GEOMETRICAL CONSTRAINTS IN BAYESIAN WAVELET FILTERING OF IMAGES

Ing. Tomáš ČERVINKA, Doctoral Degree Programme (1)
Dept. of Biomedical Engineering, FEEC, BUT
E-mail: xcervi07@stud.feec.vutbr.cz

Supervised by: Dr. Ivo Provazník

ABSTRACT

This paper describes a new method for the suppression of noise in images based on wavelet transform [3]. The method relies on two criteria. The first is a traditional criterion of smoothness of the image based on an approximation of the local Hőlder exponent via the wavelet coefficients. The second, novel criterion takes into account geometrical constraints, which are generally valid for natural and also simulated images. The smoothness measure and the geometrical constraints are combined in the described method in Bayesian probabilistic formulation, and are implemented as a Markov random field (MRF) image model. The manipulation of the wavelet coefficients is consequently based on the obtained probabilities. This method is proposed to quantitatively improve noise suppression comparing to classical methods based on wavelet transform. Qualitative improvement of images is also required (subjective sensation of sharpness and contrast).

1 INTRODUCTION

Although image acquisition techniques yield an ever improving quality, there is a need for post-processing methods to remove noise from images. The problem of suppressing noise in digital images is based on the model $y = f + n$, where f denotes the true noise-free pixel value, y the observed noisy pixels and n the observation noise. Other efficient methods are based on statistical modification of coefficients of wavelet transform via masks. The method proposed in this paper combines the wavelet multi-resolution concept with a Markov random field model.

2 THE PROPOSED METHOD

In the described method, we propose to use redundant wavelet transform to obtain same number of wavelet coefficients in each decomposition level. We use the following notation. $\boldsymbol{W}$ denotes a set of wavelet coefficients of one component in a particular level. For instance, $\boldsymbol{W} = \left\{ w_{j,l}^{HL} \middle| j = J \right\}$ is a vertical component at level J. Since the level and components type do not matter for discussion in other text, the coefficients will carry the position l as subscript.

The set of indices of the coefficients within a level is denoted by L. Hence, $\boldsymbol{W} = \{w_l | l \in L\}$. Then we initiate a set of binary labels $\boldsymbol{X} = \{x_l | l \in L\}$, which assign value label 0 or 1 to each coefficient, depending on the value information of the measure in comparison to computing threshold value based on estimate noise energy in image. This set is referred to as a mask. These labels x_l correspond to the coefficients w_l. The information of the measure that indicates how noisy the coefficients are, is similarly represented as a set $\boldsymbol{M} = \{m_l | l \in L\}$. In the proposed denoising method, the m_l values are based on the local Hőlder exponent [1] at the position l in the image. This value has the same form as wavelet coefficients and masks.

In contrast to other methods [3], the proposed method does not attempt to find a set of masks for the given noisy image, that are optimal according to a single criterion. The method is based on the following approach. For any particular masks $\boldsymbol{X}$, we can specify how probable it is, taking into account the given image and chosen measure. According to Bayes rule, the *a posteriori* probability is:

$$P(\mathbf{X}|\mathbf{M}) = P(\mathbf{M}|\mathbf{X}) \cdot P(\mathbf{X}) / P(\mathbf{M}) \tag{1.1}$$

We can say that there is not reason why *a priori* probability $P(\boldsymbol{M})$ will be considered as a uniform distribution over its domain. Therefore *a priori* there is no reason why a set $\boldsymbol{M}$ of Hőlder exponents should be more probable than another one. Then:

$$P(\mathbf{X}|\mathbf{M}) \propto P(\mathbf{M}|\mathbf{X}) \cdot P(\mathbf{X}) \tag{1.2}$$

The factor $P(\boldsymbol{M}|\boldsymbol{X})$ is the conditional probability. The second factor $P(\mathbf{X})$ is the *a priori* probability. Thus, $\boldsymbol{M}$ is considered as a multivariate observation, and $\boldsymbol{X}$ is an unknown parameter of which the *a priori* distribution is known.

Both the *a priori* and the conditional probability are modelled as a Gibbs probability function. This has an advantage that the variables x_l and m_l can be described directly with an image model using a Markov random field. The relation between Markov random field and Gibbs probability functions is expressed in the Hammersly-Clifford theorem.

