Wavelet Based Noise Reduction by Identification of Correlations

Anja Borsdorf, Rainer Raupach, and Joachim Hornegger

Institute of Pattern Recognition, Friedrich-Alexander-University Erlangen-Nuremberg
Siemens Medical Solutions, Forchheim

Abstract. In this paper we present a novel wavelet based method for edge preserving noise reduction. In contrast to most common methods, the algorithm introduced here does not work on single input data. It takes two or more spatially identical images, which are both impaired by noise. Assuming the statistical independence of noise in the different images, correlation computations can be used in order to preserve structures while reducing noise. Different methods for correlation analysis have been investigated, on the one hand based directly on the original input images and on the other hand taking into account the wavelet representation of the input data. The presented approach proves to be suited for the application in computed tomography, where high noise reduction rates of approximately 50% can be achieved without loss of structure information.

1 Introduction

Particularly in diagnostic imaging, data contains noise predominantly caused by quantum statistics. A common problem in image processing, therefore, is the reduction of this pixel noise. Several approaches for edge-preserving noise reduction are known. The goal of all of these methods is to lower the noise power without averaging across edges. Some popular examples are nonlinear diffusion filtering [1] and bilateral filtering [2], which directly work in the spatial domain. Additional approaches exist that reduce noise based on the frequency representation of the input data, in particular wavelet-domain denoising techniques. Most of these algorithms are based on the observation that information and white noise can be separated using an orthogonal basis in the wavelet domain, as described e.g. in [3]. Structures (such as edges) are located in a small number of dominant coefficients, while white noise, which is invariant to orthogonal transformations and remains white noise in the wavelet domain, is spread across a range of small coefficients. This observation dates back to the work of Donoho and Johnstone [4]. Using this knowledge, thresholding methods were introduced, which erase insignificant coefficients but preserve those with larger values. Several techniques have been developed to further improve the detection of edges and relevant image content, for instance by comparing the detail coefficients at adjacent scales [5,6]. Most denoising methods based on wavelets suffer from the limitation that they are only applicable to white noise. A more robust algorithm which adapts itself to several types of noise is for instance presented in [7].

Nevertheless, most existing methods for noise reduction work on single image data and their ability to distinguish between information and noise, therefore, strongly depends on the size and the contrast of image structures. In contrast, if two or more images

K. Franke et al. (Eds.): DAGM 2006, LNCS 4174, pp. 21–30, 2006.

are available, which show the same information but statistically independent noise, the differentiation between signal and noise can be further improved [8]. By comparing the input images either in spatial domain or on the basis of the wavelet coefficients, frequency dependent weighting factors can be computed. These weighting factors are then used to eliminate noise, whilst maintaining structural information in the wavelet representation of the images. Reconstruction of the modified wavelet coefficients yields an image with suppressed noise but including all structures detected as correlations between the input images.

This paper is structured as follows: After summarizing the basic concepts of the wavelet transformation in section 2 the noise reduction algorithm is introduced in detail in section 3. In section 4 the achieved results for the specific applications in computed tomography and fluoroscopy are presented.

2 Wavelet Transformation

Wavelets are generated from a single basis function $\psi(t)$ called mother wavelet by means of scaling and translation:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{|s|}}\,\psi\left(\frac{t-\tau}{s}\right); \quad s,\tau \in R, s \neq 0, \tag{1}$$

where s is the scaling parameter and τ is the translation parameter. Wavelets must have zero mean and have bandpass like spectrum. For the computation of the discrete wavelet transformation only discrete pairs of s and τ are used. Taking the discrete parameters

$$s_j = 2^{-j} \quad \text{and} \quad \tau_k = k \cdot s_j = k \cdot 2^{-j}; \quad j,k \in N^0, \tag{2}$$

where k is the translation and j the scale index, results in a dyadic sampling. Using these parameters a family of wavelets, spanning the $L^2(R)$ can be derived from a mother wavelet $\psi(t)$ as follows:

$$\psi_{j,k}(t) = \sqrt{2^j}\,\psi(2^j t - k). \tag{3}$$

The discrete wavelet transform (DWT) of a 1D function $f(t)$ can then be computed by projecting the function onto the set of wavelets:

$$c_{j,k} = \int_{-\infty}^{\infty} f(t)\psi^*_{j,k}(t)dt, \tag{4}$$

where $\psi^*_{j,k}(t)$ is the complex conjugate of $\psi_{j,k}(t)$.

