WAVELET BASED JOINT DENOISING OF DEPTH AND LUMINANCE IMAGES

Ljubomir Jovanov, Aleksandra Pižurica and Wilfried Philips

Telecommunications and Information Processing Group
Ghent University
Sint Pietersnieuwstraat 41
B 9000 Gent Belgium
ljj@telin.ugent.be

ABSTRACT

In this paper we present a new method for joint denoising of depth and luminance images produced by time-of-flight camera. Here we assume that the sequence does not contain outlier points which can be present in the depth images. Our method first performs estimation of noise and signal covariance matrices and then performs vector denoising. Two versions of the algorithm are presented, depending on the method used for the classification of the image contexts. Denoising results are compared with the ground truth images obtained by averaging of the multiple frames of the still scene.

Index Terms— Denoising, video, TOF camera, noise estimation

1. INTRODUCTION

Object recognition, autonomous navigation of robots, industrial inspection, biometric authentication are complex tasks which require reliable and clean features in order to be performed successfully. Algorithms which aim to solve these problems often rely on luminance, color and motion information in order to get an interpretation of the scene. The above-mentioned features are often not sufficient for a valid interpretation of the scene due to the occlusions and the lack of information needed for a unique interpretation.

Scene interpretation can be significantly improved by introducing range data into the feature set. Depth information makes the task of the scene interpretation more feasible and robust. Various range measuring techniques exist which are based on usage of multiple cameras. These include triangulation systems such as stereo vision (or structured light), depth-from-focus, depth-from-shape and depth-from-motion. Most recent depth sensors are based on the measuring of time of flight of the light beam. This type of depth sensors offers better accuracy, higher frame rate and lower computational requirements in order to reconstruct depth image. In Section 2, we describe noise characteristics of the sensor, and the way of getting ground truth images from noisy observations. In the Section 3, we describe the proposed noise estimation technique and denoising method. The experimental results are presented in Section 4, and the conclusions are in Section 5.

2. NOISE BEHAVIOUR OF THE DEPTH SENSOR

Depth resolution of the time-of-flight depth sensors is limited by a number of factors. The main limitation factor is shot noise noise present in depth sensor. Amount of shot noise is determined by an uncertainty in the number of the generated electrons. Other sources of noise are AD converter quantization noise, kT/C reset noise and thermal noise.

Due to the large number of factors, which affect the measured distance, each range pixel can be modelled as a Gaussian random variable with a mean value μ_i and a standard deviation σ_i where the mean value corresponds to the actual range value of pixel i. If we assume that the range value is constant over a local neighbourhood of pixel i, and that all pixels can be modelled as Gaussians with mean μ_i and standard deviation σ_i, the range value can be obtained via averaging within a neighbourhood (spatial or temporal) of pixels around pixel i.

If we define mean value over N time instants as $X = \frac{1}{N}\sum_{k=1}^{N} X_k$, then the mean value and standard deviation of the temporal average can be written as:

$$E(X) = \frac{1}{N}\sum_{k} \mu_k = \mu_i \tag{1}$$

$$Std(X) = \frac{1}{N} Std(\sum_{k=1}^{N} X_k) = \frac{1}{N}\sqrt{N}\sigma_i = \frac{\sigma_i}{\sqrt{N}} \tag{2}$$

Thw above expressions show that the signal to noise ratio, and therefore, depth measurements resolution can be increased by a factor of $\sqrt{N}$ if the range values are averaged in a spatial or temporal neighbourhood of N pixels. This is only valid if the range values are constant in the observed neighbourhood. Unfortunately, this does not hold in the most practical cases. A side effect of the averaging either in temporal or spatial domain is that details such as edges or textures

are significantly degraded. Temporal averaging creates motion blur. In order to avoid these effects it is necessary to use more sophisticated methods for noise removal.

However, in order to evaluate the performance of our denoising algorithm we use temporal averaging over 20 frames, which do not contain significant motion, in order to obtain ground truth images, since it is not possible to get exact noise-free depth image. By using temporal averaging we avoid blurring of the edges in the spatial domain.

