PDE-BASED DISPARITY ESTIMATION WITH OCCLUSION AND TEXTURE HANDLING FOR ACCURATE DEPTH RECOVERY FROM A STEREO IMAGE PAIR

Hansung Kim and Adrian Hilton

Centre for Vision, Speech and Signal Processing, University of Surrey, GU2 7XH, UK
{h.kim, a.hilton}@surrey.ac.uk

ABSTRACT

This paper presents a novel PDE-based method for floating-point disparity estimation which produces smooth disparity fields with sharp object boundaries for surface reconstruction. In order to avoid the over-segmentation problem of image-driven structure tensor and the blurred boundary problem of field-driven tensor, we propose a new anisotropic diffusivity function controlled by image and disparity gradients. We also embed a bi-directional disparity matching term to control the data term in occluded regions. We evaluate the proposed method on data sets from the Middlebury benchmarking site and real data sets with ground-truth models scanned by a LIDAR sensor.

***Index Terms*—**Disparity estimation, Variational method, Surface reconstruction, Occlusion handling

1. INTRODUCTION

Disparity estimation is one of the most important steps in image-based 3D reconstruction. Scharstein and Szeliski discussed the taxonomy of existing stereo algorithms and created a benchmarking site for the quantitative evaluation of algorithms[1]. However, most current disparity estimation algorithms solve the correspondence problem on a discrete domain such as integer, half- or quad-pixel levels which are not sufficient to recover a smooth surface. Figure 1 shows depth errors for a 0.5 pixel disparity error inherent in integer disparity fields. For example, a 0.5 pixel error for the point at a distance of 5m leads to a depth error of 6cm, and the point at 8m leads to a 15cm depth error when the image resolution is 1600x1200 and the baseline distance is 20cm.

A variational approach which theoretically works on a continuous domain and is popular in optical flow estimation can be a solution for accurate floating-point disparity estimation. In this approach, the disparity vector fields are extracted by minimizing an energy functional involving a fidelity term and a smoothing term such as:

$$\begin{aligned} E(d) &= E_f(d) + E_s(d) \\ &= \int_\Omega (I_l(x) - I_r(x+d))^2 dx + \lambda \int_\Omega \psi(\nabla d, \nabla I_l) dx \end{aligned} \quad (1)$$

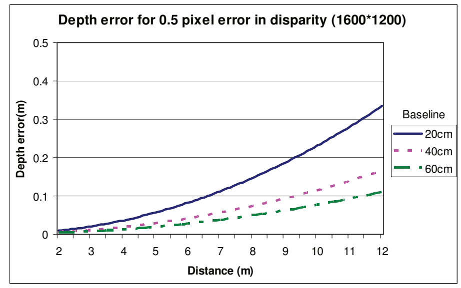


Fig. 1. Behaviour of depth errors according to distance

where $x \in \Omega$ is an open bounded set of R^2 and d is a disparity vector. The minimization problem can be solved by solving the associated partial differential equation (PDE) equation [2][3]. This method produces accurate depth fields across most regions, but it has several limitations related to occlusion around depth discontinuities and highly textured regions.

Traditional image-driven regularisation methods preserve sharp object boundaries but sometimes result in over-segmentation in highly textured regions. Slesareva et al. [4] and Zimmer et al. [5] proposed disparity-driven methods with total variation and anisotropic tensor, respectively, to avoid this over-segmentation method. However, disparity-driven methods tend to blur object boundaries. Sun et al. recently proposed joint image-/flow-driven optical flow [6] based on steerable random fields [7] to obtain sharp object boundaries without over-segmentation.

In terms of the occlusion problem, many researchers have designed energy functionals which handle occlusion regions and converge to optimal minimum solution. Ben-Ari and Sochen proposed an iterative method which repeats occlusion detection by level set, disparity estimation for visible regions and filling occluded region with anisotropic filtering [8]. Alvarez et al. proposed a symmetrical dense optical flow energy functional which includes a bi-directional disparity checking term [9]. Ince and Konrad also proposed similar bi-directional disparity checking method, but they put it into the data term [10].

