Denoising Strategies for Time-of-Flight Data

Frank Lenzen[1,2], Kwang In Kim[3], Henrik Schäfer[1,2], Rahul Nair[1,2], Stephan Meister[1,2], Florian Becker[1], Christoph S. Garbe[1,2], and Christian Theobalt[3,2]

[1] Heidelberg Collaboratory for Image Processing (HCI), Heidelberg University, Speyerer Str. 6, 69115 Heidelberg, Germany
{Frank.Lenzen,Henrik.Schaefer,Rahul.Nair,Stephan.Meister}@iwr.uni-heidelberg.de, becker@math.uni-heidelberg.de, Christoph.Garbe@uni-heidelberg.de

[2] Intel Visual Computing Institute, Saarland University, Campus E2-1, 66123 Saarbrücken, Germany

[3] Max-Planck-Institut für Informatik, Saarland University, Campus E1-4, 66123 Saarbrücken, Germany
kkim@mpi-inf.mpg.de, theobalt@mpii.de

Abstract. When considering the task of denoising ToF data, two issues arise concerning the optimal strategy. The first one is the choice of an appropriate denoising method and its adaptation to ToF data, the second one is the issue of the optimal positioning of the denoising step within the processing pipeline between acquisition of raw data of the sensor and the final output of the depth map. Concerning the first issue, several denoising approaches specifically for ToF data have been proposed in literature, and one contribution of this chapter is to provide an overview. To tackle the second issue, we exemplarily focus on two state-of-the-art methods, the *bilateral filtering* and *total variation (TV) denoising* and discuss several alternatives of positions in the pipeline, where these methods can be applied. In our experiments, we compare and evaluate the results of each combination of method and position both qualitatively and quantitatively. It turns out, that for TV denoising the optimal position is at the very end of the pipeline. For the bilateral filter, a quantitative comparison shows that applying it to the raw data together with a subsequent median filtering provides a low error to ground truth. Qualitatively, it competes with applying the (cross-)bilateral filter to the depth data. In particular, the optimal position in general depends on the considered method. As a consequence, for any newly introduced denoising technique, finding its optimal position within the pipeline is an open issue.

1 Introduction

Measurements from Time-of-Flight cameras suffer from severe noise. This noise is introduced when the raw image data are recorded by the camera sensor. It is non-linearily amplified in the subsequent post-processing, where the actual depth data are derived. For a detailed discussion on the noise we refer the reader to the first chapter of this book.

M. Grzegorzek et al. (Eds.): Time-of-Flight and Depth Imaging, LNCS 8200, pp. 25–45, 2013.
© Springer-Verlag Berlin Heidelberg 2013

Higher level computer vision algorithms are often sensitive to the noise level typically for ToF data and it is inevitable to denoise the data before applying these methods. Three major questions arise concerning the denoising task:

1. Which state-of-the-art method should be chosen for denoising the depth data?
2. At which stage of the data processing should the denoising method be applied? Two obvious alternatives are to denoise the raw data or the final depth data. Denoising of some intermediate data is also possible.
3. Which modifications can be applied to state-of-the-art methods to increase their performance with respect to ToF data?

We start this chapter with an overview over state-of-the-art denoising methods for standard gray or color images in Section 2.1, including the class of learning approaches, which are gaining importance in this field. Afterwards, we discuss approaches which are proposed in literature specifically for denoising ToF data, cf. Section 2.2.

The main focus of this chapter is the question of the optimal position of the denoising method within the data processing pipeline. Not much research has been done in this direction so far. Most of the related work solely considers denoising of the depth map provided by the camera. One reason for this might be the fact that for most cameras, the raw data is not accessible to the users. However, since having access to raw data is of interest for scientific applications of ToF cameras, in future more camera manufactures might consider to provide corresponding interfaces.

To answer the question of positioning, we exemplarily consider in Section 3 two denoising methods, which are commonly used for ToF data, the *bilateral filter* and *total variation-(TV)*-based denoising. We discuss several alternatives of how to apply these methods to the raw, intermediate and final data processed by the ToF camera. In addition, we discuss modifications to improve the restoration quality of the considered methods. These modifications consist in making the approaches *adaptive*, *anisotropic* and, in particular for the TV denoising approach, to consider *second-order* smoothing terms.

In the experimental part in Section 4 we evaluate the different approaches based on a test data set with ground truth. It turns out that for TV denoising the optimal position is at the end of the processing pipeline. For the bilateral filter, we found that applying it to the raw channels and performing a subsequent median filter provides the smallest quantitative error. Qualitatively, it competes with applying the bilateral and the cross-bilateral filter to the depth data.

2 State-of-the-Art Denoising Techniques

2.1 Denoising of Standard Images

The task of denoising faces the major problem of finding a trade-off between removing the noise and preserving the detailed structures of the original data.

For images, these details are mainly the edges and textures. Applying for example classical Gaussian convolution for images, one obtains a blurred images with unsharp edges and with textures removed.

Various approaches exist in literature, which tackle both edge and texture preservation. One edge preserving variant of Gaussian convolution is the *bilateral filter* [1,2,3]. Here, the filter kernel decreases with increasing spatial distance as well as with increasing distance in intensity. Another family of denoising methods are the *PDE-based* approaches. They built on the fact that Gaussian convolution provides a solution to the linear diffusion equation, but use modifications to guarantee edge preservation. The most prominent methods of this kind are the *nonlinear diffusion* proposed by Perona and Malik [4] and the *anisotropic diffusion* [5].

The bilateral filter as well as the mentioned PDE approaches provide solutions which are smooth in the mathematical sense. As a consequence, sharp jumps in intensity or color can only be modeled by steep but smooth slopes. There exists approaches which explicitly allow for piecewise constant solutions, where edges can be represented by sharp jumps. Among these are the *wavelet methods*, see e.g. [6]. In image processing the most commonly used wavelets are the Haar wavelets, which represent a discrete space of piecewise constant functions. Soft thresholding then is applied to the wavelet coefficients of the image to remove highly oscillating components.

In 1992 Rudin, Osher and Fatemi [7] proposed to consider a variational approach using *total variation (TV)* regularization. In particular, this approach allows for piecewise constant solutions and thus is able to restore image edges sharply. Due to the variational formulation with a data-fidelity and a regularization term, this ansatz easily extends to other applications in computer vision such as optical flow and stereo, cf. Chapter 6.The classical TV regularization faces the drawbacks of a loss of contrast and stair-casing artifacts (piecewise constant reconstruction of the data where a smooth slope would be expected). Various TV variants have been proposed to overcome these drawbacks, including *adaptive TV* [8,9], *anisotropic TV* [10,11] and *approaches of higher order TV* [12,13,14].

Another approach dealing with piecewise smooth functions has been proposed by *Mumford and Shah* [15,16].

The methods mentioned so far all share the problem that textures in the data are over-smoothed. Non-local approaches such as *non-local means* [17], *non-local TV* [7,18,19] and the *BM3D* methods (e.g. [20]) turned out to have better texture preserving qualities.

Besides image denoising techniques, which are driven by a single input image, we also discuss data-base driven methods, which are gaining importance in image processing.

