Stereo Matching with Color-Weighted Correlation, Hierarchical Belief Propagation and Occlusion Handling

Qìngxióng Yáng Liang Wang Ruigang Yang Henrik Stewénius David Nistér
Center for Visualization and Virtual Environments
Department of Computer Science, University of Kentucky
http://www.vis.uky.edu/~liiton/

Abstract

In this paper, we formulate an algorithm for the stereo matching problem with careful handling of disparity, discontinuity and occlusion. The algorithm works with a global matching stereo model based on an energy-minimization framework. The global energy contains two terms, the data term and the smoothness term. The data term is first approximated by a color-weighted correlation, then refined in occluded and low-texture areas in a repeated application of a hierarchical loopy belief propagation algorithm. The experimental results are evaluated on the Middlebury data set, showing that our algorithm is the top performer.

1. Introduction

Stereo is one of the most extensively researched topics in computer vision. Stereo research has recently experienced somewhat of a new era, as a result of publically available performance testing such as the Middlebury data set [11], which has allowed researchers to compare their algorithms against all the state-of-the-art algorithms.

In this paper, we describe our stereo algorithm, which is currently evaluating as the top performer on the Middlebury data set. The algorithm springs from the popular energy minimization framework that is the basis for most of the algorithms on the Middlebury top-list, such as graph cuts [4, 10] and belief propagation [12, 13]. In this framework, there is typically a data term and a smoothness term, where the data term consists of the matching error implied by the extracted disparity map, and the smoothness term encodes the prior assumption that world surfaces are piecewise smooth.

However, the algorithm presented in this paper departs somewhat from the normal framework, in that in the final stages of the algorithm, the data term is updated based on the current understanding of which pixels in the reference image are occluded or unstable due to low texture.

The paper is organized as follows: Section 2 gives a high-level overview of the approach. In Section 3 we then give the detailed equations for all the building blocks. Section 4 reports results supporting the claims that the algorithm is currently the strongest available on the Middlebury data set. Section 5 concludes.

2. Overview of the Approach

The algorithm can be partitioned into three blocks, initial stereo (Figure 1), pixel classification (Figure 2) and iterative refinement (Figure 3). In the initial stereo, see Figure 1, the correlation volume is first computed. A basic way to construct the correlation volume is to compute the absolute difference of luminances of the corresponding pixels in the left and right images, but there are many other methods for correlation volume construction. For instance, Sun et al. [12] use Birchfield and Tomasi's pixel dissimilarity [1] to construct the correlation volume, and Felzenszwalb [6] suggests to smooth the image first before calculating the pixel difference. In this work, we are using color-weighted correlation to build the correlation volume, in a similar manner as was recently described by Yoon and Kweon [17]. The color-weighting makes the match scores less sensitive to occlusion boundaries by using the fact that occlusion boundaries most often cause color discontinuities as well. The initial stereo is run in turn with both the left and the right image as reference images. This is done just to support a subsequent mutual consistency check (often called left-right check) that takes place in the pixel classification block. Functions E_S^L and E_S^R defining the smoothness costs in the left and right reference images, respectively, are determined based on the color gradients in the input images. The left and right smoothness costs and the left and right correlation costs are then optimized using two separate hierarchical belief propagation processes. The hierarchical belief propagation is performed in a manner similar to Felzenszwalb [6], resulting in the initial left and

Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)
0-7695-2597-0/06 $20.00 © 2006 IEEE

right disparity maps $D_L^{(0)}$ and D_R, respectively. The left disparity map is given an iteration index $i = 0$ here, because it will be repeatedly refined in the iterative refinement module. The outputs needed from the initial stereo are the initial left and right disparity maps $D_L^{(0)}$ and D_R, the left correlation volume C_L, the left image I_L and smoothness cost E_S^L.

In the pixel classification module, see Figure 2, pixels are given one out of three possible labels: occluded, stable or unstable. The occluded pixels are the ones that fail the mutual consistency check that is performed using the left and right disparity maps. The pixels that pass the mutual consistency check are then labeled stable or unstable based on a confidence measure derived from the left correlation volume, which measures if the peak in the correlation score is distinct enough that the local disparity can be considered stable. The output from the pixel classification module is simply the pixel class membership.

In the iterative refinement module, see Figure 3, the left smoothness cost E_S^L, initial left disparity map $D_L^{(0)}$, left image I_L, pixel class membership and left correlation volume C_L are all used as input. The goal here is to propagate information from the stable pixels to the unstable and the occluded pixels. This is done using color segmentation and plane fitting in a way inspired by [14]. In our work, we use color segments extracted by mean shift [5] applied to the left input image. In each color segment, the disparity values for the stable pixels are used in a plane fitting procedure. Note that the disparity values used here are taken from the current hypothesis $D_L^{(i)}$ for the left disparity map. This disparity map is first initialized with the left disparity map $D_L^{(0)}$ given by the initial stereo module. The result of plane fitting within color segments is then used together with the pixel class membership and the left correlation volume to give the current data term hypothesis $E_D^{(i+1)}$, which is used with the original smoothness cost E_S^L in hierarchical belief propagation. Effectively, the plane fitted depth map is used as a regularization for the new disparity estimation. The hierarchical belief propagation yields the updated disparity map hypothesis $D_L^{(i+1)}$, which is iteratively fed back into the plane fitting.

