Texture Enhanced Image Denoising via Gradient Histogram Preservation

Wangmeng Zuo[1,2] Lei Zhang[2] Chunwei Song[1] David Zhang[2]
[1]Harbin Institute of Technology, [2]The Hong Kong Polytechnic University
{cswmzuo, wolvesandme}@gmail.com, {cslzhang, csdzhang}@comp.polyu.edu.hk

Abstract

Image denoising is a classical yet fundamental problem in low level vision, as well as an ideal test bed to evaluate various statistical image modeling methods. One of the most challenging problems in image denoising is how to preserve the fine scale texture structures while removing noise. Various natural image priors, such as gradient based prior, nonlocal self-similarity prior, and sparsity prior, have been extensively exploited for noise removal. The denoising algorithms based on these priors, however, tend to smooth the detailed image textures, degrading the image visual quality. To address this problem, in this paper we propose a texture enhanced image denoising (TEID) method by enforcing the gradient distribution of the denoised image to be close to the estimated gradient distribution of the original image. A novel gradient histogram preservation (GHP) algorithm is developed to enhance the texture structures while removing noise. Our experimental results demonstrate that the proposed GHP based TEID can well preserve the texture features of the denoised images, making them look more natural.

1. Introduction

The goal of image denoising is to estimate the latent clean image **x** from its noisy observation **y**. One commonly used observation model is $\mathbf{y} = \mathbf{x} + \mathbf{v}$, where **v** is additive white Gaussian noise. Image denoising is a classical yet still active topic in image processing and low level vision, while it is an ideal test bed to evaluate various statistical image modeling methods. In general, we hope that the denoised image should look like a *natural* image, and therefore the statistical modeling of natural image priors is crucial to the success of image denoising.

Based on the fact that natural image gradients exhibit heavy-tailed distributions, gradient-based priors are widely used in image denoising [10, 17, 18]. The well-known total variation minimization methods actually assume Laplacian distribution of image gradients [25]. By observing that natural images can be sparsely coded over a redundant dictionary, the sparsity prior has proved to be effective in image denoising via l_0-norm or l_1-norm minimization [8, 9]. Another popular prior is the nonlocal self-similarity (NSS) prior [2, 16]; that is, in natural images there are often many similar patches (i.e., nonlocal neighbors) to a given patch, which may be spatially far from it. The joint use of sparsity prior and NSS prior has led to state-of-the-art image denoising results [7, 21]. However, the many denoising algorithms based on the above priors can still fail to preserve the image fine scale texture structures, which have certain overlap with noise in the frequency domain. The over-smoothing of those detailed texture structures makes the denoised image look less natural, degrading much the visual quality (please refer to Fig. 1 for example).

With the rapid development of digital imaging technology, the resolution of imaging sensor is getting higher and higher. On one hand, more fine texture features of the object and scene will be captured; on the other hand, the captured high resolution image is more prone to noise because the smaller size of each pixel makes the exposure less sufficient. However, suppressing noise while preserving textures is difficult to achieve simultaneously, and this has been one of the most challenging problems in natural image denoising. Unlike large scale edges, the fine scale textures have much higher randomness in local structure and they are hard to characterize by using a local model. Considering the fact that texture regions in an image are homogeneous and are usually composed of similar patterns, statistical descriptors such as histogram are more effective to represent them. Actually, in literature of texture representation and classification [13, 27, 28], global histogram of some local features is dominantly used as the final feature descriptor for matching. Meanwhile, image gradients convey most of semantic information in an image and are crucial to the human perception of image visual quality. All these motivate us to use the histogram of image gradient to design new image denoising models.

With the above consideration, in this paper we propose a novel method for texture enhanced image denoising (TEID) via gradient histogram preservation (GHP). From the given noisy image **y**, we will estimate the gradient histogram of

1063-6919/13 $26.00 © 2013 IEEE
DOI 10.1109/CVPR.2013.159

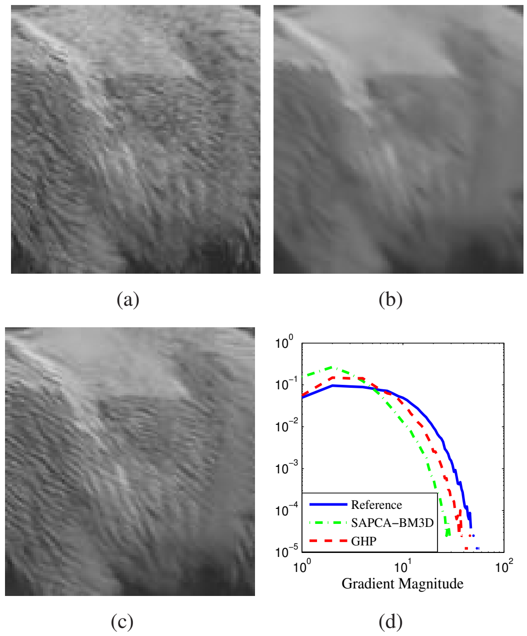


Figure 1. Denoised images and their gradient histograms. (a) A cropped image with hair textures; (b) denoised image by the SAPCA-BM3D method [16]; (c) denoised image by the proposed texture enhanced image denoising via gradient histogram preservation (GHP); (d) the gradient histograms of the denoised images. We can see that the proposed GHP method leads to better texture preservation and visual perception, and the gradient histogram of the denoised image by GHP is also closer to the reference gradient histogram estimated from the noisy image.

original image $\mathbf{x}$. Take this estimated histogram, denoted by $\mathbf{h}_r$, as a reference, we search for an estimate of $\mathbf{x}$ with GHP, i.e., the gradient histogram of the denoised image should be close to $\mathbf{h}_r$. As shown in Fig. 1, the proposed TEID method can well enhance the image texture regions, which are often over-smoothed by other denoising methods. The major contributions of this paper are as follows:

(1) A novel image denoising framework, i.e., TEID, is proposed, which preserves the gradient distribution of the original image. The existing image priors can be easily incorporated into the proposed framework to improve the quality of denoised image.

