Gradient Histogram Estimation and Preservation for Texture Enhanced Image Denoising

Wangmeng Zuo, Lei Zhang, Chunwei Song, David Zhang and Huijun Gao

Abstract—Natural image statistics plays an important role in image denoising, and various natural image priors, including gradient based, sparse representation based and nonlocal self-similarity based ones, have been widely studied and exploited for noise removal. In spite of the great success of many denoising algorithms, they tend to smooth the fine scale image textures when removing noise, degrading the image visual quality. To address this problem, in this paper we propose a texture enhanced image denoising method by enforcing the gradient histogram of the denoised image to be close to a reference gradient histogram of the original image. Given the reference gradient histogram, a novel gradient histogram preservation (GHP) algorithm is developed to enhance the texture structures while removing noise. Two region-based variants of GHP are proposed for the denoising of images consisting of regions with different textures. An algorithm is also developed to effectively estimate the reference gradient histogram from the noisy observation of the unknown image. Our experimental results demonstrate that the proposed GHP algorithm can well preserve the texture appearance in the denoised images, making them look more natural.

Index Terms—Image denoising, histogram specification, non-local similarity, sparse representation.

I. INTRODUCTION

IMAGE denoising, which aims to estimate the latent clean image **x** from its noisy observation **y**, is a classical yet still active topic in image processing and low level vision. One widely used data observation model [4], [7], [9]–[11] is **y** = **x** + **v**, where **v** is additive white Gaussian noise (AWGN). One popular approach to image denoising is the variational method, where an energy functional is minimized to search the desired estimation of **x** from its noisy observation **y**. The energy functional usually involves two terms: a data fidelity term which depends on the image degeneration process and a regularization term which models the prior of clean natural images [4], [7], [8], [12]. The statistical modeling of natural image priors is crucial to the success of image denoising.

Motivated by the fact that natural image gradients and wavelet transform coefficients have a heavy-tailed distribution, sparsity priors are widely used in image denoising [1]–[3].
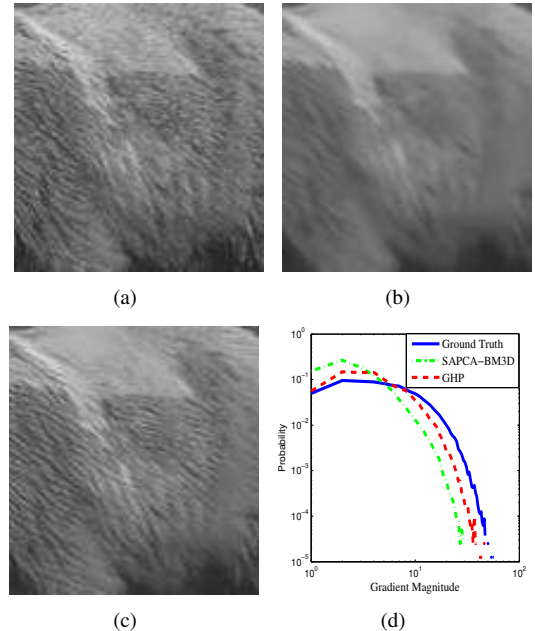
Fig. 1: Examples of denoised images and their gradient histograms. (a) A cropped image with hair textures; (b) denoised image by the SAPCA-BM3D method [11]; (c) denoised image by the proposed texture enhanced image denoising via gradient histogram preservation (GHP); (d) the gradient histograms of the denoised images. One can see that the proposed GHP method can recover more texture details than other methods, and the gradient histogram of the denoised image by GHP is also closer to the gradient histogram of ground truth image.

The well-known total variation minimization methods actually assume Laplacian distribution of image gradients [4]. The sparse Laplacian distribution is also used to model the high-pass filter responses and wavelet/curvelet transform coefficients [5], [6]. By representing image patches as a sparse linear combination of the atoms in an over-complete redundant dictionary, which can be analytically designed or learned from natural images, sparse coding has proved to be very effective in image denoising via l_0-norm or l_1-norm minimization [7], [8]. Another popular prior is the nonlocal self-similarity (NSS) prior [9]–[11], [50]; 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 connection between NSS and the sparsity prior is discussed in [11], [12]. The joint use of sparsity prior and NSS prior has led to state-of-the-art image denoising results [12]–[14]. In spite of the great success of many denoising algorithms, however, they often fail to preserve the image fine scale texture structures

[23], degrading much the image visual quality (please refer to Fig. 1 for example).

With the rapid development of digital imaging technology, now the acquired images can contain tens of megapixels. On one hand, more fine scale texture features of the scene will be captured; on the other hand, the captured high definition image is more prone to noise because the smaller size of each pixel makes the exposure less sufficient. Unfortunately, suppressing noise and preserving textures are 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 are much more complex and are hard to characterize by using a sparse model. Texture regions in an image are homogeneous and are composed of similar local patterns, which can be characterized by using local descriptors or textons [43]. Cognitive studies [35], [36], [43], [44] have revealed that the first-order statistics, e.g., histograms, are the most significant descriptors for texture discrimination. Considering these facts, histogram of local features has been widely used in texture analysis [15]–[17]. Meanwhile, image gradients are crucial to the perception and analysis of natural images [45], [46]. All these motivate us to use the histogram of image gradient to design new image denoising models.

With the above considerations, in this paper we propose a novel gradient histogram preservation (GHP) method for texture enhanced image denoising. From the given noisy image $\mathbf{y}$, we estimate the gradient histogram of original image $\mathbf{x}$. Taking this estimated histogram, denoted by $\mathbf{h}_r$, as a reference, we search an estimate of $\mathbf{x}$ such that its gradient histogram is close to $\mathbf{h}_r$. As shown in Fig. 1, the proposed GHP based denoising method can well enhance the image texture regions, which are often over-smoothed by other denoising methods. The major contributions of this paper are summarized as follows:

(1) A novel texture enhanced image denoising framework is proposed, which preserves the gradient histogram of the original image. The existing image priors can be easily incorporated into the proposed framework to improve the quality of denoised images.
(2) Using histogram specification, a gradient histogram preservation algorithm is developed to ensure that the gradient histogram of denoised image is close to the reference histogram, resulting in a simple yet effective GHP based denoising algorithm.
(3) By incorporating the hyper-Laplacian and nonnegative constraints, a regularized deconvolution model and an iterative deconvolution algorithm are presented to estimate the image gradient histogram from the given noisy image.

The rest of the paper is organized as follows. Section II provides a brief survey of the related work. Section III introduces the gradient histogram estimation and preservation framework. Section IV presents the proposed denoising model and the iterative histogram specification algorithm, while Section V describes the regularized deconvolution model and algorithm for gradient histogram estimation. Section VI presents the experimental results. Finally, Section VII concludes this paper.

II. RELATED WORK

Image denoising methods can be grouped into 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 belong to the first category. Learning-based methods attempt to learn a mapping function from the noisy image to the clean image [19], and have been receiving considerable research interests [20], [21]. Here we briefly 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 domain. One representative class of image priors is the gradient prior based on the observation that natural images have a heavy-tailed distribution of gradients. The use of gradient prior can be traced back to 1990s when Rudin et al. [4] proposed a total variation (TV) model for image denoising, where the gradients are actually modeled as Laplacian distribution. Another well-known prior model, the mixture of Gaussians, can also be used to approximate the distribution of image gradient [1], [22]. In addition, hyper-Laplacian model can more accurately characterize the heavy-tailed distribution of gradients, and has been widely applied to various image restoration tasks [2], [3], [23]–[25].