3. Detailed Description

In this section, we give a more detailed description of the building blocks outlined above. The order of description follows the above outline through Figures 1,2 and 3.

3.1. Initial Stereo

The main building blocks of the initial stereo module, see Figure 1, are color-weighted correlation, smoothness cost definition and hierarchical belief propagation.

The objective of the color-weighted cost aggregation is to initialize a reliable correlation volume. To obtain more accurate results on both smooth and discontinuous regions, an appropriate window should be selected adaptively for each pixel during the cost aggregation step. That is, the window should be large enough to cover sufficient area in textureless regions, while small enough to avoid crossing depth discontinuities. Many methods [9, 3, 15, 16, 8, 2] have been proposed to solve this ambiguity problem.

In our implementation, we use an amended version of the color based weight approach proposed recently by Yoon and Kweon [17]. In this method, instead of finding an optimal support window, adaptive support-weights are assigned to pixels in some large window with side-length α_{cw} based both on the color proximity and the spatial proximity to the pixel under consideration (the central pixel of the support window).

In Yoon and Kweon's work, the similarity between two pixels within the support window is measured in the CIELab color space. Our approach however simply measures it in the RGB color space. Due to our post-refinement process, this change does not prevent us from achieving state-of-the-art results. However, instead of using a raw pixel difference, we use Birchfield and Tomasi's pixel dissimilarity [1].

The color difference $\Delta_{\mathbf{xy}}$ between pixel $\mathbf{x}$ and $\mathbf{y}$ (in the same image) is expressed as

$$\Delta_{\mathbf{xy}} = \sum_{c\in\{r,g,b\}} |I_c(\mathbf{x}) - I_c(\mathbf{y})|, \tag{1}$$

where I_c is the intensity of the color channel c. The weight of pixel $\mathbf{x}$ in the support window of $\mathbf{y}$ (or vice versa) is then determined using both the color and spatial differences as

$$w_{\mathbf{xy}} = e^{-(\beta_{cw}^{-1}\Delta_{\mathbf{xy}} + \gamma_{cw}^{-1}\|\mathbf{x}-\mathbf{y}\|_2)}, \tag{2}$$

where β_{cw} and γ_{cw} are parameters determined empirically.

The data term is then an aggregation with the soft windows defined by the weights, as

$$C(\mathbf{x_L},\mathbf{x_R}) = \frac{\sum_{(\mathbf{y}_L,\mathbf{y}_R)\in W_{\mathbf{x}_L}\times W_{\mathbf{x}_R}} w_{\mathbf{x}_L\mathbf{y}_L} w_{\mathbf{x}_R\mathbf{y}_R} d(\mathbf{y}_L,\mathbf{y}_R)}{\sum_{(\mathbf{y}_L,\mathbf{y}_R)\in W_{\mathbf{x}_L}\times W_{\mathbf{x}_R}} w_{\mathbf{x}_L\mathbf{y}_L} w_{\mathbf{x}_R\mathbf{y}_R}},$$

where $W_{\mathbf{x}}$ is the support window around $\mathbf{x}$ and $d(\mathbf{y}_L, \mathbf{y}_R)$ represents Birchfield and Tomasi's pixel dissimilarity, $\mathbf{x}_L$ and $\mathbf{y}_L$ are pixels in the left image $\mathbf{I}_L$, $\mathbf{x}_R$ and $\mathbf{y}_R$ are pixels in the right image $\mathbf{I}_R$.

The smoothness cost should be decreased at depth edges, since these are likely to coincide with color edges, the luminance difference

$$\delta_{\mathbf{xy}} = |I(\mathbf{x}) - I(\mathbf{y})| \tag{3}$$

between neighboring pixels $\mathbf{x}$ and $\mathbf{y}$ is used to decrease the cost. The difference $\delta_{\mathbf{xy}}$ is normalized to span the interval

Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)
0-7695-2597-0/06 $20.00 © 2006 IEEE

