

Denoising of continuous-wave time-of-flight depth images using confidence measures

Mario Frank

Department of Computer Science, ETH Zurich
Haldeneggsteig 4, 8092 Zurich, Switzerland
Phone: +41 44 632 3179, Fax: +41 44 632 15 62, mario.frank@inf.ethz.ch

Matthias Plaue

Department of Mathematics, Technische Universität Berlin
Straße des 17. Juni 136, 10623 Berlin, Germany
Phone: +49 30 314 232 65, Fax: +49 30 314 792 82, plaue@math.tu-berlin.de

Fred A. Hamprecht[†]

Heidelberg Collaboratory for Image Processing, University of Heidelberg
Speyerer Straße 4, 69115 Heidelberg, Germany
Phone: +49 6221 54 88 75, Fax: +49 6221 54 52 76, fred.hamprecht@iwr.uni-heidelberg.de

[†]corresponding

Abstract

Time-of-flight range sensors with on-chip continuous-wave correlation of radio frequency modulated signals are increasingly popular. They simultaneously deliver depth maps and intensity images with noise and systematic errors that are unique for this particular kind of data. Based on recent theoretical findings on the dominating noise processes we propose specific variants of normalized convolution and median filtering, both adaptive and non-adaptive, to the denoising of the range images. We examine the proposed filters on real-world depth maps with varying reflectivity, structure, over-exposure, and illumination. The best results are obtained by adaptive filters that locally adjust the level of smoothing using the estimated modulation amplitude as a measure of confidence.

1 Introduction

Time-of-flight (TOF) 3D cameras using radio frequency (RF) modulated light sources can be applied to a whole array of problems, ranging from object tracking [1, 2], object detection and recognition [3, 4], traffic [5, 6] to robotics and automated production [7].

Since all available prototypes as well as future devices deliver range information that is subject to noise and systematic errors, for most applications, denoising/smoothing of the range images is necessary to produce reliable data for further processing. Systematic errors can be tackled by calibration [8, 9] or by avoiding the responsible mechanisms, for example by suppressing backlight illumination [10]. In this article we investigate methods for the denoising of TOF depth maps which are specially tailored to this particular kind of data. All assessed methods make use of recent theoretical findings [11] for these sensors.

The range data of the investigated 3D cameras is acquired as follows. Infrared light, modulated with radio frequency, is emitted by an LED array and reflected by objects at distance d in the scene observed by the camera. After the time of flight $t_d = \frac{2d}{c}$ the reflected light is detected on the chip by a set of n gates per pixel. These gates correlate the signal shortly after charge generation with a reference signal of the same frequency but phase-shifted with $\alpha_n = \frac{2\pi n}{N}$, $n = 0 \dots N - 1$. The results of these hardware based autocorrelation functions (ACF) give one intensity I_n for each α_n and for each pixel. From these intensities, the modulation amplitude A of the optical signal and the phase delay ϕ_d corresponding to t_d is computed with

$$A = \frac{2}{N} \left| \sum_{n=0}^{N-1} I_n e^{-2\pi i \frac{n}{N}} \right|, \quad \phi_d = \text{arg} \left(\sum_{n=0}^{N-1} I_n e^{-2\pi i \frac{n}{N}} \right). \quad (1)$$

Finally, the depth d is obtained from the phase delay by $d = \frac{c}{4\pi f} \phi_d$. Technical details of the 3D imaging system, the RF-modulated TOF method and the phase-shifting technique are discussed extensively in [12, 13]. The underlying phase estimation problem is analyzed in [14, 15, 11] and the most important results are summarized above without derivation. The resulting depth values are biased and can be subject to strong noise due to various effects. One straightforward approach that is frequently used in the context of TOF correlation detectors is to simply discard all depth measurements whose signal amplitude lies below a certain threshold. However, the information of these measurements is lost completely and the resulting depth image may contain holes that make further processing difficult.

Much previous work has focused on the denoising of 2D scalar images. Anisotropic diffusion [16, 17], wavelet denoising [18] (in particular wavelet shrinkage [19, 20]), non-local means [21, 22], total variation [23], Wiener filter [24, 25] or adaptive linear and non-linear filters [26] are well-established smoothing techniques, to only name the most prominent ones. There exist powerful (albeit computationally costly) smoothing procedures especially designed for surfaces or inspired by surface theory that use various geometric flows like mean curvature flow [27, 28], Beltrami flow [29] or Willmore flow [30]. These measures optimize certain functionals, and many other variational frameworks for surface smoothing and reconstruction have been proposed (cf. [31], also for other techniques). In [32], such a variational method is used to reconstruct noisy range data by combination with gray-level data – a problem very similar to the one described in this paper. However, we cannot expect the proposed shape-from-shading algorithm to function with TOF images of real-world objects with varying IR reflectivity.

As was noted in [33], the restoration error of various denoising methods can be quite similar where other criteria such as aesthetics often play a fundamental role in real-world applications. Furthermore, in such applications non-parametric methods are desired to minimize the need for user interaction, and limited computational resources must also be considered.

In this contribution, we focus on a limited class of low-complexity methods that allow one to explicitly take the uncertainty of the measurement in each individual pixel into account. Any such adaptive image denoising algorithm has to consider the respective noise model (see for example [34]). Fortunately, the measurement errors and accuracy of TOF systems have been studied extensively [35, 36, 11], and as an important result, the standard deviation of the range data was found to be reciprocal to the IR modulation signal amplitude [11]. Thus, an adaptive filter can be employed with the modulation amplitude as confidence value. In particular, we chose the normalized convolution as the basic technique, the optimality of which (in the mean-square sense) was proven in [37].

Our goal is to suggest and evaluate filtering methods that are tailored to this particular kind of data. To this end, we compare simple approaches that all have the common aim to make depth images more reliable while preserving as much depth information as possible. The principal idea of our method is to smooth each pixel to an extent that is determined by an estimation of its reliability. In this way, the values of reliable pixels are preserved more faithfully while pixels with a suspected high variance are rigorously corrected.

Section 2 gives a short introduction to the estimation of the reliability of the range values and introduces our denoising methods and their underlying theory.

In section 3, their performance on various scenes is evaluated and extensively discussed. Our experimental findings are summarized in section 4. Finally, we conclude in section 5.

2 Confidence Estimation and Denoising Strategies

In [11], we have shown that the probability distribution of the phase follows a special instance of the Offset Normal distribution

$$G_c(\phi_d; A/\sigma) = \frac{1}{2\pi} \exp\left[\frac{-A^2}{2\sigma^2}\right] \left[1 + \sqrt{\frac{\pi}{2}} \frac{A}{\sigma} \cos(\phi_d) \exp\left[\frac{A^2 \cos^2(\phi_d)}{2\sigma^2}\right] \left(1 + \operatorname{erf}\left[\frac{A \cos(\phi_d)}{\sqrt{2}\sigma}\right]\right)\right] \quad (2)$$

with the same quantities as defined in section 1. Here, σ is the standard deviation of the raw intensities I_n which are acquired directly on the chip. The spread of this particular instance of the Offset Normal distribution is $\frac{\sigma}{A}$. Hence, for a fixed standard deviation σ of the sensitivity of the gates to a physical light signal, this theoretical result establishes a relation between the variance σ_d^2 of the estimated range and the physical amplitude A of the modulation signal: $\sigma_d^2 \propto \frac{1}{A^2}$. Therefore, the amplitude is an optimal estimator for the reliability of a measurement. Moreover, since the modulation amplitude is calculated along with the distance at very little extra effort this estimator also comes at low computational cost.

