A Clustering Based Denoising Technique for Range Images of Time of Flight Cameras

H. Schöner, B. Moser
Software Competence Center Hagenberg
A-4232 Hagenberg, Austria
{*holger.schoener,bernhard.moser*}*@scch.at*

A.A. Dorrington, A.D. Payne, M.J. Cree
Dept. of Engineering, Univ. of Waikato
Hamilton, New Zealand
{*adrian,a.payne,cree*}*@waikato.ac.nz*

B. Heise, F. Bauer
Dept. of Knowledge-Based Math. Systems, Univ. of Linz
A-4232 Hagenberg, Austria
{*bettina.heise,frank.bauer*}*@jku.at*

Abstract

A relatively new technique for measuring the 3D structure of visual scenes is provided by time of flight (TOF) cameras. Reflections of modulated light waves are recorded by a parallel pixel array structure. The time series at each pixel of the resulting image stream is used to estimate travelling time and thus range information. This measuring technique results in pixel dependent noise levels with variances changing over several orders of magnitude dependent on the illumination and material parameters.

This makes application of traditional (global) denoising techniques suboptimal. Using free aditional information from the camera and a clustering procedure we can get information about which pixels belong to the same object, and what their noise level is, which allows for locally adapted smoothing. To illustrate the success of this method, we compare it with raw camera output and a traditional method for edge preserving smoothing, anisotropic diffusion [10, 12].

We show that this mathematical technique works without individual adaptations on two camera systems with highly different noise characteristics.

1. TOF Cameras

Principle of operation Time-of-Flight (TOF) cameras simultaneously measure distance (range) and intensity within every pixel allowing 3D information to be collected within a scene. An amplitude modulated light source, typically operated between 10–100 MHz, illuminates the field of view and the reflected light is imaged with a gain modulated camera system. The flight time causes a delay in the modulation envelope that is exhibited as a phase shift φ in the received modulation signal proportional to object distance d as given by Equation (1), where f_{mod} is the modulation frequency and c is the speed of light.

$$\varphi = \frac{4\pi f_{\text{mod}} d}{c} \tag{1}$$

To measure the phase shift φ the gain modulated camera can either operate at the same frequency (homodyne operation as used by the SwissRanger SR-3000 [9]), or at a slightly different frequency (heterodyne operation as used by the Waikato Range Imager [5]) to that of the illumination source. The camera records an intensity image I which is a cross correlation between the modulated optical signal and the receiver modulation waveforms. From a minimum of three phase offset images indexed by i and of phase θ_i, the signal amplitude A and phase φ can be determined by

$$A = \frac{\sqrt{\left(\sum_i I_i \cos\theta_i\right)^2 + \left(\sum_i I_i \sin\theta_i\right)^2}}{2} \tag{2}$$

$$\varphi = \arctan\left(\frac{\sum_i \left(I_i \cos\theta_i\right)}{\sum_i \left(I_i \sin\theta_i\right)}\right) \tag{3}$$

Homodyne operation typically uses four images, each with a phase shift of $\pi/2$ radians [9, 7], whereas in heterodyne operation, the phase is continually changing and allows range measurements of higher precision by acquiring a larger number of phase offset images. This can be used to remove the influence of harmonics that could otherwise contaminate the phase measurement.

Technical details The Waikato Range Imager [5] uses an array of four visible (658 nm) laser diodes placed around the

978-0-7695-3514-2/08 $25.00 © 2008 IEEE
DOI 10.1109/CIMCA.2008.95

IEEE computer society

imaging lens to illuminate the scene with a combined output power of approximately 300 mW. A Dalsa Pantera TF 1M60 digital video camera has been fitted with a 25 mm image intensifier from Photek Ltd (East Sussex, United Kingdom) that is employed as a high speed shutter, providing gain modulation at frequencies up to 90 MHz. The digital video camera is capable of operating at 1024×1024 resolution with frame rates up to 60 Hz, although pixel binning allows faster frame rates at the expense of lower spatial resolution.