The image gradient prior is a kind of local sparsity prior, i.e., the gradient distribution is sparse. More generally, the local sparsity prior can be well applied to high-pass filter responses, wavelet/curvelet transform coefficients, or the coding coefficients over a redundant dictionary. In [5], [6], Gaussian scale mixtures are used to characterize the marginal and joint distributions of wavelet transform coefficients. In [26], [27], the Student t-distributions are used for both basis filter learning and filter response modeling. 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., analysis and synthesis K-SVD [7], [28], task driven DL [29], and adaptive sparse domain selection [8] have been proposed and applied to image denoising and other image 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 [9] 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 [11], LSSC [12], and EPLL [30], etc. While most of the NSS based methods find similar patches on the original scale, recent studies have shown that NSS across different scales can also benefit image denoising [52]. Under the NSS framework, Levin et al. [31] investigated the inherent limit of denoising algorithms, and their empirical validation showed that the existing methods might still be improved by 1 dB.

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. [13] unified both image local sparsity and nonlocal similarity priors via clustering-based sparse representation. Recently, Jancsary et al. [32] proposed a method called regression tree fields (RTF) to integrate different priors.

However, many existing image denoising algorithms, including those local sparsity and NSS based ones, tend to wipe out the image fine scale textures while removing noise. As we discussed in Section I, considering the randomness and homogeneousness of image texture regions, we propose to use the histogram of gradient to describe image texture and design a novel image denoising algorithm with gradient histogram preservation. In [23], [24], Cho et al. used hyper-Laplacian distribution 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. [23] found that the deblurred images can have more detailed textures as well as better visual quality. However, in [23], [24] the estimation of desired gradient distribution is rather heuristic, and the iterative distribution reweighting algorithm is very complex.

III. THE TEXTURE ENHANCED IMAGE DENOISING FRAMEWORK

The noisy observation $\mathbf{y}$ of an unknown clean image $\mathbf{x}$ 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

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{x}\|^2 + \lambda \cdot R(\mathbf{x}) \right\}, \tag{2}$$

where $R(\mathbf{x})$ denotes some regularization term and λ is a positive constant. The specific form of $R(\mathbf{x})$ depends on the employed 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, whose framework is illustrated in Fig. 2.

Suppose that we have an estimation of the gradient histogram of $\mathbf{x}$, denote by $\mathbf{h}_r$. 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:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x},F} \left\{ \frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{x}\|^2 + \lambda R(\mathbf{x}) + \mu\|F(\nabla\mathbf{x}) - \nabla\mathbf{x}\|^2 \right\},$$
$$\text{s.t. } \mathbf{h}_F = \mathbf{h}_r \tag{3}$$

where F denotes an odd function which is monotonically non-descending, $\mathbf{h}_F$ denotes the histogram of the transformed gradient image $|F(\nabla\mathbf{x})|$, ∇ denotes the gradient operator, and μ is a positive constant. The proposed GHP algorithm adopts the alternating optimization strategy. Given F, we can fix $\nabla\mathbf{x}_0 = F(\nabla\mathbf{x})$, and update $\mathbf{x}$. Given $\mathbf{x}$, we can update F by the histogram specification based shrinkage operator which will be introduced in Section IV. Thus, by introducing F, we can easily incorporate the gradient histogram constraint with any existing image regularizer $R(\mathbf{x})$.

Another issue in the GHP model is how to find the reference histogram $\mathbf{h}_r$ of unknown image $\mathbf{x}$. In practice, we need to estimate $\mathbf{h}_r$ based on the noisy observation $\mathbf{y}$. In Section V, we will propose a regularized deconvolution model and an associated iterative deconvolution algorithm to estimate $\mathbf{h}_r$ from the given noisy image. Once the reference histogram $\mathbf{h}_r$ is obtained, the GHP algorithm is then applied for texture enhanced image denoising.

IV. DENOISING WITH GRADIENT HISTOGRAM PRESERVATION

A. The Denoising Model

The proposed denoising method is a patch based method. Let $\mathbf{x}_i = \mathbf{R}_i\mathbf{x}$ be a patch extracted at position i, $i = 1, 2, ..., N$, where $\mathbf{R}_i$ is the patch extraction operator and N is the number of pixels in the image. Given a dictionary $\mathbf{D}$, we sparsely encode the patch $\mathbf{x}_i$ over $\mathbf{D}$, resulting in a sparse coding vector $\boldsymbol{\alpha}_i$. Once the coding vectors of all image patches are obtained, the whole image $\mathbf{x}$ can be reconstructed by [7]:

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

where $\boldsymbol{\alpha}$ is the concatenation of all $\boldsymbol{\alpha}_i$.

Good priors of natural images are crucial to the success of an image denoising algorithm. A proper integration of different priors could further improve the denoising performance. For example, the methods in [12], [14], [32] integrate image local sparsity prior with nonlocal NSS prior, and they have shown promising denoising results. In the proposed GHP model, we adopt the following sparse nonlocal regularization term proposed in the nonlocally centralized sparse representation (NCSR) model [14]:

$$R(\mathbf{x}) = \sum_i \left\|\boldsymbol{\alpha}_i - \boldsymbol{\beta}_i\right\|_1, \text{ s.t. } \mathbf{x} = \mathbf{D} \circ \boldsymbol{\alpha}, \tag{5}$$

where $\boldsymbol{\beta}_i$ is defined as the weighted average of $\boldsymbol{\alpha}_i^q$:

$$\boldsymbol{\beta}_i = \sum_q w_i^q \boldsymbol{\alpha}_i^q, \tag{6}$$

and $\boldsymbol{\alpha}_i^q$ is the coding vector of the qth nearest patch (denoted by $\mathbf{x}_i^q$) to $\mathbf{x}_i$. The weight is defined as $w_i^q = \frac{1}{W}\exp\left(-\frac{1}{h}\left\|\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_i^q\right\|^2\right)$ ($\hat{\mathbf{x}}_i$ and $\hat{\mathbf{x}}_i^q$ denote the current estimates of $\mathbf{x}_i$ and $\mathbf{x}_i^q$, respectively), where h is a predefined constant and W is the normalization factor. More detailed explanations on NCSR can be found in [14].

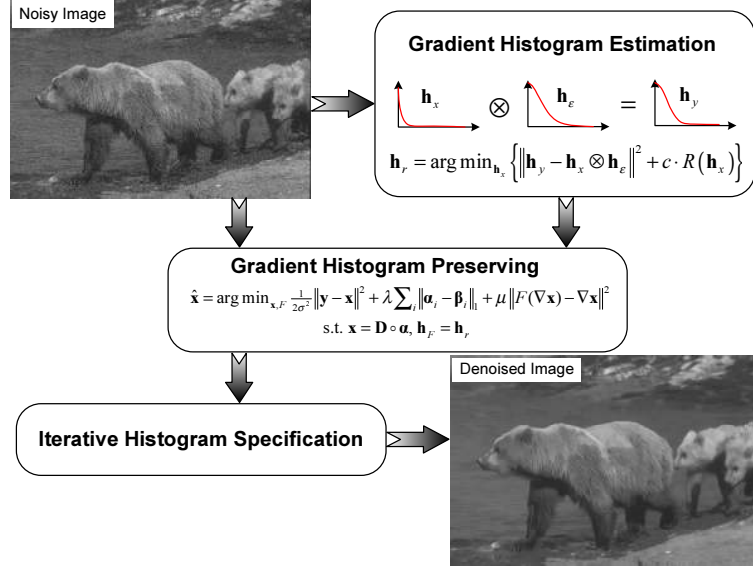By incorporating the above $R(\mathbf{x})$ into Eq. (3), the proposed

Fig. 2: Flowchart of the proposed texture enhanced image denoising framework.