3. THE PROPOSED ALGORITHM

Although depth images can be observed as ordinary images and denoised using some of the numerous image or video denoising algorithms, such as [1], [2] and [3] better denoising results can be obtained by jointly using of luminance and depth information, because of the interdependencies between them.

For example, parts of the objects which are closer to the light source will be brighter, and the luminance will decrease with increasing the distance from the light source. Besides that, in the cases of the missing data points in the depth sequence, it is possible to make more reliable interpolation using both luminance and depth from the surrounding locations.

Features appearing in one image such as edges, lines, textures etc. will probably appear in the other image, enabling a more reliable detection of signal in noise, and detection of false structures generated by noise. By including pixel neighbourhood, denoising performance can be additionally improved.

Another important observation is that the luminance sequence contains less noise than the depth measurements. It was estimated that the PSNR of the luminance image was 35,9dB, and the psnr of the depth sequence 14,9dB. This means that luminance can be used for more reliable segmentation of the depth sequence, in the case of the higher illumination, when the depth measurements become more noisy.

In this paper we use Daubechies db4 and db8 wavelet decomposition of both depth and luminance image. We have used two levels of decomposition in all experiments, to keep the computation time at acceptable level, since k-means clustering is performed for each level.

Filtering of vector-valued images has been explored by several researchers within the frameworks of multispectral image restoration [4], multichannel image restoration [5] and multiframe image restoration [6]. Vector image filtering methods perform filtering on all channels simultaneously.

Proposed method performs filtering using vectors which include 8 neighbouring and central pixels from both luminance and depth image, thereby constructing 18 dimensional noisy vector for each wavelet band and scale:

$$\boldsymbol{y^{o,s}} = [d_1^{o,s}, ..., d_9^{o,s}, l_1^{o,s}, ..., l_9^{o,s}]^T, \quad (3)$$

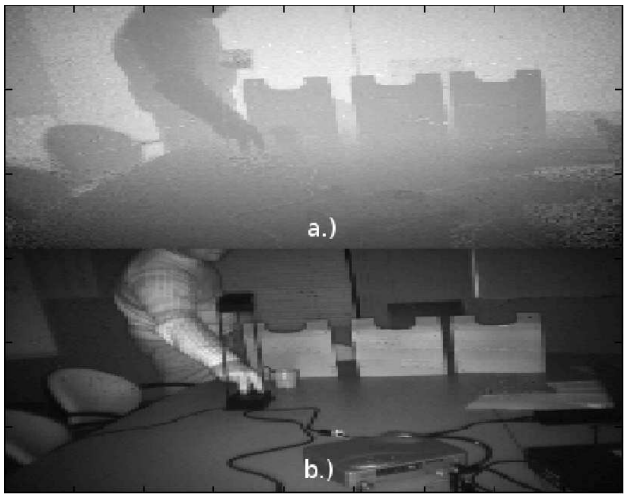

Fig. 1. a.) Noisy range image b.) Noisy luminance image

where $d_i, i \neq 5$, are the values of neighbouring wavelet coefficients , d_5 value of the current wavelet coefficient of the depth image, $l_i, i \neq 5$, are the values of neighbouring wavelet coefficients and l_5 value of the current wavelet coefficient of luminance image, for the scale s and orientation o. Same processing steps will be performed for all scales and orientations, so the superscripts denoting scale and orientation will be omitted. In this paper we assume additive noise model:

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}, \quad (4)$$

where $\boldsymbol{n}$ is Gaussian vector with zero mean and covariance matrix $\boldsymbol{C_n}$, $\boldsymbol{y}$ is a vector of wavelet coefficients contaminated by noise and $\boldsymbol{x}$ is a vector with noise-free wavelet cofficients.

Main idea present in our work is to perform segmentation of image into contexts, where the main criterion for grouping is the similarity of the 3x3 blocks containing luminance and depth values. We assume that inside each of these groups signal vectors $\boldsymbol{x}$ obbey multivariate Gaussian distribution, with covariance matrix $\boldsymbol{C_x}$, because of the properties of k-means algorithm. Similar contexts obtained by segmentation are shown in Fig. 2. Each color corresponds to the different cluster.