THEOREM (Hammersly-Clifford): Probability function is a Markov random field with respect to a neighbourhood system if and only if it is a Gibbs distribution with respect to same neighbourhood system.

The *a priori* probability $P(\boldsymbol{X})$ expresses that a priori, when $\boldsymbol{M}$ is not taken into account, the masks are not all equally probable. This knowledge about masks can be introduced and exploited. We know that relatively few, large wavelet coefficients carry the essential information of an image with little noise. These coefficients tend to be clustered around the location of important features in the image, such as edge discontinuities, peaks, and corners. We can thus a priori expect that 0 and 1 mask labels appear in more or less separated clusters. Via the *a priori* probability we can assign a higher probability to masks that have this property. The idea is to assign a higher probability to masks in which pixels with the 0-label and pixels with the 1-label appear in separated, continuous clusters. We also assume that there is no distinction between the spatial directions, since the method should be rotation invariant, and that the model does not vary over the spatial domain. In other words, it is an isotropic and homogeneous model.

An *a priori* model of distribution wavelet of coefficients in neighbourhood should be general enough to be valid for a wide class of natural and synthesized images. Such an *a priori* model, given as a Gibbs probability function and defined on a corresponding Markov random field, is described below.

The states of the Markov random field are binary mask labels x_l in neighbourhoods 3-by-3. The *a priori* probability expresses that masks in which neighbouring state vectors have the same label value, are more probable than those with different values. Computation is based on a comparison of the central state with its neighbours. The *a priori* probability is thus:

$$P(\boldsymbol{X}) = \frac{1}{Z} \cdot \exp(-V(\boldsymbol{X})) \quad with \quad V(\boldsymbol{X}) = \sum_l V_{N_l}(\boldsymbol{X}) \tag{1.3}$$

To find an expression for the conditional probability $P(\boldsymbol{M}|\boldsymbol{X})$, we need to translate a classic binary decision based on the chosen measure (the Hőlder exponent) into a probability. For the conditional probability, there is no need for interaction between neighbours in the system. Neighbourhood are thus individual sites of the wavelet coefficients. Conditional probability is then:

$$P(\mathbf{M}|\mathbf{X}) = \prod_l P(m_l|x_l) = \exp\left(-\sum_l V(m_l|x_l)\right) \tag{1.4}$$

The individual probabilities $P(m_l|x_l)$ express that if the label $x_l = 1$ (coefficients w_l is kept unchanged), there is more probable that Hőlder exponents m_l is above a threshold T than it is below. If the label $x_l = 0$, there is more probable that Hőlder exponents m_l is below T than it is above. Since the prior distributions $P(m_l)$ are assumed to be uniform, the conditional probabilities should meet the condition $P(m_l|x_l = 0) + P(m_l|x_l = 1) = \mathrm{C}_1$, where C_1 is a constant. The probability functions are usually represented by piecewise continuous potential functions $V(m_l|x_l)$, each with two constant parts and linear transition around the threshold.

A priori and conditional probabilities are specified, therefore we can compute the *a posteriori* probability $P(\boldsymbol{M}|\boldsymbol{X})$ for every $\boldsymbol{X}$. In practice, the method actually computes the marginal probability $P(x_l = 1|\boldsymbol{M})$ for each "clean" coefficient. "Clean" coefficient is coefficient without noise. The manipulation of the coefficient is a multiplication with this probability:

$$w_l^{new} = w_l \cdot P(x_l = 1|\boldsymbol{M}) \tag{1.5}$$

The method thus involves adaptive shrinkage of wavelet coefficient with a factor, that can be different for each coefficient. The identification and modification of wavelet coefficients (both "clean" and noisy) are more conservative and less radical then in other methods based on classic binary decision (coefficient is only "clean" or affected by noise).