The algorithm introduced by Mallat [9], allows a fast computation of the discrete dyadic wavelet transformation. The wavelet coefficients are computed by iteratively decomposing the singnal into its high-pass filtered details and low-pass filtered approximation, reducing the resolution of the signal in each iteration by a factor of two. It can be shown that the discrete dyadic wavelet decomposition can be computed by an iterated filter bank (see [10] for details).

When dealing with images the two-dimensional wavelet transformation needs to be used. The one-dimensional transformation can be applied to the rows and the columns

in succession, which is referred to as a separable two-dimensional wavelet transformation. After this decomposition four two-dimensional blocks of coefficients are available, including on the one hand the lowpass filtered approximation image $\boldsymbol{C}$ and three detail images $\boldsymbol{D}^x, \boldsymbol{D}^y, \boldsymbol{D}^{xy}$. Analogously to the 1D case, the multiresolution wavelet decomposition can be computed iteratively from the approximation coefficients.

At every decomposition level, the detail images include high frequency structure information in horizontal, vertical and diagonal direction together with noise in the respective frequency band. Goal of the noise suppression method is to detect those detail coefficients which represent structure information. These coefficients should be kept unchanged, while coefficients, which are due to noise should be eliminated or at least be suppressed.

3 Filtering Algorithm

Fig. 1 shows a brief overview of the different steps used for the noise reduction algorithm. Although the algorithm can also be extended to work with more than two input images, without loss of generality, only the case of two images will be considered in the following.

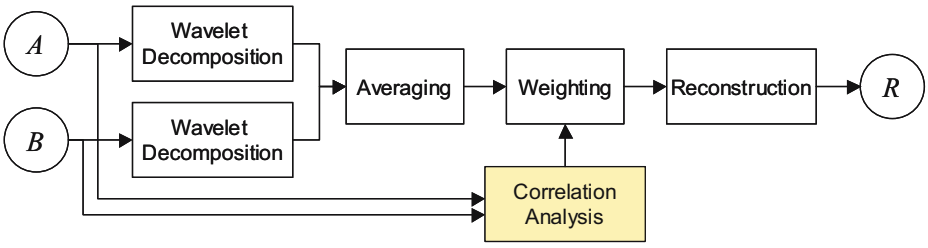


Fig. 1. Overview of the noise reduction method

The two input images A and B are both decomposed into multiple frequency bands by a 2D discrete dyadic wavelet transformation. Of course, for the reduction of high frequency noise only those decomposition levels covering the frequency bands of the noise spectrum are of interest. Therefore, it is not necessary to compute the wavelet decomposition up to the coarsest scale. In our experiments, two to four decomposition levels were sufficient. For each decomposition level a similarity matrix is computed based on correlation analysis. The frequency dependent local discrepancy measurement can be based directly on the comparison of the original input images or on the wavelet representation of the input images. By the application of a predefined weighting function to the computed similarity values a level dependent weighting factor is computed. The resulting mask should preferably include ones in regions where structure information has been detected and values smaller than one elsewhere. The averaged wavelet coefficients of the input images, i.e. the detail coefficients, can then be weighted according to this mask. Averaging in the wavelet domain allows the computation of just one inverse wavelet transformation in order to get a noise suppressed result image R.

3.1 Correlation Analysis

Goal of the correlation analysis is to estimate the probability of a coefficient to correspond to structural information. This estimate is based on the measurement of the local homology of the input images. In the following, three different methods of similarity computation will be introduced, measuring the similarity based on the original input images, secondly based on the approximation coefficients and thirdly directly from the detail coefficients. The core idea behind all of these methods is similar: For the three blocks of detail coefficients $\boldsymbol{D}_l^x, \boldsymbol{D}_l^y, \boldsymbol{D}_l^{xy}$ of the wavelet decomposition, including horizontal, vertical and diagonal details, a corresponding similarity matrix $\boldsymbol{S}_l$ is computed for every level l up to the maximum decomposition level. Then, according to the defined weighting function the detail coefficients are weighted with respect to their corresponding values in the similarity matrix.

