High Accuracy TOF and Stereo Sensor Fusion at Interactive Rates

Rahul Nair[1,2], Frank Lenzen[1,2], Stephan Meister[1,2], Henrik Schäfer[1,2], Christoph Garbe[1,2], and Daniel Kondermann[1,2]

[1] Heidelberg Collaboratory for Image Processing, Heidelberg University
[2] Intel Visual Computing Institute, Saarland University
{rahul.nair,frank.lenzen}@iwr.uni-heidelberg.de

Abstract. We propose two new GPU-based sensor fusion approaches for time of flight (TOF) and stereo depth data. Data fidelity measures are defined to deal with the fundamental limitations of both techniques alone. Our algorithms combine TOF and stereo, yielding megapixel depth maps, enabling our approach to be used in a movie production scenario. Our local model works at interactive rates but yields noisier results, whereas our variational technique is more robust at a higher computational cost. The results show an improvement over each individual method with TOF interreflection remaining an open challenge. To encourage quantitative evaluations, a ground truth dataset is made publicly available.

1 Introduction

Knowledge of scene geometry is required in many areas such as visual effects, human computer interfaces or augmented reality systems. While Lidar scanners are accurate, they usually are expensive, slow and require extensive postprocessing. Stereo matching works well on textured scenes, but has difficulties at depth boundaries, in homogeneous regions and when repetitive patterns introduce ambiguities. TOF cameras recently have become an interesting alternative source of 3D data. They deliver consistent results even on textureless surfaces at the cost of a lower resolution, systematic errors such as flyig pixels, interreflections as well as high noise - especially in regions with low infrared reflectance. By harnessing the advantages of both methods we are able to improve the reconstructed depth.

1.1 Contributions

1. We define fidelity measures to identify those regions were either TOF, stereo or both approaches are likely to fail.
2. We locally fuse both sensors by appropriately dealing with problems such as missing textures in the stereo and low IR reflectance in TOF data.
3. We transfer this model to a slower but accurate variational approach using a new higher order TV regularization with a joint stereo and TOF data term.
4. The methods upsamples the 200 x 200px TOF depth to 1082x1082px.

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part II, LNCS 7584, pp. 1–11, 2012.

5. Both local and global approaches are implemented on GPU.
6. Finally, we have created a publicly available ground truth dataset to which we compare the results of our approach.

2 Related Work

3D reconstruction is a broad field of research in measurement sciences, photogrammetry and computer vision. TOF imaging is an active illumination system where infrared light is emitted by a source and reflected by the target. In each pixel the phase shift between emitted and reflected waveform is then calculated[1, 2]. TOF imaging suffers from several systematic errors such as an offset between actual depth and measured data as well wiggling errors due to deviations of the modulation from a pure sine function. A detailed description of the TOF noise sources and model is given in [3]. An interesting approach to estimating TOF noise has recently given in [4]. Here a regression forest is used to learn the noise model given lidar data as ground truth. Depth from stereo on the other hand has a long-standing history in computer vision. An overview of the methods and their shortcomings have been discussed in [5]

TOF based sensor fusion methods can be differentiated by the camera setup employed: *a)* With a single additional camera typically edge information from the high resolution intensity image is used to guide the upsampling of the depth image [6, 7]. *b)* Methods combining TOF and stereo not only filter the data to gain a higher lateral resolution; they also refine the actual depth information in textured regions. TOF imaging is a comparatively young field with a limited number of methods that have been proposed. As in pure stereo methods [5], these can be broadly grouped into local and global methods.

Local methods [8–13] tend to be faster and parallelizable but cannot cope with locally erroneous data. [9] applies a hierarchical stereo matching algorithm directly on the remapped TOF depth data without considering uncertainties. [8, 11] compute confidences in the TOF image and let stereo refine the result in regions with low confidence. The latter are similar to our method but only use binary confidence maps based on the TOF anmplitude image and therefore only sparsely use stereo information. Instead, we use the rich information of both data sources in the form of data fidelity measures to guide the fusion process.

