

A MEDICAL IMAGE DENOISING METHOD USING SUBBAND ADAPTIVE THRESHOLDING BASED ON A SHEARLET TRANSFORM

Miroslav Petrov

ABSTRACT. The image denoising process is of great importance when analyzing images and their visualization. A major problem is finding the boundary between clearing the noise and keeping the salient features in the images. This paper proposes adaptive subband threshold image denoising in a shearlet domain based on the Shannon entropy. The method does not suppose a specific type of noise, it does not require data for its spectrum, nor does it lead to highly complex computational algorithms.

1. Introduction. A common problem in the field of computer vision and image processing is noise detection and its reduction in digital images. Medical images are a special class of images containing structural and functional information on anatomical organs. They are intended for examining their structure, as well as for diagnosing various diseases. The main source for obtaining medical

ACM Computing Classification System (1998): I.5.4, I.4.3, I.4.5.

Key words: medical image, denoising, shearlet tresholding, Shannon entropy, Rician noise.

images are the non-invasive diagnostic imaging techniques: *Computed Tomography (CT)*, *Ultrasonography*, *Magnetic Resonance Imaging (MRI)*, *Positron Emission Tomography (PET)*, *Single Photon Emission Computed Tomography (SPECT)*, *Optical Coherence Tomography (OCT)*, etc. Each of these medical imaging devices is affected by different types of noise. Obtaining good-quality images is essential for reliable clinical interpretation.

The main types of noise that exist in medical visualization are:

- *Gaussian (CT [16])*;
- *Poisson (X-ray, PET and SPECT [10, 9])*;
- *Rician (MRI and Ultrasonography [19])*;
- *Speckle (OCT [4])*.

When modelling noise in images with high signal-noise ratio, Gaussian distribution is typically used due to the central limit theorem of Lyapunov. But medical images are mainly low-contrast and the modeling of statistical data distribution is one of the most important stages in noise reduction methods. For example, the variance of Gaussian noise is constant and the variance of Poisson noise is proportional to the average noise, whereas for Rician noise this dependence is nonlinear [9, 21]. Besides, the noise type in medical images may be complex, non-stationary, etc., which further complicates the situation.

A number of methods have been developed in order to solve the basic problem of image reconstruction from noisy data, in particular, those based on multiscale transformations [13, 23, 15].

The shearlet is a new multidimensional and multiscale transform, which is optimally efficient in representing images containing edges. Shearlet systems are superior to the rest of the functions generating multiscale transformations, in terms of orientation, spatial localization and image approximation.

The present work proposes a method for denoising medical images using a Shearlet Transform based on the Shannon entropy. To assess the quality of the denoising images, the following functions are used: *PSNR (Peak Signal-to-Noise Ratio)* and *SSIM (Structural Similarity Index)*.

2. Related theories.

2.1. Minimax and wavelet thresholding estimator. The additive noise model is most commonly used in the task of recovering signal f_0 with limited power ($f_0 \in L_2(R)$) from noisy data f and has the representation $f = f_0 + n$,

where n is the noise. It must be observed that the multiplicative noise components can be considered as additive in relation to the logarithmic scale.

The noise in the signal is modeled by a random vector with a certain probability distribution, which should be known in advance. In many known models it is supposed that n is white Gaussian noise with a standard deviation σ .

The signal f_0 is evaluated by transforming noisy data f through the so-called decision operator D , and the corresponding evaluation is $\tilde{f}_0 = Df$. The operator D should minimize the error of assessment $f_0 - Df$ measured by a loss function.

Usually a mean square distance is used and the loss function is selected as the square of L_2 -norm. Then the estimator risk \tilde{f}_0 is given by the *Mean Square Error (MSE)* calculated in terms of the noise probability distribution: $r(D, f_0) = E \{ \|f_0 - Df\|^2 \}$, where E is a mathematical expectation.

In practice, we may have some prior information concerning the signal, for example, it may belong to a set $F_0 \subset L_2(R)$, but, as for complex signals, we should not expect to know their probability distribution. This means that the expected risk over F_0 cannot be calculated.

With the minimax estimator for controlling the risk for each signal $f_0 \in F_0$, it is necessary to minimize the maximum risk $r(D, F_0) = \sup_{f_0 \in F_0} E \{ \|f_0 - Df\|^2 \}$ and the minimax risk is $r(F_0) = \inf_D r(D, F_0)$. Applications usually look for a decision operator D , which is calculated easily and $r(D, F_0)$ is close to $r(F_0)$ [17].