Correlation Coefficient Based Methods: One popular method for measuring the similarity of noisy data is the computation of the empirical correlation coefficient [11]:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}, \tag{5}$$

where $x = x_1, x_2, \ldots, x_n$ and $y = y_1, y_2, \ldots, y_n$ are two sequences of data each with n data points. The mean values of x_i and y_i are denoted as $\overline{x}$ and $\overline{y}$. The empirical correlation coefficient also known as *Pearson's correlation* is independent from both origin and scale and takes values out of the interval $[-1; 1]$, whereas one means perfect correlation, zero no correlation and minus one perfect anticorrelation.

This correlation coefficient can now be used in order to compute the local homology between the input images, by taking blocks of pixels out of the two images and computing the correlation coefficient (see Fig.2(a)). Of course, the pixels used for similarity measurement at a respective position should be closely associated with the particular detail coefficient. This should later on be weighted according to the computed similarity value. Preferably all pixels from the original input image, which influenced the detail coefficient at the current position and scale, through the computation of the wavelet decomposition, should be incorporated into the similarity computations. It is clear that with increasing decomposition level the size of the pixel regions in the original image must also increase. Additionally, it is necessary to take the length m of the wavelet filters into consideration. Altogether, the number of pixels n_l of the original image influencing a coefficient at level l can be computed iteratively according to:

$$n_l = 2 \cdot n_{l-1} + m - 2; \quad \text{with } n_1 = m. \tag{6}$$

With this size adaptation the application of the algorithm in combination with arbitrary wavelets, which can be represented by FIR lowpass and highpass filters, becomes possible. However, the computational costs are quite high, because of the increasing size of the pixel regions in dependence on the decomposition level.

Improved results can be achieved with respect to performance as well as image quality if the correlation computations are not based on the original input images but on the

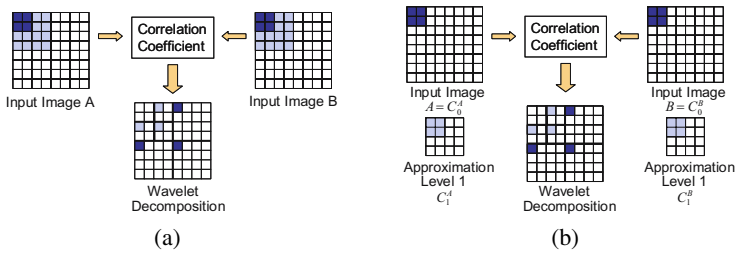


Fig. 2. Correlation computations based on correlation coefficients - (a) based on the original input images, (b) based on the approximation images of the previous decomposition level

approximation images. The multiresolution wavelet decomposition is computed iteratively. Thus the detail coefficients at level l are gained from the approximation image of the previous decomposition level. A very close connection between the detail coefficients and the computed similarity values can be obtained if the pixel regions are also taken from the approximation images of the previous decomposition level. The advantage of this approach is that the size of the pixel regions no longer depends on the decomposition level (see Fig.2(b)). Only the length of the wavelet filters needs to be considered. The disadvantage is that the approximations at all scales need to be stored for this method, although only the approximations of the maximum decomposition level are needed for perfect reconstruction.

Both of the methods mentioned so far have to deal with the same problem. The image regions are adjusted to the length of the filter used for analysis, but not to the coefficients of the filter. All intensity values within the considered pixel region are weighted equally. The result is that edges of higher contrast dominate the correlation values, as long as they occur within the region covered by the filter. However, if the filter coefficients should be taken into consideration all three blocks of detail coefficients must be treated separately, because the corresponding 2D filters are different. A third alternative method, where the similarity is directly computed from the detail coefficients circumvents this problem.

