ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 16, No. 2, 2021, *pp*.098-107

Adaptive parameter selection for preserving edges based on EPLL

Ze, Qin[1, *], Xiulan Sheng[1]
[1] College of education, Jiangsu Open University, Nanjing 210044, China
Email: qinzjsou@163.com
(Received June 21, 2021, accepted September 30, 2021)

Abstract: Though image denoising has experienced rapid development, there remain problems to be solved such as preserving the edge and meaningful details in image denoising. In this paper, we focus on this hot issue. Considering the parameter in original method is a constant, we introduce a new adaptive parameter selection based on EPLL (Expected Patch Log Likelihood) by the use of image gradient and the local variance, which varies with different regions of the image. What's more, for solving staircase effect which common in anisotropic diffusion models, we add a gradient fidelity term to release it. The experiment shows that our proposed method proves the effectiveness not only in vision but also on quantitative evaluation.
Keywords: image denoising; adaptive parameter; Expected Patch Log Likelihood; edges

1. Introduction

Images have been a more and more important carrier of plenty information, which can deliver messages intuitively. Unfortunately, images inescapably suffer from noise due to various reasons, such as motion blur and the precision in measurements of sensors. Noise removal has always been a hot issue in the past few decade years, the research that follows is also growing rapidly.

A large number of denoising methods have been proposed during hot discussion. Much work has begun on sparse representation, which regards image patches as the linear combination of some atoms based on a dictionary [1]. After this, low-rank approximation methods also achieve good results [2], [3]. The total variation (TV) has always been a hot topic [4], [5]. Establishing an appropriate prior model directly influences the outcome. So mixture models have raised much concern in image restoration due to its robustness, especially Gaussian Mixture Model [6], [7], [8]. Inspired by this, some further studies have been proposed.

When it comes to noise removal, one sharp problem is always unavoidable. How to preserve as much structural information as possible while removing noise. So preserving edges of the image has become a thorny problem among academic and industry communities. Therefore, a lot of work revolves it. Perona and Malik firstly put forward pioneering model of anisotropic diffusion [9]. This idea was immediately spread due to its powerful effect [10], [11], [12]. Of course, this great approach also has its weakness. The number of iterations has giant influence on results. It always leads to staircase effect after excessive iterations. Inspired by this, Tebini proposed a fast and efficient to speed up the convergence of algorithm [13]. In this paper, we focus on the edge-preserving, and propose a new method of adaptive regularization parameter selection. Moreover, we add a gradient-fidelity term to relieve the staircase effect and preserve more details of image.

2. Proposed method

2.1 Original EPLL Model

Given a noisy image corrupted by noise which can be described as $u_0 = u + v$ Where u represents clean image and v represents noise. We aim at separating the pure image from the noisy one. In some ways one image can be seen as a type of high dimensional data, so making clear of prior knowledge becomes a principal problem. Gaussian mixture model is considered as one of the ideal models to describe statistical characteristics of the gray image. In this paper, the Gaussian mixture model is trained by a set of clean image patches $D = \{a_1, a_2, \cdots, a_N\}$. For each patch a_i, the distribution can be described as the following:

$$p(a_i) = \sum_{i=1}^{K} \pi_k N(a_i \mid \mu_k, \Sigma_k) \tag{1}$$

where

$$N(a_i \mid \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{\frac{d}{2}} \Sigma_k^{\frac{1}{2}}} \exp\left(-\frac{(a_i - \mu_k)\Sigma^{-1}(a_i - \mu_k)}{2} \right) \tag{2}$$

Where π_k is the prior probability for *k-th* component, K is the number of mixing components, μ_k and Σ_k denote the corresponding mean vector and covariance matrix. To simplify notations, using $\Theta = \{\pi_k, \mu_k, \Sigma_k\}$ denote all these parameters. Expectation Maximization algorithm has proven its effect in estimating parameters in various models.

