Denoising Time-Of-Flight Data with Adaptive Total Variation

Frank Lenzen[1,2], Henrik Schäfer[1,2], and Christoph Garbe[1,2]

[1] Heidelberg Collaboratory for Image Processing, Heidelberg University
[2] Intel Visual Computing Institute, Saarland University

Abstract. For denoising depth maps from time-of-flight (ToF) cameras we propose an adaptive total variation based approach of first and second order. This approach allows us to take into account the geometric properties of the depth data, such as edges and slopes. To steer adaptivity we utilize a special kind of structure tensor based on both the amplitude and phase of the recorded ToF signal. A comparison to state-of-the-art denoising methods shows the advantages of our approach.

Keywords: Time-of-flight, Denoising, Adaptive Total Variation, Higher Order Regularization.

1 Introduction

In the last few years time-of-flight (ToF) cameras have become popular in order to retrieve depth information from 3D scenes.

The basic idea of ToF cameras is to actively illuminate the scene by a modulated infrared (IR) signal and calculate depth information from the phase shift between the emitted and recorded signal [17,20]. As an example, Fig. 1 shows a depth map (right) taken with a PMD Cam Cube 3, together with the IR amplitude of the recorded signal (middle) and a standard RGB image providing an overview over the scene (left). ToF recordings suffer from some drawbacks.
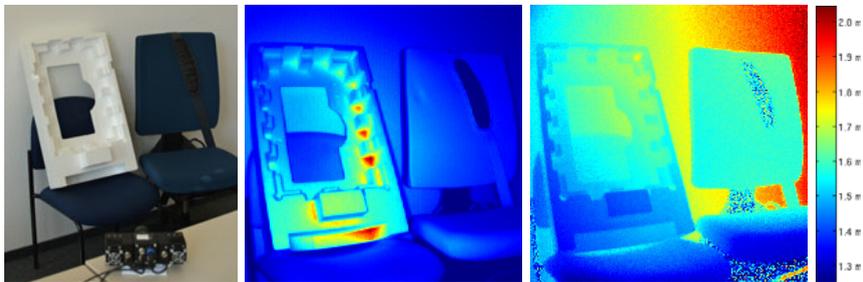
Fig. 1. Test scene and data set. Left: RGB image of test scene. Middle: amplitude of signal. Right: depth data with colorbar.

One issue is the presence of significant noise in the depth data, as can be seen in Fig. 1 right. Second, the output of ToF cameras often is of low resolution. Typical resolutions are in the range up to 200×200 pixels (PMD Cam Cube 3). Finally, if dynamic scenes are recorded, the output suffers from motion blur.

In this paper, we tackle the problem of noise contained in ToF data. The task is φ to reconstruct a noise-free and accurate depth map. This problem has already been addressed in literature. Denoising approaches for ToF data are e.g. presented in [15,10,7,6,11,1]. These approaches are mainly based on bilateral filtering or wavelet techniques, combined with appropriate noise models. In order to regularize the denoising problem, i.e. prescribing smoothness of the result, these models (except for the clustering approach in [15]) do not assume explicit geometric structures in the depth map such as regular edges or piecewise planar surfaces. In contrast, we present here a variational denoising approach based on adaptive total variation (TV), which allows to take into account such geometric properties and thus to reconstruct edges and slopes with sufficient regularity.

Denoising with TV methods has been intensively studied in literature. For the seminal work we refer to [18]; adaptive TV variants are described e.g. in [22,9,14]; higher order TV approaches are considered in [2,21,19,4]. As an alternative, *non–local* techniques have been proposed, e.g. the nonlocal-means approach in [3] and non-local TV regularization [12].

Contribution: We present a total variation (TV) based denoising approach especially tailored for smoothing depth maps. This variational approach uses penalization terms, which locally adapt to the image content and which are well suited for preserving edges and linear slopes in depth maps. Such slopes for example can be expected in depth maps of piecewise planar objects. Thus, our approach is motivated by geometrical considerations.

Organization of the paper: We start with a description of the data acquisition and transformation, cf. Sect. 2. Our denoising approach is presented in Sect. 3. A comparison to state-of-the-art methods in Sect. 4 shows the advantages of our approach. We conclude the paper with Sect. 5.