Imaging setup Scene capturing (cf. Section 3) with the Waikato Range Imager was performed using a laser modulation frequency of 80 MHz and an image intensifier modulation frequency of 80 MHz + 20 Hz. This produces a 20 Hz beat signal output from the image intensifier which is captured by the digital video camera operating at a frame rate of 100 Hz. This configuration utilises five raw intensity images to produce each range image, which is optimal to minimise systematic errors while allowing high speed acquisition and high spatial resolution [4], generating range data at 20 frames per second with 512×512 resolution.

In comparison, the SwissRanger SR-3000 uses 55 LEDs at a wavelength of 850 nm producing an optical output of approximately 1 W. The sensor resolution is 176×144, and as the gain is internally modulated at frequencies up to 30 MHz, there is no need for an external shutter. The range (and intensity) of the scene is computed by the sensor and is typically output at 25 measurements per second. The expamles in Section 3 use range images preprocessed by the camera software (median filtering, geometry correction).

Noise A typical characteristic of pixel noise as obtained from TOF cameras is the dependence on surface and illumination properties. It can be shown that it is proportional to the inverse of the intensity, $\mathrm{var}(\varphi) \propto 1/A^2$, [11].

Traditional denoising such as temporal and spatial averaging do not take this into account adequately. Their parameters can usually only be tuned appropriately for a part of the images. Furthermore, spatial averaging usually blurs edges, thus sometimes introducing considerable errors in the range images at the border between objects located at different distances from the camera. On the other hand, the clustering based denoising approach presented in the next section can make use of the additional information about surface region and object boundaries, which is present in images representing the estimated noise level at each pixel. It performs an initial segmentation of the image into regions with similar characteristics, and performs standard Gaussian smoothing limited to these regions, and tuned to their noise level. In Section 3 we show examples of the achieved denoising quality, and compare it to standard anisotropic diffusion.

Another issue present in images of many TOF cameras is a small bias on the estimated range, depending again on pixel intensity. A checker board recorded with such cameras appears to have the dark fields at a different distance than the light fields. This phenomenon is present in the images presented here, but is not topic of denoising procedures. Instead, this should be treated using additional methods.

2. Clustering Based Denoising

Considering the fact, that image regions with different reflection properties lead to different range noise levels, denoising of the data should be adaptive to the properties of each region. We propose a clustering approach to identify the regions with homogeneous properties (range values, range noise, and intensity values). The identified clusters are used to estimate the noise level inside each cluster, and smoothing is then performed on each cluster tuned to its characteristics. The details of this approach are given in [8]; the following gives an overview of this approach.

Clustering The data $\mathcal{D}$ available for clustering consist of a feature vector $\mathbf{d}_{x,y} = (\boldsymbol{\varphi}_{x,y}, \mathrm{var}_{x,y}, \mathbf{I}_{x,y}, x, y)^T$ for each pixel (x, y), with the range information $\boldsymbol{\varphi}$, the estimated noise levels var (usually smoothed using a Gaussian kernel), and the intensity $\mathbf{I}$.

The regions or objects which should be identified by the clustering algorithm do not have a predefined shape in this feature space,[1] which rules out several common clustering techniques which depend on the existence of a meaningful distance between points and cluster prototypes. We propose a variant of the Mean Shift clustering algorithm (as proposed by Comaniciu et. al [3], based on previous publications [6, 1, 2]), which can make use of a multi-dimensional feature space and is not restricted to given cluster shapes or cluster numbers. The Mean Shift clustering algorithm is a density based method. For each point it determines the corresponding cluster by following the gradient of a kernel density estimate (without actually computing the density) to the maximum of the density estimate. Points with the same or similar maxima (for the case of ridges in the density function) are put into the same cluster.

