Multiresolution Wavelet-Based Disparity Estimation for Stereo Image Compression

Chompoonuch Tengcharoen* and Ruttikorn Varakulsiripunth**

Faculty of Engineering, and Research Center for Communications and Information Technology (ReCCIT),
King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand.
E-mail: chompoonoot@hotmail.com*, kvruttik@kmitl.ac.th**

Abstract: The ordinary stereo image of an object consists of data of left and right views. Therefore, the left and right image pairs have to be transmitted simultaneously in order to display 3-dimentional video at the remote site. However, due to the twice data in comparing with a monoscopic image of the same object, it needs to be compressed for fast transmission and resource saving. Hence, it needs an effective coding algorithm for compressing stereo image. It was found previously that compressing left and right frames independently will achieve the compression ratio lower than compressing by utilizing the spatial redundancy between both frames. Therefore, in this paper, we study the stereo image compression technique based on the multiresolution wavelet transform using varied disparity-block size for estimation and compensation. The size of disparity-block in the stereo pair subbands are scaling on a coarse-to-fine wavelet coefficients strategy. Finally, the reference left image and residual right image after disparity estimation and compensation are coded by using SPIHT coding. The considered method demonstrates good performance in both PSNR measures and visual quality for stereo image.

Keywords: stereo image compression, wavelet transform, disparity estimation, SPIHT coding

1. INTRODUCTION

A stereo image pair consists of two separated views of a 3-dimensional scene captured and recorded by two cameras, corresponding to the left and right eyes of the human visual system. One scene becomes left image and the other becomes right image as shown in Fig.1. The current standard image compression techniques that used to code the two constituent images independently can reduce data storage. Nevertheless, due to a stereo image pair essentially depicts the same scene from two different points of view, independent coding of both images of a stereo pair is redundant. Therefore, the data storage of both images can be reduced more by adding the disparity estimation and compensation techniques into the coding method.

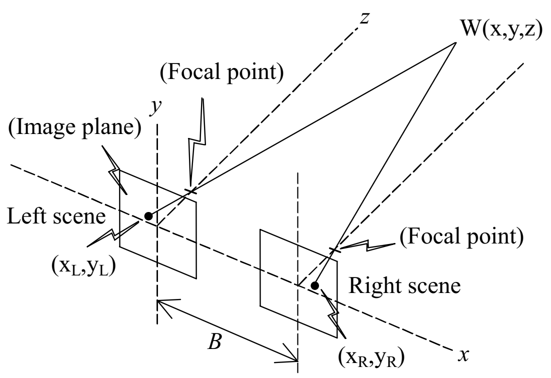


Fig.1 A stereo system.

Up to present, the motion estimation techniques are classified into four main groups [1], i.e., gradient techniques, Pel-recursive techniques, block matching techniques and frequency-domain techniques. Gradient and Pel-recursive techniques are typically used for analysis of image sequences. Frequency-domain techniques are based on the relationship between transformed coefficients of shifted image. The block matching techniques, based on the minimization of a specified cost function, are most widely used in coding applications. From these mentioned techniques, we have observed the quality of motion estimation that uses block-matching technique on the time-frequency domain. We have applied this technique to stereo image compression and studied its performance. Here, the time-frequency domain is obtained from multiresolution wavelet transform.

There are two distinct types of motion compensation. The first one is forward motion compensation and it is the result of coding a frame based on a previous frame in the actual temporal sequence. The second one is symmetrical technique that uses a subsequent frame in the sequence to act as a basis frame for prediction. These two techniques can be used separately or in conjunction with one another. The disparity estimation for monoscopic video can do by spatial disparity of two consecutive frames or by setting one frame as reference for the frame series. The essential of monoscopic video consecutive frame has majority similar background and changes only when the movement occurs. However, the disparity displacement of stereo image, unlike motion displacement, is caused by the slightly different view at the entire left and right image [2]. Thus, the compression technique for stereo image should be slightly different from monoscopic image. Fig.2 shows the block diagram of the procedure of Wavelet-based disparity estimation that used in our performance study in this paper.

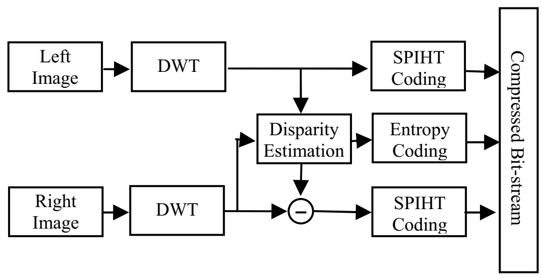


Fig.2 Procedure of Wavelet-based disparity estimation.