Gradient Approximation: The core idea behind the similarity measurement based on detail coefficients at level l is to use the fact that horizontal and vertical detail coefficients $\boldsymbol{D}_l^x$ and $\boldsymbol{D}_l^y$ can be regarded as approximations of the partial derivatives of the approximation image $\boldsymbol{C}_{l-1}$. Coefficients in $\boldsymbol{D}_l^x$ show high values at positions where high frequencies in x-direction are present and $\boldsymbol{D}_l^y$ where sudden changes in contrast in y-direction can be found. If these two aspects are considered together, we get an approximation of the gradient field of $\boldsymbol{C}_{l-1}$:

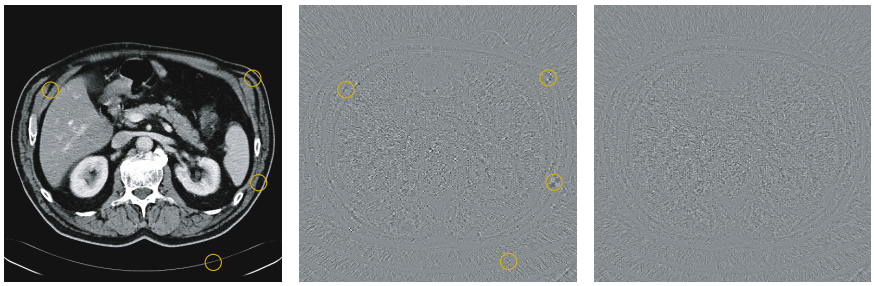
$$\nabla \boldsymbol{C}_{l-1} = \begin{pmatrix} \partial \boldsymbol{C}_{l-1}/\partial_x \\ \partial \boldsymbol{C}_{l-1}/\partial_y \end{pmatrix} \approx \begin{pmatrix} \boldsymbol{D}_l^x \\ \boldsymbol{D}_l^y \end{pmatrix}. \tag{7}$$

The detail coefficients in x- and y-direction of both decompositions approximate the gradient vectors with respect to equation (7). The similarity can then be measured by

computing the angle between the corresponding gradient vectors [8]. The goal is to obtain a similarity value in the range of $[-1, 1]$, analogously to the correlation computations above. Therefore, we take the cosine of the angle resulting in:

$$\boldsymbol{S}_l = \frac{\boldsymbol{D}_l^{A\,x}\boldsymbol{D}_l^{B\,x} + \boldsymbol{D}_l^{A\,y}\boldsymbol{D}_l^{B\,y}}{\sqrt{\left(\boldsymbol{D}_l^{A\,x}\right)^2 + \left(\boldsymbol{D}_l^{A\,y}\right)^2}\sqrt{\left(\boldsymbol{D}_l^{B\,x}\right)^2 + \left(\boldsymbol{D}_l^{B\,y}\right)^2}}, \tag{8}$$

where the superscript A refers to the first and B to the second input image. The gradient approximation method and the more time consuming computation of the correlation coefficients explained above are closely related. Nevertheless, the approaches are not identical and do not generally lead to the same results. The application of the algorithm for noise reduction based on the gradient approximation, as introduced so far, sometimes leads to visible artifacts in the resulting images. Fig. 3(b) and the difference image Fig. 3(b) give an example where this problem can be seen in case of using the Haar wavelet. Noticeably, the artifacts predominantly emerge where diagonal structures



(a) (b) (c)

Fig. 3. Artifacts due to weighting down correlated diagonal coefficients with gradient approximation method - (a) noise suppressed image with gradient approximation without separated treatment of diagonal coefficients, (b) difference image to average of input images, (c) difference image after special treatment of diagonal coefficients

appear in the image, and their shape generally enforces the assumption that diagonal coefficients at different decomposition levels are falsely weighted down. Reason for this is that diagonal patterns exist, which lead to vanishing detail coefficients in x- and y-direction. If the norm of one of the approximated gradient vectors is too small or even zero, no reliable information about the existence of correlated diagonal structures can be obtained from equation (8).

