Window-based Stereo Matching Algorithm Using a Weighted Average of Costs Aggregated with Window Size Reduction

Kan'ya Sasaki, Seiji Kameda, Hiroshi Ando, Mamoru Sasaki and Atsushi Iwata
Graduate School of Advanced of Matter, Hiroshima University,
1-3-1, Kagamiyama, Higashi-Hiroshima-shi 739-8527, Japan
Phone and Fax: +81-824-22-7358, E-mail: {kanya, kameda, ando, sasaki, iwa}@dsl.hiroshima-u.ac.jp

1. Introduction

The intensity-based stereo matching produces a dense disparity map by using pixel intensities in the two images of a stereo camera system and constructs a 3-dimensional depth structure. The 3-dimensional information is very useful to computer and robot vision. Therefore, many intensity-based stereo algorithms have been proposed and applied to various applications [1].

In the intensity-based stereo matching algorithm, window-based and coarse-to-fine algorithms are known as typical approaches. The window-based algorithm matches intensity values within windows between two stereo images. The conventional window-based algorithms, however, face a trade-off between error rates of the disparity map in disparity continuity and discontinuity regions due to the window size dependence. To solve the problem, the coarse-to-fine algorithm has been proposed [2-3]. The algorithm starts the matching process by the largest window size and gradually decreases the window size with narrowing a rage of candidates. The algorithm, however, often cannot find true disparities due to a limitation of candidates of disparity. To solve these issues, we proposed a new algorithm of the window-based stereo matching.

2. Window-based stereoscopic algorithm

A fundamental process of window-based algorithm is generally divided into four steps [4]; matching cost computation, cost aggregation, disparity computation and disparity refinement. The processing flow is explained below. The first step is a matching cost computation. The matching cost means a similarity between left and right pixel intensities in two stereo images. In the next step, the matching costs within a window are aggregated. Because the cost aggregated within the window (hereinafter called an aggregated cost) allows a comparison of texture and inhibition of noise component, there is a clear difference in the aggregated cost between true and fault disparities. In the disparity computation step, the best disparity is selected by comparing the aggregated costs across all disparities. The most simple and widely used disparity computation method is a winner-take-all (WTA) optimization [1]. The WTA finds a disparity when the aggregated cost is minimum value at each pixel position. The disparity of the minimum aggregated cost is defined as the best disparity. The disparity selected for every pixel position forms the disparity map. In the last step, the sub-pixel disparity refinement is computed by fitting a curve to the aggregated costs at discrete pixel units to increase a resolution of the disparity map [5-6].

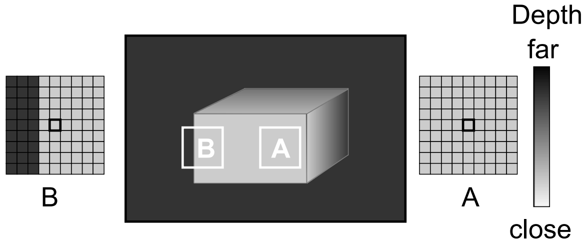


Fig. 1 Disparity map of a box, A: disparity continuity region, B: disparity discontinuity region.

3. Issue of the conventional algorithms

In the conventional window-based stereo algorithms, the optimal window size depends on variation in disparity value around a given pixel position. The dependence of the window size is explained below by using Fig. 1. Figure 1 shows a disparity map of a box. The disparity map is divided into two regions, disparity continuity region, A, and discontinuity region, B. The disparity continuity region, A, is defined as a region where the all disparities are same. In the disparity continuity region, larger window is desirable to avoid the noise influence. In contrast, the disparity discontinuity region, B, is defined as a region where some disparities are existed. If large window including some disparities is used, the incorrect disparity may be selected because the aggregated cost in the disparity is small. When small window is used, the correct disparity is normally selected. Thus, in the disparity discontinuity region, the small window is desirable to avoid including the different disparities.

