FACTA UNIVERSITATIS (NIŠ)
Series: Electronics and Energetics vol. 14, No. 3, December 2001, 000-000

WAVELET-BASED MAP IMAGE DENOISING USING PROVABLY BETTER CLASS OF STOCHASTIC I.I.D. IMAGE MODELS

Andriy Synyavskyy, Sviatoslav Voloshynovskiy and Ivan Prudyus

Abstract: The paper advocates a statistical approach to image denoising based on a Maximum a Posteriori (MAP) estimation in wavelet domain. In this framework, a new class of independent identically distributed (i.i.d.) stochastic image priors is considered to obtain a simple and tractable solution in a close analytical form. The proposed prior model is considered in the form of Student distribution. The experimental results demonstrate the high fidelity of this model for approximation of the sub-band distributions of wavelet coefficients. The obtained solution is presented in the form of well-studied shrinkage functions

Key words: Wabelet, image model, image denoising, maximum posterior estimation, Student distribution, shrinkage function.

1. Introduction

Signal and images denoising algorithms which are based on maximum likelihood (ML) estimates or which do not use any prior information are known to possess low efficiency and robustness [1]. The situation can be drastically improved in the case of MAP-estimate which provides higher accuracy of estimation. Therefore, the development of accurate and tractable stochastic image models is of great importance for many applications such as image restoration, denoising, compression and segmentation.

Manuscript received December 12, 2001. A version of this paper was presented at the fifth IEEE Conference on Telecommunication in Modern Satellite, Cable and Broadcasting services, TELSIKS 2001, September 19-21, 2001, Niš, Serbia.

A. Synyavskyy and I. Prudyus are with Lviv Polytechnic National University, Institute of Telecommunications, Radio Electronic and Electronic Techniques, S. Bandery Str. 12, Lviv 79013, Ukraine, e-mail: sinat@polynet.lviv.ua. S. Voloshynovskiy is with CUI, University of Geneva, 24 General Dufour, 4 Geneva 1211, Switzerland, e-mail: svolos@cui.unige.ch.

However, the proper choice of prior image model is a notoriously difficult problem. The problem becomes even more complicated for real world image which are known to have a high variability that should be reflected by the adequate stochastic image modeling. The non-stationary behavior of image statistics is an additional factor that complicates the situation. Oppositely, the desire to use more complex image models that capture interscale and intrascale image dependencies in wavelet domain results either in iterative estimation procedures based on Expectation Maximization method or intractable solutions that are very difficult for analytical consideration.

2. Problem Formulation

We consider the problem of stochastic image modeling on the example of stochastic image denoising. We assume the classical formulation of this problem, i.e. an image O is contaminated by additive white Gaussian noise (AWGN) n with p.d.f. $N\left(0,\sigma_n^2\right)$

$$f = O + n. \tag{1}$$

The goal is to estimate the original image O based on the noisy image f. To simplify the stochastic image modeling and sequentially the estimation problem we use multiresolution image representation [2], [3]

$$f\left(x,y\right) = \sum_k c_{j_0,k}\Phi_{j_0,k}\left(x,y\right) + \sum_{i=0}^{2}\sum_{j=j_0}^{\infty}\sum_k d_{j,k}^{i}\Psi_{j,k}^{i}\left(x,y\right), \tag{2}$$

where c and d - decomposition coefficients of two-dimensional function f; Φ and Ψ^i - two-dimensional functions presenting the tensor product result of the orthogonal functions ϕ and φ with Wavelet basis [3]. Wavelet-based image representation has a number of advantages in comparison with the other domains derived from principal component analysis (PCA) framework. First, such a representation of the images allows to locally define their degree of smoothness, because orthogonal functions Ψ in the wavelet transform (2) possess specified number of vanishing moments [3]. Secondly, it nearly decorrelates images with respect to intrascale statistics. Thirdly, it has good energy compaction. Forth, it partially reflects the anisotropy of the human visual system vortex decomposition due to 3 orientations and multiresolution image representation. Fifth, it does not produce the blocking artifacts such as DCT transform which approximates PCA for the class of first-order autoregressive models.

