Block-Based Cost Aggregation and Layered Disparity Optimization for Dense Stereo Correspondence

Mingjing Ai, Wenbo Chen
State Key Laboratory of Virtual Reality Technology and Systems
Beihang University
Beijing, China
Email: c-bob-87@hotmail.com

Abstract—**Recent top performing dense stereo algorithms based on advanced cost aggregation strategies could gain accurate disparity map. However, accuracy is achieved at the expense of high computational requirement. And the existing disparity optimization techniques enforce piecewise continuity assumption by penalizing disparity variations. Unfortunately, those pixels closed to the camera would lose some detail information. In this paper, a novel block-based cost aggregation strategy and a layered disparity optimization approach for dense stereo correspondence is proposed aimed at maximizing the speed-accuracy trade-off. The matching cost weights are computed on block basis in order to increase noise robustness and improve computing speed. And layered optimization method enables us to obtain disparity map with details well saved. Experimental comparison with reference algorithms confirms the effectiveness of our proposal.**

Keywords-stereo; disparity map; block-based; cost aggregation; layered optimization

I. INTRODUCTION

Dense disparity map is a very important requirement in many applications, such as robot navigation and control, 3D tracking, 3D Graffiti detection, 3D reconstruction and virtual reality [1][2][3][4][5]. In many areas such as image based render (IBR) and autonomous navigation, the stereo algorithms with low run-time are required. At the meanwhile there is a need for relative accurate disparity map to improve the quality of result. In spite of the efforts made in the last decades, the stereo correspondence that generates accurate dense disparity map with low run-time is still an important and open problem.

According to [6], most dense stereo algorithms are currently classified into local approaches and global approaches, which perform four steps: matching cost computation, cost aggregation, disparity computation/optimization and disparity refinement. Local approaches based on cost aggregation strategy which aggregate costs of neighboring points within a support window are typically faster than global approaches[7][8]. As reported on the Middlebury stereo evaluation site [9], the state-of-the-art cost aggregation strategies of local algorithms are based on adaptive weights or segments which yield accurate disparity map. However, the execution time is often comparable. In addition, the results of the existing cost aggregation strategies are easily affected by noise [10].

On the other hand, recently it's noteworthy that advanced stereo algorithms focus on the disparity optimization techniques which enforce piecewise smoothness assumption in vertical and horizontal directions in order to improve the quality of disparity map. There are lots of efficient disparity optimization techniques, such as Graph Cuts, Belief Propagation, Scanline Optimization, Dynamic Programming and Locally Consistent, which all can obtain reasonable and accurate disparity by modeling the behavior of piecewise continuity among points of disparity map[11][12][13][14]. However, those techniques treat the points that have different disparity (depth) with the same method. It will lose important detail information when smoothing disparities of the points with large disparities (or the points closed to camera).

The main contributions of this paper can be described as follows:

1) A novel block-based cost aggregation strategy is proposed. The matching cost weights of points within a support window are computed on block basis. And the matching cost weights are computed only in the reference image. So the aggregation strategy we proposed can increase the noise robustness with low run-time.

2) A novel layered disparity optimization approach is proposed. The layered disparity optimization approach deals with the points with large disparity in a cautious method and the points with small disparity in the conventional method. Thus, the important detail information is well saved.

Overall, the dense stereo correspondence algorithm we proposed aims at maximizing the speed-accuracy trade-off. And experimental results show the effectiveness of our proposal.

This paper is organized as follows: Section II reviews some related work on aggregation strategy and optimization technique. Section III introduces the new method we proposed. Experiments and discussion are arranged in Section IV. Finally we get the conclusion of this paper.

II. RELATED WORK

In this section, we briefly review cost aggregation strategies of local stereo algorithms and disparity optimization methods.

A. Cost aggregation strategies

According to the Middlebury stereo evaluation site, the top-banked local algorithms are based on the aggregation strategy of adaptive weights (AW) and segment support (SS). Yong and Kweon proposed the AW approach that matching cost is weighted by spatial and color constraints. This method deploys fixed support window but adaptive weights to image content and adaptive weights are computed both on the reference and target images [7]. The SS method proposed by Tomabari segments both reference and target images, and discards the geometric proximity, weights rely only on segment result and photometric variation [15]. Though they can obtain accurate result, the required computational burden is heavily. There are two noteworthy methods which can improve the traditional approaches, namely Fast Aggregation (FA) and Fast Bilateral Stereo (FBS). Inspired by SS, FA segments only the reference image, and its support window extends to the entire segment [16]. FBS improves AW by computing the weights of similarity and proximity based on block basis in support window. And the block weight is computed on both images [10]. Our aggregation strategy combines accuracy of AW and efficiency of FA and FBS to obtain high quality disparity map within low run-time.

