

# Multi-stage Branch-and-Bound for Maximum Variance Disparity Clustering

Ninad Thakoor, Venkat Devarajan  
Electrical Engineering Department  
The University of Texas at Arlington  
{ninad.thakoor,venkat}@uta.edu

Jean Gao  
Computer Science and Engineering Department  
The University of Texas at Arlington  
gao@uta.edu

## Abstract

*A split-and-merge framework based on a maximum variance criterion is proposed for disparity clustering. The proposed algorithm transforms low-level stereo disparity information to mid-level planar surface information which can be used further to carry out high-level computer vision tasks such as shape classification. Unlike conventional clustering, the proposed algorithm assumes that the number of clusters is unknown. Instead, a maximum variance criterion is applied to extract planar surfaces from the disparity image. The split phase of the algorithm creates clusters based on spatial continuity and the merge phase combines these clusters such that variance per cluster does not exceed an allowable value. For efficient maximum variance clustering, a greedy branch-and-bound procedure is introduced. Efficiency of the approach is verified through experiments.*

## 1. Introduction

The depth segmentation problem has been extensively dealt with in the terms of range image segmentation. However, due to different modality of depth estimates achieved through stereo, different treatment of the disparity segmentation problem becomes necessary. Disparity based spatial segmentation in which depth values drive spatial segmentation has been addressed by some researchers [3]. Disparity and motion information is also combined to carry out segmentation [1, 7]. However, very limited work has been done in segmentation of disparity alone [4, 5, 6].

Se and Brady [5] apply random sample consensus (RANSAC) to the disparity values to detect the ground plane in their stereo vision based algorithm. They assume that the ground plane is always visible and is the dominant plane in the image. The detected ground plane is then used to detect the small obstacles on it. Okada *et al.* [4] apply a randomized Hough transform to the Euclidean 3D data calculated from stereo. Peak points in the Hough space are selected as plane candidates. Finally, each image point is associated with the closest plane candidate. Trucco *et al.* [6] apply their method to disparity of weakly calibrated

stereo. For each image pixel, plane parameters are calculated by the least squares fitting on disparity values of the local neighborhood. Planes are found by carrying out clustering in the plane parameter space. However, for the point on the boundary of the planes, calculated parameters are inaccurate. This leads to inaccurate plane labeling in these regions.

This paper proposes an iterative split-and-merge algorithm for planar surface segmentation from disparity. An initial labeling of the disparity image based on the residuals is carried out. The initial labeling is then split to form spatially continuous regions. These regions are merged using the proposed multi-stage branch-and-bound merging. The proposed algorithm extracts a single planar surface at a time by maximizing the area of the surface under the constraint that the average residuals of the merged area are less than a fixed value. The merging process is repeated until all the regions are processed to form the planar surfaces. The number of the merging stages gives the number of the planar surfaces present in the image. With the updated number of planar surfaces and parameters, the split-and-merge process is repeated until convergence or until acceptable segmentation is attained.

The rest of the paper is organized as follows: the split-and-merge paradigm utilized is illustrated in Section 2. The proposed branch-and-bound algorithm is formulated in Section 3. Section 4 presents experimental results for a variety of stereo images. The paper concludes with suggestions for the future work in Section 5.

## 2. Methodology

A planar surface in the disparity space is given by [6]:

$$d = ax + by + c, \quad (1)$$

where  $\theta = (a, b, c)$  are the plane parameters in the disparity space and  $d$  is the disparity for the image location  $(x, y)$ . Thus, the planar surfaces can be segmented by carrying out clustering in the  $(a, b, c)$  space. This section elaborates the proposed method which follows a split-and-merge paradigm. Splitting is accomplished with spatial continuity and merging is carried out under a constraint of maximum allowable variance for a cluster.

## 2.1. Split

In the split step, it is assumed that the number of planar surfaces  $N$  and corresponding plane parameters  $\Theta = [\theta_1 \ \theta_2 \ \dots \ \theta_N]$  are known either due to the initialization or as a result of a multi-stage merge in the previous iteration. The estimate for the label of each pixel  $f_i$  can be computed as:

$$\hat{f}_i = \arg \min_f \{ \|d_i - (a_f x_i + b_f y_i + c_i)\|^2 \}. \quad (2)$$

The planar surfaces generated after above labeling are expected to be spatially continuous. However, the labels generated with (2) are not necessarily spatially continuous. The labeled image is split into  $N_s$  (which is greater than or equal to  $N$ ) regions based on the spatial continuity criterion.

