TIME-OF-FLIGHT DEPTH IMAGE DENOISING USING PRIOR NOISE INFORMATION

T. Edeler[1]*, K. Ohliger*[1]*, S. Hussmann*[1]*, and A. Mertins*[2] *Senior Member, IEEE,*

[1] Westcoast University of Applied Sciences (FHW), Heide, 25746 Germany
[2] Institute for Signal Processing, University of Lübeck, 23538 Germany
edeler@fh-westkueste.de

ABSTRACT

In this paper, we propose a novel way of using time-of-flight camera depth and amplitude images to reduce the noise in depth images with prior knowledge of spatial noise distribution, which is correlated with the incident light falling on each pixel. The denoising is done in wavelet space and the influence and implications of the extended noise model to wavelet space and common denoising methods are shown.

Index Terms— Time-Of-Flight, Denoising, Wavelet, Non-Local-Means

I. INTRODUCTION

Time-of-flight cameras provide, beside an ordinary 2D intensity image, a depth map containing gray levels proportional to the distance of objects. The depth map itself is superimposed by a considerable amount of noise, which intensity is correlated with the amount of light collected by a single pixel.

In [1] we proposed a method for depth-image super resolution with implicit noise reduction in spatial domain which exploited this correlation. In contrast to that paper we herein use this correlation in a wavelet context. For that we derive a novel noise adaptive wavelet thresholding method. It is worth noting that depth-map noise is not signal dependent in the sense that the noise level depends on the depth-signal itself. It rather depends on the intensity image. This is why the proposed method is not comparable to signal depended noise reduction methods like in [2].

In the next section of this paper we introduce the principle of time-of-flight cameras and show the signal model for the depth map. In Section 3 we first recall the wavelet denosing through thresholding and show the implications of the signal model on the thresholding coefficients in wavelet space. In Section 4 we show experimental results for our proposed scheme on simulated and real data, and Section 5 concludes this paper.

II. TIME-OF-FLIGHT PRINCIPLE AND NOISE MODEL

Each time-of-flight camera is equipped with its own source of light. An object in a distance d from the camera (and its light source) reflects photons stemming from the modulated light source. They are collected by the time-of-flight pixel as

$$s(t) = a_0 \cos(\omega_0 t - \phi) + B, \tag{1}$$

where $s(t)$ is the average number of photons per unit time at given time t, ϕ is the phase shift resulting from photons traveling to the object and back to the camera $\left(\phi = 2\omega_0 \frac{d}{c}\right)$, with c being the speed of light. Thus the phase shift has a linear dependency from the distance of the reflecting object. The average incident light and the modulation amplitude are taken into account by B and a_0 respectively.

Since the phase shift can not be measured directly, many time-of-flight systems use a pixel structure that performs some correlation

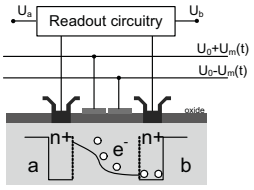

Fig. 1. Cross section of a single time-of-flight PMD pixel containing two wells.

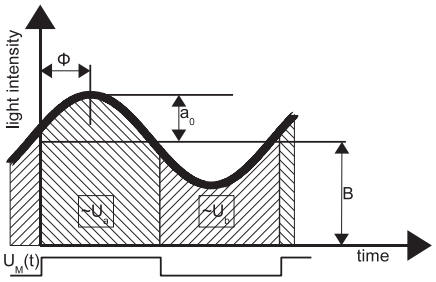

Fig. 2. Light intensity "seen" by a pixel over time. Correlation Voltage $U_m(t)$ is shifted by 0° to the modulation of the light source.

of the optical received signal with an electrical reference source. The pixel structure used for our experiments is shown in Fig. 1. The modulation signal $U_m(t)$ (see Fig. 2) directs electrons (caused by incoming photons) to either of two wells (a and b). To measure the phase shift of incoming light, four images (each taken with $U_m(t)$ shifted by 90° to its predecessor) have to be acquired. Fig. 2 shows the light intensity integrated by a pixel using $U_m(t)$ shifted by 0° to the emitted modulated light. Each pixel provides two voltages U_a and U_b. The differences ($\Delta U = U_a - U_b$), sampled at the four phase shifts, are used to calculate the modulation amplitude a_0 and phase shift ϕ of the optical echo [3]:

$$a_0 = \frac{1}{2}\sqrt{(\Delta U_{270} - \Delta U_{90})^2 + (\Delta U_0 - \Delta U_{180})^2} \tag{2}$$

$$\phi = \arctan\left(\frac{\Delta U_{270} - \Delta U_{90}}{\Delta U_0 - \Delta U_{180}}\right). \tag{3}$$

