EFFICIENT STEREO MATCHING BASED ON A NEW CONFIDENCE METRIC

Won-Hee Lee, Yumi Kim, and Jong Beom Ra

Department of Electrical Engineering, KAIST, Daejeon, Korea

ABSTRACT

In this paper, we propose a new confidence metric for efficient stereo matching. To measure the confidence of a stereo match, we refer to the curvatures around the two minimum costs of a cost curve, the size of aggregation kernel, and the occlusion information. Using the proposed confidence metric, we then design a weighted median filter, in order to refine the initially estimated disparities with a small aggregation kernel. In the design of weighted median filter, we overcome the performance degradation due to a small kernel size by utilizing the filter information of previously processed pixels. It is found that the performance of the proposed stereo matching algorithm is competitive to the other existing local algorithms even with a small size of aggregation kernel.

Index Terms**—**stereo matching, confidence metric, aggregation kernel size

1. INTRODUCTION

With the recent popularity of 3D display, various 3D displays without glasses have been developed [1]. Among those displays, a display that uses arrays of vertically oriented cylindrical lenslets shows a remarkably good actual appearance compared to the others. This lenslet-based display needs a series of multiple views to generate the 3D effect. Those multiple view images can be generated by using pixel disparities, which are obtained from two stereo images transmitted to the display. To obtain the disparities, stereo matching between the stereo images is necessary. For the TV application, however, the stereo matching should be performed in real-time. Therefore, the algorithm may need to be implemented in VLSI with consideration of low computational cost.

Several stereo matching algorithms had been implemented in VLSI [2, 3]. However, those algorithms may be improved for more efficient VLSI implementation. To design a stereo matching algorithm with low computational complexity, the aggregation kernel size is to be small because the aggregation process takes a large portion of computation load. However, a small size of aggregation kernel may cause a problem especially in a textureless area where little information for matching exists inside the aggregation kernel.

To examine the stereo matching performance according to the aggregation kernel size, a stereo matching algorithm [4] is performed by using two different sizes of aggregation kernel, 35 x 35 and 5 x 35, respectively, and the corresponding disparity maps are given in Fig. 1. As expected, the disparity map obtained with a smaller aggregation kernel includes much larger erroneous areas especially in the textureless area (see Fig. 1(d)). However, the increase is usually negligible in the texture area. Therefore, the disparities in the texture area are more reliable for refining incorrect disparities in the neighboring textureless area. Hence, for efficient refinement of the disparity map obtained with a small size kernel, we propose a new confidence metric. Our stereo matching algorithm based on the proposed metric is fully local and progressive so as to satisfy a VLSI framework.

The paper is organized as follows. In section 2, we briefly describe a stereo matching algorithm that we adopt for initial disparity estimation. We also review various confidence metrics. The proposed confidence metric and the corresponding stereo matching algorithm are described in section 3. Section 4 shows some experimental results. Finally, we conclude the paper in section 5.

2. RELATED WORK

2.1. Cross-based stereo matching algorithm

To estimate an initial disparity map, we adopt a local stereo matching algorithm, which is called the cross-based stereo matching algorithm [4]. In the algorithm, the raw matching cost is first computed for every pair of corresponding pixels. Namely,

$$e(x,d) = \min\left(\sum_{c\in\{R,G,B\}} \left|I_c^L(x) - I_c^R(x-d)\right|, T\right), \tag{1}$$

where I^L and I^R represent the left and right images, x and d denote the spatial coordinate and the disparity, respectively, and T denotes a truncation limit of matching cost. By using the obtained raw matching cost, the aggregated cost is computed as