Before applying the clustering algorithm to the data, we rescale $\mathbf{d}_{x,y}$ by the inverse of the standard deviation of each dimension, and apply a factor for weighting the importance of these dimensions. We use the factors 1 for the image dimensions, and 0.2 for the pixel location dimensions; the latter can be adapted to the size the images and the size of the objects therein, but does not seem to be very sensitive in

[1] Consider objects leading to ridges in the range image, e.g. tape rolls; the location of pixels has to be included in the feature space to obtain spacially contiguous regions.

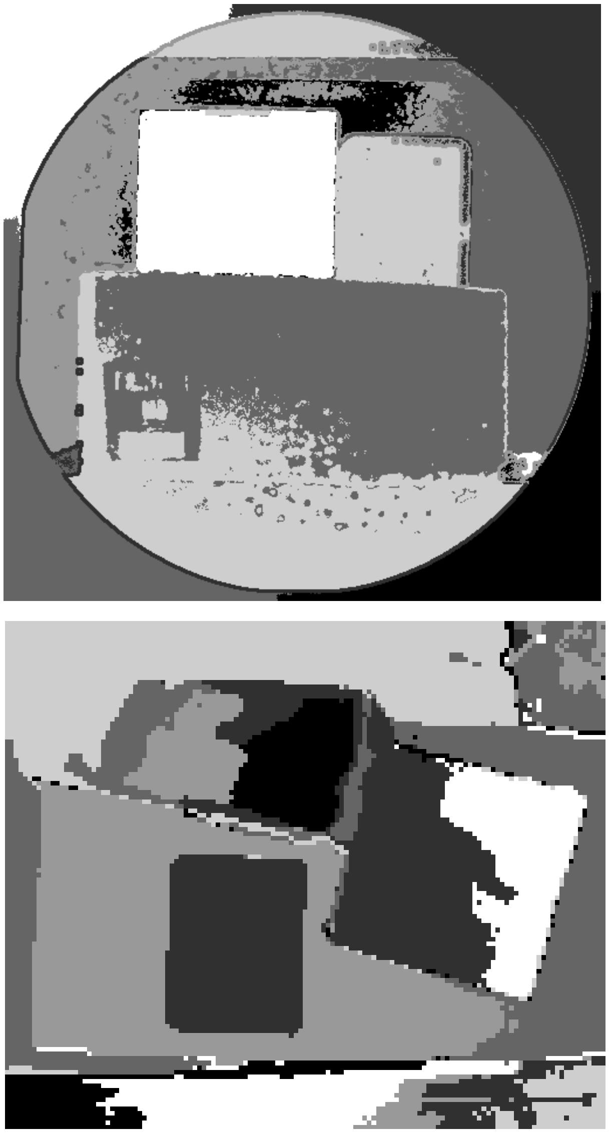

Figure 1. Clusters found for the example scenes (Section 3). Only six grey levels are used, so some clusters share the same color.

our experience. After the clustering, we merge very small clusters (containing less than 15 pixels) to the closest other cluster. An example for a clustering found for one of the example scenes is given in figure 1.

Smoothing The smoothing approch has to deal with two issues. It should be adapted to the noise properties of regions with homogeneous noise characteristics, and it should not blur edges in the range image. The clusters found by the Mean Shift algorithm allow to estimate the noise level for the corresponding image region. This information can be used to adapt the strength of smoothing, which is to be performed on the region defined by all members of a given cluster. Edge blurring is avoided as long as edges coincide with boundaries between clusters.

A remaining problem is that homogeneous regions could be split into different clusters, because they have a non-constant distance to the camera (i.e. a gradient in the range image). Even when such regions are covered by separate clusters, the smoothing should not stop at the boundaries between such clusters, otherwise artificial steps could be introduced into the smoothed range image.

We avoid this, by computing a (in general non-symmetric) similarity between each pair of clusters, and taking points from the other cluster(s) into account when smoothing one cluster, weighted according to the cluster similarity. The computed similarity takes into account mainly the range values and the average direction of the range gradients occuring in one cluster, by comparing, whether the neighboring clusters have range values, which could also occur in the given cluster. Additionally, intensity, location and noise characteristics are incorporated. Details are given in [8]. The actual smoothing uses a Gaussian kernel. It is extended such that each pixel has an assigned weight given by the similarity between a given cluster and this pixel; this weight is incorporated in the summation and normalization taking place in computing the smoothed pixel range value.