The coarse-to-fine algorithm solves the issue by using multiple costs aggregated by various window sizes. However, the conventional coarse-to-fine algorithm is difficult to find a true disparity of small object differed vastly from a disparity of background. Fig. 2(a) shows a disparity map of box with plate-like bulge at the front. The aggregated cost is computed at a center of the bulge by the coarse-to-fine algorithm. Fig. 2(b) shows the aggregated costs using a large window, A, against disparity. The aggregated cost at a disparity of the background is

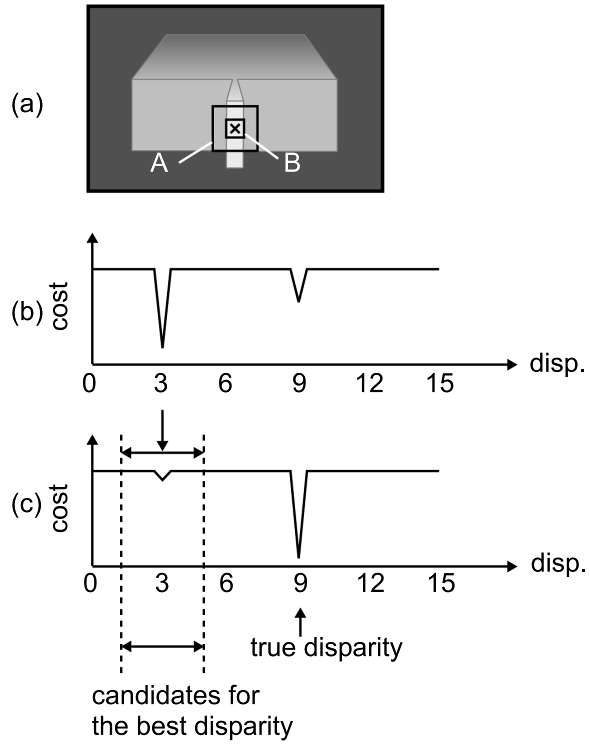


Fig. 2 Issue of the coarse-to-fine algorithm, (a) a disparity map of a box with plate-like bulge, (b) the aggregated cost using a large window "A" against disparity, (c) aggregated cost using a small window "B".

smaller than at a disparity of the bulge due to a strong influence of the background. Thus, a disparity of background, 3, is selected as the best disparity. In addition, the aggregated costs using a small window, B, as shown in Fig. 2(c). In this case, the aggregated cost at the disparity of the bulge is smaller than at a disparity of the background. In the coarse-to-fine algorithm, however, candidates for the best disparity are limited within a given area centered on the disparity computed by the large window. Therefore, the disparity of the bulge is not selected as the best disparity.

4. Proposed algorithm

To solve this issue, we propose a new window-based stereo matching algorithm. The absolute intensity difference, which is given by following equation, is used as the matching cost computation.

$$C_{mat}(x,y,d) = \left| I_r(x,y) - I_m(x+d,y) \right|. \quad (1)$$

Here, in the two stereo images, one is a reference image, and the other is a matching image. $I_r(x,y)$ is a pixel intensity at (x, y) in the reference image. And $I_m(x+d, y)$ is a intensity of pixel shifted in horizontal by the disparity value, *d*, from (x, y) in the matching image. Therefore, if *d* is true disparity, the matching cost, $C_{mat}(x,y,d)$, is reduced to almost zero because I_r is approximately equal to I_m at the disparity. The Gaussian filter is used as the cost aggregation. The aggregated cost by the Gaussian filter, $C_{agg}(x,y,d)$ is given by

$$G(i,j) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{i^2 + j^2}{2\sigma^2} \right), \quad (2)$$

$$C_{agg}(x,y,d) = \sum_{i,j} G(i,j) C_{mat}(x+i, y+j, d), \quad (3)$$

where, σ^2 is a variance of the Gaussian distribution and the filter size increased with the σ.

