Efficient Motion Estimation for Video Coding Using Depth Information

Soon-kak Kwon

Dept. of Computer Software Engineering, Dongeui University

Abstract: - This paper presents the reducing method of the complexity in motion estimation for video coding by using the depth information. We obtain two regions of each picture within video sequence from distance information calculated by depth camera. Background region has large distance, but object region has small distance. If a current block is the object region, then motion is searched within the object region of reference picture. Also if a current block is the background region, then motion is searched within the background region of reference picture. From simulation results, we can see that the proposed method compared to the full search method remains the almost same as the motion estimated difference signal and significantly reduces the searching complexity.

Keywords: - *Depth Information, Motion Estimation, Searching Complexity*

I.   INTRODUCTION

The Recently video compression standard was established by to ISO / IEC of the MPEG-4 Part 10 and ITU-U of H.264 standard [1]. In addition, the standardization work for the HEVC (High Efficiency Video Coding) as the next generation of video coding standards is in progress [2-4]. These video compression standards usually have used on motion estimation coding technique to remove the temporal redundancy. Motion estimation is the process of finding the blocks in the current picture (A bundle of neighboring pixels to spatial direction) and the neighboring referenced picture on the direction of time.

For the motion estimation process, the estimation measure to look for the block, the size of the searching area in the reference picture, the size of the current block, etc. must be considered. As for the estimation measure to look for the block, the sum of the absolute error signal between the current block and the reference block or the sum of the squared error signal is used. In terms of the complexity of the implementation, the absolute error is simpler than the squared errors. However, in terms of the accuracy of the estimation, the squared error is more accurate than the absolute error. The complexity of the motion estimation increases in proportion to the size of the searching area in the reference picture. The number of searching associated with the complexity of the motion estimation means the total number of searching in order to find the best matched block in the reference picture for each block in the current picture. The number of searching is directly proportional to the size of the searching area in the horizontal and vertical directions. As the size of the searching area gets larger, the complexity of the motion estimation will have to get larger; however, the accuracy of the estimation will likely be increased if the larger size of the searching area. Therefore, it is necessary to set an adequate size of the searching area for each video coding. Smaller the size of the block in the picture for the motion estimation is, greater the accuracy of the estimation is; however, there will be a disadvantage such as that the complexity required for the estimation will increase and the motion vector to be transferred additionally will increase. In the case of H.264, we can select variable block size, 16x16, 8x8, 4x4, etc.

Since the motion estimation accounts for more than 70 percent of the complexity of the implementation in the video coding, the several methods for high-speed motion estimation to reduce the complexity have been researched from the early stage. The existing high speed estimation methods include the searching pattern methods [5-7], the setting method of the searching starting location [8,9], the early termination method of the searching [10,11]. The searching pattern method has patterns such as diamond, hexagon for the three steps searching. These are to search in accordance with the limited pattern considering the motion vector distribution characteristic without using the whole searching method that searches all of the searching areas. However, although they can reduce the number of searching significantly, they also have a risk of reducing the accuracy of the estimation. The setting method of the searching starting location is to set efficiently the starting location to search in consideration with the temporal and spatial correlation between the blocks in the video; thereby, reducing the complexity of the estimation through searching only the limited area near to the starting location. The early termination method of the searching is the method to terminate early the comparison of the reference block in case of exceeding the criterion value upon setting the criterion value of the estimation measure.

These exiting methods obtain the information of brightness using only the color camera and perform the motion estimation procedure between the information of brightness. However, in recent days, a variety of

application fields using the depth camera as well as the color camera are on the rise. In this paper, we propose a motion estimation method that uses both the depth and color cameras. It is to group the pixels presented in a similar distance from the distance information obtained from the depth camera and extract the object through labeling. Then, it is to classify into the background region that has large depth located on a background, the object region that consists only of an object. The background region does not perform the motion searching whereas the object region limits the searching area to the internal region of the object in the reference picture; thereby, reducing the number of the searching of the motion estimation significantly.

II.   PROPOSED MOTION ESTIMATION USING DEPTH INFORMATION

It is possible to reduce the complexity of the motion estimation through separating the background region and the object region effectively in a state of that the camera is fixed, and estimating the motion only as to the object area. It is possible to classify from the depth information for the background and object regions. The region that has the depth information that is the farthest becomes a background; otherwise, it becomes the object regions. The object regions can be classified into more than one kind. A background region is a fixed one so that it does not have a change of the picture. Thus, it is not necessary to search the motion, and the difference value of the pixels located in the same spatial location of the current picture and the reference picture is used as a signal of the motion estimation.

The proposed motion estimation method obtains YUV signal and distance information between the objects from a color camera and depth camera.

