Temporal Stereo Disparity Estimation with Graph Cuts

Eu-Tteum Baek and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
123 Cheomdangwagi-ro Buk-gu, Gwangju 61005, Republic of Korea
E-mail: {eutteum, hoyo}@gist.ac.kr

***Abstract*— In this paper, we propose a temporal stereo disparity estimation method. Conventional stereo disparity estimation methods rely on matching costs regarding computation of intensity or position similarities. However, most applications do not consider the temporal dimension when estimating the disparity. In other words, previous approaches disregard potentially useful disparity information that is already estimated. Therefore, we exploit reasonable temporal disparity information for accurate disparity map estimation. First, we calculate the optical flow between two color frames. Then, the disparity from the previous frame becomes reliable information, shrinking the value of candidate disparity. Consequently, the proposed technique increases the temporal consistency of estimated disparity maps and outperforms per-frame methods when image noise is present.**

I. INTRODUCTION

The research on stereo disparity estimation has a long history and has been a prevalent topic in computer vision. Numerous state-of-art approaches have been addressed recently. Stereo vision is highly important in applications such as 3D object recognition, extraction of information from aerial surveys, geometry extraction for 3D building mapping, and feature detection.

Depth information can be acquired by several methods such as active depth cameras, passive depth cameras, and hybrid depth cameras. Active depth cameras acquire depth information with a physical sensor [1], whereas passive depth cameras measure correlation of images captured from two or more cameras [2]. Hybrid depth cameras associate two methods to generate more accurate depth data and to cover their weaknesses [3]. Active depth cameras and hybrid depth cameras ensure more accurate depth information than passive depth cameras, provide depth data much faster than passive depth cameras. However, Active and hybrid depth cameras can use only low-resolution images due to hardware limitations. Therefore, we focus on estimating a depth map with passive depth cameras.

Stereo disparity estimation algorithms can be categorized into global and local methods. Given two images, local methods calculate the disparity. It only depends on image intensity and color values within a window. The disparity with the minimum aggregated cost is selected after aggregating the cost over the window. Common local costs function include the sum of absolute differences (SAD), the sum of squared differences (SSD), normalized cross correlation (NCC), and the census transform [4].

Global methods consider stereo disparity estimation as a labeling problem where the pixels of the reference image are nodes and the estimated disparities are labels. The problem gets resolved by global optimization techniques such as dynamic programming [5], graph cuts [6], belief propagation [7], and semi-global matching [8]

The goal of this work is to enhance the matching quality by using global optimization techniques and previous disparity information. To use suitable disparity information, we apply an optical flow technique to estimate displacement vectors for each pixel in two frames of a video. When calculating energy function, the previous disparity information is multiplied by weights.

II. MOTION ESTIMATION

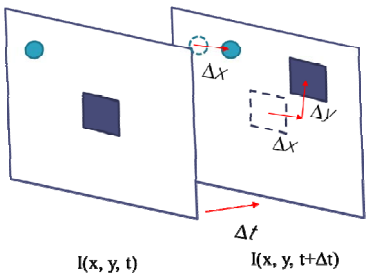


Fig. 1 Optical flow.

In order to use reasonable disparity data, we need to estimate motion difference of each pixel between two frames. Therefore, we use the motion estimation methods to find motion difference of each pixel. The optical flow methods calculate the motion difference from one frame to the next. Figure 1 illustrates the optical flow and 2D velocity.

