15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

High Accuracy Local Stereo Matching using DoG Scale Map

Masamichi Kitagawa
Tokyo Univ. of Agric. and Tech.
h, Koganei-shi, Tokyo
kitagawa@m2.tuat.ac.jp

Ikuko Shimizu
Tokyo Univ. of Agric. and Tech.
2-24-16 Naka-cho, Koganei-shi, Tokyo
ikuko@cc.tuat.ac.jp

Radim Sara
Czech Technical University in Prague
Karlovo namesti 13,121 35 Praha 2,Czech Republic
sara@cmp.felk.cvut.cz

Abstract

Local matching is one of approaches for stereo matching which needs cost aggregation. In Guided Filter based method proposed by Hosni, the cost map is smoothed by Guided Filter using original image as a guiding image. However, the Guided Filter sometimes fails when there are regions whose textures are same but disparities are different. Thus, parameter tuning for filter size of Guided Filter is difficult to obtain the best accuracy. In this paper we propose an algorithm for automatic filter size selection for each pixel of Guided Filter based stereo matching based on the response of the Different of Gaussian (DoG). In our algorithm, we generate the Filter-Size map whose pixel value for each pixel is appropriate filter size. The value of the Filter-Size map is the largest size of the filtering area around the pixel in interest calculated such that more than two edges are not included in filtering area. In our experiments, we evaluated accuracy of Guided Filter based method with our algorithm for selecting filter size compared with the original Guided Filter based method without our algorithm. By using the Middlebury datasets, the experimental results shows our algorithm's superiority in accuracy.

1 Introduction

In recent years, a lot of realtime applications using 3D geometry are developed. Therefore, stereo matching technique becomes more and more important. In computer vision research area, many stereo matching methods have been proposed in the literature. Local stereo matching methods are suitable for realtime applications because they requires much less computational time than that of global stereo matching methods in general.

Stereo matching methods consist of four steps[1, 2] : (1) cost computation, (2) cost aggregation, (3) disparity computation, and (4) optimization, disparity refinement. In local matching methods, cost computation and cost aggregation are very important. In cost computation, matching cost for each pixel with each disparity is evaluated by Sum of Squared Distance(SSD), Sum of Absolute Distance(SAD), Normalized Cross Correlation(NCC), and so on[3]. In this step, cost map is stored which contains computed cost for each pixel with each disparity. Then cost map is smoothed by some kind of filter in cost aggregation step. In the cost aggregation step, many filters were proposed[4, 5, 6] and the accuracy of the stereo matching depends on characteristics of the filter. Next, the disparity value which has minimum cost is assigned for each pixel in the disparity computation step. Finally, disparity is refined by checking the consistency between right and left images in the last step.

Among local matching methods, Guided Filter based method[4] is known as a method which gives better results than other local matching methods. Guided Filter is an edge-preserving filter[7]. In this method[4], each image itself is used as a guiding image and the disparity map is smoothed according to the guiding image. It means that weight for smoothing is larger if a pixel in interest has no discontinuity in its neighborhood, and if the pixel has a discontinuity in its neighborhood, weight for smoothing becomes smaller. Even though Guided Filter is is an edge-preserving filter, it sometimes fails to estimate the appropriate weight for the pixels in regions that include more than two edges. For a textureless area, a bigger filter size for the Guided Filter gives better results, while for an area with complex texture, if the filter size is too big, the weight for the smoothing is not estimated correctly. Therefore, it is very difficult to find an appropriate filter size of the Guided Filter.

In this paper, we propose an algorithm for finding the appropriate filter size for each pixel in Guided Filter based stereo matching methods. We generate a Filter-Size map for each image whose pixels are appropriate filter size for that pixel based on the responses of the Difference of Gaussian (DoG) filters. Using Filter-Size map, smoothing weight for the Guided Filter is correctly estimated in the cost aggregation step.