In this paper, we propose a new energy functional which handles occlusions and over-segmentation while preserving sharp object boundaries. The diffusivity function for regularisation is mainly controlled by image gradient but scaled by disparity gradient to avoid over-segmentation in highly textured region. This is simpler and more intuitive than Sun et al.'s method [6]. The data term is controlled by bi-directional disparity matching. This is similar to Ince's work, but we use it as a switch to turn on and off the data term in the regularisation process.

2. PROPOSED METHOD

If we set the gradient of the potential function $\psi(\nabla d, \nabla I_l)$ in Eq. (1) as Eq. (2), the minimization problem can be solved by the associated Euler-Lagrange equation in Eq. (3) with Neumann boundary conditions.

$$\nabla(\psi(\nabla d, \nabla I_l)) = g(\nabla I_l, \nabla d)\nabla d \tag{2}$$

$$\begin{aligned} -\nabla E(d) = \lambda div(g(\nabla I_l, \nabla d)\nabla d) \\ + (I_l(x) - I_r(x+d))\frac{\partial I_r(x+d)}{\partial d} = 0 \end{aligned} \tag{3}$$

We obtain the solution of Eq. (3) by calculating the asymptotic state ($t \longrightarrow \infty$) of PDE:

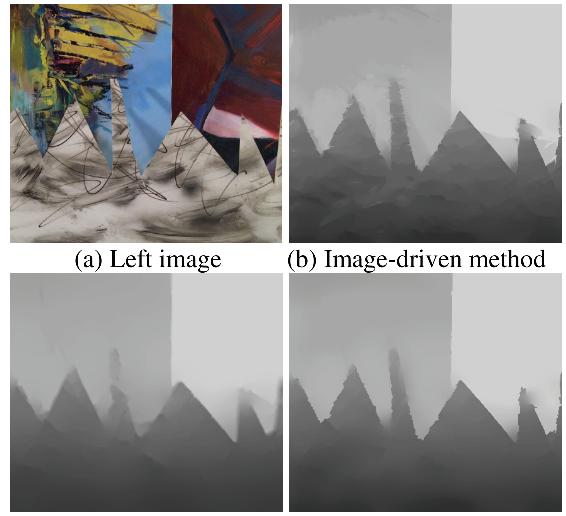
$$\begin{aligned} \frac{\partial d}{\partial t} = \lambda div(g(\nabla I_l(x), \nabla d(x))\nabla d(x)) \\ + (I_l(x) - I_r(x + d(x)))\frac{\partial I_r(x+d(x))}{\partial x} \end{aligned} \tag{4}$$

This PDE corresponds to the nonlinear diffusion equation with an additional reaction term [11], and $g(\bullet)$ is a diffusivity function which controls the direction and amount of diffusion filtering. We are interested in smoothing disparity fields for continuous surfaces while preserving sharp depth discontinuities. Therefore we design the diffusivity function as follows:

$$\begin{aligned} g(\nabla I, \nabla d) = \frac{1}{(1 + s(\nabla I, \nabla d))^2} \cdot K_\rho * (\nabla I_\sigma \ \nabla I_\sigma{}^T) \\ s(\nabla I, \nabla d) = -\ln(0.35 + 0.65 e^{-|\nabla d|}) \cdot |\nabla I|^2 \end{aligned} \tag{5}$$

where $I_\sigma = K_\sigma * I$, K_σ denote a Gaussian kernel with standard deviation σ and $*$ is a convolution operator. The term $K_\rho * (\nabla I_\sigma \ \nabla I_\sigma{}^T)$ works as a structure tensor for anisotropic diffusion filtering [11]. The first term $-\ln(0.35 + 0.65 e^{-|\nabla d|})$ in $s(\nabla I, \nabla d)$ is a monotonically increasing function and scales the diffusivity according to the gradient of the disparity field. As a result, the filtering direction is decided by the image gradient, and the amount of smoothing is decided by both image and disparity gradients.

Figure 2 shows an example of the behaviours of diffusivity functions. We can see that the result with image-driven method is not regularised enough because of strong textures. On the other hand, the disparity-driven method produces very smooth surfaces, but we can see leaks of the fields at object boundaries. Compared with the other two methods, the proposed method produces very smooth field on each plane while keeping sharp object boundaries.