Based on this result we study a set of non-parametric smoothing techniques that use the modulation amplitude of the signal to determine to what extent each image region should be smoothed in order to obtain a particular confidence of the result. The selection is restricted to non-iterative, non-parametric approaches and can be divided into median-based filters and those using normalized convolution. In the following, all used filters are described. In section 3 we report the performance of these methods measured by comparative assessment.

2.1 Normalized Convolution using Confidence Values

The simplest way to incorporate a measure of confidence to the smoothing of range maps is to weight each pixel with its confidence estimate during the averaging process: depth values that are more reliable contribute more to the average depth while unreliable pixels have lower impact. Moreover, we want to take the spatial relationship of the pixels into account. Therefore, in addition to our confidence measure, we also use a Gaussian kernel as weighting factor. As reasoned

above, the variance of the depth information is proportional to the inverse square of the amplitude. Thus, the amplitude is a well-suited measure of confidence. In many cases, other errors such as quantization noise and saturation effects due to overexposure also reveal themselves by a vanishing modulation amplitude¹ [11].

The pixel (i, j) of the smoothed depth image $d_{h;i,j}$ is computed from the raw depth map d with

$$d_{h;i,j} = \frac{\sum_{k_i=-\frac{n-1}{2}}^{\frac{n-1}{2}} \sum_{k_j=-\frac{n-1}{2}}^{\frac{n-1}{2}} d_{i+k_i,j+k_j} \cdot f_{k_i,k_j}^h \cdot A_{i+k_i,j+k_j}^2}{\sum_{k_i=-\frac{n-1}{2}}^{\frac{n-1}{2}} \sum_{k_j=-\frac{n-1}{2}}^{\frac{n-1}{2}} f_{k_i,k_j}^h \cdot A_{i+k_i,j+k_j}^2} \quad (3)$$

where f_{k_i,k_j}^h are the coefficients of the smoothing mask with bandwidth h , and n is an odd mask size. In our experiments, a Gaussian kernel with fixed bandwidth (standard deviation) of one third of the mask size n was used as filter f_{k_i,k_j}^h . The net effect of eq. 3 is to give more importance to those pixels for which a more reliable depth estimate is available. In all figures, this filter is abbreviated as Weighted Gaussian (WG).

2.2 Complex Normalized Convolution

The most natural way to represent the raw data acquired with the 4-phase-shifting technique is a vector-valued 2D image containing the measured intensity differences $(I_0 - I_2, I_3 - I_1)$. Regarded as a complex-valued image, the actual range image is proportional to the argument, and the amplitude A is the magnitude. It is a reasonable approach to work on this original directional data rather than the range image computed from it:

$$\phi_{h;i,j} = \arg \left(\sum_{k_i=-\frac{n-1}{2}}^{\frac{n-1}{2}} \sum_{k_j=-\frac{n-1}{2}}^{\frac{n-1}{2}} \exp(i\phi_{i+k_i,j+k_j}) \cdot f_{k_i,k_j}^h \cdot A_{i+k_i,j+k_j}^t \right) \quad (4)$$

For $t = 1$, the above formula describes a Gaussian filter applied to each pairwise difference of the raw intensities and subsequent computation of the phase/range. We have also computed this directional average for $t = 2$ in order to stronger penalize those values with high variance. The latter method shows similar results compared to the filter introduced in the previous subsection if adjacent pixels do not differ more than π in phase. In all figures, this filter will be denoted depending on the exponent t of the weighting factor either as “complex” or “ A^2 -complex”.

¹Unfortunately, the observed amplitude can still be high at the onset of overexposure.

2.3 Adaptive Normalized Convolution

The methods described so far can be extended to filters that locally adapt to the data quality. A real scene typically consists of regions for which the depth can be estimated with different reliabilities. An ideal filter should smooth each region only to the extent that is truly required by a specific region to obtain an estimate with variance smaller than some user-selected threshold. To this end, each pixel is first weighted with its inverse variance (with A^2), and the image is then convolved with several Gaussian kernels of different width h as shown in eq. (3). Assuming spatially uncorrelated noise, the estimated variance of a smoothed pixel is now a function of the bandwidth [38]:

$$\sigma_{h;i,j}^2 = \frac{\sum_{k_i=-\frac{n-1}{2}}^{\frac{n-1}{2}} \sum_{k_j=-\frac{n-1}{2}}^{\frac{n-1}{2}} \left(f_{k_i,k_j}^h \cdot A_{i+k_i,j+k_j} \right)^2}{\left(\sum_{k_i=-\frac{n-1}{2}}^{\frac{n-1}{2}} \sum_{k_j=-\frac{n-1}{2}}^{\frac{n-1}{2}} f_{k_i,k_j}^h \cdot A_{i+k_i,j+k_j}^2 \right)^2} \quad (5)$$

For each pixel, we choose the depth value computed from eq. (3) whose corresponding new variance – estimated with eq. (5) – is the highest variance below a user-defined threshold σ_{thresh}^2 .

This method causes every pixel to be averaged over only as many neighbors as are strictly required to obtain sufficient “confidence”. If this criterion is reached with the smallest width for a particular pixel, it will not be smoothed at all, if the criterion is not reached even with the largest mask, then the result obtained with maximal width is taken. A maximal width up to one third of the mask size is allowed in order to prevent discontinuity at the tails of the Gaussian. The number of convolutions with different Gaussians f^h determines how densely scale space is sampled. This is a crucial parameter regarding computational costs. Below we refer to this spatially adaptive filter as the “Adaptive Weighted Gaussian” (AWG) filter.

2.4 Median and Adaptive Median

In addition to the filters using variants of normalized convolution, the classical median was used as well. In order to be able to locally adapt to the image quality, we also implemented the Hampel Detector [39]. It is an adaptive non-linear filter that, in analogy to the standard deviation, computes the spatial median absolute deviation (MAD) from a median image d_{med_n} in an $n \times n$ neighborhood

$$\text{MAD}_n = \text{med}_n |d - d_{med_n}| \quad (6)$$

Using this robust measure of the outlyingness for each pixel, the filter smooths with a spatial median only those pixels with a MAD value lying above a user-defined threshold. This can be regarded as another way of allowing for a certain surface roughness in the picture. However, this filter cannot distinguish a rough surface from an even but noisy surface since it uses no measure of confidence other than the depth data itself. In order to overcome this limitation, the Hampel detector can be modified to better suit the features of the particular data type at hand: instead of the MAD value, the modulation amplitude A can be used as a threshold. In section 3 we report on the assessment of all three filters for several threshold parameters (A or MAD) and mask sizes n .