For uniformly regular and partially regular signals this evaluation is easily achieved due to the sparsity properties of wavelets, using the so-called wavelet threshold estimator. Moreover, the invariance of the white Gaussian noise on orthogonal transformations makes it preferable in most models for extracting information from noisy data.

The computation of the wavelet thresholding estimator can be described in the following manner [5]. Let $\{\psi_{j,m}\}$ be an orthogonal wavelet basis in $L_2(R)$ and $f = \sum_{j,m} \langle f, \psi_{j,m} \rangle \psi_{j,m}$.

Applying a hard wavelet threshold algorithm leads to reducing to zero all wavelet coefficients whose absolute values are not higher than a certain threshold value $T = T(\sigma)$, where the standard noise deviation σ is estimated by the signal f . The other coefficients do not change and the coefficients obtained in this way are denoted by $c_{j,m}$. The estimator \tilde{f}_0 is calculated using $Df = \sum_{j,m} c_{j,m} \psi_{j,m}$.

An alternative threshold method is the soft threshold value algorithm,

where the new wavelet coefficients are determined by a shrinkage function [4]:

$$shr(c) = \text{sgn}(c) \max(|c| - T, 0).$$

The properties of these thresholding estimators remain valid for nonorthogonal Riesz bases and frames. The frame redundancy leads to a lower risk, which makes it preferable in applications. What is more, these methods are summarized for some improved multiscale transformations based on: curvelets, bandelets, shearlets, etc. [23, 6].

2.2. Entropy and frames. In the information theory, entropy is used to quantitatively measure the information content of a random source [22]. The zero entropy corresponds to the absence of information, while its higher value reflects the increased informativeness of the relevant data. By *information* in this paper we mean the noise level in the image concerned. In the signal processing theory, using the entropy properties for the signal $x \in L_2(R)$ represented by the sequence of its coefficients $\{x_i\}_{i \in I}$ on an orthogonal basis or, more generally, on the frame in this Hilbert space, the following equality is valid: $E(x) = \sum_{i \in I} E(x_i)$, where

$$E(x_i) = -|x_i^0|^2 \cdot \log_2(|x_i^0|^2), \quad x_i^0 = x_i \cdot \left(\sum_{i \in I} x_i^2 \right)^{-0,5}.$$

Therefore, the Shannon entropy for this signal is given by $-E(x) = -\sum_{i \in I} |x_i^0|^2 \cdot \log_2(|x_i^0|^2)$ [3, 20].

2.3. Shearlet transform. For efficient representation of multidimensional data, the *Shearlet transform (ST)* is used. The Continuous Shearlet-system is generated using the following operators [15]:

- scaling D_{A_a} , where $A_a = \text{diag}(a, \sqrt{a})$, $a \in R^+$ is the matrix of the parabolic scaling;
- shearing (ensures the anisotropy of the transform) D_{S_s} , $s \in R$, where $S_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$ is the shear matrix;
- translating T_τ , where $T_\tau \phi(t) = \phi(t - \tau)$.

The continuous shearlet system for the function $\psi \in L_2(R^2)$ is the set $\left\{ \psi_{a,s,\tau}(t) = T_\tau D_{S_s} D_{A_a} \psi(t) = a^{-3/4} \psi(A_a^{-1} S_s^{-1}(t - \tau)); a \in R^+, s \in R, \tau \in R^2 \right\}$, and the associated with it *ST* of the function $f \in L_2(R^2)$ is $ST_\psi f(a, s, \tau) = \langle f, \psi_{a,s,\tau} \rangle$.

There are various algorithms for the implementation of the respective *Discrete Shearlet Transform (DST)* [15]. The present paper uses the algorithm *Fast Finite Shearlet Transform (FFST)* proposed by S. Häuser and implemented on the basis of cone-adapted continuous shearlet systems [12]. For the construction of the classical wavelet, the Meyer function is used as a scaling function [18] and the proposed algorithm is based on *Fast Discrete Fourier Transforms*. The resulting discrete shearlet system $\{\psi_{j,k,m}^{\kappa}(\omega)\}$ forms a Parseval frame of the finite Euclidean space, a subspace of $L_2(\mathbb{R}^2)$, which provides the construction of the *Inverse Discrete Shearlet Transform (IDST)*.

