initial estimation vector.

	Disparity search range	
Level	0	1
X	-2.5 ~ 2.5	-7 ~ 7
Y	-1.5 ~ 1.5	-1 ~ 1

Table.1. Disparity estimation search range when using 2-level pyramid

2. HIRARCHICAL DISPARITY ESTIMATION WITH PREDICTORS

In the stereoscopic image sequence coding system, all implemented motion estimation algorithms can be directly applied for disparity estimation. But in the strict sense, motion is temporal, and disparity is spatial. By considering the extended continuity constraint of disparity, we can use the adjacent estimated disparity vectors as the predictors for the current disparity vector in the disparity estimation process. This is under the assumption of a large object with a smooth surface shown in the stereoscopic image pair. In the proposed method, we utilize the three adjacent estimated vectors as the predictors to estimate the current vector as depicted in Fig. 2.
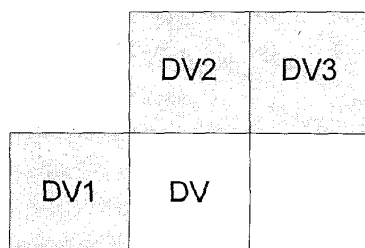

Fig. 2. DV: Current displacement vector; DV1, DV2, DV3: Predictors

The predictors are only utilized at the highest level of the pyramid. The one with the least MAD among the three predictors is taken as the initial estimation of the current vector with a rather small search area. For the blocks in the edges of the image, the number of predictors may be one or two even zero. So it is necessary to search in a larger area for these edge blocks to avoid error propagation.

The main effect of predictors is significantly reducing the computation load in the estimation process, but degrading the quality of the estimation in an acceptable range. In addition, the whole displacement vector field is becoming smoother and it brings benefits to the segmentation process.

3. DISPARITY ESTIMATION WITH RATE-DISTORTION OPTIMIZATION

A rate-distortion framework can be used to define a displacement (disparity or motion) vector field estimation (DVFE) technique for the image sequence coding. This technique achieves maximum reconstructed image quality under the constraint of a target bit rate for the coding of the displacement vector field. The main purpose is to find a compromise between minimizing the bit rate of the displacement vector filed and minimizing the displaced frame difference (DFD) between temporally or spatially adjacent frames. The effect of rate-distortion optimization would be more obvious when dealing with the low bit rate image sequence coding, since the number of bits needed for encoding the displacement vector field is rather small in comparison to that for encoding the texture of the image.

In the proposed method of DVFE using rate-distortion optimization, the rate part is composed of two components: one is the bit rate for transmitting the displacement vector field and

the other is for coding the error image. On the other hand, the distortion part is determined by means of DFD. The target rate constrained optimization problem is transformed into an equivalent unconstrained problem by merging the rate and distortion parts through the Lagrange multiplier [11-13]. The following is the detailed description of the proposed method.

3.1 Rate-distortion formulation for Disparity Estimation

Let $v_i \in V$ be the disparity vector corresponding to the block i of the image, where V is the set of all disparity vectors determined by the hierarchical block matching algorithm mentioned in Section 2. The purpose of the DVFE algorithm is to minimize the distortion D of the reconstructed image sequence, under the constraint of the target rate R_{target} for transmitting the disparity vector field and the error image. This corresponds to the following constrained optimization problem:

$$\min_{v_i \in V} \sum_{i=1}^{N} D(v_i) \quad (9)$$

subject to

$$\sum_{i=1}^{N} R_i(v_i) \leq R_{target} \quad (10)$$

where N is the total number of blocks in the image, $D(v_i)$ is the contribution of v_i to the overall distortion, and $R(v_i)$ is the contribution of the vector v_i to the total rate.

From the methodology shown in [14], the above problem can be transformed into an unconstrained optimization problem by adopting the Lagrange multiplier λ. Thus the solution $\{ v_i^*(\lambda), i = 1, \ldots, N \}$ of the unconstrained minimization of the cost function $C(\lambda)$:

$$C(\lambda) = \sum_{i=1}^{N} C_i = D(\lambda) + \lambda R(\lambda)$$

$$= \sum_{i=1}^{N} D[v_i(\lambda)] + \lambda \sum_{i=1}^{N} R[v_i(\lambda)] \quad (11)$$

is also a solution of (9) if:

