IMPROVING SUBPIXEL STEREO MATCHING WITH SEGMENT EVOLUTION

Yao-Jen Chang[1], *Hung-Hsun Liu*[2], *Tsuhan Chen*[1]

[1]School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA
[2]Telecommunication Laboratories, Chunghwa Telecom Co., Ltd.,Taoyuan, Taiwan

ABSTRACT

Segmentation-based approach has shown significant success in stereo matching. By assuming pixels within one image segment belong to the same 3D surface, robust depth estimation can be achieved by taking the whole segment into consideration. However, segmentation has been mostly used for stereo matching at integer disparities rather than subpixel disparities. One major reason is that small segments may be insufficient for estimating surfaces like slanted planes, while large segments may contain segmentation errors impacting the accuracy of depth estimation. In this work, we propose a segmentation-based scheme for subpixel stereo matching. Instead of using a fixed segmentation, segments are evolved to find a better support for reliable surface estimation. Given an initial estimation of segmentation and depth, the proposed algorithm jointly optimizes the segmentation and depth by evolving the segmentation at the pixel level and updating the plane parameters at the segment level. Justified with experiments performed on the Middlebury benchmark, we show that the proposed method achieves significant improvements for subpixel stereo matching.

Index Terms— Stereo vision, Image segmentation, Surface fitting

1. INTRODUCTION

Stereo matching is a fundamental problem in computer vision that estimates depth of a 3D scene with a pair of images. With a well-established Middlebury benchmark established by Scharstein and Szeliski [1], new approaches can be easily evaluated on a common foundation, thereby boosting the research advancement of stereo matching.

As pointed in [1], subpixel accuracy is crucial for applications like image-based rendering. However, it receives much less attentions than pixel level accuracy evaluated at integer disparities. Among the top performers on the benchmark at integer disparities, the idea of using color-based segmentation proposed by Tao et al. [2] is widely adopted for dealing with unreliable depth estimation in textureless regions. By assuming pixels within the same homogenous region belong to the same 3D surface, robust depth estimation can be achieved by taking the whole region into consideration. However, segmentation errors may also lead to erroneous depth estimation when the assumption is violated. Zitnick and Kang [3] proposed to restrict the impact of segmentation errors by over-segmentation with lots of small segments. Taguchi et al.[4] further proposed an adaptive over-segmentation approach to handle segmentation errors. Since segments are too small for surface estimation, fronto-parallel planes constraint is imposed that sacrifices subpixel accuracy. On the other hand, Bleyer and Gelautz [5] and Klaus et al. [6] proposed to group similar segments together for robust plane fitting, thus achieving subpixel accuracy. However, it suffers from the initial segmentation errors since plane fitting and segment clustering are both based on segments fixed by the initial segmentation. A matting method is further proposed [7] to alleviate the impact of small segmentation errors near segment boundaries, but large segmentation errors are left unresolved. Based on these observations, segmentation seems to be less promising for subpixel stereo matching.

For subpixel accuracy, simple methods such as curve fitting to the matching costs at discrete disparity levels have been utilized for fast computation [1, 8]. Until recently, Yang et al. [9] proposed a super-resolution scheme based on bilateral filtering for disparity refinement. The bilateral filter works like soft color segmentation that preserves discontinuities by considering the color differences in addition to spatial differences. Gehrig and Franke [10] similarly proposed to use adaptive smoothing for edge-preserving disparity smoothing, which is incorporated in the depth estimation modeled as an energy minimization problem. The effectiveness of [9, 10] encourage us to revisit segmentation for subpixel stereo matching.

Sharing the concept proposed by Hoiem et al. [11] where several sub-tasks benefit each other in a closed-loop to accomplish the scene interpretation task, we propose to jointly optimize image segmentation and depth estimation in a closed loop for subpixel stereo matching. Instead of using a fixed segmentation, segments are evolved based on the depth information to provide a better support for reliable surface estimation. In the next section, an overview of the proposed framework and detailed algorithms are presented. Experiments initiated with different stereo matching algorithms are conducted in Section 3. Finally, Section 4 concludes and addresses several possible extensions.