Global methods apply spatial regularization techniques to propagate more information to regions with low confidence. Among the global methods, [14] applied an extension of semiglobal matching [15] to additionally handle TOF data. [16] applied a graph cut approach with a discrete number of disparities to sensor fusion. The methods closest to ours were proposed in [17] and [18] who employ an MAP-MRF framework and and a energy minimization scheme with quadratic regularization terms respectively. We employ adaptive first and second order total variation (TV) with L1 regularization. Such approaches already have been used for optical flow estimation [19–21]. For denoising TOF data, adaptive TV methods have been introduced in [22]. Our proposed approach differs to existing methods in the way adaptivity is defined: we include adaptivity in the differential operators, not in the TV norm.

3 Experimental Setup

The experimental setup (Fig. 1) consists of two high-resolution cameras (L, R)[1] and a low-res TOF camera[2] (T). To enable depth estimation in all areas of the primary (L) camera, the TOF camera is positioned such that the regions of occlusion between L-T and L-R are on opposite sides. Both our algorithms use the TOF intensity (TOFI) and depth (TOFD), as well as left and right images (Fig. 2). The setup is calibrated such that TOF and the right camera are registered to the left stereo image.

Fig. 1. Camera setup

Fig. 2. Input data: TOFD, TOFI (200x200), L and R (1312x1082, cropped for vis.)

4 Data Fidelity Measures

Various fidelity measures are then computed based on the input data.

1. TOF depth estimates are corrupted either when the reflectance of the illuminated surface is low or when the active light is not strong enough. A map $\boldsymbol{C^{TOFI}}$ is obtained by normalizing the inverse TOF intensity.
2. Due to flying pixels, the measurements at depth boundaries are error prone. The gradient of the TOF depth image $\boldsymbol{C^{\nabla TOFD}}$ is used to account for this.
3. In stereo matching, disparities can only be estimated when the texture within a region is sufficiently unambiguous. Regions with high horizontal gradients

[1] Photon Focus MV1-D1312-160-CL-12 with Linos Mevis-C lenses at 25mm/1.6, 1312x1082px.
[2] PMDTech Camcube, 200x200px.

can be used as indicator: the third fidelity map is the horizontal gradient of L, denoted by $\boldsymbol{C^{\nabla_x L}}$.

4. Finally, occluded areas can partly be detected using a regular photo-consistency check [15] on the TOF depth information projected on L. We denote this binary map as $\boldsymbol{C^{LOcc}}$, where occluded pixels are marked with 0.

All fidelity measures are normalized to the range $[0 \ldots 1]$ (Fig. 3). These measures enable to find regions with noisy depths, systematically inaccurate or simply missing. However, regions with corrupted data not covered by our measures (e.g. interreflection) may remain. Therefore, the next steps have to be robust.

Fig. 3. Fidelity of scene in Fig. 2. Left to right: C^{TOFI}, $C^{\nabla TOFD}$, $C^{\nabla_x L}$ and C^{LOcc}.

5 Local Fusion of TOF and Stereo

Once calibrated, each TOF depth pixel j can be projected into the L frame. Since $TOFD$ is at a lower resolution and itself has an error, each $TOFD_j$ maps to multiple pixel locations i in L, with a certainty $\alpha_1 C_j^{TOFI}$ in each pixel. Conversely, for each pixel i in L the viewing ray intersects with the TOF depth map at multiple locations. Thus there may be multiple possible depths and probabilities $(TOFD_j, p_j)$, $j = 1, \ldots, N$. With L and R rectified, these depth measurements can be converted into disparities. Using stereo block matching we can then probe the matching score of the candidate disparities and choose the best candidate. This is only feasible if the pixel at i can be used for matching at all. This depends on C_i^{LOcc} and $C_i^{\nabla_x L}$. We make two simplifications at this point to speed up the method:

1. We are using C_i^{TOFI} as an uncertainty along rays from L even though it is defined along rays from T. Since we are observing objects at a long distance this simplification is justified.
2. Instead of keeping a list of all candidate disparities in L we keep the **most probable disparity ($\mathbf{disp}^{TOF}$) and a range**.

