LETTER

Stereo Matching Using Local Plane Fitting in Confidence-Based Support Window

Chenbo SHI[†a)], ***Student Member***, **Guijin WANG**[†b)], ***Member***, **Xiaokang PEI**[†], ***Nonmember***, **Bei HE**[†], ***Student Member***, ***and*** **Xinggang LIN**[†], ***Nonmember***

SUMMARY This paper addresses stereo matching under scenarios of smooth region and obviously slant plane. We explore the flexible handling of color disparity, spatial relation and the reliability of matching pixels in support windows. Building upon these key ingredients, a robust stereo matching algorithm using local plane fitting by Confidence-based Support Window (CSW) is presented. For each CSW, only these pixels with high confidence are employed to estimate optimal disparity plane. Considering that RANSAC has shown to be robust in suppressing the disturbance resulting from outliers, we employ it to solve local plane fitting problem. Compared with the state of the art local methods in the computer vision community, our approach achieves the better performance and time efficiency on the Middlebury benchmark.

key words: *stereo matching, local based, local plane fitting, RANSAC*

1. Introduction

Stereo matching is an active research topic in the community of computer vision. It is one of effective 3D reconstruction methods since it can usually recover a dense disparity map from image pair once the inherent ambiguity is properly addressed. The main difficulty is how to accurately estimate the disparity maps from image pair under different scenarios of smooth region, obviously slant plane and occluded areas [1].

Among available methods in stereo matching, local based algorithms have lately gained much attention due to their competitive performance. In local-based schemes, the disparity at any given pixel only depends on intensity values within a finite window. Most early local based methods, such as line scanning [2] and adaptive windows [3], improved the performance by space constraints between neighboring pixels. However, since the size of the local window was small, it was hard to deal with textureless and occlusion regions. In recent years, some effective methods have been proposed with weighted support window [4], where each pixel has different weight for the evaluation of central point. The size of window could be much larger than the earlier methods. Similar to the bilateral filter, Yoon [4] weighted each pixel by combining the space and color similarity between pixels. Hosni [5] used the geodesic distance as the support weight which is defined as the shortest path connecting two pixels in the color volume. These methods were able to obtain the disparity in textureless region and kept the discontinuity property on the edge. However, these algorithms required a high computational cost. Gupta [6] focused on reducing the computation time. Binary window was proposed to choose the pixels which were similar to the color of the central pixel. But its model was not fit to the slant plane and sensitive to the input parameters.

In this paper, a novel Confidence-based Support Window (CSW) is presented to efficiently select the reliable pixels in local based scheme. Confidence is defined to measure the reliability of current disparity value by the local matching cost. Only those pixels with low confidence will be selected as a center of CSW. For each CSW, the pixels used for fitting are picked up not only according to the color similarity with the central point, but also the confidence of disparity value. Morphological operators are used to select more reliable CSW. Then we employ RANSAC, which has shown to be robust to suppress the disturbance resulting from outliers, to solve local plane fitting problem. The results in different datasets show that our algorithm can deal well with the smooth region and obviously slant plane. The ranks on Middlebury benchmark prove that our method is one of the best local based algorithms at present.

2. Algorithm

In Local window, the reliability of each pixel's disparity value is quite different. The messages from reliable pixels, like pixels on the color edge, are more important than others. If we can distinct pixels with different confidence, the message will be more efficient using high confidence pixels as sources. Consequently, we consider the confidence map to better evaluate the disparity values. Figure 1 shows the procedure flow chart of our algorithm. First we initialize the disparity map and the confidence map by local matching cost (Sect. 2.1). It is a simple but useful method for the coarse estimation. Then we select the CSW to get reliable pixel set (Sect. 2.2). Combining with CSW and initial disparity map, local plane fitting is employed to reassign the disparity map (Sect. 2.3). The reassignment in smooth region can be solved by the following two steps. Firstly, the disparity of the smooth region boundary (more reliable) is robustly reassigned. Then since our method is in situ, the plane fitting by RANSAC rejects the error results and reassigns new disparity to the inner pixels by those sparse pix-

Manuscript received April 1, 2011.
Manuscript revised August 31, 2011.
[†]The authors are with the Dept. of Electronic Engineering, Tsinghua University, ROC.
a) E-mail: scb@mails.tsinghua.edu.cn
b) E-mail: wangguijin@tsinghua.edu.cn
DOI: 10.1587/transinf.E95.D.699

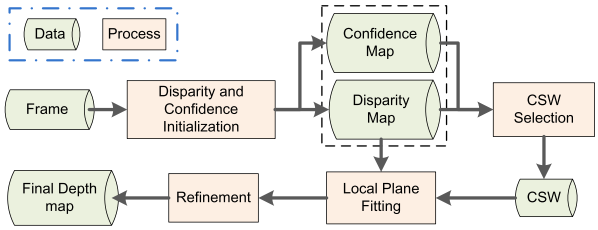

Fig. 1 The flow chart of our algorithm.