$$c(x,d) = \frac{1}{\left\|U_d(x)\right\|} \sum_{s\in U_d(x)} e(s,d), \tag{2}$$

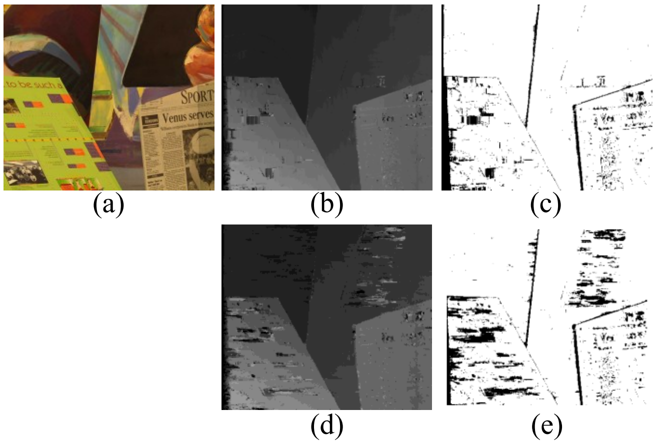


Figure 1. Disparity maps obtained by using aggregation kernels of different sizes. (a) Left image of the Venus set, (b), (c) disparity map and the corresponding bad pixel map (error > 1) obtained by using a kernel of 35 x 35, and (d), (e) disparity map and its bad pixel map (error > 1) by using a kernel of 5 x 35.

where $U_d(x)$ is a local support region [4], and $\|U_d(x)\|$ denotes the number of support pixels in $U_d(x)$. Finally, the disparity is selected according to the Winner-Takes-All strategy, as follows.

$$d^0(x) = \arg\min_d c(x,d), \quad d \in [0, d_{\max}], \tag{3}$$

where $d^0(x)$ denotes the initial disparity, and $d_{\max}$ denotes the maximum disparity.

2.2. Confidence metrics

Due to the inherent ambiguity of stereo matches, several metrics were proposed to measure the confidence level of match [5]. Those confidence metrics utilize an aggregated cost, curvature of the cost curve, and left-right consistency. The followings are some examples of existing metrics.

2.2.1. Matching score metric (MSM)
This metric uses the aggregated cost defined in Eq. (2) as a confidence measure, namely,

$$C_{MSM}(x) = -c(x, d_1). \tag{4}$$

Here, d_i denotes the disparity that reveals the i^{th} minimum cost. Intuitively, a lower cost implies a more correct match.

2.2.2. Curvature of cost curve metric (CUR)
The curvature of cost curve around the minimum cost indicates the confidence of a match, because the cost increases rapidly near the disparity having the minimum cost. Accordingly, CUR is defined as

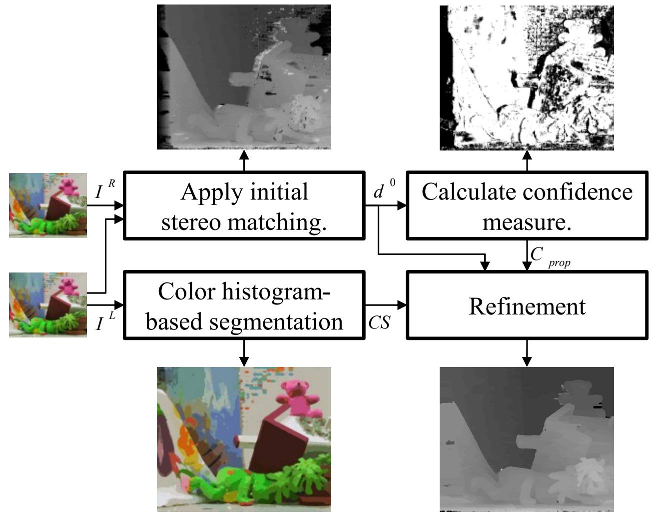


Figure 2. Overall diagram of the proposed algorithm.

$$C_{CUR}(x) = -2c(x,d_1) + c(x,d_1-1) + c(x,d_1+1). \tag{5}$$

2.2.3. Naive peak ratio metric (PKRN)
A distinctive minimum cost in the cost curve can be considered an evidence of confident match. Accordingly, PKRN is defined as a ratio of the two minimum costs, i.e.,

$$C_{PKRN}(x) = c(x,d_2)/c(x,d_1). \tag{6}$$

2.2.4. Naive winner margin metric (WMNN)
Similar to PKRN, WMNN utilizes the distinctiveness of the minimum cost. Instead of using a ratio, WMNN computes a margin between two minimum costs and normalize it with the sum of total costs.

$$C_{WMNN}(x) = \frac{c(x,d_2) - c(x,d_1)}{\sum_d c(x,d)}. \tag{7}$$

2.2.5. Left right difference metric (LRD)
Consistency between left and right disparity maps can be used to check if a match is confident. LRD is defined by using a margin between two minimum costs and the consistency of the minimum cost across the left and right disparity maps, i.e.,

$$C_{LRD}(x) = \frac{c(x,d_2) - c(x,d_1)}{\left|c(x,d_1) - \min\{c_R(x-d_1, d_R)\}\right|}. \tag{8}$$

Here, $\min\{c_R(x - d_1, d_R)\}$ implies the minimum value of a cost curve at the corresponding pixel in the right image.

