Optic Flow Goes Stereo: A Variational Method for Estimating Discontinuity-Preserving Dense Disparity Maps

Natalia Slesareva, Andrés Bruhn, and Joachim Weickert

Mathematical Image Analysis Group,
Faculty of Mathematics and Computer Science,
Saarland University, Building 27.1, 66041 Saarbrücken, Germany
{Slesareva, Bruhn, Weickert}@mia.uni-saarland.de

Abstract. We present a novel variational method for estimating dense disparity maps from stereo images. It integrates the epipolar constraint into the currently most accurate optic flow method (Brox *et al.* 2004). In this way, a new approach is obtained that offers several advantages compared to existing variational methods: (i) It preservers discontinuities very well due to the use of the total variation as solution-driven regulariser. (ii) It performs favourably under noise since it uses a robust function to penalise deviations from the data constraints. (iii) Its minimisation via a coarse-to-fine strategy can be theoretically justified. Experiments with both synthetic and real-world data show the excellent performance and the noise robustness of our approach.

Keywords: computer vision, variational methods, stereo reconstruction, differential techniques, partial differential equations.

1 Introduction

The reconstruction of 3-D information from two views is one of the key problems in computer vision. Since the prototypical approach of Marr and Poggio [11] three decades ago, a variety of algorithms have been proposed for this purpose. Depending on their strategy for solving the correspondence problem, these algorithms can be divided into four classes: *Feature-based* algorithms [3,6] that make use of characteristic image features such as corners or lines, *area-based* methods [8,15,16] that correlate image patches by aggregating local similarity measures, *phase-based* approaches [4,5,7] that estimate displacements via the phase in the Fourier domain, and *energy-based* techniques [1,9,10,12,14,15] that seek to minimise variational formulations, where deviations from data and smoothness constraints are penalised.

Variational methods offer one decisive advantage when compared to the other three strategies: They allow for an estimation of correspondences at those locations where no image information is available. Since they regularise the often non-unique solution of their data constraints by assuming (piecewise) smoothness of the result, neighbourhood information is propagated to locations where such information is missing. As a consequence, always 100% dense estimates are obtained. This so-called *filling-in effect* is one reason why variational techniques have become increasingly popular during the last few years.

W. Kropatsch, R. Sablatnig, and A. Hanbury (Eds.): DAGM 2005, LNCS 3663, pp. 33–40, 2005.
© Springer-Verlag Berlin Heidelberg 2005

A second reason is the fact that this regularisation can be adapted such that it respects discontinuities in the data (*image-driven*) or the solution (*solution-driven*). Thus, objects boundaries are preserved and the accuracy of the reconstruction increases. First approaches with image-driven regularisation go back to Mansouri *et al.* [10] who proposed an anisotropic method that smoothes along object boundaries but not across them. More recently, Kim and Sohn [9], presented a similar approach with non-convex regularisation that gives even sharper results. However, both techniques are restricted to the ortho-parallel case, where images have already been *rectified* and the correspondence problem reduces to a search along the x-axis. In Alvarez *et al.* [1] a more general variational method with image-driven regularisation is proposed: By using knowledge on the geometry of the scene this technique does not require an explicit rectification of the input images but still satisfies the so-called *epipolar constraint* that relates corresponding points in both views. A similar method, however with solution-driven regularisation was presented by Robert and Deriche [12]. Apart from using a different regularisation strategy, their approach was also the first one that considered constancy assumptions on higher order image derivatives such as the image gradient or the Laplacian. This, in general, yields a better performance under varying illumination.

While the previous methods already combine some successful concepts, recent progress in variational optic flow computation shows that there are even more useful strategies which should be integrated in the estimation. This problem is addressed in our paper: We propose a novel variational method for estimating dense disparity maps that is based on the currently most accurate optic flow technique: The optic flow method of Brox *et al.* [2]. By integrating knowledge on the geometry of the scene we obtain a general approach that introduces the following novelties to the field of variational stereo reconstruction: Firstly, a robust data term is used. Although this concept is quite common for area-based reconstruction methods (cf. [15]), it has not been considered for variational techniques so far. Secondly, we make use of the total variation (TV) [13]. During the last years this form of penalisation of both the smoothness and the data term has become very popular and achieved good results in various research fields such as deconvolution, image restoration and optic flow estimation. And finally, the proposed approach allows to extend the theoretical justification of the warping strategy given by Brox *et al.* [2] to the field of variational stereo reconstruction methods.

