Hybrid Plane Fitting for Depth Estimation

Lingfeng XU, Oscar C. AU, Wenxiu SUN, Yujun LI, Jiali LI
The Hong Kong University of Science and Technology, Hong Kong
E-mail: {lingfengxu, eeau, eeshine, liyujun, jiali}@ust.hk

***Abstract*—in this paper, a novel plane fitting algorithm with low complexity and high accuracy is proposed to refine the depth maps generated by stereo matching. We first compute the confidence coefficient for each pixel in the depth map by cross checking and stable pixel calculation. According to the outlier pixel percentage for each segment, we choose one method, either proposed weighted least square error based or RANSAC based plane fitting algorithm, to estimate the plane parameters. Experimental results show that our method outperforms other existing plane fitting algorithms.**

I. Instructions

Stereo matching is one of the most active research topics in computer vision. By capturing images/videos by stereo cameras, we can rectify the stereo images [1], and estimate the depth maps by disparity estimation [2]–[6]. The depth maps can be applied into view synthesis, object tracking, image based rendering, etc.

According to Daniel's paper [4], stereo matching generally performs (subsets of) four steps: 1. Matching cost computation; 2. Cost (support) aggregation; 3. Disparity computation/optimization; 4. Disparity refinement. The first step measures the similarities between two pixels, one is from left image and another one is from the right image. Some common used methods are absolute difference, square difference and cross-correlation, etc. In order to include more texture information and reduce the affection of the noise and color inconsistence between the stereo images, the similarity measurements are aggregated in a local region, such like square window, shiftable window, or weighted window. After that, the disparity of each pixel can be estimated by local methods like winner take all approach or some other global methods such as graph cuts and belief propagation, which include the smoothness constraint between neighboring pixels. After the three steps, an initial depth map can be generated with some outliers in the occlusion regions and textureless region. Some post-processing methods are proposed to refine the depth maps by cross checking, median filter, plane fitting, etc. In general, those methods can clean up most of the mismatches and holes in the depth maps.

Plane fitting is one of the most efficient ways for disparity refinement. Based on the assumption that pixels within one segment are co-planer, the plane coefficients are estimated. Pixels which are far away from the plane are considered as outliers and replaced by the depth of the plane. Usually, some over-segmentation methods such like mean-shift [7] are utilized for segmentation.

Many plane fitting algorithms are proposed in recent years. The most common used methods are RANSAC [8]–[10] based plane fitting. RANSAC is a minimization algorithm that can exclude the outliers. In RANSAC based plane fitting algorithm, three pixels are randomly chosen to calculate the plane function during each iteration. And the accuracy of the plane is estimated by counting the number of active pixels inside the segment. The advantage of this method is insensitive to the outliers. RANSAC can exclude the outliers and generate quite reasonable plane coefficients. However, the drawbacks are also obvious. The computation complexity is extremely high, especially when large numbers of iterations are performed. Another disadvantage is that when estimating the plane function for large and slant planes (e.g. more than 1000 pixels within a segment), randomly chosen planes can be easily trapped into local extremes and generate inaccurate results.

Another algorithm utilizes least square error (LSE) based plane fitting algorithm [6], [11]. Since the cost function of this method is convex, a close form solution can be found and computation load is very small. The method works well when there are few outliers. However, least square error is quite sensitive to noise. When it comes to the occlusion region where large number of wrong disparity exists, the plane fitting will fail.

Other algorithm, like Klaus's robust plane fitting [12]. The main idea is quite similar to RANSAC algorithm. The difference is that horizontal slant and vertical slant are estimated separately and integrated together in the end.

Our approach provides the advantage of both RANSAC based and LSE based plane fitting algorithm. We first exclude the occlusion pixels by a novel robust cross checking. And then assign each remaining pixel a confidence level by measuring its stableness. Based on the non-occlusion pixel percentage, either RANSAC or weighted LSE based fitting algorithm is selected to fit each segment. The experimental results show that our method is more accurate than existing methods while keeping a low complexity. The rest of the paper is organized as follows: section II will explain the detailed algorithm of proposed hybrid plane fitting. Experimental results will be shown in section III, followed by the conclusions and future work in section IV.

