Temporally Consistence Depth Estimation from Stereo Video Sequences

Ji-Hun Mun and Yo-Sung Ho(✉)

School of Information and Communications,
Gwangju Institute of Science and Technology (GIST),
123 Cheomdangwagi-ro, Buk-gu, Gwangju 500-712, Republic of Korea
{jhm,hoyo}@gist.ac.kr

Abstract. In this paper, we propose complexity reduction method in stereo matching. For complexity reduction in video sequences, we start from generating of initial disparity information. Initial disparity information can give a restricted disparity search range when performing the local stereo matching. As an initial disparity information, we use 4 different kinds of input images. The initial disparity information types can divide into two main streams like '*calculated'* and '*given'* materials. Iterative stereo matching method, motion prediction stereo matching method and global matching result based stereo matching method are used as a '*calculated'* initial value. Captured by depth camera image is used as '*given'* information. By using those 4 different types of disparity information, we can save the time consuming when performing the local stereo matching with consecutive image sequences. Results of the experiment prove the efficiency of proposed method. By using proposed local stereo matching method, we can finish all procedures within a few seconds and conserve the quality of disparity images.

Keywords: Local stereo matching · Stereo camera · Time complexity

1 Introduction

Binocular disparity is essential concept to interpret the three-dimensionality of human visual system. Because of the importance of human visual system, jointed two-frame stereo matching is one of the most considerably studied field in computer vision. Stereo matching algorithm attempts to generate the depth information from stereo image pairs captured by stereoscopic camera or multiple cameras. It is also the major concern in a wide variety of applications such as super multi-view virtual view synthesis, free view point image and three-dimensional virtual reality. Most stereo matching algorithm can be roughly classified into two categories: *global* and *local* matching methods [1].

Global stereo matching method calculate the matching problem using an energy function with data and smoothness constraint. To find out the minimized energy value, generally dynamic programing [2], graph cuts [3] and belief propagation [4] are used. Global optimization method can considerably reduce the stereo matching ambiguities. Normally matching ambiguity factors are coming from illumination variation, saturation region and texture region. Global stereo matching can generate more efficient matching result than local stereo matching method, because global stereo matching consider all of

Y.-S. Ho et al. (Eds.): PCM 2015, Part II, LNCS 9315, pp. 579–588, 2015.
DOI: 10.1007/978-3-319-24078-7_59

the image conditions. But this method usually has an expensive computation time due to consideration of all image pixel value for minimization. To reduce the time complexity many state-of-the-art global stereo matching methods are developed [5].

Local stereo matching basically use the window concept in matching procedure. This method compute each pixels disparity value independently over the all image region. To derive the cost function value, the matching costs are aggregated over the window region. Within the window size, minimal cost value is selected as an output of the associated pixel. The procedure of local stereo matching method basically depend on the support window user designated. Generally window size starting from 3×3 to (2n-1) ×(2n-1) size when 'n' is not zero and larger than 2. If the window size is small, then matching result has a considerably accurate disparity result in edge or boundary region. While performing the matching procedure with large window size, then conversely matching result has an accurate value on homogeneous region or similar texture region [6].

Our algorithm basically focus on the local stereo matching method due to the complexity problem in global stereo matching. Generally to solve the complexity problem when performing the local or global stereo matching, GPU process is normally used for implementation step [7]. And one of the state-of-the-art local stereo matching, adaptive supporting-weight (ADSW) [8], could deliver disparity maps close to global optimization. However, the matching method like ADSW suffer from highly computational complexity, and the following associated research attempted to accelerate it. Chang et al. [9] simplify the ADSW by using a hardware implementation. Although the previous researching for high complexity problem, but they basically use other hardware assistance like using a GPU: CUDA or onboard implementation.

