STEREO MATCHING USING ADAPTIVE BELIEF PROPAGATION ALONG AMBIGUITY GRADIENT

*Sumit Srivastava, Sang Hwa Lee, Nam Ik Cho, Sang Uk Lee, Jong-Il Park**

Department of Electrical Engineering, INMC, Seoul National University
San 56-1 Shillim-dong Gwanak-Gu Seoul, 151-742, Korea
*Department of Electronics and Computer Eng. Hanyang University, Seoul, Korea
email: sumit.xenon@gmail.com, lsh@ipl.snu.ac.kr, nicho@snu.ac.kr, sanguk@ipl.snu.ac.kr, jipark@hanyang.ac.kr

ABSTRACT

This paper proposes a stereo matching algorithm based on hierarchical belief propagation and occlusion handling. We define a new order for message passing in belief propagation instead of the scanline approach. The primary assumption is that a pixel with a well-defined minimum in its likelihood field is more likely to contain a correct disparity, when compared to a pixel having an ill-defined minimum with several local minima. The order for message passing is determined by the variance of likelihood field at each pixel. The variances evaluate the ambiguity of likelihood fields, and the messages are hierarchically updated along the gradient of ambiguity. We also propose occlusion handling method which incorporates information from color segmentation and likelihood field. The occluded pixels are detected based on the ambiguity of likelihood field and reliability of neighboring disparity field. Then, the occlusions are filled according to the statistics of neighboring disparities in the same segment. The experimental results show that the proposed method estimates the disparities correctly in the hard regions such as large occlusions and textureless regions. The proposed algorithm currently ranks second on the Middlebury stereo site.

1. INTRODUCTION

Stereo matching is one of the most researched area in the field of computer vision. The stereo matching is basically a problem of correspondence. That is to say, given two images I_{left} and I_{right}, we have to find a corresponding pixel in image I_{right} for each pixel $p \in I_{left}$. The two images I_{left} and I_{right} are usually rectified so that the relative shift of each pixel is purely horizontal. Excellent overview of the various issues involved in stereo matching is presented by Scharstein and Szeliski [1]. The algorithms proposed to solve the stereo matching problem can be broadly classified into two categories. Global algorithms are those in which a global energy function needs to be minimized to find the disparity field. Local algorithms, on the other hand, estimate disparity at a pixel using information available in a finite neighboring window.

Many effective local algorithms have been reported in literature for solving the stereo correspondence problem [11, 12, 13]. Kanade used an adaptive window mechanism to take care of both textureless and disparity discontinuity regions [11]. Yoon has used color similarity and geometric proximity to form an adaptive support which is very accurate in reducing image ambiguities [12]. Tombari modified the cost function defined by Yoon using segmentation information instead of Euclidian distance. These approaches solved the stereo correspondences by using likelihood functions with some adaptive deformable windows [13].

Recently global energy minimization methods and Markov random field (MRF) have been combined for stereo matching. Stochastic diffusion, graph cut, and belief propagation are the global energy minimization methods for Markov random field (MRF) models [1]. Stochastic or nonlinear diffusion diffuses the energy function to the stable states using probabilistic models of neighboring disparities [5, 6]. Graph cuts provided fast energy minimization scheme based on the graphs [2]. Kolmogorov and Zabih developed a method which was accurate in detecting occlusions as well as computing disparities [14]. Belief propagation (BP) is another popular inference scheme used in stereo matching. Sun *et al.* have formulated stereo matching problem as a Markov network [7]. Disparity, depth discontinuity, and occlusion are represented by three coupled MRF's. BP algorithm finds the MAP estimation in the Markov network. Various modifications of BP has been reported in literature. Yang *et al.* defined a hierarchical scheme and used BP to iteratively refine the disparity plane in the occluded and low texture areas [8]. Sun *et al.* have devised a symmetric framework to deal with occlusions, and used the conventional BP to minimize the 3-D energy field [9]. Zitnick *et al.* used image over-segmentation, and computed matches over entire segments to provide robustness against noise and intensity bias. The energy field was minimized by BP [10].

