Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure

Andreas Klaus, Mario Sormann and Konrad Karner
VRVis Research Center
8010 Graz, Austria
{klaus,sormann,karner}@vrvis.at

Abstract

A novel stereo matching algorithm is proposed that utilizes color segmentation on the reference image and a self-adapting matching score that maximizes the number of reliable correspondences. The scene structure is modeled by a set of planar surface patches which are estimated using a new technique that is more robust to outliers. Instead of assigning a disparity value to each pixel, a disparity plane is assigned to each segment. The optimal disparity plane labeling is approximated by applying belief propagation. Experimental results using the Middlebury stereo test bed demonstrate the superior performance of the proposed method.

1. Introduction

Stereo matching continues to be an active research area as is proven by a large number of recent publications dedicated to this topic [1, 2, 4, 6, 9, 12]. The goal is to determine disparities that are indicating the difference in locating corresponding pixels. The recovery of an accurate disparity map still remains challenging, mainly due to the following reasons:
(i) Pixels of half occluded regions do not have correspondences in the other image, leading to incorrect matches if not taken into account.
(ii) Images are disturbed because of sensor noise. This is especially problematic in poorly textured regions due to the low signal-to-noise-ratio (SNR).
(iii) The constant brightness or color constraint is only satisfied under ideal conditions that can only roughly be met in practice.
A comprehensive overview on stereo matching can be found in [8]. In general matching algorithms can be classified into local and global methods. Local approaches are utilizing the color or intensity values within a finite window to determine the disparity for each pixel. Global approaches are incorporating explicit smoothness assumptions and are determining all disparities simultaneously by applying energy minimization techniques such as graph cuts [2, 4, 6, 7], belief propagation [5, 9, 10, 12], dynamic programming, scanline optimization or simulated annealing.
Recently, segment-based methods [1, 2, 4, 6, 11] have attracted attention due to their good performance. They are based on the assumption that the scene structure can be approximated by a set of non-overlapping planes in the disparity space and that each plane is coincident with at least one homogeneous color segment in the reference image. Segment-based methods generally perform four consecutive steps that are illustrated in Figure 1. First, regions of homogeneous color are located by applying a color segmentation method. Second, a local window-based matching method is used to determine disparities of reliable points. Third, a plane fitting technique is applied to obtain disparity planes that are considered as a label set. Fourth, an optimal disparity plane assignment (optimal labeling) is approximated using greedy [1, 11] or graph cuts [2, 4, 6] optimization. Despite our method shares the same consecutive steps, there are three main distinguishing features:
First, a self-adapting dissimilarity measure is used to increase the number of reliable correspondences as is explained is Section 3. Second, a novel outlier insensitive approach is applied to extract the disparity planes (Section 4). Third, the labeling problem is solved using belief propagation (Section 5). These features represent the main contribution of our approach which results in superior matching quality (demonstrated in Section 6).

2. Color segmentation

The first step in the workflow is to decompose the reference image into regions of homogeneous color or grayscale. The algorithm assumes that disparity values vary smoothly in those regions and that depth discontinuities only occur on

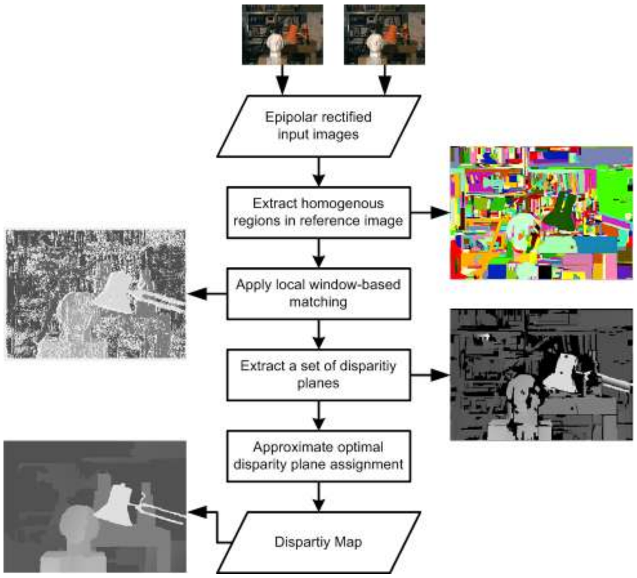


Figure 1. Block diagram of segment-based stereo matching algorithms augmented with input data, intermediate and final results of the proposed method.