2   Data Acquisition and Transformation

The data used for the experiments is acquired with a PMD CamCube 3 with 200×200 pixels, a continuously modulated light source and suppression of background illumination. The camera records 8 images per shot, correlated with the modulation signal at 4 different phase shifts. From these 8 images, the phase shift φ and the amplitude A of the reflected light can be calculated for each pixel, as described in [17]:

$$A_{i,j} = \frac{2}{N} \left| \sum_{n=0}^{N-1} I_{i,j}^{(n)} e^{-2\pi i \frac{n}{N}} \right|, \qquad \varphi_{i,j} = arg \left(\sum_{n=0}^{N-1} I_{i,j}^{(n)} e^{-2\pi i \frac{n}{N}} \right).$$

While the phase shift φ of a pixel corresponds to the distance from the scene to the camera, the desired depth map d should contain the distance of scene and camera plane, measured parallel to the optical axis. Therefore, the phase shift has to be transformed according to $d_{i,j} := \varphi_{i,j} \cos \alpha_{i,j}$, where α is the angle included by the optical axis and the current pixel.

Remark 1. In this paper, we decided to retain the original xy-grid of the camera layout and to apply only a transformation to the depth data (z-coordinate), with the disadvantage that the geometry of the scene is not optimally represented. Future work will focus on an exact handling of the 3D geometry of the data.

3   Denoising Method

In the following, we describe our approach to denoise the depth map d obtained as described in Sect. 2. Our ansatz is based on the variational problem

$$\min_{u \in \mathbb{R}^{n,m}} F(u) := \min_u \sum_{i,j} w_{i,j}(u_{i,j} - d_{i,j})^2 + \phi(u). \tag{1}$$

Here $w_{i,j}$ are local weights on the data term in order to incorporate data feasibility. These weights are determined by considering an appropriate noise model, see Sect. 3.1. The *regularization term* $\phi(u)$ is assumed to be of the form

$$\phi(u) = \sum_{i,j} \sup\{v_{i,j}^T L_{i,j}(u) \mid v_{i,j} \in \mathcal{C}_{i,j}\}, \tag{2}$$

where $L_{i,j} : \mathbb{R}^{n,m} \to \mathbb{R}^s$ are local finite difference operators and $\mathcal{C}_{i,j} \subset \mathbb{R}^s$ are closed convex constraint sets. The general concept in (2) allows for a locally adaptive L_1-penalization of derivatives of u, provided by $L_{i,j}u$, where the adaptivity is determined by the size and shape of the constraints set $\mathcal{C}_{i,j}$. This concept also covers standard TV regularization approaches.

3.1   Weighting of the Data Term

We follow the noise model presented in [8], according to which the noise $\varepsilon_{i,j}$ at each pixel (i,j) is independent Gaussian distributed. The variance $\sigma_{i,j}$ depends on the amplitude $A_{i,j}$ of the recorded IR signal, i.e. $\sigma_{i,j}^2 = \sigma_0^2/2A_{i,j}^2$ for some constant factor $\sigma_0 > 0$. The recorded depth map then is given as $d_{i,j} = u_{i,j} + \varepsilon_{i,j}$ with noise-free data u. We apply a maximum-a-posteriori (MAP) estimator:

$$\max_{u \in \mathbb{R}^{n \times m}} p(u|d) = \max_{u \in \mathbb{R}^{n \times m}} p(d|u)p(u), \tag{3}$$

where $p(u|d)$ is the conditional probability of u given d, $p(d|u)$ is the conditional probability of d given u and $p(u)$ is the (unconditioned) probability of u. $p(u)$ is commonly assumed to be known a priorily, thus $p(u)$ is also referred to as *prior* on u. From the Gaussian noise model we find that

$$p(d|u) = \frac{1}{c_1} \prod_{i,j} e^{-\frac{A_{i,j}^2}{\sigma_0^2}|u_{i,j} - d_{i,j}|^2}, \tag{4}$$

with some constant $c_1 > 0$. For the prior on u, we consider at this point the general form $p(u) = \frac{1}{c_2}e^{-\phi(u)}$ for some suitable ϕ and a constant $c_2 > 0$. With these settings and using the fact that problem (3) is equivalent to solving $\min_u - \log p(u|d)$, we end up with

$$\min_u \sum_{i,j} \frac{A_{i,j}^2}{\sigma_0^2}|u_{i,j} - d_{i,j}|^2 + \phi(u), \tag{5}$$

where additive constant terms have been omitted. Comparing (5) with (1), we find that the weights $w_{i,j}$ in (1) have to be chosen as $w_{i,j} := A_{i,j}^2/\sigma_0^2$.

