DISPARITY ESTIMATION FOR 3DTV VIDEO COMPRESSION USING HUMAN VISUAL PROPERTY

Myeong-Hoon Jo and Woo-Jin Song, *Member, IEEE*
Division of Electrical and Computer Engineering,
Pohang University of Science and Technology (POSTECH)
Pohang, Kyungbuk, 790-784, Republic of Korea
e-mail: wjsong@postech.ac.kr

Abstract - For efficient transmission of 3DTV video signals, it is necessary to eliminate the inherent redundancy between the stereo image pairs. Though disparity estimation provides a powerful tool for eliminating the redundancy, it is very time consuming. This paper presents a novel disparity estimation scheme based on the human visual property. The disparity vectors of image blocks spatially adjacent to the current block are used as initial guesses for the disparity vector of the current block. In addition, mixed-resolution coding is applied to reduce the computational complexity of disparity estimation. Through computer simulations on a stereoscopic sequence we show that the proposed method gives rise to visually pleasing results with much reduced computational complexity.

1. Introduction

Stereoscopic display systems such as three-dimensional (3D) TV receivers can reconstruct a 3D scene with slightly different two-dimensional (2D) views of a scene acquired from different perspectives. The 2D views are separately presented to the right eye and the left eye of the human observer. Such a presentation allows the perception of a depth in a scene by processing the disparity - geometric difference of the objects in stereo images - of the objects between the 2D views of stereo images [1].

To stimulate real-life applications of stereoscopic systems, it is imperative to develop an efficient compression schemes for a stereoscopic video that requires twice as many bits unless we remove the inherent redundancy in stereo images. Therefore disparity estimation (DE) is indispensable and is generally combined with motion estimation (ME) in transmitting a stereoscopic video signal for 3DTV.

Although disparity compensation (DC) provides a powerful tool for stereoscopic video coding, DE is very time consuming operation. Therefore, it is necessary to simplify the estimation algorithm without seriously degrading the quality of the reconstructed images. There are two different approaches for this purpose. One approach is the hierarchical disparity estimation (HDE) methods for stereoscopic video coding [2], [3]. These methods reduce the amount of computation needed for block matching. The other is the mixed-resolution coding based on the human visual property: a human being can easily fuse a high-resolution image and a low-resolution image into a 3D scene and perceive depth in them [4], [5]. [6] and [7] show that one motion vector can be predicted by a linear combination of other motion and disparity vectors. However, this assumption does not work for block-based matching algorithms because it is established for pixels between stereo image pair. This algorithm additionally requires the ME process and thus increases computation complexity. This novel disparity estimation algorithm reduces the computational complexity of DE process by efficiently using the HBM methods and human visual property in the 3DTV video sequences.

The remainder of this paper is organized as follows. In section 2, we discuss a hierarchical disparity estimation and in Section 3, we describe a mixed-resolution coding. In Section 4, we describe a novel disparity estimation for 3DTV. In Section 5, simulation results are presented, and finally Section 6 gives some conclusions.

2. Hierarchical Disparity Estimation

The hierarchical block matching (HMB) methods are used to reduce the computational complexity of motion estimation and disparity estimation process. For applying the HMB methods, we need L-level hierarchical image structure, which can be obtained by utilizing the Laplacian pyramid [3]. L-level hierarchical image structure is constructed by low-pass filtering and downsampling the image as follows:

$$B_{l+1}[i,j]=\sum_{m=-2}^{2}\sum_{n=-2}^{2} w[m,\ n]\cdot B_l[2i+m,\ 2j+n] \qquad (1)$$

where $B_l[i,j]$ indicates the (i,j)th pixel value in the L-layer image and $w[m,\ n]=w[m]w[n]$, $w[0]=a,\ w[1]=w[-1]=1/4,\ w[2]=w[-2]=1/4-1/2\times a$.

For an efficient disparity estimation, we must consider the differences between a stereoscopic sequence and a video sequence. First, every object of the left image of the stereo sequence is displaced with respect to the right image while only a few objects in the video sequence

change their location from frame-to-frame. Second, in a stereo sequence with the parallel axes geometry only horizontal directional displacements can exist while in video sequences, there can exist vertical and horizontal displacements in both directions. Therefore, we take care of the selection of search window range.