In the E step, estimate the posterior probability for each component

$$p(k \mid a_i, \Theta) = \frac{\pi_k N(a_i \mid \mu_k, \Sigma_k)}{\sum_{k=1}^{K} \pi_k N(a_i \mid \mu_k, \Sigma_k)} \tag{3}$$

Then, we have the likelihood function for all patches as below:

$$LL(D) = \sum_{i=1}^{N} \ln\left(\sum_{i=1}^{K} \pi_k N(a_i \mid \mu_k, \Sigma_k) \right) \tag{4}$$

In the M-step, let so as to estimate parameters, solving it then we get:

$$\mu_k = \frac{\sum_{i=1}^{N} p(k \mid a_i, \Theta) \cdot a_i}{\sum_{i=1}^{N} p(k \mid a_i, \Theta)} \tag{5}$$

Equation (5) notices that the mean vector is calculated by weighted average. Moreover, the weight represents the posterior probability.

In the same way, let $\frac{\partial LL(D)}{\partial \Sigma_K} = 0$, then:

$$\Sigma_k = \frac{\sum_{i=1}^{N} p(k \mid a_i, \Theta) \cdot a_i}{\sum_{i=1}^{N} p(k \mid a_i, \Theta)} \tag{6}$$

Considering mixing weights π_k, using Lagrange multiplier method, then get the function:

$$LL(D) + \gamma\left(\sum_{k=1}^{K} \pi_k - 1\right) \tag{7}$$

Subject to $\pi_k \geq 0$, $\sum_{k=1}^{K} \pi_k = 1$

where γ is the Lagrange multiplier.

Let the function (7) equal to 0, then take the derivative of π_k:

$$\pi_k = \frac{1}{N} \sum_{i=1}^{N} p(k \mid a_i, \Theta) \tag{8}$$

Equation (11) reveals that the averaging posterior probability determines each mixing weight.

Repeating above iterative steps until the algorithm converges.

2.2 Original EPLL Model

As is mentioned above, there is difficulty in directly building a model based on a whole image due to image's high dimensionality. For solving this, we can learn the distribution of image patches which extracted from the whole original image. EPLL (Expected Patch Log Likelihood) is such a denoising method which using image's statistical property based on Gaussian Mixture Model [14]. The central theme of this model is that every patch will have the highest log likelihood probability under our prior. So the cost function of this method is defined as:

$$f = \frac{\lambda}{2}\|u-u_0\|^2 - \sum_i \log p(R_i u) \tag{9}$$

where the first term is called data fidelity term and the second one is called regularization term. λ represents the regularization parameter and R_i is an operator that extracts patches from the original image, obviously $R_i u$ denotes the *i-th* patch we need. We must point out the second term in equation (9) denotes cumulative log probabilities of all patches that overlapped, rather than the whole image. In other word, each pixel will appear more than one time, so it will be average according to its frequency

2.3 Proposed method with adaptive regularization

When it comes to denoising, EPLL is well behaved without a doubt, but still not perfect. Because the parameter λ is a constant, that means the intensity of noise removal is uniform on the whole image. It's obvious that edge areas of the image will be over-smoothed while parameter λ is too large. Instead, if we set the value of λ too small, noise removal can hardly be thorough enough. A sharp question comes out that the balance between edge preserving and denoising is hard to maintain. For solving this, a new denoising method with adaptive regularization parameter based on image gradient is proposed as the following:

$$\frac{\lambda(x,y)}{2}\|u-u_0\|^2 + \frac{\alpha(x,y)}{2}\|\nabla u - \nabla(G_\sigma * u_0)\|^2 - \sum_i \log p(R_i u) \tag{10}$$

where the second term is a gradient fidelity term, it guarantees that the restored image and the noisy image have a similar structure, and it can effectively alleviate the problem of gradient effect. This paper [15] pointed out that the optimal solution of the new model is the same as that of the original diffusion model with this gradient fidelity term. $\lambda(x,y)$ and $\alpha(x,y)$ are adaptive parameters. $\lambda(x,y)$ devotes the value at each pixel which relied on image's gradient. $\lambda(x,y)$ can be expressed as the following:

$$\lambda(x,y) = k_1\left[1+\frac{2}{\exp\left(2\nabla u(x,y)\cdot M_N(x,y)/k_2\right)}\right] \tag{11}$$

where k_1 is a constant , k_2 is also a constant as the threshold value, $\nabla u(x,y)$ is the gradient of pixel(x,y), and $M_N(x,y)$ devotes the standardized variance at the pixel. Here, the variance of the image is introduced. In order to keep the variance occupy the same weight as the gradient, normalize the variance by:

$$M_N = 1 + \frac{\sigma^2(x,y) - \min\sigma^2}{\max\sigma^2 - \min\sigma^2} \cdot 254 \tag{12}$$