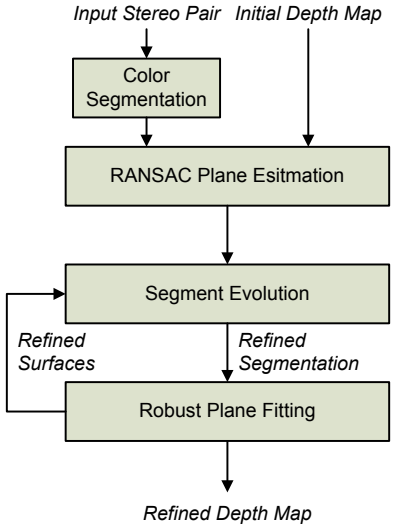

Fig. 1. The conceptual flow diagram of the proposed work.

2. THE ALGORITHM

The conceptual flow diagram of the proposed framework is depicted in Fig.1. By approximating a 3D scene as a collection of planar surfaces, depth for each pixel can be derived from its corresponding surface at subpixel precision. First of all, an initial depth map and image segmentation are provided for initialization to obtain initial plane parameters for each segment. Segment evolution and robust plane fitting are then instantiated alternatively with refined information provided by each process. Analogously, the proposed algorithm can be interpreted as a *k*-means clustering algorithm [12] in a broad sense, where initial seeds are given at first and then membership assignment and cluster update are performed alternatively to achieve optimization. Detail descriptions for each process are given in the following sub-sections.

2.1. Initialization

Instead of acting as an individual stereo matching algorithm, our framework can work with other stereo matching algorithms to refine its depth estimation. Accompanied with an initial image segmentation such as the mean-shift segmentation proposed by Christoudias et al. [13] , we perform robust plane estimation with RANSAC [14] to obtain initial plane parameters for each segment. Segment evolution and robust plane fitting are followed to obtain a better plane parameterization and image segmentation.

2.2. Segment Evolution

The goal of segment evolution is to adapt the support of each segment such that the points within a segment correspond to the same planar surface in the 3D space. This can be taken as a labeling problem where each pixel is assigned with a plane label that minimizes a global energy function. Instead of allowing all plane labels to be assigned, we restrict the candidate set of plane labels assigned for each pixel s to be the labels of neighboring pixels within a $W_p \times W_p$ window centered at s. This is equivalent to deforming a segment within a certain range from its original shape, thus adapting its support.

The global energy function can be modeled with a data term and a weighted smoothness term: $E = E_{data} + \lambda_{smooth} E_{smooth}$. The data term E_{data} is defined by the the color inconsistency cost of each pixel s on the left view image I_L with its corresponding point on the right view image I_R related by a homography h_s associated with the plane assigned to s:

$$E_{data} = \sum_{s \in I_L} (1 - o_s) \min(f(I_L(s), I_R(h_s(s))), T_f), \quad (1)$$

where $o_s \in \{0, 1\}$ indicates its occlusion state derived from the current depth map via Z-buffer testing similar to [5], the function f is the Birchfield and Tomasi's pixel dissimilarity measure [15], together with a truncation threshold T_f to form a robust error measure. The smoothness term E_{smooth} is defined by incorporating smoothness constraints imposed on three weighting functions on every two neighboring pixels s and t on the left view image I_L:

$$E_{smooth} = \sum_{s,t \in \mathcal{N}, s<t} w(o_s, o_t) e_d(s,t) D(s,t,h_s,h_t), \quad (2)$$

where the occlusion consistency weighting function $w(o_s, o_t)$ discourages neighboring pixels to be assigned to the same plane if pixel s is under occlusion but t is not. The color consistency function $e_d(s,t)$ sets a larger penalty λ_e for assigning different planes to neighboring pixels with low edge strength between them. The last term of Eqn.(2) is a plane dissimilarity measure defined by:

$$D(s,t,h_s,h_t) = \delta(h_s \neq h_t) + \min(d(s,t,h_s,h_t), T_d), \quad (3)$$

with the first term acting as a plane inconsistency model, which is set to 1 if two neighboring pixels are assigned to different planes, and the second term imposing robust error measure on the disparity differences induced by projecting s and t to both planes assigned to them, plus the disparity differences induced by projecting segment centers to both planes based on the segmentation in the previous iteration. This would encourage different plane assignments to happen at the intersection of two planes, but discourages two planes with large angle differences to be connected together.