(a) Left image (b) Image-driven method

(c) Disparity-driven method (d) proposed method

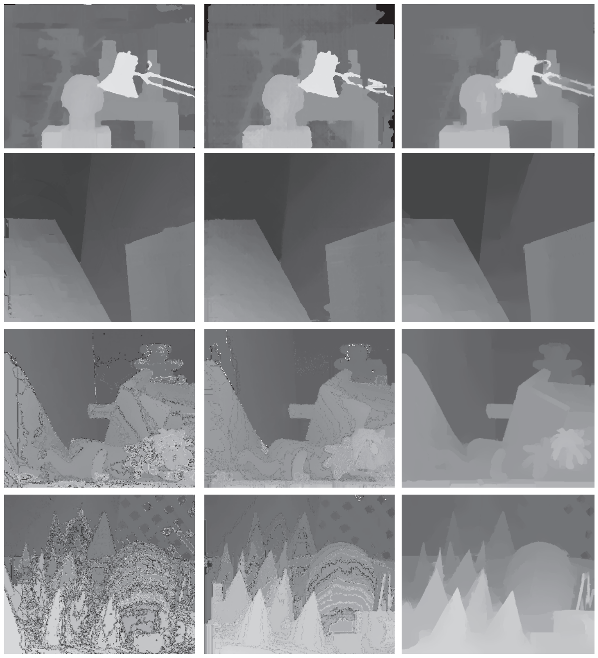
Fig. 2. Disparity fields produced by various methods

For occlusion handling, we take a similar approach to the Ince and Konrad's work [10]. They used a bi-directional disparity matching as a scaling factor for the data term in Eq. (4), but we use it as a switch to turn on/off the data term in regularisation as follows:

$$\begin{aligned} \frac{\partial d_l}{\partial t} = \lambda \ div\big(g(\nabla I_l, \nabla d_l)\nabla d_l(x,y)\big) \\ + H(1 - O(x))(I_l - I_r(x + d_l))\frac{\partial I_r(x + d_l)}{\partial x} \\ O(x) = \big|d_l(x) + d_r(x + d_l(x))\big| \end{aligned} \tag{6}$$

where $H(\bullet)$ is a unit step function. In visible regions, normal balanced diffusion equation with data term is run to find the optimised solution. However, the data term distorts the field in occlusion region because there is no real correspondence for the point. Therefore, in occlusion region, only pure anisotropic diffusion filtering is performed to smooth disparity field and propagate correct depth information from visible regions to occlusion regions.

Equation (6) can be solved explicitly or semi-implicitly with iterative method by updating the timely discretised field [12]. As the Ince's method could not guarantee the mathematical convergence of an iterative solver with the bi-directional disparity matching term, we cannot provide the proof of convergence of Eq. (6) in the iterative numerical solver because the switch can make the solver stuck and resonant between visible and occlusion modes. Therefore,



(a) Cost Aggr. (b) Semi-global (c) Proposed

Fig.3. Estimated disparity maps (Tsukuba, Venus, Teddy and Cones from the top)

we set a maximum number of iteration in the solver to avoid being trapped into an eternal loop. However, we have not experienced this problem in our experiments.

Finally, the local minimum problem is one of the most serious problems in PDE-based methods. Alvarez et al. used a scale-space approach [3] and Brox et al. used a warping method [13] to avoid the local minimum problem for large displacements. We also use a similar coarse-to-fine pyramidal structure which starts from the lowest resolution images and recursively refines the result at higher levels. This hierarchical approach has another merit of reducing computation time for large images.

3. EXPERIMENTAL RESULTS

3.1. Evaluation with benchmarking data sets

In order to verify the performance of the estimated disparity field, we used the Middlebury stereo benchmarking data sets [1]. Figure 3 shows the estimated disparity maps and comparison with state-of-the-art algorithms. Cost-Aggr [14] is an energy-based method with asymmetric occlusion detection and Semi-Global [15] is a pixel-wise mutual information-based matching with global constraints. The proposed method produced clean maps with sharp object boundaries over all images even in occlusion regions compared with other algorithms.