II. Proposed plane fitting algorithm

Plane fitting assumes that pixels within one segment are co-planar. This assumption is quite reasonable in the case of over-segmentation. By estimating the plane coefficients, we can filter the wrong disparities and replace them by the plane disparities.

Our paper focuses on the problem of plane fitting. For the first three steps in stereo matching: matching cost computation, cost (support) aggregation and disparity computation/optimization, we utilize the most common used stereo matching algorithms to estimate the initial depth maps. Our main contribution of this paper is to refine the depth maps calculated by the existing methods. A robust cross checking is first applied to filter out most of the outliers. For the remaining pixels, we calculate the stableness and assign each pixel a weighting coeffcient which can be used in the weighted LSE based plane fitting. For the segment contains too many outliers, we apply RANSAC based plane fitting to increase the robustness of fitting results.

A. *Initial the depth maps*

There are hundreds of stereo matching algorithms. In this paper, the most common used stereo matching algorithms are chosen to initial the depth map. Our method can be also applied to refine the depth maps generated by any other stereo matching algorithms. For the matching cost computation, truncated absolute difference is chosen in our algorithm.

$$m(x,y,d) = \min(|I_L(x,y) - I_R(x-d,y)|, T) \tag{1}$$

Here (x,y) is a pixel in the left image, $(x-d,y)$ is the pixel in the right image. T is a threshold that controls the maximum of the matching cost in order to reduce the noise affection.

We utilize the adaptive support weight approach for cost aggregation. [10], [11]

$$E_L(x,y,d) = \frac{\sum\limits_{(i,j)\in\Omega(x,y)} w_{L(x,y)}(i,j)\cdot w_{R(x-d,y)}(i-d,j)\cdot m(i,j,d)}{\sum\limits_{(i,j)\in\Omega(x,y)} w_{L(x,y)}(i,j)\cdot w_{R(x-d,y)}(i-d,j)} \tag{2}$$

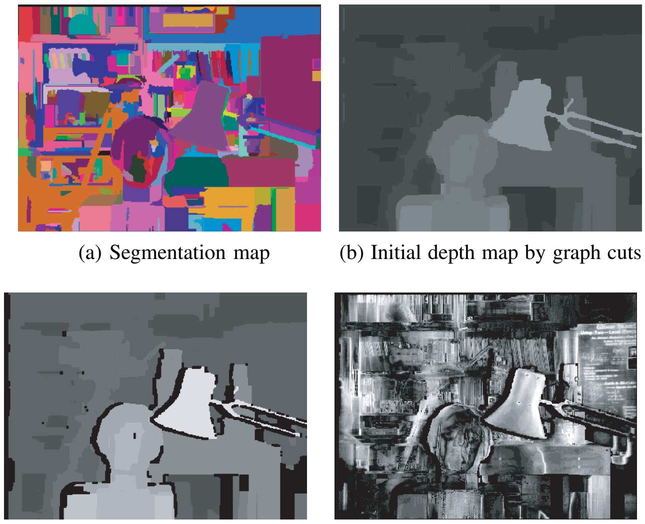
It is a weighted average of the matching cost within a window $\Omega(x,y)$. (x,y) is the center pixel and (i,j) is a pixel inside the window. The weighting coefficient assigns large weight when (i,j) shares the similar color and is close to the center pixel.

$$w_{(x,y)}(i,j) = \exp(-\alpha|I(x,y) - I(i,j)| - \beta||(x,y) - (i,j)||_{l_2}) \tag{3}$$

After aggregating the matching cost with the adaptive support, the similarity measurement can be derived. In order to reduce the noise in the texture-less region, smoothness constraint is added.

$$E(d) = \sum_p E_L(d_p) + \sum_{(p.q)} V_{(p,q)}(d_p, d_q) \tag{4}$$

Here $E_L(d)$ is the data term which we have calculated in the weighted block matching in (2). $V_{(p,q)}$ assigns smoothness penalty to adjacent pixels p and q. In our approach we assign large smoothness penalty to pixels inside a segments and small penalty to those between segments to encourage depth discontinuous around segment boundaries. Mean-shift based segmentation [7] is chosen because of its low complexity and high accuracy. Then graph cuts based minimization algorithm is applied to minimize the cost function (4). An initial depth map can be generated.