## 2.2. Merge

A multi-stage merging strategy is proposed which can detect the number of planar surfaces automatically. At each stage, a planar surface is extracted by merging regions  $P_k$  under the constraint that the variance  $\eta^2$  is less than a constant value  $B$ . The variance  $\eta^2$  for the surface  $\mathcal{P}_j$ , extracted at stage  $j$ , can be calculated as,

$$\eta^2(\mathcal{P}_j) = \frac{1}{|\mathcal{P}_j|} \sum_{i \in \mathcal{P}_j} \|d_i - (a_j x_i + b_j y_i + c_j)\|^2 \leq B, \quad (3)$$

where  $(a_j, b_j, c_j)$  are computed from the merged region  $\mathcal{P}_j$  as the least squares estimate. However, this constraint alone does not yield a unique solution. While obeying this constraint the area of the extracted surface is also maximized.

Thus, we want to choose a subset  $\mathcal{P}_j$  of  $P = \{P_1, P_2, \dots, P_{N_s}\}$  at each stage  $j$  such that its area is maximized while its variance remains under a fixed value  $B$ . Once such a subset is determined, this optimal subset is extracted as  $\mathcal{P}_j^*$  and its members are removed from the set  $P$  to update  $P$  as  $P = P \setminus \mathcal{P}_j^*$ .  $N_s$  is also updated to  $N_s = N_s - |\mathcal{P}_j^*|$ . The merging process is repeated for the updated values of  $P$  and  $N_s$  until  $P$  is empty.

After a successful merging phase, updated number of planar surfaces  $N$  is available which is same as the number of surfaces separated by the merging phase. Corresponding planar surface parameters can be calculated as the least square estimate over  $\mathcal{P}_j^*$ ,  $j = 1, 2, \dots, N$ . If labeling results do not converge or are not acceptable, then the split-and-merge procedure can be repeated using the updated  $N$  and planar surface parameters. In the next section, a branch-and-bound algorithm is formulated to solve the merging problem efficiently.

## 3. Multi-stage branch-and-bound merging

The branch-and-bound approach to global optimization splits the optimization problem into smaller subproblems. Bounds on these subproblems are used to eliminate the ones which would not lead to an optimal solution. The subproblems that survive are divided further and the process is continued until all the subproblems are explored. The rest of

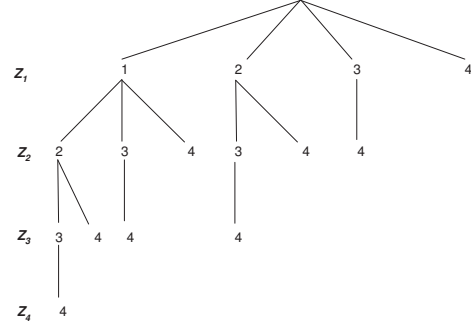


Figure 1. Solution tree for  $N_s=4$

this section constructs the branch-and-bound algorithm for the multi-stage merging.

Let  $\mathcal{P}_j = \{z_1, z_2, \dots, z_n\}$  denote the set of  $n$  elements which optimizes merging criterion at the stage  $j$ . For convenience, hereon the regions are indicated with their indices alone, i.e.,  $z_1, z_2, \dots, z_n \in \{1, 2, \dots, N_s\}$ . For each stage of merging, there are  $2^{N_s}$  possible solutions. These solutions can be searched efficiently with a branch-and-bound procedure [2]. Solutions for the problem can be represented as a rooted tree as shown in figure 1. Each node of the tree gives one possible solution for the problem. The regions included in the solution can be found by tracking from the root of the tree to the current node. If root of the tree represents an empty set, then each node encountered is added to the set of its parent node to generate the solution.