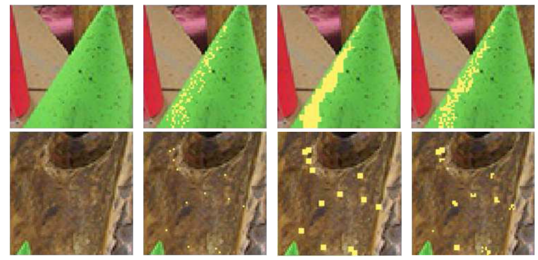

Fig. 2 Efficient selection of reliable pixels in CSW. The upper row is the smooth region while the lower row is the texture region. From left to right, (a) the original color image; (b) color similar pixels set S_p; (c) S'_p after morphology operation; (d) E_p derived from S'_p.

els with correct values. Finally in refinement, the left-right check [1] and median filter are employed to remove noise and inconsistency.

2.1 Disparity Initialization and Confidence Definition

The initialization of disparity map uses the matching cost in a small window. A widely used matching cost is the Sum of Absolute Difference (SAD) on pixel intensities between the left and right frames. The intensities distance defined in RGB color space is shown as:

$$d_{rgb}(p,p') = |r_p - r_{p'}| + |g_p - g_{p'}| + |b_p - b_{p'}|. \tag{1}$$

where p and p' are the corresponding pixels in left and right frames respectively. So the matching cost can be expressed as follows.

$$Cost(x,y,d) = \sum_{(\eta,\xi)\in\omega} \min\left(d_c(L(x+\eta, y+\xi), R(x+\eta-d, y+\xi)), T_c\right). \tag{2}$$

where L and R are the left and right images and d is the disparity of pixel (x,y). $\eta,\xi \in [-w_s/2, w_s/2]$. ω represents the local window of size w_s. T_c refers to the color distance truncation threshold. While dealing with these pixels where $x < d$ in left frame or $x + d >= width$ in right frame, the matching cost is set to T_c. Then we use the Winner Take All (WTA) [1] method to obtain the best candidate as the initialization of disparity map:

$$D(x,y) = \arg\min_d Cost(x,y,d). \tag{3}$$

The parameters for local matching are discussed in [1], accordingly, we suggest to choose $w_s = 3$ and $T_c = 150$. Although the disparity map selects a minimum value for each pixel, the quality of the value cannot be evaluated well. So we define the confidence to describe the reliability for each selection. Based on the matching cost array obtained above, the confidence map is initialized by the ratio of minimum cost and second-minimum cost.

$$Conf(x,y) = \left(1 - \frac{\min_d Cost(x,y,d)}{\min_{d \neq D(x,y)} Cost(x,y,d)}\right) \times 100. \tag{4}$$

where $\min_{d\neq D(x,y)} Cost(x,y,d)$ means the minimum cost when $d \neq D(x,y)$. If the second minimum equals to the minimum, the candidate should be unreliable. We normalize the value of confidence from 0 to 100. After that, the left-right check [1] is used to adjust the confidence map. If the corresponding pixels p and p' in left and right frames have the same disparity, the confidence of p and p' are updated to the sum of their confidence. Otherwise, the pixels' confidence should be reduced by a constant c_f.

2.2 Confidence-based Support Window Selection

Although the initialization of disparity map provides coarse estimation, there are still many error pixels, especially those with low confidence in smooth region and occluded region. Gupta [6] considered the color similarity in local window to rectify some errors, but the simple count of disparity values is noise in low confidence region. For this reason, in our algorithm, the color similarity and confidence are combined. Figure 3 shows an example of the disparity reassignment in the CSW.

Considering that those pixels with low confidence have ambiguous disparity, only they will be chosen as the center of the support window to reduce computation. For each low confident pixel p, R_p means the support window with the center at p. First we search the binary window S_p in which pixels have intensity difference less than a threshold T_s with p in LAB color space as introduced in [6].

$$S_p(q) = \begin{cases} 1 & if\ d_{Lab}(p,q) < T_s \\ 0 & otherwise. \end{cases} \tag{5}$$

where q is any point in R_p and $d_{Lab}(p,q)$ is the Euclidean distance of intensity difference.In order to weaken the effect on S_p caused by T_s and reduce the computational cost, we design the following operations:

$$S'_p = \begin{cases} Dilate(S_p, s_1)\ if\ N(S_p) < T_N \\ Close(S_p, s_1)\ otherwise. \end{cases} \tag{6}$$

As shown in Fig. 2, if the number of pixels in S_p is less than a threshold T_N, it means that the current pixel is discriminative enough (the lower row). So a morphologic DILATE is operated on S_p to pick up those separate pixels at one time. On the other hand, if the number of pixels in