3 Data and Experiments

In this section, we assess the proposed methods on real-world data. In order to account for the influence of different conditions, all experiments were done on a set of complex indoor scenes. In addition to the quantitative analysis, the figures shown in this section (Fig. 2 and Fig. 3) allow for qualitative judgment by visual inspection. As follows, we describe the experimental setting and the evaluation methods. Then, we report on our results and discuss the observed effects in detail.

3.1 Real-World Data

We used various static scenes such that the quality of the results is expected to depend on the features of the image itself. All scenes were acquired with the PMD[vision] 19k camera from PMDtec in closed rooms, and most of them under the exclusion of daylight. One scene shows a flat surface without much structure (a locker) at close distance, another one is composed of a box on a carpet of low reflectivity with distant background. In order to observe the impact of the filters on a fine depth structure, an entangled hose was filmed as well (Fig. 1 right). Here the background is a mixture of distant objects and very close pillars. There is much fine structure in this depth map like cables and neighbored twists of the hose, and also sharp and deep edges. Other scenes were a board with numerous items and finally a whole room (Fig. 1 left) with a selection of bigger objects placed over the whole unambiguous range of 7.5m, such that illumination reaches from overexposure to insufficient illumination (both depending on the integration time). The latter scene also has a region where daylight is reflected on a surface close to an open window. Various integration times were used at every scene in



Figure 1: Pictures of two of the scenes used for filter assessment. Depth maps and amplitude images together with filter results and additional information can be seen in Fig. 2 and Fig. 3. The color pictures shown here may not exactly match the depth images.

order to study the filters impact on overexposed regions and suitable illuminated areas as well as on low amplitude regions. The results shown in all figures refer to the two scenes shown in Fig. 1.

3.2 Evaluation Methods

3.2.1 Reference Depth Map

We want to investigate the performance of a particular filter on an image with given features, ideally with respect to a ground truth. We have obtained this reference depth map by taking the mean over a large number of frames from a static scene. Initially, the median across frames was used because this method is expected to be more robust. For reasons explained in section 4.4 this method turned out to be inappropriate for this particular kind of data.

3.2.2 Extended Comparison

The comparison methods used here were introduced in [40]. A one-dimensional measure of how well a filter performs is to take the mean absolute deviation from the reference depth map. For added insight, the performance can be regarded as a function of the uncertainty at a pixel, i.e. the difficulty in obtaining a good ground truth estimate of that pixel. This uncertainty grows with the spread of all depth

values observed at a given pixel over various frames. In contrast to [40] where the MAD of the depth across frames was used as an uncertainty measure, we have decided to use the inverse amplitude of the modulation signal since it is proportional to the standard deviation of the estimated distance and can be obtained as a byproduct from the camera for each pixel.

The distribution of the deviations of any image from the ground truth can now be summarized in a two-dimensional histogram (Fig. 6) mapping the inverse amplitude (the “uncertainty”) in one direction and the absolute deviation from the reference depth map in the other direction. This allows to assess the “easy” and “difficult” regions performance separately, and provides information on how serious the errors are in a particular type of region. Two filter results can be compared directly by subtracting their histograms from each other. Finally, accumulating all absolute deviations at a given inverse amplitude allows us to compare more than two filters at once (see, for instance, Fig. 5).

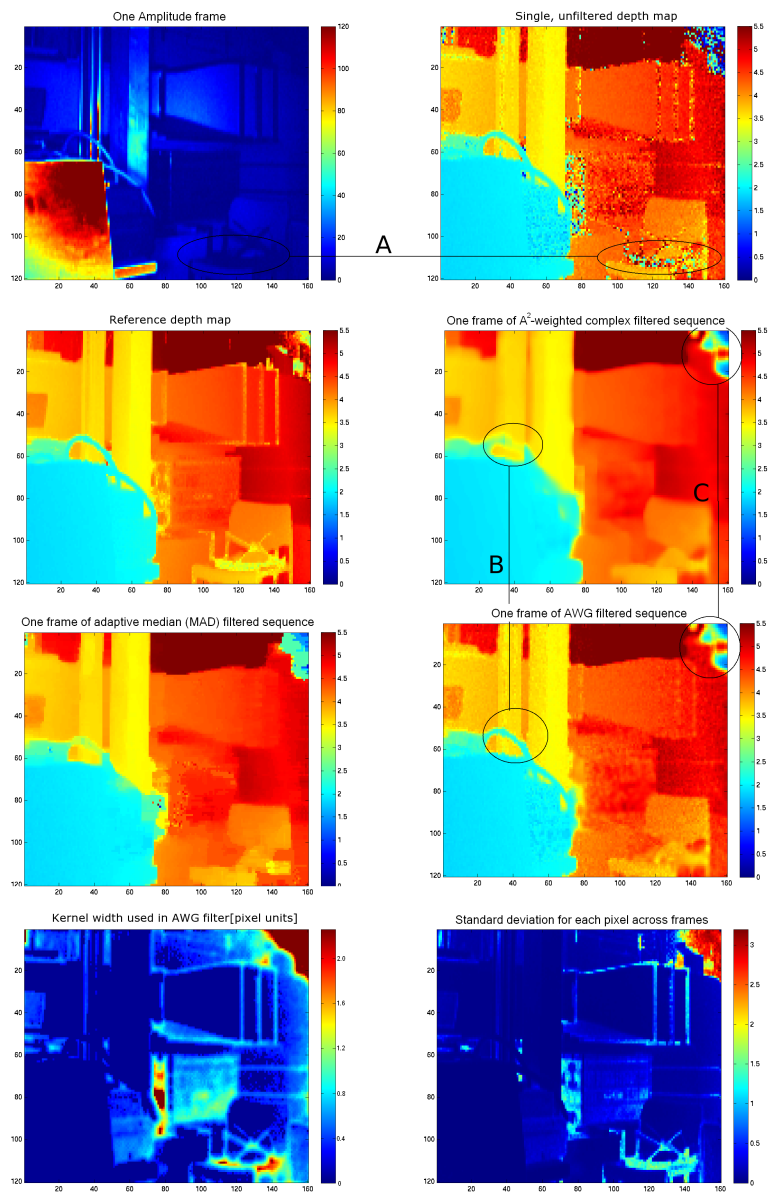


Figure 2: Results for a static scene with regions of sufficient and insufficient illumination. On the top row the amplitude image and the unfiltered depth image are shown. Then come the reference depth map and the filter results for a 7×7 filter mask. In the bottom row the Gaussian width that was used for the Adaptive Weighted Gaussian filter and the standard deviation over all frames is shown. Color scales of the depth images are in [m].