Our paper is organised as follows. In Section 2 we give a short review on the stereo problem and discuss how knowledge on the geometry of the scene can be integrated appropriately into the estimation of the correspondences. This motivates us to propose a novel variational approach in Section 3 that transfers successful concepts of the highly accurate optic flow method of Brox *et al.* [2] to the field of stereo reconstruction. In Section 4 the performance of our approach is evaluated on both synthetic and real-world data while a summary in Section 5 concludes this paper.

2 The Stereo Problem

Let us consider a stereo image pair $g_l^*(\mathbf{x})$ and $g_r^*(\mathbf{x})$, where the subscripts l and r stand for the left and the right camera, respectively, and $\mathbf{x} = (x, y)^\top$ denotes the location within a rectangular image domain Ω. Then, projective geometry tells us that we can

recover the depth of a point $\mathbf{x}$ in the left image by finding its corresponding point $\mathbf{x}'$ in the right image. In other words: If we are able to compute the displacement field $\mathbf{d}(\mathbf{x}) = \mathbf{x}' - \mathbf{x}$ between the two images (*disparity*) we can reconstruct the original scene.

2.1 Epipolar Geometry

However, the displacement field $\mathbf{d}(\mathbf{x})$ cannot be arbitrary. Due to the geometry of the scene the so called *epipolar constraint* [3] must hold. It is given by

$$\hat{\mathbf{x}}'^{\top} F \hat{\mathbf{x}} = 0 \tag{1}$$

and relates the projective coordinates $\hat{\mathbf{x}} = (x, y, 1)^{\top}$ and $\hat{\mathbf{x}}' = (x', y', 1)^{\top}$ of corresponding points in both views via a 3×3 matrix of rank two – the so called *fundamental matrix* F. For a given point $\mathbf{x}$, this constraint describes a line in the right image on which $\mathbf{x}'$ must lie: The *epipolar line* Φ. Defining the following abbreviations

$$\begin{aligned} a(\mathbf{x}) &:= f_{11}x + f_{12}y + f_{13}, \\ b(\mathbf{x}) &:= f_{21}x + f_{22}y + f_{23}, \\ c(\mathbf{x}) &:= f_{31}x + f_{32}y + f_{33}, \end{aligned}$$

with f_{ij} being the entries of the fundamental matrix F, this epipolar line Φ can be written as

$$a(\mathbf{x})x' + b(\mathbf{x})y' + c(\mathbf{x}) = 0. \tag{2}$$

2.2 Integration of the Epipolar Constraint

In order to allow for an accurate estimation of the displacement field $\mathbf{d}(\mathbf{x})$, the epipolar constraint has to be integrated in the formulation of the correspondence problem. To this end, we follow the idea of Alvarez *et al.* [1] and perform an orthogonal decomposition of $\mathbf{d}(\mathbf{x})$ with respect to the direction of the epipolar line Φ. Thus, we obtain

$$\mathbf{d}(p(\mathbf{x})) = p(\mathbf{x}) \underbrace{\frac{1}{\sqrt{a^2(\mathbf{x})+b^2(\mathbf{x})}} \begin{pmatrix} -b(\mathbf{x}) \\ a(\mathbf{x}) \end{pmatrix}}_{\text{epipolar direction } \mathbf{e}(\mathbf{x})} + q(\mathbf{x}) \underbrace{\frac{1}{\sqrt{a^2(\mathbf{x})+b^2(\mathbf{x})}} \begin{pmatrix} -a(\mathbf{x}) \\ -b(\mathbf{x}) \end{pmatrix}}_{\text{epipolar normal } \mathbf{e}^{\perp}(\mathbf{x})}, \tag{3}$$

where $p(\mathbf{x})$ and $q(\mathbf{x})$ stand for the component of the projection of the displacement field $\mathbf{d}(\mathbf{x})$ in direction of and orthogonal to the epipolar line Φ, respectively. For a point $\mathbf{x}'$ on the epipolar line Φ, however, $q(\mathbf{x})$ is known. It is the distance of the point $\mathbf{x}$ to the epipolar line Φ and can be computed via

$$q(\mathbf{x}) = \left(\frac{a(\mathbf{x})x + b(\mathbf{x})y + c(\mathbf{x})}{\sqrt{a^2(\mathbf{x}) + b^2(\mathbf{x})}} \right). \tag{4}$$

Plugging this expression in equation (3) satisfies the epipolar constraint and restricts the correspondence problem to the search of one unknown function, namely $p(\mathbf{x})$.