. As can see from Fig.2 that the Discrete Wavelet Transform (DWT) is applied to the left and right image at the first step. The DWT's results of both image will passed to disparity estimation. This part will make compensation in order to find the difference between wavelet coefficients of left and right image. As shown on the uper part of Fig.2, the DWT's result of only left image will be encoded by Set Partitioning In Hierarchical Trees (SPIHT) coding. Here, the wavelet coefficients for a given bit rate of left image is encoded. Similarly, the disparity compensation of the right image is also encoded by using SPIHT coding. Simultaneously, the disparity vector is encoded by entropy coding. The related work on stereo image compression study can be found in [3].

2. THE MULTIRESOLUTION WAVELET-BASED DISPARITY ESTIMATION

The basic technique used for disparity estimation is searching and comparing between the reference pixels in each region. At this stage, the tradeoff between the searching efficiency and the image quality is the most important. Though, an acceptable image quality is achieved when small size of the search space is used, but it takes long time for searching and uses huge memory space. On the other hand, a poor image quality is occurred when large size of search space is used, but the searching time is shorter. From this fact, it is not good to use only one disparity-block size with every wavelet subband. Therefore, we applied scaling block size for disparity estimation search in multilevel wavelet coefficients instead of using constant disparity-block size. In addition, we have also performed the comparing study on the processing time between the constant disparity-block size and the scaling disparity-block size as discussed in section 4.

2.1 Wavelet transform

The wavelet decomposition is a multiresolution representation of a signal by using a set of basic function that generated by the dilation and translation of a unique wavelet function $\psi(t)$ as shown in Eq.(1). The corresponding wavelets are

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}). \tag{1}$$

where $\psi_{a,b}(t)$ is a wavelet with scale factor a and shift factor b.

Let $\phi(t)$ be a low pass scaling function and $\psi(t)$ be an associated band pass wavelet function. Then, the two-dimensional wavelet decomposition can be constructed by using the separable products of $\phi(t)$ and wavelet function $\psi(t)$, The discrete two-dimensional wavelet transform [4] of the image size $M \times N$ pixels in function $f(x,y)$, in one level decomposition can be written as follow;

$$Af = ((f(x,y) * \phi(-x)\phi(-y))(2N,2M))_{(N,M)\in z^2} \tag{2}$$

$$D^1 f = ((f(x,y) * \phi(-x)\psi(-y))(2N,2M))_{(N,M)\in z^2} \tag{3}$$

$$D^2 f = ((f(x,y) * \psi(-x)\phi(-y))(2N,2M))_{(N,M)\in z^2} \tag{4}$$

$$D^3 f = ((f(x,y) * \psi(-x)\psi(-y))(2N,2M))_{(N,M)\in z^2} \tag{5}$$

The implementation of the decomposition in Eqs.(2)~(5) is similar to subband decomposition as shown in Fig.3. The output of the analysis system is a set of four $M/2 \times N/2$ sub-images: the so-called Af (LL), $D^1 f$ (LH), $D^2 f$ (HL) and $D^3 f$ (HH) subbands, which correspond to different spatial frequency bands in the image. The LL subband is a coarse (low resolution). The HL, LH and HH subbands contain details with vertical, horizontal and diagonal orientations, respectively. The total number of pixels in the four subband is equal to the original number of pixels, MN.

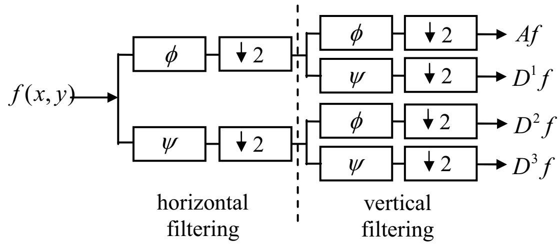


Fig.3 Wavelet decomposition of an image $f(x,y)$.

2.2 Disparity estimation and compensation

The advantage of this search technique is to reduce the number of search location and the number of search pixels. We use wavelet transform of left image I_L as a reference. Then, the wavelet coefficient of right and left images, i.e., I_R and I_L are divided into non-overlapped $m \times n$ blocks. The corresponding block in I_L is found by the applied mean square difference (MSD) from [5] as shown in Eq.(6). The entire pixels in the same block is assumed to have the same disparity estimation.

$$MSD(dx,dy) = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}[F(i,j) - G(i+dx, j+dy)]^2, \tag{6}$$

where
$F(i,j)$ is the $m \times n$ macroblock being compressed,
$G(i,j)$ is the reference $(m \times n)$ macroblock, and
(dx,dy) is the search location motion vector.

