IMPROVING SEGMENT BASED STEREO MATCHING USING SURF KEY POINTS

Gorkem Saygili, Laurens van der Maaten, Emile A. Hendriks

Department of EEMCS, Delft University of Technology

ABSTRACT

State-of-the-art stereo matching algorithms estimate disparities using local block-matching, and subsequently refine the disparity estimates by introducing smoothness constraints and performing global energy minimization. Such algorithms are hampered by the inability of local block-matching algorithms to deal with repetitive patterns. This paper presents an approach that overcomes this problem by incorporating the disparity obtained from matching SURF key points between stereo image pairs. The algorithm provides further robustness to problems with repetitive pattern by penalizing the discrepancy between the initial and final disparity estimates in the global energy minimization. Evaluation of our approach on the Middleburry data set results shows that the our approach is more robust against repetitive patterns than existing approaches.

Index Terms— Stereo Matching, Global Energy Minimization, SURF key points

1. INTRODUCTION

The goal of stereo matching is to find an estimate of the depth information inside a scene. This depth estimate can be used for, among others, 3D image reconstruction, virtual view rendering, and 3D object classification.

State-of-the-art stereo matching algorithms (see [1, 2] for an overview) are based on locally matching small image patches to determine the disparity at each image location [3, 4]. In many algorithms, the results of the local matching algorithm are refined by incorporating smoothness constraints that suppress noisy disparity estimates [5, 6, 7, 8, 9]. Although these algorithms work well in many cases, they have severe problems in the presence of repetitive patterns (see Fig. 1): when a repetitive pattern is present, the disparity estimates of the neighbouring patches may be widely varying. Employing smoothness constraints is generally insufficient to resolve the *repetitive pattern problem*. To our knowledge, there is no study that specifically aims to solve this problem for state-of-the art segment based methods.

In this paper, we introduce a novel stereo disparity estimation algorithm that tries to resolve the repetitive pattern problem by employing SURF key points to refine local matching algorithms. Incorporation of matching key points between rectified stereo pairs provides robustness against repetitive patterns by restricting the search space of the local matching. To further reduce the problems with repetitive patterns, we propose a global energy minimization formulation that penalizes large discrepancies between the initial and the refined disparity estimates. As a result, any improvement in the initial disparity estimation has a direct influence on the final disparity estimation.

Our approach comprises four main stages which will be further discussed in Section 2. Section 3 presents the experimental evaluation of our approach. We draw our conclusions in Section 4.

2. DISPARITY ESTIMATION

Similar to other state-of-the-art stereo matching approaches, our approach comprises four main stages: (1) colour segmentation, (2) initial disparity estimation with SURF key points, (3) plane fitting, and (4) disparity plane assignment using graph cuts. These four stages are described separately below.

Colour Segmentation. The reference image is segmented into non-overlapping homogeneous color regions using mean-shift segmentation [10] resulting in a set of segments T. The main assumption of segmenting the image is that disparity discontinuities can only occur at segment boundaries. Therefore, the disparity within a segment can be modelled by a planar surface in later stages of our approach.

Initial Disparity Estimation with SURF Key Points. In this work, we propose to incorporate information obtained by key points into the initial estimation of the disparity. Our approach decreases the noise that is caused by repetitions of pattern using a restricted disparity search space, since it reduces the tendency of many algorithms to estimate the disparities wrongly. The details of the algorithm are as follows: Since the images are already rectified for stereo-matching, the matching key points should lie on the same epipolar line. Therefore, the vertical positions of key points should satisfy:

$$\forall s \in S : |y_{sL} - y_{sR}| < 0.5, \tag{1}$$

where S is the set of matched key points, and y_{sL} and y_{sR} are the vertical positions of those points in left and right views, respectively. The information on the disparity of key points and on the segments in which the key points are located can be used to obtain a new lower and upper bound, $d_{t,low}$ and