3. Application Results

The scenes selected for comparing the different denoising approaches consist of a collection of differently colored boxes, which are arranged in part parallel to the image plane, and in part tilted. Several characteristics in these scenes allow a good comparison: The different colors of the boxes lead to different noise levels, surfaces parallel to the image plane allow to judge the amount of noise and the effect of the smoothing techniques. The tilted planes test the ability of the smoothing approaches to keep smooth surfaces, even if they stretch across a distance range. Furthermore, there are sharp edges in the range images at the boundaries of the boxes. The two cameras were available at two different places (Austria and New Zealand), thus we just imaged similar scenes (Figure 2).

For computation of the intensity, range and noise images, two successive frames for the Waikato scene, and five successive frames for the SwissRanger scene are used, averaging them for intensity and range images, and taking the (log-)variance of the range as estimate of the noise. The resulting images, which are only temporally averaged over two and five frames, are shown in Figure 3.

For the Waikato scene, Figure 4 shows a cut through

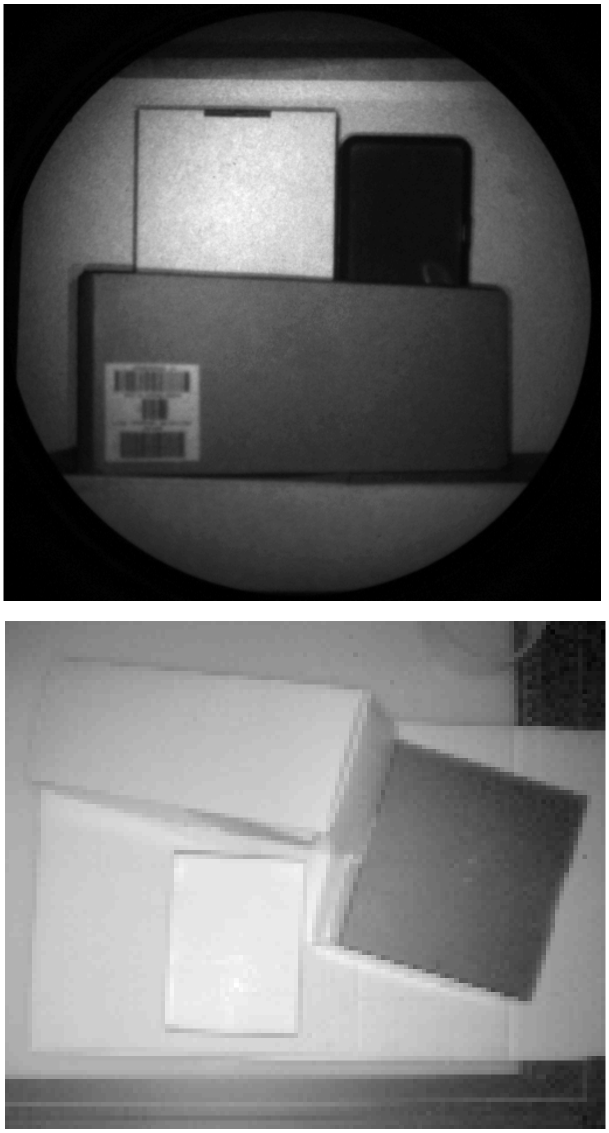

Figure 2. Intensity image of the scenes recorded with the Waikato Range Imager (top) and Swiss Ranger (bottom).

the light and the dark boxes. The lines correspond to the original (temporally averaged) range, the cluster-based smoothed range, and (only in the two subplots) the range obtained from anisotropic diffusion, as a widely used algorithm for edge preserving smoothing. The left subplot shows part of the lighter box with low noise level (on the left) and part of the black box with high noise level (on the right). The performance of anisotropic diffusion and clustering based smoothing is relatively similar in the low noise region, where both improve the signal slightly. For the black box on the right, both methods show a clear improvement, but only the clustering based smoothing yields a very smooth range estimate. This estimate is also quite realistic, as the edges of the black box are indeed slightly rounded.