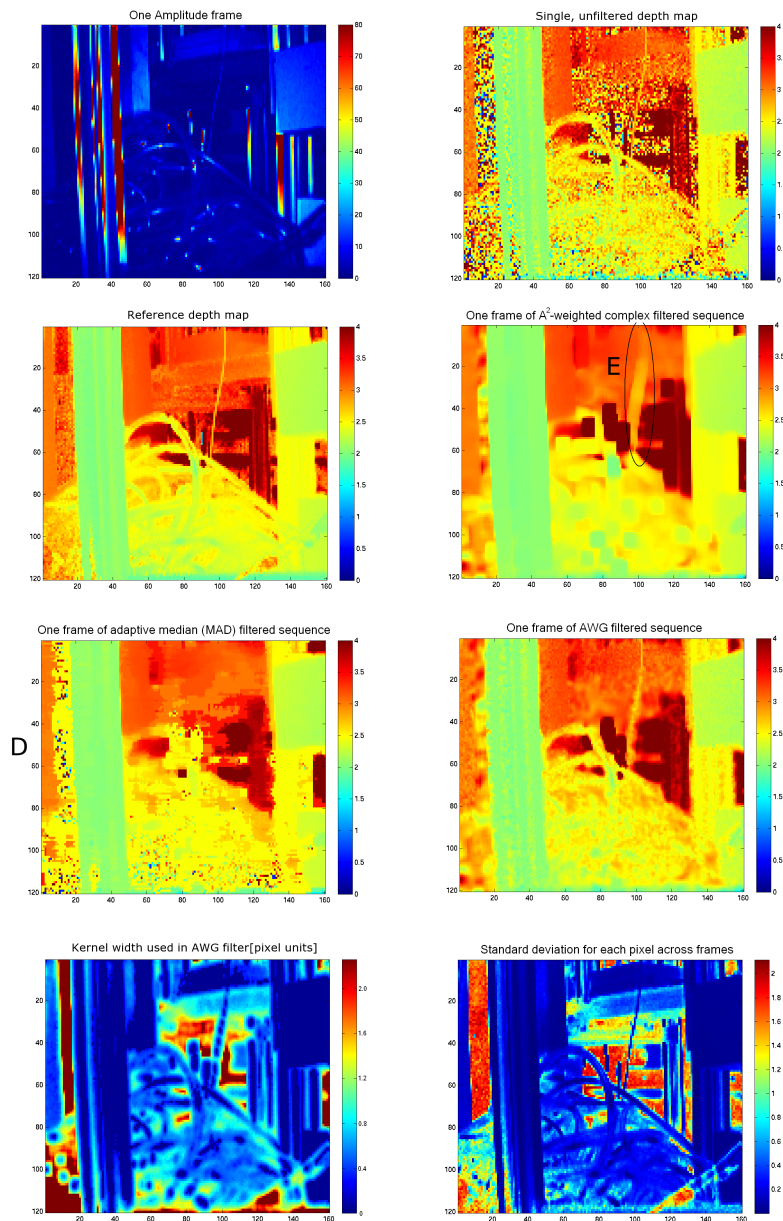


Figure 3: Results for a static scene with regionally poor illumination. The order of the images is the same as in Fig. 2. Again, color scales of the depth images are in [m]. Bottom left: Note that there are whole regions filtered with maximum width (red) as well as sufficiently illuminated regions not smoothed at all (blue). The thin dangling cable is bloated by the non-adaptive WG filter (region marked "E"), and falsely eliminated by the MAD filter ("D").

4 Results and Analysis

For the two scenes described above, the filter results are illustrated in Fig. 2 and Fig. 3. Both results were obtained by running the filters with a parametrization that was optimal in terms of the average error per pixel (epp) for the particular scene. In both figures, the top left image shows the amplitudes of a single frame. Similar to a spotlight, the illumination fades radially from the center since the light source can be approximated as a point (for larger distances). The objects have highly varying reflectivities (note the chair cushion marked by “A”) which makes computing the correct depth very challenging. Comparing this with a single depth frame of the original scene (top right), one can directly observe a relation between the amplitude and the confidence. In this context, also note the standard deviation across repeated measurements of the scene (bottom right). The bottom left picture shows the width of the Gaussian that was used with the Adaptive Weighted Gaussian (AWG) filter. Note that there are whole regions filtered with the maximum width of one third of the full mask size (red) as well as sufficiently illuminated regions not smoothed at all (blue).

The reference depth map is shown on the left side of the second row. It was constructed by taking the mean over 300 repeated measurements of the static scene. Adjacent to this are the results of three filters. The outcome of the AWG and the complex WG filter can be best compared by using the regions marked with B and C together with the bottom left picture. AWG and WG give exactly the same result for pixels smoothed with maximal width (red) and differ more and more as the AWG filter uses smaller widths. The adaptive median filter using MAD as the threshold performs well in preserving the edges but eliminates small structures (compare markers D and E in Fig. 3). Moreover, it leaves surprisingly numerous outliers that intuitively should have been removed (see Fig. 3 D). We will refer to this effect later.

4.1 The Effect of Directed Blurring/Dilation

All filters using the amplitude as a measure of confidence have to cope with artifacts that appear as directed blurring at the edges. This problem is due to the fact that the amplitude is itself a function of distance: Objects further away have a much lower amplitude than objects close to the camera. Spatially adjacent pixels may show objects of different distance, and as a result, of different amplitudes. When convolving with $A^t; t > 0$ as a weighing factor, the pixels with smaller depth value and larger amplitude will always have a much higher impact on the

result: closer objects grow at their edges and occlude the weakly illuminated objects in the background. The impact of the effect grows with t and the mask size. This can be seen in Fig. 3 where the cable (marker E) and the pillow are much broader after filtering. Thus, in the context of the described effect, the selection of t is a crucial choice. In assessing the complex filters introduced in section 2.2 we observed that weighting with A^2 performs better than using A . This is due to the fact that the former method penalizes values exactly according to the variance. However, with mask sizes of 11×11 or higher (and therefore also with higher maximal Gaussian width, which is restricted to one third of the mask size) the boosted effect of “directed blurring” produces errors that outbalance the benefit of this penalization. This effect can dominate the mean absolute error per pixel (epp) obtained from comparing the resulting image with a reference depth map as described in section 3.2.1.

The choice of the variance threshold for AWG and the maximal allowed Gaussian width have a large impact. One could argue that the AWG filter should allow for a higher maximal width in order to decrease the often very high fraction of pixels convolved with that width. However, one has to consider that scenes with deep depth edges would be strongly distorted, then. Since we restrict the maximal Gaussian width to one third of the mask size n we can observe this effect by comparing filters with different mask sizes. Fig. 4 shows the average error per pixel (epp) against the AWG variance threshold for various mask sizes. With increasing variance threshold, the fraction of pixels smoothed with maximal Gaussian width decreases to a small number of pixels having exactly zero amplitude. These pixels are responsible for the constant differences at the right tails between curves for different mask size n (top left in Fig. 4) because in this regime the AWG filter always smooths with maximal allowed width, which increases with n . At very small variance threshold, almost every pixel is smoothed with maximal width and thus the error of larger masks rises very fast with vanishing threshold.