3. Proposed method. Wavelet representations are non-optimal in case of piecewise regular multidimensional functions and in this situation the wavelet thresholding estimator does not provide a minimax risk. It has been shown that a denoising estimator based on the thresholding of shearlet coefficients essentially possesses the minimax optimality for images with edges [11, 7].

Noise in medical images is in effect too complicated to describe in an absolutely correct way by a given model of statistical presentation of unknown data. Motivated by the complicated conditions of noise in medical images, O. Tischenko et al. [24] and A. Borsdorf et al. [2] proposed a denoising method by wavelet thresholding based on the correlation between the two *CT* medical images containing almost identical information. The basic idea is that, in contrast to the structural information, noises in these images are not in correlation with the passage of time.

3.1. Entropy-based shearlet thresholding. The present paper considers the generalized model $f = F(f_0, n)$, where f and f_0 are respectively the noisy image and the reconstructed (denoised) image, n is the noise component and the function F describes the relation between the two images. It does not use certain statistics of the random variable n . The reduction of the noise component is performed in the frequency domain, using a *DST*.

The proposed shearlet thresholding comprises the following steps:

- an adapted selection of the maximum scale;
- shearlet decomposition;
- determining the adapted thresholding rule;
- computing the affine thresholding estimator;
- reconstructing the denoised image.

The adapted choice in this algorithm is based on qualitative assessment obtained by the criterion of the relative change (rate of change) of the Shannon entropy $-\frac{dE}{E_j} = \frac{|E_{j+1} - E_j|}{|E_j|}$, where E_j is the entropy at the j th level of the decomposition of the image [1]. *CT* images of different groups of anatomical organs have been used as surveyed medical images.

The test image f is subjected to *FFST* for all permitted scales. The criterion for the selection of the relevant scale j_a adapted to the image is determined by the dependence of the entropy change rate on the level of the shearlet decomposition. The calculation of the shearlet coefficients is performed in terms of the frame $\Phi = \{\psi_{j,k,m}^\kappa\}$, $j = 0, 1, \dots, j_a$, i. e., $\langle f, \psi_{j,k,m}^\kappa \rangle$.

The rule for determining the threshold criterion in the algorithm is based on the change of $\frac{dE}{E_i}(j)$, ($i = 1, 2, \dots; j = 0, 1, \dots, j_a$) along an indicative graded scale of threshold values. In accordance with the determined threshold T_j , threshold processing is applied to all shearlet coefficients of this scale, with the exception of those with the lowest frequency energy ($\kappa = 0$). The quality of the noise reduction in the image directly depends on the choice of the threshold value. The low threshold keeps the background in the coefficients and its high values lead to the loss of coefficients containing structural information.

The resultant affine shearlet thresholding estimator of f_0 in the Φ frame has the representation $Df = \sum_{\kappa,j,k,m} c_{j,k,m}^\kappa \psi_{j,k,m}^\kappa$. The image f_0 itself is reconstructed by *IDST* using the coefficients $c_{j,k,m}^\kappa$.

3.2. Experimental results. In order to conduct the study and the analysis of the proposed image denoising algorithm, *CT* images of four groups of anatomical organs have been used: *Head, Spine, Knee* and *Cardiac*. Ten slides have been picked at random from a randomly selected subfolder for each study. The dimensions of the medical images are 1024×1024 and the output format is *DICOM (Digital Image and Communications in Medicine)*. The images have been obtained by a *Siemens CT scanners—Somatom Definition* and *Somatom Spirit*, without any additional processing. The proposed image denoising algorithm based on *ST* is a program implemented in *Matlab* using the additional package *FFST*.

In order to evaluate the scale levels of decomposition, the images are subjected to *FFST*, up to the maximum possible level of decomposition. The resulting average results of the calculation of the relative entropy change for the corresponding scale levels are represented in the graphs in Fig. 1.