3 From Optic Flow to Stereo

Recently, Brox *et al.* [2] proposed a highly accurate variational method for computing the displacement field between two images. However, their method was designed to be used in the context of *optic flow estimation*, where correspondences can be *arbitrary*. Since we have seen that the epipolar constraint can be satisfied if a suitable decomposition of the displacement field $\mathbf{d}(\mathbf{x})$ is performed, we can modify this approach such that it meets all requirements for a stereo reconstruction method. In the following, we present the new approach in detail.

3.1 The Variational Model

Let $g_r(\mathbf{x})$ and $g_l(\mathbf{x})$ be presmoothed versions of the original images $g_r^*(\mathbf{x})$ and $g_l^*(\mathbf{x})$ that have been obtained by convolution with a Gaussian kernel of standard deviation σ. Furthermore, let α and β be nonnegative weights. Then we propose to compute $p(\mathbf{x})$ as minimiser of the energy functional

$$E(p) = E_D(p) + \beta\, E_S(p), \tag{5}$$

where the data term is given by

$$E_D(p) = \int_\Omega \Psi_D\Big(|g_r(\mathbf{x}+\mathbf{d}(p)) - g_l(\mathbf{x})|^2 + \alpha\,|\nabla g_r(\mathbf{x}+\mathbf{d}(p)) - \nabla g_l(\mathbf{x})|^2\Big)\, dx\, dy, \tag{6}$$

and the smoothness term reads

$$E_S(s) = \int_\Omega \Psi_S\left(|\nabla p|^2\right)\, dx\, dy. \tag{7}$$

While the first part of the data term models the assumption of a constant grey value in both views, the second one renders the approach more robust against varying illumination. This is achieved by assuming constancy of the spatial image gradient $\nabla g = (g_x, g_y)^\top$. In accordance with equation (3) both assumptions are modified such that the epipolar constraint is satisfied. Moreover, no linearisation of the data term is performed to allow for a correct estimation of large displacements. Finally, in order to render the approach more robust with respect to outliers, a robust function Ψ_S is applied to the whole data term. As proposed in [2] we consider a regularised version of the total variation (TV) [13] for this purpose that is given by $\Psi(s^2) = \sqrt{s^2+\epsilon^2}$ with $\epsilon := 10^{-3}$.

In the smoothness term we follow a different strategy: Instead of regularising the total displacement field $\mathbf{d}(p(\mathbf{x}))$ we penalise deviations from the unknown function $p(\mathbf{x})$ directly. Also in this case, the regularised version of the total variation with $\epsilon = 10^{-3}$ is used as non-quadratic function Ψ_S. This solution-driven regularisation models a piecewise smooth result and thus preserves discontinuities in the disparity map.

3.2 The Euler-Lagrange-Equation

Let us now derive the Euler-Lagrange equation that is a necessary condition for the minimiser of the proposed energy functional. For better readability we use the following

abbreviations for derivatives and differences:

$$\begin{array}{lll} g := g_r(\mathbf{x}+\mathbf{d}(p)) & , & g_{xx} := \partial_{xx} g_r(\mathbf{x}+\mathbf{d}(p)), \\ g_x := \partial_x g_r(\mathbf{x}+\mathbf{d}(p)) & , & g_{xy} := \partial_{xy} g_r(\mathbf{x}+\mathbf{d}(p)), \\ g_y := \partial_y g_r(\mathbf{x}+\mathbf{d}(p)) & , & g_{yy} := \partial_{yy} g_r(\mathbf{x}+\mathbf{d}(p)), \\ g_z := g_r(\mathbf{x}+\mathbf{d}(p)) - g_l(\mathbf{x}) & , & g_{xz} := \partial_x g_r(\mathbf{x}+\mathbf{d}(p)) - \partial_x g_l(\mathbf{x}), \\ & & g_{yz} := \partial_y g_r(\mathbf{x}+\mathbf{d}(p)) - \partial_y g_l(\mathbf{x}). \end{array}$$