The values taken by (dx,dy) are specific to the definition of the search space. The search space parameter is generally called **p**. The search space is then defined as an (n+2**p**) by (m+2**p**) area. It is therefore an area **p** above, **p** below, **p** to the left, and **p** to the right of the block being compressed. We generally represent the search space in set notation given by dx={-**p**,+**p**} and dy={-**p**,+**p**}.

Since each block contains objects as various distances, the estimated disparity vector for any feature will be in error.

$$d_x = d_x + d_x, \tag{7}$$

where d_x is the estimated disparity, d_x is the true or actual disparity, and d_x is the disparity error [6].

2.3 SPIHT coding

A wavelet image can be represented as a tree structured spatial set of coefficients. A wavelet coefficient tree is defined as the set of coefficients form different bands that represent the same spatial region in the image. Fig. 4 shows a tree-level wavelet coefficient tree structure. Arrows in Fig.4 identify the parent –children dependencies in a tree. The lowest frequency band of the decomposition is represented by the root nodes (top) of the tree. Similarly, the highest frequency bands is represented by the leaf nodes (bottom) of the tree. Each parent node represents a lower frequency component than its children.

The SPIHT coding is based on the idea of using multipass zero-tree coding to transmit the largest wavelet coefficients (in magnitude) at first. A set of tree coefficients is significant if the largest coefficient magnitude in the set is greater than or equal to a certain threshold; otherwise, it is significant. Similarly, a coefficient is significant if its magnitude is greater than or equal to the threshold; otherwise, it is insignificant. In each pass the significance of a larger set in the tree is tested at first: if the set is insignificant, a binary “zero-tree” bit is used to set all coefficients in the set to zero; otherwise, the set is partitioned into subsets (or child sets) for further significance test. After all coefficients are tested in one pass, and the threshold is obtained before the next pass [7].

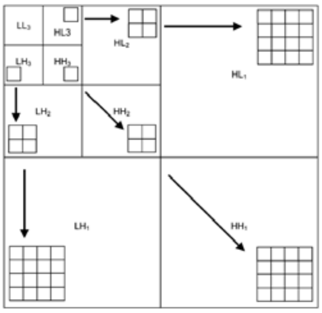


Fig. 4 Inter-band spatial dependency in wavelet decomposition.

Then, the output disparity vector of right image is encoded by entropy encoder and the others is encoded by using the set partitioning in hierarchical trees.

3. CONCEPT THEORY

As mentioned in section 2.2 that the advantage of this search technique is to reduce the number of search location and the number of search pixels. Then the main idea in the hierarchical search algorithm is to form several low-resolution versions of the frame to be compressed.

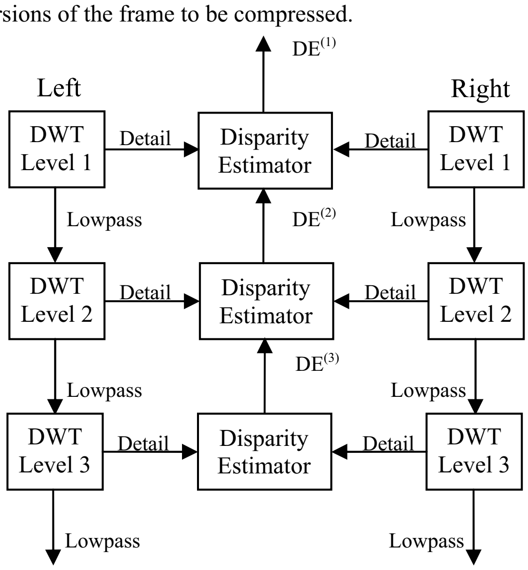


Fig.5 Disparity estimation for multiresolution wavelet transform.