(a) Color signal
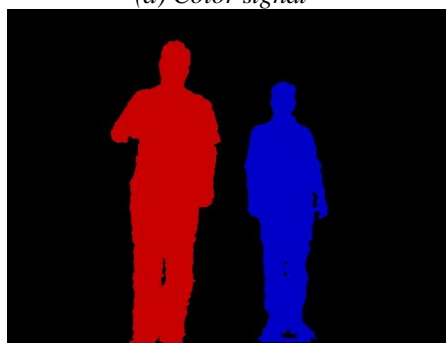
(b) Depth signal

Fig. 1: Color and depth signals obtained from the Kinect depth camera

The depth camera is implemented as an infrared-pulsed laser operating in the time of flight (TOF: Time-of-Flight). The infrared-pulsed laser includes an infrared laser diode, and the pulsed light emitted from the diode is investigated spatially. And the infrared-pulsed lasers are scattered on the surface of the subject, the optical-detector detects the returning time of each lasers. It is possible to measure the distance from the subject through measuring this time. The distance information of the measured subject consists of a picture that will be composed of each pixel.

Since the same object in the picture exists in a similar distance, it performs grouping of the pixels within several tens of centimeter and the labeling process. The pixel value of the labeling picture is stored in the size value equivalent to the number of the objects such as "0" in the background region, "1" in the first object region, "2" in the second object region, etc. The two blocks that are necessary to do the motion estimation based on this labeling picture, in other words, the motion estimation block of the current labeling picture and the motion searching block of the reference-labeling picture, are determined as being classified into the background, an object.

In the background region, all of the pixel values of the block in the labeling picture exist as "0". Therefore, if the sum of all of the pixel values in the block is "0", then the block will be determined that is within the background region

The internal region of an object is the case in which all of the pixel values in the labeling block is not "0" while all of them have the same value; thus, it will be to determine as an internal region of an object if it is the same value but non-zero when comparing all of the pixel values in the block.

It is necessary to have the current labeling picture and the reference-labeling picture when estimating the motion. The current labeling picture is same as the current color picture to estimate the motion. The reference-labeling picture is same as the color picture that before or after the time to need the motion estimation. As for the current labeling picture, the location of the block in the current picture is determined when the size of the block for the motion estimation is determined and the searching area for each block is determined. If a block in the labeling picture is determined to be a background region, it will not be to determine a searching area as to the reference block since it will not be necessary to search in most of the reference-labeling picture. If the current block is an internal area of an object, it will be to determine a searching area for each searching block in the reference-labeling picture.

Since the motion estimation is to reduce the information for coding through reducing the size of the signal of the motion estimation difference, the actual motion estimation should utilize the color information. The blocks of the current and the reference color picture are utilized at the same time when estimating; the motion estimation is being conducted based on the determination of the searching area of the block in each labeling-picture. The block in the current color picture varies as follows depending on the determination of the two searching areas of the labeling block.

• Background region

If all blocks are judged to background region as on same direction with current block. then it will not be to perform the motion searching, and the signal of the motion estimation difference is calculated as the difference between the values of the pixels in the same spatial location of current and reference blocks

The current block is within the background region, and if the reference block located in the same place with the spatial direction is not within the background region, then it will be to determine whether it will be within a background region as to the all of the reference blocks in the searching area. The motion searching is performed as to the reference block determined to be within a background region, but for any other reference block, the motion searching is not performed.

• Internal area of object

If the current block is within the internal area of the object and each block to be searched in the reference-labeling picture is determined to be within a background region, it will not be to perform the motion searching. If it is determined to be within an internal area of the object, it will be to perform the motion searching as to the block. The motion searching process obtains the motion vector to minimize the searching criterion between the block of the current color picture and the block of the reference color picture.

In a case of that an object suddenly appears in the current picture, it will not be to perform the motion searching as assuming the motion vector as "0", since an internal region of an object does not exist in the reference picture.

III.　　IMPLEMENT RESULTS OF PROPOSED METHOD

The In this paper, the Kinect [12] developed by Microsoft as depth camera is used. The Microsoft company developed the Kinect to apply in games as by merging the companies of the Canesta and the 3DV Systems; it uses the GestureTek's patented licensing and uses the PrimeSense's motion sensor. Kinect technology using the laser projects the infrared light, and measures the intensity of reflected waves from infrared dots; thereby divides the pictures. The dot with weak reflection is considered to be reflected from the object located far away, and the dot with strong reflection is presumed to have come from the user.

Labeling picture is shown in Fig. 1, obtained from object information from color picture and distance information using PrimeSense's OpenNI [13]. Each picture consists of 640 and 480 pixels in the horizontal and vertical directions, respectively. A total of 90 pictures, 30 pictures per second were used. The test video includes two objects and background. The motion estimation block size is 16 pixels and 8 pixels on the horizontal and vertical directions, respectively and the searching area is 10 pixel, 15 pixel, or 20 pixels for comparing the performance of the existing and proposed methods. The two kinds of methods are estimated the motions as integer unit. The reference picture is considered one of previous-picture. Table 1 shows that the proposed method is almost same level the signal of the motion estimation difference with the existing method. The number of the searching is reduced to under the 12 percent.

Table 1: Comparison of motion estimation performance of the proposed method with the conventional method

Estimated block size	Search region size	Conventional method		Proposed method	
		average searching per picture	MSE	average searching per picture	MSE
16	10	500200	35.54	57960	35.59
	15	1089000	33.90	116820	33.94
	20	1881744	31.62	190032	31.65
8	10	2046816	23.88	226803	23.92
	15	4423776	22.91	452693	22.94
	20	7660704	21.41	821792	21.43

The reason that the signal of the motion estimation difference increases very small compared to the existing method is that the proposed method do not search the motion in the background region and the motion vector is "0".