The proposed algorithm is a global stereo matching algorithm using an adaptive likelihood function and hierarchical BP. The likelihood function is based on the color segmentation and geometric distance. The proposed hierarchical BP scheme first determines the order of pixels to propagate their messages with priority. The beliefs of pixels are passed to the neighborhood along the ambiguity gradient which specify the order and the direction of propagation. The proposed stereo matching algorithm has the main distinguishing features that separate our approach from the earlier ones.

1. All the earlier methods based on BP involved updating and passing of messages along the scanline. However, we propose a novel scheme in which the messages are updated along the ambiguity gradient.

2. In all the algorithms, the strength of BP is fixed by specifying a set of parameters which remains same for all pixels across all images. In our scheme, these parameters have been automatically determined for each pixel. Thus the strength of BP is adaptive and optimized.

3. A novel refinement stage is devised to take care of the occluded pixels. The occluded pixels are detected based on the likelihood energy field, and they are corrected according

to the statistical models of disparities in the segment.

The rest of paper is organized as follows. Section 2 describes the likelihood function based on the color segmentation and geometric distance. The proposed hierarchical BP scheme based on ambiguity gradient is explained in Section 3. We describe how to handle the occlusion in Section 4. We show the experimental results in Section 5, and finally conclude this paper in Section 6.

2. LIKELIHOOD MODELS

2.1 Color Segmentation

The first step in our scheme is to perform multi level segmentation on both images I_{left} and I_{right}. For performing color segmentation we have used the mean shift algorithm developed by Comanaciu *et al.* [15]. The basic assumption behind using the segmentation clue is that disparity is smooth in a given segment and that discontinuity occurs at segmentation boundaries. Consequently that over segmentation is preferred in comparison with under segmentation as it better preserves the assumption. Also the result of color segmentation is exploited to calculate the likelihood field.

2.2 The Likelihood Field

The likelihood model used in our method is based on the algorithm proposed by Yoon and Kweon [12]. Their method creates an adaptive support by assigning a weight to each pixel in the current correlation window of both the target and reference image. The matching cost was given by the following equation,

$$E(p_c,q_c) = \frac{\sum\limits_{p_i\in w_i, q_i\in w_i} w_r(p_i,p_c).w_t(q_i,q_c).e(p_i,q_i)}{\sum\limits_{p_i\in w_i, q_i\in w_i} w_r(p_i,p_c).w_t(q_i,q_c)}, \quad (1)$$

where w_r and w_t are the correlation window in the reference and target images respectively. p_c and q_c are the central pixels of reference and target images. We modified the weight function in [12] as follows,

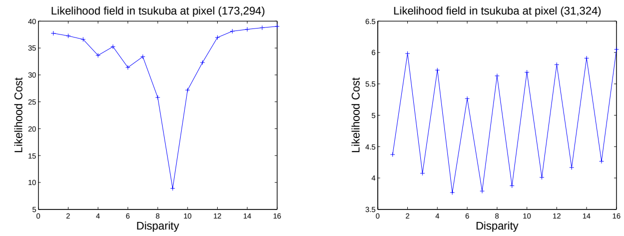
$$w_r(p_i,p_c) = \begin{cases} 1, & p_i \in S_c, \\ \exp\left(-\frac{\alpha(p_i,p_c)}{\gamma_p} - \frac{\beta(p_i,p_c)}{\gamma_c}\right), & p_i \notin S_c. \end{cases} \quad (2)$$

where the term $\beta(p_i,p_c)$ represents Euclidian distance between two RGB triplets at p_i and p_c. In (2), $\alpha(p_i,p_c)$ is the Euclidian distance between the coordinates of p_i and p_c. γ_p and γ_c are the parameters of the algorithm, and S_c is the segment containing the central pixel p_c. w_t is calculated in a similar fashion in the target image as in (2). We assign higher weights for the pixels inside the same segment. This modified likelihood function increases the robustness of the likelihood model.