Putting all together, we compute a **joint per pixel search range map** $[\mathbf{disp}_i^{TOF} - C_i, \mathbf{disp}_i^{TOF} + C_i]$ by combing the measures as follows:

$$C_i = (\alpha_1 C_i^{TOFI} + \alpha_2 C_i^{\nabla TOFD} + \beta_1 C_i^{\nabla_x L}) C_i^{LOcc}. \quad (1)$$

The parameters $\{\alpha_1, \alpha_2, \beta_1\}$ can be easily fixed for a given camera setup and lighting conditions. With the most probable depth and range standard block-matching is applied using a SSD cost on a square support window. This search space reduction not only speeds up the process: Also, repetitive pattern regions (multiple likely matches) are biased towards the TOF depth.

6 Variational Fusion of TOF and Stereo

In the following we describe a variational approach for the estimation of a smooth disparity map from both stereo and TOF data. A regularization term based on total variation (TV) of first and second order guarantees that the solution has the required smoothness.

Our approach considers two fidelity terms: the first one incorporates the stereo data and is based on a linearized 1D optical flow constraint (BCCE) [23], the second one incorporates the upsampled TOF data. Since the linearization of the BCCE assumes small displacements, we perform an incremental update (cf. [24]) of the targeted disparity map, i.e. $d^{k+1} := d^k + \delta^k$, starting with the TOF disparity $d^0 := \text{disp}^{TOF}$. δ^k is found by solving the variational problem

$$\delta^k := \underset{\delta}{\text{argmin}}\ \mathcal{S}_1(\delta) + \mathcal{S}_2(\delta, d^k) + \mathcal{R}(\delta, d^k) \tag{2}$$

with the two fidelity terms $\mathcal{S}_1, \mathcal{S}_2$ and a regularization term $\mathcal{R}$. We describe the terms $\mathcal{S}_1, \mathcal{S}_2$ and $\mathcal{R}$ in detail below.

The first fidelity term $\mathcal{S}_1$ is to penalize derivations from the linearized BCCE. To be robust against outliers, we use the L^1-norm. Since we already know from the map $\boldsymbol{C^{\nabla_x L}}$ where to expect matchable structures, we only apply $\mathcal{S}_1$ in regions where $\boldsymbol{C^{\nabla_x L}}$ is large. To this end we define $g_i := C_i^{\nabla_x L} \cdot C_i^{LOcc}$ and set

$$\mathcal{S}_1(\delta) := \sum_i g_i \, |(D_x R^{\text{wrp}})_i \cdot \delta_i + R_i^{\text{wrp}} - L_i)|, \tag{3}$$

where the summation is performed over all pixel indices i and D_x denotes a discrete differential operator for the horizontal derivative. R^{wrp} is the right stereo image warped according to the current disparity map d^k using bi-cubic interpolation (cf. [24]). When no matchable structures are present, we rely on the disparity map disp^{TOF} given by the TOF camera and make use of the uncertainty map $\boldsymbol{C^{TOFI}}$ and occlusion map $\boldsymbol{C^{LOcc}}$. The second fidelity term $\mathcal{S}_2$ penalizes the deviation from the data disp^{TOF} with weights $w_i := (1/C_i^{TOFI} - 1) \cdot C_i^{LOcc}$. Again, to be robust against possible outliers, we use the L^1-norm:

$$\mathcal{S}_2(\delta, d^k) = \sum_i (1 - g_i)\, w_i \, |\delta_i + d_i^k - \text{disp}_i^{TOF}|. \tag{4}$$

We now specify the regularization term $\mathcal{R}$. We use first and second order TV, which is adapted to local edges obtained as follows: Edges are defined as the maximum of the eigenvalue differences of a local structure tensor of an upsampled

gradient image (see [25]) above a certain threshold. Inspired by Canny[26] edges below a certain length are rejected. Unlike Canny's double thresholding, we check for an angular alignment of the normal vectors of neighboring pixels. The orientation of the edge can be determined since the upsampling gradient filters yield edge locations with subpixel accuracy. We denote the location of such edges by binary data terms c_x and c_y, which are zero if an horizontal/vertical edge is present. For TV regularization we consider finite difference operators D_x and D_y for calculating horizontal and vertical derivatives. We adapt the finite differences to the edge information described above and use $\overline{D}_x := c_x D_x$ instead of D_x, and $\overline{D}_y := c_y D_y$ instead of D_y, respectively. Second order finite differences $\overline{D}_{xx}$ and $\overline{D}_{yy}$ then can be defined based on $\overline{D}_x$ and $\overline{D}_y$ (see [27]). In the regularization term $\mathcal{R}(\delta, d^k)$ we then penalize the norms of the first and second order derivatives of the targeted disparity map $d^k + \delta$. Thus, we set $\mathcal{R}(\delta, d^k) := \mathcal{R}(d^k + \delta)$, where