Therefore, for all groups of studied organs, the entropy change rate undergoes a qualitative change on the third level of the scale. This can be considered

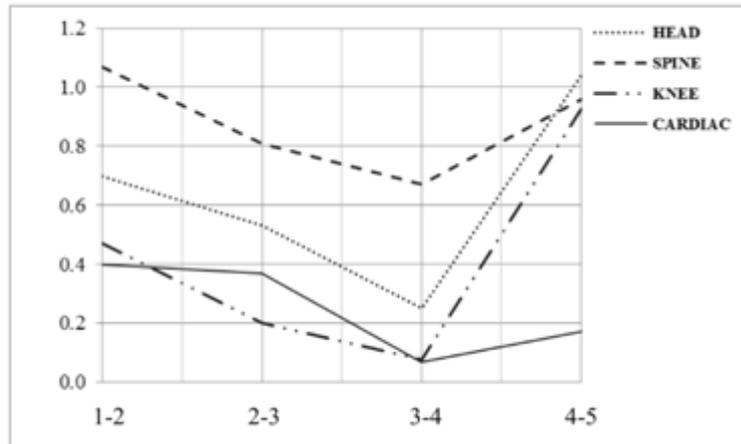


Fig. 1. The dependence of the relative entropy change rate on the level of decomposition

as an indicator for changes in the type of the removed information, which gives grounds to choose $j_a = 3$. Using higher scale levels will lead to image smoothing and, respectively, to loss of some local peculiarities.

The choice of the adapted threshold T_j , ($j = 0, 1, 2, 3$) is determined on the basis of the dependence of $\frac{dE}{E_i}(j)$ on an eight-grade scale and a step unit. The graph of this dependence for $j = 3$ is given in Fig. 2. Practically, there is no point in choosing a threshold value greater than three, because the rate of entropy change after the third scale can be assumed to be equal to zero.

4. Results and comments.

4.1. Test images. Usually, as a performance indicator for the algorithms for noise reduction, a standard peak signal-to-noise ratio is used. It is measured in decibels and is defined by $PSNR = 20 \lg \frac{255 \cdot N}{\|f - f_0\|_F}$, where $\|\bullet\|_F$ is the Frobenius norm. Standard Lena and Barbara images noised by different levels of Rician noise are used as test images to perform a comparative analysis of the proposed algorithm and other denoising methods.

Fig. 3 shows the results obtained with the proposed method for the two test images noised with 5% Rician noise.

The results of the comparative analysis for the given method with *Classical Shearlet Algorithm* and *Shearlet-MOGA* [8] for different levels of Rician noise are

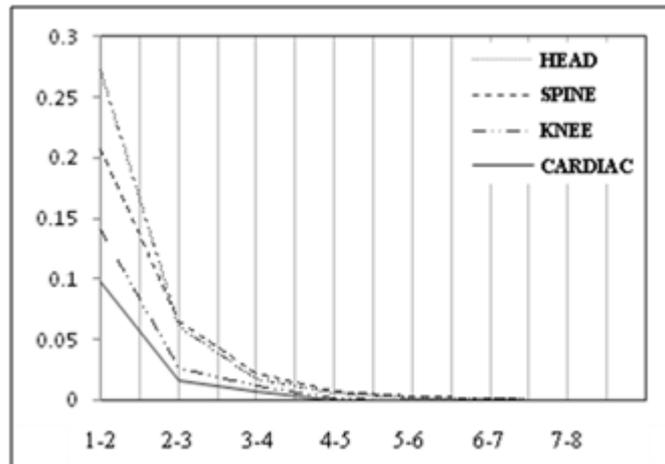


Fig. 2. The dependence of the relative entropy change rate on the threshold value



Fig. 3. Results for 5% Rician Noise

presented in Table 1. The data obtained show that our algorithm gives higher values of *PSNR* in more than 83% of the cases.

Table 1. Results for a Different Noise Level and Different Algorithm

Image	σ [%]	PSNR		
		Shearlet	Shearlet-MOGA	Proposed algorithm
Lena	10	34.37	35.13	33.49
	20	31.80	33.12	32.00
	30	29.23	30.09	31.64
Barbara	10	33.17	33.11	34.10
	20	29.42	29.30	32.72
	30	26.33	27.54	31.44

4.2. Real medical images. The developed algorithm for noise reduction is also applied to real computer tomography images from the anatomic groups mentioned in section 3.2. In addition, the metric function *SSIM* is used in order to assess the quality of the denoised images. Unlike *PSNR*, this index is compatible with human visual perception. *SSIM* is defined as follows: $SSIM = [l(f, f_0)]^\alpha \cdot [c(f, f_0)]^\beta \cdot [s(f, f_0)]^\gamma$, where $l(f, f_0)$, $c(f, f_0)$ and $s(f, f_0)$ are respectively luminance, contrast and structural comparison functions. The positive constants α , β and γ are used to weigh each comparison function.