Traditional gradient fidelity term adaptively gets larger in homogeneous areas, leading to over-convergence. There still remains residual noise in gray gradient areas. So fidelity term's coefficient can be reduced appropriately in homogeneous areas to avoid over-smoothing. Then $\alpha(x,y)$ can be described as the following:

$$\alpha(x,y) = \frac{1}{1 + \left(\nabla\left(G_\sigma * u_0(x,y)\right)/k_3\right)^{\eta(\nabla u(x,y))}} \tag{13}$$

where k_3 is a threshold value, G_σ is a Gaussian filter operator, $\eta(\nabla u(x,y))$ can be described as the following:

$$\eta(\nabla u(x,y)) = 2 + \frac{2}{1 + 0.5\left|\nabla u(x,y)\right|^2} \tag{14}$$

The gradient ∇u and the variance can be seen as indications of edges. Hence, the regularization parameter function mainly related to them. $\lambda(x,y)$ is small while large gradient and variance are detected at edges of the image, and $\alpha(x,y)$ is large so edges will be preserved. Instead, if small gradient and variance are detected in smooth regions, $\lambda(x,y)$ is small so as to keep the strength of removing noise. Meanwhile, $\alpha(x,y)$ is small, which aims to maintain as much similar structure as possible between the restored image and the observation..

We introduce a method named Half Quadratic Splitting algorithm to solve equation (10), rather than solving it directly. A series of patches $\{z_i\}$ are introduced that each one refers to the corresponding overlapping patch $R_i u$, then equation (10) turns into:

$$\frac{\lambda(x,y)}{2}\left\|u - u_0\right\|^2 + \frac{\alpha(x,y)}{2}\left\|\nabla u - \nabla\left(G_\sigma * u_0\right)\right\| - \sum_i \left(\frac{\beta}{2}\left\|R_i u - z_i\right\|^2 - \log p(z_i)\right) \tag{15}$$

Where β is the penalty parameter, it is obvious that auxiliary variable $\{z_i\}$ equals to the patch $R_i u$ when β tends to infinite.

For minimizing equation (15), firstly, we choose the Gaussian component that has the highest conditional mixing weight $k_{\max}$ for each patch, then alternately optimize equation(15) by updating z_i and u alternately:

$$z_i^{n+1} = \left(\Sigma_{j\max} + \frac{1}{\beta} I \right)^{-1} \cdot \left(R_i u^n \Sigma_{j\max} + \frac{1}{\beta} u_{j\max} I \right) \tag{16}$$

$$\begin{aligned} u^{n+1} = u^n + \Delta t[\lambda(x,y)\left(u_0 - u^n\right) - \sum_i \beta R_i^T \left(R_i u^n - z_i^n\right) \\ + \alpha(x,y)\left(u_{xx}^n + u_{yy}^n\right) - \alpha(x,y)\left(G_\sigma \cdot \left(u_{0xx} + u_{0yy}\right)\right)] \end{aligned} \tag{17}$$

where Δt denotes time step and I is the unit matrix.

Generally speaking, the proposed method can be implemented as the following:

Table 1: Proposed algorithm

Input:	corrupted image u_0, penalty parameter β, the time step Δt, regularization parameter functions $\lambda(x,y)$, $\alpha(x,y)$
Step	Choose the most likely Gaussian mixing weights for each patch $R_i u$
	Calculate z_i^{n+1} using equation (16);
	Pre-estimate image u^{n+1} by equation(17);
	Repeat above steps for 4-5 times.