3.2   Edge Detection

To supply edge information for the proposed method, we use the enhanced structure tensor proposed in [13]. While calculating the derivatives for the structure tensor, the data is upsampled by a factor 2, to prevent information loss.

After obtaining the structure tensor, smoothed with an ordinary Gaussian, it is recalculated but with an hourglass-shaped Gaussian filter, aligned to the previously detected edges. This prevents too much smoothing in cross-edge direction, to distinguish close-by parallel edges. We use all available data, i.e. structure tensors for both A and d. Then the sum of both is evaluated to acquire the eigenvectors. This way, the normal orientation of the edges (eigenvectors $v_{i,j}$) and a value for their distinctness (differences of eigenvalues $s_{i,j}$) are obtained.

To reduce the noise in the edge image, a smoothing function is applied. Weighted with the ℓ_1 distance, the surrounding 24 pixels are checked for aligned edges by utilizing the inner product of the two relevant vectors. A Gaussian-like function is used to distinguish better between (almost) aligned and unaligned edges: $s_{i,j}^*(A,d) = \frac{1}{\text{norm}} \sum_{(k,l) \in \mathcal{N}_{24}} \left| \frac{1}{(i,j)-(k,l)} \right| \exp\left(-\left(1 - \|v_{i,j}^T \cdot v_{k,l}\| \right) / (2\sigma^2) \right)$, where norm is a normalizing constant.

Finally, to have matching edge and depth data, the edge images are downsampled to the original size.

3.3   Regularizer $\phi(u)$

Our regularization is based on discrete derivatives of first and second order, using finite differences on the pixel grid. Let $D_x u$, $D_y u$ denote right-sided finite differences for the first, and $D_{xx}u, D_{yy}u, D_{xy}u$ central finite differences for the second order, respectively. At the boundary of the pixel grid, u is constantly extended, i.e. we assume (discrete) homogeneous Neumann boundary conditions.

We are aiming at an anisotropic total variation (TV) penalization of u (cf. [9]), coupled with an isotropic L^1-penalization of the Hessian of u (cf. [19]),

$$\phi(u) := \sum_{i,j} \left(c_{i,j} \sqrt{ \begin{pmatrix} D_x u_{i,j} \\ D_y u_{i,j} \end{pmatrix}^T G_{i,j} \begin{pmatrix} D_x u_{i,j} \\ D_y u_{i,j} \end{pmatrix} } + (1-c_{i,j})\gamma_{i,j} \left\| \begin{pmatrix} D_{xx}u_{i,j} \\ D_{yy}u_{i,j} \\ D_{xy}u_{i,j} \end{pmatrix} \right\|_F \right), \tag{6}$$

where

- $c_{i,j} \in [0,1]$ provide a local weighting of the first and second order terms,
- matrix $G_{i,j} = \alpha_{i,j}^2 v_{i,j} v_{i,j}^T + \beta_{i,j}^2 (\mathrm{Id} - v_{i,j} v_{i,j}^T)$ for a given unit vector $v_{i,j} \in \mathbb{R}^2$ defines the anisotropy for the first order, with $\alpha_{i,j} > 0$, $\beta_{i,j} > 0$ being the local regularization parameters parallel and orthogonal to $v_{i,j}$, respectively,
- the term

$$\left\| \begin{pmatrix} D_{xx}u \\ D_{yy}u \\ D_{xy}u \end{pmatrix} \right\|_F := \sqrt{(D_{xx}u)^2 + (D_{yy}u)^2 + 2(D_{xy}u)^2},$$