3. THE PROPOSED ALGORITHM

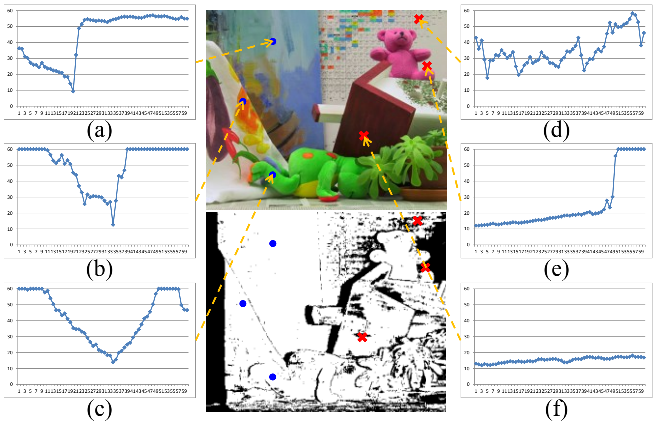


Figure 3. (a), (b), (c) Cost graphs of correctly estimated dot pixels; (d), (e), (f) cost graphs of incorrectly estimated cross pixels.

The overall diagram of the proposed algorithm is described in Fig. 2. In the algorithm, we first estimate an initial disparity map by using the cross-based stereo matching algorithm with a limited size of aggregation kernel. We then examine the confidence of initial disparity at each pixel by using the proposed confidence metric. Here, the metric is calculated by using the cost graph obtained during the initial disparity estimation process. To refine the disparities of low confidence, we determine neighboring pixels that belong to the same object, because they tend to have correct disparities. To determine those neighboring pixels, we perform the color segmentation by adopting a simple and efficient histogram-based algorithm [6], and assign the corresponding color segment index, *CS*, to each pixel. We then perform the refinement by using the initial disparity, confidence measure, and color segment index.

3.1. Proposed confidence metric

Since the limitation on the aggregation kernel size produces a considerable amount of bad pixels, it is important to discriminate them from the correctly estimated ones. To design a new confidence metric under various conditions, we examine many different types of cost graphs. Figs. 3(a), (b), and (c) show the cost graphs for three correctly estimated pixels, while Figs. 3(d), (e), and (f) show the graphs for the other three incorrectly estimated pixels. Since the correctly estimated pixels in Figs. 3(a) and (b) commonly provide a large curvature around the minimum cost, CUR can provide a large confidence value. However, CUR cannot provide a large confidence value for a small curvature in Fig. 3(c), since it is calculated by using only three cost values. Therefore, a new confidence metric is needed so that a cost graph with a small curvature in Fig. 3(c) can also provide a high confidence value. The metric should also discriminate from flat cost graphs shown in Figs. 3(e) and (f).

In addition, a new metric needs to provide a low confidence value for a periodic pattern. However, since the periodic pattern can have a large curvature around the minimum cost as shown in Fig. 3(d), its confidence value may become large. To alleviate this problem, a new metric may need to be designed using two minimum cost values as in PKRN and WMNN metrics. It is observed that a periodic pattern has high curvatures around at least two minimum costs. This observation may be utilized for designing a new confidence metric so that it can successfully handle periodic patterns.

Accordingly, a new metric is proposed as

$$C_{\mathrm{prop}}(x) = \sqrt{\left\|U_d(x)\right\|}\left\{(\mathrm{LoG} * c)(x, d_1) - (\mathrm{LoG} * c)(x, d_2)\right\}, \quad (9)$$

where LoG represents a Laplacian of Gaussian filter of n-taps. Note that C_{prop} includes an occlusion check as in LRD so that it becomes zero if a pixel is found to be occluded through a cross-check between left and right disparity maps.