The underlying idea of database-driven methods is to learn a map from low-quality (noisy) images to high-quality images based on example pairs of low- and high-quality images. Burger et al. [21] proposed denoising images using multilayer Perceptrons (MLPs): A given noisy image is divided into an overlapping set

of image patches (small sub-windows). For each noisy patch, the corresponding clean patch is predicted using an MLP that is trained based on a large collection of pairs of input noisy and the corresponding output clean image patches. Given the patch predictions, the final image-valued output is reconstructed by taking averages for overlapping windows. A similar approach has also been proposed by Jain and Seung [22] in which convolutional networks are adopted.

An important advantage of database-driven methods is that they relieve the user from the extremely difficult task of designing an analytical noise model. This is especially important when the underlying noise generation process is non-Gaussian or, in general, not well-studied or modeled. Accordingly, conventional analytic noise models cannot be straightforwardly applied. Database-driven approaches enable building a denoising system (and general image enhancement system) by preparing a set of example pairs of clean and noisy images and learning specific degradation models from such training data. This has been demonstrated by the success of database-driven approaches for the related problems of single-image super-resolution and artifacts removal in compressed images, in which no analytical noise models are available. The reported results in these domains were superior to the state-of-the-art algorithms [23,24,25,26]. Even for the extensively studied Gaussian noise case, the reported performances were comparable to state-of-the-art image denoising algorithms [22,21].

One major drawback is that these algorithms are 'black boxes': due to the non-parametric nature of modeling, the trained denoising algorithms do not assist understanding the underlying noise generation or image degradation processes. Another limitation that is especially relevant for ToF image denoising is that they require pairs of clean and noisy images. Please see the next section for a more detailed discussion.

2.2 Denoising Techniques for Time-of-Flight Data

We start this section with a discussions of the **challenges**, which arise with denoising ToF data compared to denoising standard images.

- As already discussed in Chapter 1, the noise in ToF data varies depending on the amplitude of the recorded signal. A Gaussian distribution with variance proportional to $A^{-2}(\boldsymbol{x})$, where $A(\boldsymbol{x})$ is the amplitude of the recorded signal at pixel $\boldsymbol{x}$, provides a efficient approximation, cf. [27]. Standard denoising models, however, often assume identically distributed Gaussian noise and thus can only be applied after adapting to the locally varying noise variance.
- Due to their low spatial resolution, textures are not as dominant as in standard images and the issue of texture preservation is less relevant. As a consequence, the texture preserving properties of non-local methods are of less importance for denoising ToF data.
- To model depth data, it is common to assume piecewise smooth data with salient depth edges. Depending on the scene recorded, planar surfaces might dominate, which could be considered in the denoising approach, e.g. by regularization methods which favor piecewise affine reconstructions. However,

one has to keep in mind that the depth maps provided by ToF cameras are actually the radial distances of the objects to the camera center. We refer to this as *radial depth*. As a consequence, surfaces which are flat in 3D are represented by curved surfaces on the camera grid. Calculating for each pixel the scene depth parallel to the viewing direction (*z-depth*) without adapting the (x,y)-pixel positions reduces the projective distortion, but does not completely compensate it (cf. [28]). An alternative would be to generate a 3D point cloud from the depth map, project these points onto the image plane and associate each of these 2D sampling points with its z-depth. The drawback for such an approach is, that these sampling points in general are no longer equally distributed. However, in our experiments we experienced that when just using the z-depth stored on the original pixel grid, the projective distortion of planar surfaces can be neglected compared to other systematic errors of the ToF systems.

- Finally, we want to stress the fact that the quality of ToF data is evaluated different to natural images. While for natural images the visual impression often is used for evaluation, for ToF data their precision is the most important criterion. Denoising methods might reveal effects that do not significantly change the visual appearance of the outcome, for example a loss of contrast. On depth maps such effects instead might significantly falsify the data. Therefore, when selecting appropriate denoising methods for ToF data, care has to be taken to preserve the accuracy of the depth data.

Let us now give a short overview over the **methods** discussed in literature for denoising ToF data.

2.2.1 Image Driven Methods

In this subsection, we consider *image driven* methods, i.e. methods which as input require only the data, which are to be denoised. Opposed to these are the *database-driven* (or learning) methods, which require a training phase with additional input prior to their actual application.

Clustering approaches for ToF denoising have been proposed by Schöner, Moser et al. [29,30]. Frank et al. [31] have considered *adaptive weighted Gaussian* as well as *median filtering*. For these approaches, they consider different positions within the depth acquisition pipeline. They come to the conclusion that, among the alternatives considered, adaptive weighted Gaussian filtering on the final depth in general gives the best results. However, it is not clear if this statement can be generalized to other denoising methods. *Wavelet denoising* of ToF data has been considered by Moser [30] and by Edeler et al. [32,33]. A popular denoising method used for ToF data is the already mentioned *bilateral filter* (see e.g. [34]). A *joint- or cross-bilateral filter* on both the depth and intensity data shows good denoising capabilities. We give a short overview over the standard and the cross-bilateral filter below. Schöner et al. [35] recently applied *anisotropic diffusion* to ToF data. In [28] we considered *total variation* regularization for ToF denoising.

In order to deal with the low spatial resolution of ToF data, fusion of multiple data sets has been proposed. In principle, ToF data can be fused with data

from any other imaging device. The most prominent variants are multiple ToF data [36,37,38], fusion of ToF with rgb data (*rgbd*) [39,40,41,42] and fusion of ToF with stereo data. For the latter, we refer to Chapter 6 for a detailed discussion.

In all these approaches, denoising of the data is also an issue. Denoising techniques considered within the fusion approaches are for example bilateral TV regularization [38], (cross-)bilateral filtering [41,39] and adaptations of non-local means [42,40].

2.2.2 Database-Driven Methods

Despite the success of database-driven approaches for image denoising and enhancement, their applications to ToF data have not been actively explored. This can partially be attributed to the difficulty in generating the training data: The performance of database-driven algorithms rely heavily on the availability of large-scale and high-quality data. However, unlike the images, generating the ground truth data is highly non-trivial, as it requires measuring the 3D geometry of the scene of interest. One way of generating example pairs is to scan the scene with a laser scanner as well as with the ToF camera. However, accurate registration between ToF and laser scan data is necessary [43].

Although this chapter does not evaluate this class of algorithms, recent work breaks the limits of ground truth generation in this respect. Mac Aodha et al. [44] proposed an algorithm for single-depth image super-resolution. Similar to existing approaches for image denoising and enhancement, they adopt a local patch-wise prediction combined with a global image prior where the patch prediction step takes account of the training data. Unlike typical example-based approaches, the training examples are generated from synthesized 3D geometries, i.e., an example pair is generated by capturing a view of a synthesized scene followed by the corresponding degradation which, in the context super-resolution, is the down-sampling. This approach can facilitate applying well-developed database-driven image enhancement algorithms to ToF data without having to set up a laser scanning studio or involve other expensive hardwares.

For denoising ToF data, this approach requires building the corresponding noise model which may invalidate an important advantage of database-driven approaches: There is no need to analyze the noise characteristics. Nevertheless, database-driven approaches have certain potential advantages over conventional approaches that we believe justify future investigation: 1) Sampling and adding noise to synthetic data is still easier than constructing an algorithm that explicitly inverts the noise generation process; 2) It is easy to reflect a certain type of a priori knowledge into database-driven approaches. For instance, if it is known that the scene of interest shows a specific class of objects (e.g., faces), one could train an algorithm on examples generated from this specific class. As exemplified in Fig. 1, this strategy can significantly improve the performance over using generic databases for the case of super-resolution [26] and may show promise for denoising. This type of a priori knowledge can not be straightforwardly exploited in conventional approaches.