Before the branch-and-bound algorithm is formulated, we estimate the bounds on the best solution which can lead from the current node. Consider a solution at a node to be  $\mathcal{P}_j = \{z_1, z_2, \dots, z_n\}$ . The best solution in terms of area, i.e. the maximum area which this node can lead to, is given by sum of the area of the regions at the current node and the area of all the regions with index greater than  $z_n$ .

$$A_{\max}(\mathcal{P}_j) = A(\mathcal{P}_j) + A(z_n + 1, z_n + 2, \dots, N_s). \quad (4)$$

Clearly, if  $A_{\max}$  is smaller than the present optimal value for the area  $A_j^*$  then the current node cannot lead to a better solution and child nodes of this node can be safely abandoned.

On the other hand, the best solution in terms of the variance is the one which minimizes the average residuals. Although average residuals do not have property of monotonicity or linearity which can be used to derive a bound on its value, the sum of residuals (before normalization by area) is monotonic and can be used to construct a bound. A bound on the average residuals is given by the ratio of the minimum possible value of residuals of any child node and the maximum possible area of any child node,  $A_{\max}$ .

$$\eta_{\min}^2(\mathcal{P}_j) = \frac{|\mathcal{P}_j| \eta^2(\mathcal{P}_j) + \min\{r_{z_n+1}^2, r_{z_n+2}^2, \dots, r_{N_s}^2\}}{A_{\max}}. \quad (5)$$

Here,  $r_k^2$  indicates the residuals after the least square fit for the region  $P_k$ . If  $\eta_{\min}^2$  is greater than  $B$  then the current parent node will not lead to an optimal solution and search for the optimal solution can be terminated along this node.

An additional bound can also be derived when at least one stage of the merging has finished, i.e., when  $j > 1$ . As  $\mathcal{P}_j^*$  is extracted by maximizing the area under the same constraint as the one used for  $j - 1$ , the area of the optimal solution in any stage  $A_j^*$  must be less than or equal to the optimal area achieved in the previous stage  $A_{j-1}^*$ .

$$A_{\text{upper}} = A_{j-1}^*. \quad (6)$$

A multi-stage branch-and-bound merging algorithm based on the above bounds is listed below.

1. Overall initialization: Set stage  $j = 1$ , number of planar surfaces  $N = 1$ , and upper bound on area  $A_{\text{upper}} = \infty$ .
2. Stage initialization: Set spatially continuous regions  $N_s = |P|$ , optimal area  $A_j^* = 0$ , optimal merging  $\mathcal{P}_j^* = \emptyset$ , the tree level  $i = 1$  and current node  $z_0 = 0$ .
3. Generate Successors: Initialize  $LIST(i)$ ,

$$LIST(i) = \{z_{i-1} + 1, z_{i-1} + 2, \dots, N_s\}.$$

4. Select new node: If  $LIST(i)$  is empty, go to step 6. Otherwise, set  $z_i = k$  where  $k \in LIST(i)$ . Set current solution  $\mathcal{P}_j = \{z_1, z_2, \dots, z_i\}$ . Delete  $k$  from  $LIST(i)$ .
5. Check bounds: Compute  $A_{\max}(\mathcal{P}_j)$ . If  $A_{\max}(\mathcal{P}_j) < A_j^*$ , go to step 6. Calculate  $\eta_{\min}^2(\mathcal{P}_j)$ . If  $\eta_{\min}^2(\mathcal{P}_j) > B$ , go to step 6. Compute  $A(\mathcal{P}_j)$ . If  $A_{\text{upper}} < A(\mathcal{P}_j)$ , go to step 6. Compute  $\eta^2(\mathcal{P}_j)$ . If  $A^* \leq A(\mathcal{P}_j)$  and  $\eta^2(\mathcal{P}_j) \leq B$ , set  $A^* = A(\mathcal{P}_j)$  and  $\mathcal{P}_j^* = \mathcal{P}_j$ . Set  $i = i + 1$  and go to step 3.
6. Backtrack to lower level: Set  $i = i - 1$ . If  $i > 0$ , go to step 4. If  $A_j^* = 0$ , set  $\mathcal{P}_j^* = \{1\}$ . Set  $A_{\text{upper}} = A_j^*$ . Update  $P = P \setminus \mathcal{P}_j^*$ . If  $P = \emptyset$ , terminate the algorithm. Set  $j = j + 1$ ,  $N = N + 1$  and go to step 2.