region boundaries. Over-segmentation is preferred, since it helps to meet this assumption in practice. Therefore mean-shift color segmentation recently successfully applied to image segmentation by Comaniciu and Meer [3] is used. The mean-shift analysis approach is essentially defined as a gradient ascent search for maxima in a density function defined over a high dimensional feature space. The feature space include a combination of the spatial coordinates and all its associated attributes that are considered during the analysis. The main advantage of the mean-shift approach is based on the fact that edge information is incorporated as well.

3. Local matching in pixel domain

In the proposed method the scene structure is modeled by a set of planar disparity planes. A disparity plane is specified by the three parameters c_1, c_2, c_3 that determine a disparity d for each reference image pixel (x,y): d=$c_1 x + c_2 y + c_3$

Due to the huge number of possible disparity planes the number is reduced by extracting a set of disparity planes that is sufficient to represent the scene structure. This is done by applying local matching in the pixel domain followed by a disparity plane estimation step.

Local matching requires to define a matching score and an aggregation window [8]. The most common dissimilarity measures are squared intensity differences (SD) and absolute intensity differences (AD) that are strictly assuming the constant color constraint. Other matching scores such as gradient-based and non-parametric measures are more robust to changes in camera gain and bias or non-lambertian surfaces at the cost of a low discriminating power. In our approach we are using a self-adapting dissimilarity measure that combines sum of absolute intensity differences (SAD) and a gradient based measure that are defined as follows:

$$C_{SAD}(x,y,d) = \sum_{(i,j)\in N(x,y)} I_1(i,j) - I_2(i+d,j)$$

and

$$C_{GRAD}(x,y,d) = \sum_{(i,j)\in N_x(x,y)} |\nabla_x I_1(i,j) - \nabla_x I_2(i+d,j)| + \sum_{(i,j)\in N_y(x,y)} |\nabla_y I_1(i,j) - \nabla_y I_2(i+d,j)|,$$

where $N(x,y)$ is a 3×3 surrounding window at position (x,y), $N_x(x,y)$ a surrounding window without the rightmost column, $N_y(x,y)$ a surrounding window without the lowest row, ∇_x the forward gradient to the right and ∇_y the forward gradient to the bottom. Color images are taken into account by summing up the dissimilarity measures for all channels.

An optimal weighting ω between C_{SAD} and C_{GRAD} is determined by maximizing the number of reliable correspondences that are filtered out by applying a cross-checking test (comparing left-to-right and right-to-left disparity maps) in conjunction with a winner-take-all optimization (choosing the disparity with the lowest matching cost). The resulting dissimilarity measure is given by:

$$C(x,y,d) = (1-\omega) * C_{SAD}(x,y,d) + \omega * C_{GRAD}(x,y,d)$$

Furthermore we are utilizing the reliable correspondences to predict the SNR that is used to normalize our dissimilarity measure. Because of the normalization a fixed truncation threshold can be set right above the noise level to obtain a robust matching score.

4. Disparity plane estimation

The reliable correspondences are used to derive a set of disparity planes that are adequate to represent the scene structure. This is achieved by applying a novel robust plane fitting method and a consecutive refinement step.

Robust plane fitting Despite only reliable disparities of each segment are used to derive a corresponding disparity plane, the estimated plane may be disturbed due to remaining outliers. A straightforward way to determine the disparity plane parameters is to solve a least square system. As is generally known least square solutions are very sensitive to outliers and that linear or median solutions are much more robust.