GHP model can be formulated as:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x},F}\left\{\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{x}\|^2 + \lambda\sum_i\left\|\boldsymbol{\alpha}_i-\boldsymbol{\beta}_i\right\|_1 + \mu\|F(\nabla\mathbf{x})-\nabla\mathbf{x}\|^2\right\}$$
$$\text{s.t. } \mathbf{x}=\mathbf{D}\circ\boldsymbol{\alpha},\ \mathbf{h}_F=\mathbf{h}_r \tag{7}$$

From the GHP model with sparse nonlocal regularization in Eq. (7), 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. In the next subsection, we will see that there is an efficient iterative histogram specification algorithm to solve the model in Eq. (7).

B. Iterative Histogram Specification Algorithm

The proposed GHP model in Eq. (7) can be solved by using the variable splitting (VS) method, which has been widely adopted in image restoration [40]–[42]. By introducing a variable $\mathbf{g}=F(\nabla\mathbf{x})$, we adopt an alternating minimization strategy to update $\mathbf{x}$ and $\mathbf{g}$ alternatively. Given $\mathbf{g}=F(\nabla\mathbf{x})$, we update $\mathbf{x}$ (i.e., $\boldsymbol{\alpha}$) by solving the following sub-problem:

$$\min_{\mathbf{x}}\left\{\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{x}\|^2 + \lambda\sum_i\left\|\boldsymbol{\alpha}_i-\boldsymbol{\beta}_i\right\|_1 + \mu\|\mathbf{g}-\nabla\mathbf{x}\|^2\right\}$$
$$\text{s.t. } \mathbf{x}=\mathbf{D}\circ\boldsymbol{\alpha} \tag{8}$$

We use the method in [14] to construct the dictionary $\mathbf{D}$ adaptively. 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 to, and then use the PCA dictionary of this cluster as $\mathbf{D}$. Although in Eq. (8) the l_1-norm regularization is imposed on $\left\|\boldsymbol{\alpha}_i-\boldsymbol{\beta}_i\right\|_1$ rather than $\|\boldsymbol{\alpha}_i\|_1$, by introducing a new variable $\boldsymbol{\vartheta}_i = \boldsymbol{\alpha}_i - \boldsymbol{\beta}_i$, we can use the iterative shrinkage / thresholding method [33] to update $\boldsymbol{\vartheta}_i$ and then update $\boldsymbol{\alpha}_i = \boldsymbol{\vartheta}_i + \boldsymbol{\beta}_i$. This strategy is also used in [14]

to solve the problem with this regularization term, and thus here we omit the detailed deduction process.

To get the solution to the sub-problem in Eq. (8), we first use a gradient descent method to update $\mathbf{x}$:

$$\mathbf{x}^{(k+1/2)} = \mathbf{x}^{(k)} + \delta\left(\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{x}^{(k)}) + \mu\nabla^T\left(\mathbf{g}-\nabla\mathbf{x}^{(k)}\right)\right), \tag{9}$$

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

$$\boldsymbol{\alpha}_i^{(k+1/2)} = \mathbf{D}^T\mathbf{R}_i\mathbf{x}^{(k+1/2)}. \tag{10}$$

By using Eq. (6) to obtain $\boldsymbol{\beta}_i$, we further update $\boldsymbol{\alpha}_i$ by

$$\boldsymbol{\alpha}_i^{(k+1)} = S_{\lambda/d}\left(\boldsymbol{\alpha}_i^{(k+1/2)}-\boldsymbol{\beta}_i\right)+\boldsymbol{\beta}_i, \tag{11}$$

where $S_{\lambda/d}$ is the soft-thresholding operator, and d is a constant to guarantee the convexity of the surrogate function [33]. Finally, we update $\mathbf{x}^{(k+1)}$ by

$$\mathbf{x}^{(k+1)} = \mathbf{D}\circ\boldsymbol{\alpha}^{(k+1)} \triangleq \left(\sum_{i=1}^N\mathbf{R}_i^T\mathbf{R}_i\right)^{-1}\sum_{i=1}^N\mathbf{R}_i^T\mathbf{D}\boldsymbol{\alpha}_i^{(k+1)}. \tag{12}$$

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

$$\min_{\mathbf{g},F}\|\mathbf{g}-\nabla\mathbf{x}\|^2 \text{ s.t. } \mathbf{h}_F=\mathbf{h}_r, \mathbf{g}=F(\nabla\mathbf{x}). \tag{13}$$

Considering the equality constraint $\mathbf{g}=F(\nabla\mathbf{x})$, we can substitute $\mathbf{g}$ in $\|\mathbf{g}-\nabla\mathbf{x}\|^2$ with $F(\nabla\mathbf{x})$, and the sub-problem becomes

$$\min_F\|F(\nabla\mathbf{x})-\nabla\mathbf{x}\|^2 \text{ s.t. } \mathbf{h}_F=\mathbf{h}_r. \tag{14}$$

To solve this sub-problem, by introducing $\mathbf{d}_0=|\nabla\mathbf{x}|$, the standard histogram specification operator [34] can be used to obtain the only feasible monotonic non-parametric transform T which makes the histogram of $T(\mathbf{d}_0)$ the same as $\mathbf{h}_r$. Note that $(x-y)^2 \leq ((-x)-y)^2$ if the signs of x and y are the same. Since $\hat{F}(|\nabla\mathbf{x}|) = T(|\nabla\mathbf{x}|)$, to minimizing the squared error $\|F(\nabla\mathbf{x})-\nabla\mathbf{x}\|^2$, we should require that the sign of $\hat{F}(\nabla\mathbf{x})$

is the same as that of $\nabla\mathbf{x}$. Thus, we define $\hat{F}(\nabla\mathbf{x})$ as

$$\hat{F}(\nabla\mathbf{x}) = \text{sgn}(\nabla\mathbf{x}) T(|\nabla\mathbf{x}|). \qquad (15)$$

Given $\hat{F}(\nabla\mathbf{x})$, we then let $\mathbf{g} = \hat{F}(\nabla\mathbf{x})$.

The proposed iterative histogram specification (IHS) based GHP algorithm is summarized in **Algorithm 1**. It should be noted that, for any gradient based image denoising model, we can easily adapt the proposed GHP to it by simply modifying the gradient term and adding an extra histogram specification operation.