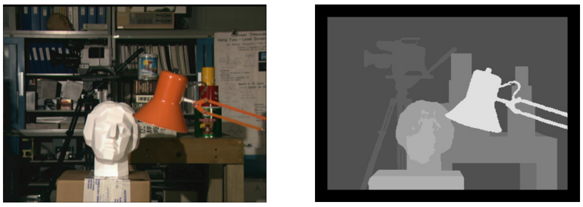
In the proposed algorithm, the aggregated cost maps are computed in sequential order while the window size is reduced gradually. And a new cost map is computed using a weighted average of the aggregated cost maps recursively and given by

$$C[n] = \begin{cases} C_{agg}[n], & n = 1, \\ \dfrac{w_1 \cdot C[n-1] + w_2 \cdot C_{agg}[n]}{w_1 + w_2}, & n \geq 2. \end{cases} \quad (4)$$

Here, C[n] and $C_{agg}[n]$ are the averaged and aggregated costs in n-th iteration, respectively. And w_1 and w_2 are weights of the averaged and aggregated costs, respectively. At first, an aggregated cost, $C_{agg}[1]$, is computed using the largest window for every possible disparity value at each pixel and is equal to the averaged cost in first iteration, C[1]. Then, a next aggregated cost, $C_{agg}[2]$, is computed at each pixel by using a window whose size is reduced compared to the first iteration. The averaged cost, C[2], is renewed at each pixel using the weighted average of the present aggregated cost, $C_{agg}[2]$, and the previous averaged cost, C[1]. These processes are computed recursively while the window size is reduced gradually. The final averaged cost map, C[N], is computed when the window becomes the minimum size and has every characteristic of aggregated costs using various window sizes.

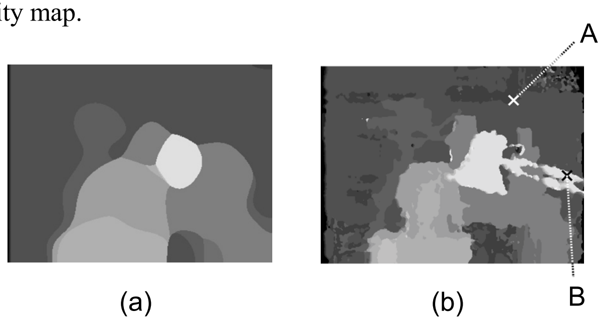
In the next step, the WTA optimization is used as the disparity computation. The WTA finds a disparity, *d*, when the averaged cost, C[N](x,y,d), is minimum value at each pixel, (x,y). And the disparity map is formed by the disparity of the minimum averaged cost. The disparity computed by the averaged cost is the best of disparities selected by every window sizes because the averaged cost expresses compressed information about aggregated costs using various window sizes from large to small. In the last step, a parabolic approximation is used as the sub-pixel disparity refinement. The parabola fits three values that are the averaged costs at the selected disparity by the WTA and both adjacent disparities. And a disparity where the parabola is minimum value is a refined best disparity.

In the proposed algorithm, because the aggregated costs computed for every possible disparity value, the



(a) (b)

Fig. 3 Simulation images: (a) reference image, (b) true disparity map.



(a) (b)

Fig. 4 Disparity map by our algorithm at the first iteration (a) and at the final iteration (b).

issue of limitation of candidates in the coarse-to-fine algorithm is solved. In addition, to construct a dense disparity map, the weight of the aggregated cost increases with reducing the window size due to the recursive formula (4).

5. Simulation

We have designed C++ programs of the proposed algorithm in order to evaluate the performance compared with the other algorithms. We used a stereo image data from the Middlebury stereo evaluation page [7] for our simulation as shown in Fig. 3. Fig. 3(a) and (b) show a reference color image, which is one of the stereo images, and a true disparity map, respectively.