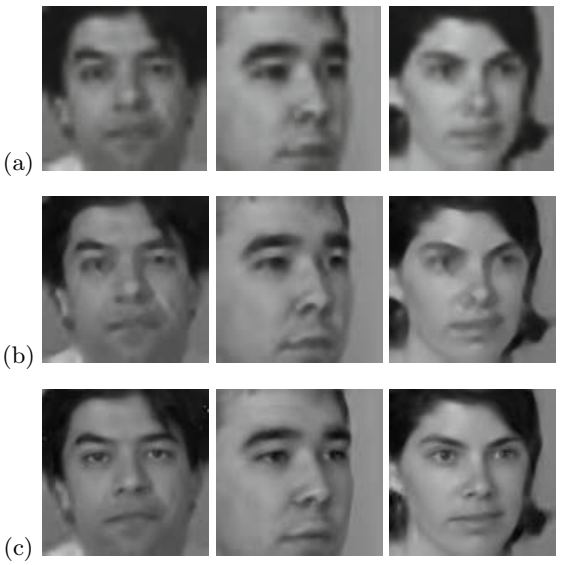


Fig. 1. The improvement made possible when training on a specific class of objects, here demonstrated for face image super-resolution (magnification factor 4). (a) bicubic resampling, (b) super-resolution results of Kim and Kwon's algorithm trained based on a generic image database [25], and (c) super-resolution results of Kim et al.'s algorithm trained based on a face database [26]. We expect a similar behavior for database driven *denoising* of ToF data.

3 Denoising Strategies

3.1 Methods under Consideration

The methods we consider here use as input some of the data provided by the ToF camera, which are the raw data, the amplitude, the intensity and/or the depth data. We exemplarily focus on the bilateral filter and total variation (TV) denoising approach and compare different modifications of both working on specific subsets of the available data. We start this section with a review of the standard versions of the bilateral filter and the TV denoising.

3.1.1 Bilateral Filter

The bilateral filter was first introduced by Aurich and Weule in [1] as edge preserving smoothing. Its actual name was conceived later by Tomasi and Manduchi [2] in 1998. The idea of the bilateral filter is to have a second domain, usually the intensity data, that weakens the smoothing of a standard Gaussian at intensity

discontinuities. A Gaussian weighting in this second domain is commonly used. The bilateral filter, providing filtered data u from input v, is given as

$$u(\boldsymbol{x_0}, v) = \frac{1}{a_{Norm}} \int_{\Omega} v(\boldsymbol{x}) G_s(\|\boldsymbol{x_0} - \boldsymbol{x}\|) G_i(|v(\boldsymbol{x_0}) - v(\boldsymbol{x})|) d\boldsymbol{x}, \tag{1}$$

where $\Omega \subset \mathbb{R}^2$ is the image domain, G_s and G_i are the Gaussian convolution kernels in spatial and intensity domain, respectively, and a_{Norm} is a normalization factor. Image coordinates are denoted by $\boldsymbol{x}$.

Regarding ToF-depth data, it is hard to find a suitable σ for the second domain, since the noise level varies strongly over the image, depending on the intensity or amplitude of different regions. The results are either smeared edges in bright parts or unsmoothed noise in darker areas. But the filter can be applied to the four different raw-images, which are basically intensity images.

Still, there are different ways to apply a bilateral filter to the depth data by incorporating other information as well. So called joint- or cross-bilateral filters [45] do not use the primary data to determine the weight in the second domain but calculate it from an additional image, which is less prone to noise (cf. [46,47]). In case of a ToF-camera, this second image could be the intensity or amplitude data. As mentioned already in Section 2.2, a different image with higher resolution can even be used to achieve super-resolution directly in the denoising step.

An alternative is to use both the intensity or amplitude image and the depth image for a combined bilateral filter, following [28]. This method especially preserves edges which are visible in both data sets. Applying the bilateral filter to the complex representation of the data has a similar effect. In the complex representation, the angle of each point towards the x-axis corresponds to the phase shift of the signal, while the distance to the origin is the amplitude. As a second weighting for the bilateral filter, the distance of points in the complex plane is used. We finally remark that the bilateral filter can be efficiently implemented on a GPU.

3.1.2 Denoising with Total Variation

Standard Total Variation denoising (the Rudin-Osher-Fatemi (ROF) model [7]) follows the classical form of a regularization approach, where the objective function to be minimized consists of a data-fidelity term combined with a regularization term. We describe the approach in a discrete framework. Let $\mathcal{N}$ denote the set of nodes of the pixel grid with grid size h. We denote image coordinates by $\boldsymbol{x} = (x, y)$. The optimization problem to be solved to obtain smoothed data u from noisy input f is given as

$$\min_u \left[\left(\sum_{\boldsymbol{x} \in \mathcal{N}} \tfrac{1}{2} w(\boldsymbol{x}) \left(u(\boldsymbol{x}) - f(\boldsymbol{x}) \right)^2 \right) + \lambda \mathcal{R}(u) \right]. \tag{2}$$

We refer to the first term in (2) as the *data (fidelity) term* and to $\lambda\mathcal{R}(u)$ as the *regularization term*. For the ROF model, the latter is given as

$$\begin{aligned}\mathcal{R}(u) :&= \sum_{(x,y)} \|\nabla u(x,y)\| \\ &= \sum_{(x,y)} \sqrt{\big(u(x+h,y)-u(x,y)\big)^2 + \big(u(x,y+h)-u(x,y)\big)^2}.\end{aligned} \tag{3}$$

The *regularization parameter* $\lambda > 0$ in (2) controls the amount of smoothing. $w(\boldsymbol{x})$ is a weighting term, which is used to account for the locally varying noise variance. For independent and identically distributed Gaussian noise with zero mean, one would use a constant weighting $w(\boldsymbol{x}) \propto \frac{1}{\sigma^2}$. After rescaling the parameter λ, $w(\boldsymbol{x}) = 1$ can be assumed. For ToF data, which show a locally varying noise variance proportional to $\frac{1}{A^2}$, we propose to use

$$w(\boldsymbol{x}) = \tfrac{1}{c}\min(c, A^2(\boldsymbol{x})), \tag{4}$$

where we cut off the weighting function above some constant $c > 0$, and rescale it to $\max_{\boldsymbol{x}} w(\boldsymbol{x}) = 1$, so that the regularization parameter λ can be chosen in the same range as in the case of constant $w(x) = 1$.

In order to solve the optimization problem (2), we propose to use a primal-dual approach as for example described in [48]. Such a primal-dual approach is able to handle the non-differentiability of $\mathcal{R}(u)$ and thus leads to a better edge preservation (in terms of sharpness) than for example methods approximating $\mathcal{R}(u)$ by smooth functions. We remark that also primal-dual approaches can be efficiently implemented on GPUs.