(a) Segmentation map (b) Initial depth map by graph cuts

(c) Robust cross checking result (d) Rotation Confidence map

Fig. 1: Intermediate results for 'tsukba'

B. *Robust cross checking*

Cross checking is a common used method to exclude the outliers in occlusion region. It requires that the disparities from left and right disparity maps are consistent:

$$d_L(x,y) = d_R(x - d_L(x,y), y) \tag{5}$$

In our algorithm, we allow the disparity to have a small variation, which can help include more reliable pixels for plane fitting. The pixel is treated as non-occlusion pixel if and only if:

$$|d_L(x,y) - d_R(x - d_L(x,y), y)| \leq 1 \tag{6}$$

Based on the observation that most of the wrong disparities exist around the occlusion region, we further dilate the occlusion region to exclude more outliers.

The occluded pixels are marked as black in Fig. 1.c. The outliers are further excluded by dilating the occlusion regions.

C. *Confidence level calculation*

For the remaining non-occluded pixels, there are still some outliers. Here we introduce a new term named confidence level, to estimate how accurate of each pixel. The confidence level is related with the sharpness of the matching cost curve and the distance from graph cuts results and local minimum disparities:

$$C(i) = \frac{E_{\min 2}(i) - E_{\min 1}(i)}{E_{\min 2}(i)} \cdot \exp(-|d_{global}(i) - d_{local}(i)|) \tag{7}$$

$Emin1$ is the minimum of matching cost curve derived by (1). And $Emin2$ is the second minimum. The first part of (7) calculates the sharpness of the cost curve. If the value is larger, the disparity value is more trustable. The second term measures the distance from global disparity and local disparity. The former disparity is calculated with the smoothness constraint and the latter one is derived by winner take all approach. If the two disparities are close to each other, we have a more confident level.

As shown in Fig. 1.d, the whiter the pixel is, the higher the confidence is. High confidence pixels usually existed in the high texture region because the texture can provide more information for correspondence. Those high confidence pixels are more accurate and will be assigned larger weight during the following plane fitting procedure.

D. Hybrid plane fitting

By proposed cross checking and occlusion region dilation, most of the wrong disparities are detected. The plane coefficients can be estimated by the remaining non-occluded pixels. Both RANSAC and least square error based methods can be applied to fit the planes. Generally speaking, RANSAC based method is more accurate when there are some outliers, but has a much higher complexity. Least square error based method can estimate accurate plane coefficients with extremely low computation but is sensitive to the noise. Here we propose a hybrid fitting algorithm. For those with only a few outliers, we utilize our proposed weighted least square error based method. Each pixel is assigned a weight which is based on its confidence level. By this procedure, weighted LSE based method can be not only fast, but also robust to noise. For some segments contain many wrong disparities, RANSAC based method is applied to filter out the noise.

The noise in the depth map is feature based. Most of the outliers are around the depth discontinuous regions. Some of them are excluded during the cross checking procedure. We further remove the errors by dilating the occlusion regions. After that, for each segment, we can calculate its non-occluded pixel percentage (NOP) by:

$$NOP(i) = \frac{NNO(i)}{N(i)} \tag{8}$$

Here $NOP(i)$ is the non-occluded pixel percentage. The numerator is the number of non-occluded pixel for segment i. And $N(i)$ is the total pixel number for that segment. NOP measures the reliability of a segment. If NOP is small, that means the segment is close to the depth discontinuous region where noise can easily appear, RANSAC based method are chosen. When NOP is large, the segment is not in the textureless region or depth discontinuous region, so the chance of weighted LSE survives is high. Even if there are some outliers, because we have large number of reliable pixels, the outliers will not affect the plane function too much.

A threshold is chosen to determine whether RANSAC based or weighted LSE based method should be applied. If it is set as 1, then RANSAC is performed to all the segments. If threshold is 0, then all the plane coefficients are estimated by weighted LSE based algorithm. In our experiment, we set it as 0.7.