The number of photons collected during integration is underlying a Poisson distribution (even with perfectly constant intensity) [4]. In practice, when collecting several hundreds of photons, the distribution can be approximated by the normal distribution with the same value for mean and variance. This photon shot noise is responsible for the fact that a pixel collecting more light also outputs more noise (even though the SNR gets better). Since the

978-1-4244-5899-8/10/$26.00 ©2010 IEEE

phase shift of the optical echo does not depend on the total amount of light (but phase noise does), the depth-map SNR lowers when the non-modulated light ($B - a_0$) gets brighter or the modulated light (a_0) gets darker. The dependency of depth-map noise is derived in [5]:

$$\sigma^2_{depth} \propto \frac{B}{a_0^2}, \tag{4}$$

where σ^2_{depth} is the variance of the depth signal.

This leads to an image model for time-of-flight depth images with a spatial variation of noise variance:

$$\mathbf{y} = \mathbf{f} + \mathbf{v}_{\mathbf{\Theta}}, \tag{5}$$

where $\mathbf{f}$ is the ideal depth image[1], $\mathbf{y}$ is the measured depth image and $\mathbf{v}_{\mathbf{\Theta}}$ represents zero mean Gaussian noise with a covariance matrix which is known from (4) up to a scaling factor:

$$\Sigma_v = \mathrm{diag}[\sigma_1^2, \sigma_2^2, ..., \sigma_{\mathrm{N}}^2] = \mathrm{diag}[\mathbf{\Theta}] = \xi \, \mathrm{diag}\left[\mathbf{B} \bullet \mathbf{a}_0^{-2}\right], \tag{6}$$

where $\mathbf{\Theta}$ is called noisemap, which holds the spatial variances. $\mathbf{B}$ and $\mathbf{a}_0$ are images holding the respective value of B and a_0 for each pixel (see (1)). A potentiation of the form $\mathbf{a}^b$ and a multiplication of the form $\mathbf{a} \bullet \mathbf{b}$ are understood as element-wise operations in this paper.

III. WAVELET BASED DENOISING

III-A. Thresholding

Wavelet thresholding (also known as shrinkage) has been introduced by Donoho and Johnson in 1994 [6]. In literature many signal recovery methods have been published since then based on their initial idea in [7], [8], [9] and references therein.

In this section we concentrate on the basic two concepts proposed by Donoho and Johnson called universal (i) hard and (ii) soft thresholding. We show the effect of the model (5) to the noise distribution in wavelet space and its successful combination with the well known methods. The basic motivation for wavelet thresholding is the fact that natural images (signals) tend to be sparse in wavelet space (e.g. concentrate their energy locally), while additive noise is distributed uniformly. If one could classify the wavelet coefficients in *relevant* and *nonrelevant* for noise-free signal description the resulting signal reconstruction using only the (modified) *relevant* coefficients leads to a less noisy signal. Wavelet thresholding consists therefore of 3 steps:

1) Transformation of noisy input signal to wavelet space.
2) Modification of wavelet coefficients.
3) Transformation to signal space.

Consider a noisy sampling process, according to the model

$$\mathbf{y} = \mathbf{f} + \mathbf{v}, \tag{7}$$

with $\mathbf{f} = [f(t_i)]_{i=1}^n$ and ($t_i = i \cdot T \cdot n^{-1}$) being the vector representation of a band limited function $f(t)$ we wish to reconstruct. $\mathbf{y}$ is the measured data, and $\mathbf{v}$ is a white Gaussian noise vector with zero mean and covariance matrix $\sigma^2\mathbf{I}$. Denoising can be described as finding an estimate $\hat{\mathbf{f}}$ which minimizes $||\mathbf{y} - \hat{\mathbf{f}}||_2$ subject to some prior knowledge about the expected signal. One way of incorporating a reasonable prior is to transform the measured data into the wavelet space where natural signals tend to concentrate their energy locally. If we use an orthonormal wavelet basis $\mathcal{T}_w = [\mathbf{g}_i]_{i=1}^n$, with $\mathbf{g}_i$ being the basis vectors, the coefficients $\mathbf{w}$ are calculated by

$$\mathbf{w} = \mathcal{T}_w \mathbf{y} = \mathcal{T}_w(\mathbf{f} + \mathbf{v}) = \mathcal{T}_w \mathbf{f} + \mathcal{T}_w \mathbf{v} = \mathbf{w}_f + \mathbf{w}_v. \tag{8}$$

