

CONTENT ADAPTIVE WAVELET BASED METHOD FOR JOINT DENOISING OF DEPTH AND LUMINANCE IMAGES

Ljubomir Jovanov, Nemanja Petrović, Aleksandra Pižurica and Wilfried Philips

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

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. Luminance image is segmented into similar contexts using k-means algorithm, which are used for calculation of covariance matrices. Denoising results are compared with the ground truth images obtained by averaging of the multiple frames of the still scene.

Keywords: depth images, wavelets, noise estimation, denoising

1. INTRODUCTION

Object recognition, autonomous navigation of robots, industrial inspection and 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. One of the important characteristics of time-of-flight sensors is the fact that luminance images contain less noise than the corresponding depth images. The main idea and novelty of our paper is to use luminance image, which is of better quality to improve the denoising of the depth image. Using this approach significant details in the depth images can be recovered from noise. 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.

Further author information: (Send correspondence to Ljubomir Jovanov)

Ljubomir Jovanov: E-mail: ljj@telin.ugent.be, Telephone: +32 9 264 34 12, A. Pižurica is a postdoctoral researcher of FWO Flanders, Belgium

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 present in depth sensor. Shot noise originates from 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:

$$E(X) = \frac{1}{N} \sum_k \mu_k = \mu_i \quad (1)$$

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

The 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 most of the 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.

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 [1], [2] and [3] better denoising results can be obtained by jointly using 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 28.56dB. 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.

Filtering of vector-valued images has been explored by several researchers within the frameworks of multispectral image denoising [4], [5], multichannel image denoising [6], [7] and multiframe image restoration [8]. Vector image filtering methods perform filtering on all channels simultaneously.

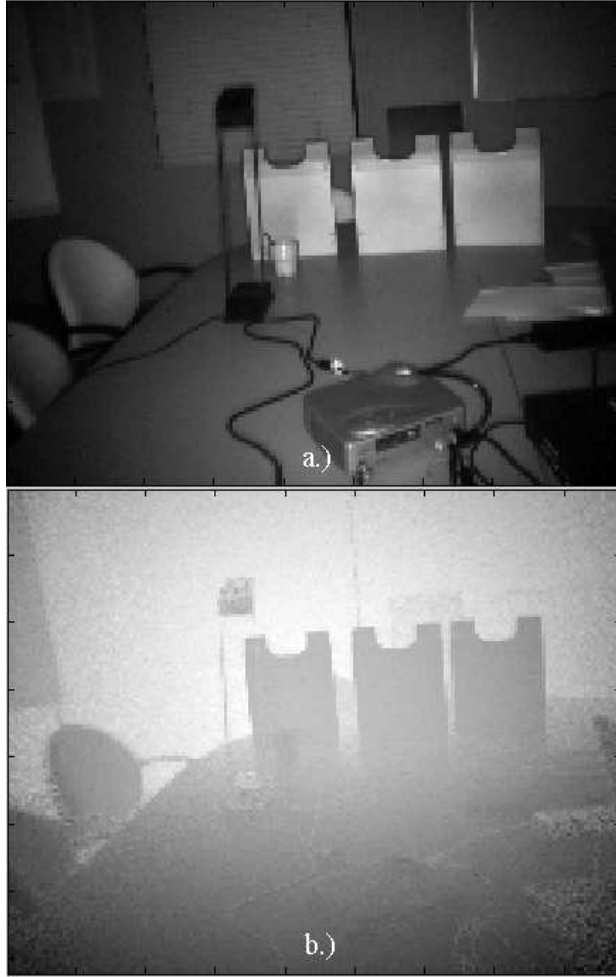


Figure 1. a.) Noisy range image b.) Noisy luminance image

The proposed method filters 18 dimensional vectors which include 8 neighbouring and central pixels from both luminance and depth image, for each wavelet band and scale:

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

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 . The 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:

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

where \mathbf{n} is Gaussian vector with zero mean and covariance matrix \mathbf{C}_n , \mathbf{y} is a vector of wavelet coefficients contaminated by noise and \mathbf{x} is a vector with noise-free wavelet coefficients.