$d_{t,high}$, for the disparity search range of the pixels inside that segment rather than manually set d_{min} and d_{max}:

$$\forall t \in T, \forall (x,y) \in t : d_{t,low} \leq d(x,y) \leq d_{t,high}. \quad (2)$$

where $d_{t,low}$ and $d_{t,high}$ are given by:

$$d_{t,low} = \begin{cases} \max\{\lfloor \theta_1 - \theta_2 \rfloor, d_{min}\} & \text{if } K_t \neq \emptyset \\ d_{min} & \text{otherwise,} \end{cases} \quad (3)$$

$$d_{t,high} = \begin{cases} \min\{\lceil \theta_3 - \theta_2 \rceil, d_{max}\} & \text{if } K_t \neq \emptyset \\ d_{max} & \text{otherwise,} \end{cases} \quad (4)$$

where, θ_1, θ_2 and θ_3 are given by:

$$\begin{aligned} &\theta_1 = \min\{\forall (x,y) \in K_t : |x_L - x_R|\}, \\ &\theta_2 = \alpha \times (d_{max} - d_{min}), \\ &\theta_3 = \max\{\forall (x,y) \in K_t : |x_L - x_R|\}, \\ &K_t : \forall (x,y) \in t \cap S. \end{aligned} \quad (5)$$

Herein, $d(x,y)$ is the disparity of the pixel at location (x,y) and α is a scaling coefficient that ranges between 0 and 1. Since the disparity search space is bounded by $d_{t,low}$ and $d_{t,high}$ rather than by d_{min} and d_{max}, the algorithm does not consider disparities corresponding to the repetitions as a result of which the correct disparity is more likely to be found by local matching.

Local pixel matching is based on a matching cost function and an aggregation window around the pixel of interest. As matching costs, we used the sum of absolute differences between images, and between their image gradients:

$$C_I(x,y,d) = |I_L(x,y) - I_R(x+d,y)| \ , \quad (6)$$

$$\begin{aligned} C_\bigtriangledown(x,y,d) &= |\bigtriangledown_x I_L(x,y) - \bigtriangledown_x I_R(x+d,y)| \\ &+ |\bigtriangledown_y I_L(x,y) - \bigtriangledown_y I_R(x+d,y)| \ , \end{aligned} \quad (7)$$

$$d^L(x,y) = \text{argmin}_d\Big(\sum_{\forall x,y \in t} C_I(x,y,d) + C_\bigtriangledown(x,y,d)\Big).$$

The cost of a pixel inside the box is aggregated if and only if that pixel resides in the same segment as the center pixel. The non-occluded pixels are used in the global energy minimization as trustful pixels in terms of disparity and similar to [7], these pixels are found by using cross-check validation. Figure 1 illustrates the enhancement in quality by using SURF key points for initial matching. The red circle indicates a region in which there is repetition of pattern.

Plane Fitting. Based on the initial disparity estimate, we model each segment by a plane and estimate the parameters of this plane using RANSAC. Since RANSAC works best when there are at least 50 percent of inliers and because large regions provide larger clusters of reliable disparities than smaller regions, we opt to apply the RANSAC to segments that contain more than 100 pixels, and of which at least 50 percent of the pixels are non-occluded. The planes that have similar surface normals and mean disparities are eliminated heuristically, which results in a small set of planes that are sufficient to represent the scene.

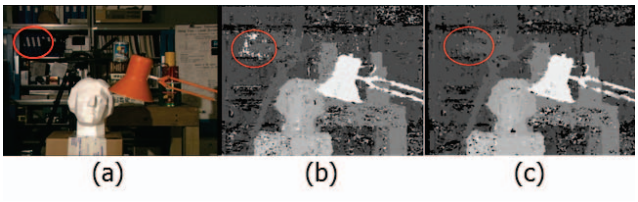

Fig. 1. The initial disparity result; (a) reference image (repetition of pattern encircled), (b) without SURF, (c) with SURF.

