Segment-Based Stereo Matching Using Energy-Based Regularization

Dongbo Min, Sangun Yoon, and Kwanghoon Sohn

Dept. of Electrical and Electronics Eng., Yonsei University
134 Shinchon-dong, Seodaemun-gu, Seoul, 120-749, Korea
khsohn@yonsei.ac.kr

Abstract. We propose a new stereo matching algorithm through energy-based regularization using color segmentation and visibility constraint. Plane parameters in the entire segments are modeled by robust least square algorithm, which is *LMedS* method. Then, plane parameter assignment is performed by the cost function penalized for occlusion, iteratively. Finally, disparity regularization which considers the smoothness between the segments and penalizes the occlusion through visibility constraint is performed. For occlusion and disparity estimation, we include the iterative optimization scheme in the energy-based regularization. Experimental results show that the proposed algorithm produces comparable performance to the state-of-the-arts especially in the object boundaries, un-textured regions.

1 Introduction

Stereo matching is one of the most important problems in computer vision. Dense disparity map acquired by stereo matching can be used in many applications including view synthesis, image-based rendering, 3D object modeling, etc. The goal of stereo matching is to find corresponding points in the different images taken from same scene by several cameras.

An extensive review of stereo matching algorithms can be found in [1]. Generally, stereo matching algorithms can be classified into two categories based on the strategies used for the estimation: local and global approaches. Local approaches use some kind of correlation between color or intensity patterns in the neighboring windows. The approaches can easily acquire correct disparity in highly textured regions. However, they often tend to produce noisy results in large untextured region. Moreover, it assumes that all pixels in a matching window have similar disparities resulting in blurred object borders and the removal of small details. Global approaches define energy model which applies various constraints for reducing the uncertainties of the disparity map and solve it through various minimization technique, such as graph cut, belief propagation [2][3].

Recently, many stereo matching algorithms use color segmentation for large untextured regions handling and accurate localization of object boundaries [4][5][6]. The algorithms have the assumption that the disparity vectors vary smoothly inside homogeneous color segments and change abruptly on

B. Gunsel et al. (Eds.): MRCS 2006, LNCS 4105, pp. 761–768, 2006.
© Springer-Verlag Berlin Heidelberg 2006

the segment boundaries. Thus, segment-based stereo matching can produce smooth disparity fields while preserving the discontinuities resulting from the boundaries.

Variational regularization approaches have been increasingly applied to stereo matching method. The regularization method used by B. Horn and B. Schunck introduces the edge-preserving smoothing term to compute the optical flow [7]. In addition, L. Alvarez modified the regularization model to improve the performance of edge-preserving smoothness [8]. In this paper, we propose a segment-based stereo matching method, which yields accurate and dense disparity vector fields by using energy-based regularization with visibility constraint.

2 Disparity Plane Estimation

2.1 Color Segmentation

Our approach is based on the assumption that the disparity vectors vary smoothly inside homogeneous color segments and change abruptly on the segment boundaries. By using this assumption, we can acquire the planar model of the disparity inside each segment [4][5][6]. We strictly enforce disparity continuity inside each segment, therefore it is proper to oversegment the image. In our implementation, we use the algorithm proposed in [9].

2.2 Initial Matching

In a rectified stereo images, the determination of disparity from I_1 to I_2 becomes finding a function $d(x, y)$ such that:

$$I_1(x, y) = I_2(x - d(x, y), y) \tag{1}$$

Initial dense disparity vectors are estimated hierarchically using region-dividing technique [10]. The criterion of determining disparity map is sum-of-absolute-difference (SAD). The region-dividing technique performs stereo matching in the order of feature intensities to simultaneously increase the efficiency of the process and the reliability of the results. In order to reject outliers, we perform a cross-check method for the matching points.