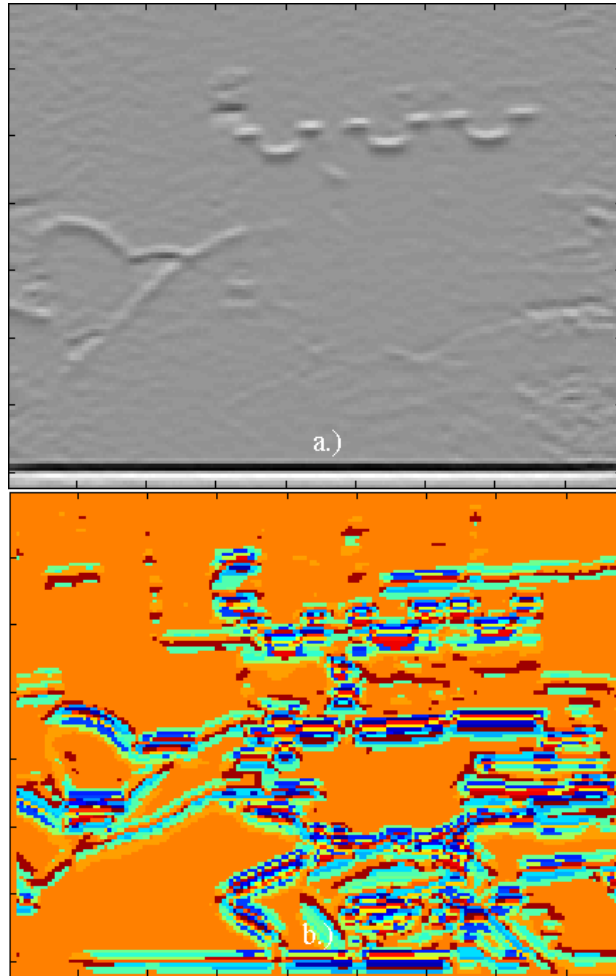


Figure 2. a.) LH2 wavelet band of the depth image b.) Segmentation of luminance contexts using k-means algorithm

The main idea of this work is to perform segmentation of image into contexts, where the main criterion for grouping is the similarity of the 3×3 blocks containing luminance values. We assume that inside each of these groups signal vectors \mathbf{x} obey multivariate Gaussian distribution, with covariance matrix $\mathbf{C}_{\mathbf{x}}$, because of the properties of the k-means algorithm, which generates groups of points with Gaussian distribution. Similar contexts obtained by segmentation are shown in Fig. 2. Each color corresponds to the different cluster.

3.1. Segmentation of luminance image into similar contexts

In our initial approach we have used the k-means algorithm to group similar 3×3 neighbourhoods into clusters. This method had one serious drawback: number of clusters had to be set in advance, which means that it could not adapt to the statistics of the each image. The solution to this problem is to determine the number of clusters automatically prior to clustering by using non-parametric clustering methods. In this paper we use method based on subtractive clustering to determine the number of clusters present in data. The number of clusters is estimated for each wavelet decomposition level and orientation.

The subtractive clustering is an extension of the mountain clustering proposed in [9]. This clustering

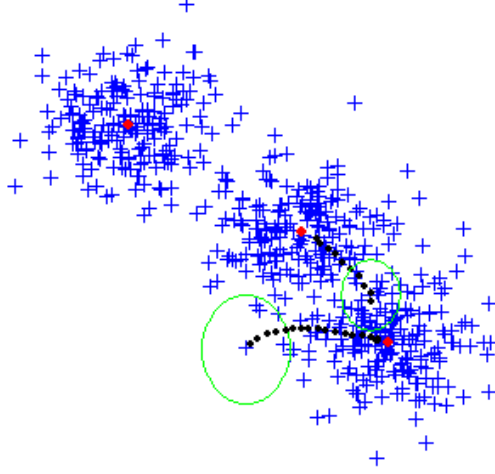


Figure 3. Mode finding based on subtractive clustering

algorithm considers each data point as a potential cluster center and calculates a likelihood that data point defines the cluster center, taking into account the density of surrounding data points. Usually, the actual cluster centres are not located at one of the data points, but in the most cases it is a good approximation, which saves computation time. Taking this into account, a *density measure* at data point \mathbf{x}_i is defined as:

$$D_i = \sum_{j=1}^n e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{(r_a/2)^2}}, \quad (5)$$

where r_a is a positive constant representing a neighbourhood radius. In the case when data point is surrounded with a large number of neighbours, density will have large value.

The first cluster center x_{c1} is chosen as the point which have the largest density value D_{c1} . Next, the density measure of each data point \mathbf{x}_i is revised as follows:

$$D_i = D_i - D_{c1} e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_{c1}\|^2}{(r_a/2)^2}}, \quad (6)$$

where r_b is a positive constant which defines a neighbourhood that has measurable reductions density measure. The data points near the first cluster center x_{c1} will have significantly reduced density measure, after the calculation of the first center. After updating the density function, the next cluster center is selected as the point which has the greatest density value.