3.2 Positioning within the Processing Pipeline

We start with a short review of depth acquisition process of a ToF camera as already discussed in detail in Chapter 1:

- Four individual raw images $A_j(\boldsymbol{x})$ at $\tau_j = \frac{\pi}{2}j$, $j = 0, \ldots, 3$, are recorded with the camera sensor. Here we denote with $\boldsymbol{x}$ the pixel position. Typically, the measurements are obtained using multiple taps. To deal with individual tap characteristics, recordings from corresponding taps are averaged [49, Sect. 5.2.]. We assume that $A_j(\boldsymbol{x})$ are already the averaged values.
- These raw data are related to the signal $\frac{A(\boldsymbol{x})}{2}\cos(\tau_j + \phi(\boldsymbol{x})) + I(\boldsymbol{x})$ amplitude $A(\boldsymbol{x})$, phase shift $\phi(\boldsymbol{x})$ and intensity $I(\boldsymbol{x})$. Optimal values for $A(\boldsymbol{x}), \phi(\boldsymbol{x})$ and $I(\boldsymbol{x})$ can be found by minimizing the least-squares error

$$\sum_{j=0}^{3} \Big(\tfrac{A(\boldsymbol{x})}{2}\cos(\tau_j + \phi(\boldsymbol{x})) + I(\boldsymbol{x}) - A_j(\boldsymbol{x})\Big)^2. \tag{5}$$

  In particular, this optimization problem is independent in each pixel position. The standard approach is to transform it into a quadratic minimization

problem by a change of variables. The analytic solution of the transformed problem is given as

$$\begin{aligned} I(\boldsymbol{x}) &= \tfrac{1}{4}\sum_{j=0}^{3} A_j(\boldsymbol{x}), \\ A(\boldsymbol{x}) &= \tfrac{1}{2}\sqrt{(A_0(\boldsymbol{x}) - A_2(\boldsymbol{x}))^2 + (A_3(\boldsymbol{x}) - A_1(\boldsymbol{x}))^2}, \\ \phi(\boldsymbol{x}) &= \arctan\left(\frac{A_3(\boldsymbol{x}) - A_1(\boldsymbol{x})}{A_0(\boldsymbol{x}) - A_2(\boldsymbol{x})}\right) \end{aligned} \tag{6}$$

(cf. Chapter 1 and [27]). We remark that ϕ is the phase of the complex-valued signal z with $Re(z) = A_0 - A_2$ and $Im(z) = A_3 - A_1$. One of the denoising strategies discussed below considers smoothing of this complex-valued signal z.

- The depth map is retrieved by

$$d(\boldsymbol{x}) = \frac{c}{4\pi f_m}\phi(\boldsymbol{x}), \tag{7}$$

where c is the speed of light and f_m is the modulation frequency.

- Depending on the respective ToF camera, post-processing for correcting systematic errors is applied.

Let us now turn to the optimal location of the denoising method within the processing pipeline. The various positions within the pipeline, where total variation denoising and bilateral filtering can be applied, are

Smoothing the Raw Data: We apply the ROF model given by (2) and (3) and the bilateral filter to each of the four raw images to obtain the filtered images. Denoting the individual results by $\tilde{A}_j$, we then proceed in the processing pipeline with $\tilde{A}_j$ instead of A_j.

Filtering the Complex Data: In this approach, we consider the vector valued data

$$z(\boldsymbol{x}) = \begin{pmatrix} z_1(\boldsymbol{x}) \\ z_2(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} A_0(\boldsymbol{x}) - A_2(\boldsymbol{x}) \\ A_3(\boldsymbol{x}) - A_1(\boldsymbol{x}) \end{pmatrix}. \tag{8}$$

$z_1(\boldsymbol{x})$ and $z_2(\boldsymbol{x})$ can be interpreted as the real and imaginary part of a complex-valued signal $z(\boldsymbol{x})$. We have to keep in mind that the depth $d(\boldsymbol{x})$ we are actually interested in is related to the *phase* $\phi(\boldsymbol{x})$ of this complex signal $z(\boldsymbol{x}) = r(\boldsymbol{x})e^{i\phi(\boldsymbol{x})}$ by (7). For smoothing data $z(\boldsymbol{x})$, we consider again two alternatives, the bilateral filtering on vector-valued data and a TV-based approach consisting in the minimization of the objective function

$$\tfrac{1}{2}\left(\sum_{\boldsymbol{x}}(\|z_1(\boldsymbol{x}) - (A_0(\boldsymbol{x}) - A_2(\boldsymbol{x}))\|^2 + \|z_2(\boldsymbol{x}) - (A_3(\boldsymbol{x}) - A_1(\boldsymbol{x}))\|^2)\right) + \lambda\mathcal{R}(z), \tag{9}$$

with some regularization parameter $\lambda > 0$. As regularization term $\mathcal{R}(z)$ we choose isotropic total variation for vector valued data (see e.g. [50]). The term

isotropic here refers to the fact that the filtering in the complex domain does not favor any direction. As an alternative one could consider a filtering which smooths the phase stronger than the amplitude of the image. We refer to such an approach as anisotropic.

Combining the Cosine Fit with Spatial Regularization: Here the approach is to find $A(\boldsymbol{x}), \phi(\boldsymbol{x})$ and $I(\boldsymbol{x})$ minimizing

$$\left(\sum_{\boldsymbol{x}} \sum_{j=0}^{3} \left(\tfrac{A(\boldsymbol{x})}{2} \cos(\tau_j + \phi(\boldsymbol{x})) + I(\boldsymbol{x}) - A_j(\boldsymbol{x}) \right)^2 \right) + \mathcal{R}(A, \phi, I). \tag{10}$$

For the regularization term $\mathcal{R}$, we propose to consider the total variation of each of the unknowns independently, i.e.

$$\mathcal{R}(A, \phi, I) = \lambda_1 TV(A) + \lambda_2 TV(\phi) + \lambda_3 TV(I), \tag{11}$$

for some $\lambda_1, \lambda_2, \lambda_3 > 0$. Note that $\mathcal{R}(\cdot)$ couples the local optimization problems considered in (5). The optimization problem (10) has the advantage that the spatial regularity of the solution compensates for local distortions of the data A_j. The drawback of (10) is its non-convexity. The existence of a unique solution is not guaranteed and, even if, it is likely that the numerical optimization gets stuck in a local minimum. As a consequence, the retrieved numerical solution depends on the initialization and might not be the global minimum. The standard approach to cope with this non-convexity is to find a convex reformulation of the data term in (10) by applying a change of variables from (A, ϕ, I) to $(z, \overline{z}, I) = (\frac{A}{2} Z, \frac{A}{2} \overline{Z}, I)$, where $Z := e^{i\phi}$ (dependency on $\boldsymbol{x}$ omitted for simplicity). Then

$$\tfrac{A}{2} \cos(\tau_j + \phi) + I = \tfrac{1}{2}(e^{i\frac{\pi j}{2}} z + e^{-i\frac{\pi j}{2}} \overline{z}) + I, \qquad j = 0, \ldots, 3. \tag{12}$$

Moreover, standard calculus shows that the data term in (10) locally can be split into terms depending only on either z or I:

$$\sum_{j=0}^{3} \left(\tfrac{A}{2} \cos(\tau_j + \phi) + I - A_j \right)^2 = T_1(z) + T_2(I) + T_3, \tag{13}$$

where

$$T_1(z) := 2(Re(z) - \tfrac{1}{2}(A_0 - A_2))^2 + 2(Im(z) - \tfrac{1}{2}(A_3 - A_1))^2, \tag{14}$$

$$T_2(I) := 4(I - \tfrac{1}{4} \sum_{j=0}^{3} A_j)^2, \tag{15}$$

$$T_3 := \tfrac{1}{4} \sum_{j=0}^{3} A_j^2 - \tfrac{1}{2} \left(A_0 A_1 - A_0 A_2 + A_0 A_3 + A_1 A_2 - A_1 A_3 + A_2 A_3 \right). \tag{16}$$

In particular, (13) can be optimized with respect to z and I independently. We remark that we are mainly interested in z and $\phi = \arg(z)$. For z we retrieve the complex-valued data term already considered in (9). However, the regularization terms $\mathcal{R}(A, \phi, I)$ and $\mathcal{R}(z)$ differ. The strong advantage of (9) compared to (10) with respect to numerical treatment is the strong convexity of optimization problem. In particular, a unique solution is guaranteed.