## 4. Experimental results

The proposed planar surface segmentation process was tested with a variety of synthetic and real data.

The first set of experiments was carried out with stereo image pairs and their ground truth disparity, available at Middlebury stereo vision research page (<http://www.middlebury.edu/stereo>). For each image, the number of planar surfaces is known and this information can be used to verify the success of the proposed method.

Figure 2(a) shows the left image of the ‘‘barn’’ image sequence. Each image in this sequence is  $432 \times 381$  pixels.

**Table 1. Solutions explored in the first stage for calculated ‘‘Barn’’ disparity**

Iteration	$N$	$N_s$	Solutions explored	Fraction explored
1	7	40	1462	1.33e-9
2	6	40	4092	3.72e-9
3	6	23	224	2.67e-5
4	6	21	146	6.96e-5

The image shows six planar surfaces of various size, shape and orientation. First the ground truth disparity (shown in Figure 2(b)) was segmented with bound  $B = 0.1$ . The number of surfaces was initialized to be  $N_{\text{init}} = 10$ . To initialize the planar surface parameters, the cumulative histogram of the disparity image was split into  $N_{\text{init}}$  equal segments. For each segment, planar parameters were estimated which were used for the spatial continuity based splitting in the first iteration of the algorithm. Planar surfaces extracted after four iterations of split-and-merge are shown in Figure 2(c).

Figure 2(d) shows the estimated disparity for the sequence. The small dark areas around the edges of the planes are the occluded areas for which a reliable estimate of disparity is not available. Detected planar surfaces are shown in Figure 2(e) after some postprocessing. The postprocessing involved size filtering on the detected regions. Any regions of size less than 50 pixels were filtered out. These regions along with the occluded regions were then labeled by nearest neighbor interpolation. The foreground planes are bigger compared to the ones in the segmented ground truth. This is due to the use of a square window to calculate the disparities. Additionally note that disparities at extremities of the image cannot be estimated. For this reason, calculated disparity images are smaller than the ground truth disparity images.

Table 1 shows the number of solutions explored to extract the largest planar surface from the calculated disparity. The table demonstrates the effectiveness of the proposed branch-and-bound algorithm which explores a very small fraction of the solution space to reach the optimal solution. Before the multi-stage merging was carried out, all the spatially continuous regions were sorted according to decreasing area. To speed up the solution, any regions which were smaller than 50 pixels were removed from the search. If the number of regions left in the search was greater than 40 after this, then only 40 largest areas were considered for the merging operation to further speed up the process. Note that, even with 40 areas the number of possible solutions are  $2^{40} \approx 1.099\text{e}12$ . The tables also show that the number of planar surfaces is correctly identified as six in both cases.

Images chosen in the first experiment contained only pla-

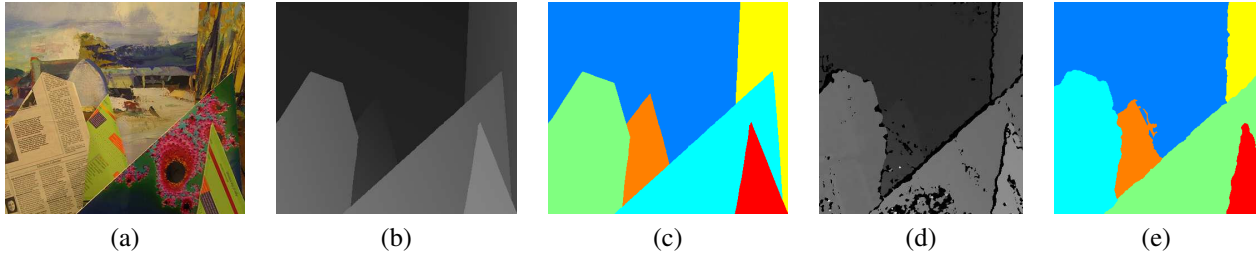


Figure 2. (a) Left image of pair “Barn”, (b) Ground truth disparity, (c) Detected planar surfaces ( $B = 0.1$ , iterations=4), (d) Calculated subpixel disparity, (e) Detected planar surfaces ( $B = 0.05$ , iterations=4)

nar objects and scene was “artificial”. The second set of experiments was carried out with some of the JISCT stereo images (<http://vasc.ri.cmu.edu/idb/html/jisct/>) which contain a few real life sequences with objects which are not necessarily planar. The sequence we tested is called “shrub” and its left image is shown in Figure 3 (a). Dimensions of the image sequence are  $512 \times 480$ . The images contain a hedge in front of a building wall along with a parking sign and part of a road at the bottom.