2 Related work

There are many stereo matching methods and their results have been compared in the literature[3][8]. Among these methods, Guided Filter based method[4] is one of the methods which give high accuracy among local stereo matching methods[9]. In addition, this method works in real-time because it is easy to parallelize it by GPGPU, and it is implemented by using C++ and CUDA[4]. Therefore, it is applicable to realtime applications.

Figure 1 shows an outline of Guided Filter-based method by Hosni[4] and our method. First, in the cost computation step, SAD of window size 1x1 (AD) and gradient of x-direction to neighboring pixel are used. The cost maps are computed based on the weighted sum of these two values for each pixel with respective disparities. Then, in the cost aggregation step, cost

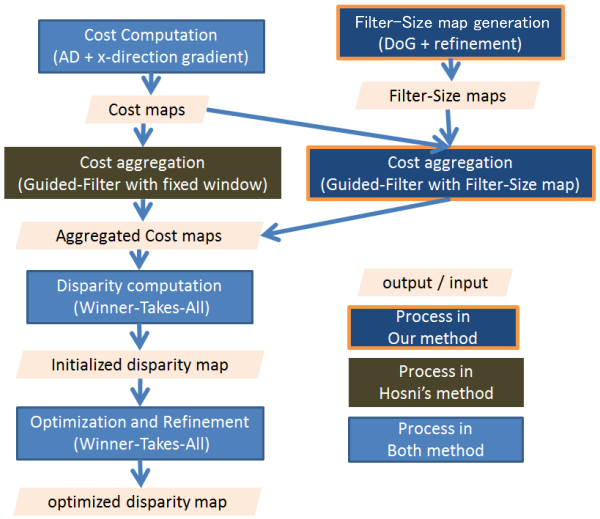

Figure 1. Outline of our method and Hosni's method.

maps are smoothed by fixed-sized Guided Filter. In the disparity computation step, disparity is assigned to each pixel by Winner-Takes-All strategy to aggregated costs and assigned disparities are verified based on the consistency of right and left images. Finally, verified disparity map is smoothed by the Bilateral-Filter in the optimization and disparity refinement step.

However, to obtain the best accuracy, this method requires parameters adjustment. Especially the size of Guided Filter is very important and it is different for each input image. The size of the Guided Filter should be changed depending on edge of texture in image. Therefore the best parameter is found by heuristic tuning by user.

The Guided Filter is one of edge-preserving smoothing filters. Its weight $W(i,j)$ of the filter is given by

$$W(i,j) = \frac{1}{|\omega|^2} \sum_{i \in \omega_k, j \in \omega_k} (1 + \frac{(I_i - \mu_k)(I_j - \mu_k)}{\sigma_k^2 + \epsilon}), \quad (1)$$

where I is a guiding image, i is a 2-dimensional index of the pixel in interest, j is index of the pixels in the neighborhood ω_k, μ is an average of pixel values in the neighborhood and σ is its variance. We call the size of ω_k "filter size".

Guided Filter-based stereo matching methods sometimes fail to estimate disparity when there are more than two edges between the pixel $I(i)$ and the pixel $I(j)$. Figure 2 is an example of the case when the window includes three different regions A, B, and C. The pixel in interest belongs to region A. The pixel values (Fig.2 (a)) of A and C are very similar but their disparity values (Fig.2 (c)) are quite different. Because the pixel values are similar, the weight $W(i,j)$ is large as shown in Fig.2 (b). When the filter is used for simple smoothing or denoising, it is not a problem. However, because disparity of region A and that of region C is different the weight is not appropriate for cost aggregation of stereo matching.

In general, large filter size is appropriate for smoothing for textureless regions. However, areas with many edges, filter size is better to be small for cost aggregation. In the conventional method[4], users have to find a filter size as large as possible such that the window does not include more than two edges. In our method, appropriate filter size is estimated automatically for each pixel. We find appropriate filter size for each pixel automatically by using DoG scale.

