An Adaptive Motion Estimation Algorithm Based on Mutual Information for Depth Information Estimation

Xin Lu, Maoliu Lin, Xuesong Jin
School of Electronics and Information Engineering
Harbin Institute of Technology
Harbin, China
E-mail: luxin5321@gmail.com

Zhijie Zhao, Shanshan Yang
School of Computer and Information Engineering
Harbin University of Commerce
Harbin, China
E-mail: zhaozj@ieee.org

***Abstract*—Left and right eye's views could be derived from the monoscopic image and corresponding depth map. Presenting these two slightly different views of every scene to the individual eyes with an appropriate disparity, a correct 3D perception can be realized within the mind. Taking the advantage of motion vector which reflects the movement between neighboring frames, the depth map could be reconstructed. The motion vector calculated by traditional motion estimation algorithm is not appropriate for depth information estimation and the computational requirement is huge. So we introduce the parameter of mutual information (MI). The block of motion estimation is classified into one of the two different types corresponding to background and active moving block depending on MI. The proposed adaptive method exploits the mutual information and adjusts the size of search area and block size. Simulation results show that the proposed algorithm can save the coding time by 10% compared with the full search algorithm and the quality of the depth map is improved significantly.**

Keywords-Motion estimation; Mutual information; Depth map estimation; Adaptive algorithm.

I. INTRODUCTION

3-Dimensions (3D) video display is nowadays often seen as one of the most important representations of future information. Experts and general public regard 3D as the next major step in video technologies [1]. Compared with traditional 2D video, 3D video could provide more impressive perception of multimedia contents to customers. Not only has the higher quality of video, but also having an immersed sense made 3D video more and more popular. With the advantages which 2D video does not possess, 3D video has a broad application prospect, such as video display, medicine image reconstruction, virtual reality, computer animation, video game, etc.. Thus 3D video generation technology emerges as the times require. For example, several special cameras have been designed to generate 3D model directly, which is the major source of 3D video. However, producing such video sequences are high cost. Another way is to convert current and past 2D video data into 3D format. Obviously, the 2D to 3D conversion method is a convenient and economical way. That is the point which we are focusing on.

With the development of study on conversion between 2D and 3D video, we found that the relative motion between neighboring frames provides an important cue to depth information. The motion may be seen as a form of "disparity over time", represented by the concept of motion vector. The basic assumptions for the method extracts depth information from motion vector are that the objects do not deform and their movements are linear. In addition, near objects move faster across the retina than far objects do. However, the motion vector calculated by traditional motion estimation algorithm could not meet the requirements of accurate depth information estimation, because the method does not take video content into consideration. In order to resolve this problem, we introduce the parameter of mutual information (MI), and classify blocks into one of the two different types corresponding to background and active moving block depending on image content. Then we propose an adaptive motion estimation algorithm, which is appropriate to recover video depth information. This algorithm could reduce the computing complexity and improve depth map estimation performance at the same time.

The rest of this paper is organized as follow: we first describe the scheme of conversion between 2D and 3D video in section II. In section III, an adaptive motion estimation algorithm based on mutual information is presented. Section IV illustrates how to recover depth map from motion vector. We give a computation analysis in section V. The experiment results are given in section VI. Finally, we conclude this paper in the last section.

II. THE SCHEME OF CONVERSION BEWTEEN 2D AND 3D VIDEO BASED ON DEPTH MAP

According to the stereo visual imaging mechanism, human vision system can interpret the 3D information of objects. Stereoscopy is based on exploiting the human perception. Simply, two views, taken at two slightly different angles are then guided to left and right eyes. The two eyes, receiving the two different views of the same scene from two different angles, namely binocular disparity, provide the visual signals to the brain; and then, the brain recovers the depth information of objects and interprets the scene as 3D[1] If depth information of the objects is available, we could deduce the binocular disparity, then left and right eye's views could be derived from the disparity. Based on this idea,

978-1-4244-8106-4/$26.00

this paper proposes a scheme of conversion bewteen 2D and 3D video(see figure 1).