(2) A histogram specification operator is developed to ensure the gradient histogram of denoised image being close to the reference histogram, resulting in a simple yet effective GHP based TEID algorithm.

(3) A simple but theoretically solid algorithm is presented to estimate the gradient histogram from the given noisy image, making TEID practical to implement.

2. Related work

Generally, image denoising methods can be grouped in two categories: model-based methods and learning-based methods. Most denoising methods reconstruct the clean image by exploiting some image and noise prior models, and they belong to the first category. Learning-based methods attempt to learn a mapping function from the noisy image to the clean image [26], and have been receiving considerable research interests [3]. Numerous image denoising algorithms have been proposed, and here we only review those model-based denoising methods related to our work from a viewpoint of natural image priors.

Studies on natural image priors aim to find suitable models to describe the characteristics or statistics (e.g., distribution) of images in some transformed domain. One representative class of image priors is the gradient priors based on the observation that natural images generally have a heavy-tailed distribution of gradients. The use of gradient prior can be traced back to 1990s, when Rudin et al. [25] proposed a total variation (TV) model for image denoising, where the gradients are actually modeled by Laplacian distribution. Another well-known prior model, the mixture of Gaussians (GMM), can also be used to approximate the distribution of gradient magnitude [10, 19]. In addition, the hyper-Laplacian model can more accurately model the heavy-tailed distribution of gradients, and has been widely applied to various image restoration tasks [4, 5, 15, 17, 18].

The image gradient prior is basically a kind of sparsity prior, i.e., the gradient distribution is sparse. More generally, the sparsity prior has been well applied to filter responses, wavelet/curvelet transform coefficients, or the coding coefficients over a redundant dictionary. In [23, 29], Gaussian scale mixtures are used to characterize the margin and joint distributions of wavelet transform coefficients. In [24, 31], the Student t-distributions are used for both learning basis filters and modeling filter responses. By assuming that an image patch can be represented as a sparse linear combination of the atoms in an over-complete dictionary, a number of dictionary learning (DL) methods (e.g., K-SVD [9], task driven DL [20], and ASDS [8]) have been proposed and applied to image denoising and other restoration tasks.

Based on the fact that a similar patch to the given patch may not be spatially close to it, another line of research is to model the similarity between image patches, i.e., the image nonlocal self-similarity (NSS) priors. The seminal work of nonlocal means denoising in [2] has motivated a wide range of studies on NSS, and has led to a flurry of NSS based state-of-the-art denoising methods, e.g., BM3D [16], LSSC [21], and EPLL [32], etc.

Different image priors characterize different and complementary aspects of natural image statistics, and thus it is possible to combine multiple priors to improve the denoising performance. For example, Dong et al. [7] unified both image local sparsity and nonlocal similarity priors via clustering-based sparse representation. Recently, Jancsary et al. [14] proposed a method called regression tree fields

(RTF) to integrate different priors.

However, many existing image denoising algorithms, including those sparsity and NSS priors based ones, tend to wipe out the image detailed textures while removing noise. As we discussed in the Introduction section, considering the randomness and homogeneousness of image texture regions, we propose to use the histogram of gradient to describe the image texture and design new image denoising algorithm with gradient histogram preservation. In [4, 5], Cho et al. used hyper-Laplacian to model gradient, and proposed a content-aware prior for image deblurring by setting different shape parameters of gradient distribution in different image regions. By matching the gradient distribution prior, Cho et al. found that the deblurred images can have more detailed textures as well as better visual quality. However, in [4, 5] the estimation of desired gradient distribution is rather heuristic, and the gradient histogram matching algorithm is very complex.

3. Denoising with gradient histogram preservation (GHP)

In this section, we first present the image denoising model by gradient histogram preservation with sparse nonlocal regularization, and then present an effective histogram specification algorithm to solve the proposed model for texture enhanced image denoising.

3.1. The denoising model

Given a clean image $\mathbf{x}$, the noisy observation $\mathbf{y}$ of $\mathbf{x}$ is usually modeled as

$$\mathbf{y} = \mathbf{x} + \mathbf{v}, \tag{1}$$

where $\mathbf{v}$ is the additive white Gaussian noise (AWGN) with zero mean and standard deviation σ. The goal of image denoising is to estimate the desired image $\mathbf{x}$ from $\mathbf{y}$. One popular approach to image denoising is the variational method, in which the denoised image is obtained by $\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{x}\|^2 + \mu \cdot R(\mathbf{x}) \right\}$, where $R(\mathbf{x})$ denotes some regularization term and μ is a positive constant. The specific form of $R(\mathbf{x})$ depends on the used image priors.

One common problem of image denoising methods is that the image fine scale details such as texture structures will be over-smoothed. An over-smoothed image will have much weaker gradients than the original image. Intuitively, a good estimation of $\mathbf{x}$ without smoothing too much the textures should have a similar gradient distribution to that of $\mathbf{x}$. With this motivation, we propose a gradient histogram preservation (GHP) model for texture enhanced image denoising (TEID).

Our intuitive idea is to integrate the gradient histogram prior with the other image priors to further improve the denoising performance. Suppose that we have an estimation of the gradient histogram of $\mathbf{x}$, denoted by $\mathbf{h}_r$ (the estimation method will be discussed in Section 4). In order to make the gradient histogram of denoised image $\hat{\mathbf{x}}$ nearly the same as the reference histogram $\mathbf{h}_r$, we propose the following GHP based image denoising model:

$$\begin{aligned} \hat{\mathbf{x}} = \arg\min_{\mathbf{x},F} \left\{ \tfrac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{x}\|^2 + \lambda R(\mathbf{x}) + \mu \|F(\nabla \mathbf{x}) - \nabla \mathbf{x}\|^2 \right\}, \\ \text{s.t. } \mathbf{h}_F = \mathbf{h}_r \end{aligned} \tag{2}$$

where F denotes an odd function which is monotonically non-descending in $(0, +\infty)$, $\mathbf{h}_F$ denotes the histogram of the transformed gradient image $|F(\nabla \mathbf{x})|$, and ∇ denotes the gradient operator. By introducing the transform F, we can use the alternating method for image denoising. Given F, we can fix $\nabla \mathbf{x}_0 = F(\nabla \mathbf{x})$, and use the conventional denoising methods to update $\mathbf{x}$. Given $\mathbf{x}$, we can update F simply by the histogram operator introduced in Section 3.2. Thus, with the introduction of F, we can easily incorporate gradient histogram prior with any existing image priors $R(\mathbf{x})$.

