Fuzzy logic-based approach to wavelet denoising of 3D images produced by time-of-flight cameras

Ljubomir Jovanov,* Aleksandra Pižurica, and Wilfried Philips

Telecommunications and Information Processing Department, Ghent University, Sint Pietersnieuwstraat 41, 9000 Ghent, Belgium

ljj@telin.ugent.be

Abstract: In this paper we present a new denoising method for the depth images of a 3D imaging sensor, based on the time-of-flight principle. We propose novel ways to use luminance-like information produced by a time-of flight camera along with depth images. Firstly, we propose a wavelet-based method for estimating the noise level in depth images, using luminance information. The underlying idea is that luminance carries information about the power of the optical signal reflected from the scene and is hence related to the signal-to-noise ratio for every pixel within the depth image. In this way, we can efficiently solve the difficult problem of estimating the non-stationary noise within the depth images. Secondly, we use luminance information to better restore object boundaries masked with noise in the depth images. Information from luminance images is introduced into the estimation formula through the use of fuzzy membership functions. In particular, we take the correlation between the measured depth and luminance into account, and the fact that edges (object boundaries) present in the depth image are likely to occur in the luminance image as well. The results on real 3D images show a significant improvement over the state-of-the-art in the field.

OCIS codes: (100.0100) Image processing; (110.0110) Imaging systems; (100.2980) Image enhancement; (100.6890) Three-dimensional image processing; (100.3175) Interferometric imaging; (110.6880) Three-dimensional image acquisition.

References and links

1. C. L. Zitnick and S. B. Kang, "Stereo for image-based rendering using image over-segmentation," Int. J. Comput. Vis. **75**, 49–65 (2007).
2. W. Miled, J.-C. Pesquet, and M. Parent, "A convex optimization approach for depth estimation under illumination variation," IEEE Trans. Image Process. **18**, 813–830 (2009).
3. S. K. Nayar and Y. Nakagawa, "Shape from focus," IEEE Trans. Pattern Anal. Mach. Intell. **16**, 824–831 (1994).
4. A. Torralba and A. Oliva, "Depth estimation from image structure," IEEE Trans. Pattern Anal. Mach. Intell. **24**, 1226–1238 (2002).
5. S. Soatto and P. Perona, "Reducing "structure from motion": A general framework for dynamic vision part 1: Modeling," IEEE Trans. Pattern Anal. Mach. Intell. **20**, 933–942 (1998).
6. R. Lange and P. Seitz, "Solid-state time-of-flight range camera," IEEE J. Quantum Electron. **37**, 390 –397 (2001).
7. R. G. J. S. D. V. Nieuwenhove, W. van der Tempel and M. Kuijk, "Photonic demodulator with sensitivity control," IEEE Sens. J. **7**, 317–318 (2007).
8. J. Shah, H. Pien, and J. Gauch, "Recovery of surfaces with discontinuities by fusing shading and range data within a variational framework," IEEE Trans. Image Process. **5**, 1243 –1251 (1996).

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

9. S. B. Gokturk, H. Yalcin, and C. Bamji, "A time-of-flight depth sensor - system description, issues and solutions," in CVPRW '04: Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04) Volume 3, (IEEE Computer Society, 2004), p. 35.
10. S. Schuon, C. Theobalt, J. Davis, and S. Thrun, "High-quality scanning using time-of-flight depth superresolution," CVPR Workshop on Time-of-Flight Computer Vision (2008).
11. M. Frank, M. Plaue, and F. A. Hamprecht, "Denoising of continuous-wave time-of-flight depth images using confidence measures," Opt. Eng. **48** (2009).
12. T. Schairer, B. Huhle, P. Jenke, and W. Straßer, "Parallel non-local denoising of depth maps," in International Workshop on Local and Non-Local Approximation in Image Processing (EUSIPCO Satellite Event) (2008).
13. L. Jovanov, A. Pižurica, and W. Philips, "Wavelet based joint denoising of depth and luminance images," in 3D TV Conference, Kos Island, Greece (2007).
14. Lj.Jovanov, N. Petrović, A. Pižurica, and W. Philips, "Content adaptive wavelet based method for joint denoising of depth and luminance images," in SPIE Wavelet Applications in Industrial Processing V, (Boston, Massuchusetts, USA, 2007).
15. S. Schulte, B. Huysmans, Pižurica, E. Kerre, and W. Philips, "A new fuzzy-based wavelet shrinkage image denoising technique," in Advanced Concepts for Intelligent Vision Systems (Acivs 2006), (Antwerp, Belgium, 2006).
16. S. De Backer, A. Pižurica, B. Huysmans, W. Philips, and P. Scheunders, "Denoising of multicomponent images using wavelet least-squares estimators," Image Vision Comput. **26**, 1038–1051 (2008).
17. A. Benazza-Benyahia and J. Pesquet, "Building robust wavelet estimators for multicomponent images using Stein's principle," IEEE Trans. Image Process. **14**, 1814–1830 (2005).
18. "3D TV Production [online]," (2009).
19. P. Seitz, "Quantum-noise limited distance resolution of optical range imaging techniques," IEEE Trans. Circuits Syst., I: Regul. Pap. **55(8)**, 2368–2377 (2008).
20. I. Daubechies, *Ten Lectures on Wavelets* (SIAM, Philadelphia, 1992).
21. J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using Gaussian scale mixtures in the wavelet domain," IEEE Trans. Image Process. **12**, 1338–1351 (2003).
22. S. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," IEEE Trans. Image Process. **9**, 1522–1531 (2000).
23. A. Pižurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising," IEEE Trans. Image Process. **15**, 654–665 (2006).
24. S. I. Olsen, "Estimation of noise in images: an evaluation," CVGIP: Graph. Models Image Process. **55**, 319–323 (1993).
25. M. Ghazal, A. Amer, and A. Ghrayeb, "A real-time technique for spatiotemporal video noise estimation," IEEE Trans. Circuits Syst. Video Technol. **17**, 1690–1699 (2007).
26. A. Amer and E. Dubois, "Fast and reliable structure-oriented video noise estimation," IEEE Trans. Circuits Syst. Video Technol. **15**, 113–118 (2005).
27. V. Zlokolica, A. Pizurica, and W. Philips, "Noise estimation for video processing based on spatio-temporal gradients," IEEE Signal Process. Lett. **13**, 337 – 340 (2006).
28. R. Bracho and A. Sanderson, "Segmentation of images based on intensity gradient information," in Proc. IEEE Computer Soc. Conf. on Computer Vision, 341–347(1985).
29. D. Donoho, I. Johnstone, and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," Biometrika **81**, 425–455 (1993).
30. A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for bayesian wavelet based image denoising," IEEE Trans. Image Process. **11**, 545–557 (2002).
31. "Swissranger sr4000 overview [online]," `http://www.mesa-imaging.ch/prodview4k.php` (2009).
32. J. A. Guerrero-Colon, L. Mancera, and J. Portilla, "Image restoration using space-variant gaussian scale mixtures in overcomplete pyramids," IEEE Trans. Image Process. **17**, 27–41 (2008).
33. G. J. Iddan and G. Yahav, "G.: 3d imaging in the studio (and elsewhere," Proc. SPIE **4298**, 48–55 (2001).
34. D. De Silva, W. Fernando, and S. Yasakethu, "Object based coding of the depth maps for 3d video coding," IEEE Trans. Consum. Electron. **55**, 1699–1706 (2009).
35. L. Zhang and W. Tam, "Stereoscopic image generation based on depth images for 3d tv," IEEE Trans. Broadcast. **51**, 191–199 (2005).

1. Introduction

Clean and reliable image features are of fundamental significance for complex tasks such as object recognition, autonomous navigation of robots and biometric authentication. With this, luminance, colour and motion information are often used as features for the scene interpretation task. In practice, these features, extracted from a two-dimensional (2D), or 2D plus temporal

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

scene representation, are often insufficient for unambiguous interpretation of the scene, due to occlusions and the lack of information along the third dimension (depth) of the scene.

The introduction of depth (i.e. range) information into the feature set makes the scene interpretation task more feasible and robust, where various depth estimation techniques exist, based on the use of one or multiple cameras (see [1] for an overview). Most frequently, two cameras are employed for depth estimation, based on the disparity measured between the left and the right camera image [2]. Alternative techniques include depth-from-focus [3], depth-from-shape [4] and depth-from-motion [5]. Some of the latest developments in this area are based on measuring the *time-of-flight* (TOF) [6, 7] of the infra-red modulated light beam.

In TOF imaging, the "luminance" channel (similar to the one in classical video) is accompanied by a "depth" channel containing distance measurements. The main characteristics of TOF depth sensors are high frame rate, good consistency of the produced depth map (with the real scene in most cases), good accuracy and low computational requirements. While this technology has already demonstrated excellent potential, some problems still have to be solved in order to make it widely usable within practical applications. An important problem here is interference between ambient light and the infra-red light source in the TOF camera, which causes errors in the measured depth and therefore makes outdoor usage of TOF cameras difficult. Another significant problem is the presence of noise, which reduces the precision of distance measurement. For TOF sensors, luminance images typically contain much less noise than the corresponding depth images. Hence, the main idea of our approach is to make use of the luminance image, which is of better quality, to improve the performance of the depth image denoising.

The topic of noise reduction in TOF depth images has not been well studied in literature yet (since the TOF technology itself is relatively new) and the reported studies are scarce. One of the first methods for denoising depth images which use the luminance information was presented in [8].

Later, in [9] the first methods for the denoising of depth images acquired by a time of flight camera were presented. In this paper authors present three methods for the denoising of depth images suitable for hardware implementation The first method presented is a standard median filter, where the denoised value is replaced by the median value of the pixels in the neighbourhood. The second is uniform low-pass filtering of the depth image, a type of filter replacing the noisy pixel value with the weighted average of neighbouring pixels. Finally, the third is based on recursive temporal filtering, where the value of each noisy pixel is replaced by an average of the current noise and the previous value of the temporal average. While these methods significantly reduce the variance of the depth measurements, they also produce spatial and motion blur.

In [10], temporal averaging is used to reduce the amount of noise prior to the super-resolution of the depth map. Denoised values here are obtained by calculating the average value of depth for each pixel in 200 frames. Although this method efficiently removes noise, it is only applicable for denoising of the depth images of the static scene. Otherwise, in the case of non-static scene, it would cause severe motion blur.