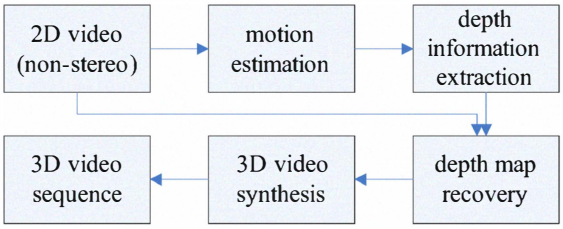

Figure 1. Scheme of conversion bewteen 2D and 3D video.

This conversion system is composed of three modular-motion estimation, depth map recovery, and 3D video synthesize. Motion vectors are extracted from video sequence in the modular of motion estimation. Conversion between motion vectors and depth map is completed in the modular of depth map recovery. 3D video synthesis calculates left and right eye's views using depth map and original 2D video.

III. Adaptive Motion Estimation Algorithm Based On Mutual Information

A. Mutual information

Let X, Y be discrete random variables with sets of possible outcome $A_X=\{a_1,a_2,...,a_N\}$, $A_Y=\{b_1,b_2,...,b_M\}$, with possibilities $\{P_{X1},P_{X2},...,P_{XN}\}$, $\{P_{Y1},P_{Y2},...,P_{YM}\}$ respectively, where $p_X(x=a_i)=P_{Xi}$, $p_Y(y=b_j)=P_{Yj}$, $P_{Xi}\geq 0$, $P_{Yj}\geq 0$ and $\sum_{x\in A_X} p_X(x)=1$, $\sum_{y\in A_Y} p_Y(y)=1$.

The MI between the random variables X and Y is given by[2]

$$I(X,Y)=-\sum_{x,y\in A_X,A_Y} p_{XY}(x,y)\log\frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}$$

and measures the amount of information conveyed by X about Y. It can be used as a measure of correlation between X and Y. MI is not only relevant to the information provided by X or Y, but also takes into account the information at their overlap. By these means, MI decreases when the amount of shared information between X and Y is small, vice versa.

Let us consider a gray level video sequence with intensity value ranging from 0 to N-1. $p_{t,t+1}(i,j)(0\leq i,j\leq N-1)$ is the possibility that a pixel with gray level i in a block of frame F_t has a gray level j in corresponding block of frame F_{t+1}. so the MI could be obtained as shown below[3]:

$$MI_{t,t+1}=-\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p_{t,t+1}(i,j)\log\frac{p_{t,t+1}(i,j)}{p_t(i)p_{t+1}(j)}$$

Where

$$p_t(i)=\sum_{j=0}^{N-1} p_{t,t+1}(i,j) \quad \text{and} \quad p_{t+1}(j)=\sum_{i=0}^{N-1} p_{t,t+1}(i,j).$$

MI can be used to detect motion degree between blocks in neighboring frames. A large difference between blocks (corresponding to the high motion activity) leads to a low MI value, while a small change between blocks responds to a high MI value.

B. Adaptive motion estimation algorithm

So far most of the algorithms developed for motion estimation are block-based techniques, called block-matching algorithm (BMA). In this technique, the current frame is divided into fixed size of blocks, then each block is compared with candidate blocks in reference frame within the search area. The widely used approach for the BMA is the full search BMA (FSBMA), which examines all candidate blocks within the search area to obtain a motion vector(MV). The most commonly used matching criteria is MAE(Mean Absolute Error) and MSE(Mean Square Error).

$$MAE(u,v)=\frac{1}{MN}\sum_{k=0}^{M-1}\sum_{l=0}^{N-1}|R(x+k,y+l)-P(x+u+k,y+v+l)|$$

$$MSE(u,v)=\frac{1}{MN}\sum_{k=0}^{M-1}\sum_{l=0}^{N-1}(R(x+k,y+l)-P(x+u+k,y+v+l))^2$$

From the perspective of mathematics, matching blocks obtained by above criterias are optimal. It would be beneficial to video compression. However, determining which block will be the optimal one merely by these criterias is unreasonable. Traditional motion estimation algorithm has some defects in such situation. For example, although two blocks belong to the same object, according to MAE/MSE criteria, the motion vectors may be quite different. So the motion vectors obtained would be not satisfactory. Thus in order to resolve this problem, we take video content into consideration, propose an adaptive motion estimation algorithm based on the type of block.

In traditional motion estimation algorithm, the size of search area and block size are fixed. The result of such algorithm is that, when we assign the block size is large, the motion information of object edge would be indistinct. On the contrary, when the block size is small, the contrast between motion vectors belong to the same object is quite sharp, due to block mismatch. Above phenomenon must bring adverse effects into the recovery of depth information.

