Smoothing Filter with Adaptive Weights for Disparity Estimation

Ji-Hun Mun, Yo-Sung Ho
Gwangju Institution of Science and Technology (GIST)
School of Electrical Engineering and Computer Science (EECS)
Gwangju, Republic of Korea
jhm@gist.ac.kr , hoyo@gist.ac.kr

***Abstract*— In recent years, several stereo matching cost functions and cost aggregation methods have been developed to accurately estimate the disparity value. In this paper, we use the sum of absolute differences (SAD) and image gradient information to calculate the matching cost. For matching cost aggregation, we apply the guided image filtering; however, it provides different results depending on the guidance image. Since the input image has accurate pixel values in terms of the guidance image, we can aggregate the cost value while preserving object boundary region. However, guided image filtering is executed with pre-defined parameters, such as spatial and range variances value; thus, the filtered result has user dependency characteristics. In this paper, we propose a new edge detection method using smooth filtering to diminish the user dependency while preserving the object boundary region. On the contrary to raw input image, smoothed image can extract the edge information without unnecessary information of the object region. Based on the extracted edge information, we apply an adaptive weighting factor especially on the edge region for matching cost aggregation. We apply different weight factors on the extracted edge region and combine them with the general guided filtering result for the final cost aggregation result. Our experimental results show that the proposed method has a better performance than other filter based disparity estimation methods.**

Keywords—Real time stereo matching; Error propagation; feature points; Temporal domain

I. Introduction

Generating dense disparity map techniques have been developed for several years. The disparity map from stereo images is very critical issue in computer vision, since it works as an essential part in many computer vision applications, such as a 3D reconstruction, obstacle detection, and multi-view generation. The main problem is finding a correct corresponding pixel between rectified stereo images. To solve that kind of issues, large number of methods have been proposed. The main purpose of proposed technique is finding a disparity value with high accuracy and low computational complexity. Generally stereo matching strategies are classified into two types [1]; local and global methods. Local stereo matching method uses a predefined kernel to find a disparity value between stereo images. Even though the local method can estimate a disparity value within short execution time, it is very vulnerable to the window size. When the kernel size is small, it has high computational complexity and vice versa. Additionally, depending on the disparity search range, the estimated disparity map quality can be changed [2]. The global matching method does not use a kernel for disparity estimation, but it considers all the pixel values in stereo images. Because of that procedure, global disparity estimation result shows accurate disparity value than the local strategy result.

To handle the conventional problems, which occurred by image noise and blurring effect in estimated disparity map, this paper proposes a smoothing filter based disparity estimation technique. Basically, we follow the traditional stereo matching procedure as insisted in [1] as follow; 1) computing an initial cost value for each pixel; 2) aggregating an initial cost values along with the support region; 3) determine the final disparity values at each pixel; 4) Applying a post-processing to remove the remaining artifacts or errors in the final disparity map. Especially, at the initial cost aggregation Step 2, we focus on the edge preserving cost aggregation strategy. Because the local matching method computes a cost value using predefined kernel size, the object edge information is not considered when computes an initial disparity cost value. Since the edge region provides important information, it has to be reserved in the disparity map.

Estimating an accurate disparity value around the edge region is difficult because of the insufficient stereo image information. Usually the edge region, which across each objects disparity value is inaccurately estimated. Since the occlusion area affects to the disparity estimation procedure, many object boundary area has a different disparity value. In this paper to solve that problem, we smooth the input image using Gaussian and joint bilateral filter. Based on that filtered input image, we extract an edge region. The extracted edge region has clearer object boundary region than non-smoothed image based edge detection result. Then, by using the guided image filtering [3] the estimated cost is aggregated. Since the guided image filtering can aggregate the cost value using a guidance image, the edge information is well preserve than other conventional cost aggregation approaches. Additionally, we apply the different weighting value while performing the cost aggregation procedure except for non-edge region. As a result of that we can estimate an accurate disparity map especially improved object boundary region.