3. BELIEF PROPAGATION ALONG THE AMBIGUITY GRADIENT

3.1 Preliminaries

Belief propagation is one of the most extensively used energy minimization scheme. Its capacity for asymmetric transfer of entropy across the 3-D energy field allows it to deal with depth discontinuity and textureless areas. The local evidence



(a) (b)

Figure 1: Examples of likelihood fields at pixels in tsukuba image. (a) The well defined likelihood field has only one dominant minimum, (b) ill defined likelihood field has multiple minima.

and the smoothness prior used for passing of messages are both represented by the following potential functions,

$$\rho_d(d_s) = -\ln\left((1-e_d)\exp\left(-\frac{e(p,d_s)}{\sigma_d}\right)+e_d\right), \quad (3)$$

$$\rho_p(d_s,d_t) = \frac{|d_s-d_t|}{\sigma_p}, \quad (4)$$

where $\rho_d(d_s)$ and $\rho_p(d_s,d_t)$ are the robust functions for the local evidence and the smoothness prior respectively. $e(p,d_s)$ is the likelihood of pixel p having disparity d_s. σ_d and σ_p are the parameters which can be used to control the strength of local evidence and smoothness prior.

3.2 Definition of Ambiguity Gradient

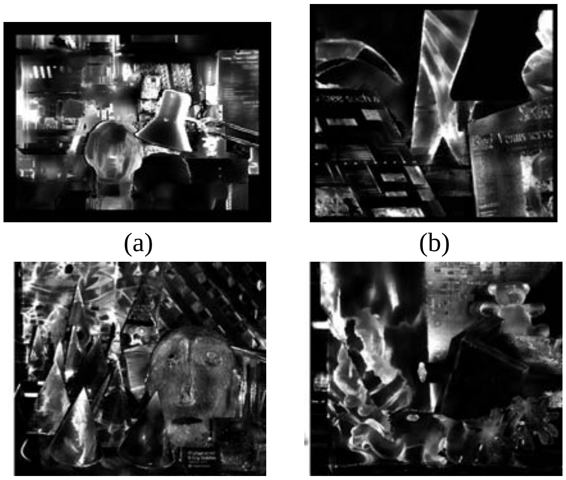
The profile of the likelihood field at any particular pixel can be a good evidence to see if the minimum cost is representing the correct disparity. For example, a pixel likely to contain a correct disparity at minimum cost will have a well defined minimum in the likelihood field as shown in Fig. 1. (a), whileas a pixel with high ambiguity in its likelihood field is less likely to contain the correct disparity as shown in Fig. 1. (b). Thus it is very essential to devise a method which separates the unambiguous pixels from the ambiguous ones.

In order to put a quantitative perspective to our discussion, we define a confidence index (CI) measure to quantify the profile of likelihood field,

$$CI(p) = v_{lm}(p) + \frac{1}{n_{lm}(p)}, \quad (5)$$

where n_{lm} is the number of local minima in the likelihood field at a pixel p. v_{lm} is the variance of all the local minima that occur in the likelihood field at pixel p. The higher is the confidence index CI at pixel p, the lower is the ambiguity at that pixel. Next, we perform a L-R consistency check to find out the occluded pixels. Let the occluded pixels be denoted by p_o and the unoccluded pixels be denoted by p_{uo}.

The hierarchical belief propagation is performed according to the index and occlusion. For each color segment, we first find the unoccluded pixel with the highest confidence index, and update and propagate its message to the neighborhood. Then, we select the unoccluded pixel with the next highest confidence index, and repeat the message updating



(a) (b) (c) (d)

Figure 2: Ambiguity Gradient for test images, (a) tsukuba, (b) venus, (c) cones, and (d) teddy. The dark pixels have the high ambiguity.

and propagating procedure. This hierarchical BP process continues until we reach the pixel with the lowest confidence index in the segment. Next, we repeat the same procedure for all the occluded pixels in the segment. This ordering of message passing effectively utilizes information with highest reliability. In Fig. 2, the confidence field for all the test images are shown. The bright pixels have low ambiguity, whileas dark ones have high ambiguity.