Algorithm 1: Iterative Histogram Specification (IHS) 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 $\mathbf{x}$:
   $$\mathbf{x}^{(k+1/2)} = \mathbf{x}^{(k)} + \delta\left(\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{x}^{(k)}) + \mu\nabla^T(\mathbf{g} - \nabla\mathbf{x}^{(k)})\right)$$
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 $\boldsymbol{\alpha}$:
   $$\boldsymbol{\alpha}_i^{(k+1)} = S_{\lambda/d}\left(\boldsymbol{\alpha}_i^{(k+1/2)} - \boldsymbol{\beta}_i\right) + \boldsymbol{\beta}_i$$
8. Update $\mathbf{x}$
   $$\mathbf{x}^{(k+1)} = \mathbf{D} \circ \boldsymbol{\alpha}^{(k+1)}$$
9. $F(\nabla\mathbf{x}) = \text{sgn}(\nabla\mathbf{x}) T(|\nabla\mathbf{x}|)$
10. $k \longleftarrow k + 1$
11. $\mathbf{x} = \mathbf{x}^{(k)} + \delta\left(\mu\nabla^T(\mathbf{g} - \nabla\mathbf{x}^{(k)})\right)$

The GHP model in Eq. (7) is nonconvex, and thus the proposed algorithm cannot be guaranteed to converge to a global optimum. However, it is empirically found that our GHP algorithm converges rapidly. Fig. 3 shows an example convergence curve of the proposed GHP algorithm on image Bear (in Fig. 2). One can see that GHP converges within 15 20 iterations.
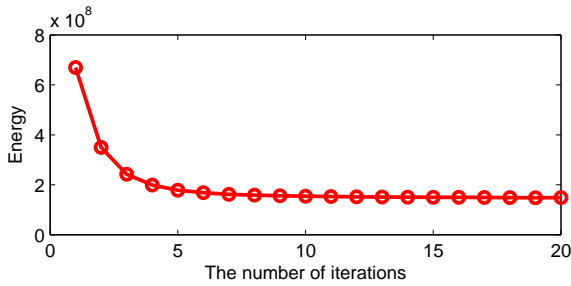

Fig. 3: The convergence curve of the proposed GHP algorithm on image *Bear*.

C. Region-based GHP

The histogram constraint in Eq. (7) is global. If the image consists of different regions with different textures, GHP may generate some false textures in the less textured areas. To address this problem, we can partition the noisy image into several regions, estimate the reference gradient histogram of each region, and then apply GHP to each region for denoising.

As shown in Fig. 4, we suggest two schemes to partition the noisy image, resulting in two region-based GHP variants. The first scheme (Fig. 4(a)), namely S-GHP, is to employ k-means clustering method to roughly partition the image into K homogeneous regions, while the second scheme (Fig. 4(b)), namely B-GHP, simply partitions the noisy image into $K = \sqrt{K} \times \sqrt{K}$ blocks with equal size. Denote by $\{\Omega_1, \ldots, \Omega_k, \ldots, \Omega_K\}$ the partitioned regions. Each region Ω_k has the corresponding reference gradient histogram $\mathbf{h}_{r,k}$, and we have a function F_k to process the pixels within region Ω_k:

$$\min_{F_k} \sum_{(i,j)\in\Omega_k} \left(F_k\left((\nabla\mathbf{x})_{ij}\right) - (\nabla\mathbf{x})_{ij}\right)^2 \text{ s.t. } \mathbf{h}_{F_k} = \mathbf{h}_{r,k}. \qquad (16)$$

We define an indicator function

$$\mathbf{1}_{\Omega_c}(i,j) = \begin{cases} 1, \text{if}(i,j) \in \Omega_k \\ 0, \text{else} \end{cases}. \qquad (17)$$

The $F(\nabla\mathbf{x})$ for S-GHP/B-GHP can then be defined as

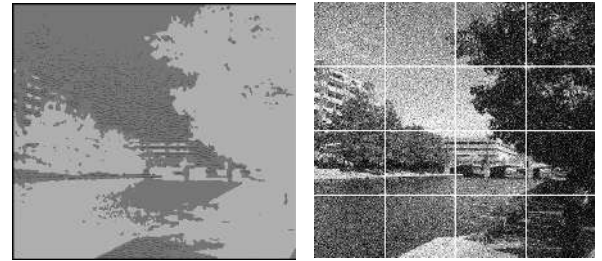$$F(\nabla\mathbf{x}) = \sum_k F_k(\nabla\mathbf{x})\mathbf{1}_{\Omega_k}. \qquad (18)$$

(a)   (b)

Fig. 4: Two image partition schemes. (a) The noisy image is partitioned into K homogeneous regions by k-means clustering. (b) The noisy image is partitioned into $\sqrt{K} \times \sqrt{K}$ blocks.

V. REFERENCE GRADIENT HISTOGRAM ESTIMATION

To apply the model in Eq. (7), we need to know the reference gradient histogram $\mathbf{h}_r$ of original image $\mathbf{x}$. In this section, we propose a regularized deconvolution model to estimate the histogram $\mathbf{h}_r$. Assuming that the pixels in 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 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 $\varepsilon = \nabla v$, ε can then be well approximated by the i.i.d. Gaussian with PDF [38]

$$p_\varepsilon = \frac{1}{2\sqrt{\pi}\sigma} \exp\left(-\frac{\varepsilon^2}{4\sigma^2}\right). \qquad (19)$$

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 $y = x + \varepsilon$. Let p_x be the PDF of x, and p_y be the PDF of y. Since x and ε are independent, the joint PDF $p(x, \varepsilon)$ is

$$p(x,\varepsilon) = p_x \times p_\varepsilon. \qquad (20)$$

Then the PDF p_y is

$$p_y(y = t) = \int_a p_x(x = a) \times p_\varepsilon(\varepsilon = (t - a))\, da. \qquad (21)$$

If we use the normalized histogram $\mathbf{h}_x$ and $\mathbf{h}_y$ to approximate p_x and p_y, we can rewrite Eq. (21) in the discrete domain as:

$$\mathbf{h}_y = \mathbf{h}_x \otimes \mathbf{h}_\varepsilon, \qquad (22)$$

where $\otimes$ denotes the convolution operator. Note that $\mathbf{h}_\varepsilon$ can be obtained by discretizing p_ε, 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:

$$\mathbf{h}_r = \arg\min\nolimits_{\mathbf{h}_x} \left\{ \left\| \mathbf{h}_y - \mathbf{h}_x \otimes \mathbf{h}_\varepsilon \right\|^2 + c \cdot R(\mathbf{h}_x) \right\}, \qquad (23)$$

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. We consider two kinds of constraints on $\mathbf{h}_x$. First, it has been shown that p_x (i.e., the continuous counterpart of $\mathbf{h}_x$) can be approximated by hyper-Laplacian distribution [3], [23], [24]. Considering that the real $\mathbf{h}_x$ might deviate from the hyper-Laplacian distribution to some extent, we only require that $\mathbf{h}_x$ should be close to the hyper-Laplacian distribution:

$$p_x \approx C \cdot \exp\left(-\kappa |x|^\gamma\right), \qquad (24)$$

where C is the normalization factor, γ and κ are the two parameters of the hyper-Laplacian distribution. More specifically, we let $\kappa \in [0.001, 3]$ and $\gamma \in [0.02, 1.5]$. Second, each element of $\mathbf{h}_x$ should be nonnegative. Based on these two constraints, gradient histogram estimation can be formulated as the following regularized deconvolution problem:

$$\mathbf{h}_r = \arg\min\nolimits_{\mathbf{h}_x, C, \kappa, \gamma} \left\| \mathbf{h}_y - \mathbf{h}_x \otimes \mathbf{h}_\varepsilon \right\|^2 + c \left\| \mathbf{h}_x - C \cdot \exp(-\kappa |x|^\gamma) \right\|^2,$$
$$\text{s.t. } \mathbf{h}_x \geq 0, \qquad (25)$$

which can be re-written as:

$$\mathbf{h}_r = \arg\min\nolimits_{\mathbf{h}_x, \mathbf{h}'_x, C, \kappa, \gamma} \left\{ \begin{array}{l} \left\| \mathbf{h}_y - \mathbf{h}_x \otimes \mathbf{h}_\varepsilon \right\|^2 \\ + c \left\| \mathbf{h}_x - C \cdot \exp(-\kappa |x|^\gamma) \right\|^2 \\ + \eta \left\| \mathbf{h}_x - \mathbf{h}'_x \right\|^2 \end{array} \right\}. \qquad (26)$$
$$\text{s.t. } \mathbf{h}'_x \geq 0$$