As a result, this paper proposes an adaptive motion estimation algorithm for depth information estimation. Firstly, calculate the mutual information between corresponding blocks in the neighboring frames. Based on the analysis of mutual information, the current frame is classified into one of the two types corresponding to the background area and active moving region. Small search area and large block size are assigned to the background block and the large search area and small block size for the active moving block. In this way, we take video content into consideration so as to make the motion estimation more precise. Furthermore, the depth map will be more accurate.

C. Thresholds selection

As mentioned above, we classify the blocks into two types based on mutual information. so the threshold which decide the type of block should be selected according to the global characters of the frame. We draw up the statistic histogram of the mutual information in a frame, and then select the most frequent value as the threshold.

IV. DEPTH MAP RECOVERY

In order to synthesize stereo video sequence, the depth information should be recovered. Motion vectors obtained in the encoder stage is useful for depth information estimation. In some cases, motion vector maps can be approximated as depth maps. This approximation holds when the two images of the stereo pair are taken in parallel viewing or when they are acquired in small ranges of disparity in the case of epipolar acquisition. This is usually the case in computation of depth maps of 2D video. However, there are some cases where this approximation does not hold. This happens when the motion is either too rapid in terms of camera rotation or in the case of camera zoom. Such cases can be detected by analysis of the motion vectors and dealt with.

In the simulation, motion vector was treated as the only depth cue, that is, depth information is calculated depending on the values of the X and Y motion vector. The depth value could be calculated by[4]:

$$D(i,j)=\sqrt{MV(i,j)_x^2+MV(i,j)_y^2}$$

Where $D(i,j)$ is the depth value of point (i,j) and $MV(i,j)_x$, $MV(i,j)_y$ are the X and Y motion vectors values of point (i,j) respectively.

V. COMPUTATION ANALYSIS

In this paper, the proposed adaptive algorithm adds the computation of mutual information to the traditional one. However, the computation complexity is reduced. This is because, compared with the time cost by calculating mutual information, motion estimation takes much more time.

To take the block size of 4 by 4, and search area of 8 by 8 for an example. The motion estimation computation of one block in traditional motion estimation is:

$$ME=B\sum_{i=-7}^{7}\sum_{j=-7}^{7}C_{ij}$$

Where C_{ij} is the computation of calculating the mean absolute error (MAE) at the position of (i,j). B is the total number of block in each frame.

In addition, the computation of mutual information is[5]:

$$MI\approx 2N^2$$

N is intensity level, where is 256.

And the computation of the proposed adaptive motion estimation algorithm is:

Background: $ME=B_1\sum_{i=-3}^{3}\sum_{j=-3}^{3}C_{ij}$

Active moving area: $ME=B_2\sum_{i=-7}^{7}\sum_{j=-7}^{7}C_{ij}$

The block size of background and active moving ones are set to 8×8 and 4×4 respectively.

To take the video sequence named Akko&Kayo(format: 640×480) for an example, the proportion of background block is about 55%~60%, and the rest 40%~45% locate in the active moving area.

Based on above analysis, we could conclude the proposed adaptive algorithm is more time-efficient.

VI. EXPERIMENTS AND RESULTS

The proposed algorithm is simulated using the Silent, Football, Foreman, and Bus video sequences. The video format is 352×288. The comparison of full-search and adaptive algorithm is consolidated in table1.

TABLE I. COMPARISON OF TRANDITIONAL AND ADAPTIVE ALGORITHM

Video Sequences (Frames)	**Full-search motion estimation**		**Proposed adaptive motion estimation**	
	Encoding Time(S)	***PSNR (dB)***	***Encoding Time(S)***	***PSNR (dB)***
Slient(64)	167.34	55.88	153.73	55.87
Football(80)	202.64	56.20	179.94	56.04
Foreman（80）	205.96	55.69	181.85	55.34
Bus（64）	165.88	56.24	150.44	55.81

The depth maps of the video sequence of Akko&Kayo are showed in the figure2.

Figure3 shows 3D images created using the techniques detailed above.