A. Related Works

Generally used raw stereo images contain many kind of noise such as radio metric differences and blurring artifacts. To compensate a radio metric error problem in a disparity map, background subtraction method is developed [4]. Additionally, the median filtering is widely used as a basic filtering to remove the noise factors in raw image. The guidance image based image filtering method is developed, which uses the raw image as a guidance image. The guided image filtering [3] performed well especially preserving object boundary region. For this kind of advantage, the guided image filtering is widely used many image processing algorithms, e.g., cost aggregation [5, 6] in Step 2. If error contained raw image was used as a guidance image, then the quality of filtering result is degraded. Herein, we adopt the new smooth filtering to suppress the small scale factors in raw image.

The initially calculated matching cost value from Step 1 is aggregated in Step 2 using a cost aggregation method. Usually a cost aggregation method assumes that, homogeneous regions have little difference disparity values and their neighbor pixels. Because of that reason, the major homogeneous regions share their disparity values after the cost aggregation. The cost aggregation method types are generally divided into three types: kernel based method, filter based method and segment based method. In this paper, we adopt a filter based cost aggregation method, especially the guided image filtering based aggregation technique [7]. Estimating an accurate disparity value is quite important issues especially near the object boundary regions. To generate an improved disparity map, we apply the weighted cost aggregation method.

B. Contributions

We estimate a disparity map while following the general disparity map generation procedure as indicated in [1]. Many powerful disparity estimation techniques are provided in each disparity generation steps. However, we only focus on Step 1 and Step 2 to improve the quality of disparity map. In Step 1, we apply the smooth filtering to remove small scale factors. That factors affect to disparity cost computation even though it is not important factors while generating a final disparity map. The final goal of this smooth filtering is preserving the object boundary region in the disparity map. Additionally, we propose the weighted cost aggregation method based on guided image filtering. Basically the guided image filtering method well preserve the edge region while aggregating cost values. However, by applying adaptive weighting method at the cost aggregation stage, we can improve the disparity map quality. The weighting factors only affect to extracted edge region and the other region is not affected by weighting factors.

II. Matching Cost Model and Smoothing Filter

A. Mixed Matching Cost Model

The stereo matching methods mainly divided into two types: local and global strategies. Local approach uses kernel for computing a matching cost between stereo images. To measure the difference between corresponding pixels, sum of absolute differences (AD) is applied in our stereo matching system. AD is easy to implement and the color difference help to reduce the matching ambiguities in repetitive texture region. Because of those several advantage, AD is commonly used in local stereo matching procedure. The AD matching is performed with three RGB channel and considers the pixel value within the pre-defined kernel *N* as indicated in (1).

$$\mathrm{Cost_{AD}}(\mathrm{x,y}) = \sum_{(i,j)\in N_{(x,y)}} |I_L(i,j) - I_R(i+d,j)| \qquad (1)$$

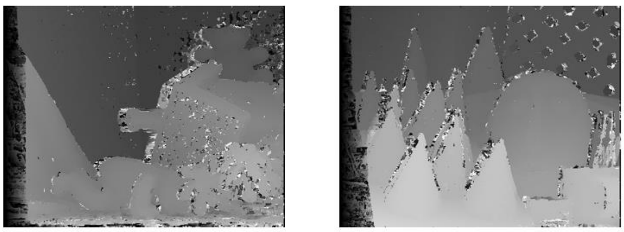
Since the AD matching procedure does not consider the object boundary information, AD based disparity map has inaccurate disparity value especially near the object boundary regions. To compensate the weakness of AD matching method, image gradient information is mixed with AD matching function. The image gradient contains plentiful structural information and it is not affected by illuminance variation condition. The gradient information is taken from the intensity of input image, and obtaining a gradient based cost matching procedure is represented in (2).