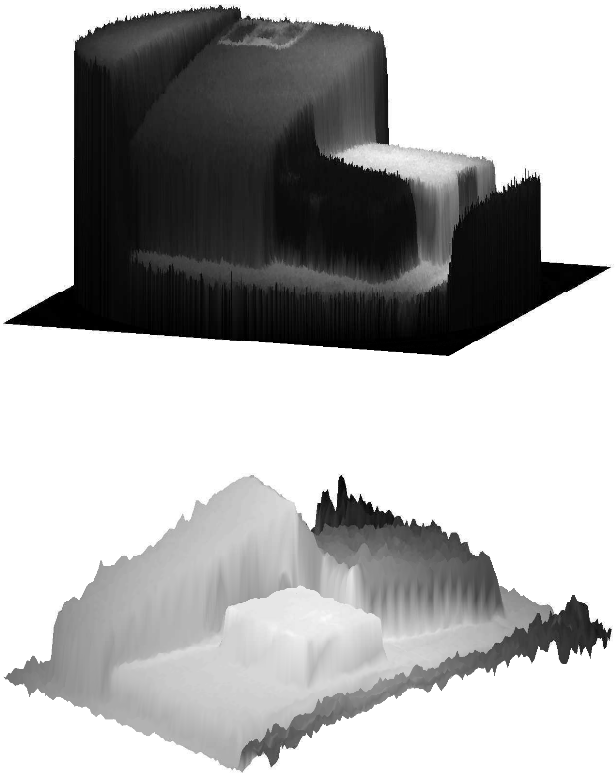

Figure 3. Temporally averaged example scenes. Height corresponds to the range image, grey scale to the intensity image. The camera is located above these scenes.

The second subplot enlarges a region around the edge of the black box. It shows, that the clustering based range estimate closely follows the steep distance change present in the original temporally averaged curve. On the other hand, the anisotropic diffusion curve shows some blurring of this edge. Effects like these are typical for many of the edges present in this range image, and are just very pronounced in this place, because there are two sharp edges very close together.

For the scene recorded by the SwissRanger, some easily identifiable dimensions were measured, and are available to quantify the smoothing results. The range image from the SwissRanger is not in a defined scale known to us, and still contains some geometrical distortions probably due to illumination differences and imperfect correction for non-parallel rays to the camera. To account for this, we first performed a relatively rigid fit of the estimated range image to the dimensions known to us. This transformation in-

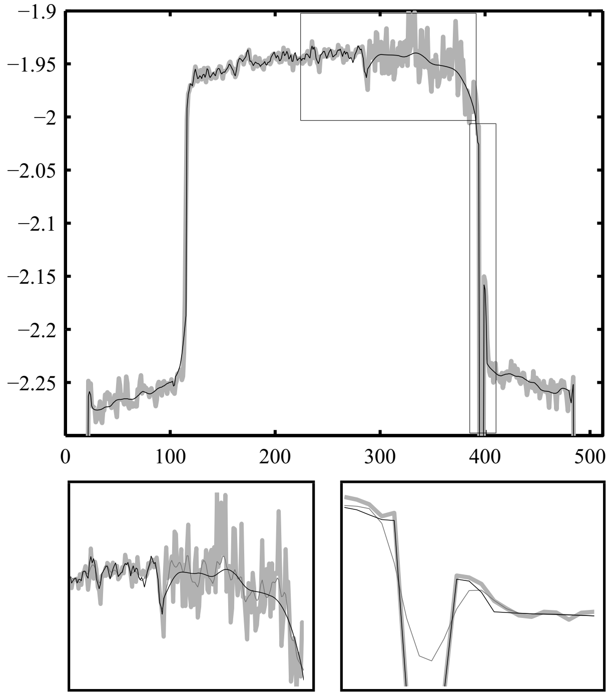

Figure 4. Waikato scene, a cut through the light and dark boxes shown in the right part of the upper image in Figure 3. The two lower graphs are zoomed versions of the rectangles indicated in the upper graph. The top graph shows the averaged range (thick light line) and the range obtained with the clustering-based smoothing (thin solid black line). The lower two plots additionally show the range obtained from edge preserving smoothing (anisotropic diffusion, thin solid grey line).