The sparsity and NSS priors have shown promising performance in denoising, and thus we integrate them into the proposed GHP model. Specifically, we adopt the sparse nonlocal regularization term proposed in the centralized sparse representation (CSR) model [7], resulting in the following denoising model:

$$\begin{aligned} \hat{\mathbf{x}} = \arg\min_{\mathbf{x},F} \left\{ \begin{array}{l} \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{x}\|^2 + \lambda \sum_i \left\|\boldsymbol{\alpha}_i - \boldsymbol{\beta}_i\right\|_1 \\ +\mu \|F(\nabla \mathbf{x}) - \nabla \mathbf{x}\|^2 \end{array} \right\}, \\ \text{s.t. } \mathbf{x} = \mathbf{D} \circ \boldsymbol{\alpha},\ \mathbf{h}_F = \mathbf{h}_r \end{aligned} \tag{3}$$

where λ is the regularization parameter, $\mathbf{D}$ is the dictionary and $\boldsymbol{\alpha}$ is the coding coefficients of $\mathbf{x}$ over $\mathbf{D}$.

Let's explain more about the model in Eq. (3). Let $\mathbf{x}_i = \mathbf{R}_i \mathbf{x}$ be a patch extracted at position i, $i = 1, 2, \ldots, N$, where $\mathbf{R}_i$ is the patch extraction operator and N is the number of pixels in the image. Each $\mathbf{x}_i$ is coded over the dictionary $\mathbf{D}$, and the coding coefficients is $\boldsymbol{\alpha}_i$. Let $\boldsymbol{\alpha}$ be the concatenation of all $\boldsymbol{\alpha}_i$, and then $\mathbf{x}$ can be reconstructed by

$$\mathbf{x} = \mathbf{D} \circ \boldsymbol{\alpha} \triangleq \left(\sum_{i=1}^{N} \mathbf{R}_i^T \mathbf{R}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{R}_i^T \mathbf{D} \boldsymbol{\alpha}_i. \tag{4}$$

The physical meaning of Eq. (4) is that we use $\hat{\mathbf{x}}_i = \mathbf{D}\boldsymbol{\alpha}_i$ to reconstruct each patch $\mathbf{x}_i$, and then put all reconstructed patches together as the denoised image $\hat{\mathbf{x}}$ (the overlapped pixels between neighboring patches are averaged).

In Eq. (3), $\boldsymbol{\beta}_i$ is the nonlocal means of $\boldsymbol{\alpha}_i$ in the sparse coding domain. With the current estimate $\hat{\mathbf{x}}$, we use the blocking matching method as in [7] to find the non-local neighbors of $\hat{\mathbf{x}}_i$, denoted by $\hat{\mathbf{x}}_i^q$. Denote by $\boldsymbol{\alpha}_i^q$ the coding coefficients of $\hat{\mathbf{x}}_i^q$. Then $\boldsymbol{\beta}_i$ is computed as the weighted average of $\boldsymbol{\alpha}_i^q$,

$$\boldsymbol{\beta}_i = \sum_q w_i^q \boldsymbol{\alpha}_i^q, \tag{5}$$

where the weight w_i^q is defined as

$$w_i^q = \frac{1}{W} \exp\left(-\frac{1}{h} \left\| \hat{\mathbf{x}}_i - \hat{\mathbf{x}}_i^q \right\|^2 \right), \tag{6}$$

where h is a parameter to control the decay rate and W is a normalization factor to guarantee $\sum_q w_i^q = 1$. Clearly, the regularization term $\sum_i \left\|\boldsymbol{\alpha}_i - \boldsymbol{\beta}_i\right\|_1$ enforce the coding coefficients $\boldsymbol{\alpha}_i$ to approach to its nonlocal means $\boldsymbol{\beta}_i$ so that noise can be removed, while the l_1-norm comes from the fact that $\left\|\boldsymbol{\alpha}_i - \boldsymbol{\beta}_i\right\|_1$ follows Laplacian distribution [7].

From the GHP model with sparse nonlocal regularization in Eq. (3), one can see that if the histogram regularization parameter μ is high, the function $F(\nabla\mathbf{x})$ will be close to $\nabla\mathbf{x}$. Since the histogram $\mathbf{h}_F$ of $|F(\nabla\mathbf{x})|$ is required to be the same as $\mathbf{h}_r$, the histogram of $\nabla\mathbf{x}$ will be similar to $\mathbf{h}_r$, leading to the desired gradient histogram preserved image denoising. Next, we will see that there is an efficient iterative histogram specification algorithm to solve the model in Eq. (3).

3.2. Iterative histogram specification algorithm

Eq. (3) is minimized iteratively. As in [7], the local PCA bases are used as the dictionary $\mathbf{D}$. Based on the current estimation of image $\mathbf{x}$, we cluster its patches into K clusters, and for each cluster, a PCA dictionary is learned. Then for each given patch, we first check which cluster it belongs, and then use the PCA dictionary of this cluster as the $\mathbf{D}$.