$$\mathrm{Cost_{Grad}}(\mathrm{x,y}) = \sum_{(i,j)\in N_{x|(x,y)}} |\nabla_x I_L(i,j) - \nabla_x I_R(i+d,j)| + \sum_{(i,j)\in N_{x|(x,y)}} |\nabla_x I_L(i,j) - \nabla_x I_R(i+d,j)| \qquad (2)$$

Instead of using each matching cost method (AD, gradient), combined matching cost function measurement achieves better performance. Thus, we applied combined cost approach for initial matching cost. The combined matching cost function is composed with weighting parameter α.

$$Cost(\mathrm{x,y}) = (1-\mathrm{\alpha}) \cdot C_{SAD}(x,y) + \alpha \cdot C_{Grad}(x,y) \qquad (3)$$

The initially estimated disparity map using combined matching cost is indicated in Fig. 1. Even though we adopt the gradient information, the edge region pixel values in estimated disparity map is still inaccurate.



Teddy Cones

Fig. 1. Initial disparity map with combied cost function

B. Smooth Filtering

A Texture region contains typically oscillated small-scale factors. Since the small-scale factors have different color channel values with their neighbor pixel values, initially estimated disparity map has some errors in most of texture regions. When those texture regions are very close to the object boundary region, the estimated disparity value is not accurate. Since the cost aggregation with this kind of errors derives an improper cost aggregation result, we apply the smooth filtering to extract a clear edge region from raw image.

Most of all, removing the small-scale factor is important issues in smooth filtering. Such that, we apply the Gaussian filtering as a first step. The Gaussian filtering is widely used to remove a noise from raw input image. We express this operation with input image I and output image G as shown in (4).

$$G(\mathrm{p}) = \sum_{q \in N(p)} \exp\left(-\frac{\|p-q\|^2}{2\sigma_s^2}\right) \cdot I(q) \tag{4}$$

where p and q indicate index pixel in input raw image, σ is the standard deviation and $N(p)$ represents the set of neighboring pixels of p. The Gaussian filter completely removes the small-scale factors, which smaller than σ as indicated in (4). This function efficiently implemented by separating kernels in perpendicular directions. After the Gaussian filtering, the raw image no longer contains small-scale factors, but it still has burring problem not only homogeneous region but also the object boundary region as shown in Fig. 2 (a).

Fig. 2. Filtering results for image smoothing. (a) Gaussian filtering, (b) Joint bilaterl filtering on (a) result.

To compensate the burring artifact in smooth filtering result, we apply the joint bilateral filter on the Gaussian filtered image. Since the joint bilateral filter is traditionally used in cost aggregation and filtering method to preserve the edge region, we apply this filter on the blurred image. The spatial and range values of input and Gaussian filtering result image are computed to recover the edge region. In this case, the original input raw image was used as a guidance image as formulated in (5).

$$J(p,q) = \sum_{q \in N(p)} \exp\left(-\frac{\|p-q\|^2}{2\sigma_s^2} - \frac{\|I(p)-I(q)\|^2}{2\sigma_r^2}\right) \tag{5}$$

where I indicates the input raw image, σ_s and σ_r represent the spatial and range standard deviation respectively. From (5), we notice that the input I is guided by the structure of overall function. The joint bilateral filtering result is displayed in Fig. 2 (b). When it compares with the Fig. 2 (a), the burring artifacts are removed and such that the object boundary region is more clearly represented in Fig. 2 (b). From this result, we can get a noise free image especially in the object inside area. Simultaneously, the edge information is also well preserved.