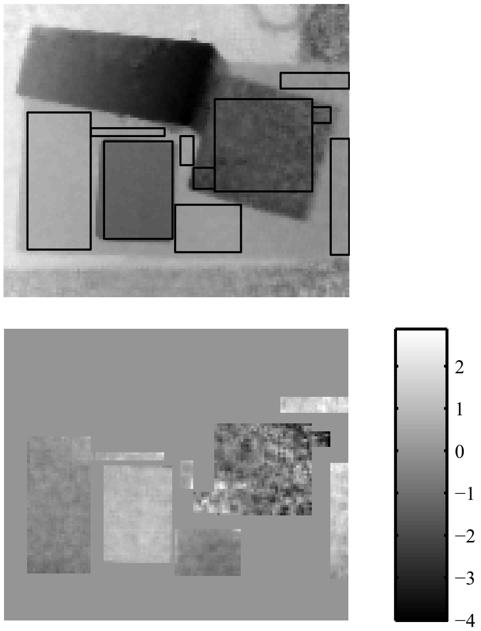

Figure 5. Temporally averaged range image of the SwissRanger scene (top), errors with respect to the known true distance values (bottom). Regions selected for comparison with known true dimensions are indicated by rectangles in the upper image. The errors are the difference of estimated and known range values for each pixel, and are measured in cm; the colorbar indicates the range of errors present. The background color corresponds to an error of 0 cm.

cluded a global range offset, a global range scale factor, and a pixel-location dependent scale (linear and quadratic terms for horizontal and vertical dimensions). Some regions were then selected for their well known true range values (background, two boxes parallel to image plane), including light and dark surfaces. The boundaries of these regions are as close as possible to the edges of the objects, to allow detection of edge blurring introduced by smoothing. These regions do not touch, because the camera-preprocessed images already contain slightly blurred edges, which should not be included in the selected regions. Figures 5 shows the selected regions, and the errors of the temporally averaged range. Figure 6 shows the errors of the anisotropic diffusion smoothed and the cluster-based smoothed range images. Without introducing more serious artifacts, the cluster-based smoothing achieves the smoothest results. When averaging over the selected areas, the root mean squared error (RMSE) of the estimated ranges are 0.2825 cm for the temporally averaged image, 0.2549 cm for the anisotropic diffusion based image, and 0.2375 cm for the cluster-based smoothing range image. Most of these errors are due to the imperfect geometrical correction of the images; but as none of the presented methods does any geometrical correction, the reduction in RMSE is mainly due to less noise in the smoothed images.

4. Conclusion

We have demonstrated the effect of temporal averaging, a traditional adaptive smoothing method (anisotropic diffusion), and our cluster-based smoothing approach on range images acquired from two TOF cameras of different types. The presented results show a clear advantage of the adaptation of smoothing strength to local noise characteristics (visible especially in the lower left graph in Figure 4), and