In addition to the original candidate set of plane labels for each pixel, a set of fronto-parallel planes within the disparity range of the original candidate set are also included to handle missing disparity planes caused either by initial segmentation error or depth estimation error. To handle large segmentation errors, the segment evolution can be performed multiple times before the next step of plane fitting. The global optimization is carried out with the Graph Cuts algorithm proposed by Boykov et al. [16].

(a) (b) (c) (d) (e) (f)

Fig. 2. Results of the proposed algorithm on the Middlebury stereo dataset: (a-c) initial segmentation obtained by mean-shift segmentation [13], depth map initiated with WarpMat [7], and associated error map, (d-f) refined segmentation, depth map, and associated error map. Black and gray pixels in (c) and (f) indicate error > 0.5 in unoccluded and occluded regions, respectively.

2.3. Robust Plane Fitting

With refined segmentation provided by segment evolution, the plane parameters of each segment is re-estimated via plane estimation with RANSAC [14] based on the depth information as done in the initialization stage. A robust plane fitting is then performed for each segment by using the gradient descent optimization with an iteratively re-weighted least squares framework proposed by Baker et al. [17], with which the forward-additive algorithm is utilized to estimate the homography warps from the left view to the right view of image pairs. To further speed up the image segmentation in the next iteration, adjacent planes are merged if fitting error is small. The depth map derived from the plane fitting is also quantized at the quarter-pixel precision to prevent over-fitting.

3. EXPERIMENTAL RESULTS

To evaluate the proposed framework for stereo matching, we test our algorithm initiated with the depth map obtained by several performers in the benchmark. The segmentation information are not available even for the segmentation-based approaches. Therefore, we utilize mean-shift segmentation with its default parameters and the minimal region size set to 64 pixels. The values of parameters used in our experiments are: $\lambda_{smooth} = 5, T_f = 15, \lambda_e = 5, T_d = 2, W_p = 9$, which are fixed for all initializations. The iterations of the segment evolution and robust plane fitting can be carried on until convergence, at the expense of computation load grows almost linearly with the number of iterations. Empirically, three iterations reach reasonable results.

Fig.2 shows one example of our algorithm initiated with the depth map generated by [7]. The initial segmentation contains segmentation errors and lots of small segments, while our refined segmentation provides better segment support for reliable surface estimation. Significant improvements can be observed by comparing the error maps of the refined depth map and the original depth map. Experiments with other stereo algorithms in the Middlebury benchmark are also conducted. Depending on whether segmentation is utilized and whether subpixel disparity is targeted, one or more representative performers in each category are investigated:

- Pixel-level without segmentation: GC+occ [18].
- Pixel-level with segmentation: DoubleBP [19], OverSegmBP [3], and AdaptOvrSegBP [4].
- Subpixel without segmentation: C-SemiGlob [8] and ImproveSubPix [10].
- Subpixel with segmentation: AdaptingBP [6], Segm+visib [5], WarpMat [7], and SubpixelDoubleBP [9].

By measuring average percentage of bad pixels over all four datasets, we compare the original performance and the

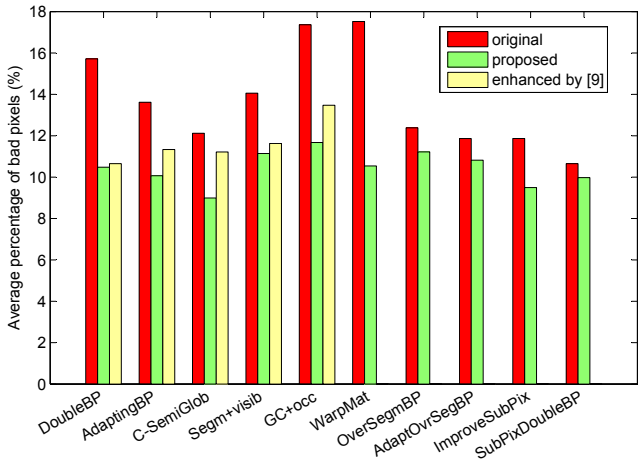

Fig. 3. Subpixel performance comparisons for various algorithms measured by average percentage of bad pixels with absolute disparity error > 0.5. The proposed algorithm significantly improves original performance of each algorithm, and outperforms the performance enhanced by [9].