We have computed disparity maps according to the proposed algorithm. The iteration count of the recursive formula (4) was five and the window size was gradually reduced, $\sigma = 24,12,6,3,1.5$. And both of the weights, w_1 and w_2 set to 1. The matching cost, C_{mat}, is defined as a summation of absolute intensity difference in each color channel (R, G and B). Fig. 4(a) shows the disparity map computed by the averaged cost at the first iteration. The disparity map was broadly correct compared with the true disparity map and inhibited the noise component the noise component significantly though detailed characteristics of objects in the disparity discontinuity region could not be detected because the computation at the first iteration used only the largest window. Fig. 4(b) shows the final disparity map. The detailed characteristics in the discontinuity region, such as poles and edges of the lump, were detected and there was a little terrible error in the disparity continuity region since the final averaged cost computed by Equ. (4) contained every characteristic of aggregated costs using various window sizes from large to small.

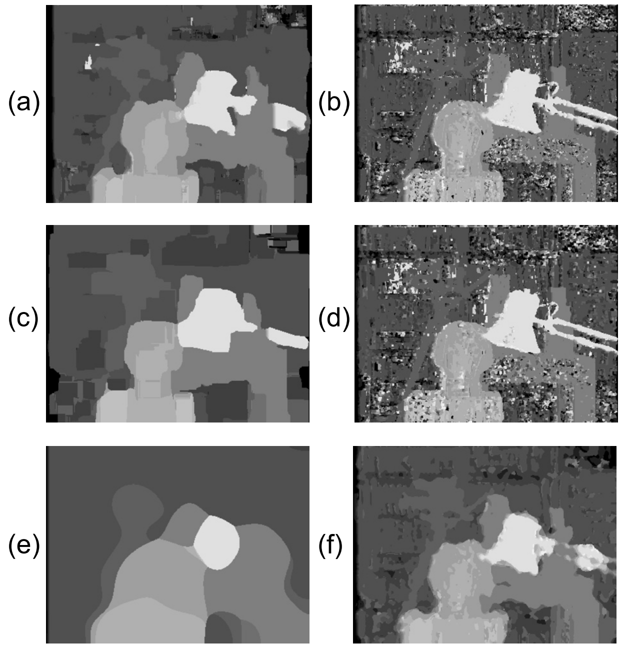


Fig. 5 Disparity maps: (a) box filter (window size = 15), (b) box filter (window size = 3), (c) shiftable window (window size = 21), (d) shiftable window (window size = 3), (e) coarse-to-fine (the first iteration), (f) coarse-to-fine (the final iteration).

We compared our algorithms with the conventional algorithms. We selected the box filter and the shiftable window for the cost aggregation as the conventional window-based algorithms. The box filter is the most simple aggregation method, which computes average of matching costs within a window. The shiftable window is introduced as the best aggregation method in the review [1]. Additionally, the conventional coarse-to-fine algorithm was simulated to the comparison.

Fig. 5(a) and (b) show disparity maps by the box filter using the large and small windows, respectively. And Fig. 5(c) and (d) show disparity map by the shiftable window using the large and small windows, respectively. The sizes of the large box filter and shiftable window were selected to get the best results in each algorithm. As shown in Fig. 5(a) and (c), when these algorithms used the large box filter and shiftable window, these algorithms has similar characteristics to the proposed algorithms at first iteration (Fig. 4(a)). However, when these algorithms used the small box filter and shiftable window, the disparity maps had many terrible errors

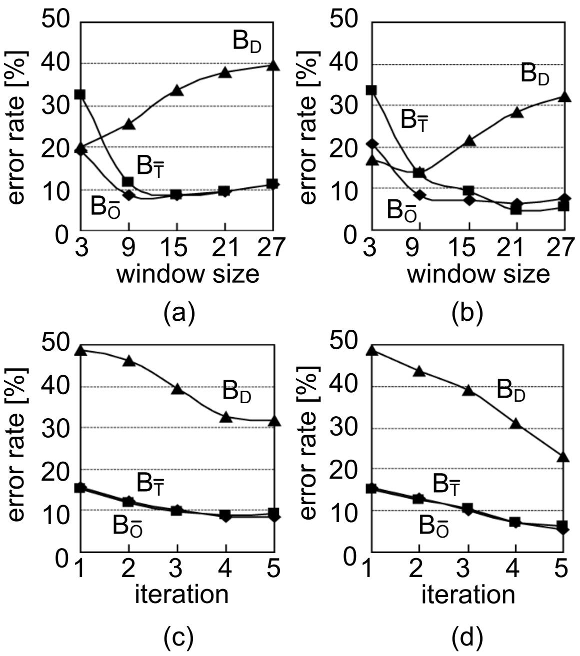


Fig. 6 Plots of the three evaluation measures of (a) box filter, (b) shiftable window, (c) coarse-to-fine algorithm and (d) proposed algorithm.