  is the Frobenius norm of the discrete Hessian of u, and
- $\gamma_{i,j} > 0$ defines the local regularization parameter for the second order term.
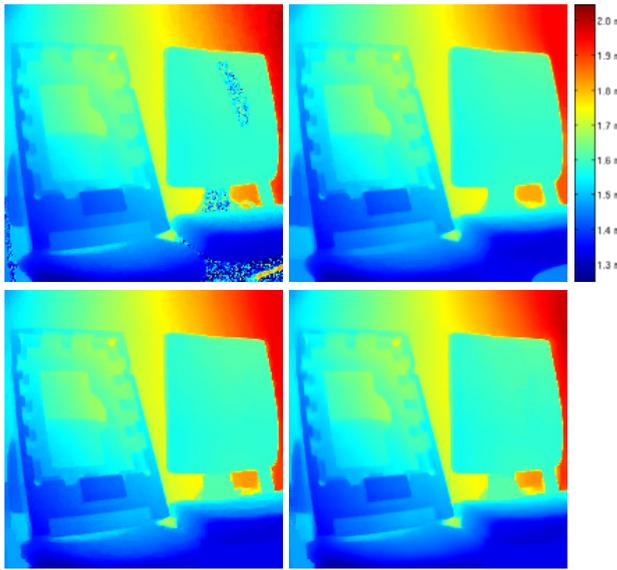
Fig. 2. Top left: standard TV. Top right: standard TV with weights. Bottom left: anisotropic TV. Bottom right: anisotropic TV with second order terms. By considering a weighted data term, we are able to cope with the local varying noise variance. Anisotropic TV ensures a better preservation of edges, while the higher order penalization term regularizes the slopes.

Remark 2. We note that the regularization term $\phi(u)$ in (6) is of the form (2), as can be seen by defining $L_{i,j}(u) := (D_x u_{i,j}, D_y u_{i,j}, D_{xx} u_{i,j}, D_{yy} u_{i,j}, D_{xy} u_{i,j})^\top$ and the constraint set $\mathcal{C}_{i,j} = \mathcal{C}(\alpha_{i,j}, \beta_{i,j}, \gamma, c_{i,j}, v_{i,j}) \subset \mathbb{R}^5$ by

$$\mathcal{C}(\alpha, \beta, \gamma, c, v) := \big\{ p \in \mathbb{R}^5 :$$
$$\frac{1}{\alpha^2} \| v^T \left(\begin{smallmatrix} p_1 \\ p_2 \end{smallmatrix} \right) \|^2 + \frac{1}{\beta^2} \| (\mathrm{Id} - vv^T) \left(\begin{smallmatrix} p_1 \\ p_2 \end{smallmatrix} \right) \|^2 \leq c^2, \frac{1}{\gamma^2} \left\| \left(\begin{smallmatrix} p_3 \\ p_4 \\ p_5 \end{smallmatrix} \right) \right\|_F^2 \leq (1-c)^2 \big\}.$$

Fig. 3. Left: investigated close-ups in the first data set. Middle: second test scene. Right: cross section (black line) through second data set.

Remark 3. For denoising ToF data, we propose to use $v_{i,j} = v_{i,j}(u)$ and $s_{i,j}^*(A, u)$ defined as in Sect. 3.2. Moreover, we use $c_{i,j} := g(s_{i,j}^*(u))$ for some continuous mapping $g : [0, 1] \to [0, 1]$ and fixed $\alpha_{i,j}, \beta_{i,j}, \gamma_{i,j} > 0$. In case that $s_{i,j}^* = 0$, i.e. no edge is present at pixel (i, j), and thus $v_{i,j}$ is not containing useful information, we additionally assume $\alpha_{i,j} = \beta_{i,j}$, leading to an isotropic TV penalization.

Note that we choose $v_{i,j}(u)$ and $s_{i,j}^*(A, u)$ depending on u instead of the noisy data d. In particular, the adaptivity of the regularizer is determined by the unknown solution u. In this case existence theory becomes more involved. Existence of a minimizer of $F(u)$ can be shown for $w_{i,j} > 0$, see [14, Prop. 1].