enhanced performance in Fig.3. Among these algorithms, enhanced performance of DoubleBP, AdaptingBP, C-SemiGlob, Segm+visib, and GC+occ conducted by Yang et al. [9] are also included. For these five algorithms, the relative improvement of the proposed method reaches 27.74% in average compared to 19.23% provided by [9]. While the enhanced DoubleBP is the top performer reported in [9], the proposed method worked best with C-SemiGlob, which is a method targeted at subpixel precision without using segmentation. For the rest 5 algorithms with no performance reported in [9], the enhanced WarpMat gets most performance gain with our approach. The enhanced ImproveSubPix achieves the lowest error among these five algorithms, but not as good as the enhanced C-SemiGlob. Note that the proposed method can still improve SubpixelDoubleBP, which is the enhanced DoubleBP provided by [9].

4. CONCLUSIONS

In this work, we proposed a segmentation-based scheme for subpixel stereo matching. Significant improvements justifies that the incorporation of the depth information can lead to a better segmentation with segment evolution, which in turn helps surface estimation for providing more accurate depth for a 3D scene. Right now, we only take the left view of an image pair as the reference image for segment evolution. As suggested in several stereo matching algorithms [9, 19], symmetric treatment of both views may also improve segment evolution for finding a even better support. Moreover, the planar surface assumption is not always true in reality, especially for curved surfaces with low texture. The use of more sophisticated surfaces such as quadratic surfaces can be investigated for further improvement.

5. REFERENCES

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV*, vol. 47, pp. 7–42, 2002, `http://vision.middlebury.edu/stereo/`.

[2] H. Tao, H. Sawhney, and R. Kumar, "A global matching framework for stereo computation," in *ICCV*, 2001, pp. 532–539.

[3] C. L. Zitnick and S. B. Kang, "Stereo for image-based rendering using image over-segmentatiaon," *IJCV*, vol. 75(1), pp. 49–65, 2007.

[4] Y. Taguchi, B. Wilburn, and L. Zitnick, "Stereo reconstruction with mixed pixels using adaptive over-segmentation," in *CVPR*, 2008.

[5] M. Bleyer and M. Gelautz, "A layered stereo algorithm using image segmentation and global visibility constraints," in *ICIP*, 2004.

[6] A. Klaus, M. Sormann, and K. Karner, "Computing visual correspondence with occlusions using graph cuts," in *ICPR*, 2006, vol. 3, pp. 15–18.

[7] M. Bleyer, M. Gelautz, C. Rother, and C. Rhemann, "A stereo approach that handles the matting problem via image warping," in *CVPR*, 2009.

[8] H. Hirschmüller, "Stereo vision in structured environments by consistent semi-global matching," in *CVPR*, 2006.

[9] Q. Yang, R. Yang, J. Davis, and D. Nistér, "Spatial-depth super resolution for range images," in *CVPR*, 2007, `http://vis.uky.edu/~liiton/publications/super_resolution/`.

[10] S. Gehrig and U. Franke, "Improving sub-pixel accuracy for long range stereo," in *ICCV VRML Workshop*, 2007.

[11] D. Hoiem, A. A. Efros, and M. Hebert, "Closing the loop on scene interpretation," in *CVPR*, June 2008.

[12] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *BSMSP*, 1967, vol. 1.

[13] C. M. Christoudias, B. Georgescu, and P. Meer, "Synergism in low level vision," in *ICPR*, 2002, vol. 4.

[14] 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.

[15] S. Birchfield and C. Tomasi., "A pixel dissimilarity measure that is insensitive to image sampling," *IEEE Trans. PAMI*, vol. 20(4), pp. 401–406, 1998.

[16] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. PAMI*, vol. 23(11), pp. 1222–1239, Nov. 2001.

[17] S. Baker, R. Gross, I. Matthews, and T. Ishikawa, "Lucas-kanade 20 years on: A unifying framework: Part 2," Tech. Rep. CMU-RI-TR-03-01, Robotics Institute, Pittsburgh, PA, February 2003.

[18] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *ICCV*, 2001, vol. 2, pp. 508–515.

[19] Q. Yang, R. Yang, H. Stewénius, and D. Nistér, "Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling," *IEEE Trans. PAMI*, vol. 31(3), pp. 492–504, 2008.