Equation showed how the method uses marginal probabilities $P(x_l = 1|\boldsymbol{M})$ to manipulate the coefficients. They are derived from the probability function $P(\boldsymbol{X}|\boldsymbol{M})$, in theory by calculating a weighted sum over the masks, in which the weights are the *a posteriori* probabilities. The sum for $P(x_1 = 1|\boldsymbol{M})$ incorporates the masks in which the label $x_1 = 1$ and is defined as:

$$P(x_l = 1|\boldsymbol{M}) = \sum_{\boldsymbol{X}} f_l(\boldsymbol{X}) \cdot P(\boldsymbol{X}|\boldsymbol{M}), \quad where \quad f_l = \begin{cases} 0, & if \quad x_l = 0, \\ 1, & if \quad x_l = 1. \end{cases} \tag{1.6}$$

The above formula implies that the method does not employ a binary mask, but a (marginal) averaged mask. Reliable estimations can be computed with special integrators for large probability spaces, called stochastic samplers. Stochastic samplers used in the described method for noise reduction is a classical Metropolis algorithm [2]. The aim is to compute an estimation of selected samples from space of possible masks. The Metropolis sampler is

called Markov chain – Monte Carlo method. The stochastic sampler generates a sample masks in consecutive series (Markov chains). In addition, the generation new samples is based on a random number generator, which corresponds Monte Carlo method. The sample masks are not uniformly selected, as in ordinary Monte Carlo method, but in proportion to their posterior probability $P\left(\hat{\mathbf{X}}_j \middle| \mathbf{M}\right)$. States that have a higher probability, thus have a higher probability of being selected as a sample state. Generation obtains sample masks as follow. A new candidate sample mask $\hat{\mathbf{X}}'_{j+1}$ is generated from previous sample $\hat{\mathbf{X}}_j$ in one or a few positions l, as in practice means that binary labels x_l previous sample $\hat{\mathbf{X}}_j$ are switching from 0 to 1, and vice versa. Then it depends on the probability ratio $r = P\left(\boldsymbol{X}'_{j+1}\right) / P\left(\boldsymbol{X}_j\right)$, whether the candidate state is accepted as a next sample state. The probability that the candidate state is accepted, is higher, when the probability ratio is higher. Therefore, in the stochastic samplers is not necessary to explicitly compute the probabilities $P\left(\hat{\mathbf{X}}'_{j+1} \middle| \mathbf{M}\right)$ and $P\left(\hat{\mathbf{X}}_j \middle| \mathbf{M}\right)$, which is a significant advantage comparing to ordinary Monte Carlo method. The probability ratio to decide upon candidate state acceptance can be efficiently calculated because of the Markov Random Field-Gibbs type probability [1].

In case of acceptance is the candidate state chosen as actual state mask $\hat{\mathbf{X}}_{j+1}$. This random perturbation and acceptance tests iterated in different positions up to updated all position *l* in *L*. The all estimates are updated in this point, and one iteration is complete.

3 OVERVIEW OF THE PROPOSED METHOD

1) Compute the redundant wavelet transform of the noisy image.
2) For each component $\boldsymbol{W}$ compute approximations of the local Hőlder exponent m_l for all $l \in L$.
3) Generate an initial mask by applying the threshold T to m_l and run the stochastic sampling procedure with *a posteriori* probability
4) Modify the wavelet coefficients $\boldsymbol{W}$ with the marginal probability.
5) Reconstruct the image with suppressed noise from the modified coefficients.

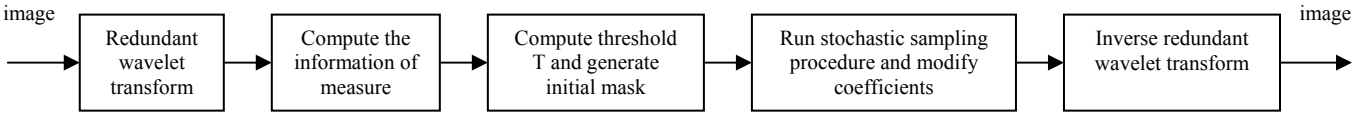

Fig. 1: *The block diagram of the proposed method*

4 SUMMARY AND CONCLUSIONS

The described method respects geometrical constraints of a processed image using *a priori* knowledge about distribution of wavelet coefficients in relation to edges of image. The method does not use classical binary decision (coefficient is only "clean" or affected by noise). The described method achieve these properties using a stochastic sampler based on Metropolis method.