The main idea of RANSAC based algorithm is to randomly choose three points to construct a plane and count the pixel numbers which are on the planes.

$$\begin{aligned} &[\hat{a}, \hat{b}, \hat{c}] = \arg\max_{[a,b,c]} \sum_i f_i(a, b, c) \\ &\text{where} \\ &f_i(a, b, c) = \begin{cases} 1 & \text{if } |d_i - ax_i - by_i - c| < Th \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{9}$$

(x_i, y_i) is coordinate for pixel i while d_i is the disparity value generated by previous stereo matching algorithm. We want to derive the plane coefficients (a, b, c) to maximum the reliable pixel number. The function is an NP hard problem. We minimize it by randomly choose some points to construct the plane and calculate the cost. The plane with the maximum number of reliable pixels is chosen to represent the segment. RANSAC method can excludes the outliers very efficiently. However, because of the limitation of iteration time, the final result is just a local maximum of the cost function. Sometimes the plane function is not correct when the local maximum is far away from the global maximum.

The target of our proposed weighted LSE based algorithm is followed:

$$[\hat{a}, \hat{b}, \hat{c}] = \arg\min_{[a,b,c]} \sum_i w_i (d_i - ax_i - by_i - c)^2 \tag{10}$$

The weighting coefficient w_i is the confidence level we calculated above. Compare with traditional LSE based algorithm which is quite sensitive to the noise. Our method assigns small weight to the noise so more accurate plane coefficients can be derived. Since the square error function is derivative, close form solution can be solved:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix}^T = \begin{bmatrix} \sum_{i=1}^{m} w_i x_i d_i \\ \sum_{i=1}^{m} w_i y_i d_i \\ \sum_{i=1}^{m} w_i d_i \end{bmatrix}^T \begin{bmatrix} \sum_{i=1}^{m} w_i x_i^2 & \sum_{i=1}^{m} w_i x_i y_i & \sum_{i=1}^{m} w_i x_i \\ \sum_{i=1}^{m} w_i x_i y_i & \sum_{i=1}^{m} w_i y_i^2 & \sum_{i=1}^{m} w_i y_i \\ \sum_{i=1}^{m} w_i x_i & \sum_{i=1}^{m} w_i y_i & \sum_{i=1}^{m} w_i \end{bmatrix}^{-1} \tag{11}$$

The computation load is extremely small comparing with RANSAC based one and the global minimum is found in the algorithm. In the implementation, when NOP is large, only weighted LSE is utilized. When NOP is small, RANSAC based algorithm is applied. To avoid the RANSAC generates 'bad' local maximum. We also calculate the plane cost by the plane coefficients generated by weighted LSE based algorithm and compare it with the local cost generated by RANSAC. Larger cost will be selected in the end.

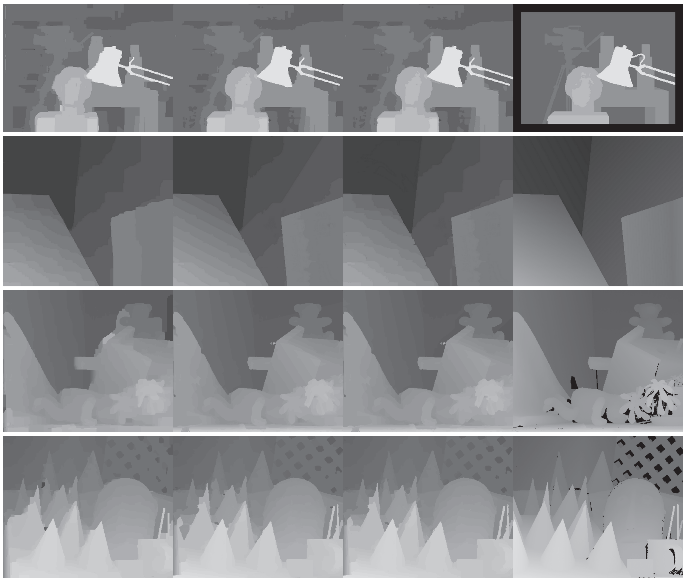
III. EXPERIMENTAL RESULTS

In this section, we present the experimental results of proposed algorithm on the Middlebury benchmarks $http : //vision.middlebury.edu/stereo/$. The simulations are performed in MATLAB 7.10 using an Intel Core i3, 3.06- GHz processor with 8GB of memory.