Disparity Plane Assignment Using Graph Cuts. A disparity plane is assigned to each image segment by minimizing an energy function that incorporates both data costs and smoothness constraints. The energy minimization problem is solved using a graph-cut approach in which each node corresponds to a segment. Let P be the set of disparity plane parameter labels. Our aim is to find a labelling f that assigns each segment $t \in T$ to its plane label $p \in P$ by minimizing the following energy function:

$$E(f) = E_{data}(f) + E_{smooth}(f), \quad (8)$$

where $E_{data}(f)$ is the cost of assigning plane labels to the segments.

In most of the state-of-the-art algorithms, such as [6, 7], the matching cost of Eq. 6 and Eq. 7 is used as the data term. In this work, we propose to use the following Disparity Estimate Discrepancy Cost (DEDC) instead:

$$E_{data}(f) = \sum_t \sum_{(x,y)} \lambda |d^{f(t)}(x,y) - d(x,y)| e^{-n/m}, \quad (9)$$
$$\forall t \in T, \quad \forall (x,y) \in t - O_t,$$

in which O_t is the set of occluded pixels in t, n is the number of non-occluded pixels that have the same initial disparity as the disparity after plane fitting, m is the number of non-occluded pixels inside the segment, λ is the scaling coefficient, and $d^{f(t)}$ represents the disparity of the pixel (x,y) after fitting a plane with label $f(t)$ on pixels for segment t:

$$d^{f(t)}(x,y) = a_{f(t)}x + b_{f(t)}y + c_{f(t)}. \quad (10)$$

By penalizing large discrepancies between the initial and the final disparity estimates, $E_{data}(f)$ favors solutions that are close to the initial estimate. As a result, the final estimate will not "jump" to the wrong part of repetitive patterns, which is what traditional data costs would do. $E_{smooth}(f)$ is a smoothness term that penalizes the discontinuities in plane labels of neighbouring segments. We define $E_{smooth}(f)$ as:

$$E_{smooth}(f) = \sum_t \sum_q \gamma(t,q)(1 - \delta(f(t), f(q))), \quad (11)$$
$$\forall t \in T, \quad \forall q \in N(t).$$

Herein, $N(t)$ is the set of neighbors of t and $\gamma(t,q)$ is:

$$\gamma(t,q) = w\beta e^{(-\tau^2/\sigma^2)}, \quad (12)$$

where w and σ are scaling parameters, β and τ are the boundary length and the mean colour difference between t and q, respectively.

3. EXPERIMENTS

Experimental Setup. To evaluate the performance of our algorithm, we performed experiments on the Middleburry benchmark data set. We evaluate the algorithm by measuring the percentage of pixels that have erroneous disparity values. Herein, a disparity value is defined to be erroneous if the absolute difference from ground truth is larger than 1. As in common practice in the evaluation of stereo algorithm, we look at results for (1) non-occluded pixels only (nonocc), (2) all pixels (all), and (3) pixels in image regions that are close to a disparity discontinuity (disc). In all experiments, we set α in Eq. 5 equal to 0.25, α in Eq. 10 equal to 10 and w = 25, σ = 150. Additionally we choose two mean-shift parameters, h_s and h_r as 5 and 4 respectively.

Experimental Results. In order to show the effect of using key points (KP) and DEDC, all four possible variants of our algorithm on the Tsukuba scene are evaluated. Table 1 presents the quantitative results and Fig. 2 shows the corresponding disparity images. The experimental results of the algorithm without KP and DEDC illustrate the problems of current stereo-matching algorithms with repetitive patterns. In particular, the disturbance of repetitions of patterns is clearly recognizable in the final disparity image in Fig. 2. When key points are used, the quantitative results get better because the disturbance of repetition of pattern is suppressed. When both KP and DEDC are incorporated, the best performance is obtained. These results illustrate the ability of our approach to deal with the repetitive pattern problem. Fig. 3 shows the performance of the best variant of our algorithm (KP+DEDC) on all four Middleburry scenes and Table 2 compares the performance of our algorithm with four state-of-the-art disparity matching algorithms. The results in Figure 3 and Table 2 indicate that our algorithm performs on par with the state-of-the-art on the first three scene; and that it outperforms all other algorithms on the Venus scene.