$$\mathcal{R}(v) := \sum_i \Big(\gamma_1 \sqrt{(\overline{D}_x(v))_i^2 + (\overline{D}_y(v))_i^2} + \gamma_2 \sqrt{(\overline{D}_{xx}(v))_i^2 + (\overline{D}_{yy}(v))_i^2} \Big). \quad (5)$$

Mixed derivatives of second order were omitted as no significant changes could be observed in experiments by [27]. Parameters $\gamma_1 \geq 0$ and $\gamma_2 \geq 0$ can be chosen pixel-wise in order to locally switch between the two orders. However, our experiments indicate that fixed positive values for the whole data set suffice. The mixture of first and second order TV in particular avoids stair-casing and provides piecewise affine reconstructions (cf. Section 7).

The numerical solution of (2) is obtained by a reformulation using variable splitting [20, 28]: after introducing auxiliary variables $\tilde{\delta}$ and $\overline{\delta}$ and penalizing their difference to δ with additional cost terms, problem (2) can be approximated by

$$\min_{\delta, \tilde{\delta}, \overline{\delta}} \mathcal{S}_1(\tilde{\delta}) + \mathcal{S}_2(\overline{\delta}, d^k) + \mathcal{R}(\delta, d^k) + \frac{1}{2\lambda_1} \|\tilde{\delta} - \delta\|_2^2 + \frac{1}{2\lambda_2} \|\overline{\delta} - \delta\|_2^2, \quad (6)$$

with $\lambda_1, \lambda_2 > 0$ tending to zero. In practice, however, we use fixed small values $\lambda_1 = \lambda_2 = 0.01$. Problem (6) then is solved by alternating minimization w.r.t. $\delta, \tilde{\delta}, \overline{\delta}$, where for the minimization w.r.t. $\tilde{\delta}, \overline{\delta}$ closed forms can be provided using the pseudo-inverse of $(D_x R^{\mathrm{wrp}})_i$ and thresholding. For minimizing (6) w.r.t. δ we apply the primal-dual algorithm proposed by Chambolle & Pock [29].

7 Experiments and Results

7.1 TOF Only versus Stereo Only versus Our Approaches

Both proposed methods were implemented in C++ and make use of the CUDA framework. Fig. 4 shows a comparison of depth maps obtained from Fig.2 : using TOF only, SGM [15] stereo with rank filtering [30], and our local/TV fusion. The local method was parametrized with $(\alpha_1, \alpha_2, \beta_1) = (0.7, 1.5, 0.1)$, the global method with $(\gamma_1, \gamma_2) = (5, 1)$.

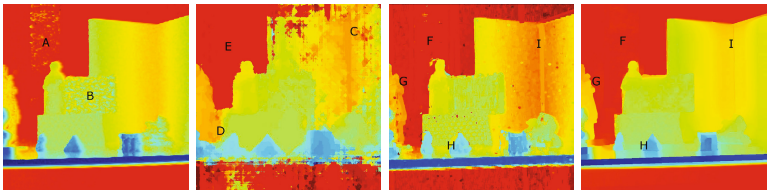

Fig. 4. Comparison of our approach. TOF only, SGM stereo, local and global fusion.

Besides the low resolution a considerable amount of noise can be observed in the TOF image, esp. in the dark regions of the poster (a) and the foam plate (b). Stereo matching fails on the wooden plates (c) due to lack of texture. Bleeding of disparities between the two statues (d) is also observable. Due to the fine texture on the poster (e) the SGM estimates the right disparity in that region.