B. Mehtods for disparity optimization

There are lots of efficient disparity optimization methods. All the methods can be classified into 2D and 1D approach. On the one hand, the 2D approaches, such as Graph Cuts (GC) and Belief Propagation (BP), enforce the smoothness assumption in vertical and horizontal directions [11][12]. On the other hand, the approaches of 1D, such as Dynamic Programming (DP) and Scanline Optimization (SO), minimize the energy function on a subset of neighboring points [6][13]. Locally Consistent (LC) is another optimization approach that enforces the local consistency of disparity map by emphasizing the mutual relationship among neighboring points in disparity map [14]. As discussed above, though they are efficient to optimize disparity, important detail information of the points will be lost with large disparity. We propose a different but interesting optimization technique pertinent with the idea of treating different layer points with different manners.

III. PROPOSED APPROACH

This paper focuses on binocular stereo system and images are in standard form. In local stereo correspondence algorithms, the homologous point is searched along the scanline in the target image for the point in the reference image. As surveyed in [6], our proposed dense stereo consists of three steps:

(1) Matching cost computation,

(2) Bock-based cost aggregation strategy,

(3) Layered disparity optimization approach.

A. Matching cost computation

In order to eliminate the influence of outliers, the matching cost in support window is computed with Sum of Truncated Absolute Differences (STAD),

$$STAD(x,y,d)=\sum_{x=-n}^{x=n}\sum_{y=-n}^{y=n}\min\{|I_R(x,y)-I_T(x+d,y)|,T\} \qquad (1)$$

Where $STAD(x,y,d)$ indicates the matching cost of point (x,y) at disparity d. The size of support window is $(2n+1)\times(2n+1)$. $I_R(x,y)$ and $I_T(x,y)$ are respectively reference and target image pixels. And T is the truncation value.

B. The novel Block-based cost aggregation strategy

As already pointed out in [7], similarity and proximity are used to compute support-weights. The more similar the color of a pixel, the larger its support-weight, and the closer the pixel is, the larger the support-weight is. Hence, the point that is similar to the central point play an important role in correspondence. As experimental result shows, when computing the support-weight between two pixels with similar color the image noise heavily affects the result in the uniform regions or ambiguous areas. To solve this problem, the proposed cost aggregation strategy takes the average value of pixels within blocks in support window instead of the single pixel. FBS method has used a similar strategy to deal with the problem. The similarity between average value of pixels within block and the central pixel is referred in that method. The approach we proposed takes into account the similarity between two blocks that neighboring points and central point respectively located in.

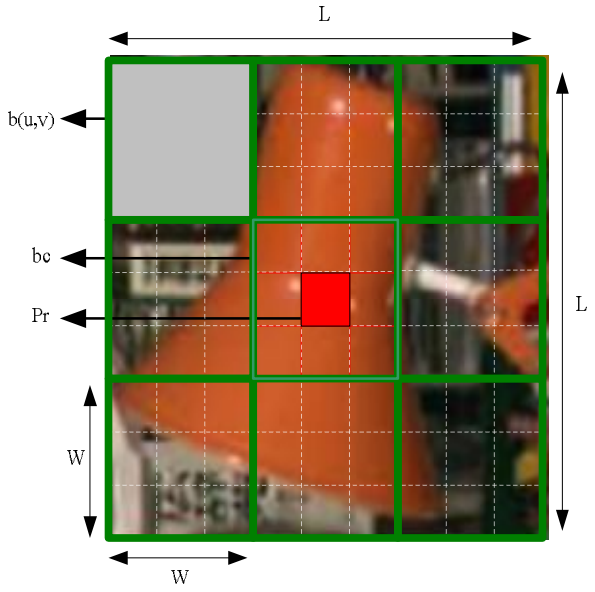

Figure 1. Proposed block-based cost aggregation strategy

In addition, inspired by FS approach, the method we proposed computes the support-weights of blocks only in the reference image. So, it's a release for the computational burden comparing with the conventional adaptive weights strategy.

As shown in Figure1, the size of support window is $L \times L$, which is divided into several blocks of size $W \times W$. Pr is the central point which is within the central block bc. And block $b(u,v)$ is a neighboring block. We independently assign spatial weight and photometric weight for each neighboring block within the support window.