Denoising the Depth Data: Finally, we consider the approach of filtering the depth data $d(\boldsymbol{x})$. This is the most commonly used strategy for denoising ToF data. Here we exemplarily consider total variation filtering, bilateral filtering and cross-bilateral filtering using both depth and intensity as input.

We remark that the approaches considered above differ in their numerical effort, which is approximately proportional on the number of channels (unknown variables) which have to be filtered. These are *four* in the case of filtering the raw data, *three* in the case of the cosine fit, *two* for filtering the complex data and *one* for smoothing the depth map. Thus, regarding numerical efficiency, the filtering of the depth map is preferable.

3.3 Restoration Quality

Since the basic aim of ToF cameras is to provide the depth of objects in the scene, the most important issue of filtering ToF data is to preserve the accuracy of the measured depth. This also concerns the location of depth edges, the depth difference at those edges and the optimal reconstruction of the slopes of surfaces.

Various techniques exist to improve given denoising schemes. We recall some particular, which concern the bilateral filter as well as the TV denoising approach. One important modification is to introduce *adaptivity* of the smoothing parameters. At edges, these parameters can be reduced to improve the edge preservation properties of the methods. This requires additional information about the edge location. For TV denoising, in particular, adaptivity of the regularization parameter significantly reduces the unfavorable loss of contrast.

Another way to improve denoising methods by local information is introduce directional dependency or *anisotropy* (also being a form of adaptivity). The basic idea goes back to the anisotropic diffusion approach presented in [5]. The aim is to provide a stronger smoothing parallel to edges than in normal direction. In the bilateral filter, the convolution mask can be made directionally depended. In the TV approaches, the regularization term can be made anisotropic, see e.g. [11]. In both cases, additional information on the location and orientation is required.

In particular, for TV approaches aiming at denoising depth data it has proven successful to include *second-order regularization* terms. Instead of piecewise constant data, these methods then favor piecewise planar structures.

We remark that with planar surfaces the following issue arises: As already mentioned above, ToF cameras provide the radial depth. After projection into 2D, planar 3D surfaces show up with a certain curvature. Using the z-depth reduces this projection effect. Thus, the model of piecewise planar, which second-order TV assumes, is fulfilled only approximately. The most accurate way to deal

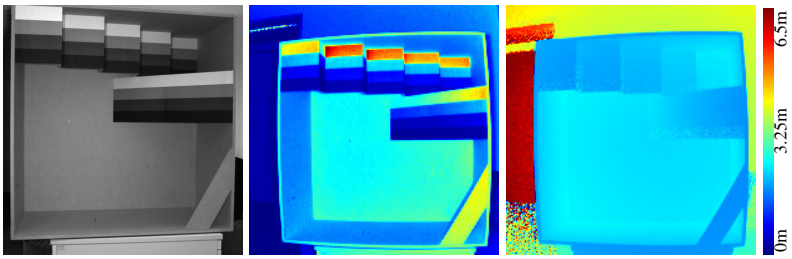


Fig. 2. The HCI box: recorded scene (left), ToF amplitude (middle) and ToF depth map (right) recorded wit a PMD Cam Cube 3

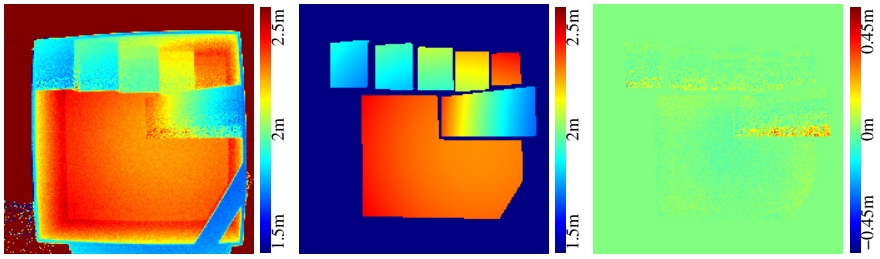


Fig. 3. Depth map with color bar clipped from 1.5 to 2.5 m, corresponding ground truth (dark blue areas are void) and difference image

with planar surfaces would be to directly work in 3D coordinates and consider the surface curvature of the objects, with the drawback that the numerical effort increases. However, the mentioned projection effect is relatively weak compared to systematic errors occurring in ToF data, such as the multi-path problem. Thus the dominant systematic errors should be tackled first before accounting for this effect.

For a detailed discussion on how edge information from both intensity/amplitude and depth data can be used to steer adaptivity, and for details on higher order TV denoising, we refer to [28].

4 Experiments and Evaluation

In this section we experimentally compare the methods presented in Section 3, applied at different positions in the processing pipeline.

As test data set, we use a recording of the *HCI box*[1] with a PMD Cam Cube 3, see Fig. 2. The box is made of medium-density fiberboard and shows different kinds of planar surfaces. Some of the surfaces are covered with paper

[1] `http://hci.iwr.uni-heidelberg.de/Benchmarks/document/hcibox/`

sheets painted in different gray tones, thus the reflectivity varies in the respective regions.

For our evaluation, we require ground truth to determine the error of each method considered. Therefore, we start with a discussion on appropriate ground truth and a description on how it is obtained. In addition, we refer to Chapter 4 for further discussion on this topic.

There is a virtual grid model of the box available, from which, after registration to the real scene, a synthetic depth map in view of the real camera can be rendered. Comparing, however, the recorded depth map with the synthetic one, the difference between both reveals not only the noise of the ToF camera but also all other kinds of systematic errors such as multi-path or an intensity dependent error. These systematic errors even dominate compared to the noise. Since the denoising methods considered above are not designed for the removal of all systematic errors, the difference between their result and the synthetic depth map will still be dominated by the systematic errors. As a consequence, the denoising capability can not be evaluated upon these differences.

We therefore use an alternative approach to obtain ground truth, such that the difference between ground truth and test data contains mainly noise. Here we make use of the fact that the HCI box consists of planar surfaces. We select those surfaces which are only weakly effected by the multi-path error. In particular, the side walls of the HCI box are left out for this reason.