3.3 Adaptive Belief Propagation

During our experimentation, we found that one of the most critical factor was the strength of belief propagation. Intuitively it makes sense to strengthen propagation of belief in regions of high ambiguity and to weaken it in the region of low ambiguity. Ambiguity mainly occurs due to small support size of correlation window, or due to presence of depth discontinuity. In both the cases, increasing strength of belief propagation as compared to likelihood field helps in better and faster removal of ambiguity. In our framework, the strength of belief propagation is adaptively determined by the parameter σ_p and the strength of likelihood field is determined by the parameter σ_d. We devise a method to automatically compute the value of σ_p while σ_d was kept fixed. The method is described by the following procedure.

1. First, we compute the average variance in the likelihood field for each segment,

$$AV(i) = \frac{\sum\limits_{p\in i, p\in p_{uo}} v_{lm}(p)}{\sum\limits_{p\in i, p\in p_{uo}} 1}, \tag{6}$$

$AV(i)$ is the average variance of segment i and p is the set of all pixels belonging to segment i. The occluded pixels are excluded from the determination of average variance since the energy field is very unstable in the occluded region.
2. At each pixel p, value of σ_p is determined by the following equations,

if $AV(p) > v_{lm}(p)$

$$\sigma_p(p) = \max\left(1, 16 - \exp\left(\frac{AV(p) - v_{lm}(p)}{\alpha_p}\right)\right), \tag{7}$$

otherwise

$$\sigma_p(p) = \min\left(16, \exp\left(\frac{v_{lm}(p) - AV(p)}{\alpha_p}\right)\right). \tag{8}$$

3. For each occluded pixel p_o, value of σ_p has been fixed to unity.

In (7) and (8), α_p is the parameter to control the range of values taken by σ_p. σ_p is calculated for each and every pixel. Basically we are trying to reduce the value of σ_p in high ambiguity region so that the strength of belief propagation increases. This automatic determination of σ_p makes belief propagation adaptive to variation of ambiguity.

Our primary aim is to find reliable pixels in each segment whose ambiguity in the likelihood field is as low as possible. We first propagate belief of such reliable pixels to unreliable ones. In Fig. 3, the belief is propagating from the brighter region to the darker region in each segment, and the strength of belief propagation is stronger in the darker region.

3.4 Energy Formulation

Global algorithms generally formulate an energy minimization scheme for stereo matching,

$$E(d, I_L, I_R) = E_d(d, I_L, I_R) + E_s(d), \tag{9}$$

where E_d is the data term derived from likelihood model, and E_s represents the smoothness assumptions for the disparity plane. In the proposed framework, the term E_d is represented by the following equations,

$$\begin{aligned} E_d(d, I_L, I_R) = &\sum_p ((1-o_p)\rho_d(d_p) + \alpha_o.o_p) \\ &+ \sum_p (1-o_p)|(d_p - (\alpha_p x_p + \beta_p y_p + \gamma_p))|, \end{aligned} \tag{10}$$

where the last term in (10) represents the soft constraint imposed by segmentation. The term α_p, β_p, γ_p represent parameters of a 3-D plane computed for each segment. The term α_o is the penalty for the occluded pixels. The smoothness energy is defined as follows,

$$E_s(d) = \sum_p \left(\sum_{t\in n(p)} \exp(-\rho_p(d_p, d_t))\right), \tag{11}$$

where the term $t \in n(p)$ represents all pixels t that belong to a first order neighborhood of pixel p, and the term $\rho_p(d_p, d_t)$ has been defined in (4). At the end of each iteration, the 3 plane parameters and σ_p to represent the strength are recomputed to better reflect the effect of belief propagation.