In addition, for the block-based techniques of disparity estimation, we must properly choose a matching block size. Using large block leads to inaccurate disparity estimation results. On the other hand, small block decreases the reliability of disparity estimation. Using hierarchical block matching method can solve this problem. The size of the block for matching varies with the resolution level. For example, the block size is $b_x \times b_y$ at level l, and it becomes $2b_x \times 2b_y$ at level l-1.

3. Mixed-Resolution Coding

When the applications of stereoscopic sequences are intended for human observers, we can exploit the stereoscopic sequences considering the human visual property. This makes the compression efficient with subjectively acceptable distortions. We call the compression technique considering human visual property as the mixed-resolution coding.

In Julesz's experiments with random dot stereogram, he found that binocular fusion depended on the identity of the low or high frequency spectrum in the two images [4]. We can perceive the depth perception in all the distorted stereograms [4], [5]. The compression is accomplished by presenting one eye with an original image and the other eye with a subsampled image. Then, it is easy to fuse such stereoscopic pairs and perceive depth in them. Furthermore, the resolution is similar to the resolution of the high-resolution picture rather than that of the low-resolution picture [4], [5].

4. Novel Disparity Estimation For 3DTV Sequence

As shown in the previous section, the hierarchical disparity estimation does not consider the human visual property to reduce the computational complexity of the disparity estimation (DE) process. Therefore, it is required to fully use the human visual property for efficient compression of stereoscopic video sequences

To achieve the goal, we propose a novel disparity estimation technique that combines the hierarchical disparity estimation methods and mixed-resolution coding. When one view of stereo sequences is low-pass filtered and subsampled, we know that the depth is still perceived and the sharpness is maintained due to the details in the other view sequence.

In the left view sequences, we apply the hierarchical block matching for the motion estimation. In B-, P-frames, the motion estimation is performed as two levels. The motion vector for the second level (the subsampled image) is used as an initial vector for the first level as follows

$$mv_1(i,j) = 2mv_2(i/2, j/2) \qquad (2)$$

where $mv_l(i,j)$ is the motion vector of the block at level l of the Laplacian pyramid

In the right view sequences, we utilize both HMB method and mixed-resolution coding for the motion estimation and disparity estimation process. For using the human visual property, we obtain the subsampled images of the stereoscopic sequences by utilizing the Laplacian pyramid method discussed Section 2. The disparity estimation is performed between the subsampled left and right image sequences instead of original stereoscopic sequences. In addition, when we perform the DE process, we use disparity predictors, which are the disparity vectors of spatially adjacent blocks – upper, left, upper left block. The Fig. 1 showed the disparity vector of current block and its predictors, where the DV_c indicates the current disparity vector of macroblock to be encoding, and the DV_1, DV_2, DV_3 indicate the disparity vector of spatial adjacent blocks. The disparity vector with the least mean absolute difference (MAD) among the predictors is used initial disparity vector of current block. The MAD used as the matching criterion is defined as

$$MAD(d_1,d_2) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|X(i,j,r) - Y(i+d_1, j+d_2, l)\right| \qquad (3)$$

where $X(i,j)$ and $Y(i,j)$ indicate the (i,j)th pixel value of block X, Y and (d_1,d_2) indicates a candidate disparity vector, and r and l indicate the left view image and the right view image, respectively. After we determine the initial disparity vector, we search the disparity vector in the search window centered on the initial disparity vector. By using the initial vector, we can make the search window size smaller, which makes the computation load of the DE process lower. When the adjacent blocks are intra-coded, there do not exist adjacent motion vectors. Then, we can not use the predictor and do not change the search window size. For

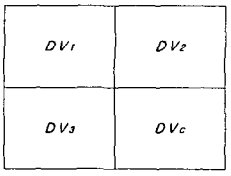


Fig. 1. The spatial relation of the current disparity vector and its predictors.

the forward prediction, the motion estimation is performed like the disparity estimation in the only right view sequence, that is, uses the predictors and makes the search search size smaller if the predictor exists.