In this paper, we propose a new local stereo matching algorithm in video sequences which based on the initial disparity information. Different from the previous complexity reduction research, our algorithm provide a significant solution only using a software method. As an initial disparity information we use 4 different type of disparity generation method. In consecutive image sequence, between the neighbor image frames, they has a small disparity differences. Because of that reason from the previous stereo images matching result information, current frame stereo images just consider smaller disparity range than previous stereo images disparity range.

To testify proposed algorithm, we use four different computer graphic video sequences with depth ground truth image that provided by Cambridge computer laboratory. We provide full description of the proposed local stereo matching algorithm in Sect. 2. In Sect. 3, we discuss about the experiment result of proposed algorithm. In experiment result we provide a 4 different test sequences matching results and compare the matching results with given depth ground truth image for efficiency. After showing the experiment results with analysis of time complexity, we conclude this paper in Sect. 4.

2 Temporal Domain Stereo Matching

In this section, we will discuss about temporal domain local stereo matching method for time complexity reduction. The proposed temporal domain stereo matching methods can divide into two main idea. Firstly we use general stereo matching method

for initial disparity information which comes from the first frame image pairs. And the other method use given depth information which captured by ToF depth camera or Kinect depth camera etc. Because we basically use general local stereo matching method for several proposed method, we will briefly explain about that method.

2.1 Local Stereo Matching

Generally used local stereo matching method basically use pre-designated window size. The basic local stereo matching use restricted disparity search range. To find out the most similar pixel value between image pairs, local stereo matching restrict the disparity search range. Figure 1 show general local stereo matching procedure.

In Fig. 1, each consecutive image frame pairs has different disparity result. But between that results it just has small different disparity value. We focus on that point and will apply this properties in temporal domain stereo matching method.

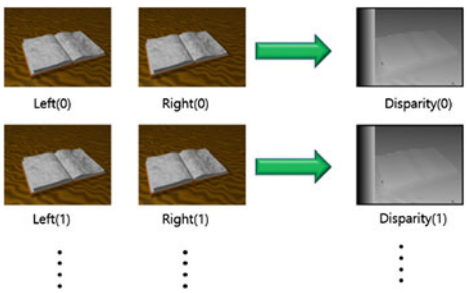

Fig. 1. General stereo matching procedure

2.2 Iterative Stereo Matching

Among the consecutive image frames, disparity value has small difference between closest neighbor image frames. As stated in previous section we use this properties in iterative stereo matching method procedure. Iterative stereo matching method consider temporal domain disparity information differences. But this method need only one time general local stereo matching method on the beginning. Because as an initial disparity information we need initial stereo image matching result. Figure 2 represent the iterative stereo matching procedure.

In iterative stereo matching method, initially generated disparity information 'Disparity (0)' used for following image pairs stereo matching procedure. As mentioned in previous section, we use following image pair disparity result value and previous disparity result value differences. In this paper we apply smallest disparity searching range with difference of adding disparity sign. As represented in Fig. 2, following image pair disparity result has similar to previous frame image pair result. Iterative stereo matching method continuously used in video sequences, from frame (1) pair, frame (2) pair to frame (n) pair.

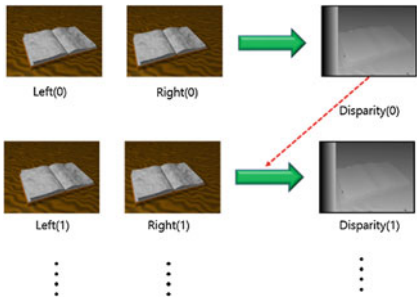

Fig. 2. Iterative stereo matching procedure

Based on the properties of previous and current disparity value differences, Eq. 1 can be derived.

$$\begin{aligned} \mathrm{Min}_{\mathrm{disparity}} &= disp_{pre} - n \\ \mathrm{Max}_{\mathrm{disparity}} &= disp_{pre} + n \end{aligned} \tag{1}$$

In Eq. 1 value n represent the disparity search range differences in following frame image pair stereo matching procedure. And $disp_{pre}$ mean previous image pair disparity result value. When generate disparity result using image pair, disparity search range has to be defined before starting that procedure. If we use original minimum and maximum disparity range, then it will take amount of time consuming compare to restricted disparity search range. Because of that reason we propose iterative stereo matching method to solve time complexity problem. When we apply iterative stereo matching method, we can more efficiently compute the disparity value than general stereo matching method. Even we diminish the disparity search range, iterative stereo matching method can consider the significant disparity search range while performing the stereo matching.