Moreover, we define the components of the direction $\mathbf{e}$ of the epipolar line Φ by $\mathbf{e} = (e_1, e_2)^\top$. Then, the Euler-Lagrange-equation can be written as

$$\begin{aligned} & \Psi_D' \left(g_z^2 + \alpha\left(g_{xz}^2 + g_{yz}^2\right)\right)\Big(g_z\left(g_x e_1 + g_y e_2\right)\Big) \\ +\alpha\, & \Psi_D' \left(g_z^2 + \alpha\left(g_{xz}^2 + g_{yz}^2\right)\right)\Big(g_{xz}\left(g_{xx} e_1 + g_{xy} e_2\right) + g_{yz}\left(g_{xy} e_1 + g_{yy} e_2\right)\Big) \\ -\beta\ & \mathrm{div}\left(\Psi_S'\left(|\nabla p|^2\right)\nabla p\right) = 0 \end{aligned} \tag{8}$$

with reflecting boundary conditions.

As proposed in [2] this coupled system of PDEs is solved by means of two nested fixed point iterations. Thereby a coarse-to-fine warping strategy with downsampling factor η is used. As for the original optic flow method it can be theoretically justified as an approximation strategy to the continuous energy functional.

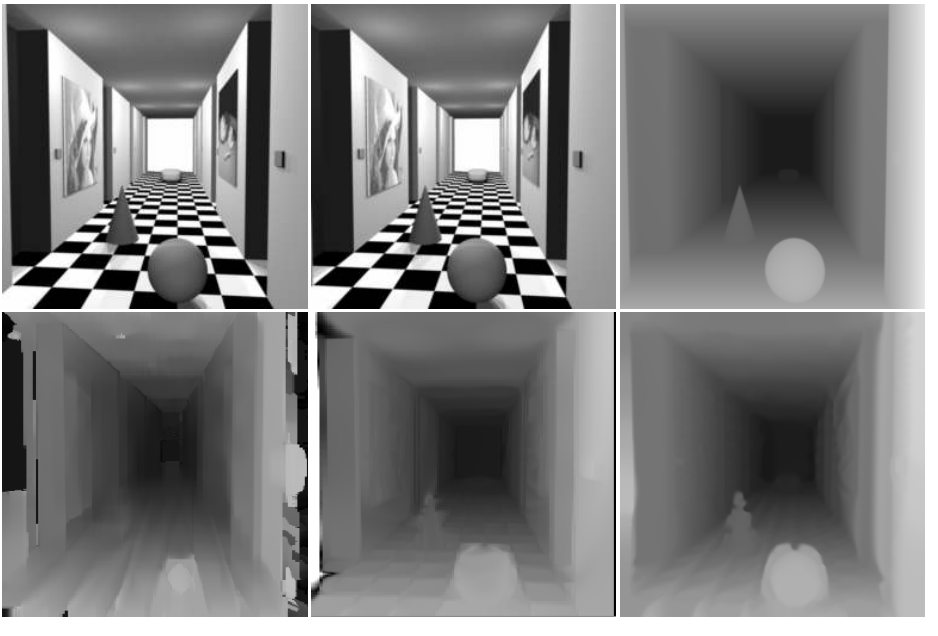

Fig. 1. *Corridor* stereo data set (`http://www-dbv.cs.uni-bonn.de/stereo_data/`). *(a) Top Left:* Left frame. *(b) Top Center:* Right frame. *(c) Top Right:* Ground truth disparity map. *(d) Bottom Left:* Correlation method. *(e) Bottom Center:* Method of Alvarez *et al.* [1]. *(f) Bottom Right:* Our method (σ = 1.55 , α = 1.1 , β = 5, η = 0.95). Computing time on a standard PC with 3.06 GHz Pentium4 CPU: 21 seconds.

Table 1. Results for the *Corridor* scene. AADE = Average absolute disparity error.

(a) Overall performance

Technique	AADE
Correlation method [1]	0.4978
Alvarez *et al.* [1]	0.2639
Our method	**0.1731**

(b) Impact of noise

Technique	Noise level σ_n^2	AADE
Our method	1	0.1952
Our method	10	0.2519
Our method	100	0.3297

4 Experiments

The proposed algorithm has been evaluated on two commonly used stereo test pairs: The synthetic *Corridor* scene from the University of Bonn, and the areal photos of the *Pentagon* building from the CMU image data base. In the case of the *Corridor* scene the known ground truth allowed us to determine the estimation quality quantitatively. This was done by computing the *average absolute disparity error* via

$$\mathbf{AADE} = \frac{1}{N} \sum_{i=1}^{N} |d_i^{\mathbf{truth}} - d_i^{\mathbf{estimate}}|. \tag{9}$$

where N is the number of pixels.