In a recent approach [11] the authors propose three approaches for depth image denosing. Here we describe the best performing approach from [11], named "Adaptive Normalized Convolution". This approach performs denoising by calculating the weighted average of depth values inside the spatial neighbourhood of the denoised depth value. The depth values that are more reliable here contribute more towards the average depth, while unreliable pixels are given lower significance. The inverse of the amplitude of the received infrared light is used as a confidence measure of depth values. This method takes the spatial relationship of the pixels into account by using Gaussian kernel as a weighting factor, where the depth value at a spatial

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

location (i, j) of the denoised depth image $d_{h,i,j}$ is computed from the raw depth map d using the equation:

$$d_{h;i,j} = \frac{\sum_{k_i=-\frac{n-1}{2}}^{\frac{n-1}{2}} \sum_{k_j=-\frac{n-1}{2}}^{\frac{n-1}{2}} d_{i+k_i,j+k_j} \cdot f^h_{k_i,k_j} \cdot A^2_{i+k_i,j+k_j}}{\sum_{k_i=-\frac{n-1}{2}}^{\frac{n-1}{2}} \sum_{k_j=-\frac{n-1}{2}}^{\frac{n-1}{2}} f^h_{k_i,k_j} \cdot A^2_{i+k_i,j+k_j}}, \tag{1}$$

where $f^h_{k_i,k_j}$ are the coefficients of the smoothing mask with bandwidth h and n is an odd mask size. In order to achieve good adaptation to the local noise variance and the level of detail, depth values are further filtered using a Gaussian filter with several different widths h, creating multiple denoised images with a different level of remaining noise. Since the authors assume spatially uncorrelated noise, the estimated variance of a smoothed pixel is:

$$\sigma^2_{h;i,j} = \frac{\sum_{k_i=-\frac{n-1}{2}}^{\frac{n-1}{2}} \sum_{k_j=-\frac{n-1}{2}}^{\frac{n-1}{2}} \left(f^h_{k_i,k_j} \cdot A^2_{i+k_i,j+k_j}\right)^2}{\left(\sum_{k_i=-\frac{n-1}{2}}^{\frac{n-1}{2}} \sum_{k_j=-\frac{n-1}{2}}^{\frac{n-1}{2}} f^h_{k_i,k_j} \cdot A^2_{i+k_i,j+k_j}\right)^2}. \tag{2}$$

The denoised depth value is estimated as the value whose corresponding new variance calculated using the Eq. (2) has the highest variance below a user-defined threshold σ^2_{thresh}. Thanks to this criterion, every pixel is only averaged over as many neighbours as are strictly required to obtain sufficient "confidence". This way, unnecessary smoothing is avoided.

Another recent depth images denoising method is presented in [12], where authors develop a variant of the non-local method for depth images denoising implemented on GPU. This algorithm restores a depth value by calculating a weighted average of similar pixels:

$$v'(i) = \sum_{j \in \mathbf{W}_i} w(i, j) v(j), \tag{3}$$

where $\mathbf{W}_i$ is a large search window around pixel at the location i, $w(i, j)$ is the weight of pixel j when pixel i is denoised and $v(j)$ is the pixel at the location j. In this paper weights are derived both from the depth and the luminance:

$$w(i, j) = \frac{1}{Z_i} e^{-\frac{1}{h} \sum_{k \in N} \xi_{ik} G_a(||k||_2)(v(i+k)-v(j+k))^2}, \tag{4}$$

where $\xi_{ik} = e^{-\frac{||(\mathbf{v}(i)-\mathbf{v}(i+k))||}{h}}$, G_a is a Gaussian function with standard deviation a and $\mathbf{v}(i)$ is a vector containing depth and luminance values from the neighbourhood of the pixel i and Z_i is the normalization constant. The term ξ constrains the similarity comparison to regions of similar depths belonging to the same surface.

In our previous work [13], we used luminance information for denoising depth data through segmentation into self-similar parts. First we form 3x3 neighbourhoods in both the depth and the luminance wavelet bands, using these neighbourhoods to form feature vectors for k-means clustering, which produces a set of segments with similarly oriented edges. Furthermore, a number of clusters was set manually. This clustering and segmentation were used to obtain parameters of a vector denoiser in each segment, where the denoiser itself was a vector Wiener filter applied to jointly luminance and depth measurements in the wavelet domain. In [14] we present the improved version of [13] where the number of clusters is automatically determined, an approach demonstrating good results in depth image denoising, but which is rather complex and hence less suitable for real time processing.

In this paper we propose a multi-resolution method for denoising depth images that makes use of the correlation between luminance and depth data. An important component of this method is a novel noise standard deviation estimation method using luminance information.

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

The present method employs no clustering. Instead, we integrate the luminance features directly into a novel estimation formula for wavelet coefficients of the depth image, an approach that significantly improves the recovery of the geometric features hidden in noise.

Two main novelties of the proposed approach are the following: (1) a novel noise estimator for depth images acquired using TOF sensors and (2) a novel wavelet domain denoiser for TOF depth images making use of the spatial context and the dependencies between the luminance and depth data. Noise in TOF depth images here is spatially variant and classical noise estimation techniques are usually insufficiently accurate, while often being overly complex for this task. However, our method makes use of the correspondence between the reflected optical signal and noise standard deviation, resulting in a simple and elegant estimator of the noise standard deviation in each pixel. The proposed wavelet denoiser builds further on the fuzzy-logic based denoiser FuzzyShrink for classical images from [15]. We make a generalization of the FuzzyShrink denoiser in order to simultaneously handle two different but correlated imaging modalities; luminance and depth from a TOF sensor. While the membership functions in [15] are 1D functions, in our case they are 2D functions, depending on both depth and luminance data. Moreover, we introduce optimization over anisotropic neighbourhoods in order to better detect and reconstruct edges hidden in noise.

The results show a clear improvement over other recently reported methods for depth image denoising [11], including non-local approach [12] and our previous work [13, 14]. The results also demonstrate an improvement over some recent wavelet domain denoisers for multivalued images [16, 17].

The paper is organised as follows: section 2 gives the necessary background for this work,in particular reviewing the basic principles of TOF sensors in section 2.1 and their noise characteristics in subsection 2.2. Wavelet denoisers for classical images that are in the basis of our work are reviewed briefly in subsection 2.3. In section 3, we present a new method for the estimation of the noise standard deviation in depth images. In section 4, we describe the proposed denoising method, and experimental results are given in the section 5. More specifically, the evaluation of the proposed algorithm in terms of objective image quality measures is presented in subsection 5.1 and the evaluation in terms of the number of occlusions in the artificially created views is presented in subsection 5.2. The conclusions are given in section 6.

2. Background and related work

2.1. Time-Of-Flight principle

Recent techniques for 3D imaging based on the Time-of-flight principle are gaining in popularity as a technology that will be used for 3D capture in future 3D TV systems [18], for 3D biometric authentication, 3D human-machine interaction tool etc.

TOF sensors are based on measuring the time that the light emitted by an illumination source requires to travel to an object and the back to the detector. The system for measuring the distance typically consists of two main components: a light emitter and a light detector, as shown in Fig. 1. The source emits the optical signal, which is modulated in amplitude. In the case of the Swiss Ranger sensor, continuous-wave modulation is used. This signal is reflected from the objects in the scene and received by the detector of the camera. To measure the time of arrival, a Swiss Ranger sensor measures the phase delay between the emitted and detected signals. This type of sensor contains an array of 176x144 pixels, each of which is capable of measuring the phase and the amplitude, where the phase is measured by each pixel in the sensor, thereby creating the complete phase and distance map.

TOF systems based on continuous wave modulation detect phases of the received optical signal using the synchronous demodulation, one of which is the cross-correlation of the received optical signal with the emitted modulated signal [6]:

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

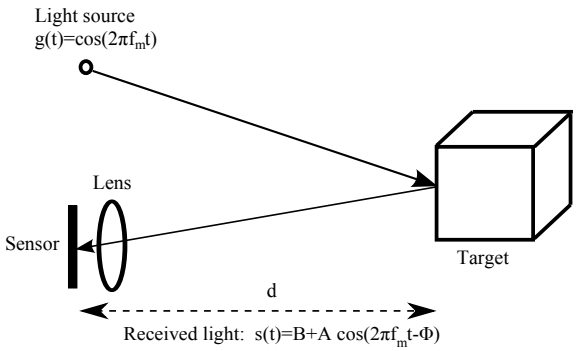


Fig. 1. 3D camera using the Time-Of-Flight principle.

$$C(\tau) = s(t) \otimes g(t) = \lim_{T->\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} s(t) \cdot g(t+\tau) dt. \tag{5}$$

The signal received by the sensor consists of an offset signal (e.g., due to the background illumination) and a sinusoidally modulated signal, due to the reflected waveform from the objects in the scene, where the emitted signal and a reflected signal from the scene are given as:

$$\begin{aligned} s(t) &= B + A \cdot cos(2\pi f_m t - \phi), \\ g(t) &= cos(2\pi f_m t), \end{aligned} \tag{6}$$

where f_m is the modulation frequency, A is the amplitude of the received optical signal, B is its bias and ϕ is the phase offset corresponding to the object distance.

The received modulated signal is demodulated by sampling input optical signal synchronously at four different time instants per period: Ai; i = 0, ..., 3:

$$\phi = \arctan \frac{A_3 - A_1}{A_0 - A_2}, \tag{7}$$

$$B = \frac{A_0 + A_1 + A_2 + A_3}{4 \cdot \Delta t}, \tag{8}$$

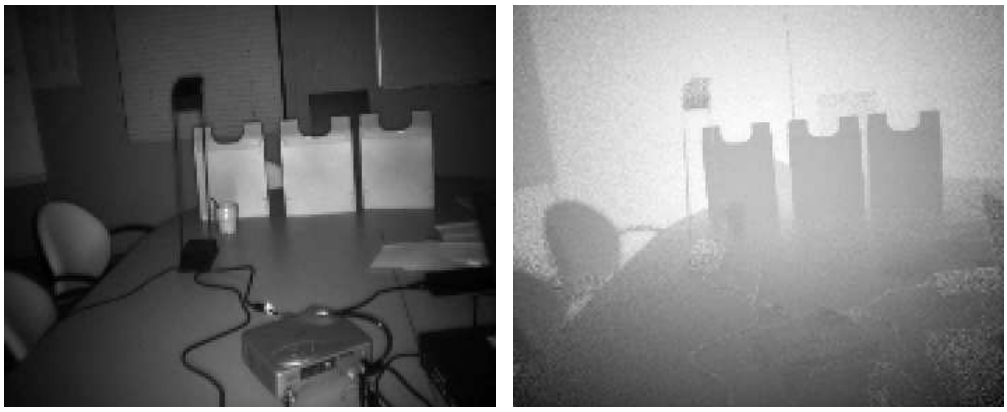
$$A = \frac{\delta}{\triangle t \cdot sin\delta} \cdot \frac{\sqrt{(A_3 - A_1)^2 + (A_0 - A_2)^2}}{2}, \tag{9}$$

where A_i denotes the amplitude at sampling point (temporal instance) i, Δt denotes the integration interval for sampling values A_1, A_2, A_3 and A_4 while $\delta = \frac{\pi \Delta t}{T}$ where T is the modulation period.