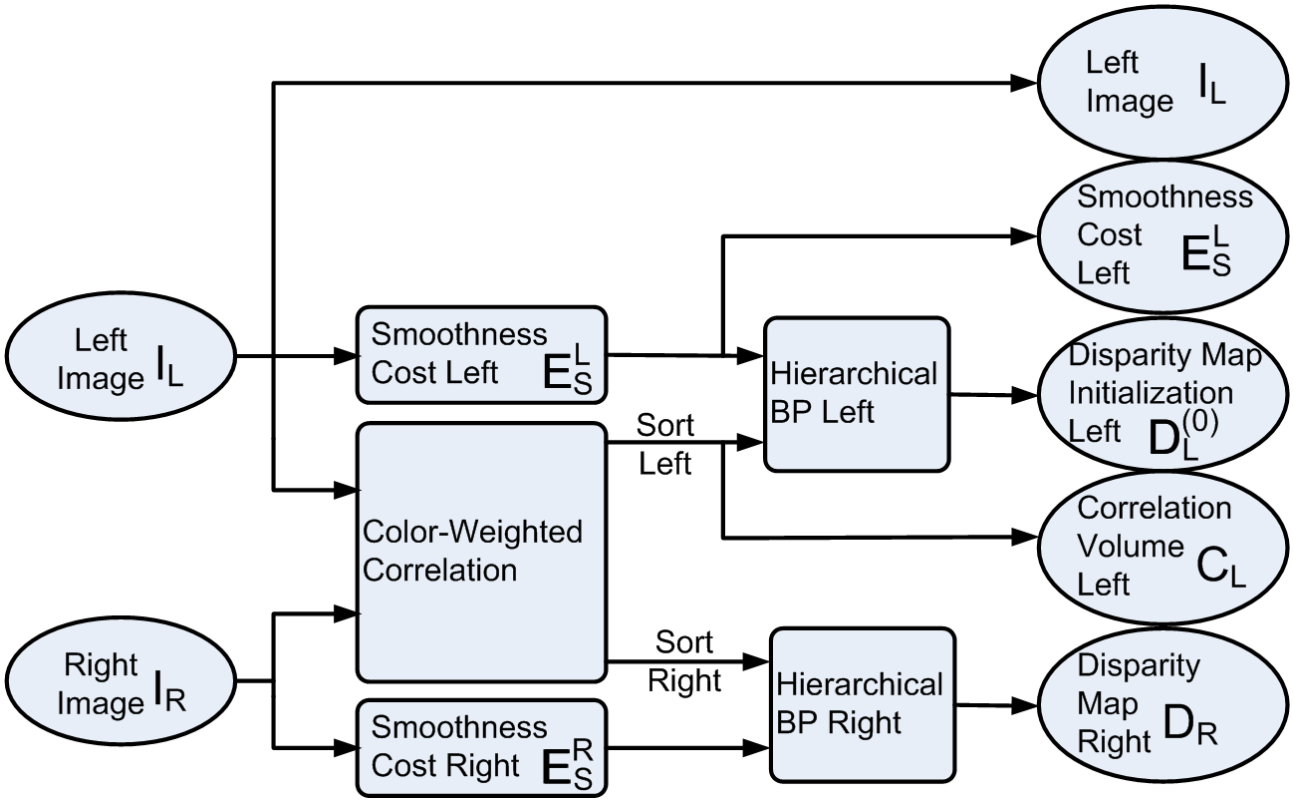


Figure 1. The initial stereo module. Hierarchical belief propagation is run with both the left and right images as reference image. The data term used is based on the color-weighted correlation, and the smoothness term is computed based on the color gradients in the reference image, see the text for more details.

$[0, 1]$. The average over the whole frame is then subtracted out to yield the normalized difference δ_{norm}. Defining the cost coefficient

$$\rho_s = 1 - \delta_{norm}, \tag{4}$$

the cost assigned to the pixel pair $(\mathbf{x}, \mathbf{y})$ is then

$$E_S = \rho_{bp}\rho_s|D(\mathbf{x}) - D(\mathbf{y})|, \tag{5}$$

where ρ_{bp} is set empirically and $D(\mathbf{x})$ and $D(\mathbf{y})$ are the disparities of $\mathbf{x}$ and $\mathbf{y}$.

Hierarchical loopy belief propagation [6] is employed to realize the iterative optimization that trades off between the data and the smoothness term. The difference between the hierarchical BP and general BP is that the hierarchical BP works in a coarse-to-fine manner, first performing BP at the coarsest scale, then using the output from the coarser scale to initialize the input for the next scale. Two main parameters s_{bp} and n_{bp} define the behavior of this hierarchical belief propagation algorithm, s_{bp} is the number of scales and n_{bp} is the number of iterations in each scale.

3.2. Pixel Classification

The main building blocks of the pixel classification, see Figure 2, are the mutual consistency check and the correlation confidence measure.

The mutual consistency check requires that on the pixel grid that the left and right disparity maps are computed, they are perfectly consistent, i.e. that

$$D_L(\mathbf{x}_L) = D_R(\mathbf{x}_L - D_L(\mathbf{x}_L)) \tag{6}$$

for a particular pixel $\mathbf{x}_L$ in the left image. If this relation does not hold, the pixel is declared occluded. If it does hold, the pixel is declared unoccluded and passed on to the correlation confidence measure.

The correlation confidence is measuring how distinct the best peak in the correlation is for a particular pixel. Assume that the cost for the best disparity value is C_1, and the cost for the second best disparity value is C_2. The correlation confidence is then

$$|\frac{C_1 - C_2}{C_2}|. \tag{7}$$

If it is above a threshold α_s the pixel is declared stable, otherwise unstable.

Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)
0-7695-2597-0/06 $20.00 © 2006 IEEE

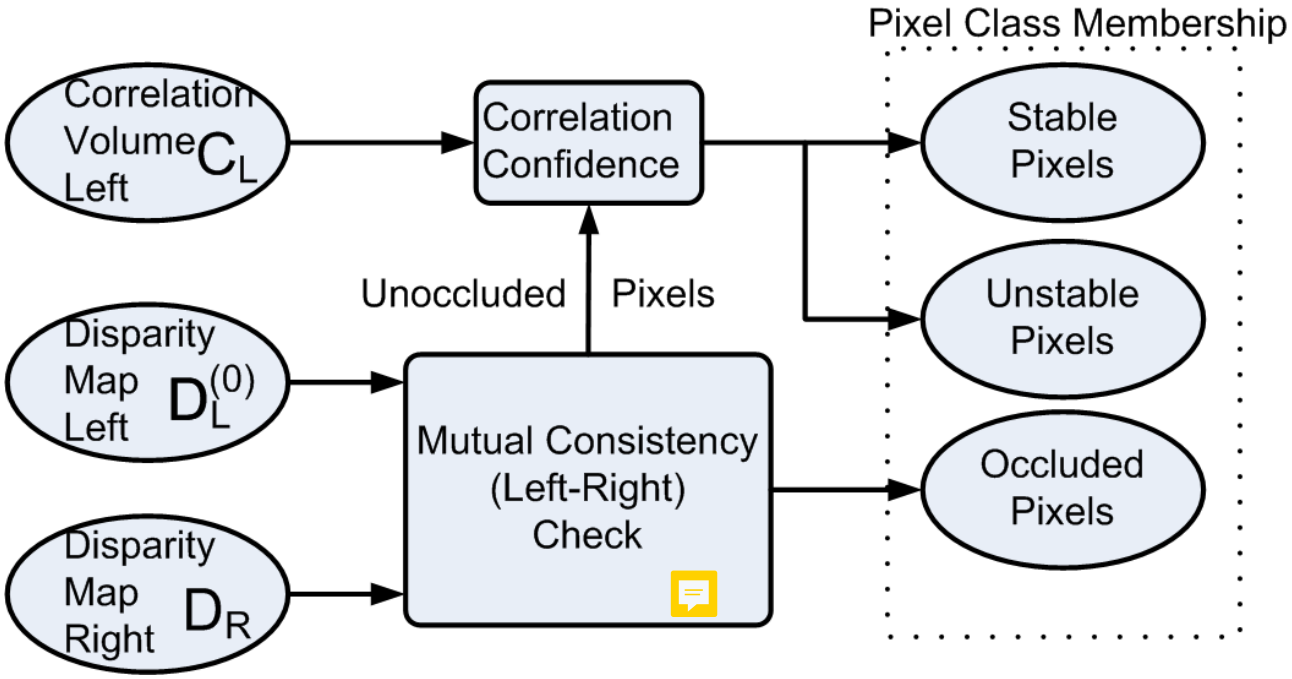


Figure 2. The pixel classification module. Pixels are classified into occluded pixels, unstable pixels and stable pixels. The occluded pixels are the ones that fail a mutual consistency check. The unoccluded pixels are then further divided into stable and unstable pixels based on a confidence measure derived from the correlation volume.