4. CONCLUSION

In this paper, we presented a novel stereo disparity estimation algorithm with two main contributions. The results indicate that the proposed algorithm perform on par with the state-of-the-art algorithms on all four scenes of Middleburry and that it outperforms all state-of-the art algorithms on the Venus scene.

Table 1. Percentage of erroneous disparity values of the disparity estimations for Tsukuba scene.

Algorithm	nonocc	all	disc
baseline	2.64	3.26	11.8
KP	1.56	2.23	7.42
DEDC	1.25	1.75	6.28
DEDC and KP	1.08	1.59	5.82

ACKNOWLEDGEMENTS

We would like to thank Dr David Tax for for his proof-reading of this paper and for his valuable contributions.

5. REFERENCES

[1] D. Scharstein and R. Szelinski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *IJCV*, vol. 47, pp. 7-42, 2003.

[2] D. Scharstein and R. Szelinski, "Middleburry Stereo Vision Page," *http://vision.middleburry.edu/stereo/eval*.

[3] R. Zambih and J. Woodfill, "Non-parametric Local Transforms for Computing Visual Correspondence," *ECCV*, vol.2, pp. 151-158, 1994.

[4] Y. Boykov, O. Veksler, R. Zambih, "A Variable Window Approach for Early Vision," *TPAMI*, vol.20, pp. 1283-1294, 1998.

[5] Z. Wang and Z. Zheng, "A Region Based Stereo Matching Algorithm Using Cooperative Region Optimization," *CVPR*, 2008.

[6] A. Klaus, M. Sourmann and K. Karner, "Segment Based Stereo Matching Using Belief Propagation and Self Adapting Dissimilarity Measure," *ICPR*, pp. 15-18, 2006.

[7] L. Hong and G. Chen, "Segment Based Stereo Matching Using Graph Cuts," *CVPR*, vol. 1, pp. 74-81, 2004.

[8] M. Tappen and W. Freeman, "Comparison of Graph Cuts with Belief Propagation for Stereo," *ICCV*, vol. 1, pp. 508-515, 2003.

[9] Q. Yang, L. Wang, R. Yang, H. Stewenius and D. Nister. "Stereo Matching with Color-Weighted Correlation, Hierarchical Belief Propagation, and Occlusion Handling," *TPAMI*, vol. 3, pp. 492-504, 2009.

[10] D. Comaniciu and P. Meer, "Mean-Shift: A Robust Approach Toward Feature Space Analysis," *IEEE PAMI*, vol. 5, pp. 603-619, 2002.

[11] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang and X. Zhang, "On Building an Accurate Stereo Matching System on Graphics Hardware," *GPUCV*, 2011.

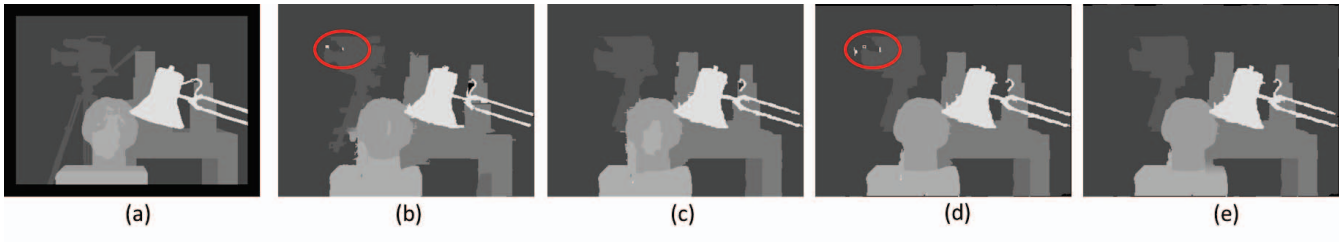

Fig. 2. Estimated final disparities: (a) Ground truth, (b) without KP and without DEDC, (c) with only KP, (d) with only DEDC, (e) with DEDC and KP.