4.2 Runtime

Depending on the application, the computation time may be a very important factor for the choice of a particular filter. The fixed width filters are the fastest, with the complex one being a bit slower due to the conversion from depth values to the complex plane. Apart from the mask size, their runtime is independent of any parameters and of the image quality. The same holds for the median filter. The AWG filter does not depend on the image quality either (if implemented noniteratively) but depends strongly on the number of performed convolutions between

minimal and maximal Gaussian width in the scale selection process. There is a tradeoff between speed and ensuring that every pixel is filtered with its appropriate Gaussian kernel. Experience shows that it suffices to sample the scale space at only a few Gaussian widths to achieve satisfactory results. The reason is that the amplitude varies a lot such that most pixels are either filtered with minimal or maximal width and only a few need to be filtered with intermediate kernel sizes.

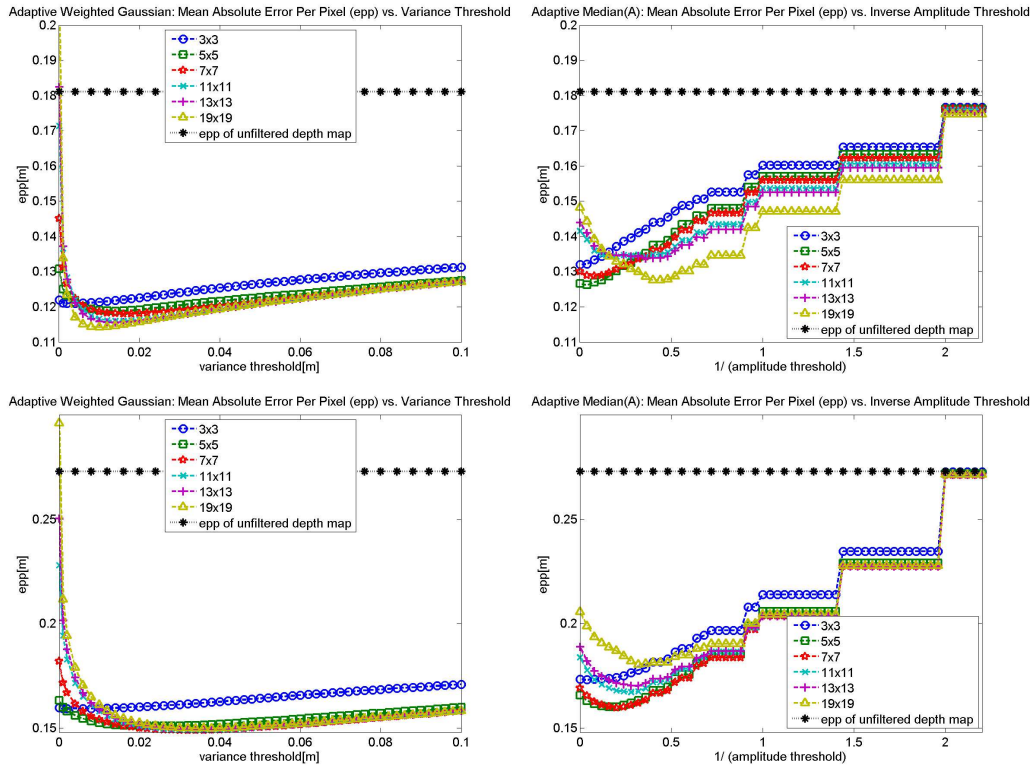


Figure 4: The mean absolute Error Per Pixel (epp) as a function of the threshold for the AWG and the Adaptive Median filter (weighted by amplitude A). The graphs within a row can be directly compared since they have the same scale for epp. The results of the first row were obtained by smoothing a room scene with mixed illumination and big objects (see Fig. 2). The second row shows the results from a highly structured scene (see Fig. 3). The steps in the right graphs result from the fact that higher values of $\frac{1}{A}$ correspond to small changes of the threshold A . Since the amplitude in the images is quantized, a small change of the threshold has no impact on the filter behavior over somewhat large intervals of $\frac{1}{A}$.

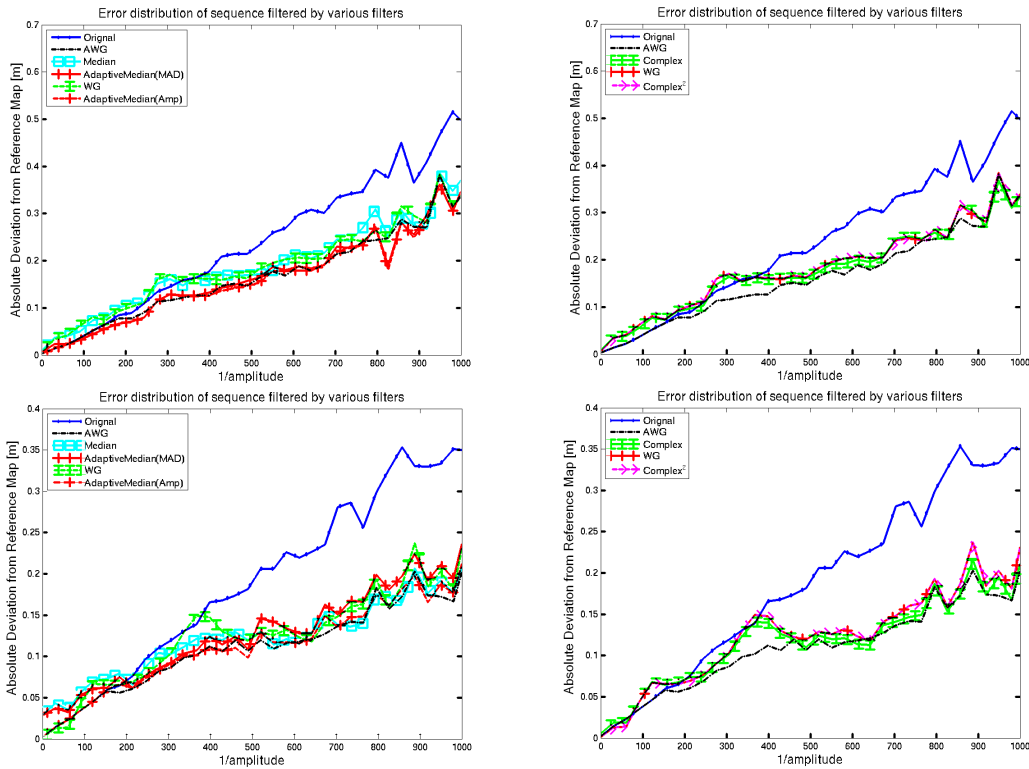


Figure 5: The absolute deviation from the reference depth map versus the expected standard deviation ($\frac{1}{A}$). Top row: results from the scene showing a room (Fig. 2). Left: The adaptive filters lead to similar results. In the second row are results from the scene showing a hose (Fig. 3). Right: The WG graph and the $complex^2$ graph are equal. For high amplitudes the filters with fixed width lead to errors higher than the original image. The composition of the error can be analyzed in more detail by employing Fig. 6.