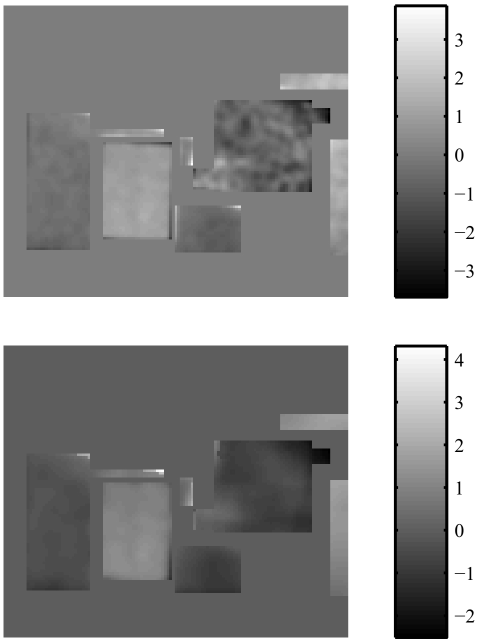

Figure 6. Errors in estimated range values for the selected regions, for anisotropic diffusion (top) and clustering based smoothing (bottom).

they show the advantage of using a segmentation of the scene to avoid edge blurring (lower right graph in Figure 4). For this approach we made use of the fact, that in contrast to most other imaging methods we simultaneously measure not just the range, but also local noise levels and the almost noise free intensity image.

The clustering-based smoothing might be sensitive to the kind of scene analysed. Problematic could be objects having strong gradients in the range image, but still being smooth. We showed that this can be controlled by using a smoothing, which takes into account, whether neighboring clusters belong to the same object.

A limitation of the clustering-based smoothing is (at least in the current non-optimized implementation in Matlab) that its computations require considerably more time. Depending on the resolution, clustering and smoothing parameters, they currently take tens of seconds to a few minutes on current PC hardware. However, a significant optimization of computation time should be possible, and several application scenarios could afford these computation times, and could even benefit from the initial image segmentation already returned by the clustering.

A topic worth of further investigation is the bias present in the estimated range of darker or lighter objects. This is not an issue of noise, and cannot be corrected using smoothing algorithms, but it might be desirable to remove such artifacts before further analysis of range images.

Acknowledgments

The authors gratefully acknowledge the support of the Upper Austrian Technology and Research Promotion, and of the Austrian COMET program.

References

[1] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, August 1995.

[2] E. Choi and P. Hall. Data sharpening as a prelude to density estimation. *Biometrika*, 86:941–947, 1999.

[3] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), May 2002.

[4] A. Dorrington, M. Cree, D. Carnegie, A. Payne, R. Conroy, J. Godbaz, and A. Jongenelen. Video-rate or high-precision: A flexible range imaging camera. In *Proc. of SPIE*, volume 6813, page 681307. SPIE, 2008.

[5] A. Dorrington, M. Cree, A. Payne, R. Conroy, and D. Carnegie. Achieving sub-millimetre precision with a solid-state full-field heterodyning range imaging camera. *Meas. Sci. Technol.*, 18:2809–2816, 2007.

[6] K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21:32–40, 1975.

[7] S. Hsu, S. Acharya, A. Rafii, and R. New. Performance of a Time-of-Flight range camera for intelligent vehicle safety applications. *Advanced Microsystems for Automotive Applications*, pages 205–214, 2006.

[8] B. Moser, F. Bauer, P. Elbau, B. Heise, and H. Schöner. Denoising techniques for raw 3D data of TOF cameras based on clustering and wavelets. In B. D. Corner, M. Mochimaru, and R. Sitnik, editors, *Proc. of SPIE, Three-Dimensional Image Capture and Applications 2008*, volume 6805, pages 68050E–68050E–12. SPIE, 2008.

[9] T. Oggier, M. Lehmann, R. Kaufmannn, M. Schweizer, M. Richter, P. Metzler, G. Lang, F. Lustenberger, and N. Blanc. An all-solid-state optical range camera for 3D-real-time imaging with sub-centimeter depth resolution (SwissRanger). In *Proceedings of the SPIE*, volume 5249, 2004.

[10] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.

[11] H. Rapp. Experimental an theoretical investigation of correlating TOF camera systems. Master's thesis, Universität Heidelberg, 2007.

[12] J. Weickert. *Anisotropic diffusion in image processing*. European Consortium for Mathematics in Industry. B. G. Teubner, Stuttgart, 1998.