Fig. 4 and Fig. 5 show one representative of each studied group of anatomical organs: an input image (corrupted with noise), the same image after applying the proposed method (denoised), the error between the two images and the corresponding values of the functions *PSNR* and *SSIM*.

The results show that the proposed algorithm can effectively reduce the noise in *CT* images while retaining the relevant structural information.

5. Conclusion. This article proposes a method for constructing an affine shearlet thresholding estimator, which does not require a concrete noise model in the algorithm for reducing image noise. It shows its possibility for adaptation to specific medical images databases through the Shannon entropy. Another advantage, in comparison with some shearlet-based thresholding methods, is the low computational value of the algorithm due to averaging calculations in the different frequency bands. This method can be used as a preliminary step in solving some tasks related to classification and recognition in large medical databases.

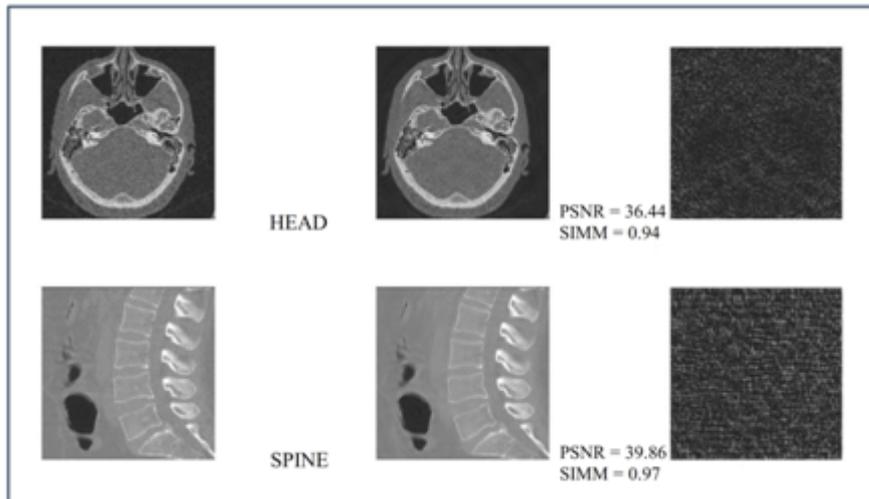


Fig. 4. The results of the proposed image denoising method (Head, Spine)

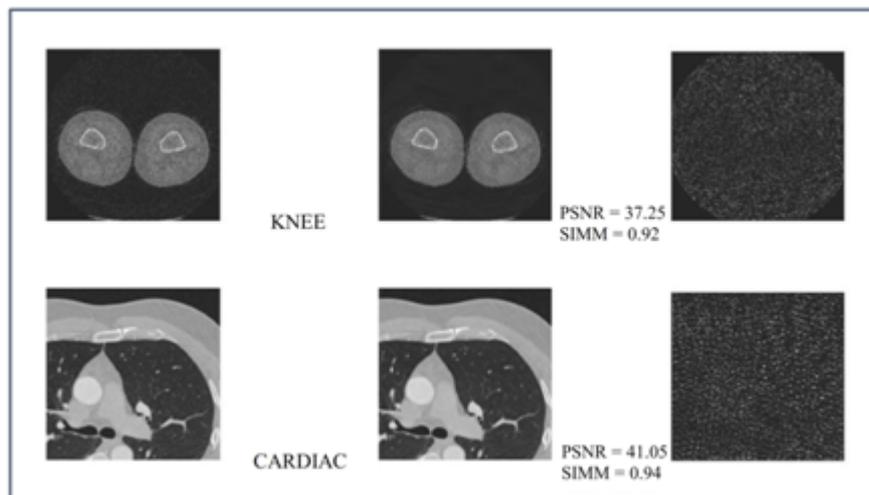


Fig. 5. The results of the proposed image denoising method (Knee, Cardiac)