We propose an alternating minimization method to solve the problem in Eq. (3). Given the transform function F, we introduce a variable $\mathbf{g} = F(\nabla\mathbf{x})$, and update $\mathbf{x}$ (i.e., $\boldsymbol{\alpha}$) by solving the following sub-problem:

$$\begin{array}{c}\min_{\mathbf{x}} \left\{\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{x}\|^2 + \lambda\sum_i \left\|\boldsymbol{\alpha}_i - \boldsymbol{\beta}_i\right\|_1 + \mu\|\mathbf{g}-\nabla\mathbf{x}\|^2\right\} \\ \text{s.t. } \mathbf{x} = \mathbf{D}\circ\boldsymbol{\alpha}\end{array}. \quad (7)$$

To get the solution to the above sub-problem, we first use a gradient descent method to update $\mathbf{x}$:

$$\mathbf{x}^{(k+1/2)} = \mathbf{x}^{(k)} + \delta\left(\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{x}^{(k)}) + \mu\nabla^T\left(\mathbf{g}-\nabla\mathbf{x}^{(k)}\right)\right), \quad (8)$$

where δ is a pre-specified constant. Then, the coding coefficients $\boldsymbol{\alpha}_i$ are updated by

$$\boldsymbol{\alpha}_i^{(k+1/2)} = \mathbf{D}^T\mathbf{R}_i\mathbf{x}^{(k+1/2)}. \quad (9)$$

By using Eq. (5) to obtain $\boldsymbol{\beta}_i$, we further update $\boldsymbol{\alpha}_i$ by

$$\boldsymbol{\alpha}_i^{(k+1)} = S_{\lambda/d}\left(\begin{array}{c}\frac{1}{d}\mathbf{D}^T\left(\mathbf{R}_i\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}_i^{(k+1/2)}\right) \\ +\boldsymbol{\alpha}_i^{(k+1/2)} - \boldsymbol{\beta}_i\end{array}\right) + \boldsymbol{\beta}_i, \quad (10)$$

where $S_{\lambda/d}$ is the soft-thresholding operator, and d is a constant to guarantee the convexity of the surrogate function [6]. Finally, we use Eq. (4) to update the whole image and let it be $\mathbf{x}^{(k+1)}$.

Once the estimate of image $\mathbf{x}$ is given, we can update F by solving the following sub-problem:

$$\min_F \|F(\nabla\mathbf{x}) - \nabla\mathbf{x}\|^2 \text{ s.t. } \mathbf{h}_F = \mathbf{h}_r. \quad (11)$$

To solve this sub-problem, we let $\mathbf{d}_0 = |\nabla\mathbf{x}|$, and use the standard histogram specification operator [12] to obtain the monotonic non-parametric mapping function F so that the histogram of $|F(\nabla\mathbf{x})|$ is the same as $\mathbf{h}_r$.

Finally, we summarize our proposed iterative histogram specification based GHP algorithm in **Algorithm 1**. It should be noted that, for any gradient based image denoising model, we can easily incorporate the proposed GHP in it by simply modifying the gradient term and adding an extra histogram specification operation.

In [1], Attouch et al. showed that: for a nonconvex function $L(\mathbf{x},\mathbf{y}) = f(\mathbf{x}) + Q(\mathbf{x},\mathbf{y}) + g(\mathbf{y})$, if L satisfies the Kurdyka-Lojasiewicz inequality, proximal alternating minimization would converge to a critical point of L. Note that our model has a similar form to the one discussed in [1], and we also adopted an alternating minimization method. Thus the conclusions in [1] ensure the convergence of the proposed GHP algorithm, and we empirically found that our algorithm converges well.

Algorithm 1: Iterative Histogram Specification for GHP

1. Initialize $k = 0$, $\mathbf{x}^{(k)} = \mathbf{y}$
2. Iterate on $k = 0, 1, ..., J$
3. Update $\mathbf{g}$:
   $\mathbf{g} = F(\nabla\mathbf{x})$
4. Update $\mathbf{x}$:
   $\mathbf{x}^{(k+1/2)} = \mathbf{x}^{(k)} + \delta\left(\begin{array}{c}\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{x}^{(k)}) \\ +\mu\nabla^T(\mathbf{g}-\nabla\mathbf{x}^{(k)})\end{array}\right)$
5. Update the coding coefficients of each patch:
   $\boldsymbol{\alpha}_i^{(k+1/2)} = \mathbf{D}^T\mathbf{R}_i\mathbf{x}^{(k+1/2)}$
6. Update the nonlocal mean of coding vector $\boldsymbol{\alpha}_i$:
   $\boldsymbol{\beta}_i = \sum_q w_i^q \boldsymbol{\alpha}_i^q$
7. Update $\boldsymbol{\alpha}$:
   $\boldsymbol{\alpha}_i^{(k+1)} = S_{\lambda/d}\left(\begin{array}{c}\frac{1}{d}\mathbf{D}^T(\mathbf{R}_i\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}_i^{(k+1/2)}) \\ +\boldsymbol{\alpha}_i^{(k+1/2)} - \boldsymbol{\beta}_i\end{array}\right) + \boldsymbol{\beta}_i$
8. Update $\mathbf{x}$
   $\mathbf{x}^{(k+1)} = \mathbf{D}\circ\boldsymbol{\alpha}^{(k+1)}$
9. Update F via histogram specification by Eq. (11)
10. $k \longleftarrow k+1$
11. $\mathbf{x} = \mathbf{x}^{(k)} + \delta\left(\mu\nabla^T(\mathbf{g}-\nabla\mathbf{x}^{(k)})\right)$

4. Reference gradient histogram estimation

To apply the model in Eq. (3), we need to know the reference histogram $\mathbf{h}_r$, which is supposed to be the gradient histogram of original image $\mathbf{x}$. In this section, we propose a one dimensional deconvolution model to estimate the histogram $\mathbf{h}_r$. Assuming that all pixels in the gradient image $\nabla\mathbf{x}$ are independent and identically distributed (i.i.d.), we can view them as the samples of a scalar variable, denoted by x. Then the normalized histogram of $\nabla\mathbf{x}$ can be regarded as a discrete approximation of the probability density function (PDF) of x. For the additive white Gaussian noise (AWGN) $\mathbf{v}$, we can readily model its elements as the samples of an i.i.d. variable, denoted by v. Since $v \sim N\left(0,\sigma^2\right)$ and let $g = \nabla v$, one can obtain that g is also i.i.d. Gaussian with PDF [22]

$$p_g = \frac{1}{2\sqrt{\pi}\sigma} \exp\left(-\frac{g^2}{4\sigma^2}\right). \quad (12)$$