5. Simulation Results

We applied the proposed algorithm on stereo image sequences called 'Aqua', 'Football'. The sequences are available as two views, a left view and a right view. The left view is coded independently by using MPEG-2 coding using frame-pictures with a prediction distance of M=3 (two B-frames) between pairs of reference frames and I-frame distance of N=12. The right view is coded with the hybrid disparity/motion compensation and is coded dependently with respect to the decoded left view frames. Three methods are tested in the experiment: the exhaustive search method, the hierarchical block matching (HBM) with predictors, the proposed algorithm. In our simulation, the search range for motion estimation of the proposed method is [-8: +7.5] horizontally and vertically, and the search range for disparity estimation of the proposed method is [-32: +31.5] horizontally and [-2: +1.5] vertically in the subsampled right view sequence of the stereoscopic sequence. In the exhaustive method and HBM method with predictors, we use the larger search window. For obtaining the disparity vectors and motion vectors, Table 1 shows the search range of a motion estimation and a disparity estimation. Table 2 compares the required computation complexity of various estimation methods. The required complexity means the required operations, such as, a subtraction, an absolute-value calculation, and an addition used in calculating the mean absolute difference criterion. In table 2, we normalize the computation complexity of the exhaustive search method equal to 100. The proposed algorithm outperforms all other methods. Both Fig. 2 and Fig. 3 show the disparity vector fields produced by the three different methods. It is evident that the proposed algorithm results in a smoother disparity vector field distribution than the others do, which the segmentation process easier.

We tested the images processed by these methods on the the 3D display. We compared the quality of stereoscopic images and the perception of a depth among these methods. Three people who experienced the 3D sequence and knew the three methods, did not find the differences among the methods.

6. Conclusions

The proposed disparity estimation (DE) combines the hierarchical disparity estimation with human visual property for reducing the computation complexity of the DE process. Additionally, using the predictor can reduce the search range of the disparity estimation and therefore can reduce the computation complexity of the DE.

The simulation results have shown that the proposed algorithm can compress the 3DTV sequences with subjectively acceptable distortions.

Acknowledgments

The authors wish to acknowledge the financial support of Electronics and Telecommunications Research Institute, Republic of Korea. The authors have been partially supported by BK21 program from Ministry of Education, Republic of Korea.

References

[1] A. Puri, R. V. Kollarits, B.G. Haskell, "Basics of Stereoscopic Video, New Compression Results and MPEG-2 and a Proposal for MPEG-4," *Signal Processing: Image Communication, Vol. 10,* pp201-234, 1997.

[2] Tzovaras, M. G. Strintzis, and H. Sahinoglou, "Evaluation of multi-resolution block matching techniques for motion and disparity estimation," *Signal Processing: Image Communication, Vol. 6,* pp.59-67, 1994.

[3] C.-W. Lin, E.-Y. Fei and Y.-C. Chen, "Hierarchical Disparity Estimation Using Spatial Correlation," *IEEE Trans. On Consumer Electronics, Vol.44,* No.3, Aug. 1998.

[4] B. Julesz, "*Foundations of cyclopean perception,*" Chicago, IL, the university of Chicago press, 1971.

[5] M. G. Perkins, "Data Compression of Stereo pairs," *IEEE Trans. On Communications, Vol. 40,* No. 4, pp684-696, Apr. 1992

[6] Y.-R. Yang and Y.-C. Chen, "MPEG-based Coding Algorithm for 3D-TV," *Signal Processing of HDTV, IV,* pp333-342, 1995

[7] A. Tamtaouri and C. Labit, "Constrained disparity and motion estimators for 3DTV image sequence coding," *Signal Processing: Image Communication, Vol. 4,* No. 1, Nov. 1991, pp.45-54.

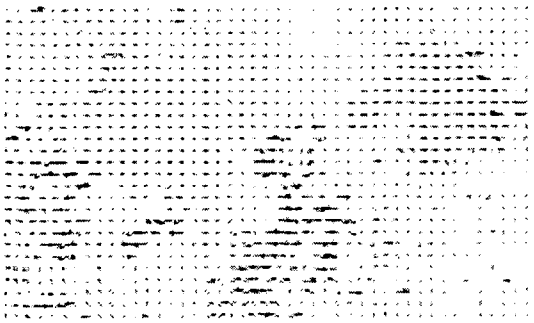
Table 1. Search range of a disparity estimation and a motion estimation.