2.3 Robust Plane Fitting

The initial disparity map is used to derive the initial planar equation of each segment [4]. We model the disparity of a segment, i.e. $d(x, y) = ax+by+c$, where $P = (a, b, c)$ are the plane parameters and d is the corresponding disparity of (x, y). (a, b, c) is the least square solution of a linear system [4]. Although we eliminate the outliers by cross-check explained in 2.2, there may be a lot of unreliable valid points in occluded and untextured regions. In order to decrease the effects of outliers, we compute the plane parameters by using $LMedS$ method, which is one of the robust least square methods [11]. Given m valid points in the segment, we select n random subsamples of p valid points. For each subsample indexed by j, we compute the plane parameter $P_j = (a_j, b_j, c_j)$. For each P_j, we can compute the median value of the square residuals, denoted by M_j, with respect to the whole set of valid points. We retain the P_j for which M_j is minimal among all n M_j's.

$$P_S^{\min} = \arg\min_{j=1.2.\cdots n} \left(\underset{i=1,2,\cdots m}{med} |a_j x_i + b_j y_i + c_j|^2 \right), \tag{2}$$

where x_i and y_i are the coordinates of valid points. The number of subsamples m is determined according to the following probability model. For given values of p and outlier's probability ϵ, the probability P_b that at least one of the n subsamples is good is given by, [11]

$$P_b = 1 - [1 - (1 - \epsilon)^p]^n \tag{3}$$

When we assume $\epsilon = 0.3$, $p = 10$ and require $P_b = 0.99$, thus $n = 170$. The estimation is performed in the entire segment in the same manner. In order to enhance the robustness of the algorithm, the segment which has very small reliable valid points is skipped as they do not have sufficient data to provide reliable plane parameter estimation. The plane parameters of the skipped segments are estimated by using the plane parameter of the neighbor segments. Though we estimate the plane parameters through robust $LMedS$ method, there may be still erroneous points due to the error of initial matching. Moreover, the plane model estimation does not consider the occluded part of the segment, especially, in the case for which all parts of the segment is occluded by a foreground object. It is necessary to handle the occluded part in the segment for improvement of the performance. We perform the plane parameter assignment for each segment with the plane parameter of neighbor segment. The cost function for assignment process is given by

$$C(S,P) = \sum_{(x,y)\in S-O} e^{1-\frac{s}{n}} \left| I_1(x,y) - I_2(x - d^P(x,y), y) \right| + \sum_{(x,y)\in O} \lambda_{OCC} \tag{4}$$

$$d^P(x,y) = a^P x + b^P y + c^P,$$

where S is a segment, $P = (a^P, b^P, c^P)$ is a plane parameter, O is an occluded part in the segment, and λ_{OCC} is a constant penalty for occlusion. In order to classify the segment into occluded and non-occluded part, we use a crosscheck method. However, the cross-check method may consider the non-occluded point as occluded in textureless regions. Thus, we perform the cross-check method and determine whether the valid point is occluded or not, only in the vicinity of the segment boundary, because only the vicinity of the segment boundary can be occluded part, in the assumption that a segment is the section of a same object. q is the number of pixels that are non-occluded in the segment and have initial disparity value estimated in 2.2, and s is the number of supporting pixels to a disparity plane P in the non-occluded part of the segment S [5]. Supporting means that the distance between the disparity computed by a plane parameter and initial estimated disparity is smaller than d_{th}(The threshold is set to 1 here.). The cost function is similar to that of [5], however, we use occlusion penalty and segment boundary as occlusion candidate region. The final plane parameter is determined as follows;

$$P_S = \underset{S_i \in N(S)+S}{\arg\min} C(S_i, P_{S_i}) \tag{5}$$

$N(S)$ is the set of neighbor segments, and P_S is the computed plane parameter in the segment S. The assignment process is repeated until the plane parameter does not change in the entire segment. In order to avoid error propagation, all the plane parameters are updated after all the segments are checked in each iteration. Moreover, we only check the segments in which their neighbor segments change in the previous iteration for the reduction of computational load [4]. In the experiment, the process is usually terminated in the $3th$ iterations.