Since $\mathbf{y} = \mathbf{x} + \mathbf{v}$, we have $\nabla\mathbf{y} = \nabla\mathbf{x} + \nabla\mathbf{v}$. It is ready to model $\nabla\mathbf{y}$ as an i.i.d. variable, denoted by y, and we have $y = x + g$. Let p_x be the PDF of x, and p_y be the PDF of y. Since x and g are independent, the joint PDF $p(x, g)$ is,

$$p(x, g) = p_x \times p_g. \quad (13)$$

Then the PDF p_y is

$$p_y(y = t) = \int_a p_x(x = a) \times p_g(g = (t - a))\, da. \quad (14)$$

If we use the normalized histogram $\mathbf{h}_x$ and $\mathbf{h}_y$ to approximate p_x and p_y, we can rewrite Eq. (14) in the discrete domain as:

$$\mathbf{h}_y = \mathbf{h}_x \otimes \mathbf{h}_g, \quad (15)$$

where $\otimes$ denotes the convolution operator. Note that $\mathbf{h}_g$ can be obtained by discretizing p_g, and $\mathbf{h}_y$ can be computed directly from the noisy observation $\mathbf{y}$.

Obviously, the estimation of $\mathbf{h}_x$ can be generally modeled as a deconvolution problem:

$$\mathbf{h}_r = \arg\min_{\mathbf{h}_x} \left\{ \left\| \mathbf{h}_y - \mathbf{h}_x \otimes \mathbf{h}_g \right\|^2 + c \cdot R(\mathbf{h}_x) \right\}, \quad (16)$$

where c is a constant and $R(\mathbf{h}_x)$ is some regularization term based on the prior information of natural image's gradient histogram. Considering that $\mathbf{h}_x$, i.e., the discrete version of p_x, can be well modeled as hyper-Laplacian distribution [4, 5, 17], in this paper we use a simple parametric method to estimate p_x and then discretize it into $\mathbf{h}_x$.

The hyper-Laplacian modeling of p_x is:

$$p_x = k \exp(-\kappa |x|^\gamma), \quad (17)$$

where k is the normalization factor. The estimation of p_x is converted into the estimation of parameters κ and γ. Considering the fact that for natural images, κ and γ will have a relatively narrow range, we preset a range of each of the two parameters, and then search for the pair (κ, γ) which makes $\left\| \mathbf{h}_y - \mathbf{h}_x \otimes \mathbf{h}_g \right\|^2$ the smallest. Specifically, we let $\kappa \in [0.001, 3]$ and $\gamma \in [0.02, 1.5]$. In addition, in our experiments the Nelder-Mead method is used to make the searching more efficient. Fig. 2 shows an example of reference gradient histogram estimation. It can be seen that our method can obtain a good estimation of $\mathbf{h}_x$.

5. Experimental results

We first give the parameter setting in our GHP based TEID algorithm, and then conduct experiments to validate its performance in comparison with state-of-the-art denoising algorithms. Finally, we make some discussion of its potential improvements. The Matlab source code of our algorithm can be downloaded at `http://www4.comp.polyu.edu.hk/~cslzhang/code.htm`.

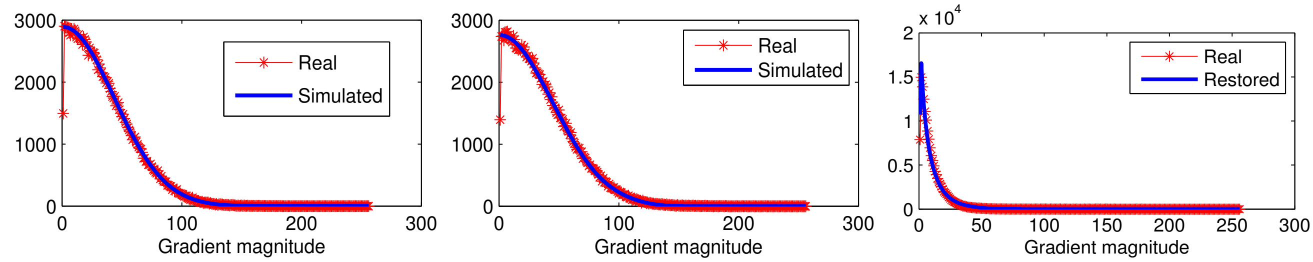
5.1. Parameter setting

Our algorithm involves a few parameters to set, including the regularization parameters λ and μ in Eq. (7) to balance the effect of gradient preservation, constant δ in Eq. (8) and d in Eq. (10) to ensure convexity. For the parameter λ, we use the same strategy as in [8] to adaptively update it according to the maximum a posterior (MAP) principle. Based on our experimental experience, we set the parameter μ to 5, and δ to 0.23 for noise level less than 30 while 0.26 for other noise levels. Based on the analysis in [6], to guarantee the convexity of surrogate function, d should be larger than the spectral norm of dictionary $\mathbf{D}$. Since in our algorithm $\mathbf{D}$ is an orthonormal PCA matrix, any d greater than 1 will be fine, and we set it to 3 by experience. Note that these parameters are fixed to all images in our experiments.

5.2. Denoising results

To verify the performance of our proposed GHP based TEID method, we apply it to ten natural images with various texture structures. The scenes of these images can be found in Fig. 3. Some state-of-the-art denoising methods are used for comparison, including shape-adaptive PCA based BM3D (SAPCA-BM3D) [16], the learned simultaneously sparse coding (LSSC) [21] and the CSR [7] methods. The codes of all the competing methods are provided by the authors and we used the recommended parameters by the authors. Considering the fact when noise is too strong, all methods cannot recover the fine scale texture structures in the image, and in practice the noise is often moderate or below, we set the AWGN noise level $\sigma \in \{20, 25, 30, 35, 40\}$ in the experiments.