For spatial constraint, the proposed strategy relies on the consumption that the closer the block is, the larger the support-weight is. It can be expressed as:

$$w_s(b(u,v)) = \begin{cases} \exp(-\dfrac{d(b(u,v),p_r)}{\gamma_s}) & b(u,v) \neq bc \\ 1 & b(u,v) = bc \end{cases} \quad (2)$$

Where $w_s(b(u,v))$ is the spatial weight of pixels within block $b(u,v)$ of the reference image, $d(b(u,v),p_r)$ is the Euclidean Distance between blocks of $b(u,v)$ and bc. γ_s is a parameter to adjust the spatial weights.

For color constraint, in order to reduce the influence of noise pixels, the mean value of pixels within the $W \times W$ block takes the place of single pixel. The photometric weight assigned to pixels within a block of the reference image is computed by means of:

$$w_p(b(u,v)) = \begin{cases} \exp(-\dfrac{d(\bar{I}(b(u,v)),\bar{I}(bc))}{\gamma_p}) & b(u,v) \neq bc \\ 1 & b(u,v) = bc \end{cases} \quad (3)$$

Where $\bar{I}(b(u,v)), \bar{I}(bc)$ represent the average value of pixels within block $b(u,v)$ and bc only in the reference image respectively. $d(\bar{I}(b(u,v)),\bar{I}(bc))$ is the difference between $\bar{I}(b(u,v))$ and $\bar{I}(bc)$. γ_p is a parameter to adjust the photometric weight.

According to equation (2) and equation (3), the overall weight assigned to all points within the same block of the reference image is computed by means of:

$$w(b(u,v)) = w_s(b(u,v)) \cdot w_p(b(u,v)) \quad (4)$$

Given a block $b(u,v)$ matching cost $STAD(b(u,v))$ for point (x,y) at disparity d, with the support window size $L \times L$ and the block size $W \times W$, the final cost with the point (x,y) at disparity d is:

$$C(x,y,d) = \frac{\sum_{i=1}^{\frac{L}{W}\times\frac{L}{W}-1} w_p(b_i(u,v)) \cdot w_s(b_i(u,v)) \cdot STAD(b_i(u,v))}{\sum_{i=1}^{\frac{L}{W}\times\frac{L}{W}-1} w_p(b_i(u,v)) \cdot w_s(b_i(u,v))} \quad (5)$$

As mentioned above, a single weight is assigned to each point within block of size $W \times W$, compared to the AW method the number of computing of weight is reduced by factor $W \times W$. Besides, computation of weights is carried out only in the reference image which halves the computational burden of conventional approach. Moreover, it is noteworthy that $STAD(b(u,v))$ and $\bar{I}(b(u,v))$ can be reused by the matching cost computation of other points.

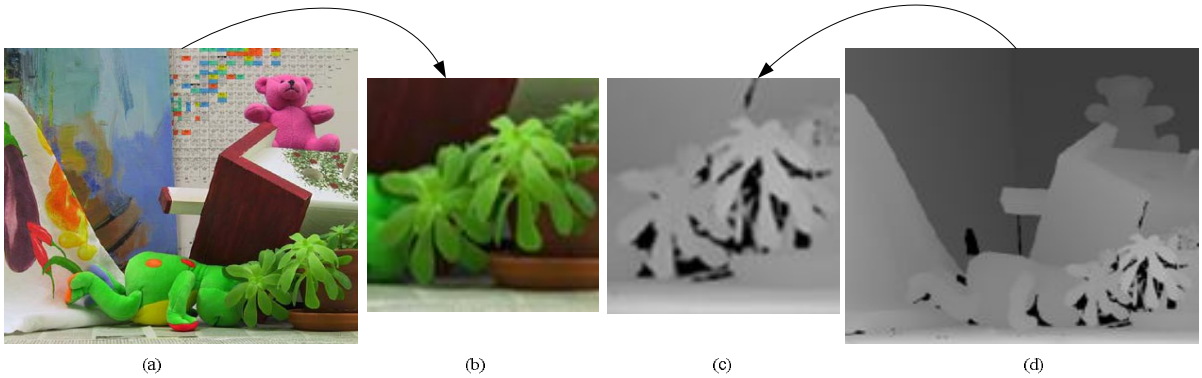

Figure 2. Example of behaviour that the section with larger disparity: (a) is original Tsukuba image, (d) is groundturth of (a), (b) is the section of image that closer to camera, (c) is the groundturth of (b). There are lots of important information about detail in (c) which is useful in human vision system.

C. *Proposed layered disparity optimization approach*

As we all know, with the stereo images in standard form we can infer depth, by means of triangulation:

$$depth = \frac{B \cdot f}{d} \tag{6}$$

Where B is the baseline of the binocular stereo camera system, f is the focus of the camera. As shown in equation (6), the larger the disparity is, the closer to camera the point is. As shown in Fig. 2, it will lose important detail information when smoothing the (b) segment of image. The proposed layered disparity optimization technique treats the points of large disparity in a cautious method.