3 Regularization by Color Segmentation and Visibility Constraint

In the disparity plane estimation, we can estimate the reliable and accurate disparity vectors which have good performance in large untextured regions and object boundaries. However, spatial correlation between neighbor segments is not considered. Moreover, detecting and penalizing the occlusion through cross-check method have limitation in the untextured region, and uniqueness constraint is not appropriate when there is correspondence between unequal numbers of pixels. Thus, we propose an energy-based regularization which considers the smoothness between the segments and penalizes the occlusion through visibility constraint.

$$\begin{aligned} E_D(d) = &\int_\Omega c(x,y)(I_l(x,y) - I_r(x+d,y))^2 dxdy \\ &+ \lambda \int_\Omega (\nabla d)^T D_S(\nabla I_{ls})(\nabla d) dxdy \end{aligned} \tag{6}$$

E_D refers to the energy functional of disparity. Ω is an image plane, λ is a weighting factor. ∇I_{ls} is the gradient of I_l, which considers the color segment.

$$\nabla I_{ls}(x,y) = \begin{cases} \nabla I_l(x,y) & if\ (x,y)\ is\ segment\ boundary \\ 0 & otherwise \end{cases} \tag{7}$$

$D_S(\nabla I_{ls})$ is an anisotropic linear operator, which is a regularized projection matrix in the perpendicular aspect of ∇I_{ls}[8]. The operator is based on the segment boundary and can be called by segment-based diffusion operator. An energy model that uses the diffusion operator inhibits blurring of the fields across the segment boundaries of I_1. This model suppresses the smoothing at the segment

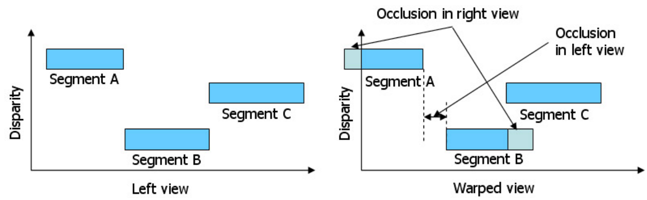

Fig. 1. Occlusion detection with a Z-buffer proposed in [6]

boundaries according to the gradients for both the disparity field and reference image. $c(x, y)$ is an occlusion penalty function and is given by

$$c(x,y) = \frac{1}{1+k(x,y)} \qquad k(x,y) = \begin{cases} 0 & if\ (x,y)\ is\ non-occluded \\ K & otherwise \end{cases} \tag{8}$$

$c(x, y)$ is similar to that proposed in [12], but it is different in the sense that our penalty function uses visibility constraint, and the function in [12] uses uniqueness constraint by cross-check method. In order to detect the occlusion through visibility constraint, we use Z-buffer that represents the second view in the segment domain [6]. Fig. 1 shows the occlusion detection with Z-buffer. By using disparity plane information estimated in 2.3, we warp the reference image to the second view. If a Z-buffer cell contains more than one pixel, only the pixel with the highest disparity is visible and the others are occluded in the second view. Empty Z-buffer cells represent occlusions in the reference image. In our energy function, we penalize the occlusions for the second image, i.e., in the case that is visible in the reference image and not visible in the second image.

The occlusion penalty function $c(x, y)$ is determined by the disparity information which should be estimated. Therefore, we propose the iterative optimization scheme for occlusion and disparity estimation, as shown in Fig 2. We compute occlusion penalty function $c(x, y)$ with Z-buffer, given current disparity information. Then, we perform the disparity regularization process in Eq. (6), and estimate the occluded region with the updated disparity information, iteratively. The minimization of Eq. (6) yields the following associated Euler-Lagrange equation. We obtain the solutions to the Euler-Lagrange equations by calculating the asymptotic state ($t \to \infty$) of the parabolic system.

$$\begin{aligned} \frac{\partial d(x,y)}{\partial t} &= \lambda div(D_S(\nabla I_{ls}(x,y))\nabla d(x,y)) \\ &\quad + c(x,y)(I_l(x,y) - I_r(x+d,y))\frac{\partial I_r(x,y)}{\partial x} \end{aligned} \tag{9}$$

We also discretize Eq. (9) using a finite difference method. All the spatial derivatives are approximated by forward differences. The final solution can be found in a recursive manner.