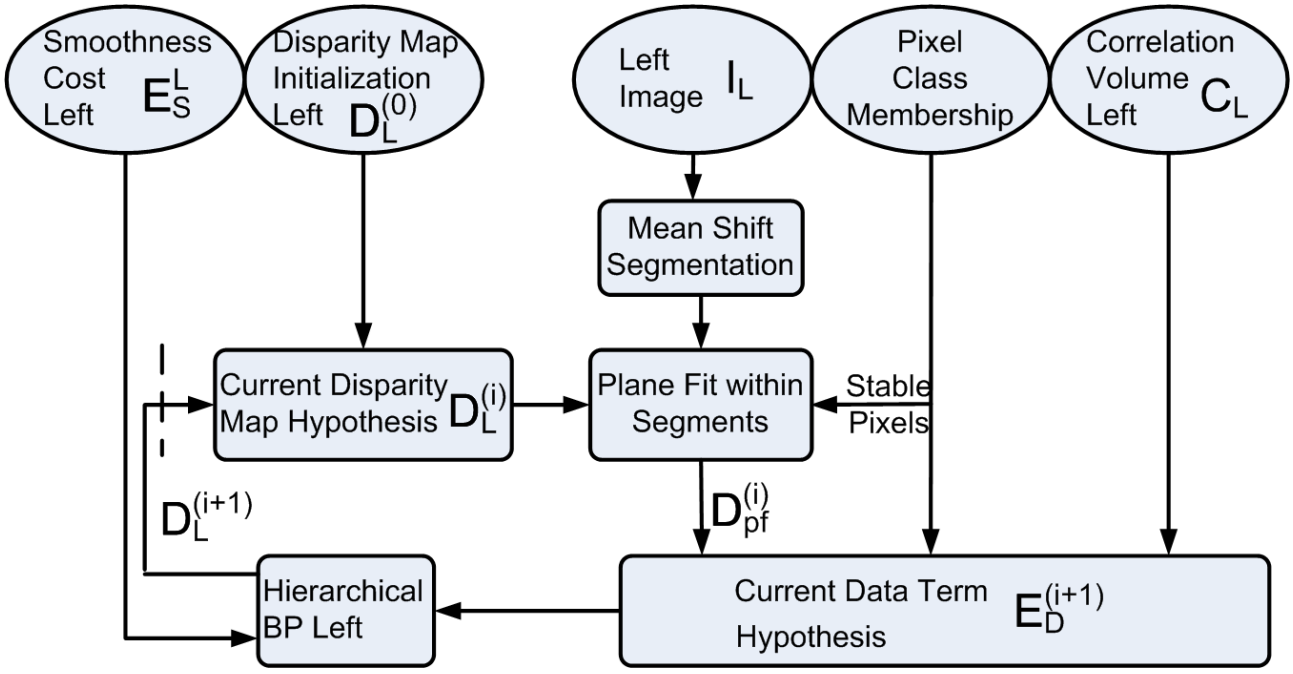


Figure 3. The iterative refinement block, where the goal is to propagate information from the stable pixels to the unstable and the occluded pixels. Mean shift color segmentation is used to derive segments. Within each segment plane fitting is then applied to the stable pixels, using the depth values from the current disparity map hypothesis. The result $D_{pf}^{(i)}$ from the plane fitting is then used together with the correlation volume and the pixel class membership to produce a new approximation $E_D^{(i+1)}$ of the data term. The data term is used with the original smoothness term in another round of hierarchical belief propagation. This gives a new disparity map hypothesis $D_L^{(i+1)}$, which is fed back into the process.

Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)
0-7695-2597-0/06 $20.00 © 2006 IEEE

3.3. Iterative Refinement

The main building blocks of the iterative refinement, see Figure 3, are the mean shift color segmentation, plane fitting within segments, the data term formulation, and another hierarchical belief propagation process identical to the previous ones.

The mean shift color segmentation is performed as described in [5].

The plane fitting is performed in the disparity space, and is applied per segment. This is done robustly using RANSAC [7] on the disparity values of the stable pixels only. The output $D_{pf}^{(i)}$ from this step is computed individually for each segment and depends on on the ratio of stable pixels of this segment. If the ratio of stable pixels is above a parameter value η_s, this means most of the current disparity values for the segment are approximated accurately so we use $D_L^{(i)}$ for the stable pixels, and for the unstable and occluded pixels we use the result of the plane fitting. If the ratio of stable pixels is below η_s we use the result of the plane fitting for all pixels.

The data term is formulated differently for the occluded, unstable and stable pixels. The absolute difference

$$a_i = |D_L^{(i+1)} - D_{pf}^{(i)}| \tag{8}$$

between the new disparity map $D_L^{(i+1)}$ and the plane fitted disparity map $D_{pf}^{(i)}$ is used to regularize the new estimation process. The difference is used to define the data term at the occluded, unstable and stable pixels as

$$E_D^{(i+1)} = \kappa_o a_i, \tag{9}$$

$$E_D^{(i+1)} = C_L + \kappa_u a_i, \tag{10}$$

$$E_D^{(i+1)} = C_L + \kappa_s a_i, \tag{11}$$

respectively. The constants $\kappa_o, \kappa_u, \kappa_s$ reflect the fact that the unstable and occluded pixels need the most regularization.

3.4. Parameter Settings

In this section, we provide all the parameter settings used in the algorithm. The same parameter settings were used throughout.

The parameters are shown in Table 1 and separated into 4 parts: 3 parameters (α_{ms}, β_{ms}, γ_{ms}) for the mean shift segmentation, 3 parameters (α_{cw}, β_{cw}, γ_{cw}) for color-weighted correlation, 6 parameters (α_{bp}, η_{bp}, ρ_{bp}, λ_{bp}, s_{bp}, n_{bp}) for hierarchical belief propagation, and 6 parameters (κ_s, κ_u, κ_o, α_s, η_s, n_s) for iterative refinement.

For mean shift color segmentation, α_{ms} is spatial bandwidth, β_{ms} is color bandwidth, and γ_{ms} is the minimum region size.

Mean Shift	α_{ms}	β_{ms}	γ_{ms}			
Segmentation	7	6	50			
Color-Weigh.	α_{cw}	β_{cw}	γ_{cw}			
Correlation	33	10	21			
Hierarchical	α_{bp}	η_{bp}	ρ_{bp}	λ_{bp}	s_{bp}	n_{bp}
BP	$n_d/8$	$2\overline{c}$	1	0.2	5	5
Iterative	κ_s	κ_u	κ_o	α_s	η_s	n_s
Refinement	0.05	0.5	2	0.04	0.7	5

Table 1. Parameter settings used throughout. n_d is the number of disparity levels. $\overline{c}$ is the average of the values in the correlation volume.

For color-weighted correlation α_{cw} is the size of the support window and β_{cw} and γ_{cw} are defined in Equation (2).

For hierarchical BP, α_{bp} and η_{bp} are truncations of the smoothness and data terms, respectively. The parameter ρ_{bp} is the constant weight factor applied to the smoothness term and λ_{bp} is a constant weight factor applied to the data term after the truncation. The parameter s_{bp} is the number of scales and n_{bp} is the number of iterations, as defined in Section 3.1.