Both our methods eliminate most of the noise around the poster (f) by using the available texture from stereo. The silhouettes (g) are reconstructed more precisely than in either TOF or stereo alone and fine details retained (e.g. pyramid (h)). Also the corner between the plates (i) that was corrupted due to interreflections is reconstructed properly as stereo cues are present there. The TV approach furthermore eliminates artifacts that could not be removed by the local method.

7.2 Ground Truth Evaluation

In the style of the Cornell Box [31], we have created a target box containing objects of different complexity (Fig. 5). The box has a size of (1 x 1 x 0.5)m. A synthetic model was then created with an accuracy within 1mm. For the shown scene we estimated the extrinsic camera parameters using manually selected 2D-to-3D correspondences. A ground truth (GT) depth map was then rendered with values that are an order of magnitude more precise than the TOF/stereo values. With few exceptions the reprojection error is lower than one pixel [31].

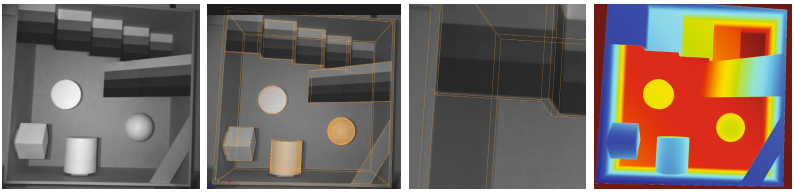

Fig. 5. *Left to right:* Box, Overlay of the GT mesh, largest alignment error, GT

To evaluate our measurements in a quantitative fashion, depth maps of the box were computed using the both approaches (Fig. 6). The variational approach was parameterized with $(\gamma_1, \gamma_2) = (5, 1)$ and the local method method with $(\alpha_1, \alpha_2, \beta_1) = (0.05, 0.05, 1.6)$. The difference between the GT image and

Table 1. Summary of GT evaluation on regions without interreflections

Data	Method	Mean	Std. dev.	1^{st} Quart.	Median	3^{rd} Quart.
TOF-data	upsampling	2.75	3.76	0.81	1.69	2.95
TOF-data	glob. meth./std. TV	2.92	3.85	0.81	1.72	3.14
TOF-data	glob. meth./adapt. TV	2.73	3.85	0.80	1.63	2.86
Stereo	SGM	2.92	3.89	0.84	1.78	3.24
Fusion	local method	2.78	3.81	0.81	1.70	2.98
Fusion	glob. meth./adapt. TV	2.57	3.53	0.75	1.58	2.78

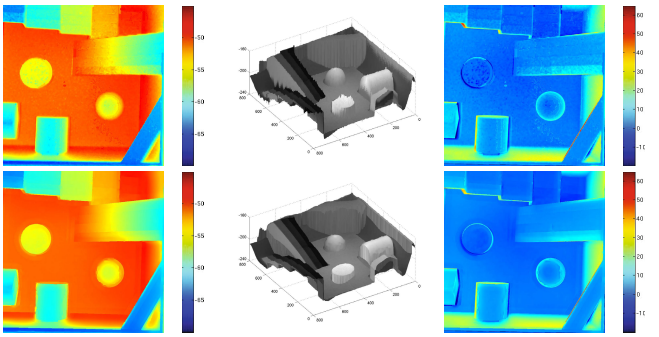

Fig. 6. *Top:* local method. *Bottom:* variational method. *From left to right:* disparities, 3D reconstruction and difference between reconstruction and GT in cm.

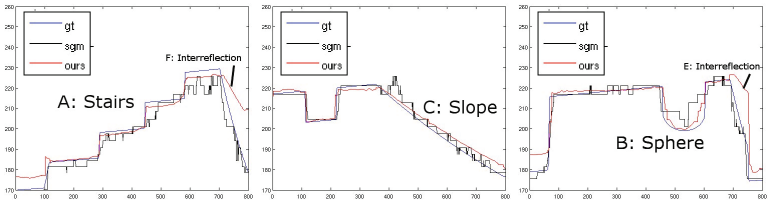

Fig. 7. Reliefs of row 50, 280 and 500. Regions of importance (a-f) are discussed in 7.2.