The quantitative experimental results by the competing methods are shown Table 1. Apart from PSNR, we also list the results by using the perceptual quality metric SSIM [30]. From this table, we can see that the proposed GHP method has similar PSNR/SSIM measures to SAPCA-BM3D, LSSC and CSR. Nonetheless, the goal of our GHP method is to preserve and enhance the image texture structures, and let's compare the visual quality of the denoised images by these methods. Fig. 4 shows an example. In this image, there are different texture regions, such as the sky, tree, water and building. We can see that SAPCA-BM3D, LSSC and CSR smooth much the textures in tree, water and building areas, while SAPCA-BM3D introduces some artifacts in the smooth sky area. Though they have good PSNR and even SSIM indices, the denoised images by them look somewhat unnatural. In contrast, the proposed GHP method



(a) (b) (c)

Figure 2. An example of reference gradient histogram estimation. (a) Real and simulated AWGN gradient histograms (noise level $\sigma = 30$); (b) real and simulated gradient histograms of noisy image; and (c) real and estimated gradient histograms of the clean image.

Table 1. PSNR (dB) and SSIM results by different methods.

σ	SAPCA-BM3D[16]					LSSC[21]					CSR[7]					GHP				
	20	25	30	35	40	20	25	30	35	40	20	25	30	35	40	20	25	30	35	40
1	30.83	29.66	28.75	28.02	27.41	30.69	29.56	28.62	27.91	27.32	30.59	29.46	28.58	27.76	27.19	30.49	29.35	28.40	27.31	26.49
	0.876	0.849	0.825	0.803	0.784	0.872	0.846	0.820	0.800	0.781	0.869	0.843	0.820	0.793	0.776	0.864	0.837	0.811	0.792	0.775
2	28.07	26.99	26.18	25.54	25.02	27.98	26.94	26.14	25.51	24.98	27.91	26.87	26.08	25.37	24.87	27.80	26.68	25.81	24.85	24.16
	0.817	0.773	0.734	0.699	0.668	0.815	0.773	0.734	0.700	0.670	0.807	0.764	0.727	0.681	0.651	0.810	0.768	0.731	0.689	0.656
3	28.39	27.43	26.66	26.01	25.46	28.46	27.52	26.66	26.03	25.47	28.11	27.16	26.39	25.64	25.10	28.09	27.14	26.36	25.46	24.88
	0.755	0.721	0.692	0.667	0.647	0.762	0.728	0.696	0.670	0.647	0.736	0.702	0.675	0.640	0.621	0.756	0.721	0.691	0.656	0.635
4	26.86	25.68	24.79	24.08	23.50	26.75	25.61	24.76	24.06	23.48	26.65	25.52	24.64	23.84	23.26	26.59	25.43	24.51	23.62	22.91
	0.803	0.758	0.715	0.677	0.641	0.803	0.758	0.717	0.678	0.643	0.782	0.737	0.697	0.640	0.604	0.796	0.752	0.715	0.673	0.637
5	30.88	29.96	29.21	28.58	28.06	30.75	29.81	29.04	28.41	27.90	30.64	29.68	28.91	28.27	27.76	30.56	29.54	28.63	27.66	26.75
	0.812	0.780	0.754	0.730	0.709	0.809	0.776	0.744	0.718	0.696	0.802	0.770	0.742	0.710	0.690	0.805	0.773	0.742	0.714	0.688
6	28.59	27.32	26.35	25.59	24.97	28.47	27.26	26.33	25.59	24.98	28.49	27.24	26.30	25.49	24.90	28.35	27.11	26.11	25.16	24.46
	0.888	0.856	0.824	0.794	0.765	0.883	0.850	0.825	0.795	0.769	0.882	0.851	0.820	0.788	0.761	0.874	0.844	0.816	0.798	0.776
7	30.17	29.14	28.35	27.71	27.18	30.18	29.18	28.40	27.81	27.32	30.13	29.14	28.38	27.71	27.22	30.07	28.98	28.13	27.11	26.37
	0.839	0.803	0.771	0.744	0.721	0.840	0.807	0.775	0.751	0.729	0.833	0.799	0.770	0.738	0.717	0.840	0.806	0.776	0.746	0.722
8	31.58	30.48	29.64	28.94	28.37	31.38	30.33	29.54	28.86	28.32	31.41	30.35	29.52	28.79	28.24	31.19	30.04	29.09	27.87	27.05
	0.900	0.879	0.861	0.843	0.828	0.894	0.872	0.858	0.840	0.826	0.897	0.877	0.860	0.841	0.827	0.889	0.865	0.844	0.832	0.820
9	27.58	26.37	25.44	24.73	24.15	27.58	26.40	25.48	24.77	24.19	27.34	26.18	25.31	24.47	23.92	27.26	26.09	25.18	24.18	23.53
	0.821	0.778	0.740	0.707	0.677	0.822	0.782	0.748	0.716	0.687	0.804	0.764	0.729	0.683	0.655	0.809	0.769	0.737	0.700	0.671
10	31.23	30.28	29.53	28.92	28.42	31.04	30.08	29.36	28.75	28.24	30.98	30.03	29.30	28.76	28.28	30.85	29.73	28.78	27.73	26.83
	0.823	0.791	0.763	0.740	0.721	0.818	0.787	0.755	0.732	0.712	0.813	0.781	0.755	0.728	0.710	0.814	0.780	0.749	0.723	0.699
Avg	**29.42**	**28.33**	**27.49**	**26.81**	**26.25**	**29.33**	**28.27**	**27.43**	**26.77**	**26.22**	**29.23**	**28.16**	**27.34**	**26.61**	**26.07**	**29.13**	**28.01**	**27.10**	**26.10**	**25.34**
	0.833	**0.799**	**0.768**	**0.740**	**0.716**	**0.832**	**0.798**	**0.767**	**0.740**	**0.716**	**0.823**	**0.789**	**0.760**	**0.724**	**0.701**	**0.826**	**0.792**	**0.761**	**0.732**	**0.708**

preserves much better these fine texture areas, making the denoised image look more natural and visually pleasant.

Due to the limit of space, here we cannot put more visual results. More examples can be found in the supplementary file attached to this paper.