Since $||\mathbf{g}_i||_2 = 1$, it is easy to see that $\mathbf{w}_v$ is again a white Gaussian noise vector with zero mean and covariance matrix $\sigma^2\mathbf{I}$. And as the main "clean" signal energy is concentrated to some coefficients, thresholding can be applied. Donoho and Johnson proposed a hard thresholding method

$$\hat{w}_{i,h} = \eta_h(w_i) = \begin{cases} w_i & |w_i| > \sigma\lambda_i \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

and a soft thresholding method

$$\hat{w}_{i,s} = \eta_s(w_i) = \begin{cases} \mathrm{sign}(w_i) \cdot (|w_i| - \sigma\lambda_i) & |w_i| > \sigma\lambda_i \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

where w_i and $\hat{w}_i$ are the elements of $\mathbf{w}$ and $\hat{\mathbf{w}}$ respectively. The latter vector is used to calculate the estimate $\hat{\mathbf{f}} = \mathcal{T}_w^{-1}\hat{\mathbf{w}}$. σ is the noise level and λ is a positive number. In literature there have been different approaches to estimate an optimal λ, either universal or dyadic level adaptive (e.g. [10], [11], [12], and [13]). We chose the optimal value (in the *minimax* sense) by the dyadic level adaptive SUREShrink procedure, proposed in [12].

III-B. Thresholding depth data

To the best of our knowledge the model (5) has never been used in the context of wavelet denoising algorithms. The difficulty is, however, to determine the power distribution in wavelet space from a known distribution in image space for successful application of wavelet thresholding. For the model (7) with white Gaussian noise, the variance does not change by transforming to an orthogonal wavelet space but this does not hold true in the case of model (5).

Given a signal $\mathbf{u} = \mathbf{H}\mathbf{v}_{\mathbf{\Theta}}$, where $\mathbf{v}_{\mathbf{\Theta}}$ is a nonstationary and uncorrelated Gaussian noise process with zero mean and covariance matrix $E[\mathbf{v}_{\mathbf{\Theta}} \cdot \mathbf{v}_{\mathbf{\Theta}}^T] = \mathrm{diag}[\mathbf{\Theta}]$. $\mathbf{H}$ is a linear FIR filter operator with impulse response $\mathbf{h}$. The covariance matrix of the signal $\mathbf{u}$ is then $E[\mathbf{u}\mathbf{u}^T] = E[\mathbf{H}\mathbf{v}_{\mathbf{\Theta}} \cdot \mathbf{v}_{\mathbf{\Theta}}^T\mathbf{H}^T] = \mathbf{H}E[\mathbf{v}_{\mathbf{\Theta}} \cdot \mathbf{v}_{\mathbf{\Theta}}^T]\mathbf{H}^T = \mathbf{H} \cdot \mathrm{diag}[\mathbf{\Theta}] \cdot \mathbf{H}^{\mathrm{T}} = \Sigma_{\mathrm{u}}$, where the average powers of the elements of $\mathbf{u}$ are given by $\mathrm{diag}[\Sigma_{\mathrm{u}}] = \mathbf{G}\mathbf{\Theta}$ with $\mathbf{G}$ being a linear FIR operator with impulse response $\mathbf{g} = \mathbf{h}^2$.