3 Our method

In Fig.1, outline of our method is shown. In cost aggregation step, we propose an algorithm for generating the Filter-Size map based on DoG scale. Each pixel in Filter-Size map describes appropriate size of the Guided Filter of each pixel in cost map. Appropriate filter size means the largest filter size such that the window does not include more than two edges.

We employ DoG scale to find the appropriate filter size for each pixel because it gives a size such that the region associated with the DoG scale contains one or zero edge in it. By using Filter-Size map, the weight using by the Guided Filter is computed appropriately when one edge is contained in considering area because such area has completely differences texture color usually.

In the followings, our algorithm for generating the Filter-Size map is described.

3.1 Initialization of Filter-Size Map

First, DoG is computed by differentiating output of two Gaussian filters with different variances. In this paper, we call DoG size which is defined by two Gaussian filters as kernel size. The kernel size is valuable in a scale space. we note that DoG value describe differences of pixel pattern between two Gaussian Filter. If DoG value become maximum or minimum in some scale space, it means that significant differences of pixel pattern occur for example an edge begins to be included. Then, we select the kernel size for each pixel which gives the maximum absolute value of DoG to obtain the scales which gives the maximum differences between two Gaussian filters in scale space. This value for each pixel is an initial value of Filter-Size map. Example of the initial values of Filter-Size map is shown in Fig. 3(b).

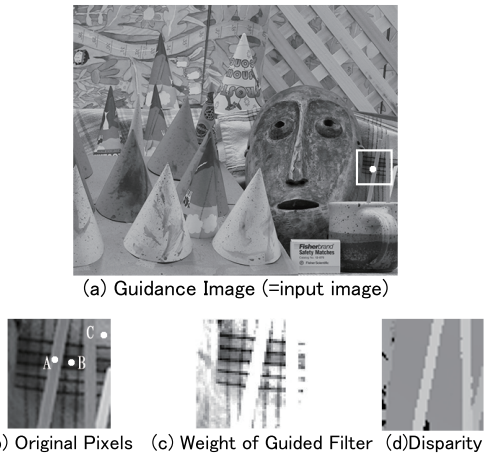

Figure 2. Example of failed to weight for a pixels.

3.2 Refinement of minimum value of Filter-Size map

In cost aggregation, the size of Guided Filter have to be as large as possible. Therefore, we would like to set the minimum value of the Filter-Size map is as large as possible.

However, if the window size is bigger than 5×5, it would contain more than two edges. On the other hand, if the windows size is 3×3, the number of edges in the area is one or zero. Thus the minimum value of the Filter-Scale map is 3. If there are some pixels with the initial values of the Filter-Size map are less than 3, the values is set to be 3.

3.3 Removing noises in Filter-Size map

Input image may have noisy pixels which have different pixel value in an around area even included in same texture region. If there are noisy pixels exceptionally in the region around the pixel in interest, the value of DoG may be affected by the noisy pixels. In such a case, the computed scale size becomes smaller than true scale size. To reduce such estimation errors, $E \times E$ window around each pixel is searched for the largest number in the window and the value of the pixel in interest is replaced with the largest number in the window. This example is also shown in figure 3(c).

4 Experiment

In this section, we show experimental results using Middlebury dataset[9]. We selected four images "cone", "teddy", "tsukuba", and "venus" because the results for these images by Hosni ' s method are reported at the Middlebury evaluation web site[9]. Figure 4 shows original images of dataset. We used input images, the ground truth image and the occlusion maps in Middlebury dataset[2] for evaluation. In this experiment, we set our method parameter such as DoG kernel size is 1×1 to 65×65 and noise reduction window size $E = 5$ or 7

Figure 5 shows the Filter-Size map for each image. The darker pixel represent smaller filter size. It shows that the filter sizes of pixels in textureless area especially in the image "venus". On the other hand, the filter size of pixels in regions of rich texture and pixels around the edges are small.