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \quad (1)$$

where I(x, y, t) is the image intensity at time t, and (Δ x, Δ y) is the 2D velocity. Taylor series expansion yields

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t + Higher\ order\ terms \quad (2)$$

From (1) and (2), it follows that

$$\frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t = 0$$
$$\frac{\partial I}{\partial x}V_x + \frac{\partial I}{\partial y}V_y + \frac{\partial I}{\partial t} = 0 \quad (3)$$

where V_x, V_y are the x and y of the optical flow of I(x, y, t), and $\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}, \frac{\partial I}{\partial t}$ are the partial derivatives of the intensity I.

$$I_x V_x + I_y V_y = -I_t$$
$$\nabla I^T \cdot \vec{V} = -I_t \quad (4)$$

where I_x, I_y, and I_t are the derivatives of the image at (x, y, t). when ∇I^T in (4) is rank deficient one cannot solve for $\vec{V}$. This is called the aperture problem which is a problem of under determination that arises in motion perception. Given some additional constraint, we need for another set of equations to estimate the optical flow. In this paper, we employ a pyramidal Lucas-Kanade flow determination. To calculate Lucas-Kanade method, we select small regions in the whole image [9-10].

III. Temporal Stereo Disparity Estimation

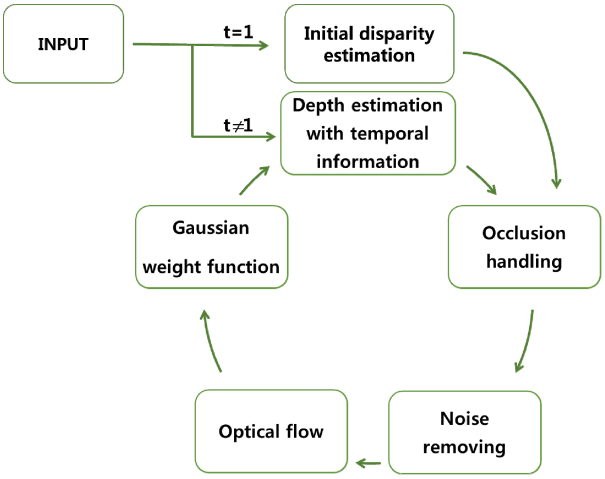


Fig. 2 Overall framework of out method.

The goal of the proposed method is to improve the conventional stereo disparity estimation methods using the temporal information. Fig. 2 shows the overall framework of the proposed algorithm. First, initial disparity is obtained by optimizing via graph cuts. Using the initial disparity map, we apply occlusion handling and noise filtering. After finding optical flow, estimate disparity estimation with Gaussian weight function. Lastly, we generate a temporal disparity map by following the process.

A. Initial Disparity Estimation

The stereo problem in terms of the MRF as the following energy function.

$$E(d) = \sum_s D_s(d_s) + \sum_{s,t \in N(s)} S_{s,t}(d_s, d_t) \quad (5)$$

where $D_s(\,)$ is the data term and $S_{s,t}(\,)$ is the smoothness term. d_s represents disparity or label for pixels. The data term is the term for how well the pixels match up for different disparities and is generally defined by intensity consistency of pixel correspondences for hypothesized disparity. The matching cost as data term is represented as

$$D_s(d_s) = \min(SAD(x, y, d), T_d)$$
$$SAD(x, y, d) = \sum_{i,j \in W} |I_r(x+i, y+j) - I_t(x+i+d, y+j)| \quad (6)$$

where I_r, I_t are the reference and target images, and d is disparity. T_d is the truncation value to control the limitation of the data cost. The smoothness term is the term which neighboring pixels have similar disparities. The smoothness term is defined as

$$S_{s,t}(d_s, d_t) = \min(\lambda |d_s - d_t|, T_s) \quad (7)$$

where T_s is the truncation value to constrain the high cost increase. The smoothness weight λ is generally represented by a scalar constant.