III. Cost Aggregation with Adpative Weight

The initially measured cost value from (3) is usually contains noise, which influence the final disparity map. To refine the initial matching cost value, the guided image filtering is conventionally used [4, 5]. The set of matching image slices $L = \{1,2,\ldots \mathrm{N}\}$, which derived by applying the winner-takes-all (WTA) from initial matching cost image volumes. The cost volume L slices are filtered to derive a more precise matching cost value such as (6).

$$C'_{i,j} = \sum_j W_{i,j}(I) C_{i,j} \tag{6}$$

where $C_{i,j}$ is a set of slice, $C'_{i,j}$ indicates a filtered matching cost volume and $W_{i,j}$ represent the cost filtering weight. Herein, we adopt the guided image filtering for filtering weight as described in (7).

$$W_{i,j} = \frac{1}{|w|^2} \sum_{K:(i,j)\in w_k} \left(1 + (I_i - \mu_k)^T \left(\textstyle\sum_k + \epsilon U\right)^{-1} (I_j - \mu_k)\right) \tag{7}$$

In (7), μ_k and $\sum_k$ are the mean vector and covariance of input image I in a squared window w_k centered at pixel k respectively. U denotes the intensity matrix. The other parameter $|w|^2$ is number of pixels in the window and ϵ represents the smoothness factor. Herein, we propose a weighting function while performing the guided filtering based cost aggregation. Firstly, we extract the edge information from the smooth filtered image. As indicated in Fig. 3 (b), the critical object edge regions are clearly extracted than non-filtered image edge result in Fig. 3 (a).

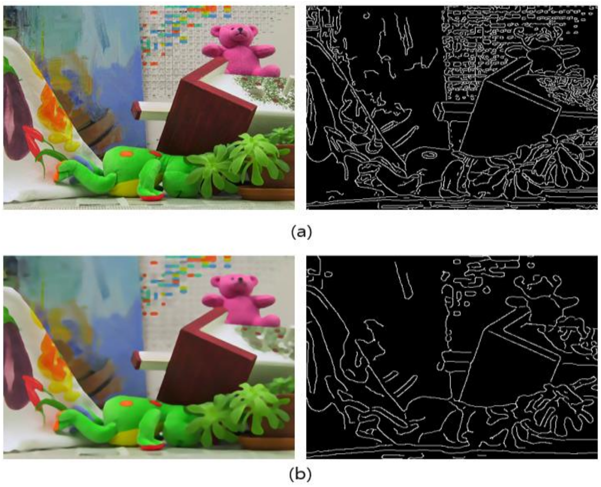


Fig. 3. Extracted edge region from raw input. (a) raw input image and extracted edge (b) smoothed input image and extracted edge.

The weight factor is applied to only on the extracted edge region, and the other region performs the guided image filtering for cost aggregation as indicated in (8).

$$C_{Agg} = \sum_{i=1}^{disp} \alpha \cdot G(I, C_{vol(i),Edge}) + (1-\alpha) \cdot G(I, C_{vol(i)}) \tag{8}$$

When the extracted edge region has same corresponding pixels in the initial matching cost, then that pixel adopt the Gaussian filtering to enhance that pixel value while performing the cost aggregation. Such that, the mean(μ_k) and covariance($\sum_k$) values are changed in guided image filtering procedure. Since we apply the Gaussian filtering on the corresponding pixel value with the edge information, a neighbor pixel values are smoothed than that pixel value.

IV. Experiment Results and Evaluation

We conducted experiments using *Middlebury* test sequences to verify an efficiency of the proposed disparity estimation method. The test benchmark contains four image sets, such as *Teddy*, *Cones*, *Venus*, and *Tsukuba*. The performance of each disparity map is evaluated using bad pixel rate (BPR) value. Since the proposed method focus on the edge preserved disparity estimation, we also evaluate two regions; whole(*all*) and discontinuities(*disc.*) region. The disparity results are shown in Fig. 4. To evaluate the performance of proposed method, we compare other filtering based stereo matching results.