For each spatial location vectors y are formed. Besides vectors which contain both luminance and depth, vectors containing only luminance contexts are formed. One way to obtain segmentation is to perform k-means clustering of these vectors. Optimal number of clusters was determined experimentally as 20. The choice of this value can be justified by the fact that we have used only one scene configuration. This number depends on image content. To overcome the problem of optimal determination of the number of clusters, unsupervised clustering method should be used.

Another way of getting segmentation is to use averaged sums of absolute values of surrounding pixels. This approach

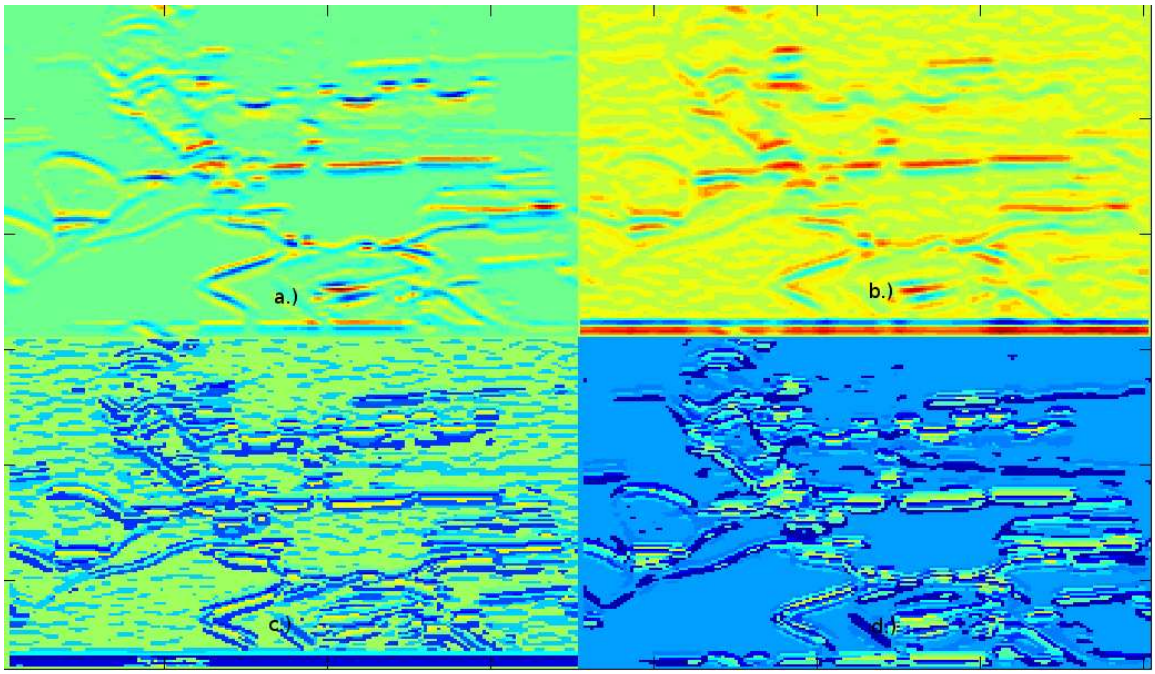

Fig. 2. a.) LH2 wavelet band of the depth image b.) Quantized values of the spatial activity indicator c.) Segmentation of depth and luminance contexts using k-means d.) Segmentation of luminance contexts using k-means

is much faster than k-means clustering. Spatial indicator value for the spatial location l is calculated as:

$$\boldsymbol{s_l} = \sum_{i \in N_l} |y_i|, \tag{5}$$

where N_l denotes set of the 9 neighbouring pixels around spatial location l. Spatial indicator values are then normalized to have values ranging from 1 to the number of clusters (20). Each of these values is rounded to the first greater integer value. Values obtained in this way are considered as labels of image contexts. Once we obtain segment labels it would be possible to calculate center values for each of the contexts. In the case of k-means clustering, centroid values are returned together with labelling. As can be seen in Fig. 2, edges of the similar orientation and similar image features are grouped in the segments.