B. Disparity Refinement

Occlusion is an important and challenging problem in stereo depth estimation. The simplest method for occluded pixel detection and disparity estimation uses cross-checking [11]. For each pixel, cross-checking tests the consistency of disparity values from left and right disparity maps, determining occluded pixels for occlusion handling. Using the estimated disparity, we apply cross-checking test to detect occlusion, and fill the occlusion. The occlusion handling equation is defined as

$$O(s, d) = \arg\min_t \frac{1}{dist(s,t)} \exp\left(-\frac{dif_{s,t}}{\sigma^2}\right)$$
$$dif_{s,t} = \sum_{c \in \{R, G, B\}} |I_c(s) - I_c(t)| \quad (8)$$

In order to enhance disparity, we use guided image filtering [12]. The filter weights $W_{p,q}$ are expressed as

$$W_{s,t} = \frac{1}{|w|^2} \sum_{k:(s,t) \in w_k} (1 + (I_s - \mu_k)(\Sigma_k + \varepsilon U)^{-1}(I_t - \mu_k)) \quad (9)$$

where |w| is the total number of pixels in a window w_k centered at pixel k, and ε is a smoothness parameter. $\sum_k$ and μ_k are the covariance and mean of pixel intensities within w_k. I_s, I_t and μ_k are 3 × 1 vectors, while $\sum_k$ and the unary matrix U are of size 3 × 3.

C. Temporal Disparity Estimation

Features or pixels are moving in the input sequences. Therefore, it is inappropriate to use the disparity information about the same position directly. We employ the pyramidal Lucas-Kanade method to estimate displacement vectors for each pixel in two frames of a video. The MRF energy function of the temporal stereo problem is represented as

$$E(d) = \sum_s g(x) D_s(d_s) + \sum_{s,t \in N(s)} S_{s,t}(d_s, d_t)$$

$$g(x) = 1 - \frac{1}{\sqrt{2\pi\sigma^2}} exp^{-\frac{(x-d_{t-1})^2}{2\sigma^2}} \quad (10)$$

where g(x) is the Gaussian weight function which indicates how important a disparity is for temporal disparity estimation. d_{t-1} is the correspondent disparity of the previous frame. Fig. 3 shows the Gaussian weight function.

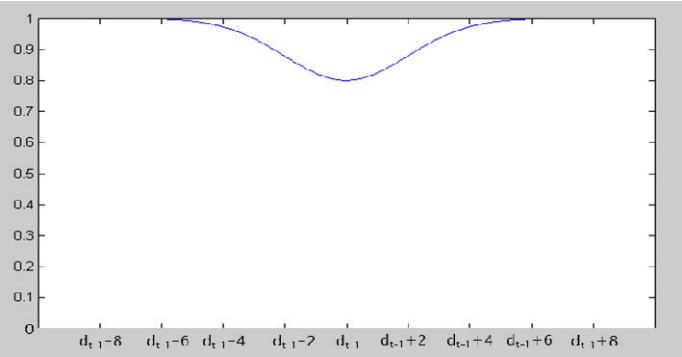
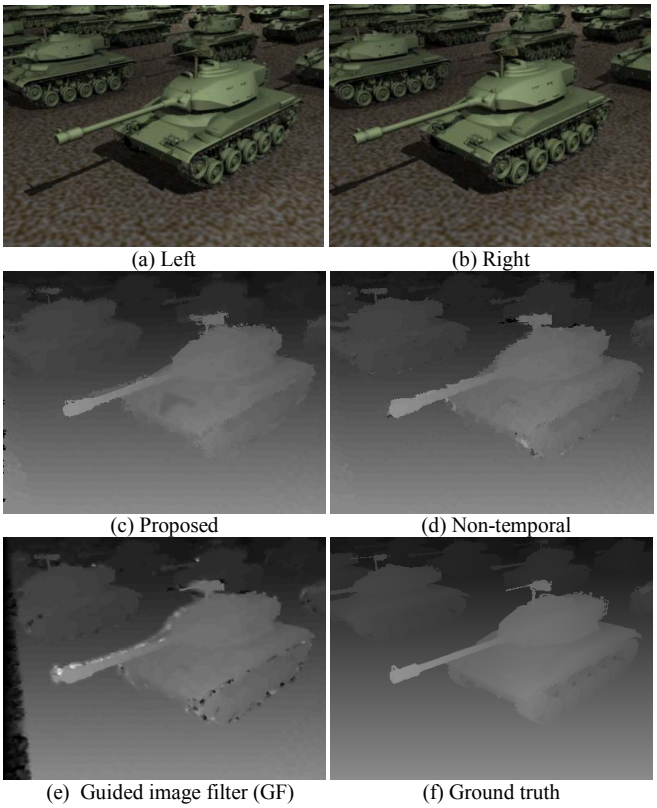


Fig. 3 Gaussian weight function.