Parameters κ_s, κ_u and κ_o for the iterative refinement are defined in Equations (9), (10) and (11), respectively. α_s is the threshold on correlation confidence defined in Section 3.2. Parameter η_s is related to the plane fitting process, as defined in Section 3.3. The parameter n_s is the number of iterations for the iterative refinement process.

4. Experimental Results

We evaluate our algorithm on the Middlebury data set and we show in Table 2 that our algorithm outperforms all the other algorithms listed on the Middlebury homepage. The result on each data set is computed by measuring the percentage of pixels with an incorrect disparity estimate. This measure is computed for three subsets of the image:

- The subset of non-occluded pixels, denoted "nonoccl".
- The subset of pixels near the occluded regions, denoted "disc".
- The subset of pixels being either non-occluded or half-occluded, denoted "all".

For the first two categories our algorithm takes the first place for all four test sets. For the third category we take first or second place for all test sets. By consistently performing first or second on all test subsets our average rank is 1.3.

In Figure 6 the results after different intermediate stages are shown. This provides a visual explanation of how the different stages in the pipeline improves the results. For

Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)
0-7695-2597-0/06 $20.00 © 2006 IEEE

Algorithm	Avg. Rank	Tsukuba nonocc	Tsukuba all	Tsukuba disc	Venus nonocc	Venus all	Venus disc	Teddy nonocc	Teddy all	Teddy disc	Cones nonocc	Cones all	Cones disc
Our Algorithm	**1.3**	**0.88** $_1$	**1.29** $_1$	**4.76** $_1$	**0.14** $_1$	0.60 $_2$	**2.00** $_1$	**3.55** $_1$	8.71 $_2$	**9.70** $_1$	**2.90** $_1$	9.24 $_2$	**7.80** $_1$
Segm+visib [2]	3.3	1.30 $_5$	1.57 $_2$	6.92 $_6$	0.79 $_4$	1.06 $_3$	6.76 $_6$	5.00 $_2$	**6.54** $_1$	12.3 $_2$	3.72 $_3$	**8.62** $_1$	10.2 $_4$
SymBP+occ [13]	3.4	0.97 $_2$	1.75 $_3$	5.09 $_2$	0.16 $_2$	**0.33** $_1$	2.19 $_2$	6.47 $_4$	10.7 $_3$	17.0 $_4$	4.79 $_7$	10.7 $_6$	10.9 $_5$
AdaptWeight [17]	4.4	1.38 $_6$	1.85 $_4$	6.90 $_5$	0.71 $_3$	1.19 $_4$	6.13 $_4$	7.88 $_5$	13.3 $_5$	18.6 $_6$	3.97 $_5$	9.79 $_4$	8.26 $_2$
SemiGlob [8]	5.8	3.26 $_{10}$	3.96 $_9$	12.8 $_{13}$	1.00 $_5$	1.57 $_5$	11.3 $_{10}$	6.02 $_3$	12.2 $_4$	16.3 $_3$	3.06 $_2$	9.75 $_3$	8.90 $_3$

Table 2. Comparison of results on the Middlebury data set. The numbers are the percentage of pixels with misestimated disparities on the different subsets of the images. The subscript of each number is the rank of that score. Our algorithm has rank 1 for most categories and rank 2 at worst. This gives an average rank of 1.3.

comparison we also give the ground truth. The scores for the intermediate results are given in Figure 7 along with $D_L^{(5)}$ SPECIAL, which is the same as $D_L^{(5)}$ except that we do not use the colors of the reference image to define the smoothness cost, which has a strong impact on the Teddy and Cones data sets.

In Figure 4 and Figure 5, it is shown how an increased number of iterations in estimating E_D improves the result. Zero iterations in Figure 4 means that we use $D_L^{(0)}$, the initial disparity map. Based on this we chose to use five iterations in our method.

5. Conclusions

In this paper, a stereo model based on energy minimization, color segmentation, plane fitting, and repeated application of hierarchical belief propagation was proposed. Typically, one application of the hierarchical belief propagation brings the error down close to its final value, so that the algorithm could perhaps be used as a two step approach, where occlusions and untextured areas are first detected and then filled in from neighboring areas.

The parameters provided constitute a good setting for the algorithm, but are not entirely optimized. More studies are needed to fully understand the behavior of our algorithm. Our algorithm is outperforming all other algorithms on the Middlebury data set, but there is space left for improvement. For instance, in our algorithm, we only refined the disparity map for the reference image, but [13] suggests that by generating a good disparity map for the right image, the occlusion constraints can be extracted more accurately. Another question that was left for further study is how to use the algorithm in a multi-view setting.

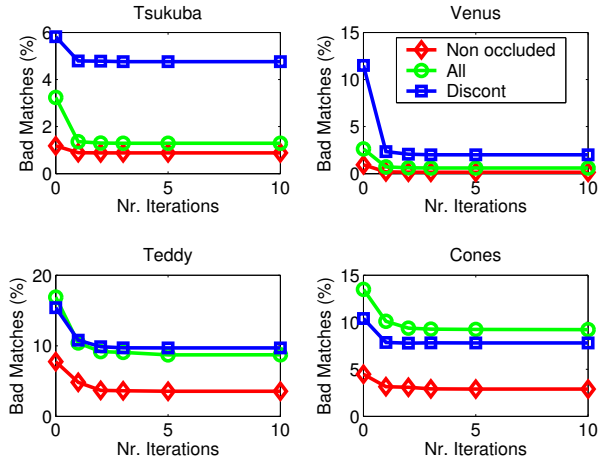


Figure 4. The iterated computations of $E_D^{(i+1)}$ improves the result. In most cases one iteration is enough for convergence. After five iterations, the result has always converged.