2.3 Motion Prediction Stereo Matching

In consecutive image sequences, following frame like 'Left (0)' and 'Left (1)' has a small difference between that frames. Even human eyes hardly perceive the difference of two images, but it has a difference value in terms of image pixels. Considering motion flow for stereo matching research has been actively performed [10, 11]. In previous works about motion prediction, proximity and similarity have been considered in the computation of image disparity map. However, that information insufficient for video disparity estimation because motion cues are very important for accurate disparity calculation near edge of moving objects. We include motion flow to compute disparity value more clearly in objects edge or moving area. Figure 3 indicate that stereo matching procedure with considering motion flow in consecutive image frames.

In Fig. 3, *Moving difference* represent difference of following image frames difference result. Comparing to iterative stereo matching method, motion prediction stereo matching method added motion information. But when compute motion information, it

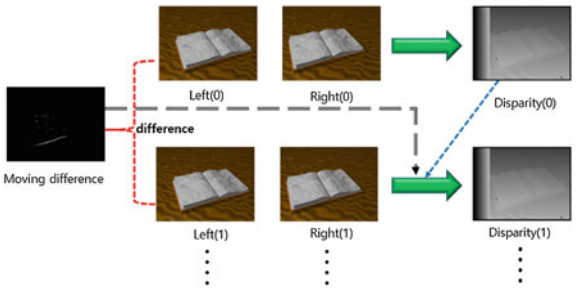

Fig. 3. Motion prediction stereo matching procedure

has different result depending on threshold value to determine 0 and 255 pixel value. Equation 2 represent the motion difference determining method.

$$\begin{aligned} Color_{diff} \geq th \quad diff_{map} = 255 \\ Color_{diff} < th \quad diff_{map} = 0 \end{aligned} \tag{2}$$

If difference of following image value is bigger than threshold value then Moving difference has 255 pixel value, otherwise it has 0 pixel value. So pre-determined threshold value effect on motion estimation result. Figure 4 show different result of motion estimation of consecutive image frames.

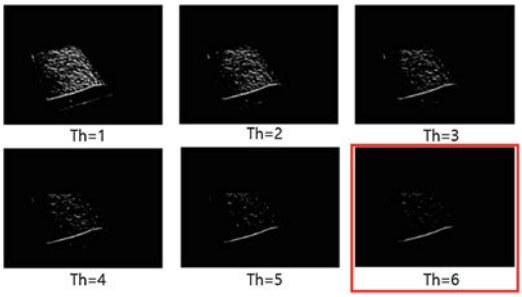

Fig. 4. Difference of motion estimation result for threshold value

Indicated in Fig. 4, as the threshold value is increased, the motion estimation results change into blurred image. Because we focusing on the object moving area or boundary region, texture region and inside of object information is unnecessary. In this paper we use fixed threshold value 6. If we use bigger threshold value, then object boundary region will be removed. Then it will spoil the stereo matching result.

The main framework same as iterative stereo matching, but to accurately compute the object boundary or moving region we use motion prediction stereo matching method. Comparing to time consuming of iterative stereo matching method, it will takes more time because of the motion predicted region. During the stereo matching,

window face to prediction region, then it have to search for original minimum and maximum disparity range. As a result of that we can get more accurate disparity information.