According to the À-Trous algorithm, wavelet coefficients $\mathbf{w}^{d,J}$ of an image $\mathbf{x}$ on the dyadic level J can be calculated by $\mathbf{w}^{d,J} = \mathbf{D}_J\mathbf{H}_{d,J}\mathbf{x}$, where $\mathbf{D}_J$ is an operator performing a subsampling of factor 2^J in both dimensions, and $\mathbf{H}_{d,J}$ is an FIR filter operator with impulse response $\mathbf{h}_{d,J}$. d can be HH, LH, or HL indicating the diagonal, horizontal, and vertical details, respectively. The impulse response $\mathbf{h}_{d,J}$ can be calculated by recursive convolution of upsampled analysing filters or in z-space

$$\mathcal{Z}\{\mathbf{h}_{d,J}\} = \begin{cases} H_1^d(\mathbf{z}), & \text{for } J = 1, \\ H_1^d(\mathbf{z}) \prod_{j=0}^{J-2} H_0\left(\mathbf{z}^{2^j}\right), & \text{for } J > 1, \end{cases} \tag{11}$$

where $H_1^d(\mathbf{z})$ and $H_0(\mathbf{z})$ are the high- and lowpass filters for the dyadic wavelet analysis.

Consequently the noise power propagation through the wavelet transform of a nonstationary process like model (5) is

$$\mathbf{w}_{\mathbf{\Theta}}^{d,J} = \mathbf{D}_J \cdot \mathrm{diag}[\mathbf{H}_{\mathrm{d,J}} \cdot \mathrm{diag}[\mathbf{\Theta}] \cdot \mathbf{H}_{\mathrm{d,J}}^{\mathrm{T}}] = \mathbf{D}_{\mathrm{J}}\mathbf{G}_{\mathrm{d,J}}\mathbf{\Theta}, \tag{12}$$

where $\mathbf{G}_{d,J}$ is a linear operator with impulse response $\mathbf{g}_{d,J} = \mathbf{h}_{d,J}^2$. This means that highpass and lowpass filters in MRA decomposition act as filters with lowpass characteristic $\langle \mathbf{1}, \mathbf{g}_{d,J} \rangle > 0$ on the noise propagation, where $\mathbf{1}$ is a vector of same size as $\mathbf{g}_{d,J}$ filled with ones and $\langle \cdot, \cdot \rangle$ denotes the inner product. This fact is shown in Fig. 3. The result of a noise propagation analysis is shown in Fig. 3(c). For the analysis we used a D4 wavelet basis and 100 images (Fig. 3(b)) with mean value of 128 and a noisemap as shown in Fig. 3(a). From now on we denote the transformed 2D noisemap as $\mathbf{w}_\Theta = [w_{\Theta,1}, w_{\Theta,2}, \cdots, w_{\Theta,N}]$ without denoting dyadic or detail levels.

[1] All images are represented as lexicographic ordered vectors in this paper

Fig. 3. Noise propagation through wavelet transform. (a) Noisemap (variance of 250 in white areas and 0 in black areas). (b) Single input image for wavelet transformation with Gaussian noise variance according to noisemap (a) and mean of 128. (c) Empirical noisemap in D4 wavelet space measured over 100 transformed input images. (d) Dydic wavelet levels for noise propagation.

The results of the noise propagation can then be used to replace the thresholding value $\sigma\lambda$ in (9) and (10) by $\sigma_i\lambda$, where σ_i directly corresponds to the noise variance of the i-th wavelet coefficient.

IV. EXPERIMENTAL RESULTS

IV-A. Setup

In this section we show results of our denoising methods for depth images on simulated and real data. The simulated data are generated with the model (5) and images from the Middlebury stereo database[2] where we used from each dataset the gray scaled image `view1.png` as amplitude image ($\mathbf{a}_0$) and the image `disp1.png` as depth map. All images are scaled to 128×128 pixel and we assumed no non-modulated light ($\mathbf{a}_0 = \mathbf{B}$). To avoid undefined values for the noise map the $\mathbf{a}_0$-images were modified to not drop below 5% of full scale value. As noise scaling factor ξ (see (6)) we used four different values (0.01, 0.03, 0.05, and 0.10).

The real data in our test setup is a scene from our laboratory taken with a *CamCube 2.0* from PMDtec[3]. A single shot (which includes the $\mathbf{a}_0$ and depth image) has been taken without any non-modulated light. The noisemap has been calculated with the $\mathbf{a}_0$-image and denoising was done on the depth image.

IV-B. Methods

To show the superiority of our noise adaptive thresholding (AT) over the conventional thresholding (CT), we used the universal soft and hard thresholding (ST and HT) with a constant λ over the whole wavelet space. This parameter is chosen for AT and CT by the SUREShrink procedure [12] on variance-normalized data. For CT we choose the parameter σ to yield the best PSNR results. The optimization was done with Matlab's *fminsearch* and the start parameter $\sigma_0 = \sqrt{N^{-1}\sum^N w_{\Theta,i}}$. For our AT the parameter σ was taken directly from the noisemap for each wavelet coefficient individually, without any optimization towards PSNR. As we don't have ground truth for the real data, no PSNR can be calculated, therefore we omitted the optimization step on that data for CT. As a wavelet basis we chose the D4-basis.