REFERENCES

- [1] BABICHEV S., N. BABENKO, O. DIDYK, V. LYTVYNNENKO, A. FEFELOV. Chromatogram filtering using wavelet analysis with use of the criterion of entropy. *System Technology*, **6** (2010), No 71, 17–32.
- [2] BORSODORF A., R. RAUPACH, T. FLOHR, J. HORNEGGER. Wavelet based Noise Reduction in CT-Images using Correlation Analysis. *IEEE transactions on medical imaging*, **27** (2008), No 12, 1685–1703.
- [3] COIFMAN R. R., M. V. WICKERHAUSER. Entropy based algorithms for best basis selection. *IEEE Trans. on Inform. Theory*, **38** (1992), No 2, 713–718.
- [4] DONOHO D. L., I. M. JOHNSTONE. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, **81** (1994), 425–455.
- [5] DONOHO D. L., I. M. JOHNSTONE. Denoising by soft-thresholding. *IEEE Trans. on Inform. Theory*, **41** (1995), 613–627.
- [6] DONOHO D. L., I. M. JOHNSTONE. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Stat. Assoc.*, **90** (1995), No 432, 1200–1224.
- [7] EASLEY G., D. LABATE, W. LIM. Sparse directional image representations using the discrete shearlet transform. *Appl. Comput. Harmon. Anal.*, **25** (2008), No 1, 25–46.
- [8] FAN Z., Q. SUN, F. RUAN, Y. GONG, Z. JI. An Improved Image Denoising Algorithm based on Shearlet. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, **6** (2013), No 4, 475–483.
- [9] GRAVEL P., G. BEAUDOIN, J. A. DE GUISE. A method for modeling noise in medical images. *IEEE Trans. Med. Imag.*, **23** (2004), No 10, 1221–1232.
- [10] GUDBJARTSSON H., S. PATZ. The Rician distribution of noisy MRI data. *Magn. Reson. Med.*, **34** (1995), 910–914.
- [11] GUO K., D. LABATE. Optimally sparse multidimensional representation using shearlets. *SIAM J. Math. Anal.*, **39** (2007), 298–318.
- [12] HÄUSER S., G. STEIDL. Fast finite shearlet transform: a tutorial. Preprint. University of Kaiserslautern, 2014.
- [13] JIN Y., E. ANGELINI, A. LAINE. Wavelets in medical image processing: denoising, segmentation and registration. In: J. S. Suri, D. L. Wilson, S. Laxminarayan (eds). *Handbook of Biomedical Image Analysis. International Topics in Biomedical Engineering*. Springer, Boston, MA, 2005, 305–358.

- [14] KARAMATA B., M. LAUBSCHER, T. LASSER. Speckle statistics in optical coherence tomography. *Journal of the Optical Society of America*, **22** (2005), No 4, 593–596.
- [15] KUTYNIOK G., D. LABATE (EDS). Shearlets: Multiscale Analysis For Multivariate Data. *Applied and numerical harmonic analysis*, Birkhäuser/Springer, New York, 2012.
- [16] LEI T., W. SEWCHAND. Statistical approach to X-ray CT imaging and its applications in image analysis—Part I: Statistical analysis of X-ray CT imaging. *IEEE Trans. Med. Imag.*, **11** (1992), No 1, 53–61.
- [17] MALLAT S. A wavelet tour of signal processing. New York, Academic Press, 1998.
- [18] MEYER Y. Oscillating patterns in image processing and nonlinear evolution equations. *University Lecture Series*, **22**. American Mathematical Society, Providence, RI, 2001.
- [19] OLLINGER J. M., J. A. FESSLER. Positron emission tomography. *IEEE Signal Process. Mag.*, **14** (1997), No 1, 43–55.
- [20] RICAUD B., B. TORRÈSANI. Refined support and entropic uncertainty inequalities. *IEEE Trans. on Inform. Theory*, **59** (2013), No 7, 4272–4279.
- [21] SANHES J. M., J. C. NASCIMENTO, S. MARQUES. Medical image noise reduction using the Sylvester-Lyapunov equation. *IEEE Transactions on Image Processing*, **17** (2008), No 9, 1522–1539.
- [22] SHANNON C. E. A mathematical theory of communication. *The Bell System Technical Journal*, **27** (1948), No 3, 379–423; No 4, 623–656.
- [23] STARCK J. L., E. J. CANDÈS, D. L. DONOHO. The curvelet transform for image denoising. *IEEE Trans. Image Proc.*, **11** (2002), 670–684.
- [24] TISCHENKO O., C. HOESCHEN, E. BUHR. An artefact-free structure-saving noise reduction using the correlation between two images for threshold determination in the wavelet domain. *Proc. SPIE*, **5747** (2005), 1066–1075.

Miroslav Petrov

St. Cyril and St. Methodius University of Veliko Turnovo

Faculty of Mathematics and Informatics

Department of Computer Systems and Technologies

5000, Veliko Turnovo, Bulgaria

e-mail: m_petrov75@abv.bg

Received May 29, 2017

Final Accepted June 21, 2017