We iteratively update $\mathbf{h}_x$, $\mathbf{h}'_x$, C, γ, and κ alternatively. Let $\mathbf{h}_0 = C \cdot \exp(-\kappa |x|^\gamma)$, $\mathbf{h}_x$ is updated by

$$\mathbf{h}_x = \frac{\overline{FFT(\mathbf{h}_\varepsilon)} \bullet FFT(\mathbf{h}_y) + cFFT(\mathbf{h}_0) + \eta FFT(\mathbf{h}'_x)}{FFT(\mathbf{h}_\varepsilon) \bullet \overline{FFT(\mathbf{h}_\varepsilon)} + c + \eta}, \qquad (27)$$

where "$\bullet$" denotes the element-wise multiplication, "$\overset{*}{\div}$" denotes the element-wise division, and "$\overline{*}$" denotes the complex conjugate operator. $\mathbf{h}'_x$ is updated by

$$\mathbf{h}'_x(i) = \max\left(\mathbf{h}_x(i), 0\right). \qquad (28)$$

C is updated by

$$C = \frac{\sum_i \exp(-\kappa |i|^\gamma)}{\sum_i \mathbf{h}_x(i)}. \qquad (29)$$

γ and κ are updated based on gradient decent

$$\kappa^{(t+1)} = \kappa^{(t)} + \tau \sum_i C|i|^\gamma \exp(-\kappa^{(t)}|i|^\gamma)\left(C \cdot \exp(-\kappa^{(t)}|i|^\gamma - \mathbf{h}_x(i)\right), \qquad (30)$$

$$\gamma^{(t+1)} = \gamma^{(t)} + \rho \sum_{|i \neq 0|} \left\{ \begin{array}{l} C\kappa^{(t)}|i|^\gamma \ln |i| \exp(-\kappa^{(t)}|i|^\gamma) \\ \cdot \left(C \cdot \exp(-\kappa^{(t)}|i|^\gamma - \mathbf{h}_x(i)\right) \end{array} \right\}. \qquad (31)$$

Fig. 5 shows an example of reference gradient histogram estimation. It can be seen that our method can obtain a good estimation of $\mathbf{h}_x$.

For region based B-GHP and S-GHP, the regularized deconvolution method can be directly applied to each region to estimate the corresponding reference gradient histogram.

VI. EXPERIMENTAL RESULTS

To verify the performance of the proposed GHP based image denoising method, we apply it to ten natural images with various texture structures, whose scenes are shown in Fig. 6. All the test images are gray-scale images with gray level ranging from 0 to 255. We first discuss the parameter setting in our GHP algorithm, and then compare the performance of global based GHP and its region based variants, i.e., B-GHP and S-GHP. Finally, experiments are conducted to validate its performance in comparison with the state-of-the-art denoising algorithms. In the following experiments we set the AWGN standard deviation from 20 to 40 with step length 5.[1]

Fig. 6: Ten test images. From left to right and top to bottom, they are labeled as 1 to 10.

A. Parameter setting

There are 4 parameters in our GHP algorithm and 4 parameters in the reference histogram estimation algorithm. All these parameters are fixed in our experiments.

1) Parameters in the GHP algorithm: The proposed GHP algorithm has two model parameters: λ, and μ. We use the same strategy as in the original NCSR model [14] to determine the value of λ. The parameter μ is introduced to balance the nonlocally centralized sparse representation term and the histogram preservation term. If μ is is set very large, GHP can ensure that the gradient histogram of the denoising result is the same as the reference histogram. Considering that in practice the reference histogram is estimated from the noisy image and there are certain estimation errors, μ cannot be set too big. We empirically set μ to 5 based on our experimental experience.

The GHP algorithm involves two more algorithm parameters: δ and d. Following [14], when the noise standard deviation is less than 30, we set δ to 0.23; when else we set δ to 0.26. Based on [14], [33], to guarantee the convexity of the surrogate function, d should be larger than the spectral norm of dictionary $\mathbf{D}$. Since in our algorithm $\mathbf{D}$ is an adaptively

---

[1]We also evaluated the denoising performance of S-GHP under lower and higher noise standard deviations, i.e., $\sigma = 5, 10, 15$, and $\sigma = 50, 80, 100$. The detailed results can be found in the supplementary file.
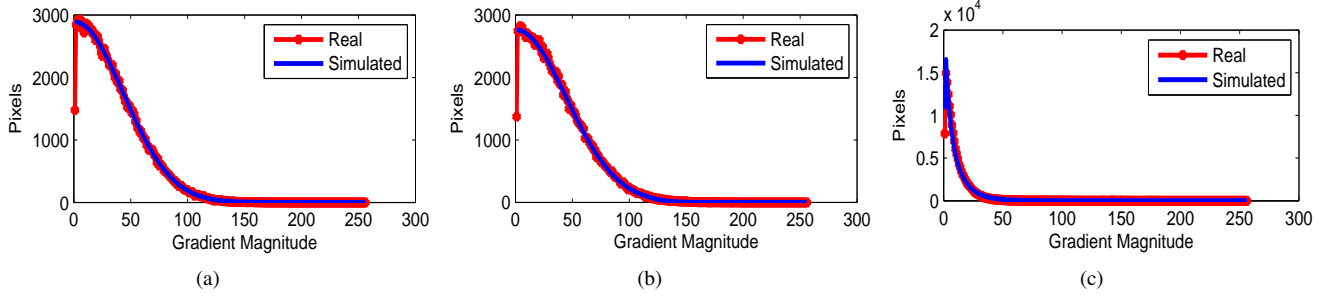
Fig. 5: An example of reference gradient histogram estimation. (a) Real and simulated AWGN gradient histograms (noise standard deviation $\sigma = 30$); (b) real and simulated gradient histograms of noisy image; and (c) real and estimated gradient histograms of the clean image.

selected orthogonal PCA dictionary, any $d \geq 1$ will be fine. According to [14], [33] we choose a little higher d ($d = 3$) for numerical stability of the algorithm.

In summary, compared with NCSR, GHP only introduces one extra model parameter μ, and we set it to 5 by experience in all the experiments.

2) Parameters in reference histogram estimation: Our reference histogram estimation method involves two model parameters, i.e., c and η. Image gradients are generally assumed to follow hyper-Laplacian distributions [2], [3]. We choose a relatively large c value, i.e., $c = 10$, to ensure that the estimated histogram should be close to a hyper-Laplacian distribution. The parameter η is introduced to ensure the non-negative property of the estimated histogram, and a large η value should be set to guarantee that the estimated histogram is non-negative. Thus we also choose a large η value, i.e., $\eta = 10$, in the implementation.

There are also two algorithm parameters, τ and ρ, in our reference histogram estimation method. τ and ρ denote the step sizes in the gradient descent algorithm to update κ and γ, respectively. If the step size is sufficiently small, the gradient descent algorithm would converge to a local optimum [53]. Thus we set two smaller values to τ and ρ, i.e., $\tau = 0.01$ and $\rho = 0.01$.

In Fig. 5, we have shown an example of gradient histogram estimation, and it can be seen that the estimated gradient histogram is very close to the ground-truth. Table I lists the K-L divergence between the estimated and the ground-truth (obtained using the noiseless image) gradient histograms for the ten test images with different standard deviations of noise. From Table I, one can see that the average K-L divergence is less than 0.1 and the standard deviation is less than 0.08, indicating that the proposed gradient histogram estimation method can obtain satisfactory estimation results.