2.2. *Noise in depth images*

The optical power of the depth sensor has to meet a compromise between image quality and eye safety. The larger the optical power, the more generated photoelectrons there are per pixel, and hence the higher the signal-to-noise ratio and measurement error of the depth measurements. Due to limited optical power, the depth images are rather noisy and therefore of relatively poor distance measuring error (see Fig. 2). The upper limit on the optical power of the infrared source also limits the detected optical power, boosting noise.

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010



(a) (b)

Fig. 2. (a) Luminance image of the scene. (b) Depth map of the scene.

The most important sources of noise in CCD sensors and photodiode arrays are electronic and optical shot noise, thermal noise, reset noise, $1/f$ noise, and quantization noise [6, 19]. Photon shot noise gives the largest contribution towards the total noise power and shot noise models the statistical nature of the arrival of photons and generation of the electron-hole pairs. Using (7) and taking into account that the noise in samples $A_0,..,A_3$ is uncorrelated and has a standard deviation of $\triangle A_i = \sqrt{A_i}$, quantum noise limited phase error $\triangle\phi$ can be obtained as in [6]:

$$\triangle\phi = \sqrt{\sum_{i=1}^{3}(\frac{\partial\phi}{\partial A_i})^2 \cdot A_i}. \tag{10}$$

If we solve the above equation for certain phase values $(0^o, 45^o, 90^o, 135^o, 180^o, ...)$ we obtain the range measurement error $\triangle L$ as in [6]:

$$\triangle L = \frac{L}{\sqrt{8}} \cdot \frac{\sqrt{B}}{2 \cdot A}, \tag{11}$$

where L is the non-ambiguity distance range. In other words, ΔL represents the random error of depth measurements.

As we can see, the measurement error of the TOF sensor is inversely proportional to the demodulation amplitude A, which depends on the modulation depth, demodulation contrast of the pixel (efficiency of the pixel on CMOS sensor), optical power of the modulated light source and on the distance and the reflectivity of the target (see [6] for further analysis).

In terms of imaging science, the measurement error of a TOF camera corresponds to the variance of noise in depth images; the lower the measurement error (the larger $\triangle L$) is, the higher noise level in the depth image, and vice versa. Moreover, the noise in depth images is highly non-stationary, as we will see in section 3.

Since the noise standard deviation is significantly higher in the depth image, many significant details and features are lost in the noise, while they are still present in the luminance image. Furthermore, we observed that, in the case of static scenes, temporal averaging of the depth sequences indeed reveals details and features hidden in the noise and hence not visible when looking at a single frame. In real situations, the scene is rarely static and the estimated motion

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

is not perfect, which means that, in reality, we cannot remove noise by temporal averaging of only the depth frames.

The most important observation is that the luminance component contains much less noise than the corresponding depth measurements. A typical range image, acquired with a Swiss Ranger camera, has a PSNR of about 34-37 dB, while PSNR values of luminance are typically about 54-56 dB. Both PSNR values for depth and the luminance were measured using ground truth images obtained by temporal averaging of the 200 static frames. The fact that the luminance image contains much less noise than the depth enables us to exploit the luminance information for more reliable denoising of the depth sequence, especially in the areas subjected to the higher illumination, which are especially noisy.

2.3. Wavelet denoisers

For a comprehensive review of the wavelet transform readers are referred to [20]. A large number of wavelet denoising methods have been proposed over the past few years, such as [21], [22], [23], [15], [17] and [16]. In [21] wavelet coefficients are modelled using the Gaussian scale mixtures model and a minimum mean square error estimator is derived for the denoising. An algorithm presented in [22] uses generalized Gaussian distribution to model the statistics of the wavelet coefficients and derives the data-driven shrinkage function. The approaches of [17] and [16] are particularly interesting, since they are applied to the denoising of multivalued signals and therefore for joint denoising of the depth and luminance. The method of [17] uses linear expansion of thresholds optimized using Stein's unbiased risk estimate for vector variables. In the method of [16] vectors containing wavelet coefficients of multivalued variables are modelled using a Gaussian scale mixtures model. The proposed method builds further on two wavelet denoisers for classical images introduced previously by some authors of this paper: ProbShrink [23] and FuzzyShrink [15].

Let

$$w_k = y_k + n_k \tag{12}$$

denote the observed noisy coefficients in a given wavelet sub-band at particular scale and orientation (suppressed in the notation for compactness) and at spatial position k. y_k is the unknown noise-free component and n_k a sample of zero mean white Gaussian noise with standard deviation $\sigma : n_k \sim \mathcal{N}(0, \sigma^2)$. ProbShrink estimator from [23] defines the hypothesis H_1 : signal of interest is present as: $|y_k| > \sigma$ and the opposite hypothesis H_0 : signal of interest is absent as $|y_k| \leq \sigma$ and estimates the noise-free signal component as: $\hat{y}_k = P(H_1|w_k, z_k)w_k$, where z_k is a local spatial activity indicator calculated from the surrounding coefficients (in particular, locally averaged coefficient amplitude).

The FuzzyShrink estimator of [15] is a fuzzy-logic version of ProbShrink, where the estimation of the required probability density functions is replaced by fuzzy-logic rules and fuzzy membership functions. It was shown in [15] that this estimator achieves a similar denoising performance as ProbShrink, while it is simpler to implement and faster. The method uses the same activity indicator z_k as ProbShrink and puts it in a fuzzy logic formalism with the following fuzzy rule:

> IF z_k is a **large activity indicator** AND w_k is a **large coefficient**
> OR z_k is a **large activity indicator** (13)
> THEN w_k is a **signal of interest**.

The fuzzy membership functions for "large coefficient" and "large activity indicator" are shown in Fig. 3. The FuzzyShrink estimator is explicitly written as:

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

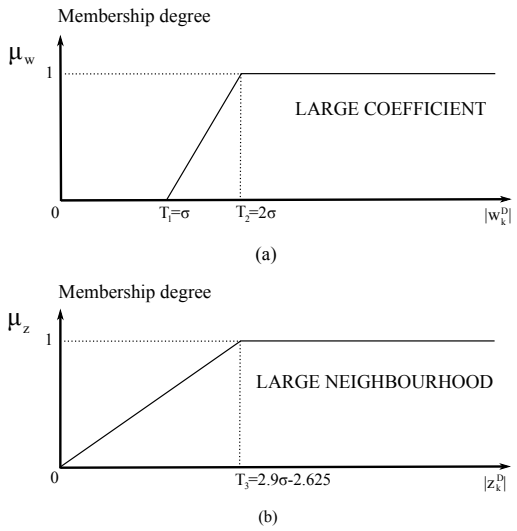


Fig. 3. (a) The membership function LARGE COEFFICIENT denoted as μ_w for the fuzzy set *large coefficient* and (b) the membership function LARGE NEIGBOURHOOD denoted as μ_z for the fuzzy set *large variable*.

$$\hat{y}_k = \gamma(w_k, z_k) \cdot w_k, \tag{14}$$

where $\gamma(w_k, z_k)$ is the degree of activation of fuzzy rule for the wavelet coefficient w_k defined as:

$$\begin{aligned} \gamma(w_k, z_k) &= \alpha + \mu_z(|z_k|) - \alpha\mu_z(|z_k|) \\ \text{with} \quad \alpha &= \mu_z(|z_k|) \cdot \mu_w(|w_k|). \end{aligned} \tag{15}$$

μ_z and μ_w are the membership functions "large activity indicator" and "large coefficient" respectively. As can be seen from Fig. 3, the shape of these membership functions depends on the noise standard deviation σ (for details, see [15]).

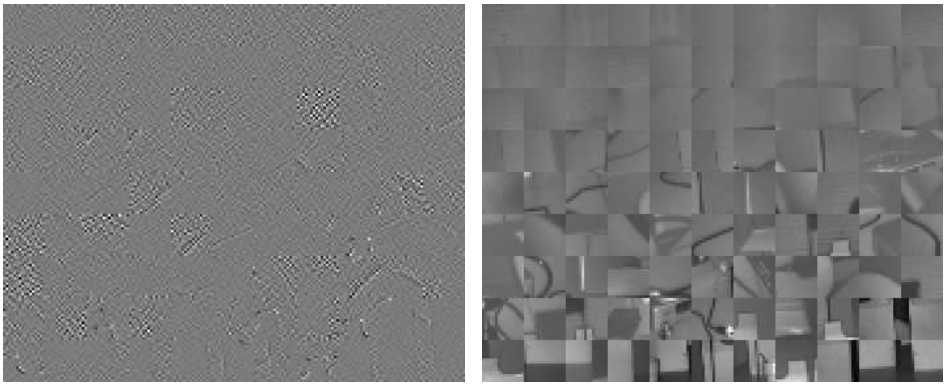
The FuzzyShrink estimator derived in this section was proposed for classical images, and in this paper we extend it for the denoising of depth images, by introducing the features from both luminance and depth images into new membership functions.

3. The proposed noise estimation method for TOF depth images

One of the main problems in noise standard deviation estimation is to find the image regions where the signal of interest (i.e. the edges and textures) is not present, so that the noise is clearly separated from the useful content. Most of the existing noise estimation methods, like [24, 25, 26, 27, 28] estimate a global noise variance in the whole image, while the noise in TOF images is highly non-stationary (see Fig. 2(b)).

Our novel noise estimation method is developed specifically for TOF images. The method is inspired by [22], where signal and noise variance are determined by finding the most similar neighbourhoods to the current one. The main idea of the proposed method is using the amplitude of the reflected signal for direct noise estimation at every pixel. A similar idea was used

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010



(a) (b)

Fig. 4. (a) Segments of depth image. (b) Segments of luminance image (both ordered by standard deviation of luminance segments).