Our method determines a robust solution by applying a decomposition method to solve each parameter separately. First, the horizontal slant is estimated using a set of all combinations of reliable disparities that are lying in the same

horizontal line within the segment. The derivations $\delta d/\delta x$ are inserted to a list and a robust estimation of the horizontal slant is determined by sorting the list and applying convolution with a Gaussian kernel.
Second, the vertical slant is estimated in a similar manner by considering all combinations lying on the same vertical line.
Third, the determined slant is used to obtain a robust estimate of the disparity value in the center of the segment. Therefore corresponding center disparities for each reliable point, that are calculated by considering the estimated slant, are inserted to a list and a robust estimate is obtained as explained before.

Disparity plane refinement The purpose of this step is to increase the accuracy of the disparity plane set by repeating the plane fitting for grouped regions that are dedicated to the same disparity plane. Similar as in [6] the following steps are processed:
First, a matching cost is calculated for each segment-to-plane assignment. It is computed by summing up the matching cost for each pixel inside the segment S:

$$C_{SEG}(S,P) = \sum_{(x,y)\in S} C(x,y,d),$$

where P is a disparity plane that defines disparity d.
Second, the disparity plane with the minimum matching cost is assigned to each segment. Third, segments that are assigned to the disparity plane are grouped. Finally the plane estimation is repeated for all grouped segments.

5. Disparity plane assignment

In the final step an optimal solution for the segment-to-disparity plane assignment is searched. Therefore the stereo matching is formulated as an energy minimization problem for the labeling f that assigns each segment $s \in R$ a corresponding plane $f(s) \in D$. The energy for a labeling f is given by:

$$E(f) = E_{data}(f) + E_{smooth}(f),$$

where

$$E_{data}(f) = \sum_{s\in R} C_{SEG}(s, f(s))$$

and

$$E_{smooth}(f) = \sum_{(\forall(s_i,s_j)\in S_N \mid f(s_i)\neq f(s_j))} \lambda_{disc}(s_i,s_j).$$

S_N represents a set of all adjacent segments and $\lambda_{disc}(s_i,s_j)$ is a discontinuity penalty that incorporates the common border lengths and the mean color similarity as proposed in [2].
An optimal labeling with minimum energy is approximated using Loopy Belief Propagation [5] where the message passing takes place between adjacent segments.

6. Experimental results

The proposed method was evaluated using the Middlebury test bed (http://cat.middlebury.edu/stereo/) provided by the authors of [8]. Qualitative results are shown in Figure 2. Quantitative results of the ten best performing methods are given in Table 1, where the percentage of pixels with an absolute disparity error greater than one pixel are shown for different regions: non-occluded (nonoccl.), whole image (all) and pixels near discontinuities (on disc.). Our method processed all four stereo pairs with a fixed parameter set and was ranked at the first place. The calculation on a 2.21GHz Athlon 64 computer takes between 14 and 25 sec, whereas the mean-shift segmentation is the most time consuming step.