In this paper, we introduce a 3-level biorthogonal 4.4 wavelet transform in two cases comparing with constant disparity-block size. In the first case; we implement the size of square block disparity (SBD) estimation and compensation by varying its values from 4x4 (**p**=2), 8x8 (**p**=4), and 16x16 (**p**=8), pixels for wavelet detail coefficients level 3, 2 and 1, respectively, as shown in Fig.5 and Fig 6. And because of ideal stereo geometry, the disparity displacement between the left and right image is horizontal difference, which means that the disparity displacement is more consistent along the horizontal direction than the vertical direction [2]. Thus, we also perform the experiment on the second case. That is we implement the size of rectangle block disparity (RBD) by varying its values from 4x8 (**p**=2), 8x16 (**p**=4), and 16x32 (**p**=8), pixels for wavelet detail coefficients level 3, 2 and 1, respectively.

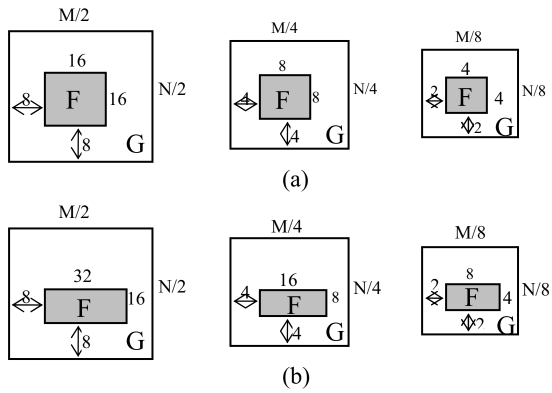


Fig.6 The search area G and block-matching size F in disparity estimation techniques for each level of wavelet detail coefficients in case : (a) SBD (b) RBD.

4. IMPLEMENTATION AND RESULTS

The 128x128 pixels of “Fruits” and, 128x128 and 256x256 pixels of “Room” are used as the test stereo image in our experiment. The peak signal-to-noise ratio (PSNR) is choosed as the performance metric for investigating the compression quality of reconstructed image. The average PSNR is defined as follow.

$$PSNR_{avr} = 10\log_{10}\left(\frac{255^2}{(D_L + D_R)/2}\right), \tag{8}$$

where D_L and D_R denote the mean-square-errors (MSE) applied from[6] of the reconstructed left and right images, respectively. The MSE is defined as Eq. (9).

Table 1 Comparison time for disparity estimation process of constant disparity-block size 8x8 pixels, SBD and RBD techniques at 0.7 bpp.

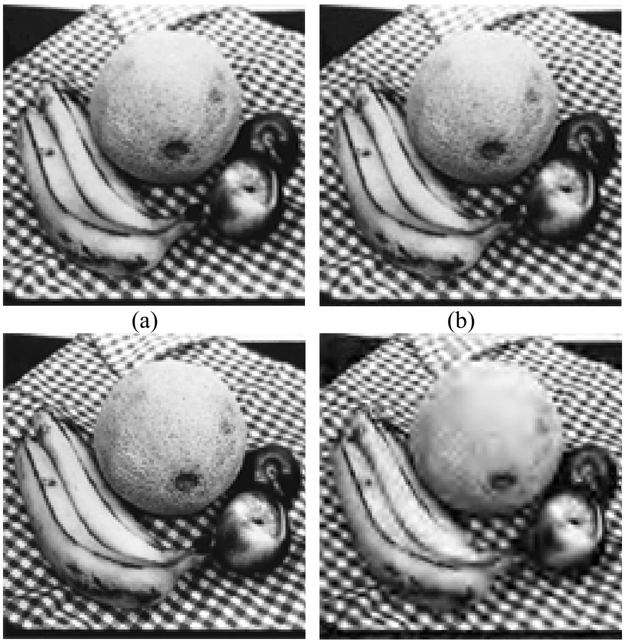
Images	Constant disparity-block size (8x8 pixels) (sec.)	SBD time (sec.)	RBD time (sec.)
Fruits 128x128	0.8006	0.7500	0.4060
Room 128x128	0.8120	0.7631	0.4220
Room 256x256	3.5940	3.5160	2.1720

$$MSE = \frac{1}{NM}\sum_{x=1}^{N}\sum_{y=1}^{M}[f(x,y) - \widehat{f}(x,y)]^2 \ , \quad (9)$$

where $f(x,y)$ and $\widehat{f}(x,y)$, $1 \le x \le N$, $1 \le y \le M$ are the original image and the reconstructed image with size $N \times M$, respectively.