3. Implementation and experiment results

In this paper, we train the GMM with 200 mixture components by a set of 2×10^6 image patches that sampled from the Berkeley Segmentation Database (BSDS300). Parameters mentioned above are set as the following: the image patch $\sqrt{L} = 8$, the penalty parameter $\beta = 1/\sigma^2 * [1,2,3,4,5]$, time step $\Delta t = 0.004$, constants are set as the following: $k_1 = 0.00025 * 1/\sigma^2$, $k_2 = 30$, $k_3 = 80$. Our methods compared with the original model are as follows:

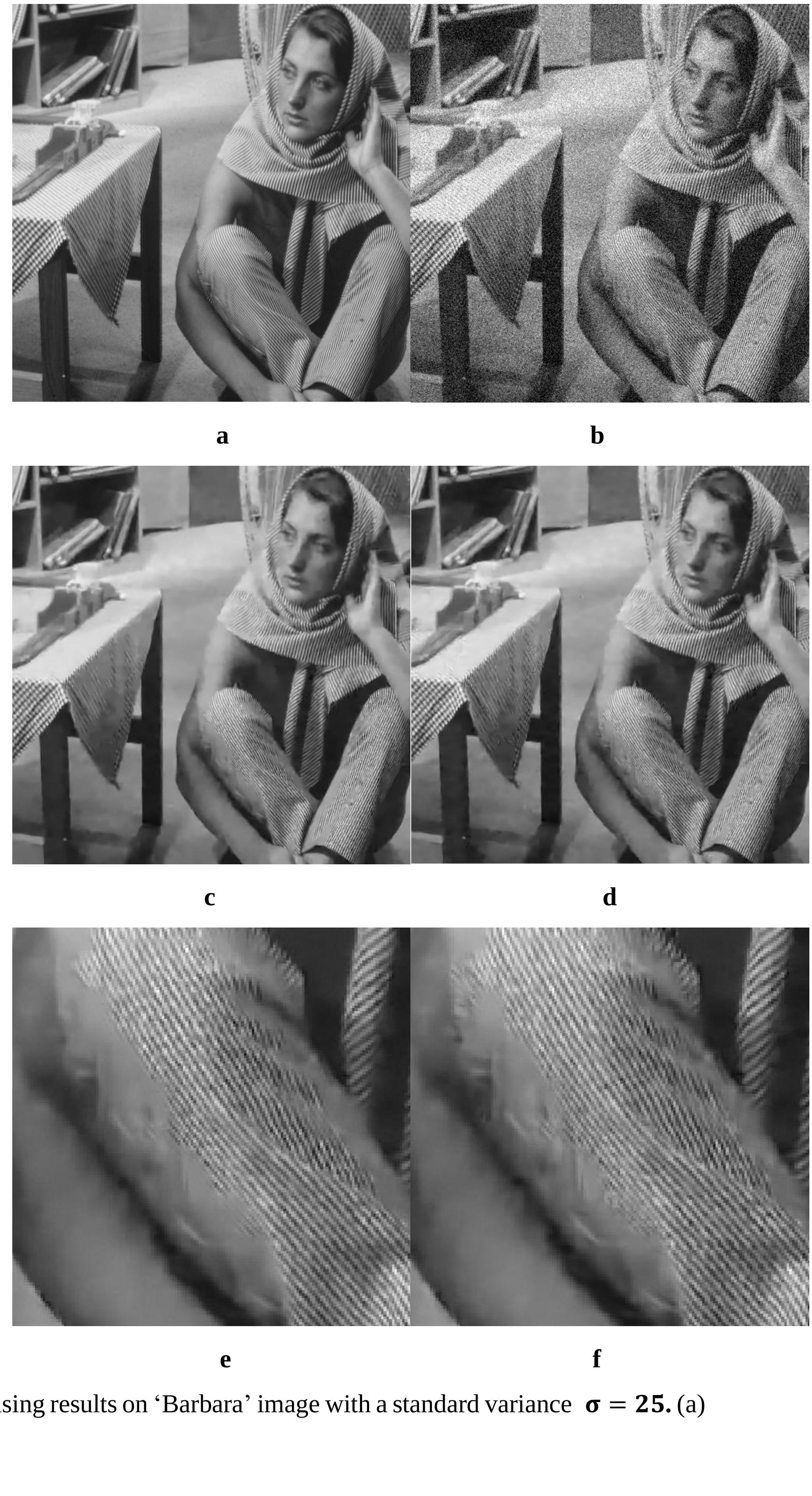

Figure1: Denoising results on ‘Barbara’ image with a standard variance $\boldsymbol{\sigma = 25.}$ (a) Original image