The simplest possibility for eliminating the artifacts is to weight only the detail coefficients $\boldsymbol{D}_l^x$ and $\boldsymbol{D}_l^y$ based on the similarity measurement $\boldsymbol{S}_l$ and leave the diagonal coefficients $\boldsymbol{D}_l^{xy}$ unchanged. Of course this avoids artifacts in the resulting images, but, unfortunately, noise included in the diagonal coefficients remains unchanged, leading to a lower signal-to-noise ratio for the denoised image. From equation (8), we can recognize that the similarity value is computed only with respect to $\boldsymbol{D}_l^x$ and $\boldsymbol{D}_l^y$. The

diagonal coefficients do not influence the computation. However, the idea to extend the approximated gradient vector (see equation (7)) by the diagonal coefficients to a three dimensional vector does not lead to the desired improvements. In case of vanishing detail coefficients in x- and y- direction, no quantitative relation between the diagonal coefficients can be gained. Moreover, the extension of the approximated gradient vector by the diagonal coefficient is not a suitable solution. A diagonal coefficient can be interpreted as second order derivative and is therefore very sensitive to noise. Mixing it with the detail coefficients in x- and y-direction generally leads to less reliable similarity measurements.

In order to avoid artifacts while still reducing noise in the diagonal coefficients, only the detail coefficients $\boldsymbol{D}_l^x$ and $\boldsymbol{D}_l^y$ are weighted, depending on the similarity measurement computed from equation (8). The diagonal detail coefficients are treated separately. The weighting function for the diagonal coefficients is based on the correlation analysis between $\boldsymbol{D}_l^{A\,xy}$ and $\boldsymbol{D}_l^{B\,xy}$:

$$\boldsymbol{S}_l^{xy} = \frac{2\boldsymbol{D}_l^{A\,xy}\boldsymbol{D}_l^{B\,xy}}{\left(\boldsymbol{D}_l^{A\,xy}\right)^2 + \left(\boldsymbol{D}_l^{B\,xy}\right)^2}. \tag{9}$$

Using this extension for separated weighting of the diagonal coefficients, denoising results are free of artifacts (see Fig.3(c)).

3.2 Weighting Function

The simplest possible method for weighting the coefficients is to use a thresholding approach. If the similarity value $\boldsymbol{S}_l$ at position (x, y) is above a defined value τ_l, the detail coefficients are kept unchanged, otherwise they are set to zero [8]. The weighting function can be defined as

$$W_l(\boldsymbol{S}_l(x,y)) = \begin{cases} 1 & \text{if } \boldsymbol{S}_l(x,y) \geq \tau_l \\ 0 & \text{otherwise} \end{cases}. \tag{10}$$

However, the choice of an appropriate threshold very much depends on the noise level of the input images. Therefore, with increasing noise level in the input images the threshold should be set less strictly and with some tolerance. Preferably the threshold should be chosen level dependent, meaning that the threshold should be abated for higher decomposition levels. Generally the use of continuous weighing functions like

$$W_l^{powN}(\boldsymbol{S}_l(x,y)) = \left(\frac{1}{2}\left(\boldsymbol{S}_l(x,y)+1\right)\right)^N \in [0,1]\,, \tag{11}$$

where no hard decision about the maintenance or the discardal of coefficients is required, leads to better results. The power N can also be chosen level adaptive.

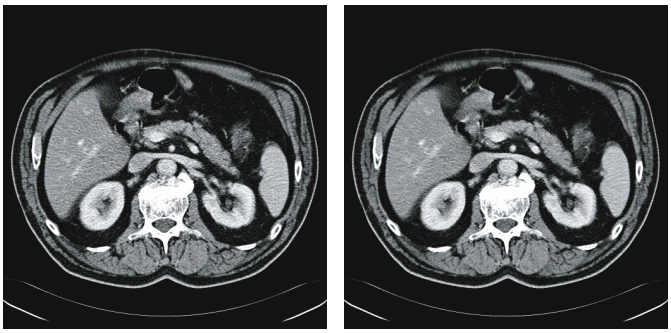
4 Experimental Evaluation

4.1 Computed Tomography

One important application of the noise reduction algorithm introduced above can be found in X-ray computed tomography (CT). In CT always a tradeoff between pixel

noise, dose of radiation and image resolution must be found. Reducing the dose of radiation for example by a factor of two increases the noise level in the images by a factor of $\sqrt{2}$. Goal of the application of the noise reduction algorithm to CT images is to achieve improved image qualtiy without increasing the dose of radiation, or, the other way round, to reduce the dose of radiation without impairing image quality.