It is already well established practice to solve the image denoising problem using shrinkage data filtering in the wavelet domain that consists of the sequence of steps:

- image decomposition using orthogonal basis of wavelet functions $d = Wf$,
- image estimation based on the non-linear processing of the decomposed coefficients $\widehat{a} = Td$;
- inverse wavelet transform of $\widehat{O} = W^{-1}\widehat{a}$.

Here W and W^{-1} denotes operators of direct and inverse wavelet transform; a are wavelet coefficients of the original image O in the given wavelet basis.

The choice of shrinkage function T is very important [2], [4] and can be derived based on the MAP estimation. Taking into account the distributive property of orthogonal transformation, the mixture (1) can be represented as a corresponding sum after the wavelet decomposition

$$d = a + n, \tag{3}$$

where $d = Wf$ is the wavelet transform of the original image f. The orthonormality of wavelet transforms preserves the statistical properties of noise component n that will have the same p.d.f. $N\left(0, \sigma_n^2\right)$ as in Eq. (3). In the above formulation, the denoising problem is reduced to the estimation of $\widehat{a}$ based on the observed image d.

3. Image Prior Model

The MAP estimators require the definition of adequate stochastic image model. In the above case, an appropriate stochastic model should be specified for wavelet domain.

The orthogonal functions ϕ and φ used in multiresolution analysis [3] are close to eigenfunctions of covariance image matrix which characterizes the short-range correlation in real-world images. Thus, the multiresolution image representation according to Eq. (2) can be considered as an approximation of Karhunen-Loeve transform [1].

In the case of stationary AR process the wavelets can provide almost complete decorrelation. Unfortunately, it is not a case for non-stationary processes such as real images. Nevertheless, the tractability of independent identically distributed (i.i.d.) models and the low complexity of the obtained solution motivate to consider this class of model more deeply.

Another important feature of the orthogonal representation (Eq. 2) consists in a specific distribution of the decomposition coefficients $a_{j,k}$. It is commonly known fact in image processing community, that the distribution of wavelet coefficients is peaked near zero, heavy-tailed and has non-Gaussian joint statistics. This feature can be explained by the presence of a significant amount of flat regions in images that contribute to the peak near zero, and comparative small number of large amplitude coefficients corresponding to edges and textures. The last ones are more significant for image simulation and image quality preservation in comparison with small amplitude coefficients.

The feature of the wavelet coefficients $a_{j,k}$ presented above are widely used for image compression and corresponding rate-distortion analysis as well as for non-linear image denoising and restoration. The most used i.i.d. stochastic models of wavelet coefficients $p(a)$ are Laplacian distribution [5], stationary Generalized Gaussian distribution (GGD) [5] (Laplacian and Gaussian distributions are two particular cases of the GGD), Gaussian mixture distribution [5] and exponential power distribution [3]. The examples of application of more complex stochastic image models are presented for Poisson process and Hidden Markov models [6]. This list is not definitively complete and can be essentially extended.

The GGD model is the most used in practice [5]

$$p(a) = \frac{\beta\eta(\beta)}{2\alpha\Gamma(1/\beta)} \exp\left\{-\left[\frac{\eta(\beta)\,|a|}{\alpha}\right]^{\beta}\right\}, \tag{4}$$

where $\eta(\beta) = \sqrt{\Gamma(3/\beta)/\Gamma(1/\beta)}$, $\Gamma(\,\cdot\,)$ is the Gamma function, α and β are scale and shape parameters.

Although, the GGD models are used for denoising applications and the close form solutions are reported for the particular shape parameters such as soft-shrinkage (Laplacian model $\beta = 1$) and hard-thresholding (asymptotically $\beta \to 0$) the theoretical analysis of Rao-Cramer bounds of this estimators remains still a challenging task. Moreover, as it will be shown in this paper, there is a class of stochastic models that provides even more accurate approximation of marginal statistics of image coefficients and simultaneously results in the close form solution of image denoising problem.