5.3. Discussions

It is worth noting that, to further enhance the noise removal and texture preservation performance of our method, region-based GHP could be implemented. Since natural images often consist of different regions with different textures, the gradient distributions in these regions will also vary. Therefore, with the help of image segmentation methods such as mean-shift [11], we can partition the noisy image into several homogeneous regions, and apply the GHP method to each region. Fig. 5 shows an example. One can see that without segmentation, the proposed GHP method may generate some false textures in the less textured area (e.g., cloud) due to the influence of other texture areas (e.g., trees). With roughly segment the image into 2 regions, as shown in Fig. 5(c), GHP leads to very satisfying denoising results in all regions.

Figure 3. Ten test images. From left to right and top to bottom, they are labeled as 1 to 10.

6. Conclusion

In this paper, we presented a novel gradient histogram preserving (GHP) model for texture-enhanced image denoising (TEID). The GHP model can preserve the gradient distribution by pushing the gradient histogram of the denoised image toward the reference histogram, and thus is promising in enhancing the texture structure while removing random noise. To implement the GHP model, we proposed an efficient iterative histogram specification algorithm. Meanwhile, we presented a simple but theoretically solid algorithm to estimate the reference gradient histogram from the noisy image. Experimental results verify the effectiveness of GHP based TEID. The proposed GHP has similar PSNR/SSIM measures to state-of-the-art denoising methods such as SAPCA-BM3D, LSSC and CSR; however, it leads to more natural and visually pleasant denoising results by preserving better the image texture areas. In the future, we will extend GHP to image deblurring, superresolution and other image reconstruction tasks.

Acknowledgements

This work is supported by NSFC under Grant No. 61271093, the Hong Kong Scholar Program, and the program of ministry of education for new century excellent talents.

References

[1] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438–457, 2010.

[2] A. Buades, B. Coll, and J. Morel. A review of image denoising methods, with a new one. *Multiscale Model. Simul.*, 4(2):490–530, 2005.

[3] H. C. Burger, C. J. Schuler, and S. Harmeling. Image denoising: can plain neural networks compete with bm3d? *Proc. CVPR*, 2012.

[4] T. S. Cho, N. Joshi, C. L. Zitnick, S. B. Kang, R. Szeliski, and W. T. Freeman. A content-aware image prior. *Proc. CVPR*, 2010.

[5] T. S. Cho, C. L. Zitnick, N. Joshi, S. B. Kang, R. Szeliski, and W. T. Freeman. Image restoration by matching gradient distributions. *IEEE T-PAMI*, 34(4):683–694, 2012.

[6] I. Daubechies, M. Defriese, and C. DeMol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.*, 57(11):1413–1457, 2004.

[7] W. Dong, L. Zhang, and G. Shi. Centralized sparse representation for image restoration. *Proc. ICCV*, 2011.

[8] W. Dong, L. Zhang, G. Shi, and X. Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE T-IP*, 20(7):1838–1857, 2011.

[9] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE T-IP*, 15(12):3736–3745, 2006.

[10] R. Fergus, B. Singh, A. Hertzmann, S. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. *Proc. ACM SIGGRAPH*, 2006.

[11] B. Georgescu, I. Shimshoni, and P. Meer. Mean shift based clustering in high dimensions: A texture classification example. *Proc. ICCV*, 2003.

[12] R. C. Gonzalez and R. E. Woods. Digital image processing. *Beijing: Publishing House of Electronics Industry*, 2005.

[13] E. Hadjidemetriou, M. D. Grossberg, and S. K. Nayar. Multiresolution histograms and their use for recognition. *IEEE T-PAMI*, 26(7):831–847, 2004.

[14] J. Jancsary, S. Nowozin, and C. Rother. Loss-specific training of non-parametric image restoration models: a new state of the art. *Proc. ECCV*, 2012.

[15] N. Joshi, C. L. Zitnick, R. Szeliski, and D. Kriegman. Image deblurring and denoising using color priors. *Proc. CVPR*, 2009.

[16] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola. From local kernel to nonlocal multiple-model image denoising. *IJCV*, 86(1):1–32, 2010.

[17] D. Krishnan and R. Fergus. Fast image deconvolution using hyper-laplacian priors. *Proc. NIPS*, 2009.

[18] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. *Proc. ACM SIGGRAPH*, 2007.

[19] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Efficient marginal likelihood optimization in blind deconvolution. *Proc. CVPR*, 2011.

[20] J. Mairal, F. Bach, and J. Ponce. Task-driven dictionary learning. *IEEE T-PAMI*, 32(4):791–804, 2012.

[21] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. *Proc. ICCV*, 2009.

[22] J. K. Patel and C. B. Read. Handbook of the normal distribution. *New York: Marcel Dekker*, 1982.

[23] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using a scale mixture of gaussians in the wavelet domain. *IEEE T-IP*, 12(11):1338–1351, 2003.

[24] S. Roth and M. J. Black. Fields of experts: a framework for learning image priors. *Proc. CVPR*, 2005.

[25] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268, 1992.

[26] K. Suzuki, I. Horiba, and N. Sugie. Efficient approximation of neural filters for removing quantum noise from images. *IEEE T-SP*, 50(7):1787–1799, 2002.

[27] M. Varma and A. Zisserman. Classifying images of materials: achieving viewpoint and illumination independence. *Proc. ECCV*, 2002.

[28] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *IJCV*, 62(1-2):61–81, 2005.

[29] M. Wainwright and S. Simoncelli. Scale mixtures of gaussians and the statistics of natural images. *Proc. NIPS*, 1999.

[30] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE T-IP*, 13(4):600–612, 2004.

[31] M. Welling, G. Hinton, and S. Osindero. Learning sparse topographic representations with products of student-t distributions. *Proc. NIPS*, 2002.

[32] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. *Proc. ICCV*, 2011.