After computation of disparity of each point by means of Winner Take All (WTA):

$$d = \underset{d\min \le d \le d\max}{\arg\min} \{C(x, y, d)\} \tag{7}$$

We use a layered disparity optimization approach to smooth the disparity map. The disparity map is divided into two layers in the novel approach, the foreground layer and the background layer. The threshold that divides disparity map is simply assigned as:

$$T_{layer} = \frac{d_{\min} + d_{\max}}{2} \tag{8}$$

Finally, the outlier detection and disparity replacement takes different manners to deal with the two layers. The pseudo-code of the proposed layered disparity optimization algorithm is shown at Fig. 3.

```
for u from 1 to width
  for v from 1 to height
    if d(u,v) ≥ T_layer ,do
        d^new(u,v) = d(u,v)
    else, do
      if d(u−1,v) = d(u+1,v) = d_w, do
        d^new(u,v) = d_w
      else if d(u,v−1) = d(u,v+1) = d_h, do
        d^new(u,v) = d_h
      else, do
        d^new(u,v) = d(u,v)
end
```

Figure 3. Pseudo-code of the layered disparity optimization algorithm

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we discuss and compare the experimental results between the approach we proposed and other typical dense stereo algorithms. The disparity maps yielded by both our method and other approaches are evaluated with the framework in [6]. And all approaches execute on the standard stereo images of the Middlebury dataset: Tsukuba (384×288) and Teddy (450×375). Since this paper aims at maximizing the speed-accuracy trade-off, as a term of comparison we select the state-of-the-art methods of high accuracy (AW, SS)， and the improved method FBS. Our result has been computed by Intel processor (Core(TM)2 Quad CPU 2.83GHz) and 4G EMS memory, and the codes are compile by VS2008. The results of referred algorithms are available on the Vision-deis site [17]. Our result has been obtained with the parameters were deployed as: L=15, W=3, γ_s =30, γ_p =40. And the optimal parameters of the referred algorithms are available on the Vision-deis site.

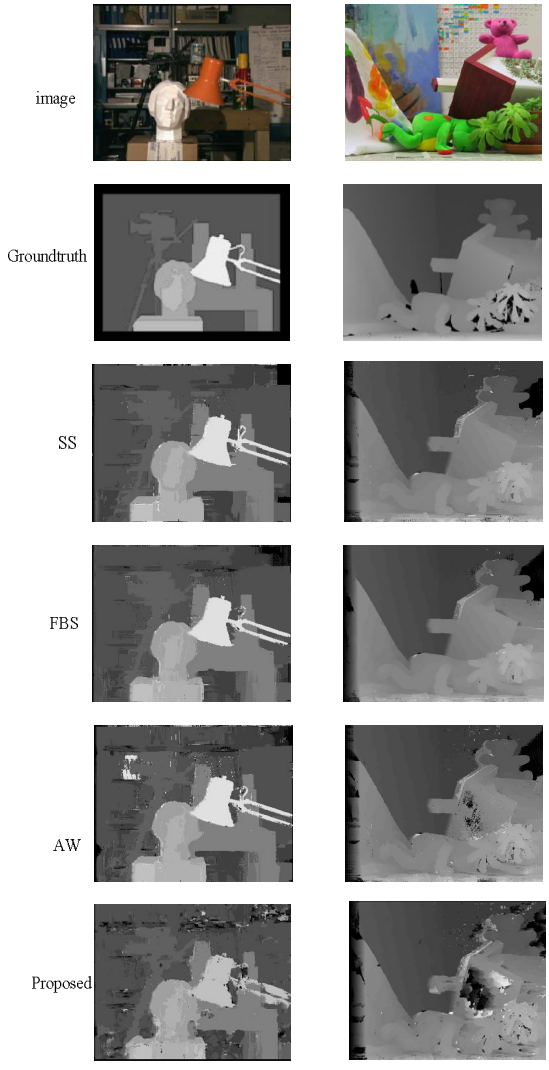

Figure 4. From top to bottom: reference standard image, Groundtruth, disparity map yielded by AW, SS, FBS and our proposed approach.