Both the adaptive median filter using the spatial MAD value in order to decide if to filter or not, and the one using the amplitude, depend on the image quality itself (if implemented such that the unnecessary computations for “good” pixels are avoided). “Difficult” images are filtered more intensively and therefore need more time. The amplitude-controlled median is always faster since its variance measure is available for free in terms of computational costs. All computations were performed with MATLAB on an Athlon 64 2.2 GHz processor with 2 GB memory. The most illustrative results are summarized in Tab. 1 and Tab. 2.

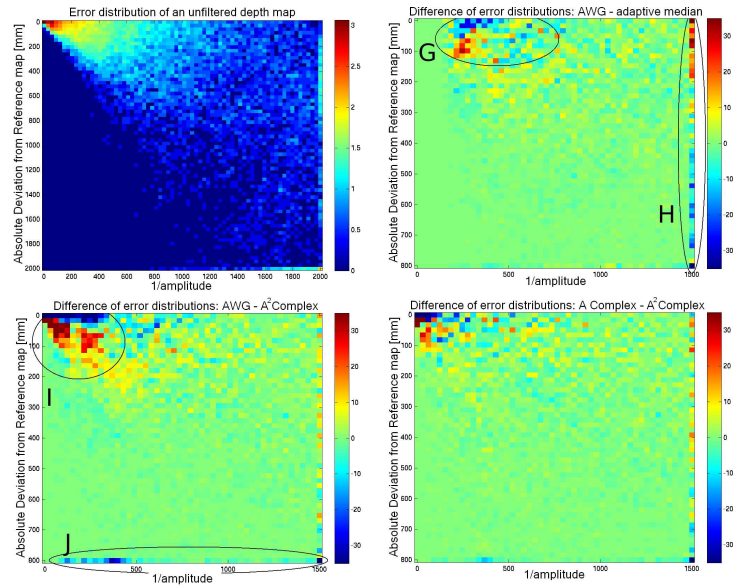


Figure 6: Top left: error histogram of the original hose scene (Fig. 3) with logarithmic frequencies (color scale). At very high amplitudes all errors are small whereas with decreasing amplitude the errors become very high. The other plots show subtracted pairwise filtering histograms. Note that the ranges of the difference histograms are smaller. Counts outside the displayed axes are summarized in the highest bins respectively. G: The adaptive median (weighted by amplitude A) reduced the error of well illuminated pixels more than AWG but was worse at very bad pixels (H). Smoothing with maximal width improves the quality of flat surfaces (I) but can lead to high errors at the edges due to directed blurring (J).

4.3 Comparison of the filters using Normalized Convolution

The filters with fixed width were the fastest and produced very similar results. All investigations have shown that the weighted Gaussian filter (WG) with fixed width leads to the same result as directional averaging in the complex plane with A^2 . Both filters use the same technique (normalized convolution) and the same weighting factors. A difference only occurs if many adjacent pixels differ by more than π in phase. In most cases, the complex filter using A as a weighting factor performed worse than the WG and the AWG filter. Only with mask sizes of 11×11 or higher did the weaker penalization of a low amplitude pay off since the effect of directed blurring was not that strong. This filter corresponds to weighting

Table 1: Computation time in [s] for an image with 160 by 120 pixels of medium quality (the room scene) and a large number of convolutions (steps) for AWG in the scale selection process. The time is averaged over 100 runs and the empirical standard deviation is given.

n	AWGsteps	AWG	AdMed (MAD)	AdMed (A)	WG	Complex- A^2
3	45	1.27±0.03	2.60±0.04	0.98±0.02	0.0054 ±0.0001	0.023±0.001
5	78	3.18±0.09	2.64±0.04	1.00±0.02	0.0066 ±0.0003	0.042±0.001
7	112	6.0±0.5	2.70±0.07	1.02±0.02	0.0073 ±0.0006	0.072±0.007

Table 2: Computation time in [s] for an image with 160 by 120 pixels of poor quality (the box, see Fig. 8) and a small number of convolutions (steps) for AWG in the scale selection process. The time is averaged over 100 runs and the empirical standard deviation is given.

n	AWGsteps	AWG	AdMed(MAD)	AdMed(A)	WG	Complex- A^2
3	9	0.16±0.01	2.61±0.03	0.99±0.01	0.005±0.001	0.024±0.001
5	16	0.32±0.01	2.67±0.03	1.00±0.02	0.007±0.001	0.044±0.001
7	22	0.53±0.01	2.70±0.04	1.03±0.02	0.007±0.001	0.071±0.001

directly on the raw data which implicitly makes use of the amplitudes and takes place in the complex plane as well.

In slightly overexposed regions where the modulation amplitude is still high, all fixed width filters performed better than the AWG since the latter does not smooth at all at these amplitudes. Adjacent to slightly overexposed pixels, there are either pixels without overexposure but sufficient illumination (and higher amplitude) or others with even more saturation effects (and lower amplitude [11]). In these cases, non-adaptive amplitude-weighted filtering always produces more reliable information (see Fig. 7, F). If the overexposed regions are too large, only the edges to non-saturated regions benefit from filtering.

Considering the effect of directed blurring, the adaptive approach is superior to the non-adaptive. This is clearly illustrated at the dangling cable in Fig. 3 (marker E). Overall, adaptivity pays off in all observed scenes: the epp is always somewhat

lower as one can see in Fig. 4 left (WG corresponds to a variance threshold of 0)