The LoG filter in Eq. (9) extracts the curvature information across a range larger than that including three cost values in the CUR metric. This filter characteristic improves the performance of the confidence metric especially for a cost graph with a small curvature. Introducing a multiplication factor relying on the number of support pixels, we also improve the metric performance for a cost graph with a small curvature. This is because the number of support pixels corresponding to a small curvature graph is usually large. In addition, subtracting the curvature of the second minimum cost from that of the first minimum cost, we can successfully distinguish a periodic pattern from a low texture area.

3.2. Refinement

Since the computational cost is our main consideration in designing a stereo matching algorithm, local approaches such as a weighted median filter [7], a joint bilateral filter [8], etc., may be desirable rather than a global one. Among them, we adopt a weighted median filter that is performed only for the disparities of neighboring pixels that belong to the same color segment.

$$d(x, y) = MED\left[w_1 \Diamond d_1^0, w_2 \Diamond d_2^0, \ldots, w_N \Diamond d_N^0\right], \quad (10)$$

where d_i^0 represents the initial disparity of neighboring pixels whose color segment index is the same as that of the center pixel. Meanwhile, $\Diamond$ denotes a duplication operator, and weight, w_i, is defined as a sigmoid function of confidence metric, i.e.,

$$w = \left(1 + \exp\left(-\left(C_{\mathrm{prop}} - \tau\right)/\sigma\right)\right)^{-1}, \quad (11)$$

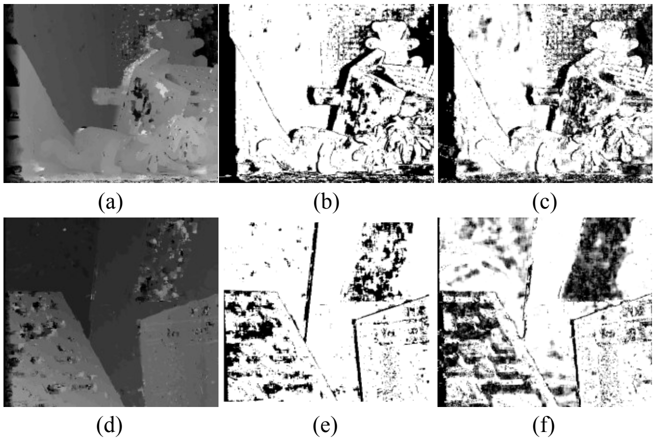


Figure 4. (a) Initial disparity map, (b) bad pixel map, and (c) proposed confidence map for Teddy; (d) initial disparity map, (e) bad pixel map, and (f) proposed confidence map for Venus.

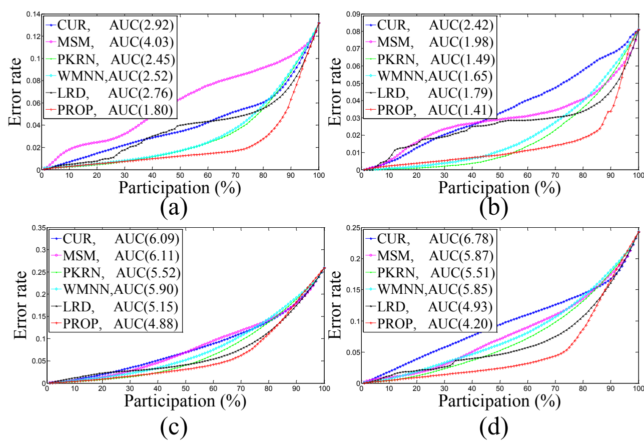


Figure 5. Performance of confidence metrics for (a) Venus, (b) Tsukuba, (c) Cones, and (d) Teddy.

where τ denotes an offset and σ determines a slope of the sigmoid function.

Due to the small size of filtering kernel, the filtering is applied only to the limited number of pixels. Hence, to effectively enlarge a filtering range, we vertically propagate the filtered result of a current pixel. The propagated data consist of a filtered disparity, weight, and color segment index, *CS*. If the propagated disparity has a larger weight than the current filtered disparity, the current weighted median filtered disparity is replaced with the propagated disparity. Otherwise, all the propagated data are replaced with the current ones.