(a) (b) (c)

Fig. 5. (a) Noise standard deviation estimate using Donoho's noise estimator. (b) Noise estimation using the proposed approach. (c) Noise in the depth map.

in [11], where the authors use values of the amplitude as a confidence measure for weighted averaging. Here we use the amplitude image to directly estimate the noise standard deviation as detailed below.

Theoretical analysis of the time-of-flight sensors precision is given in [6] and also in section 2.2. Based on that analysis and on our experiments, we propose a model for the non-stationary noise standard deviation as a function of the inverse amplitude $\frac{1}{A_l}$:

$$\hat{\sigma}_l = C \cdot \frac{1}{A_l}, \tag{16}$$

where C is a constant for a given camera.

We have experimentally validated the model on different depth images and in all cases it fits the experimental data very well, as illustrated on the scatter plot in Fig. 6, where each point represents the measured noise standard deviation at a given position in the image against the measured amplitude at the same position. The noise standard deviation was measured from 200 frames of a static scene, and the amplitude was measured from one particular image in this sequence (its variation over different images of the static scene is negligible with respect to the variation in the depth image, see Figs. 6 and 7). The proposed noise model from Eq. (16) is overlaid on top of the scatter plot and shows a very good agreement with the measured

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

data. Please note that this good agreement of the proposed model and the measurements is not surprising, because the noise model directly relates to the range measurement error model from Eq. (11) if the noise is interpreted as measurement error. In our experiments constants C estimated from various scenes do not vary significantly as illustrated in Fig. 6. In particular, the plot in Fig. 6(d) shows that the noise model with the constant C estimated from one depth image fits well the measured scatter plot of another depth image. We repeated such experiments on a multiple depth images from various scenes and this conclusion was valid in all the analyzed cases.

In the experiments discussed above we estimated the constant C from a number of frames of static scenes. However in reality we often do not have such a static scene at our disposal for the initialization of the algorithm. Our goal is to estimate the constant C from the available video sequence at hand. For doing so we will first identify the image blocks dominated by noise. Therefore, we divide the luminance image and the HH_1 band of the wavelet decomposition of the depth image into non-overlapping blocks and sort them according to the mean variance in luminance blocks, in increasing order as in Fig. 4. For each of the blocks b_l in the HH_1 band of the depth image, we estimate noise variance as:

$$\hat{\sigma}_{b_l}^2 = \frac{1}{N}\sum_{k=1}^{N}(X_{l,k} - m_l)^2, \tag{17}$$

where m_l is the mean value of the HH_1 coefficients in l-th block and $X_{l,k}$ is the k-th coefficient in the current block. Let A_{b_l} denote the mean luminance in block l. We take $P \cdot 100$ percent of the blocks from the luminance with the smallest variance and their corresponding blocks from HH_1 wavelet band of the depth image, and estimate the constant C as:

$$C = \frac{1}{P \cdot B}\sum_{l=1}^{PB}\frac{A_{b_l}}{\hat{\sigma}_{b_l}}, \tag{18}$$

where B is the number of the blocks in the image. In practice, we use 3% of the blocks for noise estimation, while the size of the blocks is 8x8 pixels. An alternative way of estimating noise standard deviation is by taking overlapping neighbourhoods centered around each pixel and applying the median absolute deviation estimator (MAD) of Donoho [29] on each block. This local MAD estimator gives a smeared estimate where all the details are lost, as can be seen in Fig. 5. The results shown in section 5 (Table 1 and Fig. 16) illustrate that using the proposed noise estimation method yields much better denoising performance. On average, depth images denoised using the proposed noise estimation method have 0.4dB better PSNR than the images denoised using the local MAD estimator. Moreover, a significant advantage of the proposed method is its computational simplicity; the fact that the estimates of the noise variance at each spatial position are immediately available after one division.

The proposed method estimates noise from one depth and one luminance frame, which makes noise estimation possible in the cases where only recorded depth sequence of the non-static scene is available.

4. The proposed denoising algorithm

Our approach builds further on the ProbShrink and FuzzyShrink estimators for classical images introduced in section 2.3. However, we now introduce an entirely new dimension in this approach, by combining two different imaging modalities. Furthermore, we propose an elegant way to introduce information from one image modality into recovering discontinuities hidden by noise in another imaging modality.

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

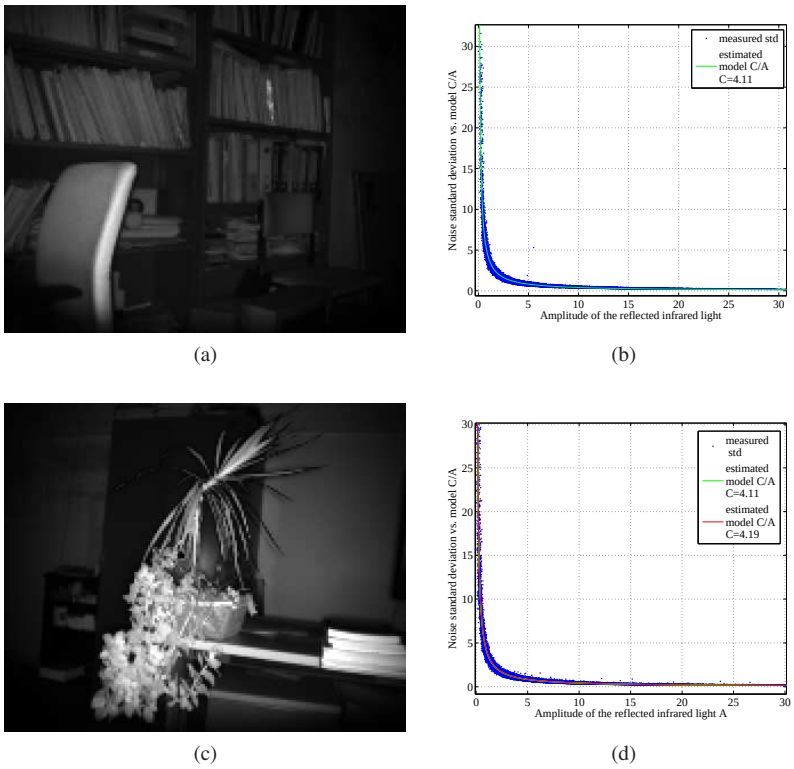


Fig. 6. (a) Amplitude image D1. (b) The corresponding scatter plot of the measured noise standard deviation σ versus the amplitude A and the fitted noise model ($C = 4.11$). (c) Another amplitude image D2. (d) The corresponding experimental $\sigma - A$ scatter plot with fitted noise models C/A using C estimated from the corresponding data ($C = 4.19$) and using C estimated from image D1 ($C = 4.11$).

4.1. Luminance driven depth denoiser

The main idea of the proposed approach is that the wavelet coefficients of the depth image should be estimated according to the probability (or evidence) that they represent the signal of interest, given *both* depth and luminance data. In the following, we use the superscript "D" to denote depth and superscript "L" to denote luminance. By adopting this criterion, we estimate the wavelet coefficient at the spatial position k as:

$$\hat{y}_k^D = \tilde{\gamma}(\mathbf{w}_k, \mathbf{z}_k) \cdot w_k^D, \tag{19}$$

where $\mathbf{z}_k = \begin{bmatrix} z_k^D & z_k^L \end{bmatrix}^T$ denotes a vector containing local spatial activity indicators (LSAI) at the position k in the depth and luminance images, $\mathbf{w}_k = \begin{bmatrix} w_k^D & w_k^L \end{bmatrix}^T$ denotes a vector containing the values of the wavelet coefficients in depth and luminance image at the location k and $\tilde{\gamma}$ denotes a wavelet estimator function.

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

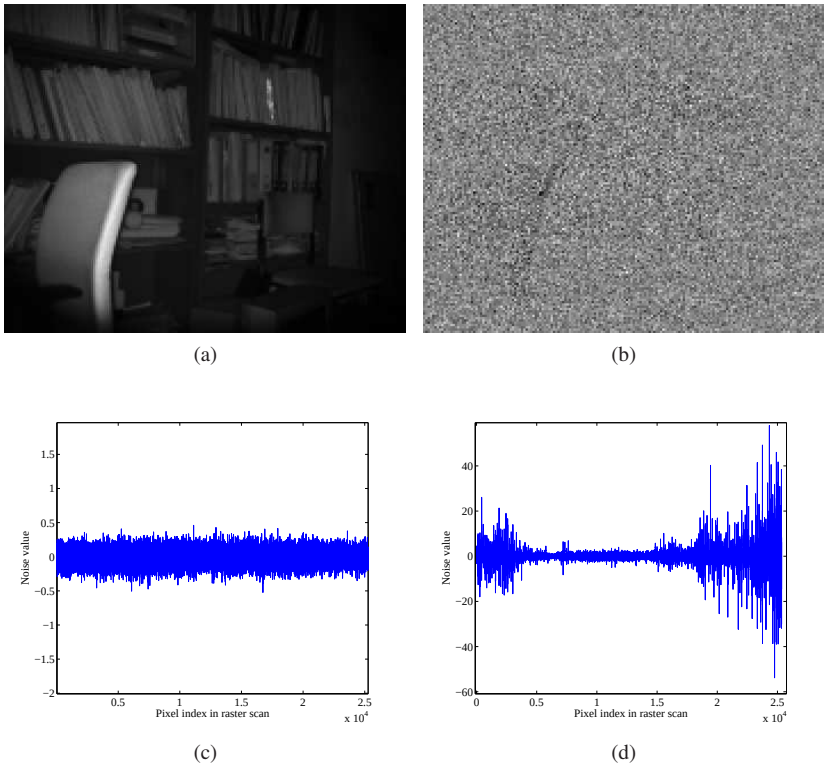


Fig. 7. (a) An amplitude image. (b) Noise in the amplitude image from (a);(c). (c) Raster scan of the noise image in (b). (d) Raster scan of the noise in the corresponding depth image.