Table 1 shows the percentages of error pixels. According to the Middlebury's web site[9], the number of pixels were counted whose disparity error is more than 1 pixel and 2 pixels respectively. Figure 6 shows error pixels of our method and that of Hosni ' s method[4]. In these images, pixels colored in black represent error

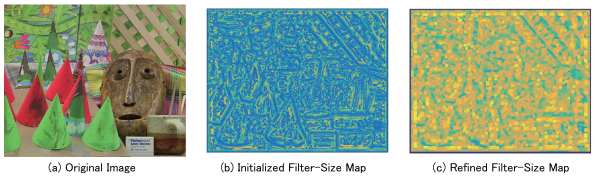

Figure 3. Example of Filter-Size map initialization and refinement.

Figure 4. Dataset images for this experiment from Middlebury.

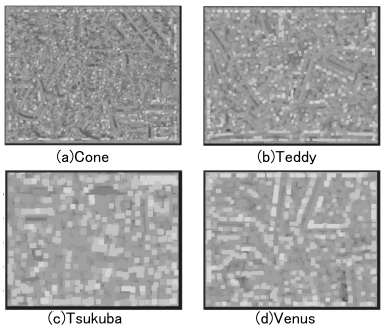

Figure 5. Filter-Size map of each image.

pixels for error > 1, and gray pixels means occluded area. As shown in Tab. 1, our method performs systematically better than Hosni's method for error > 1. Especially, our method succeeded to reduce its error more than 1% for the image "teddy" image because this image contains many edges. In addition, result on "cone" shows an 0.6% improvement in accuracy. As shown in Fig.6, error pixels in rich texture regions are removed in our method. For example, error pixels in the left side in the image "cone" is removed because there are a lot of cones and those disparities are different completely although their texture color is similar. In the image "teddy", error pixels in lower part of image are removed, most importantly, errors around the edge of green doll's arm and around the paper on the floor are removed. In those areas, it is very difficult for Guided Filter to set the appropriate fixed filter size beforehand. If fixed window is used, window size have to be set in small size. But in our method, filter size becomes small for the pixel near edges, and it becomes larger if the pixel is far from edges. Thus, even if matching cost computation failed in an area far from an edge, our method assign a large filter size for such pixel and these noises are removed by smoothing.

On the other hand, the image "venus" has simple disparity structure and texture region is similar to disparity region. In such case, the Guided Filter with fixed filter size works well. Thus, accuracy of this image almost the same as the Hosni's method. In image "Tsukuba", there are textureless regions and narrow regions, for example, the orange arm of the lamp in right part of the image. Estimating disparity in narrow region is difficult for local matching. For the pixels

Table 1. Accuracy of our method and Hosni's method.

dataset	cone	teddy
Error >1	nonocc, all, disc	nonocc, all, disc
our method	2.19, 6.94, 6.49	5.80, 9.44, 14.2
Hosni's method	2.71, 8.24, 7.66	6.16, 11.8, 16.0
Error >2	nonocc all disc	nonocc all disc
our method	1.87, 5.88, 5.54	4.80, 6.69, 11.2
Hosni's method	2.12, 6.82, 6.37	4.44, 7.63, 11.4
dataset	tsukuba	venus
Error >1	nonocc, all, disc	nonocc, all, disc
our method	1.50, 1.76, 7.73	0.20, 0.53, 2.50
Hosni's method	1.51, 1.85, 7.61	0.20, 0.39, 2.42
Error >2	nonocc, all, disc	nonocc, all, disc
our method	1.13, 1.29, 5.75	0.19, 0.46, 2.38
Hosni's method	1.19, 1.50, 5.86	0.18, 0.32, 2.09

on the edges of such region, the values of different disparity pixels around the pixel are used for cost computation of x-gradient. Consequently, cost computation failed for such pixels. In addition, assigned filter size by our Filter-Size map is small because DoG scale is small in such region. As a result, pixels have correct matching cost are not included in smoothing region in cost aggregation. However, the accuracy of our method is higher than Hosni's method for the image "tsukuba".

