

Robust segment-based Stereo using Cost Aggregation

Veldandi Muninder¹
 veldandi.muninder@nokia.com
 Ukil Soumik²
 soumik.ukil@nokia.com
 Govindarao Krishna²
 krishna.govindarao@nokia.com

¹ Nokia Technologies,
 Sunnyvale
 California, USA
² Nokia Technologies
 Bangalore
 India

Introduction Most segment based stereo methods estimate disparity by modeling color segments as 3-D planes [2]. Inherently, such methods are sensitive to segmentation parameters and intolerant to segmentation errors. Two main dependencies of these methods on the underlying segmentation algorithm are: size of segments used for estimating planes, and assignment of a single plane to the whole segment. Specifically, in the case of under-segmentation, there is a higher chance of merging multiple objects (with multiple plane surfaces) into a single segment. Consequently, planes estimated using these segments are erroneous. The effect propagates to the disparity map, wherein a larger segment encompassing multiple objects is incorrectly represented by a single disparity plane. In the over-segmentation case, which gives smaller color segments, the estimated planes may be unreliable, leading to an inaccurate disparity map. Popular segment based methods try to solve this problem by re-fitting the planes on grouped segments, in an iterative manner [2]. We propose a novel algorithm for generating sub-pixel accurate disparities on a per-pixel basis, thus alleviating the problems arising from methods that estimate disparities on a per-segment basis. The proposed method computes sub-pixel precision disparity maps using the recent minimum spanning tree (MST) [4] based cost aggregation framework. Since the disparity at every pixel is modeled by a plane equation, the goal is to ensure that all pixels belonging to a planar surface are labeled with the same plane equation. We show that using a reduced and refined set of planes as candidate labels in the aggregation framework ensures homogeneous labeling within a color segment.

Proposed Method Our method computes an initial set of plane equations (label set) by fitting planes inside a color segment using the consistent disparities from an initial disparity map. The initial disparity map may be generated using any local or global algorithm. These plane equations form the initial label set and a matching cost volume is computed over this set for every pixel. This cost volume is aggregated using MST based cost aggregation framework and a WTA over the aggregated cost volume gives the initial labeling. The number of labels in the initial set is of the order of the number of segments, with a plane estimate for every segment. The initial labeling is used along with the color segmentation to filter and generate a reduced set of planes. This framework of plane filtering followed by re-labeling leads to a more accurate disparity map. In addition, segment analysis is also used to modify the plane matching cost. We weigh the pixel matching cost by a support factor, where the support factor is derived from the distribution of plane labels within the color segment, as follows:

$$D(p, l) = \rho(p, q) e^{-\frac{n_{l,s}}{\tau n_s}} \quad (1)$$

where $\rho(p, q)$ computes the pixel dissimilarity between the pixels p and q , n_s is the number of pixels in the segment s that contains p , $n_{l,s}$ is the number of pixels in the segment s that are assigned plane label l , and τ is a constant. This cost update adds a bias towards locally dominant labels, whilst suppressing labels with smaller support. The labeling derived from modified cost volume with reduced set of labels is more locally homogeneous than previous labelings. The above matching cost modification is also used in occlusion filling step, which encourages labeling the occlusion region with a dominant plane label in the color segment the occlusion belongs to. The core algorithm block of plane labeling can be iterated on, in a feedback loop. The sub-pixel precision disparity map generated from the final plane labeling is used along with the initial color segments to re-estimate the set of planes. While a convergence criteria based on change in absolute disparities between iterations can be used, we have empirically found that convergence is reached in 3 iterations.

Results We report the experimental results using the proposed method on the Middlebury set [3] and also on natural scenes. We demonstrate the

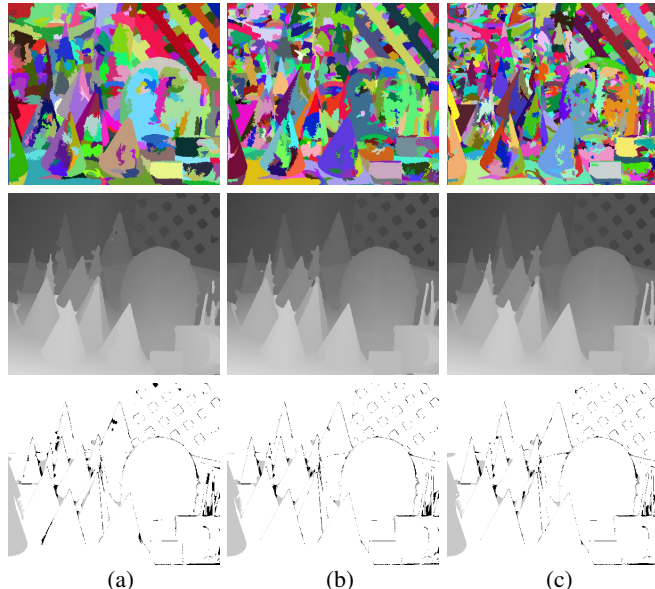


Figure 1: Effect of segmentation variance on disparity (Cones): (a) 266 segments, error = 2.58, rank = 23; (b) 507 segments, error = 2.10, rank = 3; (c) 836 segments, error = 2.16, rank = 6;

robustness of proposed method to the quality of the initial disparity map by considering two different methods for creating input fronto-parallel disparity maps. First, we initialize our method with a disparity map generated using simple WTA, without cost aggregation. The overall Middlebury rank [3] with this initialization is 21 after three iterations of our algorithm. Next, we initialize our method with the disparity generated by [4]. Three iterations of our algorithm using this initialization leads to an improvement in overall Middlebury rank from 43 to 11. Additionally, we report the lowest average percentage of bad pixels (3.58), of all methods in the Middlebury evaluation. The results indicate that our method adds a refinement step that is robust and can be added to any local or global algorithm generating fronto-parallel disparities. The recent method of Bleyer et al. [1] which also estimates a plane assignment per pixel takes 1 minute on an average to compute a disparity map on the Middlebury. The average run-time of our method on the Middlebury set is 25 seconds on a 2.67 GHz Intel Core i7 CPU with 8 GB memory.

Next, we demonstrate the robustness to segmentation parameter variation. The minimum segment size parameter in mean-shift segmentation is varied to generate varying segmentation maps. Our method is robust to these variations, resulting in accurate disparity maps in all instances as shown as shown in Fig. 1 of Middlebury Cones image. Observing the bottom right corner of the disparity maps in Fig. 1(a), 1(b), the pencils belong to a segment that spans multiple objects. Despite this leakage our algorithm is able to recover and assign correct disparities. The methods of [2] inherently generate labels on a per-segment basis, leading to a lower tolerance for such variations.

- [1] M. Bleyer, C. Rhemann, C. and C. Rother. Patchmatch stereo - stereo matching with slanted support windows. *BMVC*, pages 1–11, 2011.
- [2] A. Klaus, M. Sormann, and K. Karner. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. *ICPR*, pages 15–18, 2006.
- [3] D. Scharstein and R. Szeliski. Middlebury stereo evaluation. <http://vision.middlebury.edu/stereo/eval/>.
- [4] Q. Yang. A non-local cost aggregation method for stereo matching. *CVPR*, pages 1402–1409, 2012.