2.4 Global Matching Based Stereo Matching

As mentioned in introduction, global stereo matching has better result than local stereo matching. So we use global stereo matching information as an initial value [12]. Global stereo matching result has more accurate disparity value compare to previous initial information, consecutive frame stereo matching result also more accurate than previous one. Figure 5 represent the global matching based stereo matching method.

Likewise previously explained proposed method it has same framework, but only initial value is different as a global matching result.

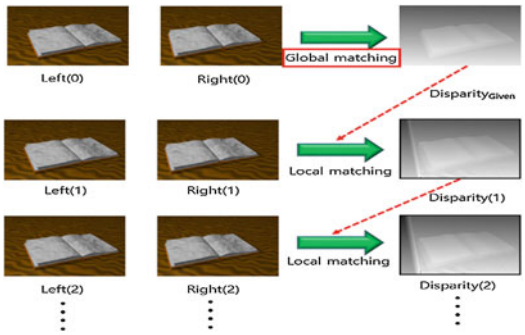

Fig. 5. Global matching based stereo matching procedure

2.5 Given Depth Based Stereo Matching

To generate depth information directly from the object, usually ToF depth camera and Kinect depth camera are used. If we use captured depth image, then it need to change depth value into disparity value with considering the Eq. 3.

$$Z_{near} = \frac{f \cdot l}{d_{max} + \Delta d'} \quad Z_{far} = \frac{f \cdot l}{d_{min} + \Delta d} \tag{3}$$

In Eq. 3, f is focal length of camera and l represent the base line of cameras. And also we already know the Z_{near} and Z_{far} value, so apply those parameter value in Eq. 3, then disparity minimum and maximum values can easily derived.

In this paper we use given computer graphics depth ground truth value, so we don't need to compute depth information to disparity value changing procedure. As like the iterative stereo matching method in Fig. 2, it also has same framework. But as an initial value it use given depth information, like a captured depth image or computer graphics ground truth image.

3 Experiment Results

We testify our proposed stereo matching method using computer graphic sequences provided by *Cambridge computer laboratory*. Test platform is a PC with Core(TM) i7-5960X 3.00 GHz CPU and 32.0 GB memory. And we use 4 different video sequence as indicated in Fig. 6: (a) *book*, (b) *street*, (c) *tanks* and (d) *temple*.

Fig. 6. Test video sequences

All of the test sequences has same resolution as 400×300 and also depth ground truth image has same resolution. Figure 7 show provided depth ground truth images.

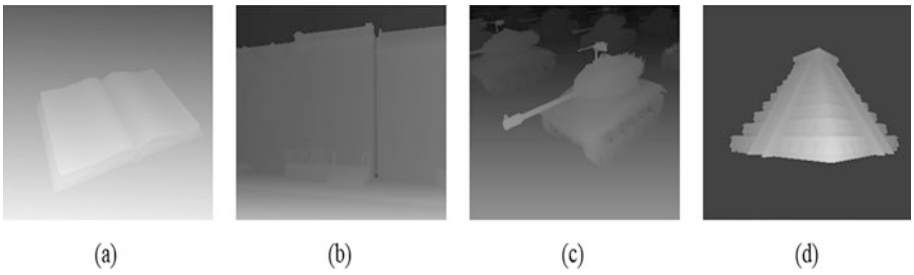

Fig. 7. Depth ground truth images for each test sequences

To compare disparity result image and depth ground truth image, we use bad pixel rate (BPR). If difference of computed disparity value and depth ground truth value is bigger than 1, then we determine that pixel as bad pixel. And also to verify the efficiency of our algorithm compare processing time for each proposed method with general local stereo matching method.

As represented in Table 1, temporal domain information based stereo matching is 10 % faster than general stereo matching method.

Figure 8 represent bad pixel rate comparison results for each test sequences. Except for test sequence ‘Street’, other test sequences has highest bad pixel rate when we use plus/minus 3 disparity search range. And stereo matching result with original minimum and maximum disparity search range has similar bad pixel rate even the frame numbers increase.