4. OCCLUSION HANDLING

Occlusion is one of the most fundamental problem encountered in the stereo matching problem. In this stage we formulate a novel effective way to find the disparity at occluded pixels. The proposed scheme for occlusion handling is explained by the following step by step procedure.

1. Calculate the average confidence index for each segment,

$$AC(i) = \frac{\sum_{p \in i} CI(p)}{\sum_{p \in i} 1}. \tag{12}$$

2. For each segment, classify pixels into two categories, reliable and unreliable. A pixel p belonging to a segment i is considered reliable if

$$CI(p) > AC(p). \tag{13}$$

Otherwise, the pixel p belonging to segment i is considered unreliable.

3. For each occluded pixel p_o, find the nearest reliable pixel in the direction of the scanline. The disparity at the nearest reliable pixel is used as the disparity of occluded pixel.

4. Find the mode of disparity for each segment using only the reliable pixels in the segment. Let d_i be the representative disparity for segment i.

5. Classify all the non-occluded pixels in the segment as confident and non-confident based upon the following criteria. A pixel p in the segment i is considered confident if

$$disp(p) \in [d_i - 1, d_i + 1], \tag{14}$$

where d_i is the representative disparity for segment i.

6. At every non confident pixel p_{nc}, find if local minimum exists in the likelihood field in following range $[d_i - 1, d_i + 1]$.

7. If a minimum does not exist, leave the disparity at pixel p_{nc} unchanged, otherwise, for each minimum, find the following belief function,

$$b(d) = m.s.(d) + hist_i(d), \tag{15}$$

$$m.s.(d) = [e(d-1) - e(d)] + [e(d+1) - e(d)], \tag{16}$$

where $b(d)$ is the belief for disparity d. The term $m.s.(d)$ is the strength of local minimum at disparity d. $e(d)$ is the likelihood value at disparity d at pixel p_{nc}. The term $hist_i(d)$ is the ratio of pixels having disparity d in the segment i.

8. Replace the disparity at p_{nc} with the disparity for which the belief function is maximum.

5. EXPERIMENTAL RESULTS

5.1 Parameter Settings

We will now provide the numerical values for all the parameters used in our algorithm. It is to be noted that the same parameters have been used for all the test images. The parameters are listed in Table 1. 3 parameters $(\gamma_{cs}, \sigma_{cs}, \alpha_{cs})$ are for the color segmentation. 3 parameters (cw, γ_p, γ_c) define the likelihood functions, and 4 parameters $(e_d, \sigma_d, \sigma_p, \alpha_p)$ are the parameters for adaptive belief propagation. Specifically, cw is the block size of the support window used for computing the likelihood values. γ_c and γ_p are the weighing factors for color and geometric proximities as defined in (2). e_d, σ_p, σ_d, and α_p are the parameters of the potential functions and have been defined in (3),(4),(7) and (8) respectively. By varying these parameters, the strength of the belief propagation can be varied. The method for determining the value of σ_p has already been discussed in Section 3.

Mean Shift	γ_{cs}	σ_{cs}	α_{cs}	
Segmentation	7	6.5	50	
likelihood	cw	γ_p	γ_c	
field	35	15	17.5	
Belief	e_d	σ_d	σ_p	α_p
Propagation	0.1	12	*automatic*	12

Table 1: Parameter values in the experiments.

5.2 Results

In order to evaluate the proposed method, we followed the methodology proposed by Scharstein and Szeliski [16]. Numerical quality of the result has been measured under three categories.

1. Non_Occ: Here only non_occluded pixels are considered for error evaluation.

2. All: All pixels where disparity in the ground_truth is known are considered.

3. Disc: Pixels belonging to region where discontinuity in the disparity map exists are considered.

The quantitative results are summarized in Table 2 and the disparity maps are shown in Fig. 3. As we can see in Table 2, the proposed algorithm performs best for *tsukuba* across all 3 categories. Our results are very close to the current best results for *venus* and *cones*. The reason for unsatisfactory performance for the *teddy* image is the presence of large areas of high ambiguity region as seen in Fig. 2 (d). Also our algorithm shows very good performance in discontinuity regions where our results are best in all test images except *teddy*. The proposed algorithm ranks second on the middlebury website [16].