4. EXPERIMENTAL RESULTS

To demonstrate the performance of the proposed algorithm, experiments were performed using the benchmark Middlebury stereo database [9]. In the experiments, the aggregation kernel is limited to 5 x 35 and the filtering kernel is set to 5 x 63. Parameters T, n, τ, and σ are set to 60, 7, 10, and 2, respectively.

To examine the performance of the proposed confidence metric subjectively, we visualize the confidence metric with a confidence map in Fig. 4. A confidence map may be compared to the corresponding bad pixel map. Since the correctly estimated disparity is to be assigned to a high value of confidence measure, a confidence map should closely correspond to the bad pixel map as shown in Fig. 4.

To examine the performance of the proposed confidence metric objectively, we adopt an objective measure, the area under the curve (AUC) [5]. To determine the AUC, we order pixels according to their confidence measure and produce a curve of error rate as a function of the participation of pixels. The area under the curve, or AUC, then represents the ability of a confidence metric to predict correct matches. A smaller value of AUC implies that the confidence metric assigns a larger confidence value to a correctly estimated disparity. Fig. 5 shows error rate graphs of four test sets, Venus, Tsukuba, Cones, and Teddy, for different confidence metrics. It is noted in the graphs that the proposed metric provides the smallest AUC for all the test sets.

The effectiveness of a confidence metric can also be examined by applying it to the refinement step of the proposed stereo matching algorithm. Table 1 shows that the proposed confidence metric produces relatively lower error rates for all the test sets than the other metrics.

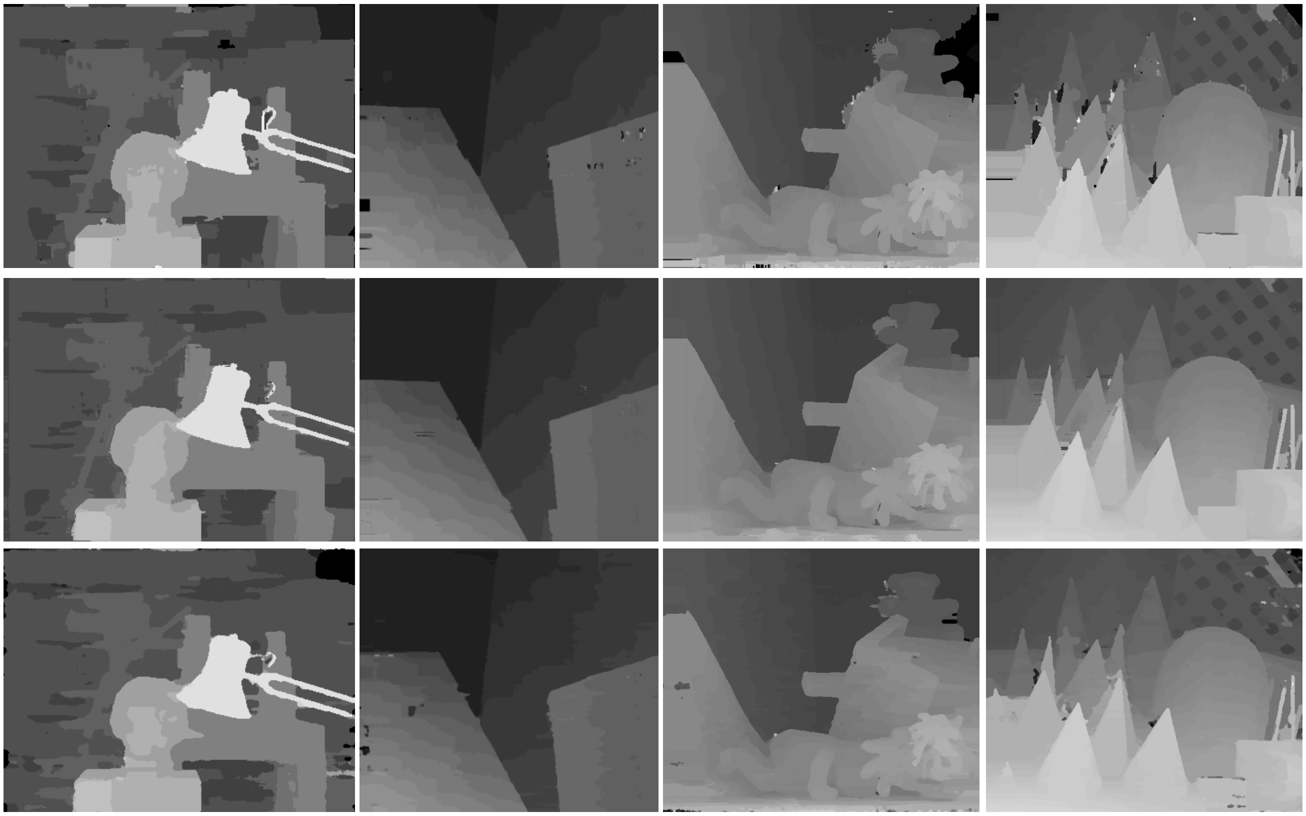
To examine the performance of the proposed local stereo matching algorithm, we compare its matching result with those of the other local algorithms reported in the Middlebury benchmark [9]. We can note in Table 2 that the proposed algorithm provides average error percentages similar to those of the others, even with a smaller size of aggregation kernel.