$$R_{target} = \sum_{i=1}^{N} R[v_i^*(\lambda)] \quad (12)$$

It was shown in [11] that $D(\lambda)$ and $R(\lambda)$ are monotonic functions of the Lagrange multiplier λ, with values ranging from zero (highest rate, lowest distortion) to ∞ (lowest rate, highest distortion). As shown in [12], the curve of $D(\lambda)$ vs. $R(\lambda)$ is not continuous, that is, there exist jumps in some intervals of value (See Fig. 3.). A value of λ corresponds to a (R, D) point. Since the relationship between $D(\lambda)$ and $R(\lambda)$ is not only one-to-one, all we have to do is to find a optimal Lagrange multiplier λ^* which makes $R(\lambda^{**})$ close to R_{target}. The corresponding solution $\{ v_i^*(\lambda^*), i = 1, \ldots, N \}$ is the optimal displacement vector field under the target rate constraint.
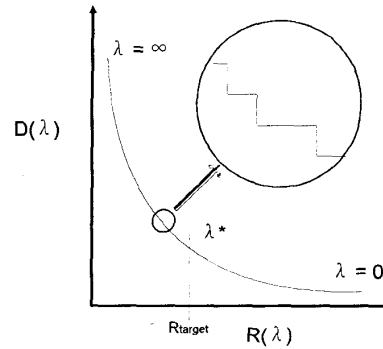

Fig. 3. Curve of $D(\lambda)$ versus $R(\lambda)$

3.2 Optimization Using Convex Hull Bisection Searching Algorithm

Each (R, D) point represents a DVFE result. The optimal DVFE result will lie on the surface of convex hull of all (R, D) points. Thus, the optimal Lagrange multiplier λ^* can be traced on the surface of the convex hull for a given rate or distortion constraint [15]. A fast method called convex hull bisection searching algorithm has been introduced in [12, 15]. The algorithm can efficiently find the optimal Lagrange multiplier λ^* and the optimal vector estimation result on the surface of convex hull of all (R, D) points by decreasing the interval of possible λ in an iteration framework

Firstly, we pick two value of λ, λ_0 and λ_1, such that their corresponding displacement vector fields, $V^*_{\lambda_0}$ and $V^*_{\lambda_1}$, satisfy $R(V^*_{\lambda_0}) > R(V^*_{\lambda_1})$ and $D(V^*_{\lambda_0}) < D(V^*_{\lambda_1})$. The next value of λ that we want to test by iteration is:

$$\lambda_{new} = \frac{D(V^*_{\lambda_1}) - D(V^*_{\lambda_0})}{R(V^*_{\lambda_0}) - R(V^*_{\lambda_1})} + \varepsilon \quad (13)$$

where ε is a vanishingly small positive number. Then a new search is needed and the search interval of possible λ is narrowed by replacing λ_0 or λ_1 with λ_{new}. If no new displacement vector field exists on the surface of the convex hull between λ_0 and λ_1, then the displacement vector field for λ_{new} is the optimal solution. We can reduce the iteration times if the difference between $R(V^*_{\lambda_{new}})$ and R_{target} is below a threshold.

A number of distortion measurements have been proposed in the literatures. The simplest and the most commonly used method is the displaced frame difference (DFD):

$$D(v_i) = \sum_{k=0}^{bx} \sum_{l=0}^{by} \left| im_1(m+k, n+l) - im_2(m+k+v_{i,x}, n+l+v_{i,y}) \right|$$
$$(14)$$

where im_1 and im_2 are temporal or spatial adjacent images, (m, n) is the upper left-hand corner coordinate of block i, $(v_{i,x}, v_{i,y})$ is the motion or disparity vector of block i, and bx, by are the dimensions of the block.
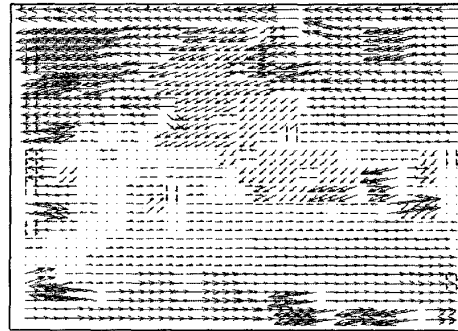
4. EXPERIMENTAL RESULTS AND CONCLUSIONS