In our first experiment we have tested the proposed approach on the *Corridor* data set *without* noise. The achieved error is shown in Table 1(a), where it is compared to results from the variational approach of Alvarez *et al.* [1] and a correlation based technique with sub-pixel accuracy from the same authors. As one can see, our method outperforms both techniques significantly.

The reason for the good performance becomes obvious in the corresponding disparity maps that are presented in Figure 1: Connected areas such as ceiling, floor, walls and objects are estimated homogeneously while boundaries between them remain relatively sharp. This is a straight consequence of using the total variation in both the data and the smoothness term. In this context one should note that in accordance with Alvarez *et al.* [1] a boundary layer of 15 pixels was omitted when computing the disparity error. From the presented maps, however, one can see that this would not have been necessary for our method.

In our second experiment we have evaluated the performance of our approach with respect to noise. To this end, we used three variants of the *Corridor* scene, where Gaussian noise of zero mean and different variance σ_n^2 has been added. While Scharstein and Szeliski [15] claim that these data sets would be too difficult for a reasonable estimation, our results in Table 1(b) show that this is not the case. As one can see, they are still very accurate. In particular, one should note that our result for the version with $\sigma_n^2 = 10$ is still more precise than the results of the other two approaches for the original data set *without* noise.

In our third experiment we have reconstructed the *Pentagon* scene using the computed dense disparity map as hightfield. This is shown in Figure 2. Evidently, the whole scene looks very realistic: The discontinuities between the different sectors of the building are well preserved, the huge bridge in the upper right corner of the image is reconstructed accurately, and there are no outliers present that require the use of postprocessing steps.

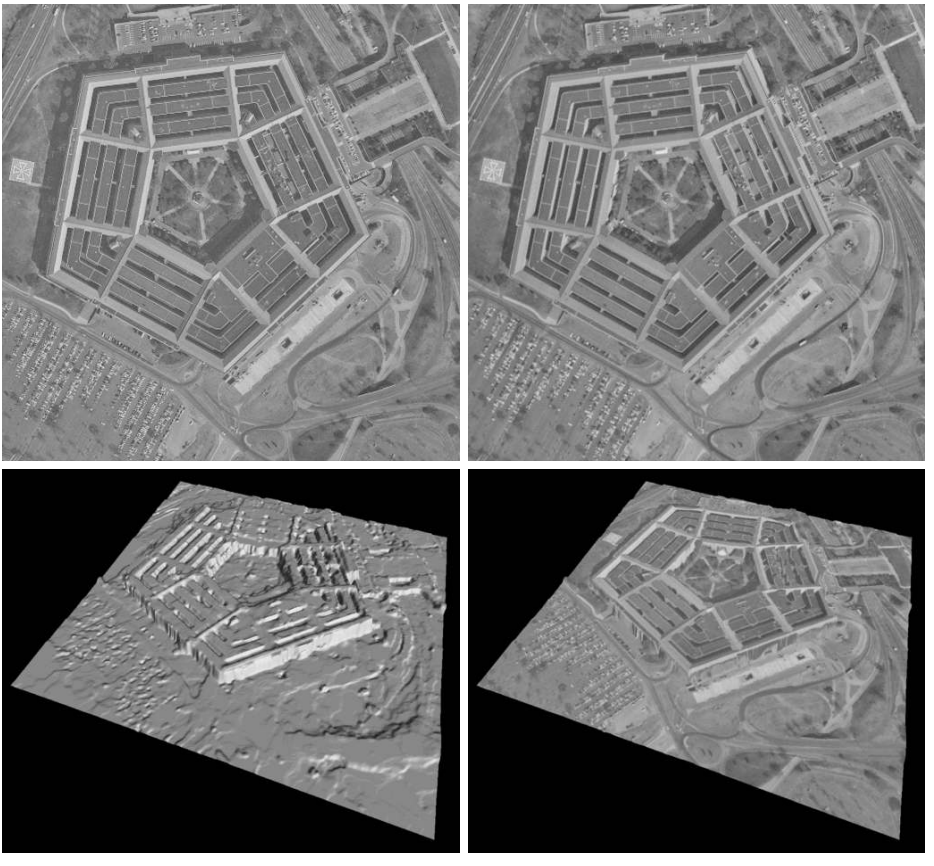

Fig. 2. *Pentagon* stereo data set (`http://vasc.ri.cmu.edu/idb/html/stereo/`). *(a) Top Left:* Left frame. *(b) Top Right:* Right frame. *(c) Bottom Left:* Computed disparity map height field with lighting (σ = 1.05 , α = 1.1 , β = 6, η = 0.95). *(d) Bottom Right:* Computed Disparity map height field with texture. Computing time on a standard PC with 3.06 GHz Pentium4 CPU: 83 seconds.