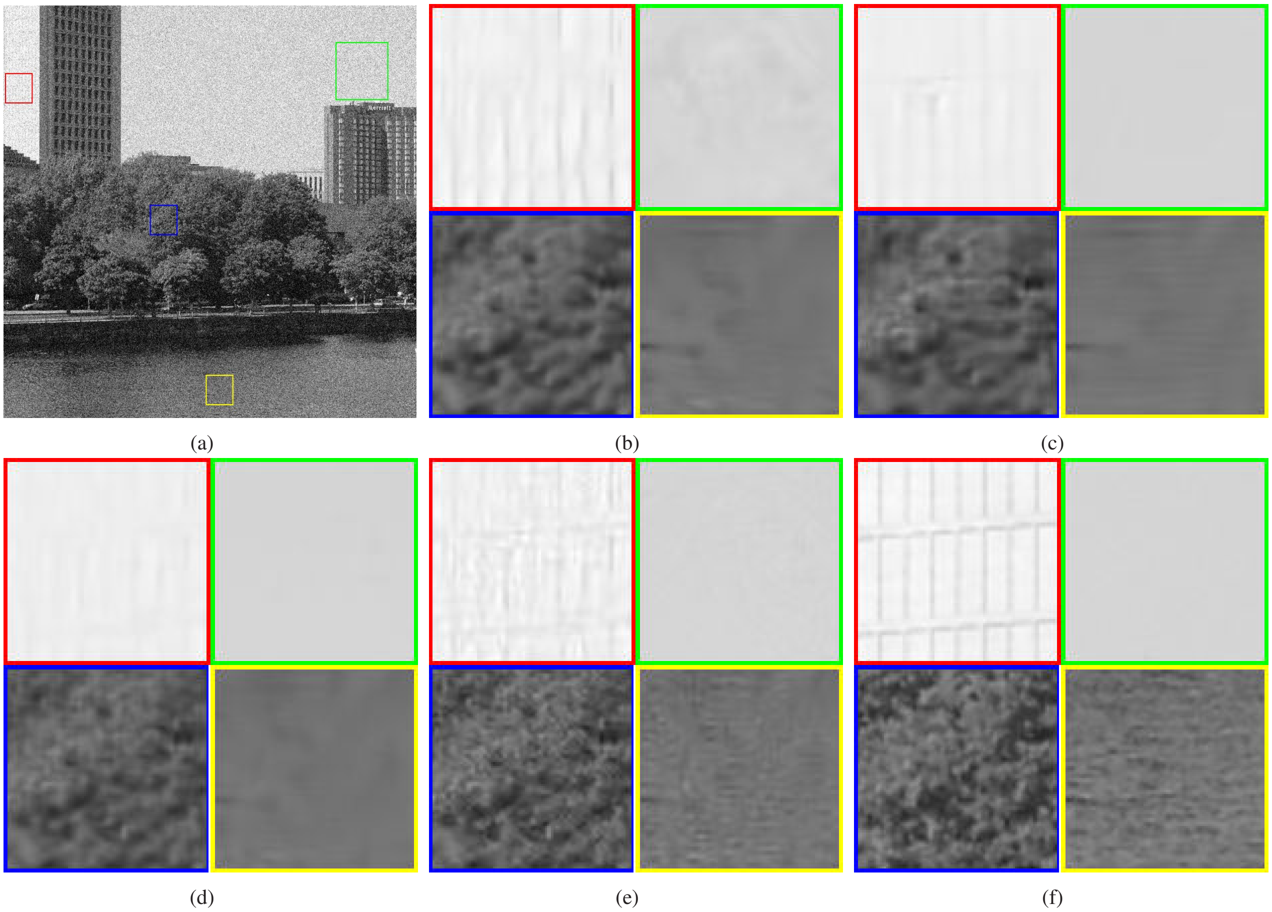


Figure 4. Methods comparison. (a) Noisy image with AWGN of standard deviation 30; (b) SAPCA-BM3D [16] restoration result; (c) LSSC [21] restoration result; (d) CSR [7] restoration result; (e) GHP restoration result; (f) ground truth.

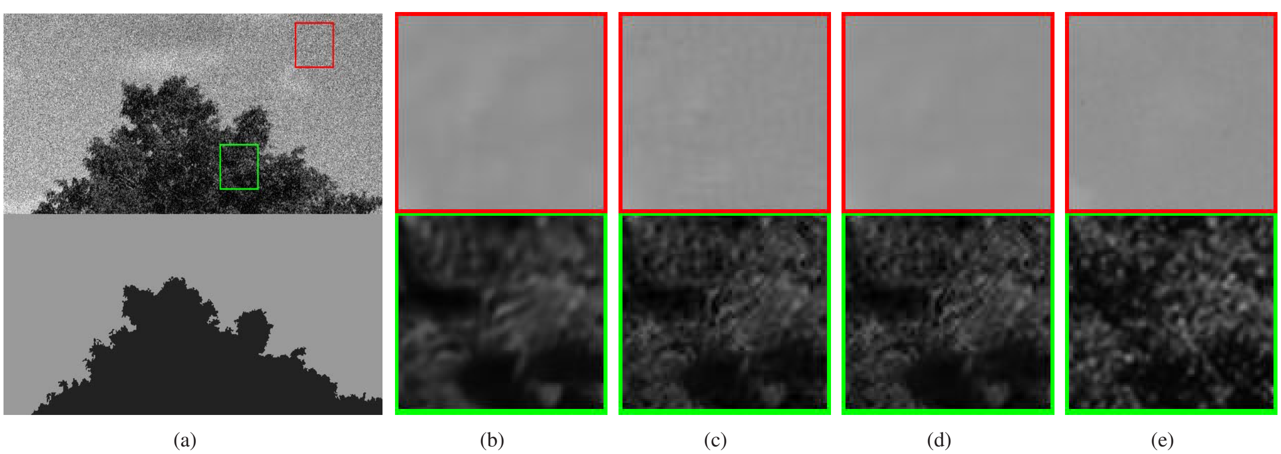


Figure 5. Results comparison with and without segmentation. (a) Top: noisy image with AWGN of standard deviation 30; bottom: a two-region segmentation of it; (b) SAPCA-BM3D [16] restoration results; (c) GHP restoration results without segmentation; (d) GHP restoration results with segmentation; (e) ground truth.