B. Comparison between the three GHP variants

By setting the AWGN standard deviation $\sigma \in \{20, 25, 30, 35, 40\}$, we evaluate the three variants of the proposed method, i.e., GHP, B-GHP, and S-GHP, in terms of PSNR and the perceptual quality index SSIM [39]. To evaluate if the proposed reference gradient histogram estimation method is effective for the final noise removal performance, we couple GHP with both the estimated

TABLE I: The K-L divergence between the estimated and ground-truth gradient histograms.

σ	20	25	30	35	40
1	0.098	0.121	0.138	0.118	0.095
2	0.036	0.036	0.027	0.020	0.014
3	0.094	0.089	0.086	0.091	0.152
4	0.014	0.017	0.014	0.014	0.017
5	0.023	0.030	0.051	0.064	0.128
6	0.233	0.261	0.254	0.218	0.151
7	0.066	0.071	0.085	0.061	0.029
8	0.123	0.167	0.192	0.161	0.103
9	0.058	0.073	0.066	0.042	0.026
10	0.015	0.028	0.038	0.071	0.140
Avg. ± Std.	0.076 ± 0.067	0.089 ± 0.076	0.095 ± 0.077	0.086 ± 0.064	0.086 ± 0.058

gradient histogram and the ground truth gradient histogram and compare the outputs. Table II lists the PSNR and SSIM values on the ten test images. One can see that GHP achieves similar PSNR/SSIM values by using the estimated gradient histogram and the ground truth gradient histogram.

We then compare the performance of GHP, B-GHP and S-GHP on the ten test images. The PSNR and SSIM indices are also listed in Table III. One can see that the region-based GHP methods, i.e., B-GHP and S-GHP, generally achieve better results than GHP in terms of both PSNR and SSIM. Fig. 7 shows an example of the denoising outputs by GHP, B-GHP, and S-GHP on test image 6. Since natural images often consist of regions with different textures and GHP uses the global gradient histogram for texture enhanced denoising, sometimes false textures can be generated in the less textured areas by GHP. By simply partitioning the image into regular blocks, the block based B-GHP can reduce the possibility of generating false textures. By segmenting the image into texture homogeneous regions, S-GHP can further achieve better denoising results than B-GHP in terms of PSNR/SSIM measures and subjective visual quality, as demonstrated in Fig. 7.

C. Comparison with the State-of-the-Arts

We then compare S-GHP with some state-of-the-art denoising methods, including shape-adaptive PCA based BM3D (SAPCA-BM3D) [11], the learned simultaneously sparse coding (LSSC) [12] and the NCSR [14] methods. The codes of

TABLE II: The PSNR (dB) and SSIM results of GHP with the estimated gradient histogram (GHP-E) and with the ground truth gradient histogram (GHP-G).

	σ = 20		σ = 25		σ = 30		σ = 35		σ = 40	
	GHP-G	GHP-E	GHP-G	GHP-E	GHP-G	GHP-E	GHP-G	GHP-E	GHP-G	GHP-E
1	**30.67**	30.59	**29.53**	29.40	**28.62**	28.47	**27.91**	27.78	**27.29**	27.01
	0.870	0.866	**0.843**	0.833	**0.817**	0.808	**0.800**	0.799	**0.781**	0.779
2	**27.93**	27.90	**26.82**	26.72	**26.00**	25.87	**25.43**	25.28	**24.90**	24.62
	0.813	0.808	**0.769**	0.761	**0.731**	0.725	**0.699**	0.697	**0.668**	0.663
3	28.13	**28.15**	**27.19**	27.12	**26.42**	26.34	**25.77**	25.71	**25.22**	25.12
	0.758	0.757	**0.722**	0.720	**0.692**	0.690	**0.662**	0.661	**0.640**	0.638
4	**26.69**	26.59	**25.51**	25.41	**24.55**	24.46	**23.93**	23.84	**23.31**	23.16
	0.803	0.795	**0.757**	0.748	**0.715**	0.708	**0.677**	0.676	**0.643**	0.640
5	**30.65**	30.54	**29.68**	29.52	**28.87**	28.61	**28.38**	28.07	**27.83**	27.26
	0.805	0.802	**0.773**	0.769	**0.743**	0.737	**0.723**	0.719	**0.701**	0.693
6	**28.54**	28.39	**27.24**	27.13	**26.22**	26.12	**25.52**	25.44	**24.90**	24.76
	0.887	0.866	**0.854**	0.833	**0.820**	0.804	**0.799**	0.797	**0.772**	0.771
7	**30.14**	30.07	**29.11**	28.99	**28.33**	28.15	**27.79**	27.58	**27.29**	26.93
	0.839	0.834	**0.804**	0.800	**0.774**	0.769	**0.752**	0.750	**0.729**	0.726
8	**31.43**	31.27	**30.33**	30.14	**29.47**	29.23	**28.83**	28.56	**28.26**	27.82
	0.893	0.884	**0.870**	0.858	**0.848**	0.835	**0.840**	0.837	**0.824**	0.821
9	**27.45**	27.31	**26.26**	26.12	**25.33**	25.21	**24.61**	24.52	**24.01**	23.88
	0.819	0.806	**0.779**	0.764	**0.743**	0.730	**0.710**	0.709	**0.681**	0.680
10	**30.93**	30.83	**29.95**	29.75	**29.19**	28.85	**28.76**	28.32	**28.26**	27.50
	0.813	0.811	**0.780**	0.776	**0.752**	0.745	**0.734**	0.727	**0.714**	0.703
Avg.	**29.26**	29.16	**28.16**	28.03	**27.30**	27.13	**26.69**	26.51	**26.13**	25.81
	0.830	0.823	**0.795**	0.786	**0.764**	0.755	**0.740**	0.737	**0.715**	0.711

TABLE III: The PSNR (dB) and SSIM results of GHP, B-GHP, and S-GHP.