This process is iterated until all of the data is within a neighbourhood radius r_a around the cluster center. The process is illustrated in Fig. 3

The optimization criterion in k-means algorithm is minimization of the sum of the squared distances between all points and the cluster centres. The main steps of the algorithm are:

1. Setting of initial values of K cluster centres $z_1(1), z_2(2), \dots, z_K(1)$.
2. Redistribution of the samples x among the clusters based on relation, $x \in C_j(k)$ if $\|x - z_j(k)\| < \|x - z_i(k)\|$ for all $i = 1, 2, \dots, K; i \neq j$, where $C_j(k)$ denotes the set of samples whose cluster centre is $z_j(k)$.
3. Computing new cluster centres $z_j(k+1), j = 1, 2, \dots, K$ such that the sum of the squared distances from all points in $C_j(k)$ to the new cluster centre is minimized. Sample mean of $C_j(k)$ minimizes the sum of the

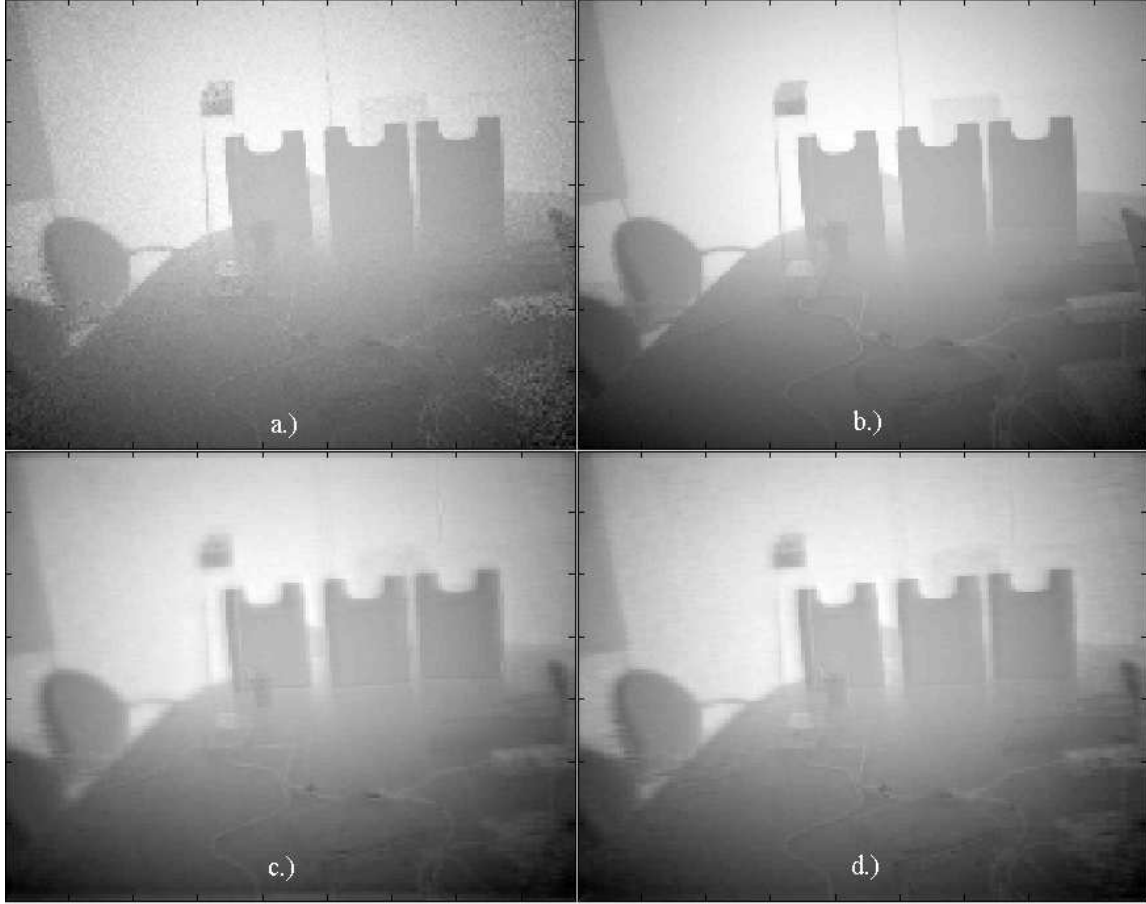


Figure 4. a.) Noisy image b.) Ground truth image c.) Denoised image for parameter value $r_a = 0.015$ d.) Denoised image for parameter value $r_a = 0.01$

squared distances. New cluster centre is given by $z_j(k+1) = \frac{1}{N_j} \sum_{x \in C_j(k)} x, j = 1, 2, \dots, K$, where N_j is the number of samples in $C_j(k)$.