	Exhaustive method & HDE method with predictors		Proposed Method	
	*H	*V	H	V
Disparity Estimation	[-64:+63]	[-8:+7]	[-32:+31.5]	[-4:+3.5]

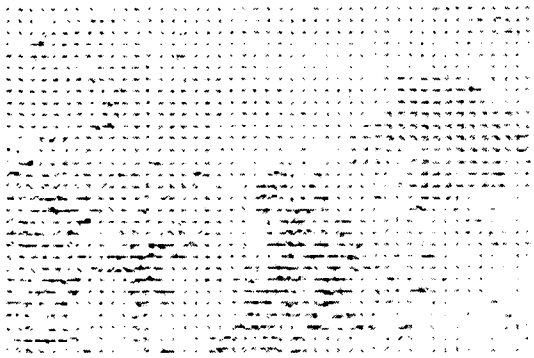
*H: horizontal, *V: vertical

Table 2. Performance comparison of various estimation methods

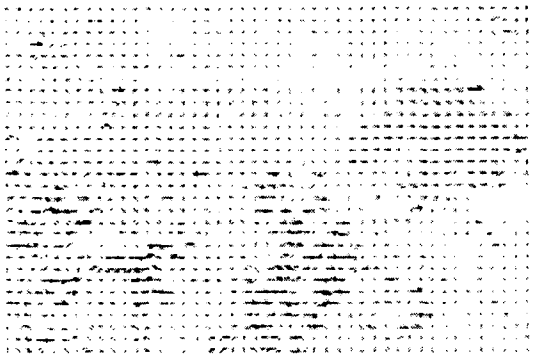
method	HDE method with predictors	Proposed method
Computation Complexity	5.75	4.8



(a) The exhaustive method

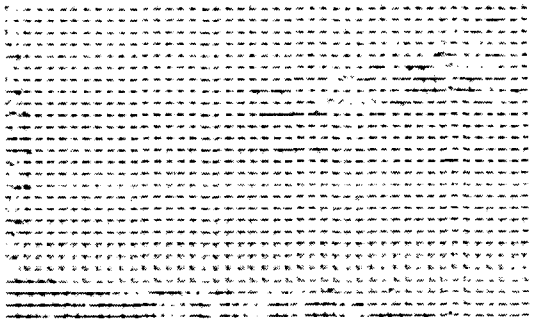


(b) HDE method with predictors [3]

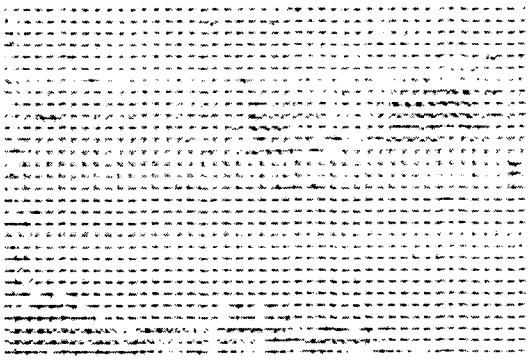


(c) The proposed method

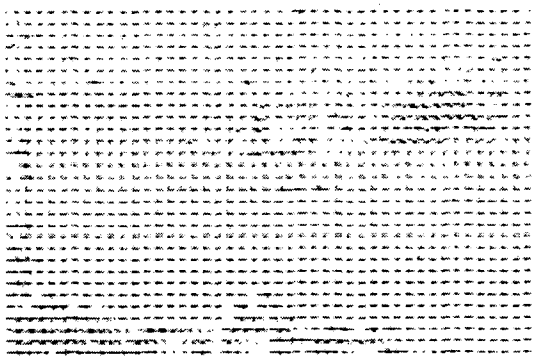
Fig. 2 Disparity vector field for sequence 'Aqua'



(a) The exhaustive method



(b) HDE method with predictors [3]



(c) The proposed method

Fig. 3 Disparity vector field for sequence 'Football'