However, the proposed algorithm is ranked at 47th among 78 algorithms in Bad Pixel Percentage (BPP) test which is used as a standard measure in the benchmarking site, while the Cost-Aggr and Semi-Global algorithms which produced visually worse results are ranked 27th and 33th, respectively. The BPP test doesn't care about the magnitude of errors but just count the number of pixels which exceed error threshold. PDE-based methods tend to spread errors into neighbours to suppress salient errors, and it caused the low ranks in the BPP test despite good subjective performance.

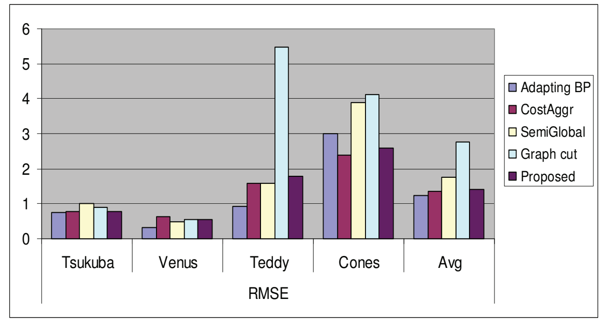
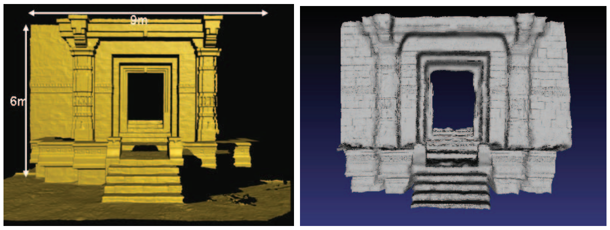


Figure 4. Comparison in RMSE

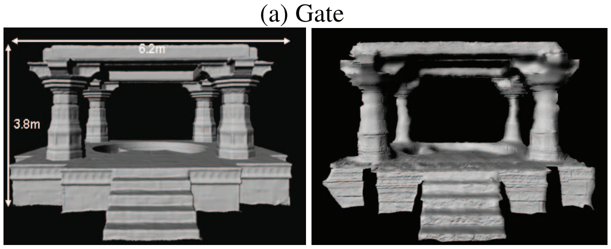
We also measured Root Mean Square Error (RMSE) to the ground truth and compared with the results from other algorithms. Figure 4 shows the result where Adapting-BP [16] is a method using belief propagation which is ranked at the top in the BPP test and Graph-cut is an alpha-expansion method with occlusion handling [17] which is ranked at 42th. All algorithms compared in this test are ranked higher than the proposed algorithm in the BPP test. We can see that the proposed method shows much better performance compared with Semi-global matching and Graph-cut, and compete with Adapting-BP and Cost-Aggregation methods in the RMSE test.

3.2. Evaluation in 3D reconstruction

We reconstruct depth fields with camera parameters from a stereo image pair captured with a line scan camera with a fisheye lens, and compared their depth with ground-truth range data scanned by a LIDAR sensor. The test stereo image pairs were captured with a baseline of 60cm, and have a resolution of 3200x5312 and a maximum disparity of 280 pixels. Figure 5 shows the ground-truth models by LIDAR scan and the reconstructed models from the proposed algorithm. We can see that the reconstructed model shows very fine structure with smooth surface and sharp edges. The surface relief pattern with features at any scale is recovered. We mentioned in the introduction that conventional disparity estimation algorithms working on discrete domain is not enough for reconstructing details. Figure 6 shows the difference in surface reconstructions from integer and floating-point disparity fields. Both disparity fields have the same BPP error but the disparity fields are calculated in integer values in Fig. 6 (a). We can see that all surface details have disappeared and it shows stepwise artefact in

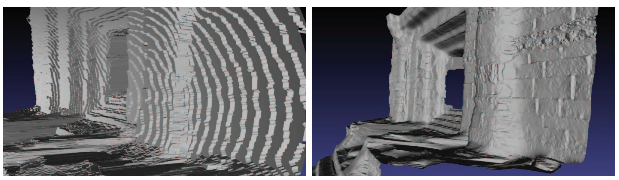


(a) Gate



(b) Coopolar

Fig. 5. Ground-truth by LIDAR and Reconstructed model



(a) Integer disparity (b) Floating-point disparity

Fig. 6. Precision in surface reconstruction

Fig. 6 (a). This artefact is more obvious because the original image pair has serious radial distortions from the fisheye lenses.