Next step in our algorithm is estimation of parameters for denoising i.e. noise and signal covariance matrices. Noise covariance matrix is estimated based on contexts which are placed closest to the centroid of the biggest segment, since we assume that those points correspond to the homogenous regions, which do not contain important image details.

Noise covariance matrix is calculated as follows:

$$\boldsymbol{C}_n = \sum_{k=1}^{N_1} (y_{1k} - E(\boldsymbol{y}_1)) \cdot (y_{1k} - E(\boldsymbol{y}_1))^T, \tag{6}$$

where $\boldsymbol{y}_1$ denotes vectors which belong to the biggest cluster, and N_1 denotes number of vectors from the biggest cluster used for covariance matrix calculation. Noise covariance matrix estimated in 6 is used for denoising of all clusters. Signal covariance matrix is calculated for each cluster separately. It was observed that significant image details are captured in the data points which are on the greatest distance from the center of the cluster. Based on that we have used 10% of the data points which are on the biggest Euclidean distance from the centroids to estimate signal covariance matrix for each image segment. Groups which have less than 2 percent of the total number of points are left intact, since it was observed that they consist of significant details. In this work, these thresholds are fixed, because of the fixed scene. In general case, they should be estimated for each image separately. Signal covariance matrix is estimated similarly as in 6, with the only difference in the sum indexes.

In this paper we use vector Wiener filtering, since it was assumed that noise-free signal inside each of the clusters obeys multivariate Gaussian distribution. Wiener filtering yields minimum mean square error:

$$\hat{\boldsymbol{x}} = \frac{\hat{\boldsymbol{C}}_{\boldsymbol{x}}}{\hat{\boldsymbol{C}}_{\boldsymbol{x}} + \hat{\boldsymbol{C}}_{\boldsymbol{n}}} \cdot \boldsymbol{y}, \tag{7}$$

where $\hat{x}$ denotes estimated value of the noise-free vector, $\hat{C}_x$ and $\hat{C}_n$ are covariance matrices of the signal and noise respectively and y is noisy vector. As a result we take mem-