IV-C. Results

Table II shows the result for simulated data. The higher PSNR of AT or CT is highlighted for hard and soft thresholding, respectively. Our AT method shows mostly better results than the best CT in the case of soft thresholding. For hard thresholding the situation is different since the CT sometimes provide much better PSNR results (Laundry $\xi = 0.05$). But as Fig. 4(a,b) shows, hard thresholding sometimes tends to produce overly smooth images, when optimizing to PSNR. And even if the PSNR of our AT is worse, the perceptual quality and usability seems to be better. Fig. 4(c,d) shows an example where the noise reduction of our method provides very good results. Around the horns of the reindeer the very good noise suppression of our method can be seen. It is worth noting that in all cases of CT, the parameters are optimized towards the best PSNR value, whereas our method is not optimized at all.

[2]The images can be downloaded from http://vision.middlebury.edu/stereo/data/scenes2005/ThirdSize/zip-2views/ALL-2views.zip

[3]http://www.pmdtec.com/products-services/pmdvisionr-cameras/pmdvisionr-camcube-20/

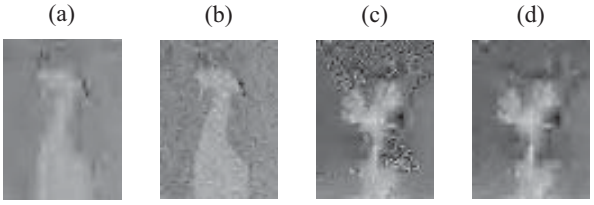

Fig. 4. Reconstruction results on simulated data for extreme cases in Table I. (Laundry $\xi = 0.05$): (a) Conventional hard thresholding (CHT), (b) Adaptive hard thresholding (AHT). (Reindeer $\xi = 0.10$): (c) Conventional soft thresholding (CST), (d) Adaptive soft thresholding (AST)

Fig. 5 shows the results on real camera data. In subfigure (a) we marked four regions and measured the spatial standard deviation (Table I). Again, the results show the superiority of our method against the convectional methods.

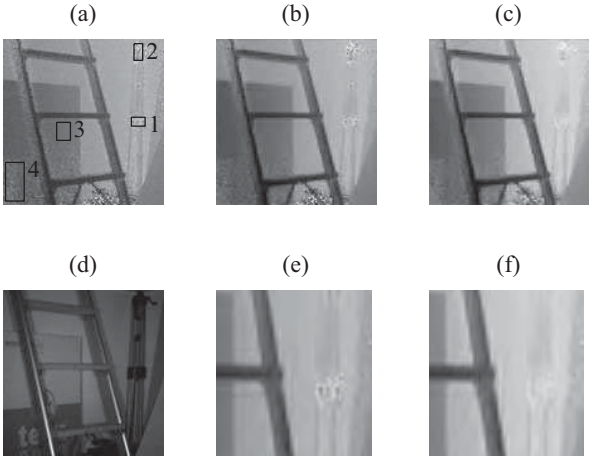

Fig. 5. Reconstruction results on real depth image. (a) Original Depth map. (b) Reconstruction using convectional soft thresholding. (c) Reconstruction using adaptive soft thresholding. (d) Original amplitude image ($\mathbf{a}_0$). (e) and (f) magnification of (b) and (c) respectively.

V. CONCLUSION

In this paper we showed the successful application of a time-of-flight depth map model with spatial variant noise variance on wavelet thresholding. We showed denoising results on real and simulated data with a universal threshold method and showed that our method outperforms existing methods. The Experiments were done by optimizing the existing methods towards a good PSNR and

Region	Noisy Measurement	CST	AST
1	25.0 cm	16.2 cm	7.0 cm
2	32.1 cm	23.7 cm	13.6 cm
3	4.2 cm	2.8 cm	2.7 cm
4	13.7 cm	8.3 cm	7.3 cm

Table I. Standard deviation in 4 regions of Fig. 5 for the original depth map provided by the camera, the reconstruction using convectional soft thresholding (CST), our adaptive soft thresholding (AST).

without optimizing our method. And as our method modifies the thresholding coefficients in wavelet domain it is neither restricted to a certain wavelet basis nor to constant thresholding parameters.