IV. Experimental Results

In order to evaluate our proposed method, we use a synthetic dataset comprising five stereo sequences with known ground truth disparity, provided by [13]. We compare our proposed method to both the frame-by-frame and the spatio-temporal methods.

First, we subjectively compare the results of our proposed method with those of the frame-by-frame method by using graph cuts [6]. Fig. 4 represent the results of the proposed method and those of the spatio-temporal algorithms. The results generated by our proposed method are temporally coherent and exhibit less artifacts than the disparity maps generated by the method applied in a frame-by-frame manner.



(a) Left (b) Right

(c) Proposed (d) Non-temporal

(e) Guided image filter (GF) (f) Ground truth

Fig. 4 Comparison of proposed and spatio-temporal. (a) and (b) are Left and Right input frames of stereo sequences. (c) is the result of the proposed method. (d) is the result of the spatio-temporal method using hole-filling without Gaussian weight function, and (e) is the results of the spatio-temporal methods using GF without Gaussian weight function . (f) is the ground truth disparity.

TABLE I
COMPARISON OF MEAN ABSOLUTE DIFFERENCE OF SEQUENCE (TANKS)

No	Proposed	GF	Non-temporal	Semi-global
1	2.717	3.644	2.717	11.643
2	2.700	3.645	2.728	11.652
3	2.722	3.640	2.746	11.6491
4	2.766	3.657	2.753	11.6675
5	2.782	3.681	2.776	11.7343
6	2.779	3.718	2.805	11.7661
7	2.773	3.750	2.839	11.9887
8	2.809	3.806	2.885	12.1494
9	2.874	3.865	2.942	12.3425
10	2.907	3.910	2.989	12.6568
11	2.987	4.003	3.057	12.961
12	3.081	4.087	3.160	13.3963
13	3.182	4.206	3.253	13.794
14	3.250	4.279	3.331	14.0393
15	3.357	4.363	3.432	14.3933
16	3.410	4.432	3.495	14.7737
17	3.463	4.504	3.556	15.0365
18	3.447	4.520	3.599	15.3484

To evaluate the performance objectively, we compared the proposed method with the spatio-temporal algorithms. We use mean absolute difference (MAD) with respect to the ground truth disparity sequences. We also obtained the superior results except a few cases, as shown in Table I. Fig. 5 represents that our results are closer to the ground truths.