4   Experiments

Results of Proposed Method. We demonstrate that the combination of noise model, anisotropic TV and higher order term is necessary in order to obtain an adequate reconstruction of the noise-free depth map. To this end, we consider variants of TV denoising, which do not address all of these issues simultaneously. First, we utilize the standard ROF model in [18], see Fig. 2, top left. Since ROF assumes a constant noise variance, this method can not cope with the

| Input | std. ROF | ROF + weights | anisotropic | anisotropic + higher order |
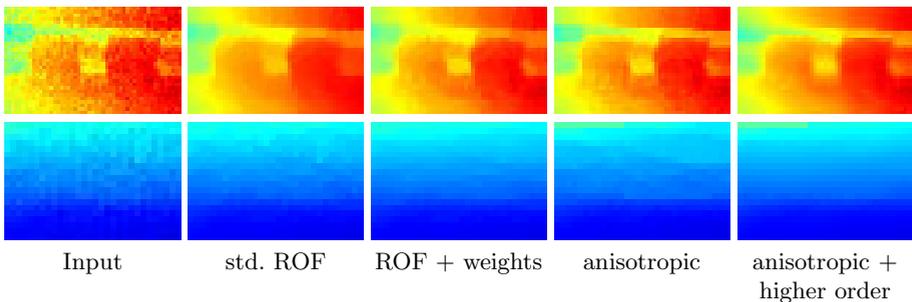
Fig. 4. Top: Edge region in the ToF image (contrast enhanced). Bottom: Slope region in the ToF image. Anisotropic TV combined with higher order terms leads to a better preservation of edges and contrast, while reconstructing slopes more regular.
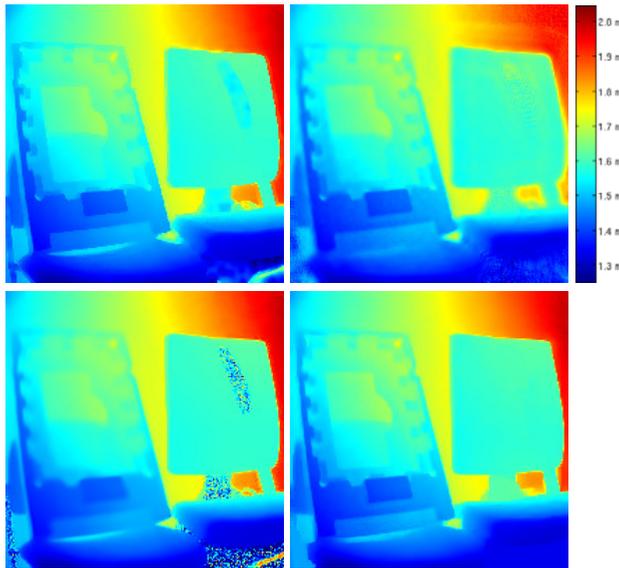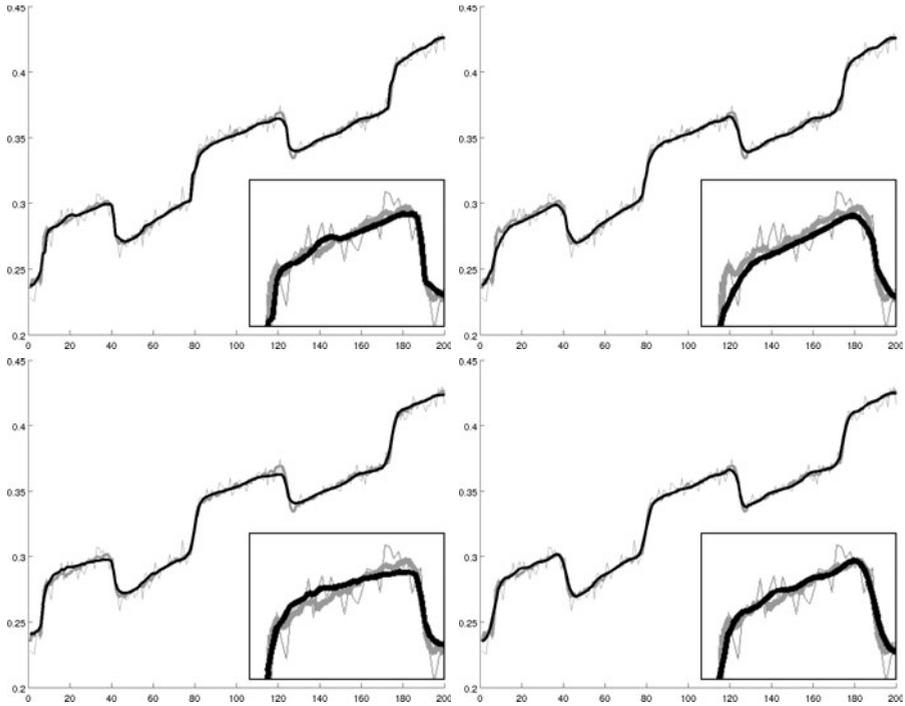
Fig. 5. Depth maps filtered with cross-bilateral filter (top left), IC (top right), NL-means (bottom left) and proposed method (bottom right). The proposed method removes noise even in regions of high noise variance, while preserving edges and slopes.