Table 2. Percentage of erroneous disparity values of proposed algorithm with top performing algorithms.

Algorithm	Avg. Rank	Tsukuba			Venus			Teddy			Cones		
		nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
Proposed	24.2	1.08	1.59	5.82	0.08	0.16	1.11	4.49	8.06	12.2	3.59	9.4	11.0
ADCensus [11]	5.8	1.07	1.48	5.73	0.09	0.25	1.15	4.1	6.22	10.9	2.42	7.25	6.95
AdaptingBP [6]	7.2	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 [5]	7.2	0.87	1.16	4.61	0.11	0.21	1.54	5.16	8.31	13.0	2.79	7.18	8.01
DoubleBP [9]	9.7	0.88	1.29	4.76	0.13	0.45	1.87	3.53	8.3	9.63	2.9	8.78	7.79

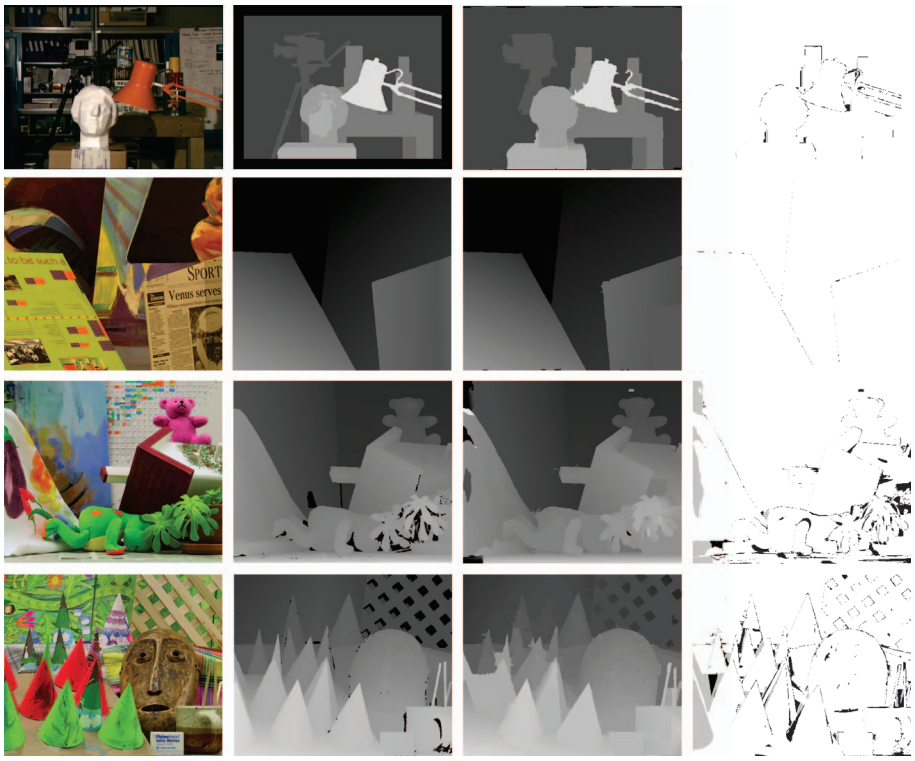

Fig. 3. Results on Middleburry scenes. From top to bottom: the Tsukuba, Venus, Teddy and Cones scenes. From left to right: reference images, ground truth disparities, the results of the proposed algorithm and the error images where the black regions represents the erroneous pixels.