Spatially identical images with uncorrelated noise can be generated through separate reconstruction from disjoint sets of projections. For example two images can be reconstructed, each using only every second projection. Specifically, one image is computed from the even, and the other one from the odd numbered projections. Due to the reconstruction with only half of the projections, the noise level of the two generated images increases by a factor of $\sqrt{2}$. By averaging the wavelet coefficients of the input images, the result image corresponds to the image reconstructed with the complete set of projections, where additionally noise is reduced. Usually, a loss of image resolution through splitting the projections into two halfs can be obviated because the overall number of projections in CT can be assured to be high enough.

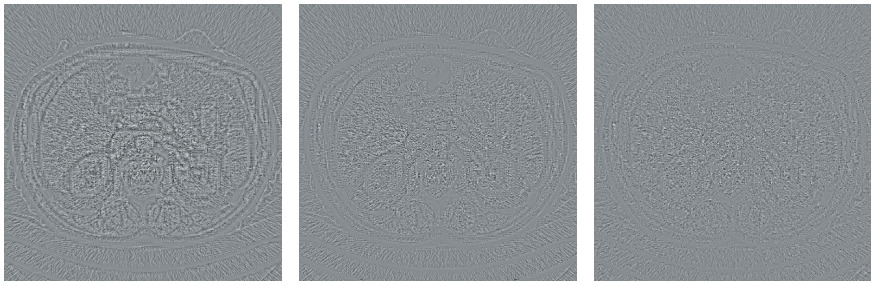


(a) (b)

Fig. 4. Application of the noise reduction algorithm to CT images - (a) average of input images (standard deviation: $\sigma \approx 52$ HU), (b) denoised result ($\sigma \approx 25$ HU)

Fig.4(b) shows the noise suppressed result image in comparison to the average of the input images Fig.4(a). It can be seen clearly that especially in homogeneous image regions, as for example in the region of the liver, noise is reduced, while structures and also small details are preserved. For clinical tests we used two CT slices, one from the abdomen, the other from the thorax. For each slice the average image and nine different configurations of denoised images were computed (see [12], [13] for details). The nine noise suppressed images were compared to the average image by two radiologists independently. All images of the tests were unlabeled. The result was that the average of the input images has never been judged superior to the noise suppressed images. In average the pixel noise σ in the noise suppressed images has been reduced by 50% in comparison to the average of input images.

The different approaches for correlation analysis can be assessed by comparing the difference images, which are presented in Fig.5. It can be seen that the correlation



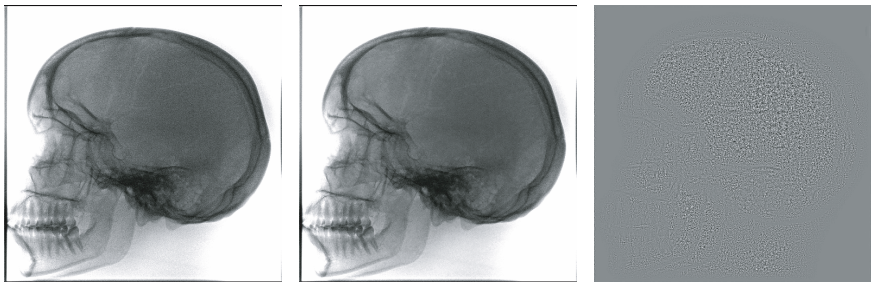
(a) (b) (c)

Fig. 5. Comparison of correlation analysis methods: difference images between result image and the average of input images - (a) CC based on original images, (b) CC based on approximation images, (c) gradient approximation

coefficient (CC) based method, where the pixel regions are taken from the original input images is less precise than the other two approaches, because structure information is also included in the difference images. The other two approaches lead to nearly the same good results. In regions of edges no noise is reduced, but the quality of the edge is kept unchanged. In the shown examples the Haar wavelet has been used. The experimental results with different wavelets showed, that especially biorthogonal spline wavelets, like the CDF9/7 wavelet [14], are well suited for the noise reduction algorithm.