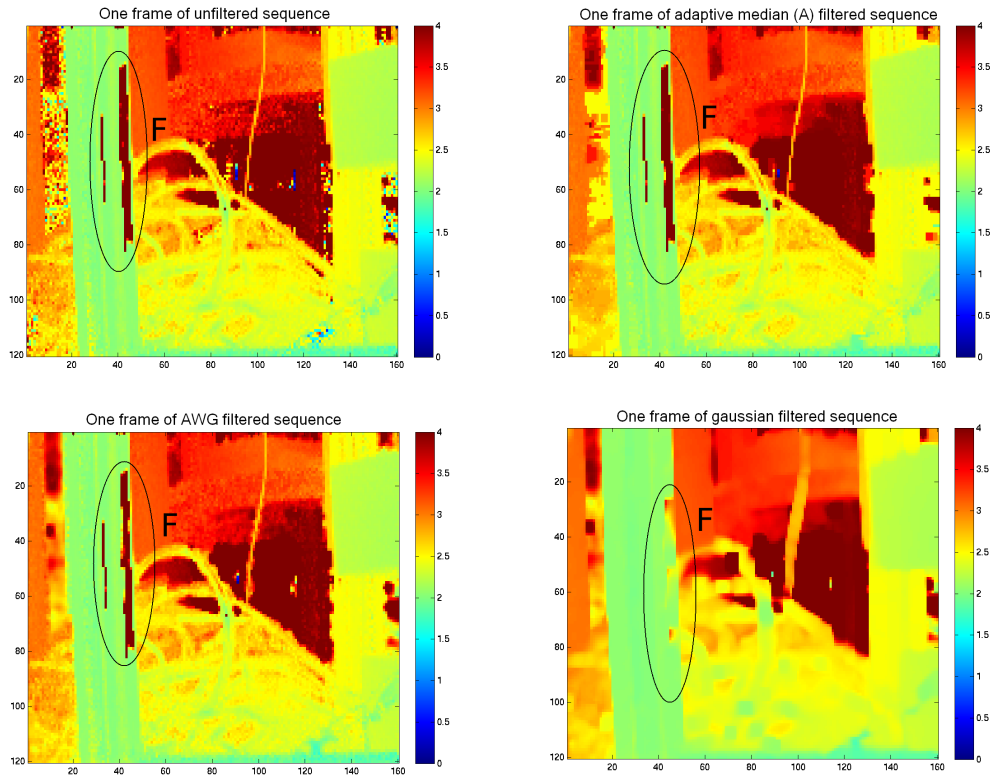


Figure 7: Four depth images of the hose scene acquired with 20 ms integration time. The region marked with F has a strong bias due to overexposure. Only the weighted Gaussian filters with fixed width were able to correct this bias.

4.4 Comparison of the Median Filters

In the first brief discussion of Fig. 3 we mentioned that the adaptive median using MAD left a surprisingly large number of outliers in the image. This occurs in regions with high standard deviation and low amplitude. The reason is that at vanishing amplitude the depth values within the unambiguous range are only sparsely populated due to strong quantization errors [11]. Taking the median of such a set of pixels always leads to the same value which in turn leads to a vanishing MAD value (see Fig. 8). These pixels are not smoothed by the adaptive median except if the detector's MAD threshold is set to zero (that is to say taking the median

nonadaptively). Therefore the simple median or the adaptive median using A as a threshold performed better in the case of very low amplitude pixels.

All median filters completely removed fine structures in the depth images such as the cable in Fig. 3 (marker E). In return, outliers are removed by the median and the adaptive median using the amplitude, too. At acceptably illuminated regions or when adapting with respect to the amplitude A , the results are comparable to the AWG filter and better than that of the WG's. In the case of very large mask sizes, the median filters performed best since they cause no directed blurring.

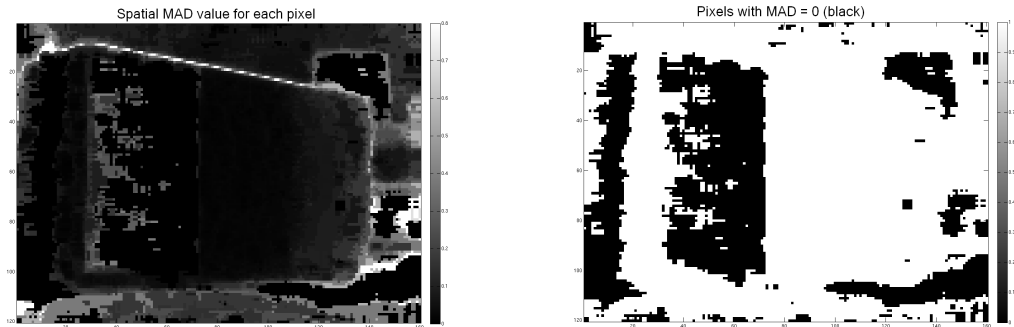


Figure 8: Regions with low amplitude are subject to strong quantization effects such that only few different depth values are possible. This leads to frequently occurring values and in consequence to a low MAD value even if the variance is high. The original scene shows an open box and background. Left: The spatial MAD value is zero for many pixels with low signal amplitude since the pixels populate only a few particular depth values there.

5 Conclusion and Outlook

We discussed various approaches to the denoising of depth maps obtained by TOF 3D cameras. Two main concepts and their variations have been investigated: normalized convolution with different weighting factors and median filtering, both in adaptive and in nonadaptive variants. An assessment has been performed using qualitative and quantitative methods. It has turned out that the inverse squared amplitude is indeed a reliable measure of confidence, as predicted by theory. The WG filters with fixed width and weighting with A^2 have performed well in all scenes if used up to a maximal mask size of 7×7 . For larger mask sizes, the

effect of directed blurring leads to large errors. They are the only ones to cope with small patches of overexposure and offer the fastest computation.

However, the fixed width filters unnecessarily blur the “good” pixels. There, the error can become larger than the error of the unfiltered image. The adaptive weighted Gaussian filter (AWG) uses smaller Gaussian widths at the edges than the WG filters and therefore has a reduced error at the edges. The adaptive median using MAD as a threshold has difficulty coping with the quantization effects of the data. However, if the modulation amplitude is used as the threshold, results are comparable to that of the AWG filter except for the drawback that small structures are completely suppressed and the advantage of better edge preserving. Considering the mean absolute error per pixel (epp) the AWG filter can reach better results even with suboptimal parametrization (see Fig. 4). Neglecting computation time, it is superior to all filters investigated.

All proposed filters can be implemented non-iteratively. Further effort could be directed towards iterative approaches or bilateral filters [42] to overcome the discussed effect of directed blurring.

References

- [1] A. Bevilacqua, L. D. Stefano, P. Azzari, “People Tracking Using a Time-of-Flight Depth Sensor”, *AVSS '06: Proc. of the IEEE Int. Conf. on Video and Signal Based Surveillance*, p.98 (2006)
- [2] S. Knoop, S. Vacek, R. Dillmann, “Sensor Fusion for 3D Human Body Tracking with an Articulated 3D Body Model”, *Proc. 2006 IEEE Int. Conf. on Robotics and Automation (ICRA 2006)*, pp. 1686-1691 (2006)
- [3] P. Breuer, C. Eckes, S. Müller, “Hand Gesture Recognition with a Novel IR Time-of-Flight Range Camera—A Pilot Study”, *Computer Vision/Computer Graphics Collaboration Techniques: Third Int. Conf. (MIRAGE 2007)*, vol. 4418/2007, pp. 247-260 (2007)
- [4] M. Haker, M. Böhme, T. Martinetz, E. Barth, “Geometric Invariants for Facial Feature Tracking with 3D TOF Cameras”, *Proc. of the IEEE Int. Symp. on Signals, Circuits & Systems (ISSCS)*, vol. 1, pp. 109-112 (2007)
- [5] M. Juberts, A. Barbera, “Status Report on Next Generation LADAR for Driving Unmanned Ground Vehicles”, *Tech. rep.*, National Inst. of Standards and Technology (2004)