Table 1. Temporal domain stereo matching methods time comparison

Matching method	Disparity range			
	Min/Max	±3	±5	±7
General	21.76(sec)	–	–	–
Iterative	–	2.43(sec)	3.77(sec)	5.12(sec)
Prediction	–	3.17(sec)	4.45(sec)	5.76(sec)
Global	–	2.09(sec)	3.50(sec)	4.86(sec)
Given	–	2.10(sec)	3.47(sec)	4.84(sec)

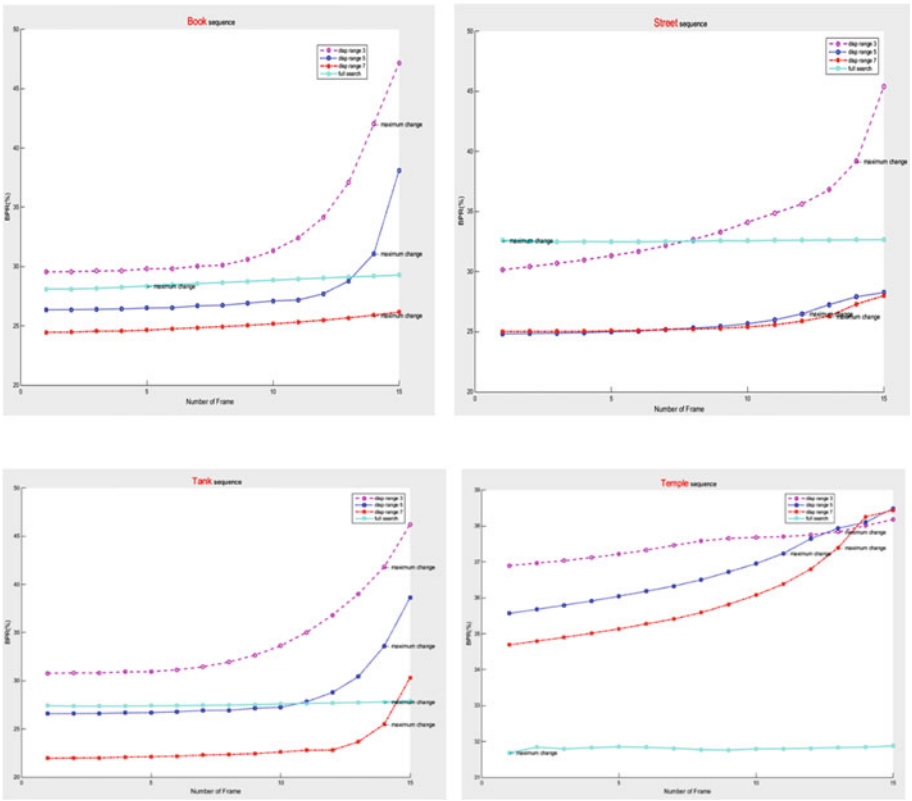

Fig. 8. BPR test for each test sequences

Proposed stereo matching results are represented in Fig. 9. When we use global matching method and given depth image as a base information, matching results has better quality than other matching methods.