the obtained depth maps were then calculated. Both methods show similar results with differences to GT on the box sides due to interreflection. Interreflection is not accounted for by our confidence measures, thus the confidence is overestimated. The results without interreflecting regions are shown in Table 1. Our fusion methods are compared with pure TOF upsampling, SGM stereo as well standard and adaptive TV regularized smoothing applied to the TOF data only. The fusion methods improve on the results of TOF or stereo alone and also of to global methods using TOF data alone. The best method having a median error of 1.58 cm compared to 1.78 cm using stereo alone and 1.69 cm using TOF alone. This justifies the combination of stereo with TOF data. Among the fusion methods the local one produces results very close to pure TOF. The first two

quartiles of the local method error are almost equal to pure TOF and the third quartile slightly higher. From this (and from visual inspection) we can assume that the slight increase in error here is introduced by a few localized outliers. In retrospect this is quite clear as this scene contains too few stereo cues to use in a purely local ansatz. On such data regularization is required to propagate good stereo cues to the other regions. When stereo cues are present the local method also improves on each individual method (cf. Fig. 4).

To give some further insight, relief plots (Fig. 7) along rows of the depth images were made. Our method produces results that are less corrupted and resemble the GT relief more closely than SGM. This can be seen on *a)* the stairs where the stereo results could be interpreted as a slope, *b)* the sphere and *c)* the slope. The effects of interreflection can be observed in *d)* and *e)* .

8 Conclusion and Future Work

We have combined stereo and TOF depth data to overcome their respective limitations. Our method is based on fidelity measures derived from TOF intensities and depth gradients as well as horizontal image gradients and stereo photo consistency. This information was fed into a local and a variational framework. Both approaches yield semantically similar results, while the local approach is faster but noisier and the TV approach is less noisy but slower. We also created a publicly available millimeter-accurate ground truth dataset for the evaluation of sensor fusion systems. With this we could show that our method improves on the results of individual methods. Yet, further investigation is needed on how to handle interreflection.

Acknowledgements. This work has been co-financed by the Intel Visual Computing Institute and the MFG Baden-Württemberg. The content is under sole responsibility of the authors.

References

1. Xu, Z., Schwarte, R., Heinol, H., Buxbaum, B., Ringbeck, T.: Smart pixel–photonic mixer device (pmd). In: Proc. Int. Conf. on Mechatron. & Machine Vision (1998)
2. Lange, R., Seitz, P.: Solid-state time-of-flight range camera. JQE 37 (2001)
3. Schmidt, M.: Dissertation, IWR, Faculty of Physics, Univ. Heidelberg (2011)
4. Reynolds, M., Dobos, J., Peel, L., Weyrich, T., Brostow, G.: Capturing time-of-flight data with confidence. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 945–952. IEEE (2011)
5. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. IJCV 47, 7–42 (2002)
6. Huhle, B., Fleck, S., Schilling, A.: Integrating 3d time-of-flight camera data and high resolution images for 3dtv applications. In: Proc. 3DTV Conf. IEEE (2007)
7. Park, J., Kim, H., Tai, Y., Brown, M., Kweon, I.: High quality depth map upsampling for 3d-tof cameras. In: IEEE Proc. ICCV (2011)