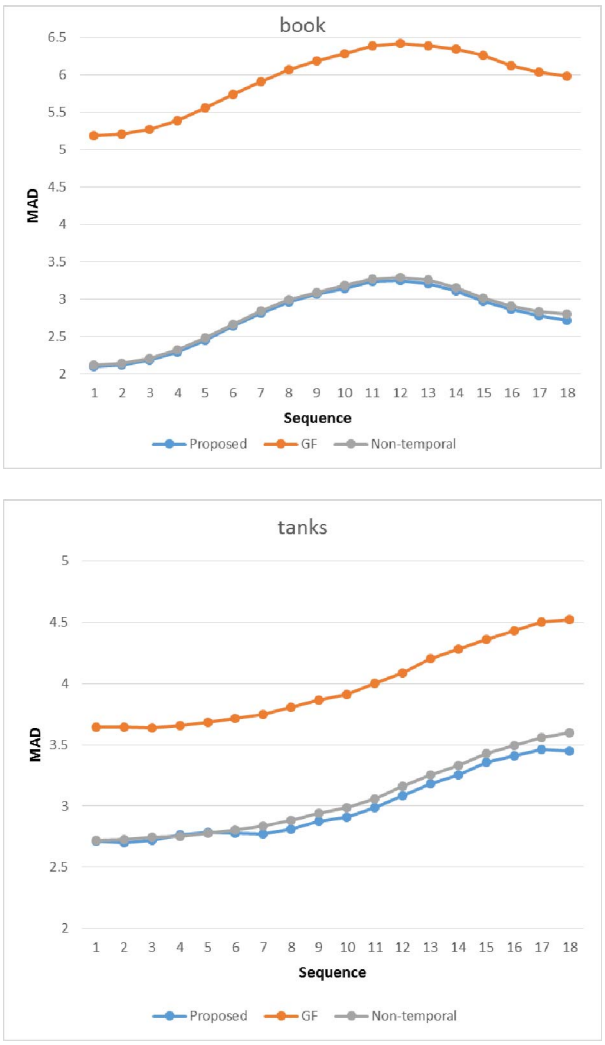


Fig. 5 Comparison of the proposed with spatio-temporal methods.

V. CONCLUSIONS

This paper proposed a method for estimating temporally coherent disparity maps from a sequence of stereo images. The proposed method used the MAP-MRF model to define an energy function considering spatial and temporal information. After optimizing the energy function via graph cuts, we applied occlusion handling and noise filtering to enhance the accuracy of disparity. From the experimental results, we have confirmed that the proposed method efficiently estimated the disparity in the temporal domain. In addition, the results produced by the proposed method outperformed conventional frame-by-frame methods in terms of mean absolute differences.

ACKNOWLEDGMENT

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] A. Frick , F. Kellner , B. Bartczak and R. Koch, "Generation of 3D-TV LDV-content with time of flight camera," IEEE Int'l Conf. on 3DTV, pp. 45–48, 2009

[2] W.S. Jang, Y.S. Ho, "Efficient disparity map estimation using occlusion handling for various 3D multimedia applications," IEEE Trans. Consumer Electronics, vol. 57, no. 4, pp. 1937–1943, 2011.

[3] E.K. Lee, Y.S. Ho, "Generation of high-quality depth maps using hybrid camera system for 3-D video," J. Visual Comm. Image Represent, vol. 22 no. 1, pp. 73–84, 2011.

[4] R. Zabih and J. Woodfill. "Non-parametric local transforms for computing visual correspondence," European Conf. Computer Vision, pp. 151–158, 1994.

[5] S. Birchfield and C. Tomasi. Depth discontinuities by pixel-to-pixel stereo. Int. J. Computer Vision, 35(3):269–293, 1999.

[6] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," Int'l Conf. Computer Vision, pp. 508–515, 2001.

[7] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations," Exploring Artificial Intelligence in the New Millenium, pp. 239–269, 2003.

[8] H. Hirschmueller, "Stereo vision in structured environments by consistent semi-global matching," Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 2386–2393, 2006.

[9] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application in stereo vision," Int'l Joint Conf. on Artificial Intelligence, pp. 674–679, 1981.

[10] J. Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker: Description of the Algorithm," OpenCV Document, Intel, Microprocessor Research Labs, 2000.

[11] G. Egnal and R. Wildes, "Detecting binocular half occlusions: empirical comparisons of five approaches," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.24 no. 8, pp. 1127-1133, 2002.

[12] K. He, J. Sun, and X. Tang, "Guided image filtering," European Conf. Computer Vision, pp. 1–14, 2010.

[13] C. Richardt, D. Orr, I. Davies, A. Criminisi, and N. A. Dodgson. "Real-time spatiotemporal stereo matching using the dual-cross-bilateral grid," European Conf. Computer Vision, pp. 6311-6316, 2010.

[14] H. Hirschmüller, "Stereo Processing by Semiglobal Matching and Mutual Information," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30, no. 2, pp. 328-341, 2008.

[15] J. Kowalczuk, E. Psota, and L. Perez, "Real-time temporal stereo matching using iterative adaptive support weights," Proc. IEEE International Conference on Electro/Information Technology, pp. 1-6, 2013.