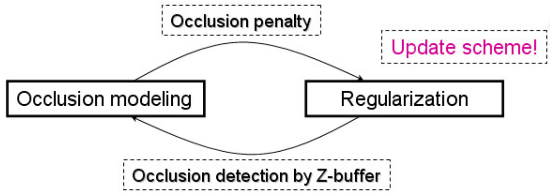

Fig. 2. Iterative optimization scheme

4 Simulation Results

To evaluate the performance of our approach, we used a test bed proposed by Scharstein and Szeliski [1]. We evaluated the proposed algorithm on these test data sets with ground truth disparity maps. The parameters used in the experiment are shown in Table 1.

Fig. 3 shows the results of stereo matching for the standard stereo images provided on Scharstein and Szeliskis homepage. We compared the performance of the proposed algorithm with other algorithms which use energy-based regularization. The results show that the proposed algorithm achieves good performance in conventionally challenging areas such as object boundaries, occluded regions and untextured regions. Especially, in the object boundaries, the proposed algorithm has the good discontinuities localization of disparity map, because it performs the segment-preserving regularization.

For the objective evaluation, we follow the methodology proposed in [1]. The performance of the proposed algorithm is measured by the percentages of bad matching (where the absolute disparity error is greater than 1 pixel). Occluded pixels are excluded from the evaluation. The quantitative comparison in Table 2 presents that the proposed algorithm is superior to other algorithms. In the 'Tsukuba' data, proposed algorithm has relatively high error percentage to graph cut algorithm, because the disparity consists of a planar surface.

Fig. 4 shows the results for new standard stereo images, 'Teddy' and 'Cone' data sets. The results include the occlusion detection of proposed algorithm. These images has very large disparity whose maximum value is 50 pixels. Though the occluded region is very large, proposed algorithm performed the occlusion detection very well. However, the error of occlusion detection in the 'Cone' image is due to an iterative scheme for disparity and occlusion estimation.

Table 1. Parameters used in simulation

Parameter	Values
Weighting factor	λ=50
Constant occlusion penalty	λ_{OCC}=30
Occlusion penalty function	K=100

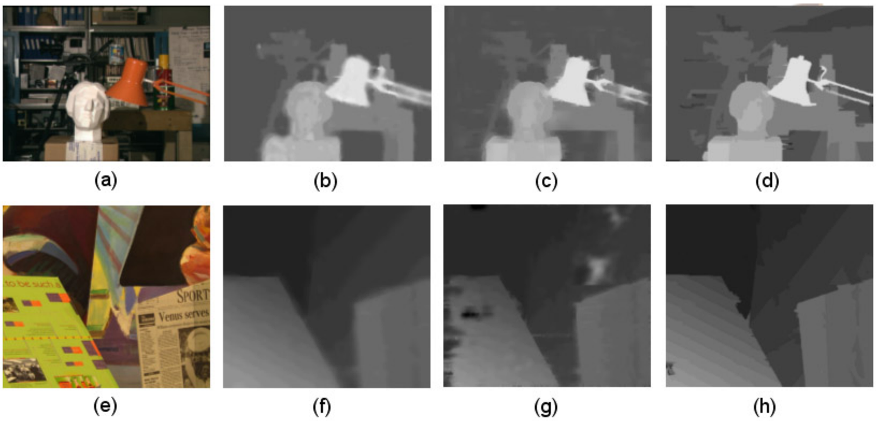

Fig. 3. Results for standard images; (a)(e) Tsukuba, Venus images, (b)(f) [13]'s results, (c)(g) [10]'s results, (d)(h) proposed results

Table 2. Comparative performance of algorithms

	Tsukuba (%)			Venus		
	nonocc	all	disc	nonocc	all	disc
Shao [13]	9.67	11.9	37.1	6.01	7.03	44.2
Hier+Regul[10]	6.17	7.98	28.9	22.1	23.4	43.4
Graph cut[2]	1.94	4.12	9.39	1.79	3.44	8.75
Proposed	3.38	3.83	14.8	1.21	1.74	13.9