The results of our analysis of wavelet image coefficients indicate that histogram of their distribution is also accurately approximated by Cauchy distribution. Moreover, Student distribution (StD), for which Cauchy distribution is only a particular case, can even more precisely characterize the behavior of wavelet coefficients for real images. The StD law can be presented as statistical model of wavelet image coefficients in the form

$$p(a) = \frac{\Gamma\left((m+1)/2\right)}{\sigma\sqrt{m\pi}\,\Gamma\left(m/2\right)}\left(1+\frac{a^2}{\sigma^2 m}\right)^{-(m+1)/2}, \tag{5}$$

where σ is a power factor, m is a number of the freedom degrees.

We have estimated the parameters of GGD and the proposed StD (Eq. 5) statistical models of wavelet coefficients using ML estimate [1]. The estimation was accomplished iteratively solving ML system of equations with two unknown variables, in which high number of investigated samples of image allows to obtain the sufficient estimation accuracy. For example, the defined parameters of GGD are equal $\widehat{\alpha} = 5.82$ and $\widehat{\beta} = 0.72$ on the level decomposition $j = 9$ for "Lena" image, while $\widehat{m} = 2.03$ and $\widehat{\sigma} = 2.75$ for StD. The example of histogram approximation of wavelet coefficients by the GGD and StD is shown in the Fig. 1. As one can observe, that the StD provides even more accurate approximation.

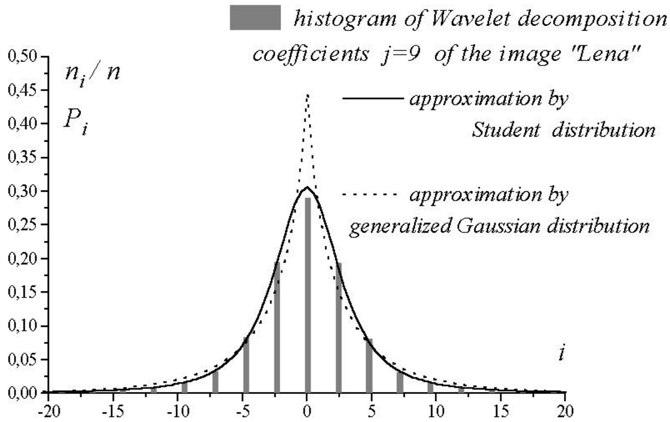

Fig. 1. The approximation of histogram of "Lena" sub-band decomposition based on GGD and StD models.

The analysis results show that the proposed StD model better describes distribution in sense of the χ^2 goodness measure [1] $\chi^2_{StD} < \chi^2_{GGD}$. The measure values are calculated using the expression $\chi^2 = n\sum_{i=1}^{R}\left((h_i - P_i)^2/P_i\right)$ for hypothetical probabilities $P_i = P\{E_i\}$ of random value appearance on i-th interval with $i = \overline{1,R}$ which can be determined accordingly with introduced assumptions about distribution law and the occurrence frequencies

$h_i = n_i/n$ that were defined on the basis of the investigated samples. The comparison performed based on Kullback-Liebler measure indicates the same behavior $D\left(h||p^{StD}\right) < D\left(h||p^{GGD}\right)$.