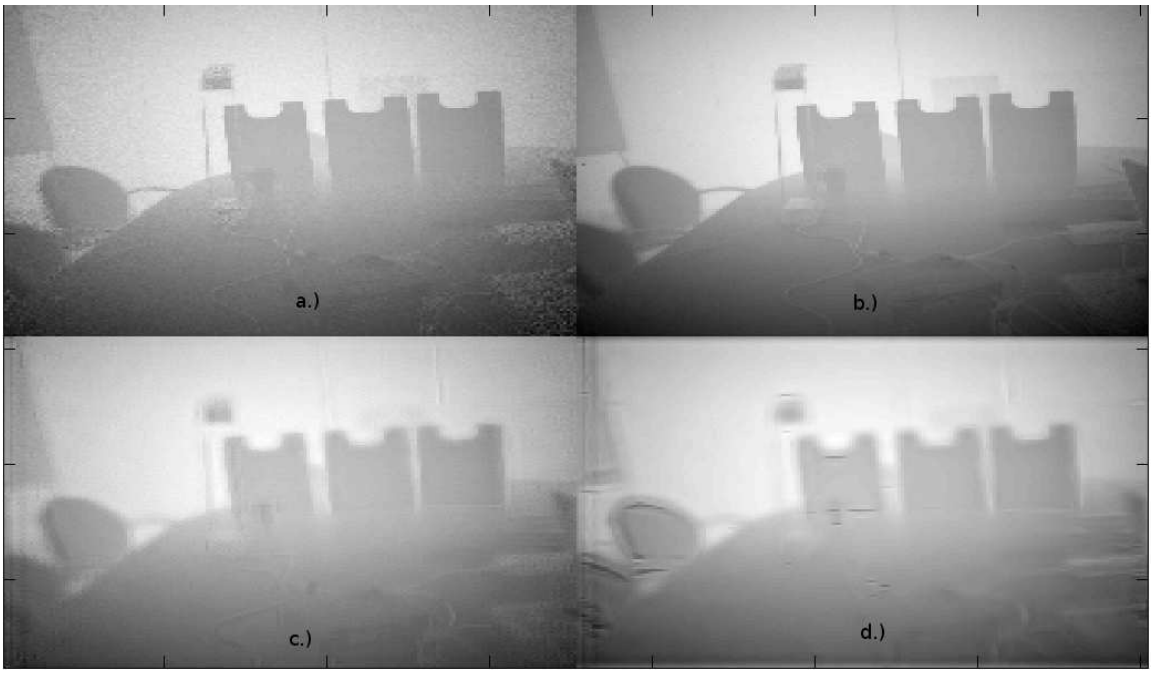

Fig. 3. a.) Noisy image b.) Ground truth image c.) Image denoised using method which relies on k-means d.) Image denoised using method which uses spatial activity indicators

bers of the vector which correspond to middle pixels. $\hat{C}_x$ is a covariance matrix, and it should be positive semi-definite. This condition can be enforced by performing singular value decomposition and setting negative singular values on small positive values, before matrix reconstruction. This effect occurs rarely, with neglectable influence on the denoising performance.

4. EXPERIMENTAL RESULTS AND DISCUSSION

In this section we will compare performance for two different versions of algorithms for joint denoising of luminance and depth images. The proposed algorithms were tested on one dataset, containing luminance and depth of the fixed scene, recorded using time-of-flight camera. Noise removal algorithms, provided with the camera were turned off in order to have realistic noisy sensor data. Since we have used depth images with real noise, and we had no noise-free images available, we had to use average of 20 frames of a still scene, as ground truth image.

Results obtained using k-means segmentation and context modelling are very close to the ground truth images, and outperforms both visually and in PSNR sense method which uses spatial indicators and wavelet image denoising method presented in [2]. PSNR for method which uses k-means is 29.7dB, which is 2.5dB better than the PSNR of the noisy image. Method which uses spatial activity indicators have PSNR which is 0.8 dB less. Method which uses k-means clustering preserves significant details in depth image better, compared with spatial indicator method. Results obtained using proposed denoising methods are shown in Fig. 3.

5. CONCLUSION

In this paper we present method for joint denoising of depth and luminance images, based on vector Wiener filtering. Proposed method preserves depth image details, because it takes luminance information into account. The effect of the smoothing of the denoised images on the quality of reconstructed images has not been investigated. Further improvements will be possible using estimated motion from depth and luminance.

6. REFERENCES

[1] A. Pižurica and W. Philips, "Estimating probability of presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Trans. on Image Processing*, vol. 15, no. 3, pp. 654–665, 2006.

[2] G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising.," *IEEE Trans. on Image Processing*, vol. 9, no. 9, pp. 1522–1531, 2000.

[3] E. J. Balster, Y. F. Zheng, and R. L. Ewing, "Feature-based wavelet shrinkage algorithm for image denoising,"

IEEE Trans. on Image Processing, vol. 14, no. 12, pp. 2024 – 2039, Dec. 2005.

[4] B. Hunt and O. Kubler, "Karhunen-loeve multispectral image restoration, part i: theory.," *IEEE Trans. on Acoust., Sp., and Sig. Proc.*, vol. 32, no. 5, pp. 592–600, 1984.

[5] N. Galatsanos and R. Chin, "Digital restoration of multichannel images.," *IEEE Trans. on Acoust., Sp., and Sig. Proc.*, vol. 37, no. 3, pp. 415–421, 1989.

[6] M. Ozkan, A. Erdem, M. Sezan, and M. Tekalp, "Efficient multiframe wiener restoration of blurred and noisy image sequences.," *IEEE Trans. on Image Processing*, vol. 1, no. 4, pp. 453–476, 1992.