VI. REFERENCES

[1] T. Edeler, K. Ohliger, S. Hussmann, and A. Mertins, "Super resolution of time-of-flight depth images under consideration of spatially varying noise variance," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, nov. 2009, pp. 1185 –1188.

[2] K. Hirakawa, "Signal-dependent noise characterization in Haar filterbank representation," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 2007, vol. 6701, p. 47.

[3] S. Hussmann and T. Liepert, "Three-dimensional tof robot vision system," *IEEE Trans. on Instrumentation and Measurement*, vol. 58, no. 1, pp. 141–146, Jan. 2009.

[4] J. Nakamura, *Image sensors and signal processing for digital still cameras*, CRC, 2006.

[5] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE Journal of Quantum Electronics*, vol. 37, no. 3, pp. 390–397, 2001.

[6] D.L. Donoho and J.M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425, 1994.

[7] M. Welk, G. Steidl, and J. Weickert, "Locally analytic schemes: A link between diffusion filtering and wavelet shrinkage," *Applied and Computational Harmonic Analysis*, vol. 24, no. 2, pp. 195–224, 2008.

[8] M. Jansen and A. Bultheel, "Empirical Bayes Approach to Improve Wavelet Thresholding for Image Noise Reduction.," *Journal of the American Statistical Association*, vol. 96, no. 454, 2001.

[9] S. Poornachandra, "Wavelet-based denoising using subband dependent threshold for ECG signals," *Digital Signal Processing*, vol. 18, no. 1, pp. 49–55, 2008.

[10] B. Vidakovic, *Statistical modeling by wavelets*, Wiley New York, 1999.

[11] A. Bruce, H.Y. Gao, and A. Bruce, *Applied wavelet analysis with S-plus*, Springer New York, 1996.

[12] D.L. Donoho and I.M. Johnstone, "Adapting to Unknown Smoothness Via Wavelet Shrinkage.," *Journal of the american statistical association*, vol. 90, no. 432, 1995.

[13] Y. Chen and C. Han, "Adaptive wavelet threshold for image denoising," *Electronics Letters*, vol. 41, pp. 586, 2005.

Dataset	ξ	CST	AST	CHT	AHT
Art	0.01	31.22	**31.51**	**30.53**	28.83
Art	0.03	28.07	**28.56**	24.37	**25.31**
Art	0.05	26.93	**27.29**	**26.02**	24.12
Art	0.10	25.17	**25.84**	20.59	**22.62**
Books	0.01	35.14	**35.40**	**34.68**	32.91
Books	0.03	31.85	**32.39**	27.88	**28.70**
Books	0.05	30.26	**31.08**	25.64	**26.67**
Books	0.10	28.21	**29.43**	**27.93**	24.67
Dolls	0.01	32.64	**33.87**	28.60	**30.35**
Dolls	0.03	29.84	**31.18**	**28.74**	26.64
Dolls	0.05	27.75	**29.99**	**27.80**	25.17
Dolls	0.10	26.77	**28.55**	**26.04**	24.17
Laundry	0.01	**32.82**	32.73	32.10	**32.25**
Laundry	0.03	30.50	**30.76**	**30.75**	28.32
Laundry	0.05	29.32	**29.68**	**29.26**	26.27
Laundry	0.10	27.86	**28.19**	**27.77**	24.84
Moebius	0.01	34.29	**34.84**	31.02	**31.66**
Moebius	0.03	31.05	**31.88**	26.70	**27.43**
Moebius	0.05	29.69	**30.53**	**28.69**	26.05
Moebius	0.10	27.89	**28.96**	**27.02**	23.80
Reindeer	0.01	30.87	**31.98**	**29.98**	28.35
Reindeer	0.03	27.90	**28.92**	**27.15**	24.20
Reindeer	0.05	25.49	**27.66**	**25.98**	22.67
Reindeer	0.10	22.91	**26.06**	17.13	**21.42**

Table II. Experimental PSNR results (in dB) on simulated data. CST: Conventional soft thresholding. AST: Proposed adaptive soft thresholding. CHT: Conventional hard thresholding. AHT: Proposed adaptive hard thresholding.