4. MAP Denoising

The statistical features of wavelet coefficients d in the mixture (3) were defined based on the assumption about independence and identity of image coefficients as well as their additive relation with Gaussian random component n. The conditional p.d.f. $p\left(d \mid a\right)$ is specified for unknown parameter a in the case of additive Gaussian random noise n

$$p\left(d \mid a\right) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{\left(d-a\right)^2}{2\sigma_n^2}\right). \tag{6}$$

In this case the MAP estimation problem can be presented via logarithm maximisation of a posterior probability function $p\left(a \mid d\right)$ [1]

$$\left.\frac{\partial}{\partial a} \ln\left(p\left(a \mid d\right)\right)\right|_{a=\widehat{a}_{MAP}} = 0, \tag{7}$$

where $\widehat{a}_{MAP}$ is statistical estimate of the original image. Taking into account the Bayesian rule [1] and expression (7) the MAP solution corresponds to the next problem

$$\left.\frac{\partial}{\partial a} \ln\left(p\left(d \mid a\right)\right) + \frac{\partial}{\partial a} \ln\left(p\left(a\right)\right)\right|_{a=\widehat{a}_{MAP}} = 0. \tag{8}$$

A solution of this equation is obtained using the above expressions for the likelihood function $p\left(d \mid a\right)$ and a priori function $p\left(a\right)$ according to the introduced prior model. The solution of Eq. (8) makes possible to determine the non-linear function T of thresholding denoising algorithm. The substitution of the expressions (4) and (6) in the MAP equation (8) allows to rewrite it in the next form

$$\left.d - a - \sigma_n^2 \beta \left[\frac{\eta\left(\beta\right)}{\alpha}\right]^{\beta} |a|^{\beta-1} \operatorname{sgn}\left(a\right)\right|_{a=\widehat{a}_{MAP}} = 0, \tag{9}$$

where

$$\operatorname{sgn}(a) = \begin{cases} 1, & \text{for } a \geq 0 \\ -1, & \text{otherwise} \end{cases}$$

is the signum function.

In the general case, the solution of equation (9) does not exist in analytical form. In image denoising problems, the partial solutions are obtained for $\beta = 1$ and 0 (soft-shinkage and hard-thresholding [2], [5] or the approximate solutions (various types of the shrinkage functions [5] and semi-soft thersholding). An arguments of the critical points can be defined by relationship of the parameters α, β and σ_n.

The series of curves (Fig. 2.a), which corresponds to the nonlinear functions for the MAP-estimation of parameter a based on the GGD prior model is obtained via numerical determination of roots of the equation (9).

The MAP-estimator based on the proposed StD prior image models was derived in this paper. After substitution of the corresponding stochastic models, we have obtained an analytical solution in the form of the cubic equation with respect to unknown variable a

$$\left(m\sigma^2 + a^2\right)(d-a) - (m+1)\, a\sigma_n^2\Big|_{a=\widehat{a}_{MAP}} = 0. \tag{10}$$

It is well-known that the cubic equation has the three roots. In the case of the MAP equation (10), each of these solutions were found in the analytical form.

Since the estimation is performed on a set of real functions and the estimation result belongs to the real set, the corresponding solution of the equation (10) is also presented analytically in the form of a conditional equality:

$$\widehat{a}_{MAP} = \begin{cases} \dfrac{d}{3} - \dfrac{3A-d^2}{9C} + C, & \mathrm{Im}\left\langle C \right\rangle = 0 \\ \dfrac{d}{3} + \dfrac{3A-d^2}{18C} - \dfrac{C}{2} + \dfrac{j\sqrt{3}}{2}\left(C + \dfrac{3A-d^2}{9C}\right), & \text{otherwise} \end{cases} \tag{11}$$

where $A = B + (m+1)\,\sigma_n^2$, $B = m\sigma^2$ and

$$C = \frac{1}{6}\Big\{4\Big[d(-9A + 27B + 2d^2) + 3\sqrt{12(A^3 + Bd^4) - 3d^2((A+9B)^2 - 108B^2)}\Big]\Big\}^{\frac{1}{3}}.$$

The series of curves corresponding to non-linear processing function T are shown in Fig. 2(b) for proposed prior model (Eq. 5) of image.