4.2. Luminance-Depth membership functions

To proceed, we need to generate the membership functions $\mu_w(\mathbf{w}_k)$ and $\mu_z(\mathbf{z}_k)$ from Fig.3 so that they depend on vectors $\mathbf{w}_k$ and $\mathbf{z}_k$, respectively. Since the original 1D functions depend on the noise standard deviation σ, a natural extension to 2D is to make them dependent on the noise covariance matrix C_n

$$\mu_{\tilde{w}}(w_k) = \mu_w(\|C_n^{-\frac{1}{2}} w_k\|). \tag{20}$$

In this way, we obtain:

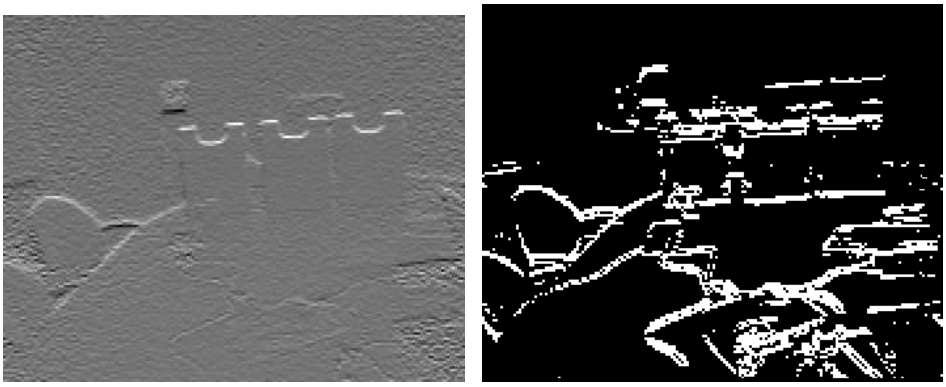
$$\mu_{\tilde{w}}(\mathbf{w}_k) = \mu_w(C_n^{-\frac{1}{2}}\mathbf{w}_k) \quad \text{and} \quad \mu_{\tilde{z}}(\mathbf{z}_k) = \mu_w(C_n^{-\frac{1}{2}}\mathbf{z}_k). \tag{21}$$

The estimator of the noise covariance matrix makes use of our noise estimator from 3, and is detailed in section 4.4. If the matrix C_n is positive semi-definite,

$$\|C_n^{-1/2}\mathbf{w}_k\|^2 = \mathbf{w}_k^T C_n^{-1}\mathbf{w}_k = T^2 \tag{22}$$

represents the equation of ellipsoid in a two dimensional space. When the point $\mathbf{w}_k$ is positioned inside the ellipsoid we consider that the signal of interest is not present at the location k. In the

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

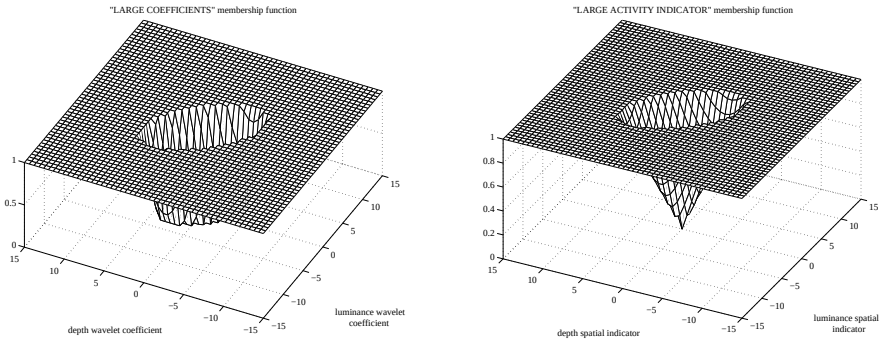


(a) (b)

Fig. 8. (a) Horizontal wavelet band of depth image. (b) Detected signal of interest.

opposite case, we assume that the signal of interest is present at the location k. The signal of interest detected in this way from a depth subband is shown in white in Fig. 8 b.

The parameter T from the Eq. (22) determines the sensitivity of the SOI detector. We found experimentally, by observing the denoising performance on multiple depth images, that the best choice for threshold is $T = 1.2$. The choice of the value of this parameter is not critical since the PSNR drops for 0.1dB if T is increased or decreased for 0.2. Then we evaluate the membership function depending on the value of the vector variable $\mathbf{w}_k$. This is analogous for the second membership function $\mu_{\tilde{w}}(\mathbf{w})$. The resulting functions $\mu_{\tilde{w}}(\mathbf{w})$ and $\mu_{\tilde{z}}(\mathbf{z})$ are shown in Fig. 9.



(a) (b)

Fig. 9. a) "LARGE COEFFICIENTS" membership function and b) "LARGE ACTIVITY INDICATOR" membership function as generalizations of the corresponding membership functions from Fig. 3.

The region defined by the ellipsoid $\mathbf{w}_k^T C_n^{-1} \mathbf{w}_k < T^2$ corresponds to the first segment of the fuzzy membership function μ_w from the Fig. 3, for the values $w_k < \sigma$. The second region defined by the area $T^2 \leq \mathbf{w}_k^T C_n^{-1} \mathbf{w}_k < 4T^2$ corresponds to the segment $\sigma \leq w_k < 2\sigma$ and finally, the area defined by the equation $\mathbf{w}_k^T C_n^{-1} \mathbf{w}_k \geq 4T^2$ corresponds to the segment $w_k \geq 2\sigma$ in the 1D membership function.

We define the "LARGE ACTIVITY INDICATOR" membership function using spatial indi-

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

cators defined in section 4.3 as:

$$\mu_{\tilde{z}}(\mathbf{z}_k) = \mu_z(\|C_w^{-\frac{1}{2}}\mathbf{z}_k\|), \tag{23}$$

where z_k is the vector containing spatial indicators for depth and luminance defined above. The shape of the membership function "LARGE NEIGBOURHOOD" is shown in Fig. 9. Here the area determined by the equation $w_k^T C_w^{-1} w_k < (2.9T + 2.625)^2$ corresponds to the segment $z_k < 2.9\sigma - 2.625$ of the membership function μ_z from the Fig. 3, and the constant part correspond to the segment of one dimensional membership function where $z_k > 2.9\sigma - 2.625$. The resulting

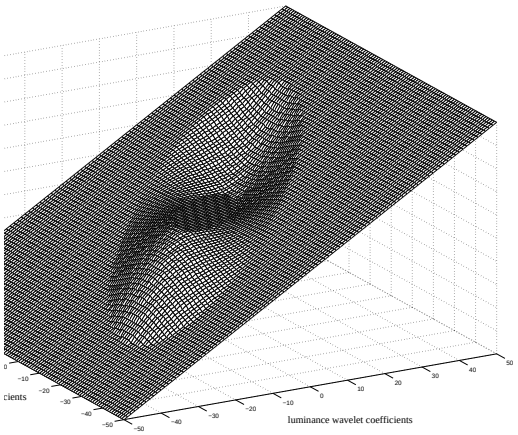


Fig. 10. An illustration of the proposed estimator functional dependence on luminance indicator and noisy coefficient value.

function is obtained by combining the above two membership functions in a following way:

$$\begin{aligned} \tilde{\gamma}(\mathbf{w}_k, \mathbf{z}_k) &= \alpha + \mu_{\tilde{z}}(|\mathbf{z}_k|) - \alpha\mu_{\tilde{z}}(|\mathbf{z}_k|) \\ \text{with} \quad \alpha &= \mu_{\tilde{z}}(|\mathbf{z}_k|) \cdot \mu_{\tilde{w}}(|\mathbf{w}_k|), \end{aligned} \tag{24}$$

where $\mu_{\tilde{z}}$ and $\mu_{\tilde{w}}$ are vector membership functions "LARGE NEIGHBOURHOOD" and "LARGE COEFFICIENT" respectively. The resulting estimator defined in the Eq. (24) is shown in Fig. 10 for constant values of noise standard deviation and spatial indicators. The estimated value of the depth wavelet coefficent depends on the value of noisy luminance and depth wavelet coefficients, and the values of spatial indicators from luminance and depth. For example, if the wavelet coefficient of the depth image has a small value and the luminance coefficient has a larger value, the value of the depth wavelet coefficient will be multiplied by the bigger value, as shown in Fig. 10. If the value of the luminance wavelet coefficient is smaller, however, the value of the depth wavelet coefficient will be multiplied by the smaller value, since the values of luminance does not indicate the existence of edge at the current location. The input-output characteristics of the proposed estimator shown in Fig. 10 facilitates the restoration of the edges in depth image by using the information from the luminance.

The estimator from Fig 10 changes its shape when the spatial indicators change their values. If the spatial indicators from the depth and the luminance increase their values, suggesting

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

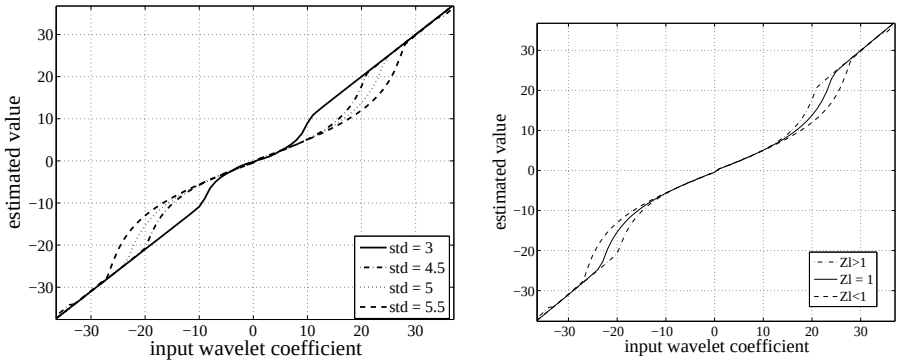


Fig. 11. (a) Shrinkage functions for different values of noise variance. (b) Shrinkage functions for different values of spatial indicator.

that the edges are present in the given neighbourhood, the elliptical part of the characteristics becomes smaller. This means that the depth wavelet coefficient will be multiplied by the bigger value, since the spatial indicator suggest the existence of the edges. The shape of the estimation function varies for each pixel depending on the standard deviation of noise and on the values of spatial indicators. For example, if, at some location, the noise standard deviation is larger, the estimator dilates, the opposite being true in the case of smaller noise.