6. CONCLUSION

This paper has proposed a stereo matching algorithm using hierarchical belief propagation and segment based occlusion handling. We have proposed a new methodology for message passing in a 3-D energy field. A criteria for ambiguity in the likelihood field is defined, and the belief is propagated along that gradient. We have handled both the textureless and the discontinuity areas effectively by varying the strength of belief propagation. The experimental results show that the proposed method estimates the disparities correctly in the hard regions such as large occlusions and textureless regions. The proposed algorithm currently ranks second on the Middlebury stereo site. The state of the art results achieved during simulation shows the effectiveness of our algorithm. More robust and effective methods to define the ambiguity gradient are possible in the future works.

REFERENCES

[1] D. Scharstein and R. Szeleiski, “A taxonomy and evaluation of dense two frame stereo correspondence algorithm,” *IJCV*, vol. 47(1/2/3), pp. 7-42, June 2002.

[2] Y. Boykov, O. Veksler, and R.Zabih, “Fast approximate energy minimization with graph cuts,” *IEEE Trans. on PAMI*, vol. 23, no. 11, pp. 1222-1239, Nov. 2001.

[3] Ze-Nian Li and Gongzhu Hu, “Analysis of disparity gradient based cooperative stereo,” *IEEE Trans. on Image Processing*, vol. 5, no. 11, pp. 1493-1506, Nov. 1996.

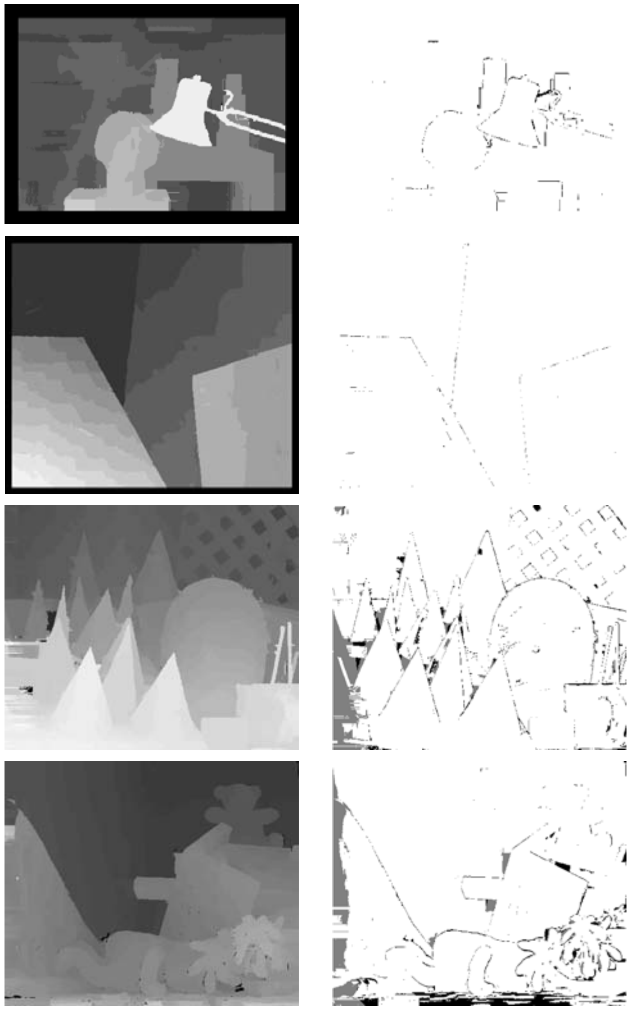


Figure 3: Disparity maps by the proposed method and errors with ground truth. (1^{st} row) tsukuba, (2^{nd}) venus, (3^{rd}) cones, and (4^{th}) teddy.