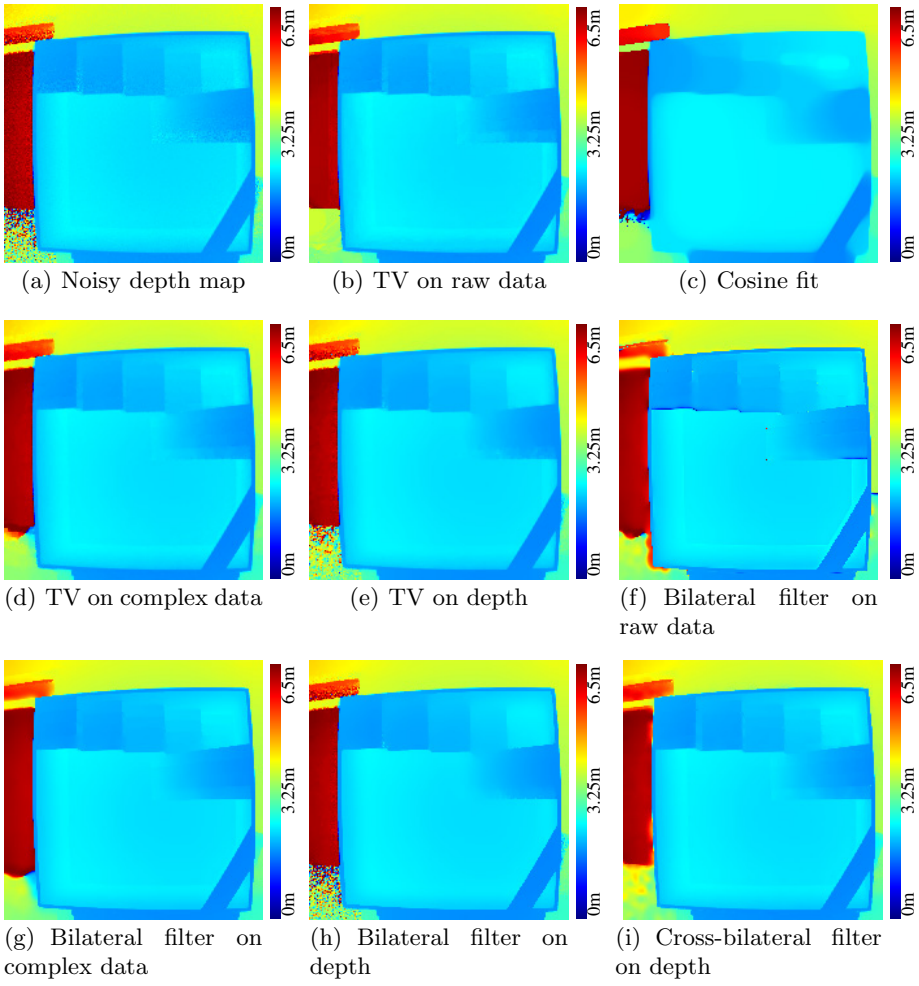
Note that, since the ToF Camera provides the radial depth to the camera center, these surfaces appear curved in the 2D depth maps. After projecting the 2D data back into 3D, a linear regression can be applied to approximate the noise free 3D surfaces. In the linear regression, regions of high noise due to low amplitude are disregarded. The ideal planar surfaces then can be projected back to retrieve the radial depth of the scene. Fig. 3 shows the result for selected regions in the depth map. We use the resulting depth in these regions as ground truth.

In order to have a fair comparison of the individual methods, care should be taken to choose the optimal parameters for each method. For our experiments we retrieve approximately optimal parameters for each method by means of the ground truth, which in practical applications of course is not at hand: for each method we seek for optimal parameters on a adaptively refined grid, so that the mean squared error (MSE) to the ground truth is minimized.

The results of the individual methods applied with these parameters are depicted in Fig. 4. Close-ups of an inner part of the HCI box are provided in Fig. 5. In addition we provide the MSE to the ground truth in Table 1.

We observe that the methods act differently on the background regions with strong noise. The cosine fitting as well as the bilateral filter applied to the depth data do barely smooth these regions at all. In order to further reduce this noise, a stronger smoothing would be preferable. Moreover, since the parameters where chosen to optimally reconstruct the planar areas where ground truth is provided, the restoration of edge regions are not as good as expected. Again, increase of the smoothing parameters would improve the regularity of the edges.

We recall that our objective is to compare methods with respect to their position in the processing pipeline. We therefore consider TV denoising and bilateral filter separately. When comparing the errors of the TV-based methods in Table 1, the cosine fitting clearly represents an outlier. This seems to be due to the non-convexity of the considered objective function, so that the optimization process most likely got stuck in a local minimum. Besides from this outlier, the TV-based methods show a clear trend. The MSE decreases the more the method is shifted to the end of the pipeline. Also the reconstruction of edges becomes better, the later the denoising methods is applied. It turns out, that the optimal strategy is to apply TV denoising at there very end of the pipeline.



(a) Noisy depth map (b) TV on raw data (c) Cosine fit

(d) TV on complex data (e) TV on depth (f) Bilateral filter on raw data

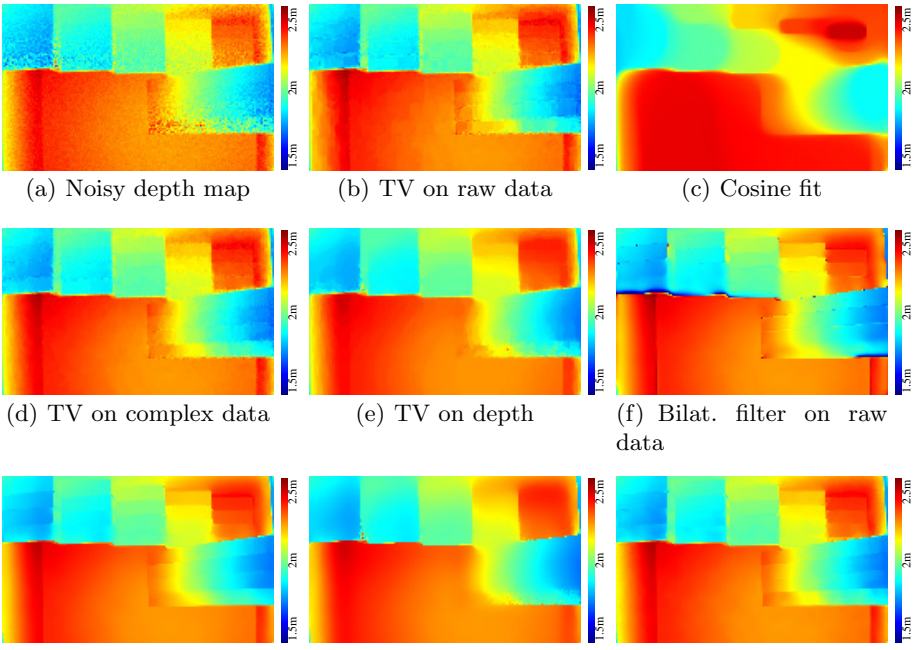
(g) Bilateral filter on complex data (h) Bilateral filter on depth (i) Cross-bilateral filter on depth

Fig. 4. Applying the bilateral filter and TV denoising at different positions of the processing pipeline. Besides the accurate restoration of surfaces (cf. MSE in Table 1) the removal of heavy noise (left region of the data) and a sharp reconstruction of edges is of importance

Concerning bilateral filter, the smallest MSE is achieved by applying the bilateral filter to the four raw channels. The result, however, reveals some distorted pixels, see Figs. 4 and 5. These distorted pixels result from the phase ambiguity after evaluating arctan(·). These distortions can be corrected by subsequently applying a median filter, which reduces the MSE further to $1.3524 \cdot 10^{-4}$. The second best result is provided by the bilateral filter applied to the depth data. Interestingly, the standard bilateral filter on the depth data slightly outperforms the cross-bilateral filter in terms of MSE.