Table 1. Error percentages according to adopted confidence metrics in the proposed algorithm. Error threshold of 1.0 is used.

Confidence metric	Venus	Tsukuba	Cones	Teddy	Avg.
CUR	1.80	2.90	11.6	13.4	7.43
MSM	8.06	5.52	14.9	20.5	12.2
PKRN	2.50	2.85	12.7	15.4	8.36
WMNN	2.90	2.67	13.1	14.4	8.27
LRD	2.56	2.86	12.0	14.4	7.96
Proposed	1.83	2.77	11.4	12.6	7.16

Table 2. Error percentages of stereo matching algorithms for several Middlebury stereo sets. Error threshold of 1.0 is used.

Algorithm	Venus	Tsukuba	Cones	Teddy	Avg.
[4]	0.96	2.65	12.7	15.1	7.85
[10]	1.19	1.85	9.79	13.3	6.53
Proposed	1.83	2.77	11.4	12.6	7.15



(a) (b) (c) (d)

Figure 7. Disparity maps obtained by using the algorithms in [4] (1st row) and [9] (2nd row), and by using the proposed algorithm, for (a) Tsukuba, (b) Venus, (c) Teddy, and (d) Cones.

5. CONCLUSIONS

In this paper, we design an efficient stereo matching algorithm with the size limitation on both aggregation and filtering kernels. Applying a weighted median filter that is based on the proposed confidence metric, we can successfully refine initial disparities. To overcome the size limitation on aggregation kernel, we effectively utilize previously obtained filtering results by propagating them. The experimental result shows the performance of the proposed algorithm with a small size of aggregation kernel is competitive to the existing algorithms with a large size of aggregation kernel.

6. ACKNOWLEDGEMENT

This work was supported by a grant from LG Display Co. Ltd., Korea.

7. REFERENCES

[1] P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, and C. V. Kopylow, "A survey of 3DTV displays: Techniques and technologies," *IEEE TCSVT,* 2007.

[2] M. Hariyama, N. Yokoyama, and M. Kameyama, "1000 frame/sec stereo matching VLSI processor with adaptive window-size control," *IEEE Asian Solid-State Circuits Conference*, 2006.

[3] S. Park, C. Chen, and H. Jeong, "VLSI Architecture for MRF based stereo matching," in *SAMOS*, *LNCS* 4599, 2007.

[4] K. Zhang, J. Lu, and G. Lafruit, "Cross-based local stereo matching using orthogonal integral images," *IEEE TCSVT*, 2009.

[5] X. Hu and P. Mordohai, "Evaluation of stereo confidence indoors and outdoors," in *CVPR*, 2010.

[6] J. Delon, A. Desolneux, J. L. Lisani, and A. B. Petro, "A nonparametric approach for histogram segmentation," *IEEE TIP*, 2007.

[7] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo, "Weighted median filters: A tutorial," *IEEE Trans. Circuits Syst. II*, 1996.

[8] S. Mattoccia, S. Giardino, and A. Gambini, "Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering," in *ACCV*, 2009.

[9] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV* 2002.

[10] K.-J. Yoon and I.-S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE TPAMI*, 2006.