spatially dependent noise variance. Accounting for the right noise model (see Fig. 2, top right), the method is able to remove the noise completely. However, both isotropic TV variants suffer from a loss of contrast, see for example Fig. 4, top row, first and second image, where the right part of the polystyrene structure is reconstructed with the wrong depth. The effect of loosing contrast is well known in literature. Anisotropic TV, however, is able to preserve the contrast (Fig. 4, top row) and, with additional higher order term, is able to regularly reconstruct the slopes of the surfaces, see Fig. 4, bottom row.

Comparison with State-of-the-Art methods. We compare the proposed approach with several state-of-the-art methods. First, we consider a cross-bilateral filter comparable to [5], using both IR phase and amplitude. The cross-bilateral filter extends the standard method by taking the intensity image into account and uses both images to calculate the local filter kernels $w_{i,j}$:

$$w_{i,j}(k,l) = \frac{1}{\text{norm}} \frac{1}{\sigma_s} e^{-\frac{\|(i,j)-(k,l)\|^2}{2\sigma_s^2}} \cdot \left(\frac{1}{2\sigma_d} e^{-\frac{|d_{i,j}-d_{k,l}|^2}{2\sigma_d^2}} + \frac{1}{2\sigma_A} e^{-\frac{|A_{i,j}-A_{k,l}|^2}{2\sigma_A^2}} \right).$$

Since a pixel is always similar to itself, we set $w_{i,j}(i,j) = 0$ before normalization, to smooth single outliers (cf. [12]). As suggested in [24], we use three iterations and as in [16] decrease σ_d and σ_A by the square root of the number of iterations.

Fig. 6. Cross section through sample piece and comparison of cross-bilateral (top left), IC (top right), NL-means (bottom left) and proposed method (bottom right). Each plot shows the input data (thin gray), high precision data (thick gray) obtained with long exposure time and the method under consideration (black). The proposed method is able to reconstruct edges and slopes with high quality. In addition, a close-up of the left part of the cross section is provided.

In addition to bilateral filtering, we apply infimal convolution (IC) [4,21] (in combination with a weighted data term) and the non-local (NL) means algorithm [3] (using the publicly available implementation by Peyre[1]).

In Fig. 5 we present the results of these methods applied to the depth map presented in Fig. 1. Because of the intensity image, the cross-bilateral filter has quite sharp edges, but these edges often have a halo. In areas with very low intensity and high noise, the cross-bilateral filter smooths better than its standard version (defined as in [1,23]) would, due to the homogeneous dark areas in the intensity image. However, cross-bilateral filtering does not compensate the false depth data completely. Infimal convolution is able to almost reduce the noise even in regions with strong variance. The NL-means algorithm faces problems in regions with strong noise. The proposed method is able to remove the noise completely, while keeping the edges sharp.

---

[1] http://www.mathworks.com/matlabcentral/fileexchange/13619

Table 1. ℓ_2-difference to high precision data obtained with long exposure time. The proposed method shows the smallest deviation from this data.

cross-bilateral	NL-means	infimal convolution	proposed
0.036866	0.036627	0.039305	0.033600

We investigate the reconstruction of edges and slopes on a second data set, see Fig. 3, where we focus on the cross section indicated by the black line. The results of the above methods along this cross section are depicted in Fig. 6. We compare each result with a depth map taken with a long exposure time (thick gray line) and thus being more accurate than the original input data (thin gray line). The proposed approach is able to reconstruct the slopes better than NL-means. Moreover, the edges are reconstructed sharper as by the infimal convolution or cross-bilateral approach, see e.g. the left part of the cross cut. The quantitative comparison of these reconstructions, see Table 1, shows that the reconstruction by the proposed method is most accurate.