Since the ground truth data only cover a part of the data set, it is inevitable to also compare the different variants in the remaining parts, especially at edges. Each of the three methods mentioned above shows a different kind of artifacts: the bilateral filter applied to the raw data shows some artifacts at the edges of the staircase, which might be due to flying pixels. The bilateral filter applied to the depth data in some regions (e.g. stairs) shows an over-smoothing, while in other regions (ramp) some noise remains. Finally, the cross-bilateral filter on the depth data provides a regular reconstruction of the true depth edges, while in the same time pronouncing false intensity-related edges. Our general conclusion is, that these three variants are competitive.

We remark that the above quantitative results are biased by the fact that we have chosen only one test scenario and that only partial ground truth is available.



(a) Noisy depth map (b) TV on raw data (c) Cosine fit

(d) TV on complex data (e) TV on depth (f) Bilat. filter on raw data

(g) Bilateral filter on complex data (h) Bilateral filter on depth (i) Cross-bilateral filter on depth

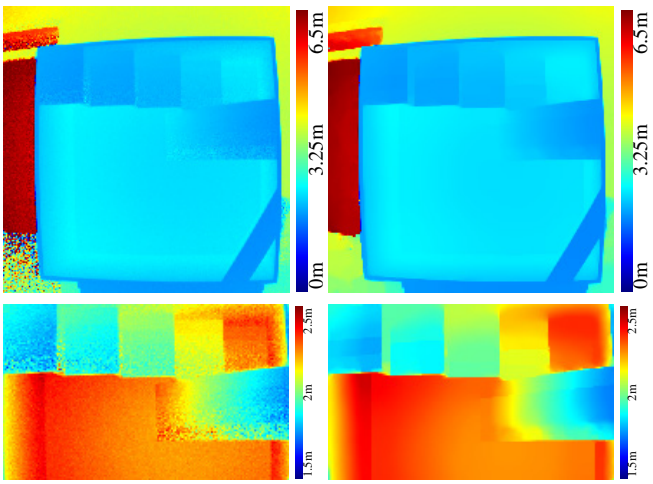
Fig. 5. Close-ups of the results in Fig. 4, where the bilateral filter and TV denoising are applied at different positions in the processing pipeline

Table 1. Mean squared error (MSE) to the ground truth (cf. Fig. 3) of the methods under consideration. The MSE strongly varies depending on the position within the processing pipeline. The bilateral filter on raw data with subsequent median filtering gives the smallest MSE.

Method	MSE ($\cdot 10^{-4}$)
Bilateral filter on raw data $A_j(x)$	1.4871
Bilateral filter on raw data plus median filter	1.3524
Bilateral filter on complex data $z(x)$	1.5444
Bilateral filter on depth map $d(x)$	1.5391
Cross bilateral filter on depth map $d(x)$	1.5819
TV denoising on raw data $A_j(x)$	1.6699
Non-convex cosine fit	7.1208
TV denoising on complex data $z(x)$	1.6320
TV denoising on depth map $d(x)$	1.5862

This stresses the need for larger data sets with highly accurate ground truth as well as a good error measure for evaluating the restoration of edges.

As mentioned in Section 3.3, additional strategies can be applied to improve the standard methods considered so far. We exemplarily consider the total variation denoising of the depth data to illustrate the potential of improvement of the methods considered so far. For TV denoising, in order to reduce the loss of contrast and prevent stair-casing, *anisotropic* total variation of first- and second-order can be applied. We refer to our work [51] for details on this approach. The result of this method is shown in Fig. 6. It achieves an MSE of $1.5105 \cdot 10^{-4}$ compared to $1.5862 \cdot 10^{-4}$ for the standard TV approach. For the bilateral filter corresponding modifications can be considered.



(a) input (b) filtered

Fig. 6. Applying adaptive first- and second-order TV on the depth map

5 Conclusion

This chapter started with an overview of state-of-the-art image denoising techniques as well as denoising algorithms especially designed for ToF data. Both image-driven and database-driven approaches were considered. As the central theme of this chapter, we discussed two alternatives for positioning the denoising algorithms in the data processing pipeline. Two well-established exemplary methods were considered and experimentally evaluated for this purpose: One is the *bilateral filtering* and the other is the *total variation-based denoising*. It turned out that for TV denoising the optimal position is at the end of the processing pipeline. For the bilateral filter, we found that applying it to the raw channels and performing a subsequent median filter provides the smallest quantitative error. Qualitatively, it competes with applying the bilateral and the cross-bilateral filter to the depth data. The general conclusion is, that the optimal position depends on the considered denoising method. As a consequence, for any newly introduced denoising technique, finding its optimal position within the pipeline is an issue which should be discussed along with the method.

Acknowledgements. This work is part of two research projects with the Intel Visual Computing Institute in Saarbrücken and with the Filmakademie Baden-Württemberg, Institute of Animation, respectively. It is co-funded by the Intel Visual Computing Institute and under grant 2-4225.16/380 of the Ministry of Economy Baden-Württemberg as well as the further partners Unexpected, Pixomondo, ScreenPlane, Bewegte Bilder and Tridelity. The content is under sole responsibility of the authors.

References

1. Aurich, V., Weule, J.: Non-linear Gaussian filters performing edge preserving diffusion. In: Proceed. 17. DAGM-Symposium (1995)
2. 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)
3. Elad, M.: On the origin of the bilateral filter and ways to improve it. IEEE Transactions on Image Processing 11(10), 1141–1151 (2002)
4. Perona, P., Shiota, T., Malik, J.: Anisotropic diffusion. In: Geometry-Driven Diffusion in Computer Vision, pp. 73–92. Springer (1994)
5. Weickert, J.: Anisotropic diffusion in image processing, vol. 1. Teubner Stuttgart (1998)
6. Donoho, D.L., Johnstone, J.M.: Ideal spatial adaptation by wavelet shrinkage. Biometrika 81(3), 425–455 (1994)
7. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D 60(1-4), 259–268 (1992)
8. Grasmair, M.: Locally adaptive total variation regularization. In: Tai, X.-C., Mørken, K., Lysaker, M., Lie, K.-A. (eds.) SSVM 2009. LNCS, vol. 5567, pp. 331–342. Springer, Heidelberg (2009)