5 Summary and Conclusions

In this paper we have demonstrated that variational stereo reconstruction methods can benefit from recent progress in optic flow computation. By embedding the currently most accurate optic flow method into epipolar geometry, we achieved dense disparity maps with high quality. They respect discontinuities and are very robust under noise.

It is our hope that this strategy serves only as a first step towards a generic and mathemetically well-founded variational framework for solving the entire class of correspondence problems with high accuracy. This is a topic of our ongoing work. Apart from improving the model, e.g. by the explicit consideration of occlusions, we will also investiagte highly efficient numerifcal methods for our framework. This in turn may allow to obtain dense deformation maps for matching problems in real-time.

Acknowledgements

Natalia Slesareva gratefully acknowledges funding by the International Max-Planck Research School. Moreover, the authors thank Oliver Vogel who developed the visualisation software.

References

1. L. Alvarez, R. Deriche, J. Sánchez, and J. Weickert. Dense disparity map estimation respecting image derivatives: a PDE and scale-space based approach. *Journal of Visual Communication and Image Representation*, 13(1/2):3–21, 2002.
2. T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optic flow estimation based on a theory for warping. In T. Pajdla and J. Matas, editors, *Computer Vision – ECCV 2004*, volume 3024 of *Lecture Notes in Computer Science*, pages 25–36. Springer, Berlin, 2004.
3. O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, Cambridge, MA, 1993.
4. M. Felsberg. Disparity from monogenic phase. In L. Van Gool, editor, *Pattern Recognition*, volume 2449 of *Lecture Notes in Computer Science*, pages 248–256, Berlin, 2002. Springer.
5. T. Fröhlinghaus and J. M. Buhmann. Regularizing phase-based stereo. In *Proc. 13th International Conference on Pattern Recognition*, pages 451–455, Vienna, Austria, Aug. 1996.
6. W. E. L. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:17–34, 1985.
7. M. Hansen, K. Daniilidis, and G. Sommer. Optimization of stereo disparity estimation using the instantaneous frequency. M. Hansen, K. Daniilidis, and G. Sommer, *editors, Computer Analysis of Images and Patterns*, volume 1296 of *Lecture Notes in Computer Science*, pages 321–328, Berlin, 1997. Springer.
8. H. Hirschmüller, P. Innocent, and J. Garibaldi. Real-time correlation-based stereo vision with reduced border errors. *International Journal of Computer Vision*, 47(1-3):229–246, 2002.
9. H. Kim and K. Sohn. Hierarchical disparity estimation with energy based regularization. In *Proc. Tenth IEEE International Conference on Image Processing*, volume 1, pages 373–376, Barcelona, Spain, Sept. 2003.
10. A.-R. Mansouri, A. Mitchie, and J. Konrad. Selective image diffusion: application to disparity estimation. In *Proc. 1998 IEEE International Conference on Image Processing*, volume 3, pages 114–118, Chicago, IL, Oct. 1998. IEEE Computer Society Press.
11. D. Marr and T. Poggio. Cooperative computation of stereo disparity. *Science*, 194:283–287, 1976.
12. L. Robert and R. Deriche. Dense depth map reconstruction: A minimization and regularization approach which preserves discontinuities. In B. Buxton and R. Cipolla, editors, *Computer Vision – ECCV '96*, volume 1064 of *Lecture Notes in Computer Science*, pages 439–451. Springer, Berlin, 1996.
13. L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
14. D. Scharstein and R. Szeliski. Stereo matching with non-linear diffusion. In *Proc. 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 343–350, San Francisco, CA, June 1996. IEEE Computer Society Press.
15. D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42, 2002.
16. J. Wötzel and R. Koch. Multi-camera real-time depth estimation with discontinuity on PC graphics hardware. In *Proc. 17th International Conference on Pattern Recognition*, volume 1, pages 741–744, Cambridge, United Kingdom, August 2004.