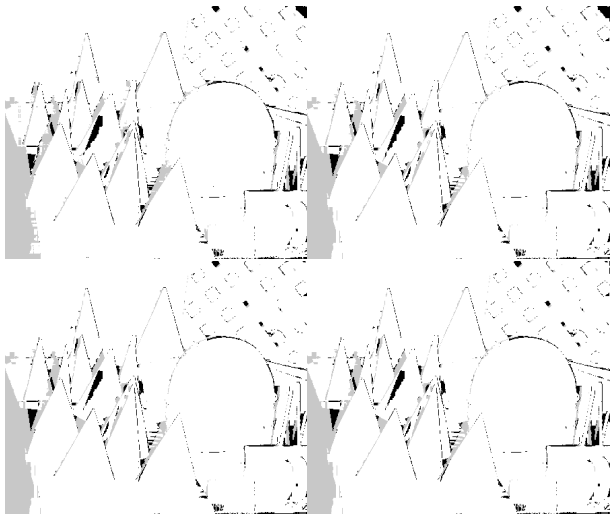
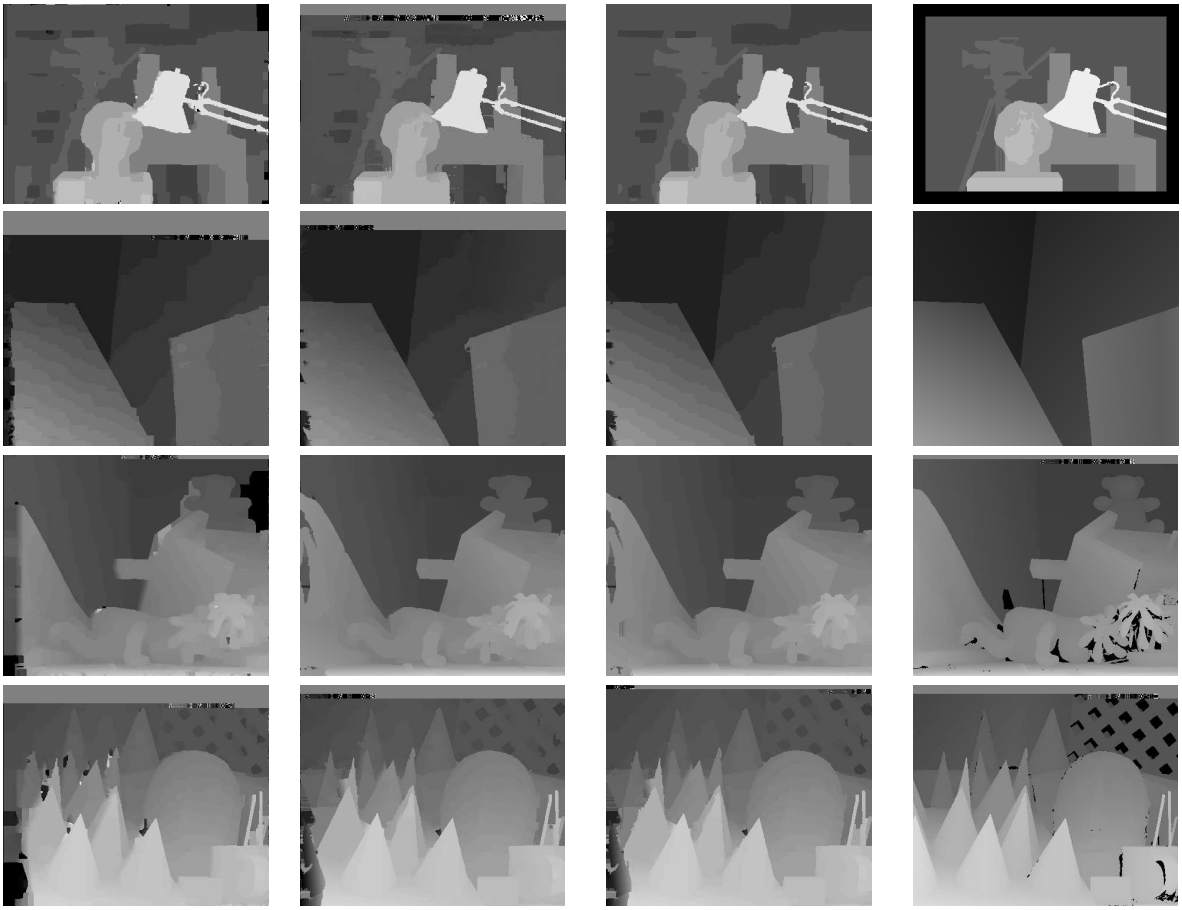


Figure 5. Pixels with incorrect disparity for the "Cones" data set. On the first row the results after 1 and 2 iterations are shown and on the second row the results after 3 and 5 iterations are shown.