	σ = 20			σ = 25			σ = 30			σ = 35			σ = 40		
	GHP	B-GHP	S-GHP	GHP	B-GHP	S-GHP	GHP	B-GHP	S-GHP	GHP	B-GHP	S-GHP	GHP	B-GHP	S-GHP
1	30.59	30.53	**30.60**	29.40	29.40	**29.47**	28.47	28.55	**28.60**	27.78	27.82	**27.85**	27.01	**27.23**	27.22
	0.866	**0.869**	0.869	0.833	0.841	**0.842**	0.808	**0.818**	0.818	0.799	0.796	**0.801**	0.779	0.779	**0.781**
2	27.90	27.89	**27.97**	26.72	26.79	**26.88**	25.87	25.98	**26.07**	25.28	25.42	**25.43**	24.62	24.86	**24.87**
	0.808	**0.815**	0.814	0.761	**0.771**	0.771	0.725	**0.735**	0.734	0.697	0.695	**0.698**	0.663	0.665	**0.666**
3	28.15	28.07	**28.17**	27.12	27.12	**27.25**	26.34	26.36	**26.43**	25.71	25.71	**25.75**	25.12	25.19	**25.21**
	0.757	0.753	0.753	**0.720**	0.720	0.720	**0.690**	0.688	0.688	**0.661**	0.657	0.659	**0.638**	0.636	0.636
4	26.59	26.60	**26.72**	25.41	25.47	**25.53**	24.46	24.58	**24.67**	23.84	23.93	**23.97**	23.16	23.31	**23.34**
	0.795	**0.801**	0.801	0.748	**0.759**	0.756	0.708	**0.719**	0.718	**0.676**	0.673	0.676	**0.640**	0.640	0.639
5	30.54	30.63	**30.65**	29.52	29.66	**29.77**	28.61	28.86	**28.99**	28.07	28.29	**28.38**	27.26	27.77	**27.88**
	0.802	**0.807**	0.807	0.769	**0.775**	0.774	0.737	0.745	**0.751**	0.719	0.716	**0.723**	0.693	0.696	**0.704**
6	28.39	28.36	**28.46**	27.13	27.08	**27.19**	26.12	26.09	**26.26**	25.44	25.41	**25.58**	24.76	24.77	**24.96**
	0.866	0.880	**0.883**	0.833	**0.848**	0.848	0.804	0.819	**0.820**	0.797	0.793	**0.802**	0.771	0.768	**0.775**
7	30.07	30.06	**30.22**	28.99	29.00	**29.12**	28.15	28.22	**28.39**	27.58	27.66	**27.81**	26.93	27.20	**27.30**
	0.834	0.839	**0.843**	0.800	0.805	**0.807**	0.769	0.777	**0.780**	0.750	0.745	**0.754**	0.726	0.725	**0.730**
8	31.27	31.15	**31.34**	30.14	29.86	**30.21**	29.23	29.14	**29.42**	28.56	28.59	**28.80**	27.82	28.16	**28.21**
	0.884	0.887	**0.895**	0.858	0.858	**0.872**	0.835	0.841	**0.853**	0.837	0.830	**0.843**	0.821	0.822	**0.829**
9	27.31	27.35	**27.40**	26.12	26.16	**26.22**	25.21	25.26	**25.31**	24.52	**24.59**	24.57	23.88	23.99	**24.05**
	0.806	0.817	**0.818**	0.764	0.777	**0.777**	0.730	0.742	**0.744**	**0.709**	0.707	0.708	0.680	0.680	**0.684**
10	30.83	30.96	**30.98**	29.75	**29.94**	29.93	28.85	**29.18**	29.15	28.32	28.67	**28.74**	27.50	**28.20**	28.15
	0.811	**0.816**	0.815	0.776	**0.783**	0.782	0.745	**0.756**	0.754	0.727	0.729	**0.736**	0.703	**0.711**	0.711
Avg.	29.16	29.16	**29.25**	28.03	28.05	**28.16**	27.13	27.22	**27.33**	26.51	26.61	**26.69**	25.81	26.07	**26.12**
	0.823	0.828	**0.830**	0.786	0.794	**0.795**	0.755	0.764	**0.766**	0.737	0.734	**0.740**	0.711	0.712	**0.716**

all the competing methods are provided by the authors and we used the recommended parameters by the authors. On a PC with two Intel CPUs (1.86 GHz) and 2GB RAM, under Matlab 2011a programming environment, GHP spends 2383s to process a 512 × 512 image, which is similar to B-GHP (2388s), S-GHP (2386s) and NCSR (2445s) but is much faster than LSSC (4287s). SAPCA-BM3D spends 420s to process a 512 × 512 image but it should be noted that SAPCA-BM3D is mainly implemented by C programming language.

The quantitative experimental results by competing methods are shown in Tables IV. One can see that the proposed S-GHP method has similar PSNR/SSIM measures to SAPCA-BM3D, LSSC and NCSR. Nonetheless, the goal of our GHP method is to preserve and enhance the image texture structures, and we further compare the visual quality of the denoised images

by these methods. Fig. 8 shows the denoising results of noisy image 7 with noise standard deviation σ = 30. In this image, there are different texture regions, such as sky, tree, water and building. We can see that SAPCA-BM3D, LSSC and NCSR smooth too much the textures in tree, water and building areas, while LSSC introduces some artifacts in the smooth sky area. Though these methods have good PSNR and even SSIM indices, the denoised images by them look somewhat unnatural. In contrast, S-GHP preserves much better the fine textures in areas of tree and water, making the denoised image look more natural and visually pleasant. It should also be noted that, the better visual quality in areas of tree and water of S-GHP might not always result in higher PSNR or SSIM values on the close-ups.

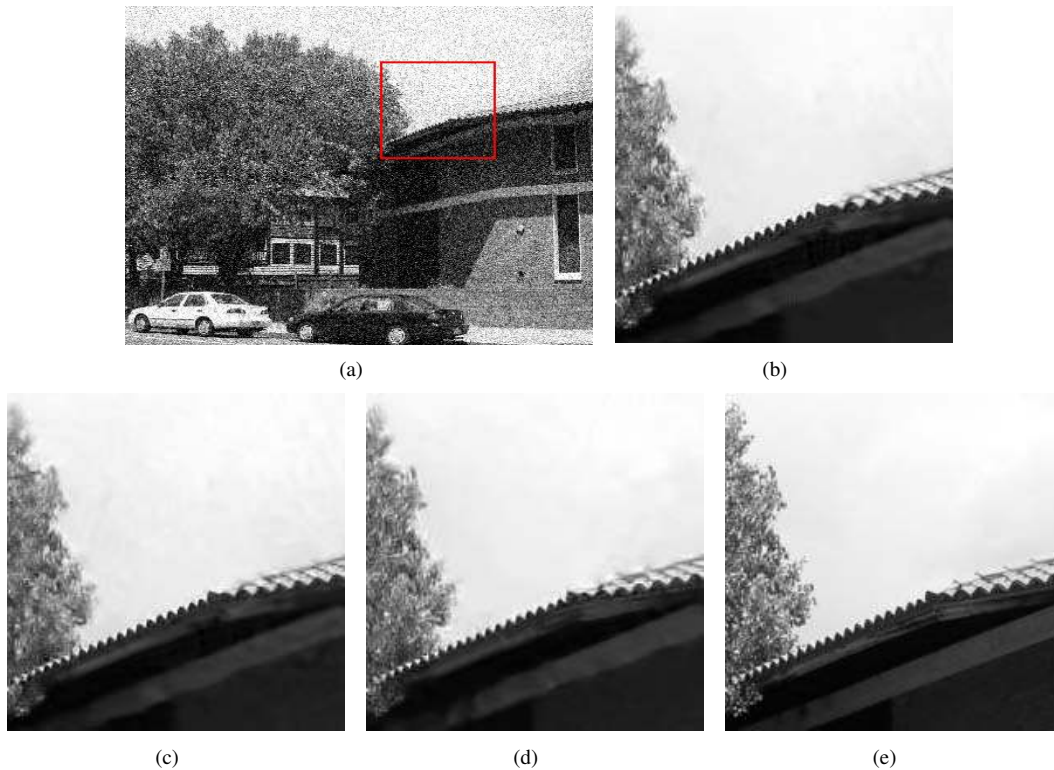Fig. 9 shows the denoising results of image 1 with noise

Fig. 7: (a) The noisy test image 6 (noise standard deviation $\sigma = 35$). From (b) to (e): the zoom-in denoising results by GHP, B-GHP and S-GHP, and the ground truth.

standard deviation $\sigma = 30$. This image consists of texture regions like cloud, tree, and water. We can see that SAPCA-BM3D, LSSC and NCSR tend to over-smooth the textures in tree and water areas. In contrast, S-GHP preserves much better the fine texture in tree and water areas, while making the cloud area look more natural.

Due to the limit of space, we do not show the full denoised images in the main paper. However, examples of full denoised images are given in the supplementary file. Similar conclusions can be made; that is, S-GHP obtains more natural denoising results and preserves better the fine textures. We also evaluated the denoising performance of competing methods under lower (i.e., $\sigma = 5, 10, 15$) and higher noise standard deviations (i.e., $\sigma = 50, 80, 100$). Please refer to the supplementary file for the detailed results. In terms of average PSNR and SSIM, S-GHP is comparable to SAPCA-BM3D, LSSC and NCSR. In terms of visual quality, when the noise standard deviation is high, S-GHP can preserve better the textures and strong edges than the competing methods. When the noise standard deviation is low, S-GHP can preserve better image fine textures.