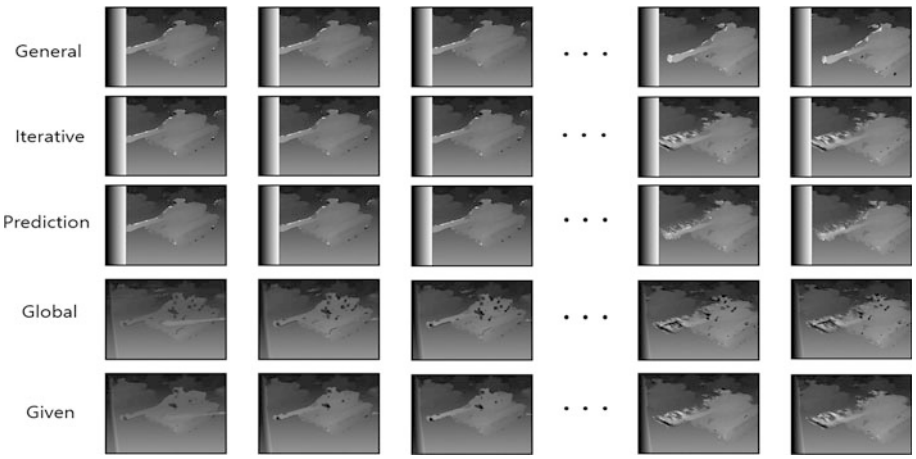

Fig. 9. Test results for sequence 'Tank'

4 Conclusion

In this paper we propose local stereo matching method in video sequences for reducing time complexity. Iterative matching, motion prediction and global matching method need 1 time general stereo matching procedure. And given depth based method directly use the temporal domain information for local stereo matching. From the experiment result we can check that proposed method is 10 % faster than general local stereo matching and wide disparity search matching has better BPR value. General local stereo matching has similar BPR value over the frame number, but proposed methods BPR values are decreased. From that result we need to refresh the reference disparity image by using minimum/maximum disparity range. We will research about that kind of problem to prevent the error propagation.

Acknowledgments. This research was supported by the 'Cross-Ministry Giga KOREA Project' of the Ministry of Science, ICT and Future Planning, Republic of Korea (ROK). [GK15C0100, Development of Interactive and Realistic Massive Giga-Content Technology]

References

1. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. IJCV **47**(103), 7–42 (2002)
2. Bobick, A.F., Intille, S.S.: Large occlusion stereo. IJCV **33**(3), 181–200 (1999)
3. Boykov, Y., Veksler, O., Zabig, R.: Fast approximate energy minimization via graph cuts. IEEE TPAMI **23**(11), 1222–1239 (2001)
4. Meltzer, T., Yanover, C., Weiss, Y.: Globally optimal solutions for energy minimization in stereo vision using reweighted belief propagation. IEEE Int. Conf. ICCV **1**, 428–435 (2005)

5. Bleyer, M., Chambon, S., Gelautz, M.: Evaluation of different methods for using colour information in global stereo matching approaches. ISPRS Congr. **1**, 1–6 (2008)
6. Hirschmuller, H., Scharstein, D.: Evaluation of cost functions for stereo matching. IEEE Conf. CVPR **1**, 1–8 (2007)
7. Zhang, K., Lu, J., Yang, Q., Lafruit, G., Lauwereins, R., Gool, L.V.: Real-Time and accurate stereo: a scalable approach with bitwise fast voting on CUDA. IEEE Trans. CSVT **21**(7), 867–878 (2011)
8. Yoon, K.J., Kweon, I.S.: Adaptive support-weight approach for correspondence search. IEEE Trans. Pattern Recogn. Mach. Intell. **2**(4), 650–656 (2006)
9. Chang, N.Y.C., Tsai, T.H., Hsu, P.H., Chen, Y.C., Chang, T.S.: Algorithm and architecture of disparity estimation with mini-census adaptive support weight. IEEE Trans. Circ. Syst. Video Technol. **20**(6), 792–805 (2010)
10. Lee, Z., Khoshabeh, R., Juang, J., Nguyen, T.Q.: Local stereo matching using motion cue and modified census in video disparity estimation. In: Signal Processing Conference (EUSIPCO), pp. 1114–1118 (2012)
11. Richardt, C., Orr, D., Davies, I., Criminisi, A., Dodgson, N.A.: Real-time spatio-temporal stereo matching using the dual-cross-bilateral grid. In: Proceedings of the ECCV, pp. 510–523 (2010)
12. Jnag, W.S., Ho, Y.S.: Discontinuity preserving disparity estimation with occlusion handling. J. Vis. Commun. Image Representation **25**(7), 1595–1603 (2014)