7. Conclusions

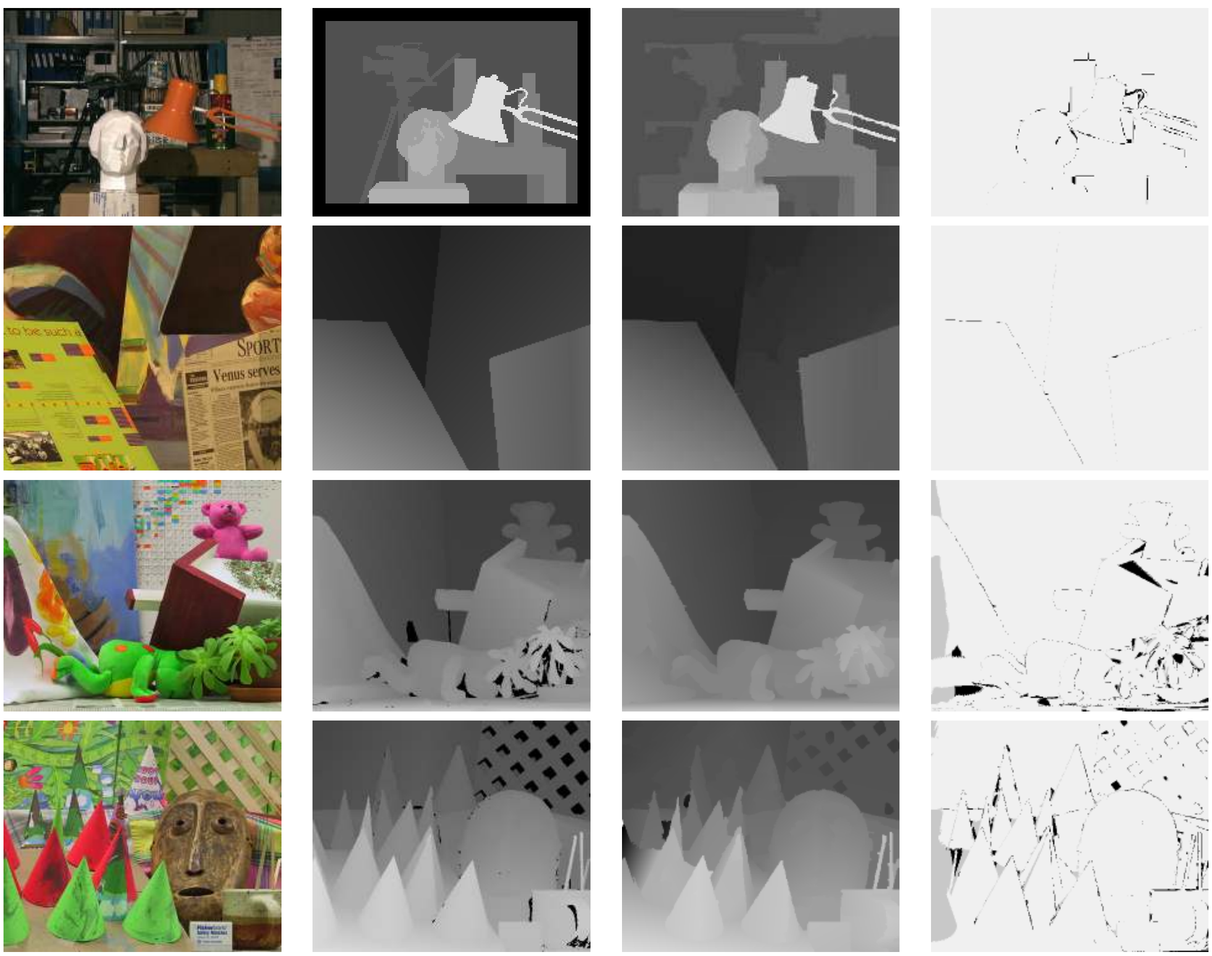
A new segment-based stereo matcher has been introduced. The conjunction of color segmentation, a self-adapting matching score, a robust plane fitting technique as well as BP-optimization yields excellent results as demonstrated on the Middlebury stereo evaluation test bed.

Acknowledgments

This work has been done in the VRVis research center, Graz and Vienna/Austria (http://www.vrvis.at), which is partly funded by the Austrian government research program Kplus.
Many thanks to the authors of [8] for providing the test images and the unique evaluation test bed.