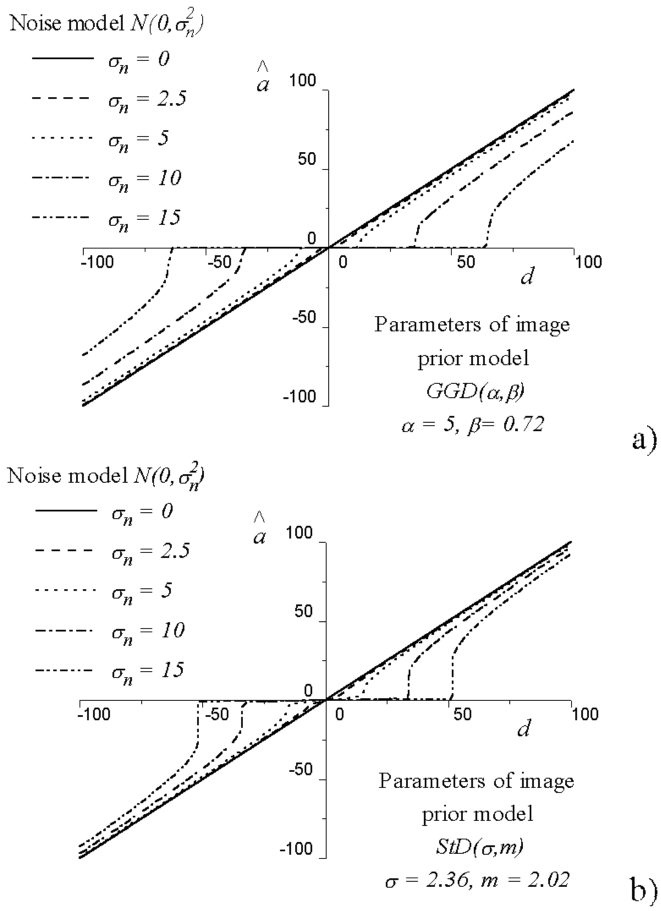

Fig. 2. Solution of MAP denoising for the GGD (a) and proposed StD (b) prior models of images.

5. Results

The proposed analytical expression (Eq. 11) of MAP denoising for prior model (Eq. 5) makes possible to simplify the estimation procedure in comparison with the MAP estimator with the GGD prior model (Eq. 4). Due to analytical form of non-linear processing function (Eq. 11) such approach allows to change softly the estimator parameters according to the parameters of used prior model and noise variance.

The efficiency of the proposed method for Gaussian noise removal from images is numerically evaluated for the series of test images. We present here results for test image "Lena". The results of AWGN removal are shown in Fig. 3 for known and proposed methods. The mean square error (MSE) was chosen for objective comparison of denoiser performance.