This is illustrated in Fig. 11b, which displays the estimator as a function of the depth coefficient for a fixed value of luminance data, different noise levels and different activity indicators. The dependence on the estimated noise level is shown in Fig. 11a. Clearly, as the noise variance increases, the shrinkage characteristics become dilated as a consequence of the larger noise dependent threshold, which is defined by the covariance matrix. Without the loss of generality in this figure, we only observe the influence of the spatial indicator from a luminance image, since both activity indicators z_k^L and z_k^D have similar influence on the shrinkage function. A "Neutral" case, where the spatial surrounding from the luminance data gives no preference to "edge" or "no edge", corresponds to a curve denoted with $z_l = 1$. When the luminance activity indicator is in favour of edge (signal of interest) presence ($z_L > 1$) the coefficient is shrunk less than in the neutral case. The opposite is true when the activity indicator favours the absence of the signal of interest ($z_L < 1$) The values of spatial indicators z_k and the locally estimated noise variance σ_k jointly affect the shape of the estimator function. Furthermore, in the most significant practical situation, high values of luminance spatial indicator compensate for high values of noise variance, facilitating a successful recovery of depth image features, masked by a high values of noise.

4.3. Introducing anisotropy into activity indicator

In order to further improve the performance of our estimator, we introduce anisotropic neighbourhoods, illustrated in Fig. 12 in the calculation of the activity indicator. The main idea is the following: in order to trace edges efficiently, we need a relatively large neighbourhood. However, if this neighbourhood is isotropic, the activity indicator will be dominated by non-edge coefficients and the edge will be falsely rejected. Hence, we will test whether an edge exists in any of the anisotropic sub-neighbourhoods and take the most indicative one. This is related to the hierarchical MRF model from [30], but tailored to a different application here and realized in a different way.

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

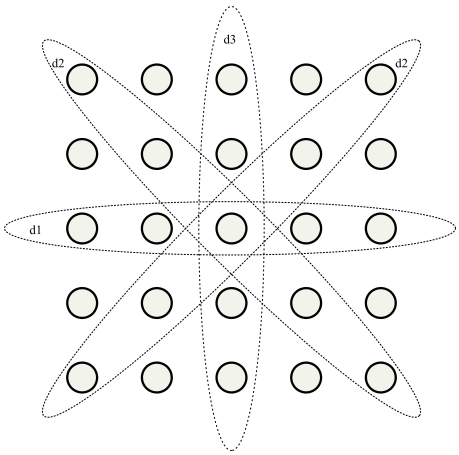
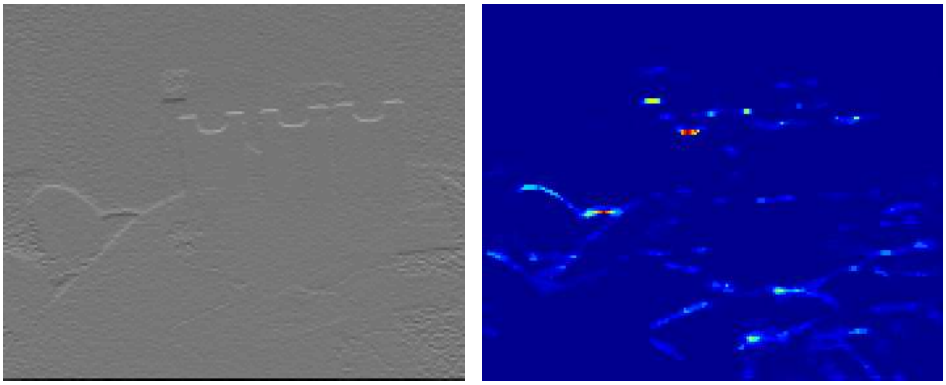


Fig. 12. An illustration of the directional windows. Spatial indicators are formed by summing along the directions d_i.



(a) (b)

Fig. 13. (a) Horizontal (LH) band of the luminance image from Fig. 2. (b) Spatial indicator obtained by directional filtering and combining depth and luminance images.

For each scale and orientation in wavelet decomposition, we form the sums of absolute values of wavelet coefficients $z_{d_i} = \sum_{k \in d_i} |w_k^D| + |w_k^L|$ for three directions d_1, d_2 and d_3 shown in Fig. 12. The pixel sets d_1 and d_3 contain the vertical and horizontal neighbours of the current pixel, while the set d_2 contains all the diagonal neighbours of the current pixel. For each pixel we choose the set d_{max} with maximal value of the sum of the coefficients inside it and form spatial indicators z_k^D and z_k^L as $z_k^D = \frac{1}{\#d_{max}} \sum_{l \in d_{max}} |w_l^D|$ and $z_k^L = \frac{1}{\#d_{max}} \sum_{l \in d_{max}} |w_l^L|$. The values of the spatial indicators obtained in this manner are shown in Fig. 13. As we can see from this figure, even though some edges are masked by noise, it is still possible to have a strong evidence of the existence of signal of interest, thanks to the luminance wavelet coefficients.

4.4. Noise covariance matrix estimation

We estimate the noise variances of depth and luminance as:

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

$$\begin{bmatrix} \hat{\sigma}_D \\ \hat{\sigma}_L \end{bmatrix} = \begin{bmatrix} \hat{\sigma}_l \\ \frac{med(|HH_1^L|)}{0.6745} \end{bmatrix}, \tag{25}$$

where $\hat{\sigma}_l$ is the locally estimated noise variance of depth image described in section 3. Noise variance in the luminance image is estimated globally since we found experimentally, by subtracting reference noise-free luminance image from the noisy luminance image, that it is constant for all spatial coordinates. Next, depth and luminance wavelet bands are scaled by the corresponding standard deviation estimates from above; first we define the scaled sub-bands of luminance and depth as $HH_{1S}^D = HH_1^D/\hat{\sigma}_D$ and $HH_{1S}^L = HH_1^L/\hat{\sigma}_L$. We estimate the cross correlation coefficients at each 16x16 block as follows:

$$\hat{\sigma}_{DL} = \hat{\sigma}_{LD} = (med(|HH_{sD1} + HH_{sL1}|) - med(|HH_{sD1} - HH_{sL1}|))^2. \tag{26}$$

Finally the cross correlation coefficient is obtained as:

$$\rho_{LD} = \hat{\sigma}_D \cdot \hat{\sigma}_L \cdot \hat{\sigma}_{LD}. \tag{27}$$

The noise covariance matrix is then:

$$C_n = \begin{bmatrix} \sigma_D & \rho_{LD} \\ \rho_{LD} & \sigma_L \end{bmatrix}. \tag{28}$$

5. Experimental results

We evaluate the proposed method on both real 3D images acquired by a Swiss Ranger SR 3100 TOF camera [31] and on clean sequences to which we add noise. In the case of real 3D images, we obtain "ground truth" for performance evaluation by temporal averaging over 80 frames of a static scene.

Table 1. PSNR values of the denoised depth images

Method	"Closet"	"Bookshelf"	"Table"
Noisy	34.07	37.15	36.18
Clust [13]	34.64	32.67	42.06
ESURE [17]	34.93	38.44	38.70
AWG [11]	35.32	38.83	40.27
Vector GSM [16]	35.29	38.99	39.90
Neighbourhood Vector GSM [16]	37.05	40.35	43.86
Non-local method from [12]	36.5	38.7	42.10
Proposed	**38.15**	**40.54**	**44.17**
Spatially adaptive vector GSM	37.63	40.03	43.80

We compare performance of the proposed algorithm with two state-of-the art methods for denoising multivalued signals from [16, 17], with the best performing approach from the recent method for depth image denoising ("Adaptive Normalized Convolution" or "Adaptive Weighted Gaussian") from [11], with the non-local depth denoising method from [12] and with our previous non-local approach presented in [14]. Vector methods used for comparison here estimate covariance matrices from the data being denoised. For all methods used for comparison we the parameter values suggested by the authors which yielded as the optimal ones.

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

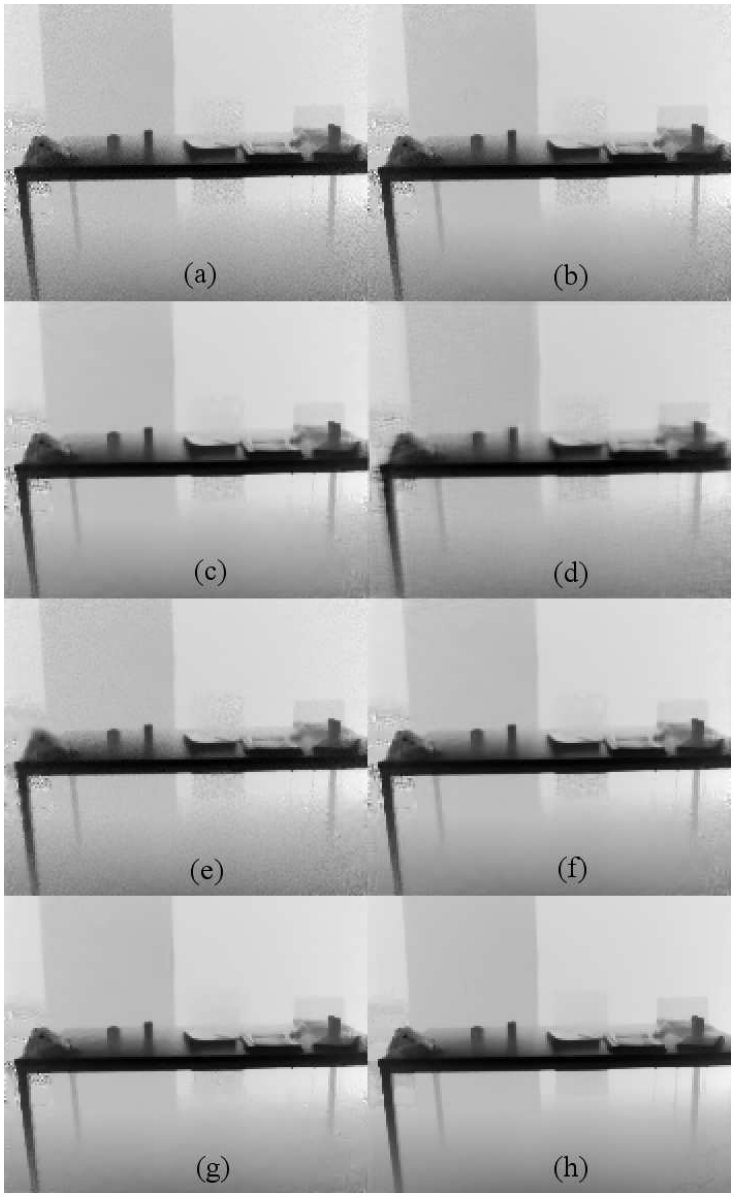


Fig. 14. Denoising result for the "Closet" depth image. (a) Noisy depth image produced by the TOF camera. (b) Image denoised using method from [17]. (c) Image denoised using method from [16]. (d) Image denoised using method from [13]. (e) Image denoised using method from [11]. (f) Image denoised using our extension of GSM vector method from [16]. (g) Image denoised using proposed method. (h) Noise-free reference image.