8. Kuhnert, K., Stommel, M.: Fusion of stereo-camera and pmd-camera data for real-time suited precise 3d environment reconstruction. In: Int. Conf. on Intelligent Robots and Systems, pp. 4780–4785. IEEE (2006)
9. Gudmundsson, S., Aanaes, H., Larsen, R.: Fusion of stereo vision and time-of-flight imaging for improved 3d estimation. IJISTA 5, 425–433 (2008)
10. Bartczak, B., Koch, R.: Dense Depth Maps from Low Resolution Time-of-Flight Depth and High Resolution Color Views. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Kuno, Y., Wang, J., Pajarola, R., Lindstrom, P., Hinkenjann, A., Encarnação, M.L., Silva, C.T., Coming, D. (eds.) ISVC 2009, Part II. LNCS, vol. 5876, pp. 228–239. Springer, Heidelberg (2009)
11. Hahne, U., Alexa, M.: Depth imaging by combining time-of-flight and on-demand stereo. In: Dynamic 3D Imaging, pp. 70–83 (2009)
12. Dal Mutto, C., Zanuttigh, P., Cortelazzo, G.M.: A probabilistic approach to tof and stereo data fusion. In: 3DPVT, Paris, France (2010)
13. Yang, Q., Tan, K.H., Culbertson, B., Apostolopoulos, J.: Fusion of active and passive sensors for fast 3d capture. In: MMSP (2010)
14. Fischer, J., Arbeiter, G., Verl, A.: Combination of time-of-flight depth and stereo using semiglobal optimization. In: Int. Conf. on Robotics and Automation (ICRA), pp. 3548–3553. IEEE (2011)
15. Hirschmuller, H.: Stereo processing by semiglobal matching and mutual information. TPAMI 30, 328–341 (2008)
16. Hahne, U., Alexa, M.: Combining time-of-flight depth and stereo images without accurate extrinsic calibration. IJISTA 5, 325–333 (2008)
17. Zhu, J., Wang, L., Yang, R., Davis, J., et al.: Reliability fusion of time-of-flight depth and stereo for high quality depth maps. TPAMI, 1 (2011)
18. Kim, Y., Theobalt, C., Diebel, J., Kosecka, J., Miscusik, B., Thrun, S.: Multi-view image and tof sensor fusion for dense 3d reconstruction. In: ICCV Workshops, pp. 1542–1549. IEEE (2009)
19. Trobin, W., Pock, T., Cremers, D., Bischof, H.: An Unbiased Second-Order Prior for High-Accuracy Motion Estimation. In: Rigoll, G. (ed.) DAGM 2008. LNCS, vol. 5096, pp. 396–405. Springer, Heidelberg (2008)
20. Wedel, A., Pock, T., Zach, C., Bischof, H., Cremers, D.: An Improved Algorithm for TV-L^1 Optical Flow. In: Cremers, D., Rosenhahn, B., Yuille, A.L., Schmidt, F.R. (eds.) Visual Motion Analysis. LNCS, vol. 5604, pp. 23–45. Springer, Heidelberg (2009)
21. Werlberger, M., Unger, M., Pock, T., Bischof, H.: Efficient Minimization of the Non-local Potts Model. In: Bruckstein, A.M., ter Haar Romeny, B.M., Bronstein, A.M., Bronstein, M.M. (eds.) SSVM 2011. LNCS, vol. 6667, pp. 314–325. Springer, Heidelberg (2012)
22. Lenzen, F., Schäfer, H., Garbe, C.: Denoising Time-Of-Flight Data with Adaptive Total Variation. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Wang, S., Kyungnam, K., Benes, B., Moreland, K., Borst, C., DiVerdi, S. (eds.) ISVC 2011, Part I. LNCS, vol. 6938, pp. 337–346. Springer, Heidelberg (2011)
23. Horn, B., Schunck, B.: Determining optical flow. AI 17, 185–204 (1981)
24. Papenberg, N., Bruhn, A., Brox, T., Didas, S., Weickert, J.: Highly accurate optic flow computation with theoretically justified warping. IJCV 67, 141–158 (2006)
25. Köthe, U.: Edge and Junction Detection with an Improved Structure Tensor. In: Michaelis, B., Krell, G. (eds.) DAGM 2003. LNCS, vol. 2781, pp. 25–32. Springer, Heidelberg (2003)
26. Canny, J.F.: A Computational Approach to Edge Detection. TPAMI 8 (1986)

27. Setzer, S., Steidl, G., Teuber, T.: Infimal convolution regularizations with discrete l1-type functionals. Comm. Math. Sci. 9, 797–872 (2011)
28. Clason, C., Jin, B., Kunisch, K.: A duality-based splitting method for l1-tv image restoration with automatic regularization parameter choice. SIAM J. Sci. Comp., 1484–1505 (2010)
29. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. JMIV 40, 120–145 (2011)
30. Hirschmüller, H., Scharstein, D.: Evaluation of cost functions for stereo matching. In: Proc. CVPR, pp. 1–8. IEEE (2007)
31. Goral, C., Torrance, K., Greenberg, D., Battaile, B.: Modeling the interaction of light between diffuse surfaces. In: ACM SIGGRAPH Computer Graphics, vol. 18, pp. 213–222. ACM (1984)