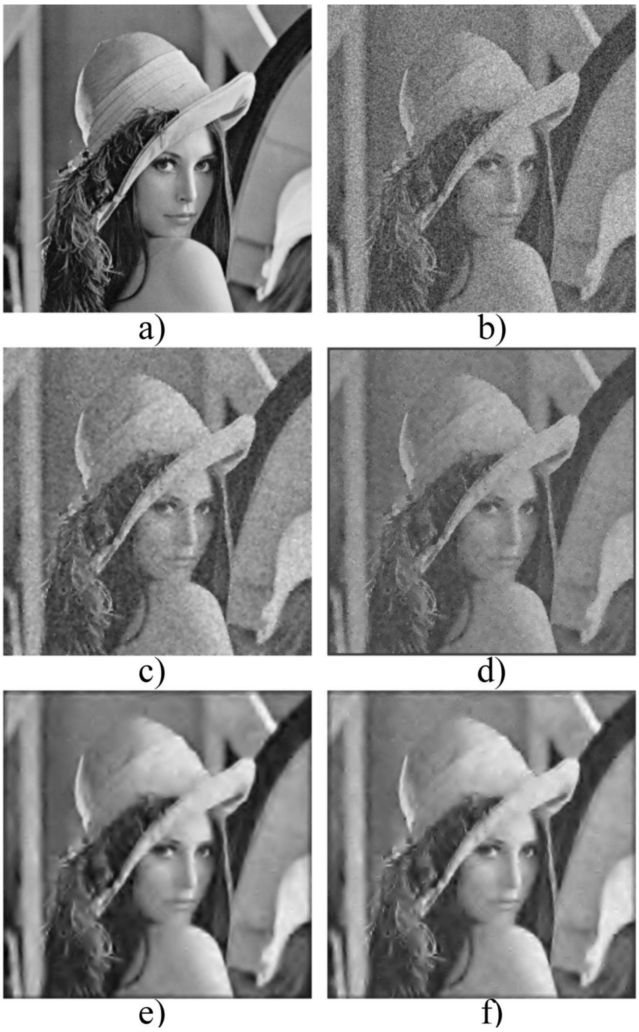

Fig. 3. The denoising methods comparison of the test image (256×256) *-a) at the presence of additive gaussian noise component* $N\left(0, \sigma_n^2\right)$ *with* $\sigma_n = 35$ *-b). The processing results of Lee's filter* $[3 \times 3]$ *-c); Sigma-filter* $[5 \times 5]$ *-d); multiresolution based MAP estimation by the known -e) and proposed -f) methods.*

The corresponding results are shown in Table 1. Sigma-filter, locally adaptive Lee's filter that corresponds to the MAP estimators for non-stationary Gaussian image prior model, as well as MAP denoisers for the GGD family and the proposed method (Eq. 11) were compared.

The results of denoising shown in Table 1 demonstrate the superior performance of the proposed denoising method.

Table 1. MSE comparison of different denoisers for "Lena" image.

Denoiser	The image distortions values MSE	
Image "Lena" without processing	($\sigma_n = 25$) 586.13	($\sigma_n = 35$) 1158.8
Lee Filter	145.21	258.82
Sigma-filter	199.06	346.74
MAP method at GGD model	130.02	184.08
MAP method at StD model	122.18	174.18

6. Conclusion

We have proposed a new class of stochastic image models within the group of i.i.d. models based on Student distribution to model the sub-band marginal statistics of wavelet image coefficients. The proposed model more accurately approximates the joint p.d.f. of intrascale coefficients in comparison with the Generalized Gaussian model. To demonstrate the advantages of the proposed model for real applications we have considered stochastic denoising problem based on the MAP estimate. In contrast to the GGD group of model we have received comparatively simple close from shrinkage solution.

We will concentrate on theoretical investigation of accuracy of the obtained estimate and compare Rao-Cramer bounds for the AWGN scenario for the GGD and Student priors in our future research.

REFERENCES

1. F. Abramovich, T. Spatinas, B.W. Silverman: *Wavelet thresholding via a Bayesian approach.* J. Roy Statist. Soc. Ser., B60, 1998, pp. 725-749.
2. A. Antoniadis, J. Bigot, T. Sapatinas: *Wavelet Estimators in Nonparametric Regression: Description and Simulative Comparison.* Technical Report Laboratoire IMAG-LMC, University Joseph Fourier, France, 78p., 2001.
3. M.S. Crouse, R.D. Nowak, R.G. Baraniuk: *Wavelet-based statistical signal processing using hidden Markov models.* IEEE Trans. Signal Proc., vol. 46, 1998, pp. 886-902.

4. D.L. Donoho: *Unconditional bases are optimal bases for data compression and for statistical estimation.* Appl. Comput. Harmon. Anal, vol.1(1), 1993, pp. 100-115,

5. S.G. Mallat: A theory for multiresolution signal decomposition: *The wavelet representation.* IEEE Trans. Patt. Anal. Mach. Intell, vol. 11, No. 7, 1989, pp. 674-693

6. H.L. Van Trees: *Detection, Estimation and Modulation Theory, Part I. - Detection.* Estimation and Linear modulation theory, Wiley, New York, 1971.