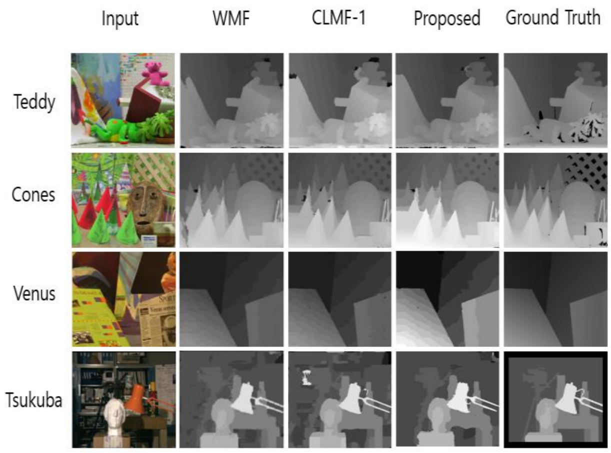


Fig. 4. Experiment result with *Middlebury* test sequencecs

We can get more accurate disparity results, when compare to the proposed result with other conventional methods. Especially, in test sequence *Tsukuba*, the background of objects is accurately estimated than other results. The other test sequences also have improved disparity quality near the object boundary regions. To numerically evaluate the performance of proposed method, Table 1 shows the BPR comparison results. Overall improvement of proposed method is about 1.2% than other conventional methods. The weighted median filtering (WMF) and cross-based local multipoint filtering (CLMF) methods are used to smoothing the input raw image for edge extraction. However, the mixed Gaussian and JBF method shows better edge extraction result than other smoothing filters. Because of that reason, the proposed adaptive weighting factors are quite well applied while performing the cost aggregation. By applying the weighting factor, the object boundary information preserving performance is improved than original cost aggregation method.

TABLE I. BPR comparison results

Sequence	WMF		CLMF-1		Proposed	
	all	***disc.***	***all***	***disc.***	***all***	***disc.***
Teddy	13.86	14.75	13.19	14.83	11.54	12.21
Cones	15.22	16.73	14.93	15.47	12.81	13.04
Venus	3.37	4.92	2.86	3.88	2.63	4.39
Tsukuba	5.45	6.73	5.15	6.51	4.28	6.10

V. Conclusion

For edge preserved disparity estimation, we propose the smooth filtering based adaptive weighting method. We basically follow the stereo matching procedure. Among that procedure, we especially concentrate on pre-processing and cost aggregation step. The small-scale factors, which in the raw input image influence to the computation of initial matching cost. To prevent the influence of small-scale factor, the Gaussian and joint bilateral filter based image smoothing is performed. Based on that smoothed image, we extract the strong object boundary edge information. Since the smoothing filtering remove the small-scale factor, the extracted edge does not contain tiny edge regions. While aggregating the matching cost, we apply the adaptive weighting factors only on the extracted edge pixels. By assigning a weight on the edge region while performing the cost aggregation, the final disparity map has accurate value than other conventional method results.

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). [GK13C0100, Development of Interactive and Realistic Massive Giga-Content Technology]

References

[1] D. scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IEEE Workshop on Proceeding in Stereo and Muti-Baseline Vision*, vol. 47, no. 1, pp. 7-42, Aug. 2002.

[2] J.H. Mun and Y. S. Ho, “Temporally consistence depth estimation from stereo video sequences,” *PCM*, pp. 579-588, Sept. 2015.

[3] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on PAMI*, pp. 1397-1409, Oct. 2012.

[4] A. Ansar, A. Castano, and L. Matthies, “Enhanced real-time stereo using bilateral filtering,” *Proceeding on 3DPVT*, pp.455-462, Sep. 2004.

[5] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," *IEEE Transaction on PAMI*, pp. 504-511, Aug. 2012.

[6] Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu, "Constant time weighted median filtering for stereo matching and beyond," *IEEE International Conf. on ICCV*, pp. 49-56, Dec. 2013.

[7] C. C. Pham and J. W. Jeon, “Domain transformation-based efficient cost aggregation for local stereo matching,” *IEEE Transaction on CSVT*, vol. 23, no. 7, pp. 1119-1130, Jul. 2013.