A Window-based Stereoscopic System Using A Weighted Average of Costs Aggregated with Window Size Reduction

Kan'ya Sasaki, Seiji Kameda 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, iwa}@dsl.hiroshima-u.ac.jp

1. Introduction

A window-based stereo matching, which matches pixel values within a window between two images, produces a dense disparity map, and as a result, constructs a dense 3-dimensional depth structure. The dense 3-dimensional information is very useful to robot vision, view synthesis, image-based rendering, etc. Therefore, many window-based stereo algorithms have been proposed and applied to various applications [1]. The conventional window-based algorithms, however, face a trade-off between accuracies of the disparity map in disparity continuity and discontinuity regions due to the window size dependence.

2. Window-based stereoscopic algorithm

A fundamental process of window-based algorithm is generally divided into four steps; matching cost computation, cost aggregation, disparity computation and disparity refinement [2]. 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 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). The WTA finds a disparity when the aggregated cost is minimum value at each pixel position. 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 [3].

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. Namely, the conventional algorithms face a trade-off between accuracies of the disparity map in disparity continuity and discontinuity regions due to the window size dependence.

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.

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

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

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

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

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

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

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. Finally, the sub-pixel disparity refinement is computed and the sub-pixel disparity map is generated.

5. Simulation

We have designed C++ programs of proposed algorithm in order to evaluate the performance compared with the other conventional algorithms. We used a stereo image data from the Middlebury stereo evaluation page [4] for our simulation as shown in Fig. 2(a) and (b).

Fig. 2(c) and (d) show disparity maps by the box filter using the large and small windows, respectively. The box filter is the most simple aggregation method, which computes average of matching costs within a window [5]. As shown in Fig. 2(c), the disparity map generated by large window was broadly correct compared with the true disparity map and inhibited the noise component. However, detailed characteristics of objects in the disparity discontinuity region could not be detected. In contrast, the disparity map generated by small window had many terrible errors though the detailed characteristics in the disparity discontinuity region were detected as shown in Fig. 2(d).

Fig. 2(e) and (f) show disparity maps by proposed algorithm. The iteration count of the recursive formula (4) was five and window size was gradually reduced, $\sigma = 24,12,6,3,1.5$. And both of the weight, w_1 and w_2 set to 1. The disparity map at the first iteration had similar characteristics to the box filter using a large window (Fig. 2(c)). In the disparity map at the final iteration, the detailed characteristics in the disparity discontinuity region were detected and there was a little terrible error in the disparity continuity region.

Fig. 3(a) and (b) show plots of the three evaluation measures $B_{\bar{O}}$, $B_{\bar{T}}$ and B_D, of the box filter and the proposed algorithm. $B_{\bar{O}}$ is the error rate in the non-occluded region. $B_{\bar{T}}$ is the error rate in the texture-less region including a part of the disparity continuity region. B_D is the error rate in the disparity discontinuity region. As shown in Fig. 3(a), in the box filter, when window size is small, $B_{\bar{T}}$ is high and B_D is low. Additionally, when window size is large, $B_{\bar{T}}$ is low and B_D is high. In contrast, $B_{\bar{T}}$ and B_D decrease with increasing the iteration as shown in Fig. 3(b).

Conclusion

We proposed a new window-based stereo matching algorithm. The disparity map was computed using a weighted average of costs aggregated by various window sizes large to small. In addition, we have designed C++ programs to evaluate the performance. The proposed algorithm solved the issue of the trade-off between accuracies of the disparity map in disparity continuity and discontinuity regions against the window size.

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

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

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

[5] T. Kanade et al. " A stereo machine for video-rate dense depth mapping and its new applications. " In *CVPR*, pages196-202, 1996.

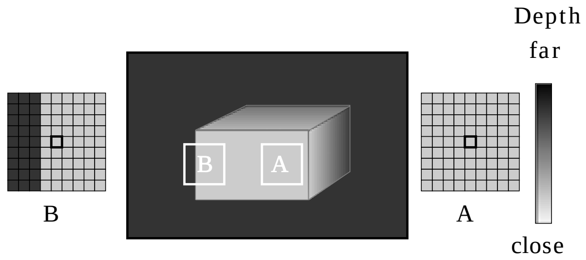


Fig. 1 Disparity map of a box, A: disparity continuity region, B: disparity discontinuity region.