TABLE I. ACCURACY AND RUN-TIME COMPARISON

	tsukuba		teddy		Run-times
	nocc	*disc*	*nocc*	*disc*	*(teddy) (mm：ss)*
SS	2.15	7.22	10.5	21.2	39:30
FBS	2.95	8.69	10.7	20.8	00:32
AW	4.66	8.25	12.7	22.4	20:35
proposed	5.67	12.1	14.6	25.0	00:08
FW	9.58	27.1	25.1	42.4	<1s

Table 1 shows the results in terms of accuracy and computation requirements yielded by the evaluated algorithms. Accuracy according to the Middlebury evaluate web site [9] and according to [6]. For what concerns computational requirement, the proposed method is significant faster than AW, SS and FBS. We noticed that our proposed method takes only 8 seconds on the Teddy stereo images while SS and AW run is about 20 minutes to 40 minutes. In addition, our method is almost 4 times faster than FBS. At the same time, as the price the accuracy of the proposed approach has not been decline too much. The table shows the method Fixed Window (FW) is significantly faster than the approach we proposed, and we also noticed that in most case the approach we proposed outperforms the accuracy of the FW.

Fig. 4 shows the disparity maps yielded by the evaluated approaches. From the figure, we noticed that the method we proposed can make a good result at dealing with the repetitive texture area compared to AW. Compared to results yielded by the top performing algorithms, our approach is a bit inferior to SS and FBS while is approximately equal to AW. Besides, foreground layer disparity is well saved in our result. Because the approach we proposed deals with the points with large disparity in a cautious method, there are some noise points in our result.

Hence, it is clear that overall the approach we proposed for dense stereo correspondence maximizes the speed-accuracy trade-off. Especially, in some areas such as IBR and autonomous navigation, experimental result shows that our proposed method can obtain relative accuracy result within low run-times.

V. CONCLUSIONS

In this paper, we have proposed a novel block-based cost aggregation strategy and layered disparity optimization technique for dense stereo correspondence. The proposed aggregation strategy showed the capabilities to increase the noise robustness with low run-times. The novel layered optimization approach obtained disparity maps with detail well saved by treating different point with respective methods. Overall, the algorithm we proposed was efficient to maximize the speed-accuracy trade-off. As our future work, we will optimize our proposed method to achieve the requirement of real-time applications by using the techniques of GPU or SIMD instructions.

REFERENCES

[1] L. Di Stefano, M. Marchionni, S. Mattoccia, A fast area-based stereo matching algorithm, Image and Vision Computing, 2004, pp 983-1005

[2] M. Harville, Stereo person tracking with adaptive plan-view templates of height and occupancy statistics, Image and Vision Computing, 2004, pp 127-142

[3] L. Di Stefano, F. Tombari, A. Lanza, S. Mattoccia, S. Monti, Graffiti detection using two views, ECCV 2008 -8th International Workshop on Visual Surveillance(VS 2008)

[4] Li Zhang, Brian Curless, and Steven M. Seitz, Spacetime Stereo: Shape Recovery for Dynamic Scenes, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003), pp. 367-374

[5] P. Azzari, L. Di Stefano, F. Tombari, S. Mattoccia, Markerless augmented reality using image mosaics, International Conference on Image and Signal Processing (ICISP 2008)

[6] D. Scharstein and R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, Int. Jour. Computer Vision, 2002, 47(1/2/3):7–42,

[7] K. Yoon and I. Kweon, Adaptive support-weight approach for correspondence search, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 28, NO. 4, pp. 650–656, 2006, APRIL 2006

[8] M. Gerrits and P. Bekaert. Local Stereo Matching with Segmentation-based Outlier Rejection, In Proc. Canadian Conf. on Computer and Robot Vision (CRV 2006), pp. 66-66, 2006

[9] D. Scharstein and R. Szeliski, http://vision.middlebury.edu/stereo/eval/

[10] S. Mattoccia, S. Giardino,A. Gambini, Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering, Asian Conference on Computer Vision (ACCV2009)

[11] V. Kolmogorovand R. Zabih, Computing visual correspondence with occlusions using graph cuts, ICCV 2001

[12] A. Klaus, M. Sormannand K. Karner, Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure, ICPR 2006

[13] H. Hirschmüller. Stereo vision in structured environments by consistent semi-global matching, CVPR 2006, 30(2):328-341, 2008

[14] S. Mattoccia, A locally global approach to stereo correspondence, 3D Digital Imaging and Modeling (3DIM 2009), pp. 1763-1770, October 2009

[15] F. Tombari, S. Mattoccia, and L. Di Stefano, Segmentation-based adaptive support for accurate stereo correspondence, PSIVT 2007

[16] F. Tombari, S. Mattoccia, L. Di Stefano, E. Addimanda, Near real-time stereo based on effective cost aggregation, International Conference on Pattern Recognition (ICPR 2008)

[17] Stefano Mattoccia, www.vision.deis.unibo.it