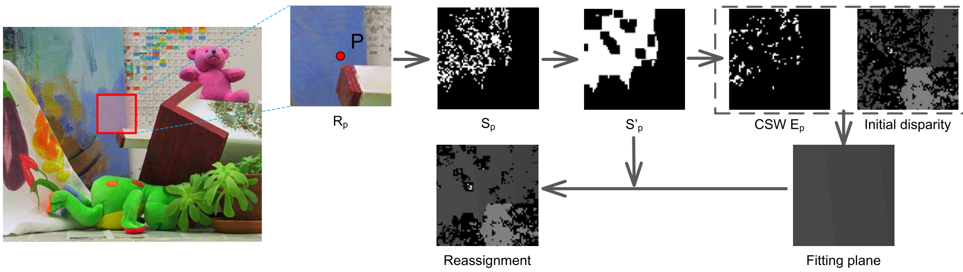

Fig. 3 The disparity reassignment in the CSW at a center of pixel p.

S_p is larger than T_N, it means that there is a smooth region in the support window (the upper row). Most pixels in this case are not reliable enough. So we operate a morphologic CLOSE on S_p to fill inside the smooth background but keep the boundary. The two operations increase the rate of reliable pixels. From S'_p, the CSW E_p is obtained by picking up all the pixels with high confidence larger than a threshold T_{conf} (we set $T_{conf} = 2c_f$).

$$E_p(q) = \begin{cases} 1 & if\ Conf(q) > T_{conf} \\ 0 & otherwise. \end{cases} \tag{7}$$

Next we will use all the pixels in E_p to calculate the equation of the disparity plane. From this equation, the disparity is reassigned for each pixel in S'_p. Those pixels which have already been reassigned will not be chosen as the center of CSW to accelerate the process.

2.3 Local Plane Fitting

To keeps the smoothness of the disparity map, the plane fitting method is usually applied for disparity map refinement. Traditionally, it requires a color segmentation which is high computation and sensitive to parameters. In our scheme, we adopt the local plane fitting on the confidence-based support window for the first time to replace segmentation. In the local area, smooth surface can be well approximated as the plane. Thus CSW has better adaptability at different positions compared with segmentation since the reference pixel is changing with the window. To suppress the disturbance resulted from outliers, we employ RANSAC [10] to solve local plane fitting problem. The equation of disparity plane is defined as $D = Ax + By + C$, where (x, y) denotes the 2-D position and D is disparity of current pixel. A, B and C is the parameter of the plane. In E_p, each three different pixels are picked up at random to determine a disparity plane. All the pixels q which are close enough to the plane will be counted to $N_R(E_p)$: $|Aq_x + Bq_y + C - D_q| < \Delta_d$, Where Δ_d is the disparity scalar. The RANSAC process will break if the number of iteration reaches the setting maximum or the rate of satisfied pixels is larger than α as follows.

$$R_s(E_p) = \begin{cases} 1 & if\ N_R(E_p) > \max(T_N, N_E * \alpha) \\ 0 & otherwise. \end{cases} \tag{8}$$

where N_E is the total number of the pixels in E_p and α is the accuracy rate. T_N is the number to reject cases of too few pixels. If the process returns 1, all the pixels in S'_p will be reassigned by the disparity plane. Furthermore, it was shown that the accuracy rate $\alpha = 0.75 \sim 0.9$ is suitable for most cases [10].

To cope with the multiple-reassignment problem, we make use of the pixels' space relationship. The assumption is that the closer pixels have the more similar disparity. In each CSW, the new distance between current pixel and the central pixel is recorded. Only if the new distance is smaller than the old value, the disparity and the distance of current pixel are updated.

3. Experiments

The proposed algorithm has been implemented using C language and under a 2.80 GHz Core 2-duo computer. In our implementation, the size of CSW is $w_b = 67$. The threshold of color similarity T_s in LAB color space is within the range of $1.2 \sim 3.5$. The confidence adjustment parameter c_f is set to 10. The minimum number of the fitting pixels T_N is 15.

Scharstein [1] presents a popular standard evaluation platform for the current stereo matching algorithm. The algorithms are evaluated by the error rate of none occlusion region, discontinuity region and the whole image. The experiments on the Middlebury dataset are shown in Fig. 4. It is shown that our algorithm performs well in both smooth region and the color edge. In Table 1, there are our results comparing with some best local based algorithms. Since the local plane fitting model is more suitable than simple counter in binary window, our performance is much better on discontinuity region and slant planes than [6]. The confidence selection strategy eliminates a lot of redundant computing. As shown in Table 2, the speed of our algorithm is about 2 seconds on Tsukuba image (size of 384×288 pixels with 16 disparities), which is about 30 times faster than [4] and [5], while the error rate is comparable. Our algorithm is a better choose in the balance of performance and speed.