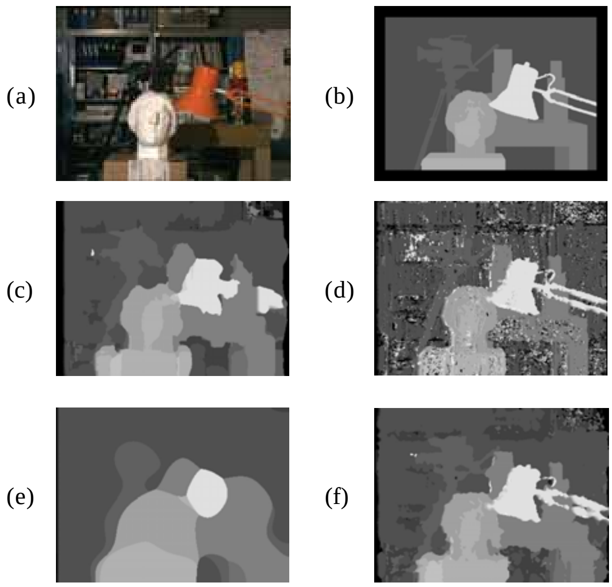


Fig. 2 Simulation images: (a) reference image, (b) true disparity map and Disparity maps: (c) box filter (win. size = 15), (d) box filter (win. size = 3), (e) proposed algorithm at the first iteration and (f) proposed algorithm at the last iteration.

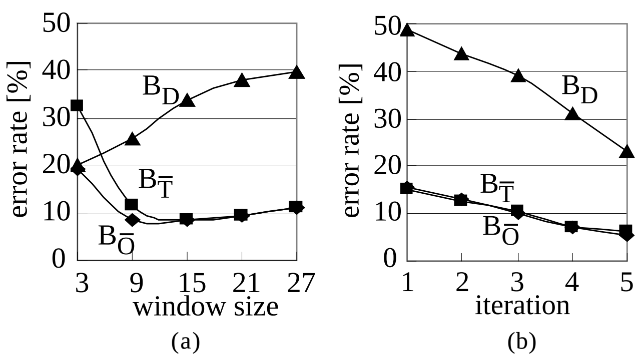


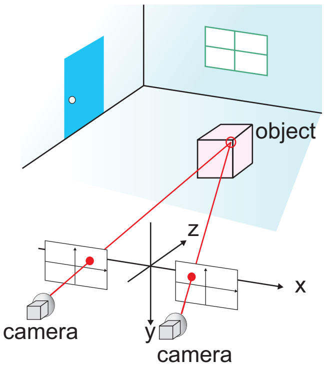
Fig. 3 Plots of the three evaluation measure of (a) box filter and (b) proposed algorithm.

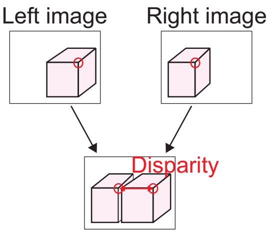
A Window-based Stereoscopic System Using A Weighted Average of Costs Aggregated with Window Size Reduction

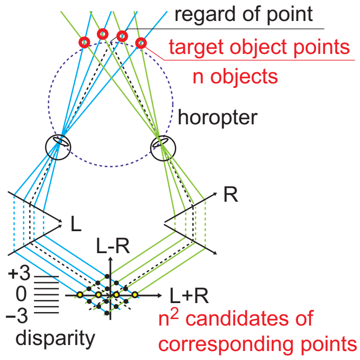
Kan'ya Sasaki, Seiji Kameda and Atsushi Iwata

Graduate School of Advanced Sciences of Matter, Hiroshima University

Introduction







Stereoscopic System

It produces a dense disparity map by using a pair of left and right images of a stereo camera system.

Window-based methods

It matches intensity values within windows between two stereo images in order to increase the accuracy of the disparity map.

Applications

robot vision, intelligent transport system (ITS)