For objective evaluation, we produced depth maps from the same viewpoint for the ground-truth and reconstructed models and measured average depth error from the viewpoints. The average depth error over the whole common area was -0.20cm with 15.2cm standard deviation for the gate scene, and -5.20cm of average depth error with 8.97cm standard deviation for the coopolar scene.

4. CONCLUSION

In this paper, we have proposed a PDE-based disparity estimation method with occlusion and texture handling for accurate and smooth surface reconstruction from a stereo image pair. We presented an anisotropic diffusivity function for regularisation which is controlled by both image and disparity gradient to generate smooth depth fields even in highly textured regions while preserving sharp depth discontinuities. In order to fill the occluded regions by pure anisotropic diffusion filtering, we embedded a bi-directional disparity matching term to control the data term. The experimental results show that the proposed PDE produces very accurate and spatially correlated disparity fields in both objective and subjective evaluations. The performance of the proposed system was also evaluated in surface reconstruction against ground-truth from a LIDAR scan.

5. ACKNOWLEDGEMENT

This research was executed with the financial support of the EU IST FP7 project i3Dpost.

6. REFERENCES

[1] D. Scharstein and R. Szeliski, “A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms,” *IJCV*, vol.47, no.1, pp.7-42, 2002.

[2] H. Kim and K. Sohn, "Hierarchical disparity estimation with energy-based regularization," Proc. ICIP, pp.373-376, Sep. 2003.

[3] L. Alvarez, R. Deriche, J. Sánchez and J. Weickert, "Dense Disparity Map Estimation Respecting Image Discontinuities: A PDE and Scale-Space Based Approach," Journal of Visual Comm. and Image Repr., vol.13, no.1, pp.3-21, 2002.

[4] N. Slesareva, A. Bruhn and J. Weickert, "Optic flow goes stereo: A variational method for estimating discontinuity-preserving dense disparity maps" Proc. DAGM, pp. 33-40, 2005.

[5] H.Zimmer, A. Bruhn, L.Valgaerts, M. Breuß, J. Weickert, B. Rosenhahn, and H.-P. Seidel, “PDE-based anisotropic disparity-driven stereo vision,” Proc. VMV, pp.263-272, 2008.

[6] D. Sun, S. Roth, J. Lewis and M. Black, “Learning optical flow,” Proc. ECCV, pp.83–97, 2008.

[7] S. Roth and M. Black, “Steerable random fields,” Proc. ICCV, 2007.

[8] R. Ben-Ari and N. Sochen, “Variational Stereo Vision with Sharp Discontinuities and Occlusion Handling”, In Proc. ICCV, pp. 1-7, 2007.

[9] L.Alvarez, R.Deriche, T.Papadopoulo and J.Sánchez "Symmetrical Dense Optical Flow Estimation with Oclussions Detection", Proc. ECCV, pp 721-736, 2002.

[10] S. Ince and J. Konrad "Occlusion-aware optical flow estimation", IEEE Trans. Image Process., vol.17, pp. 1443-1451, 2008.

[11] J. Weickert, “A Review of Nonlinear Diffusion Filtering,” LNCS, vol.1252, pp.3-28, 1997.

[12] C. Johnson, *Numerical Solution of Partial Differential Equations by the Finite Element Method*, Cambridge University Press, 1988.

[13] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High Accuracy Optical Flow Estimation Based on a Theory for Warping," Proc. ECCV, pp.25-36, 2004.

[14] D. Min and K. Sohn, “Cost aggregation and occlusion handling with WLS in stereo matching,” IEEE Trans. Image Processing, 2008.

[15] H. Hirschmüller, “Accurate and efficient stereo processing by semi-global matching and mutual information,” IEEE Trans. PAMI vol. 30, no.2, pp.328-341, 2008.

[16] A. Klaus, M. Sormann and K. Karner, “Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure,” Proc. ICPR, 2006.

[17] V. Kolmogorov and R. Zabih, “Computing visual correspondence with occlusions using graph cuts,” Proc. ICCV, 2001.