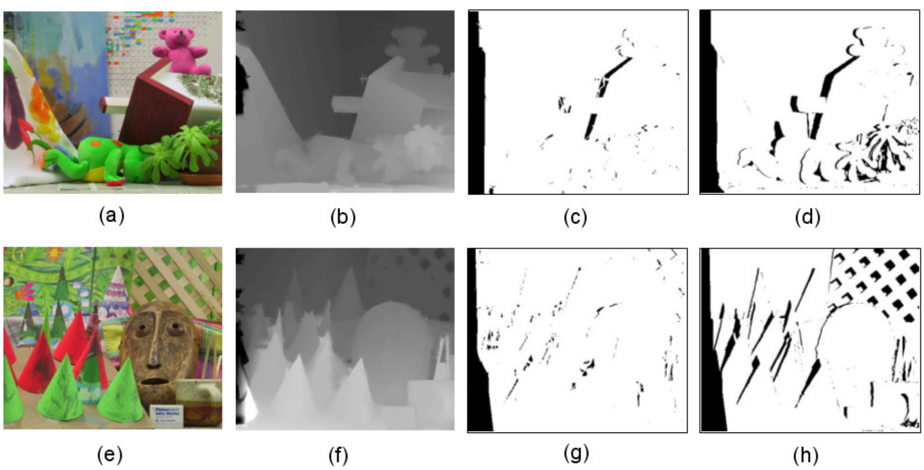

Fig. 4. Results for 'Teddy' and 'Cone' images; (a)(e) original images, (b)(f) disparity maps, (c)(g) occlusion maps, (d)(h) true occlusion maps

5 Conclusion

We proposed a new stereo matching algorithm which uses disparity regularization through segment and visibility constraint. By using the initial disparity vectors, we extracted the plane parameter in each segment through robust plane fitting method. Then, we regularized the disparity vector through segment-preserving regularization with visibility constraint. We confirmed the performance of the algorithm by applying it to several standard stereo image sequences.

Acknowledgement

This work is financially supported by the Ministry of Education and Human Resources Development (MOE), the Ministry of Commerce, Industry and Energy (MOCIE) and the Ministry of Labor(MOLAB) through the fostering project of the Lab of Excellency.

References

1. D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," IJCV, Vol. 47 (2002) 7-42.
2. Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," IEEE Trans. PAMI, Vol. 23 (2001) 1222-1239.
3. J. Sun, N. N. Zheng, and H. Y. Shum, "Stereo matching using belief propagation," IEEE Trans. PAMI, Vol. 25 (2003) 787-800.
4. H. Tao and H. Sawhney, "A global matching framework for stereo computation," Proc. ICCV (2001) 532-539.
5. L. Hong and G. Chen, "Segment-based stereo matching using graph cuts," Proc. IEEE CVPR (2004) 74-81.
6. M. Bleyer and M. Gelautz, "A layered stereo algorithm using image segmentation and global visibility constraints," Proc. IEEE ICIP (2004) 2997-3000.
7. B. Horn, B. Schunck, "Determining optical flow," Artificial Intelligence, Vol. 17 (1981) 185-203.
8. L. Alvarez, R. Deriche, J. Sanchez, and J. Weickert, "Dense Disparity Map Estimation Respecting Image Discontinuities: A PDE and Scale-space Based Approach," J. of VCIR, Vol. 13 (2002) 3-21.
9. C. Christoudias, B. Georgescu, and P. Meer, "Synergism in low-level vision," Proc. IEEE ICPR, Vol. 4 (2002) 150-155.
10. H. Kim, K. Sohn, "Hierarchical disparity estimation with energy-based regularization", Proc. IEEE ICIP, Vol. 1 (2003) 373-376.
11. Z. Zhang, R. Deriche, O. Faugeras, and Q. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," Artificial Intelligence, Vol. 78 (1995) 87-119.
12. C. Strecha, T. Tuytelaars, L.Van Gool, "Dense matching of multiple wide-baseline views," Proc. ICCV (2003) 1194-1201.
13. J. Shao, "Generation of Temporally Consistent Multiple Virtual Camera Views from stereoscopic image sequences," IJCV, Vol.47 (2002) 171-180.