Window-based Algorithm

Input images

↓

Matching cost computation

↓

Cost aggregation

↓

Disparity optimization

↓

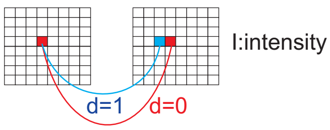
Disparity refinement

↓

Disparity map

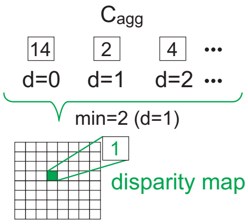
Matching cost computation

reference & matching images

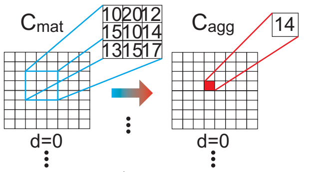


$$C_{mat}(x,y,d) = |I_r(x,y) - I_m(x+d,y)|$$

Disparity optimization



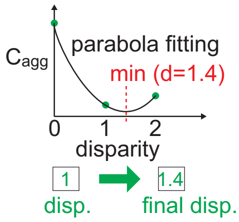
Cost aggregation



$$C_{agg}(x,y,d) = \frac{1}{i \times j} \sum_{i,j \in W} C_{mat}(x+i, y+j, d)$$

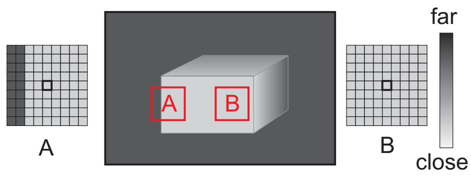
Box filter

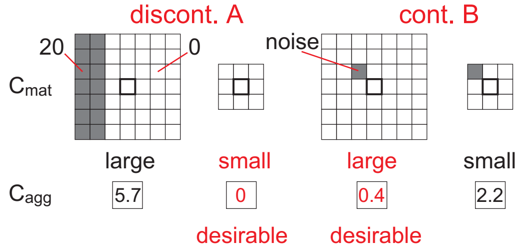
Disparity refinement



Issue of the algorithm

A trade-off between accuracies of the disparity map in disparity discontinuity region A and continuity region B due to the window size dependence





Proposed Algorithm

Input images

↓

Matching cost computation

↓

recursive calculation

Cost aggregation

↓

Cost averaging

Win. size reduction

↓

Disparity optimization

↓

Disparity refinement

↓

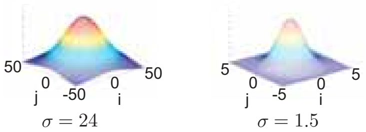
Disparity map

Cost aggregation

Gaussian filter

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

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



The Gaussian filter has a better performance than the box filter in the disparity discontinuity region because a weight of the Gaussian filter is the largest at a given pixel position.

Cost averaging

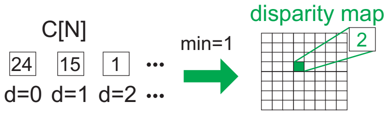
$C_{agg}[n]$:aggregated costs in n-th iteration

$C[n]$:averaged costs in n-th iteration

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

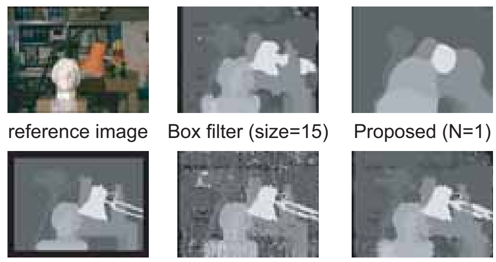
In the first iteration, matching costs are aggregated by the largest window. The averaged costs are renewed by the previous averaged costs and the present aggregated costs with window size reduction.

Disparity optimization



The averaged costs, C[N] have every characteristic of aggregated costs using various window sizes.
Therefore, the proposed algorithm can address the trade-off.

Simulation



reference image　Box filter (size=15)　Proposed (N=1)

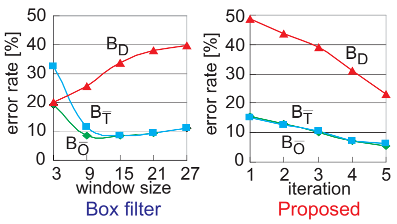
true disparity map　Box filter (size=3)　Proposed (N=5)

condition

$w_1 = w_2 = 1$

$N = 5$

$\sigma = 24, 12, 6, 3, 1.5$



The conventional algorithm faces the trade-off between $B_{\overline{T}}$ ($B_{\overline{O}}$) and B_D. In contrast, the proposed algorithm can reduce both $B_{\overline{T}}$ ($B_{\overline{O}}$) and B_D.

$B_{\overline{O}}$: error rate in the nonoccluded region

$B_{\overline{T}}$: error rate in the textureless region (part of the disparity continuity region)

B_D : error rate in the disparity discontinuity region

Conclusion

We proposed a new window-based stereo matching algorithm. The disparity map was computed using a weighted average of costs aggregated by various window sizes large to small. In addition, we have designed C++ programs to evaluate the performance. The proposed algorithm solved the issue of the trade-off between accuracies of the disparity map in disparity continuity and discontinuity regions against the window size.