- [6] J. Wikander, "Automated Vehicle Occupancy Technologies Study: Synthesis Report", *Tech. rep.*, Texas Transportation Inst. (2007)
- [7] J. W. Weingarten, G. Gruener, R. Siegwart, "A State-of-the-Art 3D Sensor for Robot Navigation", *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, vol. 3, pp. 2155-216 (2004)
- [8] M. Lindner and A. Kolb: "Lateral and Depth Calibration of PMD-Distance Sensors", *Advances in Visual Computing, Springer, 2*, pages 524-533 (2006)
- [9] T. Kahlmann, F. Remondino, H. Ingensand, "Calibration for increased accuracy of the range imaging camera Swissranger" *ISPRS Commission V Symposium 'Image Engineering and Vision Metrology' (IEVM)* (2006)
- [10] T. Möller, H. Kraft, J. Frey, "Robust 3D Measurement with PMD Sensors", *Proceedings of the 1st Range Imaging Research Day at ETH Zurich* (2006)
- [11] M. Frank, M. Plaue, H. Rapp, U. Köthe, B. Jähne, F. A. Hamprecht, "Theoretical and Experimental Error Analysis of Continuous-Wave Time-Of-Flight Range Cameras" *Optical Engineering*, vol. 48, No. 1, 013602 (2009).
- [12] T. Spirig, P. Seitz, O. Vietze, and F. Heitger, "The lock-in CCD - twodimensional synchronous detection of light", *IEEE Journal of Quantum Electronics*, vol. 31, pp. 1705-1708, (1995)
- [13] R. Schwarte, H. G. Heinol, Z. Xu, K. Hartmann, "A new active 3D-Vision system based on RF-modulation interferometry of incoherent light", *Photonics East – Intelligent Systems and Advanced Manufacturing, Proceedings of the SPIE, Vol. 2588, Philadelphia* (1995)
- [14] Z. Xu, "Investigation of 3D-Imaging Systems Based on Modulated Light and Optical RF-Interferometry (ORFI)", *ZESS Forschungsberichte*, vol. 14, pp. 214 (1999)
- [15] B. Schneider: "Der Photomischdetektor zur schnellen 3D-Vermessung für Sicherheitssysteme und zur Informationsübertragung im Automobil", Diss., *Department of Electrical Engineering and Computer Science, University of Siegen* (2003)
- [16] G. Aubert, P. Kornprobst, "Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations", Springer (2006)

- [17] J. Weickert, "Anisotropic Diffusion in Image Processing", *Diss.*, Univ. of Kaiserslautern
- [18] R.-J. Recknagel, R. Kowarschik, G. Notni, "High-resolution Defect Detection and Noise Reduction Using Wavelet Methods for Surface Measurement", *J. Opt. A: Pure Appl. Opt.*, vol. 2, pp.538-545 (2000)
- [19] R. Coifman, D. Donoho, "Translation Invariant De-noising", *Tech. rep.* (1995)
- [20] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, D. Picard, "Wavelet Shrinkage: Asymptopia?", *Tech. rep.* (1994)
- [21] A. Buades, B. Coll, J. M. Morel, "A Review of Image Denoising Algorithms, With a New One", *Multiscale Model. Simul.*, Vol. 4, No. 2, pp. 490-530 (2005)
- [22] A. Buades, B. Coll, J.-M. Morel, "A Non-local Algorithm for Image Denoising", *IEEE Comp. Soc. Conf. on Comp. Vision and Pattern Recognition, CVPR*, vol. 2, pp. 60-65 (2005)
- [23] L. I. Rudin, S. Osher, E. Fatemi, "Nonlinear Total Variation Based Noise Removal Algorithms", *Physica D* 60, pp. 259-268 (1992)
- [24] J. Portilla, V. Strela, M. J. Wainwright, W. P. Simoncelli, "Adaptive Wiener Denoising Using a Gaussian Scale Mixture Model in the Wavelet Domain", *Proc. of the 8th Int. Conf. on Image Processing*, vol. 2, pp. 37-40 (2001)
- [25] N. Wiener, "Extrapolation, Interpolation, and Smoothing of Stationary Time Series", MIT Press (1964)
- [26] T. Chen, K.-K. Ma, L.-H. Chen, "Tri-State Median Filter for Image Denoising", *IEEE Trans. on Image Processing*, vol. 8, No. 12 (1999)
- [27] K. Hildebrandt, K. Polthier, "Anisotropic Filtering of Non-Linear Surface Features", *Computer Graphics Forum*, vol. 23 (3) (2004)
- [28] C. Lange, K. Polthier, "Anisotropic Fairing of Point Sets", *Special Issue of CAGD* (2005)

- [29] Ron Kimmel, Ravi Malladi, Nir A. Sochen, "Image Processing via the Beltrami Operator", *Proc. of the Third Asian Conf. on Comp. Vision*, vol. 1, pp. 547-581 (1998)
- [30] A. I. Bobenko, P. Schröder, "Discrete Willmore Flow", *Eurographics Symp. on Geom. Processing*, pp. 101-110, eds. M. Desbrun, H. Pottmann (2005)
- [31] R. M. Bolle, B. C. Vermuri, "On Three-Dimensional Surface Reconstruction Methods", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, No. 1 (1991)
- [32] J. Shah, H. H. Pien, J. M. Gauch, "Recovery of Surfaces with Discontinuities by Fusing Shading and Range Data Within a Variational Framework", *IEEE Trans. on Image Processing*, vol. 5, No. 8, pp. 1243-1251 (1996)
- [33] M. Bertero, P. Boccacci, "Introduction to Inverse Problems in Imaging", Inst. of Phys. Publ. (1998)
- [34] A. Foi, "Pointwise Shape-adaptive DCT Image Filtering and Signal-dependent Noise Estimation", *Tampere Univ. of Technology, Publication 710* (2007)
- [35] X. Luan, "Experimental Investigation of Photonic Mixer Device and Development of TOF 3D Ranging Systems Based on PMD Technology", *Diss.*, Univ. Siegen (2001)
- [36] H. Rapp, M. Frank, F. A. Hamprecht, B. Jähne, "A Theoretical and Experimental Investigation of the Systematic Errors and Statistical Uncertainties of Time-of-Flight Cameras", *Int. J. Intelligent Systems Technologies and Applications*, vol. 5, Nos. 3/4 (2008)
- [37] H. Knutsson, C.-F. Westin, "Normalized and Differential Convolution", *Proc. CVPR*, pp. 515-523 (1993)
- [38] J. Restle, M. Hissmann, F. A. Hamprecht, "Nonparametric Smoothing of Interferometric Height Maps Using Confidence Values", *Optical Engineering*, Vol. 43 No. 4 (2004)
- [39] F. R. Hampel: "The Breakdown Points of the Mean Combined with Some Rejection Rules" *Technometrics*, Vol. 27, No. 2, pp. 95-107, (1985)

- [40] M. Hissmann, F. A. Hamprecht: “Bayesian Surface Estimation for White Light Interferometry”, *Optical Engineering, Vol. 44 No. 1* (2005)
- [41] A. Kirsch, “An Introduction to the Mathematical Theory of Inverse Problems”, Springer (1996)
- [42] C. Tomasi, R. Manduchi, “Bilateral Filtering for Gray and Color Images” *Proceedings of the IEEE International Conference on Computer Vision, Bombay* (1998)