Figure 5 shows other results on the dataset [9]. The plane model is adaptive for the spherical surface of the counter in baby data and the ball in bowling data, because the sphere with large radius can be dealt with as a plane in a local area. However, our algorithm is not suitable for the unsmooth region with similar color. One example is the wrong disparity in the rectangle at the left bottom of cones data.

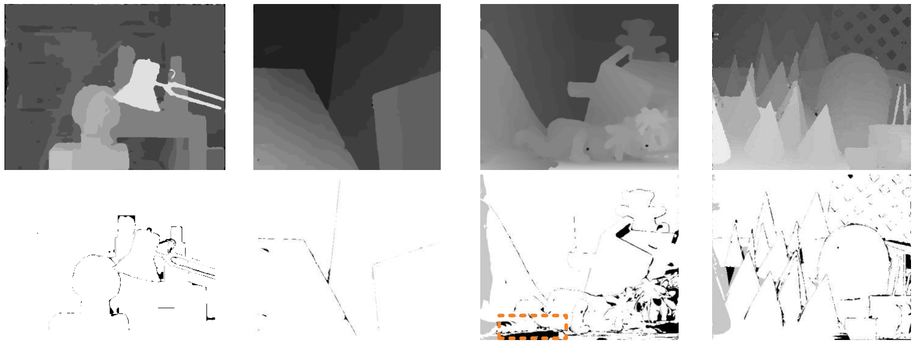

Fig. 4 Results on the Middlebury data set (Tsukuba, Venus, Cones and Teddy). The upper row is the depth map by our algorithm. The lower row shows the error pixels larger than one scale comparing with ground truth.

Table 1 Ranks of local based algorithms on [8].

Algorithm	Rank	Error Rate (All)				AER
		Tsukuba	Venus	Teddy	Cones	
GeoSup [5]	18	1.83	0.26	13.2	8.89	5.80
Ours	23	1.83	0.65	11.4	8.60	5.75
C-SG [7]	24	3.29	0.57	11.8	8.35	5.76
AdapW [4]	40	1.85	1.19	13.3	9.79	6.67
ABW [6]	49	1.67	0.65	18.3	12.6	7.90
SSD+MF [1]	97	7.07	5.16	24.8	19.8	15.7

Table 2 Time cost of the modules on [8].

Processing Time	Tsukuba	Venus	Teddy	Cones
Local Matching/ms	313	688	3110	3094
Plane Fitting/ms	2078	2856	5160	7328
Total/ms	2391	3544	8270	10422

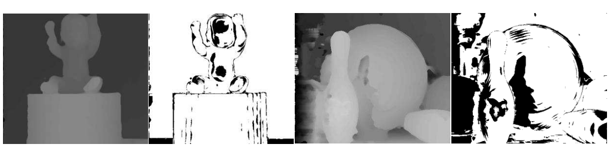

Fig. 5 Results on baby and bowling on dataset [9].

4. Conclusion

In this paper, a new local based algorithm using CSW is proposed. The CSW selects discriminative pixels not only according to the color similarity, but also the confidence of disparity value. The confidence is initialized by the matching cost and then adjusted by left right check operation. Based on these valid pixels, a local plane fitting by RANSAC is operated to estimate the disparity for low confidence pixels. The experiments show that our algorithm is one of the best local based method on Middlebury benchmark.

However, the accuracy on object boundaries should be improved and the proposed approach is not having subpixel precision. In the future work, we will design a confidence updating scheme to overcome those problems.

References

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," IJCV, vol.47, no.1, pp.7–42, 2002.

[2] S. Birchfield and C. Tomasi, "Depth discontinuities by pixel-to-pixel stereo," Proc. ICCV, pp.1073–1080, 1998.

[3] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," IEEE Trans. Pattern Anal. Mach Intell., vol.16, no.9, pp.920–932, 1994.

[4] K.-J. Yoon and I.S. Kweon, "Stereo matching with the distinctive similarity measure," Proc. ICCV, pp.1–7, 2007.

[5] A. Hosni and M. Bleyer et al., "Local stereo matching using geodesic support weights," Proc. ICIP, pp.2093–2096, 2009.

[6] R. Gupta and S.-Y. Cho, "Real-time stereo matching using adaptive binary window," Proc. 3DPVT, 2010.

[7] H. Hirschmer, "Stereo vision in structured environments by consistent semi-global matching," IEEE Trans. Pattern Anal. Mach Intell., vol.30, no.2, pp.328–341, 2008.

[8] http://vision.middlebury.edu/stereo/

[9] H. Hirschmer and D. Scharstein, "Evaluation of cost functions for stereo matching," Proc. CVPR, pp.1–8, 2007.

[10] M.A. Fischler and R.C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," Commun. ACM, vol.24, no.6, pp.381–395, 1981.