(b) Noisy image, (c) EPLL result (d), Proposed method, (e) EPLL enlarged partial result, (f) Proposed method enlarged partial result

a **b**

c **d**

e **f**

Figure2: Denoising results on 'lena' image with a standard variance $\sigma = 50$. (a) Original image (b) Noisy image, (c) EPLL result (d), Proposed method, (e) EPLL enlarged partial result, (f) Proposed method enlarged partial result

Table 2: PSNR results for test images with noise standard deviation σ=25

	EPLL	Proposed method
Barbara	28.55	28.72
House	32.15	32.41
Man	29.57	29.71
Lena	31.52	31.7
boat	29.62	29.78

As in shown in figure1 and figure2, while the original EPLL method may denoise well, but it also remains a problem that in some regions edge and structure information disappear. Instead, edges and details of the image are better preserved in our proposed method. Both visually and numerically, we can make the conclusion that obviously our method is superior to the original EPLL method. This is probably due to the fact that our method selects a proper regularization parameter and adds a gradient fidelity term.

4. Conclusion

Many denoising methods based on statistics have achieved huge success recent years, and EPLL is one of them. It performs rather well on noise removal, but still lack in preserving edges. Considering this, we select an adaptive parameter related to the image gradient, which varies with different regions of the image. In smooth regions parameter is small so that noise will be removed strongly. And at edges of the image, the parameter is set large so as to preserve edges well. Our proposed method is well behaved in restoration and shows an obvious advance compared with the original method.

References:

[1] Liu H, Zhang J, Xiong R. CAS: Correlation Adaptive Sparse Modeling for Image Denoising[J]. IEEE Transactions on Computational Imaging, 2021, PP (99):1-1.

[2] Dong W; Shi G.; Li X: Nonlocal image restoration with bilateral variance estimation: a low-rank approach[J]. IEEE Trans. Image Process, 2013, vol. 22, no. 2, pp. 700-711.

[3] Dong W, Shi G, Li X, Ma Y, Huang F: Compressive sensing via nonlocal low-rank regularization. IEEE Transactions On Image Processing[J], 2014, vol. 23, no. 8, pp. 3618-3632.

[4] Chatterjee S, Goswami S. New Risk Bounds for 2D Total Variation Denoising[J]. IEEE Transactions on Information Theory, 2021, PP (99):1-1.

[5] Yang N, Wu C, Qu H. Study on mixed noise removal by the total variation method[J]. Journal of Xi'an University of Posts and Telecommunications, 2013.

[6] Wang Y; Morel J. Sure guided Gaussian mixture image denoising. Siam Journal on Imaging Sciences[J], 2012, vol. 6, no. 2, pp. 999-1034.

[7] Yu G, Sapiro G, Mallat S. Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity. IEEE Trans Image Process[J], 2012, vol. 21, no. 5, pp. 2481-2499.

[8] Yu G, Sapiro G, Mallat S. Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity. IEEE Trans Image Process[J], 2012, vol. 21, no. 5, pp. 2481-2499.

[9] Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis & Machine Intelligence[J], 2012, vol. 12, no. 7, pp. 629-639.

[10] Chao S, Tsai, D. An improved anisotropic diffusion model for detail and edge-preserving smoothing. Pattern Recognition Letters[J], 2012, vol. 31, no. 13, pp. 2012-2023.

[11] Ma X, Shen H, Zhang L, Yang J, Zhang H. Adaptive Anisotropic Diffusion Method for Polarimetric SAR Speckle Filtering. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing[J], 2017, vol. 8, no. 3, pp. 1041-1050.

[12] Tebini S, Mbarki Z. Seddik H. Braiek E. Rapid and efficient image restoration technique based on new adaptive anisotropic diffusion function. Digital Signal Processing[J], 2016, vol. 48, no. C, pp. 201-215.

[13] Wang Y, Guo J, Chen W, Zhang W. Image denoising using modified Perona-Malik model based on directional Laplacian. Signal Processing[J], 2013, vol. 93, no. 9, pp. 2548-2558.

[14] Zoran D, Weiss Y. From learning models of natural image patches to whole image restoration[C], IEEE International Conference on Computer Vision (ICCV), IEEE, 2011: 479-486

[15] Zhu L, Pheng A, Xia D. Nonlinear Diffusion based Image Denoising Coupling Gradient Fidelity Term. Journal of Computer Research and Development[J], 2007, vol. 44, no.8: 1390-1398