Similar to the first experiment,  $N_{init}$  was chosen to be 10. The initial segmentation is shown in Figure 3(c). After four iterations of the split-and-merge with bound  $B = 0.3$ , three planar surfaces are detected. The first surface corresponds to the wall, the second corresponds to the hedge and the last represents the part of the road. Due to the increased complexity of the scene compared to the first experiment, the maximum number of solutions is explored in the first stage in first iteration, i.e., 212644 for  $N_s = 40$  (fraction  $1.93e-7$ ). This large number is result of numerous small regions and very few large regions present. As more regions are needed to form the optimal solution, the optimal solution is farther from the root, which leads to a larger number of solutions being explored.

## 5. Conclusion and future work

In this paper, we proposed an iterative split-and-merge approach for segmentation of planar surfaces in the disparity space. Spatial continuity based splitting and variance based merging was carried out iteratively to detect the number of planar surfaces and corresponding parameters automatically. An efficient multi-stage merging algorithm based on branch-and-bound was also proposed. The effectiveness of the proposed scheme and the branch-and-bound algorithm was demonstrated with experimental results for different data sets.

While impressive speedups are achieved with the proposed branch-and-bound algorithm, the algorithm can be improved with additional heuristics. Also, combination of spatial segmentation with the proposed multi-stage merging algorithm might eliminate need for the iterative process.

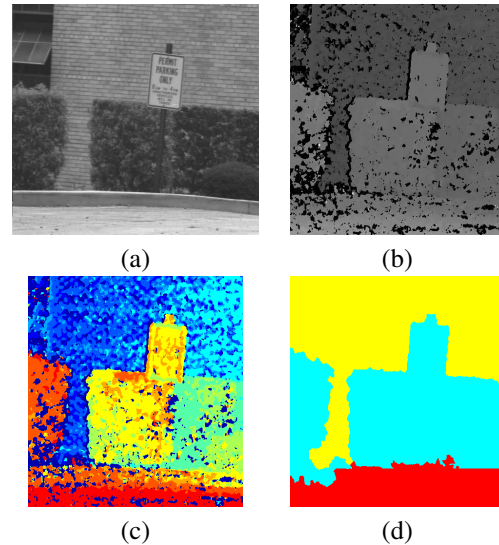


Figure 3. (a) Left image of pair “Shrub”, (b) Calculated disparity, (c) Initial segmentation ( $N_{init} = 10$ ), (d) Detected planar surfaces ( $B = 0.3$ , iterations=4)

## References

- [1] Y. Altunbasak, A. Tekalp, and G. Bozdagi. Simultaneous motion-disparity estimation and segmentation from stereo. In *IEEE ICIP*, volume 3, pages 73–77 vol.3, 1994.
- [2] M. Brusco and S. Stahl. *Branch-and-Bound Applications in Combinatorial Data Analysis*. Springer, 2005.
- [3] E. Izquierdo. Disparity/segmentation analysis: matching with an adaptive window and depth-driven segmentation. *IEEE Trans. CSVT*, 9(4):589–607, 1999.
- [4] K. Okada, S. Kagami, M. Inaba, and H. Inoue. Plane segment finder: algorithm, implementation and applications. In *IEEE ICRA*, volume 2, pages 2120–2125 vol.2, 2001.
- [5] S. Se and M. Brady. Stereo vision-based obstacle detection for partially sighted people. In *ACCV*, pages 152–159, London, UK, 1997. Springer-Verlag.
- [6] E. Trucco, F. Isgro, and F. Bracchi. Plane detection in disparity space. In *International Conference on Visual Information Engineering*, pages 73–76, 2003.
- [7] D. Tzovaras, N. Grammalidis, and M. G. Strintzis. Joint three-dimensional motion/disparity segmentation for object-based stereo image sequence coding. *Optical Engineering*, 35:137–144, Jan. 1996.