5   Conclusion and Outlook

We have proposed an adaptive TV approach to denoise ToF data, where adaptivity is determined based on a extended structure tensor using the full ToF signal (amplitude and phase). Our ansatz allows to regularize the depth map in view of its geometric properties, e.g. edges and slopes. A comparison to state-of-the-art methods shows that our approach better reconstructs the depth. In future work we will further improve the regularization in view of the true 3D geometry of the data and we will perform a detailed evaluation using ground truth.

References

1. Aurich, V., Weule, J.: Non-linear Gaussian filters performing edge preserving diffusion. In: Proceed. 17. DAGM-Symposium (1995)
2. Bredies, K., Kunisch, K., Pock, T.: Total Generalized Variation. SIAM J. Imaging Sciences 3(3), 492–526 (2010)
3. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. Multiscale Model. Simul. 4(2), 490–530 (2005)
4. Chambolle, A., Lions, P.-L.: Image recovery via total variation minimization and related problems. Numerische Mathematik 76, 167–188 (1997)
5. Chan, D., Buisman, H., Theobalt, C., Thrun, S.: A noise-aware filter for real-time depth upsampling. In: Proc. of ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications, pp. 1–12 (2008)

6. Edeler, T., Ohliger, K., Hussmann, S., Mertins, A.: Time-of-flight depth image denoising using prior noise information. In: Proceedings ICSP, pp. 119–122 (2010)
7. Frank, M., Plaue, M., Hamprecht, F.A.: Denoising of continuous-wave time-of-flight depth images using confidence measures. Optical Engineering 48 (2009)
8. Frank, M., Plaue, M., Rapp, K., Köthe, U., Jähne, Hamprecht, F.A.: Theoretical and experimental error analysis of continuous-wave time-of-flight range cameras. Optical Engineering 48(1), 13602 (2009)
9. Grasmair, M., Lenzen, F.: Anisotropic Total Variation Filtering. Appl. Math. Optim. 62(3), 323–339 (2010)
10. Schöner, H., Moser, B., Dorrington, A.A., Payne, A.D., Cree, M.J., Heise, B., Bauer, F.: A clustering based denoising technique for range images of time of flight cameras. In: CIMCA/IAWTIC/ISE 2008, pp. 999–1004 (2008)
11. Jovanov, L., Pizurica, A., Philips, W.: Fuzzy logic-based approach to wavelet denoising of 3D images produced by time-of-flight cameras. Opt. Express 18, 22651–22676 (2010)
12. Kindermann, S., Osher, S., Jones, P.W.: Deblurring and denoising of images by nonlocal functionals. Multiscale Model. Simul. 4(4), 1091–1115 (2005) (electronic)
13. 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)
14. Lenzen, F., Becker, F., Lellmann, J., Petra, S., Schnörr, C.: Variational image denoising with constraint sets. In: Proceedings SSVM (in press, 2011)
15. Moser, B., Bauer, F., Elbau, P., Heise, B., Schöner, H.: Denoising techniques for raw 3D data of ToF cameras based on clustering and wavelets. In: Proc. SPIE, vol. 6805 (2008)
16. Paris, S., Durand, F.: A fast approximation of the bilateral filter using a signal processing approach. Technical report, MIT - CSAIL (2006)
17. Plaue, M.: Analysis of the PMD imaging system. Technical report, Interdisciplinary Center for Scientific Computing, University of Heidelberg (2006)
18. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D 60(1-4), 259–268 (1992)
19. Scherzer, O.: Denoising with higher order derivatives of bounded variation and an application to parameter estimation. Computing 60, 1–27 (1998)
20. Schmidt, M., Jähne, B.: A physical model of time-of-flight 3D imaging systems, including suppression of ambient light. In: Kolb, A., Koch, R. (eds.) Dyn3D 2009. LNCS, vol. 5742, pp. 1–15. Springer, Heidelberg (2009)
21. Setzer, S., Steidl, G., Teuber, T.: Infimal convolution regularizations with discrete l1-type functionals. Comm. Math. Sci. 9, 797–872 (2011)
22. Steidl, G., Teuber, T.: Anisotropic smoothing using double orientations. In: Proceedings SSVM 2009, pp. 477–489 (2009)
23. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proceedings of the Sixth International Conference on Computer Vision (ICCV 1998), p. 839 (1998)
24. Weiss, B.: Fast median and bilateral filtering. ACM Trans. Graph. 25(3), 519–526 (2006); Proceedings of ACM SIGGRAPH 2006