4.2 Fluoroscopy

A second clinical application of the introduced method for noise reduction can be found in fluoroscopy, where sequences of x-ray projections are acquired. Therefore, achieving the maximum image quality with a minimum of radiation dose is required. In Fig.6 the initial experimental results achieved for x-ray images of a human skull are presented.



(a) (b) (c)

Fig. 6. Application of the noise reduction method to fluoroscopy images of the human skull - (a) average of input images, (b) denoised image, (c) difference image

5 Conclusion

We presented a novel edge-preserving wavelet based method for noise reduction. The algorithm works on two input images, which show the same information whereas the noise between the input images is uncorrelated. Using this property, correlation computations can be used in order to differentiate between structures and noise. Three different approaches of correlation analysis have been discussed. Especially the gradient approximation approach with the introduced separated treatment of the diagonal wavelet coefficients allows an artifact free and computationally efficient noise suppression. The application of the algorithm to computed tomography images showed that a noise reduction of approximately 50% is possible without loss of stucture information. Even fine edges and small structures are preserved. For the application to fluoroscopy images it must be assured that the patient does not move. Otherwise the method must be used in combination with image registration algorithms.

References

1. Catte, F., Lions, P.L., Morel, J.M., Coll, T.: Image selective smoothing and edge-detection by nonlinear diffusion. SIAM Journal on Numerical Analysis **29** (1992) 182–193
2. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: IEEE International Conference on Computer Vision, Bombay, India (1998) 839–846 http://www.cse.ucsc.edu/ manduchi/Papers/ICCV98.pdf.
3. Hubbard, B.: Wavelets: Die Mathematik der kleinen Wellen. Birkh"auser Verlag, Basel, Schweiz (1997)
4. Donoho, D.L., Johnstone, I.M.: Ideal spatial adaptation by wavelet shrinkage. Biometrika **81** (1994) 425–455
5. Xu, Y., Weaver, J., Healy, D., Lu, J.: Wavelet transform domain filters: A spatially selective noise filtration technique. IEEE Transactions on Image Processing **3** (1994) 747–758
6. Faghih, F., Smith, M.: Combining spatial and scale-space techniques for edge detection to provide a spatially adaptive wavelet-based noise filtering algorithm. IEEE Transactions on Image Processing **11** (2002) 1062–1071
7. Pizurica, A., Philips, W., Lemahieu, I., Acheroy, M.: A versatile wavelet domain noise filtration technique for medical imaging. IEEE Transactions on Image Processing **22** (2003) 1062–1071
8. Hoeschen, C., Buhr, E., Tischenko, O.: Verfahren zur Reduktion von Rauschstrukturen in Arrays von Pixelwerten (2004) Offenlegungsschrift DE10305221A1 2004.08.26.
9. Mallat, S.: A theory for multiresolution signal decomposition: The wavelet representation. IEEE Transactions on Pattern Analysis and Maschine Intelligence **11** (1989) 674–693
10. Strang, G., Nguyen, T.: Wavelets and Filter Banks. Wellesley- Cambridge Press (1996)
11. Weisstein, E.: Correlation coefficient (2005) http://mathworld.wolfram.com/ CorrelationCoefficient.html.
12. Borsdorf, A.: Noise reduction in CT images by identification of correlations (2005)
13. Borsdorf, A., Raupach, R., Hornegger, J.: Reduction of quantum noise in CT-images by identification of correlations. In: Proceedings of the 2nd Russian-Bavarian Conference on Bio-Medical Engineering, Moskow (2006)
14. Getreuer, P.: Filter coefficients to popular wavelets (2005) http://www.mathworks.com.