Table 2 The results of RBD coding for the 128x128 "Room" and "Fruits" image pair at several bit rates.

image; bpp	PSNR of left image	PSNR of Right image	$PSNR_{avr}$
Room; 0.5	32.1588	31.9382	32.0457
Room; 0.7	33.7944	33.8091	33.8018
Room; 0.9	35.0845	34.9542	35.0184
Room; 1.1	35.5448	36.4754	36.5098
Fruits; 0.5	33.0674	32.8890	32.9363
Fruits; 0.7	34.3004	33.9336	34.1092
Fruits; 0.9	35.4360	35.0965	35.2596
Fruits; 1.1	36.3575	35.8799	36.1056

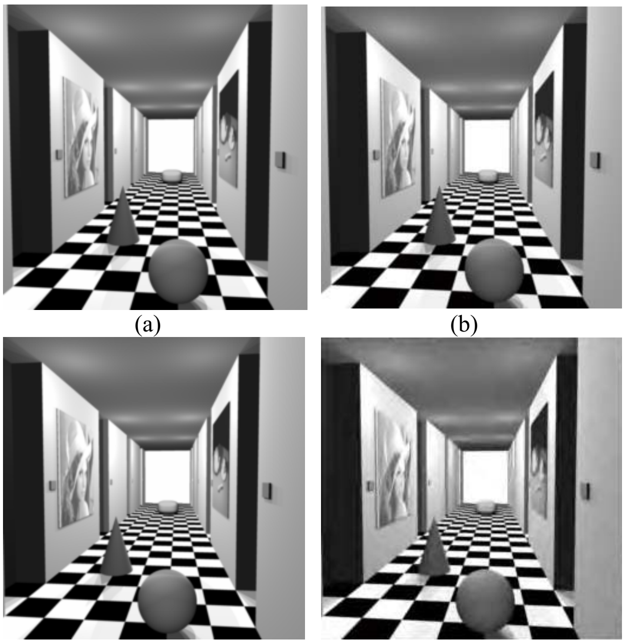


(a) (b) (c) (d)

Fig.7 RBD coding results for the "Fruits" image pair, (128x128 pixels each) at 0.9 bpp.
(a) Original left image, (b) Reconstructed left image,
(c) Original right image, (d) Reconstructed right image.

5. CONCLUSIONS

The square block of disparity estimation and compensation are wildly used for stereo image compression. In this paper, the adaptive block scan size for multiresolution wavelet coefficients are utilized. The advantage of this search technique is to reduce the number of search location and the number of search pixels. Moreover, from stereo characteristic, the left and right cameras geometric generate the disparity displacement in the horizontal correlation more than vertical direction. Then, we also apply non-square block size for disparity estimation. All of these techniques not only speed up the time of experiment but also give good quality of reconstructed image in both visual quality and measure quantities.



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

Fig.8 RBD coding results for the "Room" image pair, (256x256 pixels each) at 0.9 bpp.
(a) Original left image, (b) Reconstructed left image,
(c) Original right image, (d) Reconstructed right image.

REFERENCES

[1] F. Dufaux and F. Moscheni, "Motion Estimation Techniques for Digital TV: A Review and a New Contribution", *Proceeding of the IEEE*, Vol. 83, No.6, pp. 858-876, 1995.

[2] Jizheng Xu, Zixiang Xiong, and Shipeng Li, "High Performance Wavelet-Based Stereo Image Coding", *IEEE International Symposium on Circuits and Systems, ISCAS 2002.* , Volume: 2, 26-29 May 2002 pp. II-273 - II-276 ,2002.

[3] Chompoonuch Tengcharoen, Ruttikorn Varakulsiripunth, Tanasak Phanprasit and Manas Saworasil, "Stereo Image Compression by Quadrant Vector Quantization", *Image and Vision Computing, New Zealand (IVCNZ2003),* Massey University, Palmerston North, New Zealand, 26-28 November 2003, pp. 210-214, 2003.

[4] Qin Jiang, Joon Jae Lee and Monson H. Hayes, "A Wavelet Based Stereo Image Coding Algorithm", *ICASSP'99 Proceedings*, pp.3157-3160, 1999.

[5] Borko Furht, Joshua Greenberg and Raymond Westwater, *Motion Estimation Algorithms for Video Compression,* Kluwer Academic Publishers, USA, 1997.

[6] Mark S. Moellenhoff and Mark W. Maier, "Characteristics of disparity-compensated stereo image pair residuals",*Signal processing: Image Communication 14, ELSEVIER,* pp.55-59, 1998.

[7] Al Bovik, *Handbook of Image & Video Processing*, Academic press pubisher, 2000.