9. Dong, Y.: Multi-scale total variation with automated regularization parameter selection for color image restoration. In: Tai, X.-C., Mørken, K., Lysaker, M., Lie, K.-A. (eds.) SSVM 2009. LNCS, vol. 5567, pp. 271–281. Springer, Heidelberg (2009)
10. Steidl, G., Teuber, T.: Anisotropic smoothing using double orientations. In: Tai, X.-C., Mørken, K., Lysaker, M., Lie, K.-A. (eds.) SSVM 2009. LNCS, vol. 5567, pp. 477–489. Springer, Heidelberg (2009)
11. Lenzen, F., Becker, F., Lellmann, J., Petra, S., Schnörr, C.: A class of quasi-variational inequalities for adaptive image denoising and decomposition. Computational Optimization and Applications, 1–28 (2013)
12. Bredies, K., Kunisch, K., Pock, T.: Total Generalized Variation. SIAM J. Imaging Sciences 3(3), 492–526 (2010)
13. Setzer, S., Steidl, G., Teuber, T.: Infimal convolution regularizations with discrete l1-type functionals. Comm. Math. Sci. 9, 797–872 (2011)
14. Lenzen, F., Becker, F., Lellmann, J.: Adaptive second-order total variation: An approach aware of slope discontinuities. In: Pack, T. (ed.) SSVM 2013. LNCS, vol. 7893, pp. 61–73. Springer, Heidelberg (2013)
15. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. Communications on Pure and Applied Mathematics 42(5), 577–685 (1989)
16. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the piecewise smooth mumford-shah functional. In: IEEE International Conference on Computer Vision (ICCV), Kyoto, Japan (2009)
17. Buades, A., Coll, B., Morel, J.: A review of image denoising algorithms, with a new one. Multiscale Model. Simul. 4(2), 490–530 (2005)
18. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. Multiscale Model. Simul. 7(3), 1005–1028 (2008)
19. Kindermann, S., Osher, S., Jones, P.: Deblurring and denoising of images by nonlocal functionals. Multiscale Model. Simul. 4(4), 1091–1115 (2005) (electronic)
20. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Transactions on Image Processing 16(8), 2080–2095 (2007)
21. Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising: Can plain neural networks compete with BM3D? In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012), pp. 2392–2399. IEEE (2012)
22. Jain, V., Seung, H.S.: Natural image denoising with convolutional networks. In: Advances in Neural Information Processing Systems, pp. 769–776 (2008)
23. Freeman, W.T., Jones, T.R., Pasztor, E.C.: Example-based super-resolution. IEEE Computer Graphics and Applications 22(2), 56–65 (2002)
24. Tappen, M.F., Russel, B.C., Freeman, W.T.: Exploiting the sparse derivative prior for super-resolution and image demosaicing. In: Proc. International Workshop on Statistical and Computational Theories of Vision (2003)
25. Kim, K.I., Kwon, Y.: Single-image super-resolution using sparse regression and natural image prior. IEEE Trans. Pattern Analysis and Machine Intelligence 32(6), 1127–1133 (2010)
26. Kim, K.I., Kwon, Y., Kim, J.H., Theobalt, C.: Efficient learning-based image enhancement: application to compression artifact removal and super-resolution. Technical Report MPI-I-2011-4-002, Max-Planck-Insitut für Informatik (February 2011)
27. Frank, M., Plaue, M., Rapp, K., Köthe, U., Jähne, B., Hamprecht, F.: Theoretical and experimental error analysis of continuous-wave time-of-flight range cameras. Optical Engineering 48(1), 13602 (2009)

28. Lenzen, F., Schäfer, H., Garbe, C.: Denoising time-of-flight data with adaptive total variation. In: Bebis, G. (ed.) ISVC 2011, Part I. LNCS, vol. 6938, pp. 337–346. Springer, Heidelberg (2011)
29. Schöner, H., Moser, B., Dorrington, A.A., Payne, A., 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)
30. 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)
31. Frank, M., Plaue, M., Hamprecht, F.A.: Denoising of continuous-wave time-of-flight depth images using confidence measures. Optical Engineering 48 (2009)
32. Edeler, T.: Bildverbesserung von Time-Of-Flight-Tiefenkarten. Shaker Verlag (2011)
33. 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)
34. Seitel, A., dos Santos, T.R., Mersmann, S., Penne, J., Groch, A., Yung, K., Tetzlaff, R., Meinzer, H.P., Maier-Hein, L.: Adaptive bilateral filter for image denoising and its application to in-vitro time-of-flight data, 796423–796423–8 (2011)
35. Schöner, H., Bauer, F., Dorrington, A., Heise, B., Wieser, V., Payne, A., Cree, M.J., Moser, B.: Image processing for 3d-scans generated by time of flight range cameras. SPIE Journal of Electronic Imaging 2 (2012)
36. Schuon, S., Theobalt, C., Davis, J., Thrun, S.: Lidarboost: Depth superresolution for tof 3d shape scanning. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 343–350. IEEE (2009)
37. Mure-Dubois, J., Hügli, H., et al.: Fusion of time of flight camera point clouds. Workshop on Multi-camera and Multi-Modal Sensor Fusion Algorithms and Applications-M2SFA2 2008 (2008)
38. Edeler, T., Ohliger, K., Hussmann, S., Mertins, A.: Super resolution of time-of-flight depth images under consideration of spatially varying noise variance. In: 16th IEEE Int. Conf. on Image Processing (ICIP), Cairo, Egypt, pp. 1185–1188 (November 2009)
39. Chan, D., Buisman, H., Theobalt, C., Thrun, S., et al.: A noise-aware filter for real-time depth upsampling. In: Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications-M2SFA2 2008 (2008)
40. Huhle, B., Schairer, T., Jenke, P., Straßer, W.: Robust non-local denoising of colored depth data. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2008, pp. 1–7. IEEE (2008)
41. Yeo, D., Kim, J., Baig, M.W., Shin, H., et al.: Adaptive bilateral filtering for noise removal in depth upsampling. In: 2010 International SoC Design Conference (ISOCC), pp. 36–39. IEEE (2010)
42. Park, J., Kim, H., Tai, Y.W., Brown, M.S., Kweon, I.: High quality depth map upsampling for 3d-tof cameras. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 1623–1630. IEEE (2011)
43. Reynolds, M., Doboš, J., Peel, L., Weyrich, T., Brostow, G.J.: Capturing time-of-flight data with confidence. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, pp. 945–952. IEEE (2011)
44. Mac Aodha, O., Campbell, N.D.F., Nair, A., Brostow, G.J.: Patch based synthesis for single depth image super-resolution. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 71–84. Springer, Heidelberg (2012)

45. Kopf, J., Cohen, M.F., Lischinski, D., Uyttendaele, M.: Joint bilateral upsampling. In: ACM SIGGRAPH 2007 Papers. SIGGRAPH 2007. ACM, New York (2007)
46. Eisemann, E., Durand, F.: Flash photography enhancement via intrinsic relighting. ACM Transactions on Graphics (TOG) 23, 673–678 (2004)
47. Petschnigg, G., Szeliski, R., Agrawala, M., Cohen, M., Hoppe, H., Toyama, K.: Digital photography with flash and no-flash image pairs. ACM Transactions on Graphics (TOG) 23, 664–672 (2004)
48. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision 40(1), 120–145 (2011)
49. Schmidt, M.: Analysis, Modeling and Dynamic Optimization of 3D Time-of-Flight Imaging Systems. Dissertation, IWR, Fakultät f ür Physik und Astronomie, Univ. Heidelberg (2011)
50. Blomgren, P., Chan, T.F.: Color tv: Total variation methods for restoration of vector-valued images. IEEE Transactions on Image Processing 7(3), 304–309 (1998)
51. Schäfer, H., Lenzen, F., Garbe, C.S.: Depth and intensity based edge detection in time-of-flight images. In: Proceedings of 3DV. IEEE (in press, 2013)