Noise removal algorithms provided with the camera were turned off in order to have realistic noisy sensor data. Furthermore, the modulation frequency was set to 30 Mhz, and integration time to 40ms. In all experiments described in this paper, we use a non-decimated wavelet decomposition with Daubechies db2 wavelet, and two levels of decomposition.

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

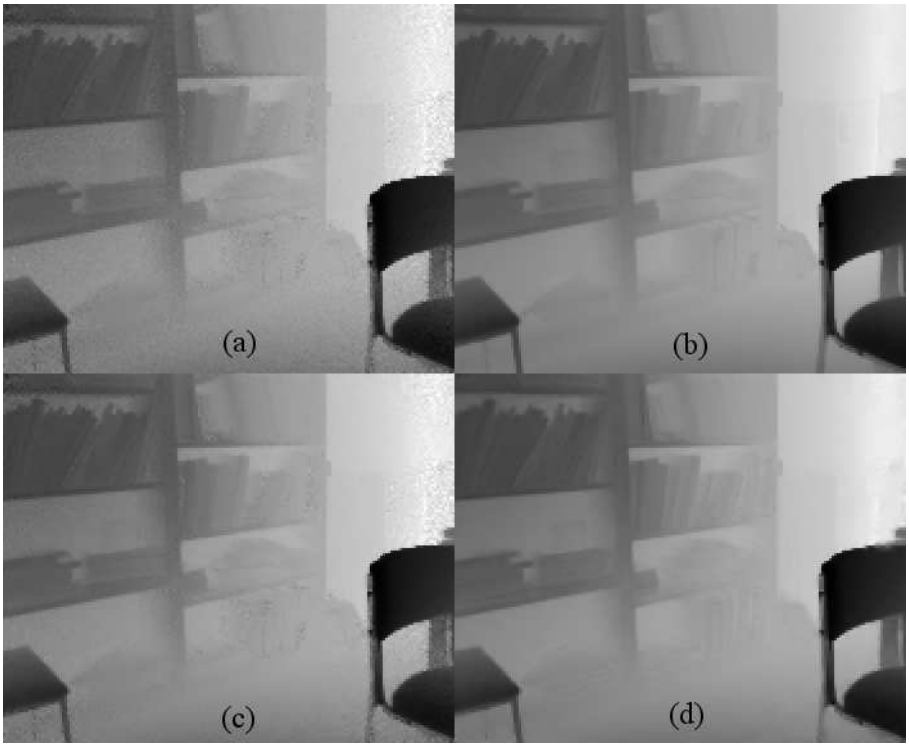


Fig. 15. Results for the "Bookshelf" depth image. (a) Noisy depth image produced by the TOF camera (b) Noise-free reference image. (c) Image denoised using method from [17]. (d) Image denoised using proposed method.

We also implement an improved version of this approach with spatially varying noise and signal covariance matrices, similar to [32]. In this case, we divide both luminance and depth images into overlapping blocks of 16x16 pixels, where the blocks are shifted by 8 pixels along each axis. For each block b_l, we estimate the noise covariance matrices as described in Subsection 4.2

The covariance matrix of the noisy wavelet coefficients is obtained by putting all the coefficients of 16x16 blocks into $N \times 1$ vectors hh_{s1}^D and hh_{s1}^L and calculating the vector product:

$$C_y = [hh_{s1}^D \quad hh_{s1}^L]^T \times [hh_{s1}^D \quad hh_{s1}^L], \tag{29}$$

where N is the number of pixels in the block. A covariance matrix of the signal is obtained by subtracting the noise covariance matrix C_n from C_y.

5.1. The evaluation of the algorithm on the real sequences

In this section we compare the results of the proposed method with the reference methods from the literature, on the sequences obtained using Swiss Ranger SR 3100 TOF camera.

The corresponding PSNR (Peak signal-to-noise ratio) values are given in Table 1. The method which does not use the spatial indicator from the luminance image (i.e. reduces to the FuzzyShrink method) on average gives 0.8dB smaller PSNR values than the proposed scheme. These results clearly show the advantage of using an activity indicator from the luminance im-

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

(a) (b)

Fig. 16. (a)Depth image denoised using Donoho's noise estimation. (b) Depth image denoised using the proposed noise estimation method.

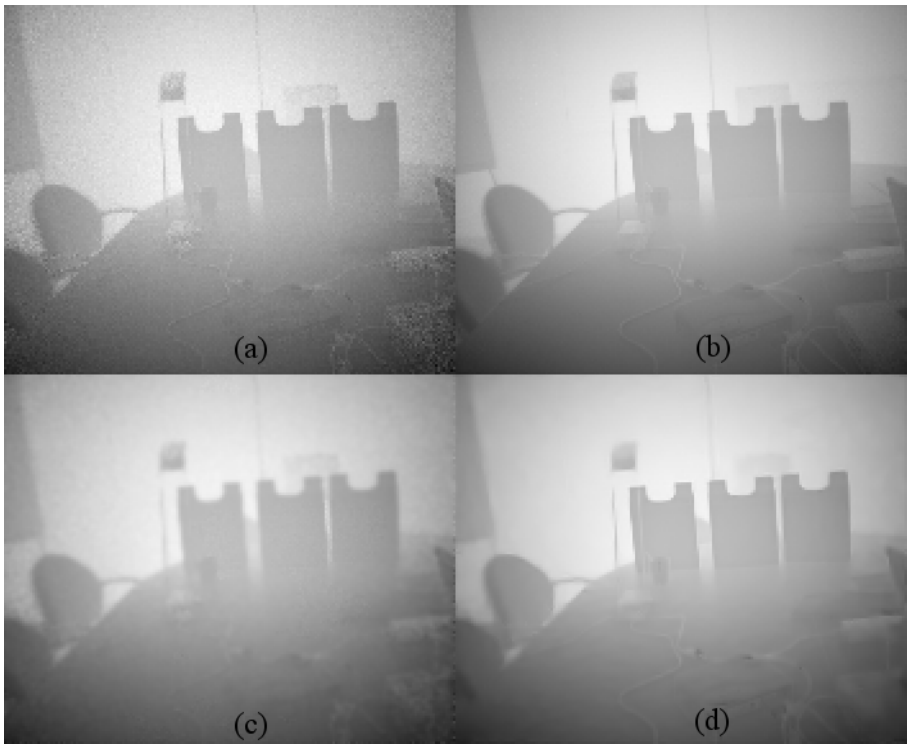
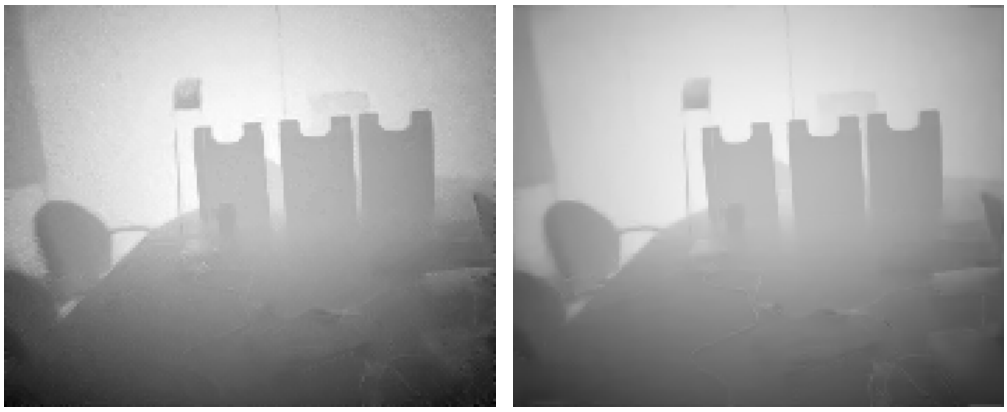


Fig. 17. Denoising results for the "Table" depth image. (a) Noisy depth image produced by the TOF camera. (b) Noise-free reference image. (c) Image denoised using method from [11]. (g) Image denoised using proposed method.

age to denoise the depth data. In Figs. 15-17, we compare the performance of the proposed method to multivalued reference methods [12, 17].

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010



(a) (b)

Fig. 18. (a)Depth image denoised non-local method from [12]. (b) Depth image denoised using the proposed method.

The results in Table 1 show that the new method significantly outperforms all reference methods in PSNR sense. The method of [17] is denoted as ESURE, [13] as Clust, [16] as vector GSM and neighbourhood vector GSM, SA VGSM denotes a spatially adaptive version of [16] described earlier in this section as Spatially adaptive vector GSM, Adaptive Weighted Gaussian method from [11] as AWG and the method from [12] is denoted as non-local. Visually, the proposed method also outperforms our previous approach [13]. Here, ESURE stands for Extended Stein Unbiased Estimator, while GSM stands for Gaussian Scale mixtures. This can be observed, e.g., in the image from Fig. 17: edges are in general sharper and details like the holder of the lamp are better reconstructed. Moreover, the proposed method is much faster. While our previous clustering based method [13, 14] takes about 1.5 minutes to process a 176x144 depth map, the new algorithm does it in real-time (30 frames per second). In other test images, the new method also outperforms the one in [13], especially in the parts of the images containing stronger noise. The most significant advantage of our proposed method is its adaptivity to noise variance in depth image, which results in much better PSNR values of denoised depth images. As stated earlier, the noise variance depends on the strength of the reflected modulated signal, which is a function of the emitted optical power, the reflectance of the materials in the scene (which is highly variable) and the distance of the objects in the scene. Hence, in order to successfully suppress noise in depth images, the denoising algorithm should be able to adapt to the changes in noise level. Otherwise, some parts of the depth image remain noisy. This can be particularly seen in Figs. 15-17 for the methods from [16, 17], which do not adapt to the spatially varying variance of noise.

Vector methods from [13, 17, 16] tend to leave a certain amount of noise on the borders of the image after filtering, where the noise standard deviation is slightly larger than in the centre of the image, as can be seen in Figs. 15, 17 and 19. The method from [17] tends to preserve the edges better than the one from [16], but it also leaves a significant amount of noise in the image. The proposed method successfully removes the noise from the whole image, especially in the parts with higher noise intensity, while preserving details (see, e.g., legs of the furniture and books on the shelf in Figs. 15 and 17).