VII. CONCLUSION

A new adaptive motion estimation algorithm is proposed based on mutual information for depth information recovery. This algorithm is simple and efficient in the term of computation complexity and the quality of depth maps. With this new motion estimation algorithm, we could improve the performance of 2D-3D video conversion system.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (No. 60772157) and Project of Harbin Science and Technology Bureau (No. 2010RFXXG017).

REFERENCES

[1] L. Onural, "An overview of research in 3DTV," 14th International Workshop on Systems, Signals and Image Processing, 2007 and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services, June 2007, pp.3-3, doi: 10.1109/IWSSIP.2007.4381083.

[2] Z. Cernekova, I. Pitas, C. Nikou, "Information theory-based shot cut/fade detection and video summarization," IEEE Transactions on Circuits and Systems for Video Technology, vol.16, no.1, Jan. 2006, pp. 82- 91, doi: 10.1109/TCSVT.2005.856896.

[3] Z. G. Liu, Y. H. Peng, C. W. Hui, W. C. Siu, "Adaptive HAAR-Like MCTF Based Wavelet Video Coding Scheme," 2007 15th International Conference on Digital Signal Processing, July 2007, pp.399-402, doi: 10.1109/ICDSP.2007.4288603.

[4] I. A. Ideses, L. P. Yaroslavsky, B. Fishbain and R. Vistuch, "3D from compressed 2D video," Proceedings of the SPIE - The International Society for Optical Engineering, Jan. 2007, doi: 10.1117/12.703416.

[5] X. Lu, M. L. Lin, X. S. Jin, Z. J. Zhao, "An adaptive fast video encoding/decoding scheme based on mutual information," Journal of Chongqing University of Posts and Telecommunications(Nature Science Edition), Vol. 22, No. 1, Feb. 2010, pp. 84-88.

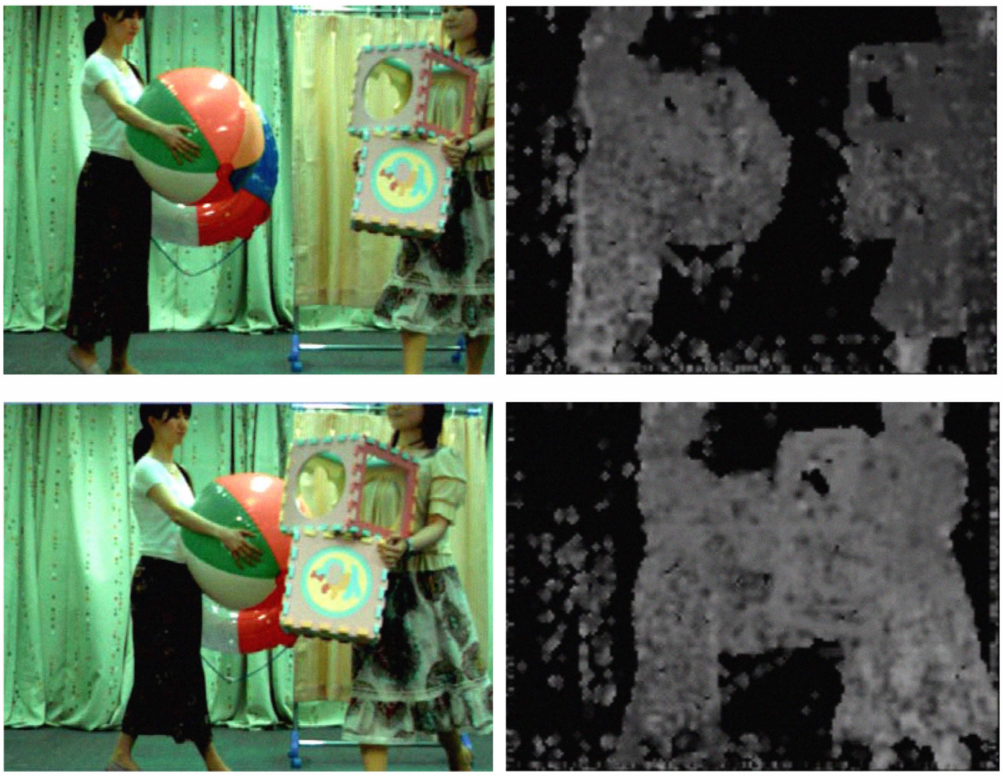

Figure 2. Original 2D images(left) and corresponding depth maps(right).

Figure 3. Original 2D images(left) and3D images(right).