In the following experiments, we take $L = 2$ and the block size is 8x8. Four methods are compared in the experiments: exhaustive search, hierarchical block matching without predictors (HBM), hierarchical block matching with predictors (HBMWP), and HBM with rate-distortion optimization (RDO). The test sequence used is "Tunnel" as shown in Fig. 4. Figs. 5(a)-(d) illustrate the disparity vector fields corresponding to the four different methods. It's evident that the HBMWP method results in smoother disparity vector field distribution than the exhaustive search and the HBM method. The rate-distortion optimized HBM (RDO) not only smoothes the vector filed distribution but also tends to minimize the difference of the disparity vector of each block with its adjacent blocks to obtain best rate-distortion trade-off. Figs. 6(a)-(d) show the depth maps resulted from the four methods.
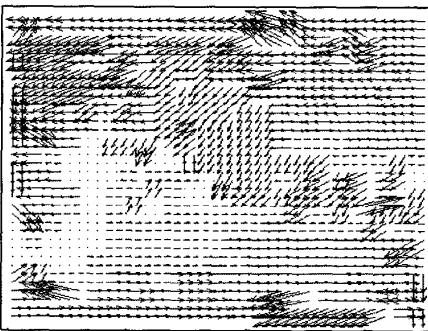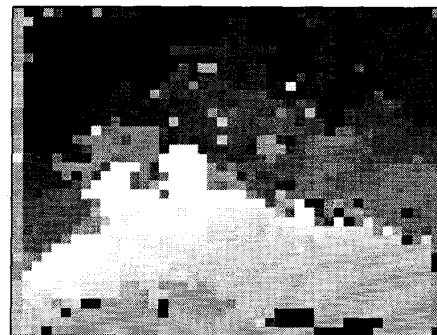

Fig.4. Test Sequence: 'Tunnel'
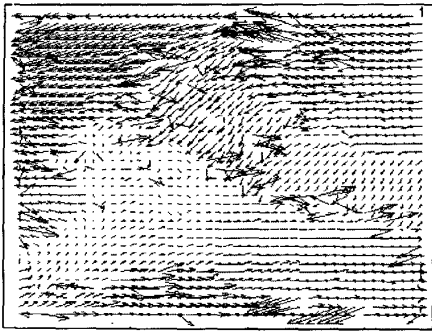
Fig. 5(a). Disparity vector field using exhaustive search for sequence "Tunnel"
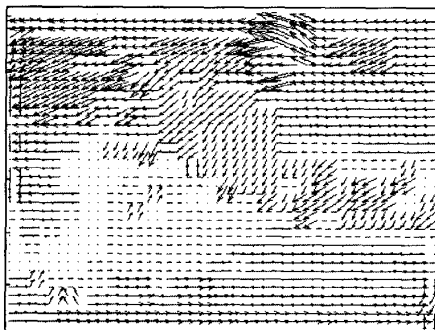

Fig. 5(b). Disparity vector field using HBM for sequence "Tunnel"

Fig. 5(c). Disparity vector field using HBMWP for sequence "Tunnel"
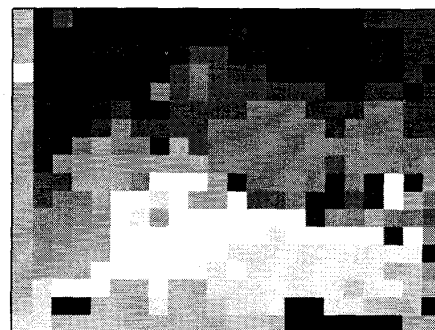

Fig. 5(d). Disparity vector field using RDO for sequence "Tunnel"

Fig. 6(a). Depth map derived from exhaustive Search for sequence "Tunnel"

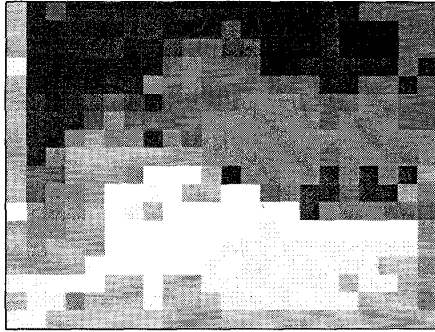**Fig. 6(b).** Depth map derived from HBM for sequence "Tunnel"

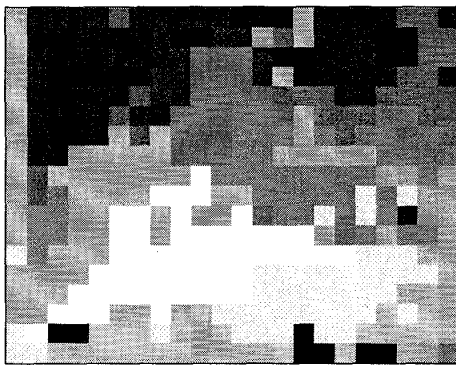**Fig. 6(c).** Depth map derived from HBMWP for sequence "Tunnel"