though the detailed characteristics in the disparity discontinuity region were detected as shown in Fig. 5(b) and (d).

The coarse-to-fine algorithm used the same filter and condition as the proposed algorithm. The range of candidate was reduced, $\pm 4, 3, 2, 1$, with reducing the window size. Fig. 5(e) shows the disparity map generated by the coarse-to-fine algorithm at first iteration. The first disparity map was exactly same as that of the proposed algorithm because of the same condition. Fig. 5(f) shows the final disparity. As shown in Fig. 5(f), the coarse-to-fine algorithm had almost similar characteristics to the proposed algorithm and there was a little terrible error in the disparity continuity region. However, the detailed characteristics in the disparity discontinuity region, such as poles and edges of the lump, were not detected due to an influence of the background.

Fig. 6(a), (b), (c) and (d) show plots of the three evaluation measures, $B_{\overline{O}}$, $B_{\overline{T}}$ and B_D, of the box filter, the shiftable window, the coarse-to-fine algorithm and the proposed algorithm, respectively. $B_{\overline{O}}$ is the error rate in the non-occluded region. $B_{\overline{T}}$ is the error rate in the texture-less region. The texture-less region includes a part of the disparity continuity region. B_D is the error rate in the disparity discontinuity region. The error rate represents the percentage of bad pixels, which mean false disparities compared with the true disparities as shown in Fig. 3(b). In the Fig. 6, the horizontal axis measures window size and the vertical axis measures the error rates. As shown in Fig. 6 (a) and (b), in the plots of the box filter and the shiftable window, when window size is small, $B_{\overline{T}}$ is high and B_D is low. In contrast, when window size is large, $B_{\overline{T}}$ is low and B_D is high. Namely, these results indicate the issue of the trade-off between accuracies of the disparity map in disparity continuity and discontinuity regions against the window size. As shown in Fig. 6(c) and (d), in the coarse-to-fine and the proposed algorithms, $B_{\overline{T}}$ and B_D decrease with increasing the iteration. However, in the coarse-to-fine algorithm, B_D does not decrease in the final iteration because of the limitation of candidates. Contrastively, in the proposed algorithm, B_D decreases more than the coarse-to-fine algorithm.

Conclusion

We proposed a new window-based stereo matching algorithm, which computes the disparity map using a weighted average of costs aggregated by various window sizes from large to small. We have designed C++ programs to evaluate the performance compared with the conventional algorithms. The proposed algorithm decreases error rates of the disparity map in both disparity continuity and discontinuity regions with the iteration. In addition, the algorithm generates a better disparity map than the coarse-to-fine algorithm.

References

[1] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *IJCV* 47(1/2/3):7-42, April-June 2002.

[2] W. E. L. Grimson, "A computational theory of visual surface interapolation," In Proc. of *Royal Soc. London B298*, pp. 395-427, 1982.

[3] A. Witkin, D. Terzopoulos and M. Kass, "Signal matching through scale space," *IJCV*, 1:133-144, 1987.

[4] D. Scharstein and R. Szeliski, "Stereo matching with nonlinear diffusion," IJCV, 28(2):155-174, 1998.

[5] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application in stereo vision," *In IJCAI-81*, pages 674-679, Vancouver, 1981.

[6] Q. Tian and M. N. Huhns, "Algorithms for subpixel registration," *GVGIP*, 35:220-233,1986.

[7] D. Scharstein and R. Szeliski, "Middlebury Stereo Vision Page," www.middlebury.edu/stereo.