The proposed method outperforms reference vector denoising methods from [16] and [17] in PSNR sense (see Table 1 and visually (see 15-17), while it is significantly simpler than the vector methods.

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

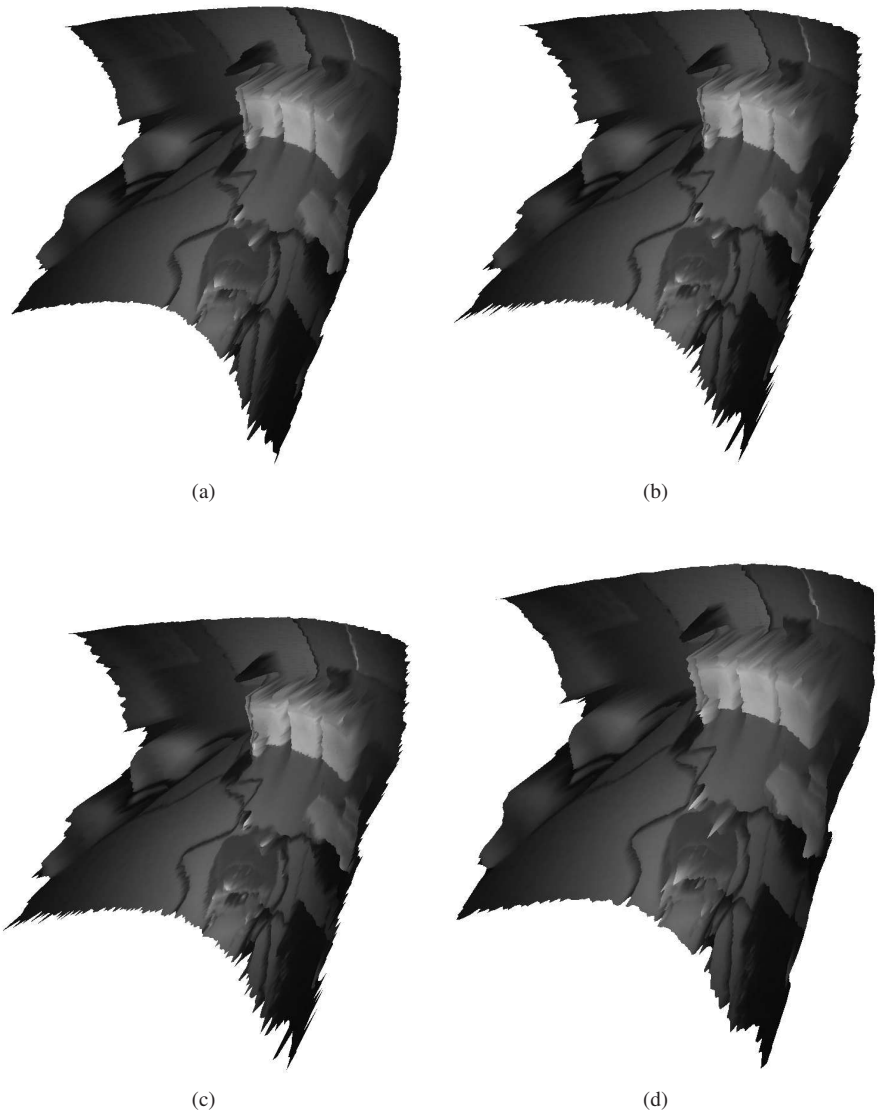


Fig. 19. (a) Rendering of a scene with noise-free reference depth map. (b) Rendering of the scene using noisy depth map. (c) Rendering of the scene using depth map denoised using the method from [11]. (d) Rendering of a scene with the depth map denoised using proposed method.

Next, we compare our results to the results of [11]. The method [11] is very successful in suppressing noise, since it adapts the strength of filtering to the amount of noise in image through several iterations. Therefore, noise is suppressed homogeneously over the whole surface. However, in depth maps that contain fine details, such as the one shown in Fig. 17 images denoised using the method from [11] tend to oversmooth the result. This is especially visible if one observes small objects on the table, for example cables, mugs and the holder of the lamp that is lost. Our proposed method preserves these details much better, as can clearly be seen in the

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

Fig. 20. Virtual views generated using (a) noisy depth image, (b) depth image denoised using method from [11], (c) depth map denoised using depth image using the proposed method, (d) noise-free depth image.

same figure, as is also confirmed by the corresponding PSNR values in Table 1. We also compare our method with the non-local method from [12] in Fig. 18. The method of [12] applies non-local denoising on depth images, which denoises each pixel by calculating a weighted sum of all pixels in the image. The weights are calculated using the exponential kernel, and depend on the Euclidean distances of the depth values inside NxN patches. We can see from the Fig. 18 here that both methods preserve the details in the depth image quite well. However, noise is not completely removed in some regions of the image denoised using the method from [12]. On the other hand, noise is removed homogeneously in the depth image denoised using the proposed method, while details, such as in the cables on the table and the boundaries of the object, are preserved to a great extent.

We also make a comparison of 3D visualizations of the results produced by different methods. Fig. 19 shows the visualizations of the reference noise-free point cloud, noisy point cloud, point cloud denoised using spatially adaptive method of [11], and the point cloud denoised using the proposed spatially adaptive algorithm. The point cloud is represented by a regular triangle mesh, with the per face textures. As can be seen in Fig. 19, z-coordinates of points from noisy point cloud differ significantly from the mean value represented by noise free image. This makes the usage of depth cameras in some applications impossible. For example, in biometric applications, noise can cause errors in recognition, and in the case of robot navigation, wrong decisions can be made and the shape of some object wrongly acquired in reverse engineering applications. He point cloud denoised using [11] shows significantly less variance, but in the regions that have higher noise variance, the shape of the objects still significantly diverges from the reference, which can be seen at the borders of the image (high variations in a region that should be flat). From Fig. 19 it can be easily seen that the point cloud denoised using our method removes almost all unwanted variations caused by noise from flat parts, while

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

preserving fine details in range intact.

5.2. Evaluation of the proposed algorithm on the sequences with artificially added noise

In addition to these sequences, we use the "Interview" sequence, where Gaussian noise with the spatially varying noise standard deviation was added to the clean depth map.

The Interview sequence is a depth sequence containing color information created using ZCAM from 3DV Systems described in [33]. This type of sensor is based on the indirect measurement of the time of flight using a fast shutter technique which yields much less noise than the standard time of flight cameras based on intensity modulation. This sequence is widely used as a perfect reference test sequence in literature, e.g. as a clean reference for coding in [34] and for creating virtual views in [35]. The resolution of the Zcam depth map and corresponding colour and segmentation images is 720x576 pixels. We add the artificial noise to this sequence that mimics the noise in images produced by Swiss Ranger sensors.

In adding artificial noise we do not take into account the exact optical models, but we rather add noise according to the simplified model proposed in this paper. In particular, we first create the amplitude image from the recorded color image by dividing its luminance by the square of the distance as it is commonly done. Then we add artificial noise to the depth image according to the model from Eq. (16) as a function of the amplitude A_i at the corresponding pixel i. The value of the constant C is set to 4.25.

Finally, we make a comparison in terms of the number of occlusions in virtual views created using depth maps. Depth images have very important application in transmission of 3D TV, since the signal can be transmitted using depth map and image of the central view on the scene. This system was adopted since depth images can be much better compressed then the ordinary images, due to their lower entropy. In this way, 3D TV program can be transmitted using much less bandwidth than would be needed in the system where two or multiple separate views were transmitted. However, 3D TV systems that use the depth images tend to create artefacts in form of occlusions, due to non-visibility of the parts of the scene that need to be reconstructed.

Here we investigate the influence of the noise on the number of occlusions in the virtual views created using noisy and denoised depth maps. Occlusions that manifest themselves as the black holes in the images of virtual views often appear in the vicinity of sudden changes in geometry in direction of z-axis at a certain scan line (x-axis). More information on these problems can be found at [35]. This can be partially prevented by smoothing the depth maps asymmetrically, which can distort the geometry of the scene. We are not proposing any new method for the correction of the virtual views here, and the only purpose of this comparison is to show the effect of noise and denoising on virtual views.

Fig. 20 shows the generated virtual views using noise-free and noisy depth maps, depth maps denoised using the method from [11] and the proposed algorithm. It is obvious from this figure that noise can create serious occlusions in the rendered views, which are distributed all over the image. Furthermore, denoising significantly removes the number of occlusions in created views, where, if some noise remains after denoising, occlusions will still be visible in virtual views images, as can be seen in Fig. 20 b). Our algorithm removes noise uniformly over the image, and there are consequently much less occlusions in Fig. 20c). To quantify the occlusions in the generated virtual views, we calculate the percentage of the missing areas (shown as the black holes) in the image. The occlusion percentage here is the largest in the virtual view generated from the noisy depth map (34%). For the image generated using the denoising method of [11] 8% of the pixels are occluded. For our method, 6% of the pixels are occluded, similar to the occlusion percentage resulting from the noise free depth map. We do not claim here that the number of resulting occlusions is the main quality measure of a depth denoising algorithm, but that these results demonstrate clearly the significance of denoising for virtual view generation.

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010

6. Conclusion

In this paper, we presente a new method for denoising depth images, which consists of two main components, which are also the two main contributions of the paper. The first novel component is an adaptive noise estimation method for depth images that makes use of the luminance information and significantly improves denoising performance compared to the method of [11]. Moreover, the denoising performance and processing speed are significantly improved when compared to the same method, with the noise estimated using the common local MAD estimator. The second component is a novel wavelet estimator that takes spatially variable noise variance into account, and uses the edges from luminance images in order to improve denoising of depth images. The experimental results on both real and artificial images show an improvement over the related state-of-the art methods. Some future improvements, in the case of image sequences, are possible using motion estimation (or searching for similar segments in the previous frames) and temporal filtering.

Acknowledgment

A. Pižurica is a postdoctoral researcher of the Fund for the Scientic Research (FWO), Flanders, Belgium. This work was funded by FWO through the project 3G002105.

#133231 - $15.00 USD Received 11 Aug 2010; revised 27 Sep 2010; accepted 2 Oct 2010; published 11 Oct 2010