However, it is because the motion vector is actually not "0", rather the block can be found by the whole searching due to noise or a change of light even though there is not a motion.

The number of the motion searching in the proposed method may be affected by the ratio comprised by the object area per picture. Smaller the ratio comprised of the object region compared to the background region gets, the searching number of the motion estimation of the proposed method will get smaller proportionally. In addition, the efficiency of the decrease of the searching number due to the proposed method affects largely to the bi-directional motion estimation since the complexity of the motion estimation increases by at least more than two times compared to the unidirectional motion estimation in case of the bi-directional motion estimation. Furthermore, the efficiency of the proposed method will increase even more if it is to perform the motion estimation for a variety of variable block sizes such as 16x16, 8x8, 4x4 at the same time just like H.264.

IV.   CONCLUSION

The object motion estimation by using the contour information has proposed from MPEG-4 Part 2, but the real time object extraction had failed. In this paper, thus, we extract the object using the depth camera, and proposed a method to decide searching area from the information of the labeling of the object. The motion searching is decided from the labeling signal in the current and referenced pictures. If the current and the referenced blocks are background region, the motion searching is not working. But if, there it is in the internal area of the object, it is performed the motion searching. This proposed method reduces to 12 percent the number of the searching than the full search method. The proposed motion estimation method can applied to the international standard video coding techniques. Furthermore, the motion estimation method using depth camera can combined with existing fast motion estimation.

V.   ACKNOWLEDGEMENTS

VI.   REFERENCES
[1]   S.-k. Kwon, A. Tamhankar, and K. R. Rao, Overview of H.264/MPEG-4 Part 10, *Journal of Visual Communications and Image Representation, 17(2)*, 2006, 186-216.
[2]   K. Ugur, K. Andersson, A. Fuldseth, G. Bjøntegaard, L. P. Endresen, J. Lainema, A. Hallapuro, J. Ridge, D. Rusanovskyy, C. Zhang, A. Norkin, C. Priddle, T. Rusert, J. Samuelsson, R. Sj¨oberg, and Z. Wu, High Performance, Low Complexity Video Coding and the Emerging HEVC Standard, *IEEE Trans. Circuit Syst. Video Technology, 20(12)*, 2010, 1688-1697.
[3]   W.-J. Han, J. Min, I.-K. Kim, E. Alshina, A. Alshin, T. Lee, J. Chen, V. Seregin, S. Lee, Y. M. Hong, M.-S. Cheon, N. Shlyakhov, K. McCann, T. Davies, and J.-H.Park, Improved Video Compression Efficiency through Flexible Unit Representation and Corresponding Extension of Coding Tools, *IEEE Trans. Circuit Syst. Video Technology, 20(12)*, 2010, 1709-1720.
[4]   F. Bossen, V. Drugeon, E. Francois, J. Jung, S. Kanumuri, M. Narroschke, H. Sasai, J. Sole, Y. Suzuki, T. K. Tan, T. Wedi, S. Wittmann, and P. Yin, Y. Zheng, Video Coding using a Simplified Block Structure and Advanced Coding Techniques, *IEEE Trans. Circuit Syst. Video Technology, 20(12)*, 2010, 1667-1675.

[5]   S. Zhu and K.-K.Ma, A New Diamond Search Algorithm for Fast Block-Matching Motion Estimation, *IEEE Trans. Circuit Syst. Video Technology, 10(2)*, 2000, 287-290.

[6]   C. Zhu, X. Lin, and L.-P.Chau, Hexagon-based Search Pattern for Fast Block Motion Estimation, *IEEE Trans. Circuit Syst. Video Technology, 12(5)*, 2002, 349-355.

[7]   X. Jing and L. P. Chau, An Efficient Three-step Search Algorithm for Block Motion Estimation, *IEEE Trans. Multimedia, 6(3)*, 2004, 435-438.

[8]   M. Ezhilarasan and P. Thambidurai, Simplified Block Matching Algorithm for Fast Motion Estimation in Video Compression, *Journal of Computer Science, 4*, 2008, 282-289.

[9]   Y. Nie and K.-K.Ma, Adaptive irregular pattern search with matching prejudgment for fast block-matching motion estimation, *IEEE Trans. Circuit Syst. Video Technology, 15(6)*, 2005, 789-794.

[10]   L. Yang, K. Yu, J. Li, and S. Li, An Effective Variable Block-size Early Termination Algorithm for H.264 Video Coding, *IEEE Trans. Circuit Syst. Video Technology, 15(6)*, 2005, 784–788.

[11]   J.-F. Yang, S.-H.Chang, and C.-Y.Chen, Computation Reduction for Motion Search in Low Rate Video Coders, *IEEE Trans. Circuit Syst. Video Technology, 12(10)*, 2002, 948–951.

[12]   http://www.microsoft.com/en-us/default.aspx

[13]   http://openni.org