References

[1] S. Birchfield and C. Tomasi, A Pixel Dissimilarity Measure That Is Insensitive to Image Sampling, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 4, April 1998.

Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)
0-7695-2597-0/06 $20.00 © 2006 IEEE



Initial Disp. ($D_L^{(0)}$) Plane Fitting ($D_{pf}^{(4)}$) Final Disparity Map($D_L^{(5)}$) Ground Truth

Figure 6. Intermediate results from our algorithm for the four different test sets compared to the ground truth. In the first column the output of the initial BP is shown. This result is denoted $D_L^{(0)}$ in Figure 2 and Figure 3. In the second column the results after fitting planes to the regions from the color segmentation are shown. These are denoted by $D_{pf}^{(4)}$ in Figure 3. In the third column the final result of our algorithm is shown. These results are denoted by $D_L^{(5)}$ in Figure 3.

Algorithm	Avg. Rank	Tsukuba nonocc	Tsukuba all	Tsukuba disc	Venus nonocc	Venus all	Venus disc	Teddy nonocc	Teddy all	Teddy disc	Cones nonocc	Cones all	Cones disc
$D_L^{(0)}$	5.7	1.18 2	3.24 8	5.82 2	0.94 4	2.63 10	11.5 10	7.75 4	16.9 7	15.4 2	4.47 6	13.5 9	10.4 4
$D_{pf}^{(4)}$	3.2	2.60 9	2.98 8	7.31 6	**0.13** 1	0.46 2	**1.79** 1	**3.93** 1	8.92 2	**9.97** 1	3.50 2	9.41 2	9.07 3
$D_L^{(5)}$	**1.3**	**0.88** 1	**1.29** 1	**4.76** 1	**0.14** 1	0.60 2	**2.00** 1	**3.55** 1	8.71 2	**9.70** 1	**2.90** 1	9.24 2	**7.80** 1
$D_L^{(5)}$ SPECIAL	**1.3**	**0.88** 1	**1.30** 1	**4.77** 1	**0.14** 1	0.60 2	**1.95** 1	**3.71** 1	9.20 2	**10.3** 1	**3.07** 2	9.33 2	**8.17** 1

Figure 7. The first three rows in the table corresponds to the first three rows in the above figure. The last row is the same as $D_L^{(5)}$ except that we do not use the colors of the reference image to define the smoothness cost.

[2] M. Bleyer and M. Gelautz, A Layered Stereo Algorithm Using Image Segmentation and Global Visibility Constraints, *IEEE International Conference on Image Processing*, pp. 2997-3000, 2004.

[3] Y. Boykov, O. Veksler and R. Zabih, A Variable Window Approach to Early Vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 12, 1998.

[4] Y. Boykov, O. Veksler and R. Zabih, Fast Approximate Energy Minimization via Graph Cuts, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 11, 2001.

[5] D. Comaniciu and P. Meer, Mean shift: A Robust Approach

Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)
0-7695-2597-0/06 $20.00 © 2006 IEEE

Toward Feature Space Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 4, May 2002.

[6] P. F. Felzenszwalb and D. P. Huttenlocher, Efficient Belief Propagation for Early Vision, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. I:261-268, 2004.

[7] M. A. Fischler and R. C. Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Communications of the ACM*, Vol. 24, pp. 381-395 1991.

[8] H. Hirschmueller, Accurate and Efficient Stereo Processing by Semi-global Matching and Mutual Information, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol II:807-814, 2005.

[9] T. Kanade and M. Okutomi, A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiments, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 9, September 1994.

[10] V. Kolmogorov and R. Zabih, Computing Visual Correspondence with Occlusions using Graph Cuts, *IEEE International Conference on Computer Vision*, Vol. I:508-515 2001.

[11] D. Scharstein and R. Szeliski, Middlebury Stereo Vision Research Page, $http://bj.middlebury.edu/\tilde{}schar/stereo/newEval/php/results.php$

[12] J. Sun, N.-N. Zheng and H.-Y. Shum, Stereo Matching Using Belief Propagation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 7, July 2003.

[13] J. Sun, Y. Li, S. B. Kang and H.-Y. Shum, Symmetric Stereo Matching for Occlusion Handling, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. II:399-406, 2005.

[14] H. Tao and H. Sawhney, Global Matching Criterion and Color Segmentation Based Stereo, *IEEE Workshop on Applications of Computer Vision*, pp. 246-253, 2000.

[15] O. Veksler, Stereo Correspondence with Compact Windows via Minimum Ratio Cycle, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 12, 2002.

[16] O. Veksler, Fast Variable Window for Stereo Correspondence using Integral Images, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. I:556-561, 2003.

[17] K.-J. Yoon and I.-S. Kweon, Locally Adaptive Support-Weight Approach for Visual Correspondence Search, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. II:924-931, 2005.

Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)
0-7695-2597-0/06 $20.00 © 2006 **IEEE**