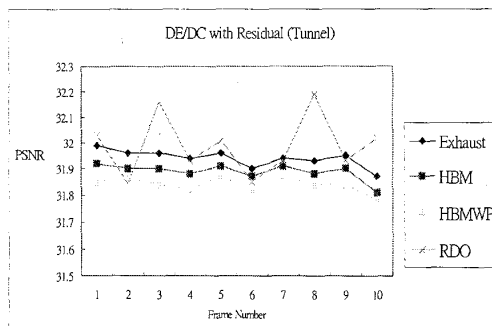**Fig. 6(d).** Depth map derived from RDO for sequence "Tunnel"

Fig. 7. PSNR comparison of the four different disparity estimation methods

The PSNR comparison result of the four disparity estimation methods is shown in Fig. 7. Table 2 shows the comparison results for average PSNR, bits/pixel, and computation time required. The number bracketed specifies the average bits required to transmit the disparity vector and error image (where M stands for disparity vector, and E for error image). The RDO method, though using the hierarchical search algorithm, outperforms all other three schemes in PSNR performance due to its optimization property. The RDO scheme, however, is the most computation intensive algorithm since it needs iterative optimization computation process.

The HPMWP scheme takes minimum computing power while maintaining comparable PSNR performance, thus well suited to the applications with real-time requirement. The rate-distortion optimized hierarchical estimation schemes has high potential in very low bit rate visual communication applications, since it can keep the disparity information as minimum as possible while keeping good video quality. But it needs further investigations to reduce the huge computational complexity required.

	Tunnel		
	PSNR	Bit/Pixel	Time
Exhaust	31.94	0.5387 (M:0.0878, E:0.4509)	100%
HBM	31.89	0.54 (M:0.0666, E:0.4734)	17.69%
HBMWP	31.84	0.5574 (M:0.0637, E:0.4937)	6.98%
RDO	31.99	0.5417 (M:0.0616, E:0.4801)	424.56%

Table.2 Performance comparison of various estimation schemes

5. REFERENCES

[1] D. Tzovaras, M.G. Strintzis, and H. Sahinoglou, "Evaluation of multiresolution block matching techniques for motion and disparity estimation," *Signal Processing:*

Image Communication, vol. 6, no. 1, pp. 59-67, March 1994.

[2] Won-Ho Kim, Jae-Young Ahn, and Sung-Woong Ra, "An efficient disparity estimation algorithm for stereoscopic image compression," *IEEE Trans. Consumer Electronics* pp. 165-171, Mar. 1997.

[3] P.J. Burt and E.H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532-540, April 1983.

[4] A. Tamtaoui and C. Labit, "Constrained disparity and motion estimators for 3DTV image sequence coding," *Signal Processing: Image Communication*, pp. 45-54, Nov. 1991.

[5] C.-H. Lee and L.-H. Chen, "A fast motion estimation algorithm based on block sum pyramid," *IEEE Trans. Image Processing.*, vol. 6, no. 11, pp. 1587-1590, Nov. 1997.

[6] Won-Ho Kim, Jae-Young Ahn, and Sung-Woong Ra, "An efficient disparity estimation algorithm for stereoscopic image compression," *IEEE Trans. Consumer Electronics*, pp.165-171, March 1997.

[7] A. Tamtaoui and C. Labit, "Constrained disparity and motion estimators for 3DTV image sequence coding," *Signal Processing: Image Communication*, vol.4, pp.45-54, November 1991.

[8] Stephen T. Barnard, and William B. Thompson, "Disparity analysis of images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. Pami-2, n0.4, pp.333-340, July 1980.

[9] T. Koga *et al.*, "Motion compensated interframe coding for video confrencing," *Proc. Nat. Telecommunication Conf.*, pp.G5.3.1-G5.3.5, 1981.

[10] M. Bierling, "Displacement estimation by hierarchical block-matching," *Proc. SPIE Conf. Visual Communications and Image Processing*, vol.1001, pp.942-951, 1988.

[11] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers,"

IEEE Trans. Acoustics, Speech, and Signal Processing, vol.36, pp.1445-1453, September 1988.

[12] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol.2, pp.160-175, April 1993.

[13] S. W. Wu and A. Gersho, "Rate-constrained optimal block-adaptive coding for digital tap recording of HDTV," *IEEE Trans. On Circuits and Systems for Video Technology*, vol.1, pp.100-112, March 1991.

[14] H. Everett, "Generalized Lagrange multiplier method for solving problem of optimum allocation of resources," *Oper. Res.*, vol.11, pp.399-417, 1963.

[15] G. J. Sullivan and R. L. Baker, "Efficient quadtree coding of image and video," *IEEE Trans. Image Processing*, vol.3, pp.327-331, May 1994.