We compare our method with traditional RANSAC based plane fitting algorithm because it is the most common used

TABLE I: Performance Comparison of the Proposed Method

	Tsukuba			Venus			Teddy			Cones		
	nonocc	all	disc.	nonocc	all	disc.	nonocc	all	disc.	nonocc	all	disc.
Hybird based	**1.12**	**1.43**	**6.03**	**0.15**	**0.34**	**1.92**	**4.41**	**9.57**	**12.4**	**2.72**	**8.00**	**7.86**
RANSAC based	1.13	1.42	5.97	0.17	0.29	1.86	5.38	10.4	13.2	2.69	7.86	7.64
No plane fitting	1.16	2.38	6.23	0.64	1.23	8.57	10.5	16.5	24.5	6.40	12.4	17.1



(a)No plane fitting (b) Hybird based (c) RANSAC based (d) Ground truth

Fig. 2: Results for (from top to bottom) Tsukuba, Venus, Teddy, and Cones image pairs

method. Some other algorithms, like Klaus's robust plane fitting [12], have similar idea and performance with RANSAC.

In our experiments, the initial depth maps are generated by graph cuts based algorithm described in the last section. The generated depth maps are compared with the groundtruth and bad pixel percentages are calculated in three different regions: non-occlusion regions, all regions and discontinuous regions. The results are shown in Table I. From the table we can see that our method can generate better results when slant planes exist (e.g. Test image: Teddy, Venus). For the other test image, the results are similar. The depth maps refined by our proposed hybrid method and RANSAC based method are quite similar. The main reason is that most of the segments are flat planes. In that case, both of the methods generate the same results. For the slant plane, hybrid algorithm can derive a more reasonable result, especially when the segment are large and slant.

TABLE II: Complexity Comparison

	RANSAC based	Hybrid based	Complexity reduction
Tsukuba	28.19 s	0.85 s	97%
Venus	41.78 s	0.98 s	98%
Teddy	47.61 s	4.48 s	91%
Cones	36.93 s	2.93 s	92%
Average	154.51 s	9.24 s	94%

In the other hand, we also compare the two algorithms from the angle of complexity. The advantage of our method is quite obvious. As shown in II, our method have more than 94% computation reduction. When the size of the segment increases, RANSAC based method will be slower, while hybrid method keeps similar complexity.

IV. Conclusions

In this paper, we proposed a novel plane fitting algorithm with weighted LSE minimization and RANSAC based outlier rejection. Experimental results show that our method has a better performance than traditional plane fitting algorithms, especially when compared with computation complexity.

V. Acknowledgement

This work has been supported in part by the Research Grants Council (RGC) of the Hong Kong Special Administrative Region, China. (GRF Project no. 610210)

References

[1] L. Xu, O. Au, W. Sun, Y. Li, S. H. Chui, and C. W. Kwok, "Image rectification for single camera stereo system," in *Image Processing , IEEE International Conference on*, 2011.

[2] J. Sun, Y. Li, S. Kang, and H.-Y. Shum, "Symmetric stereo matching for occlusion handling," in *Computer Vision and Pattern Recognition. IEEE Computer Society Conference on*, 2005.

[3] Y. Li, O. Au, L. Xu, W. Sun, S.-H. Chui, and C.-W. Kwok, "A convex-optimization approach to dense stereo matching," in *Image Processing, IEEE International Conference on*, 2011.

[4] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *Stereo and Multi-Baseline Vision, IEEE Workshop on*, 2001.

[5] Y. Kuk-Jin and K. In-So, "Adaptive support-weight approach for correspondence search," in *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2006.

[6] H. Li and G. Chen, "Segment-based stereo matching using graph cuts," in *Computer Vision and Pattern Recognition. IEEE Computer Society Conference on*, 2004.

[7] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," in *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2002.

[8] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," in *Communications of the ACM*, 1981.

[9] S. N. Sinha, D. Steedly, and R. Szeliski, "Piecewise planar stereo for image-based rendering," in *Computer Vision, International Conference on*, 2009.

[10] Q. Yang, L. Wang, R. Yang, H. Stewenius, and D. Nister, "Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling," in *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2009.

[11] T. Masayuki, F. Toshiaki, and S. Kazuyoshi, "Depth estimation reference software (ders) with image segmentation and block matching," in *ISO/IEC JTC1/SC29/WG11, MPEG2008/M16092*, 2009.

[12] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," in *Pattern Recognition, IEEE International Conference on*, 2006.