4. If $z_j(k+1) = z_j(k)$ for $j = 1, 2, \dots, K$ then the algorithm has converged and the procedure is terminated.
5. Otherwise go to Step 2.

Obviously, final clustering will depend on the initial cluster centres position and the value of K . Good choice for these parameters is the centres obtained from subtractive clustering.

For each spatial location vectors y are formed. Besides vectors which contain both luminance and depth, vectors containing only luminance contexts are formed. As in our initial approach [10], we use k-means algorithm to group similar 3×3 neighbourhoods of all wavelet bands. Here we use method based on subtractive clustering to obtain number of clusters for each wavelet band. We choose the value of r_b parameter to maximize the output PSNR of the depth image. The main difference compared to our previous approach [10] is that the optimal number of clusters will be chosen depending on the image content. Since it was observed that PSNR of the luminance image is higher than the PSNR of the depth image, luminance image is used for grouping into similar contexts.

The 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 in 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:

$$\mathbf{C}_n = \sum_{k=1}^N (\mathbf{y}_{1k} - \mathbf{E}(\mathbf{y}_1)) \cdot (\mathbf{y}_{1k} - \mathbf{E}(\mathbf{y}_1))^T, \quad (7)$$

where \mathbf{y}_1 denotes vectors which belong to the biggest cluster, and N denotes number of vectors in the biggest cluster. Noise covariance matrix estimated in 7 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. Contexts from other, smaller clusters are used to estimate signal covariance matrix. Signal covariance matrix is estimated similarly as in 7, 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:

$$\hat{\mathbf{x}} = \frac{\hat{\mathbf{C}}_x}{\hat{\mathbf{C}}_x + \hat{\mathbf{C}}_n} \cdot \mathbf{y}, \quad (8)$$

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

In this section we will compare performance of the denoising algorithm for joint denoising of luminance and depth images for different values of parameters. The proposed algorithm was 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.

In this paper we use Daubechies db4 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.

Results obtained using k-means segmentation and context modelling are very close to the ground truth images, and outperform both visually and in PSNR sense method which uses spatial indicators presented in [10] and wavelet image denoising method presented in [2].

We made several experiments with different values of r_b parameter used for determination of the number of clusters. In the first experiment we set the value of neighbourhood radius to 0.01, which resulted in the largest number of clusters, close to empirically determined optimal value (20) for the test scene. Increasing neighbourhood radius leads to smaller number of clusters. PSNR of the noisy depth image was 28.56dB, compared to ground truth image. For the smallest neighbourhood radius value denoised depth image had PSNR value of 35.47dB. Values of r_b of 0.012 and 0.015 yielded PSNR values of 34.04 and 33.38 dB

respectively. PSNR for empirically determined number of clusters (20) was 35.09dB. Results obtained for different parameter values are shown in Fig. 4. As can be seen from the Fig. 4 lower number of clusters yields artefacts around edges of objects in depth image. Since in this approach we determine number of clusters automatically, better statistical modelling is possible, and such artefacts are successfully avoided.

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. In our future work we plan to use more sophisticated methods for searching similar contexts in n-dimensional space, to achieve more precise context modelling.

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* **15**(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* **9**(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* **14**, pp. 2024 – 2039, Dec. 2005.
4. A. Pizurica, P. Scheunders, and W. Philips, "Multiresolution multispectral image denoising based on probability of presence of features of interest," in *Advanced Concepts for Intelligent Vision Systems (Acivs 2004)*, (Brussels, Belgium), 2004.
5. P. Scheunders, "Wavelet thresholding of multivalued images," *IEEE Trans. Image Proc.* **13**, pp. 475–483, Apr. 2004.
6. P. Scheunders and S. De Backer, "Wavelet denoising of multicomponent images, using a noise-free image," in *Proc. IEEE Internat. Conf. Image Proc. ICIP*, (Atlanta, USA, 2006.), 2006.
7. A. Benazza-Benyahia and J. Pesquet, "Building robust wavelet estimators for multicomponent images using stein's principle," *IEEE Trans. on Image Processing* **14**, pp. 1814–1830, 2005.
8. M. Ozkan, A. Erdem, M. Sezan, and M. Tekalp, "Efficient multiframe wiener restoration of blurred and noisy image sequences," *IEEE Trans. on Image Processing* **1**(4), pp. 453–476, 1992.
9. S. Chiu, "Fuzzy model identification based on cluster estimation," *Journal of Intelligent and Fuzzy Systems* **2**(3), 1994.
10. Lj.Jovanov, A. Pižurica, and W. Philips, "Wavelet based joint denoising of depth and luminance images," in *3D TV conference*, (Kos Island, Greece), 2007.