Algorithm	*Avg.* Rank	*Tsukuba*			*Venus*			*Teddy*			*Cones*		
		nonoccl.	all	on disc.	nonoccl.	all	on disc.	nonoccl.	all	on disc.	nonoccl.	all	on disc.
Proposed Method	1.7	1.11_3	1.37_2	5.79_3	0.10_1	0.21_1	1.44_1	4.22_2	7.06_2	11.8_2	2.48_1	7.92_1	7.32_1
Double-BP [12]	2.3	0.88_1	1.29_1	4.76_1	0.14_2	0.60_5	2.00_3	3.55_1	8.71_3	9.70_1	2.90_3	9.24_4	7.80_2
Segm+visib [1]	5.1	1.30_6	1.57_3	6.92_8	0.79_8	1.06_6	6.76_9	5.00_3	6.54_1	12.3_3	3.72_5	8.62_3	10.2_6
SymBP+occ [9]	5.1	0.97_2	1.75_5	5.09_2	0.16_3	0.33_2	2.19_4	6.47_6	10.7_4	17.0_7	4.79_{11}	10.7_8	10.9_7
C-SemiGlob	6.2	2.61_{16}	3.29_{11}	9.89_{14}	0.25_5	0.57_3	3.24_5	5.14_4	11.8_5	13.0_4	2.77_2	8.35_2	8.20_3
RegionTreeDP	7.0	1.39_9	1.64_4	6.85_6	0.22_4	0.57_3	1.93_2	7.42_9	11.9_6	16.8_6	6.31_{14}	11.9_{12}	11.8_9
AdaptWeight	7.3	1.38_8	1.85_6	6.90_7	0.71_6	1.19_7	6.13_7	7.88_{10}	13.3_9	18.6_{11}	3.97_7	9.79_6	8.26_4
SemiGlob	9.3	3.26_{17}	3.96_{14}	12.8_{19}	1.00_9	1.57_8	11.3_{14}	6.02_5	12.2_7	16.3_5	3.06_4	9.75_5	8.90_5
RealtimeBP	10.4	1.49_{10}	3.40_{13}	7.87_{10}	0.77_7	1.90_{11}	9.00_{13}	8.72_{13}	13.2_8	17.2_8	4.61_9	11.6_{10}	12.4_{13}
Layered	11.4	1.57_{11}	1.87_7	8.28_{11}	1.34_{11}	1.85_9	6.85_{10}	8.64_{12}	14.3_{10}	18.5_{10}	6.59_{16}	14.7_{15}	14.4_{15}
GC+occ	11.5	1.19_4	2.01_9	6.24_4	1.64_{14}	2.19_{13}	6.75_8	11.2_{16}	17.4_{16}	19.8_{14}	5.36_{13}	12.4_{13}	13.0_{14}
MultiCamGC	12.0	1.27_5	1.99_8	6.48_5	2.79_{19}	3.13_{17}	3.60_6	17.0_{17}	17.6_{17}	22.0_{16}	4.89_{12}	11.8_{11}	12.1_{11}

Table 1. Middlebury stereo evaluation on different algorithms, ordered according to their overall performance. The subscript numbers indicate the rank of each method in each column.



reference images | ground truths | our results | 'bad pixel' maps

Figure 2. Results using the Middlebury datasets: Tsukuba, Venus, Teddy and Cones. Pixels with a disparity error greater than one pixel are displayed in the 'bad pixel' maps, where missmatches in non-occluded areas are indicated in black, in occluded areas in gray color.

References

[1] M. Bleyer and M. Gelautz. A layered stereo matching algorithm using image segmentation and global visibility constraints. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(3):128–150, May 2005.

[2] M. Bleyer and M. Gelautz. Graph-based surface reconstruction from stereo pairs using image segmentation. In *SPIE*, pages vol. 5665: 288–299, January 2005.

[3] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE:PAMI*, 24(5):603–619, May 2002.

[4] Y. Deng, Q. Yang, X. Lin, and X. Tang. A symmetric patch-based correspondence model for occlusion handling. In *ICCV*, pages II: 1316–1322, 2005.

[5] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. In *CVPR*, pages I: 261–268, 2004.

[6] L. Hong and G. Chen. Segment-based stereo matching using graph cuts. In *CVPR*, pages I: 74–81, 2004.

[7] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions via graph cuts. In *ICCV*, pages II: 508–515, 2001.

[8] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42, Apr. 2002.

[9] J. Sun, Y. Li, S. Kang, and H. Shum. Symmetric stereo matching for occlusion handling. In *CVPR*, pages II: 399–406, 2005.

[10] J. Sun, N. Zheng, and H. Shum. Stereo matching using belief propagation. In *ECCV*, page II: 510 ff., 2002.

[11] H. Tao, H. S. Sawhney, and R. Kumar. A global matching framework for stereo computation. In *ICCV*, pages I: 532–539, 2001.

[12] Q. Yáng, L. Wang, R. Yang, H. Stewénius, and D. Nistér. Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling. Accepted to CVPR 2006.