As a whole, our method is effective for input images which have complicated edge and many different disparity region.

5 Conclusion

In this paper, we proposed a local stereo matching method using Filter-Size map for Guided Filter. The pixel value of the Filter-Size map is the appropriate filter size for the Guided Filter for each pixel.

It is difficult to set a fixed appropriate filter size for the Guided Filter because sometimes there are pixels which have the same color texture but different disparity. To overcome this problem, we propose an algorithm for estimating appropriate filter size for each pixel based on the DoG scale size. We generate Filter-Size map by using refined DoG scale map. The Filter-Size map is computed automatically. Small filter size is assigned for pixels around the edges and large filter size is assigned for pixels in the textureless area.

We compared our method with Hosni's method using Middlebury dataset. It showed that our method is better than or equal to Hosni's result in terms of accuracy. Especially, for the image "teddy" and the image "cone", the accuracy was much improved because there are a lot of areas which include different disparity but similar color area in these images.

Acknowledgement

This work was part supported by Grant-in-Aid for Scientific Research No.15X00445.

References

[1] Richard Szeliski, Computer vision: algorithms and applications, Springer Science & Business Media, 2010.

[2] Daniel Scharstein and Richard Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", International journal of computer vision, 2002.

[3] Heiko Hirschmuller and Daniel Scharstein, "Evaluation of Stereo Matching Costs on Images with Radiometric Differences", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1582–1599, 2009.

[4] Asmaa Hosni, Michael Bleyer, Margrit Gelautz, Christoph Rhemann, "Local stereo matching using geodesic support weights", The 16th IEEE International Conference on Image Processing, pp. 2093–2096, 2009.

[5] C. Cigla, "Recursive edge-aware filters for stereo matching", CVPR Embedded Vision Workshop, 2015.

[6] Jedrzej Kowalczuk, Eric T. Psota, and Lance C. Perez, "Real-time Stereo Matching on CUDA using an Iterative Refinement Method for Adaptive Support-Weight Correspondences", IEEE Transactions on Circuits and Systems for Video Technology Digital, Volume 23, Issue 1, pp.94–104, 2013.

[7] Kaiming He, Jian Sun, Xiaou Tang, "Guided Image Filtering", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 35, Issue 6, pp. 1397–1409, 2013.

[8] Deepika Kumaria, Kamaljit Kaurb, "A Survey on Stereo Matching Techniques for 3D Vision in Image Processing", International Journal of Engineering and Manufacturing, Vol.4, pp,40–49, 2016.

[9] `http://vision.middlebury.edu/stereo/eval/`

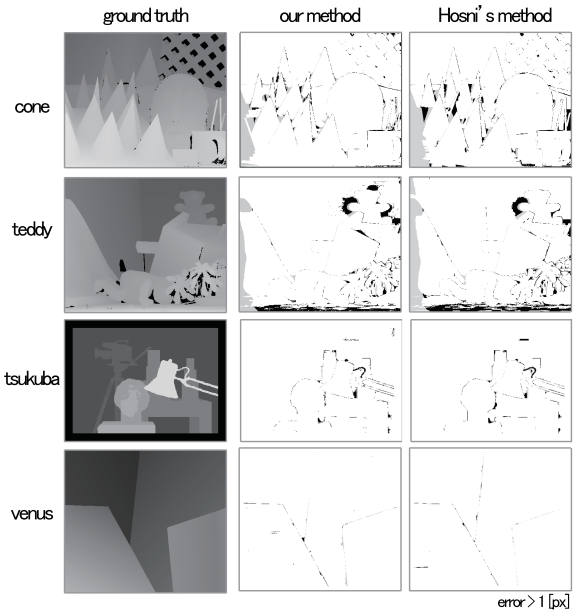

Figure 6. Error pixel map for each image.