In this method we use three distributions to generate random numbers (normal distribution with μ=0.5 and σ=0.25, beta distribution with parameters a=4 and b=0.75 and

uniformly distributed, see Figure 2, left panel). Iteration process of the method converges to optimum for three distributions as shown on Figure 2, right panel. In Figure 3, an original ultrasound image (left panel) and the resulted image with gain SNR=8.1 dB (right panel) are shown.

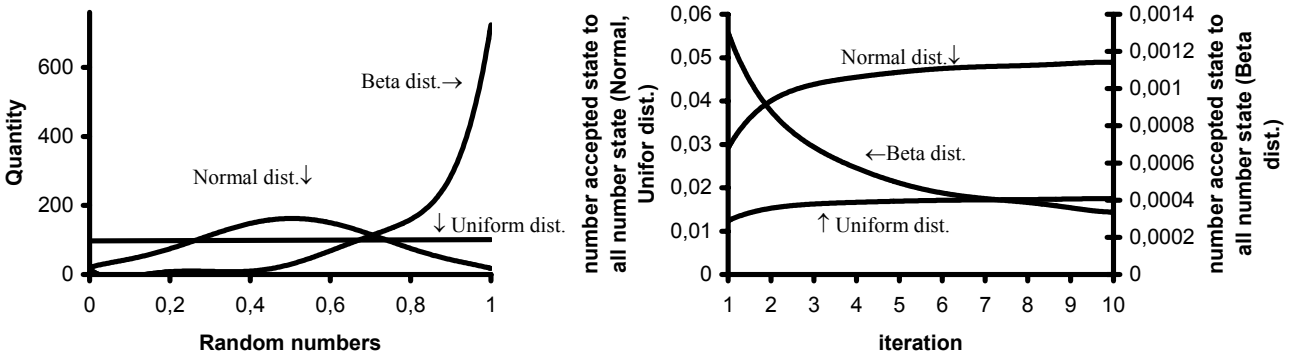

Fig. 2: *Left: Distributions used to generate random numbers. Right: Ratio of accepted states to all states depending on iteration.*

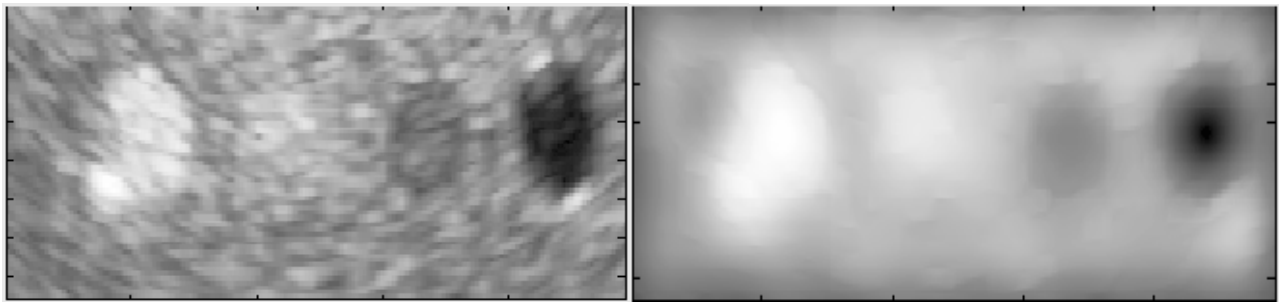

Fig. 3: *Left: Detail ultrasound (US) image of four phantoms with different absorption US wave. Right: Resulted image with gain SNR=8.1 dB for Beta distribution (see the text).*

ACKNOWLEDGEMENTS

This work has been partly supported by the Research Programme of Brno University of Technology No. J22/98:262200011.

REFERENCES

[1] Malfait, M., Roose, D.: Wavelet based image denoising II: Wavelet based image denoising using a Markov Random Field a priori model, Report TW 228, Katholieke Universiteit, Leuven, Belgium, 1995

[2] Nezbeda, I., Kolafa, J., Kotrla M.: Úvod do počítačových simulací: Metody Monte Carlo a molekulární dynamiky, Skriptum K.U., Praha 1998

[3] Aldroubi, A., Unser, M.: Wavelets in Medicine and Biology, CRC Press, 1996, Chapter 8.: Using Wavelets to Suppress Noise in Biomedical Images