[4] C. L. Zitnick and T. Kanade, "A Cooperative algorithm fr stereo matching and occlusion detection," *IEEE Trans. on PAMI*, vol. 22, no. 7, pp. 675-684, July, 2000.

[5] S. H. Lee, Y. Kanatsugu, and J.-I. Park, "Stochastic diffusion for stereo matching and line fields estimation," *IJCV*, vol. 47 (1/2/3), pp. 195-218, June 2002.

[6] D. Scharstein and R. Szeliski, "Stereo matching with non-linear diffusion," *IJCV*, vol. 28, no. 2, pp. 155-174, Feb. 1998.

[7] J. Sun, N. N. Zheng, and H. Y. Shum, "Stereo matching using belief propagation," *IEEE Trans. on PAMI*, vol. 25, no. 7, July 2003.

[8] Q. Yang, L. Wang, R. Yang, H. Stewenius, and D. Nister. "Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion Handling," *IEEE Conference on CVPR*, vol. 2, pp. 347-354, 2006.

[9] J. Sun, Y. Li, S. B. Kang, and H. Y. Shum. "Symmetric stereo matching for occlusion handling," *IEEE Conference on CVPR*, vol. 2, pp. 399-406, 2005.

[10] L. Zitnick and S. B. Kang. "Stereo for image-based rendering using image over-segmentation," *IJCV*, vol. 75, Issue 1, Oct. 2007.

[11] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiments," *IEEE Trans. on PAMI*, vol. 16, no. 9, pp. 920-932, Sept. 1994.

[12] K. J. Yoon and I. S. Kweon, "Adaptive support weight approach for correspondence search," *IEEE Trans. on PAMI*, Vol. 28, no. 4, April 2004.

[13] F. Tombari, S. Mattoccia, L. D. Stefano, "Segmentation-based adaptive support for accurate stereo correspondence," LNCS 4872, pp. 427-438, 2007.

[14] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," *IJCV*, July, 2001.

[15] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. PAMI*, vol. 2, no. 4, May 2002.

[16] *http://vision.middlebury.edu/stereo*.

[17] A. Klaus, M. Sormann, and K. Karner. "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," *International Conference on Pattern Recognition*, vol. 3, pp. 15-18, 2006.

[18] Z. Wang and Z. Zheng, "A region based stereo matching algorithm using cooperative optimization," *IEEE Conference on CVPR*, 2008.

[19] L. Xu and J. Jia. "Stereo matching: an outlier confidence approach," *Proc. of ECCV*, 2008.

algorithm	rank	tsukuba			venus			teddy			cones		
		nonocc	all	disc	nonocc	all	disc	nonoccl	all	disc	nonocc	all	disc
Proposed	3.1	**0.79**	**1.13**	**3.97**	0.13	0.24	**1.33**	5.10	8.86	13.2	2.61	7.41	**6.72**
Adapting BP[19]	3.1	1.11	1.37	5.79	**0.10**	0.21	1.44	4.22	7.06	11.8	**2.48**	7.92	7.32
CoopRegion[20]	3.1	0.87	1.16	4.61	0.11	0.21	1.54	5.16	8.31	13.0	2.79	**7.18**	8.01
Double-BP[9]	4.1	0.88	1.29	4.76	0.13	0.45	1.87	3.53	8.30	**9.63**	2.90	8.78	7.79
OutlierConf[21]	4.8	0.88	1.43	4.74	0.18	0.26	2.40	5.01	9.12	12.8	2.78	8.57	6.69
SubPixelBP[16]	6.5	1.24	1.76	5.98	0.12	0.46	1.74	**3.45**	8.38	10.0	2.93	8.53	7.91
OverSegBP	11	1.69	2.04	5.64	0.14	0.20	1.47	7.04	11.1	16.4	3.60	8.96	8.84
Segm+Visb	13.5	1.30	1.57	6.92	0.79	1.06	6.76	5.00	**6.54**	12.3	3.72	8.62	10.2

Table 2: Comparison of results. The bold numbers are best.