The proposed S-GHP method has similar overall PSNR/SSIM results to the state-of-the-arts and it leads to better visual quality of the denoised images. In terms of visual quality, it improves much the textured areas. However, the improvement in visual quality may not result in PSNR and SSIM improvements. S-GHP enforces the statistical distribution of image gradients to be close to the reference histogram, but it cannot guarantee that the restored image will be close to the real image in terms of PSNR or SSIM.

Using image 6 with AWGN of standard deviation 30 as an example, we calculate the local PSNR maps (with a sliding window of size 41×41) for the denoised images by SAPCA-BM3D and S-GHP. Denote by $\mathbf{p}_{BM3D}$ and $\mathbf{p}_{GHP}$ the two local PSNR maps of SAPCA-BM3D and S-GHP, respectively. In Fig. 10, we show the original image, the denoised images by SAPCA-BM3D and S-GHP, and the difference map ($\mathbf{p}_{BM3D} - \mathbf{p}_{GHP}$). The white values indicate the areas where $\mathbf{p}_{BM3D}$ is higher than $\mathbf{p}_{GHP}$, while the dark values indicate the areas where $\mathbf{p}_{GHP}$ is higher than $\mathbf{p}_{BM3D}$. One can see that, in terms of PSNR, S-GHP outperforms SAPCA-BM3D in many smooth areas, while SAPCA-BM3D outperforms S-GHP in many textured areas. However, the denoised image by S-GHP has better visual quality than the one by SAPCA-BM3D in most textured and smooth areas.

D. Subjective Evaluation

How to evaluate the perceptual quality of an image is a very challenging problem. Though many image quality assessment (IQA) indices (e.g., SSIM [39] and FSIM [51]) have been developed and they can well predict the perceptual quality of images with a single type of distortion such as noise corruption, Gaussian blur and compression artifacts, they are still far from satisfying to faithfully evaluate the subjective quality of denoised images, whose distortion is much more complex. This is also why the denoised images by our methods have better visual quality, but their SSIM indices are similar to the denoised images by other methods.

TABLE IV: The PSNR (dB) and SSIM values of SAPCA-BM3D (BM3D), LSSC, NCSR and S-GHP.

| | $\sigma = 20$ | | | | $\sigma = 25$ | | | | $\sigma = 30$ | | | | $\sigma = 35$ | | | | $\sigma = 40$ | | | |
	BM3D	LSSC	NCSR	S-GHP	BM3D	LSSC	NCSR	S-GHP	BM3D	LSSC	NCSR	S-GHP	BM3D	LSSC	NCSR	S-GHP	BM3D	LSSC	NCSR	S-GHP
1	**30.83**	30.69	30.59	30.60	**29.66**	29.56	29.46	29.47	**28.75**	28.62	28.58	28.60	**28.02**	27.91	27.76	27.85	**27.41**	27.32	27.19	27.22
	0.876	0.872	0.869	0.869	**0.849**	0.846	0.843	0.842	**0.825**	0.820	0.820	0.818	**0.803**	0.800	0.793	0.801	**0.784**	0.781	0.776	0.781
2	**28.07**	27.98	27.91	27.97	**26.99**	26.94	26.87	26.88	**26.18**	26.14	26.08	26.07	**25.54**	25.51	25.37	25.43	**25.02**	24.98	24.87	24.87
	0.817	0.815	0.807	0.814	**0.773**	**0.773**	0.764	0.771	**0.734**	**0.734**	0.727	**0.734**	0.699	**0.700**	0.681	0.698	0.668	**0.670**	0.651	0.666
3	28.39	**28.46**	28.11	28.17	27.43	**27.52**	27.16	27.25	26.66	**26.66**	26.39	26.43	26.01	**26.03**	25.64	25.75	25.46	**25.47**	25.10	25.21
	0.755	**0.762**	0.736	0.753	0.721	**0.728**	0.702	0.720	0.692	**0.696**	0.675	0.688	0.667	**0.670**	0.640	0.659	**0.647**	**0.647**	0.621	0.636
4	**26.86**	26.75	26.65	26.72	**25.68**	25.61	25.52	25.53	**24.79**	24.76	24.64	24.67	**24.08**	24.06	23.84	23.97	**23.50**	23.48	23.26	23.34
	0.803	**0.803**	0.782	0.801	**0.758**	**0.758**	0.737	0.756	0.715	0.717	0.697	**0.718**	0.677	**0.678**	0.640	0.676	0.641	**0.643**	0.604	0.639
5	**30.88**	30.75	30.64	30.65	**29.96**	29.81	29.68	29.77	**29.21**	29.04	28.91	28.99	**28.58**	28.41	28.27	28.38	**28.06**	27.90	27.76	27.88
	0.812	0.809	0.802	0.807	**0.780**	0.776	0.770	0.774	**0.754**	0.744	0.742	0.751	**0.730**	0.718	0.710	0.723	**0.709**	0.696	0.690	0.704
6	**28.59**	28.47	28.49	28.46	**27.32**	27.26	27.24	27.19	**26.35**	26.33	26.30	26.26	**25.59**	25.59	25.49	25.58	24.97	**24.98**	24.90	24.96
	0.888	0.883	0.882	0.883	**0.856**	0.850	0.851	0.848	0.824	**0.825**	0.820	0.820	0.794	0.795	0.788	**0.802**	0.765	0.769	0.761	**0.775**
7	30.17	30.18	30.13	**30.22**	29.14	**29.18**	29.14	29.12	28.35	**28.40**	28.38	28.39	27.71	**27.81**	27.71	**27.81**	27.18	**27.32**	27.22	27.30
	0.839	0.840	0.833	**0.843**	0.803	**0.807**	0.799	**0.807**	0.771	0.775	0.770	**0.780**	0.744	0.751	0.738	**0.754**	0.721	0.729	0.717	**0.730**
8	**31.58**	31.38	31.41	31.34	**30.48**	30.33	30.35	30.21	**29.64**	29.54	29.52	29.42	**28.94**	28.86	28.79	28.80	**28.37**	28.32	28.24	28.21
	0.900	0.894	0.897	0.895	**0.879**	0.872	0.877	0.872	**0.861**	0.858	0.860	0.853	**0.843**	0.840	0.841	**0.843**	0.828	0.826	0.827	**0.829**
9	**27.58**	**27.58**	27.34	27.40	26.37	**26.40**	26.18	26.22	25.44	**25.48**	25.31	25.31	24.73	**24.77**	24.47	24.57	24.15	**24.19**	23.92	24.05
	0.821	**0.822**	0.804	0.818	0.778	**0.782**	0.764	0.777	0.740	**0.748**	0.729	0.744	0.707	**0.716**	0.683	0.708	0.677	**0.687**	0.655	0.684
10	**31.23**	31.04	30.98	30.98	**30.28**	30.08	30.03	29.93	**29.53**	29.36	29.30	29.15	**28.92**	28.75	28.76	28.74	**28.42**	28.24	28.28	28.15
	0.823	0.818	0.813	0.815	**0.791**	0.787	0.781	0.782	**0.763**	0.755	0.755	0.754	**0.740**	0.732	0.728	0.736	**0.721**	0.712	0.710	0.711
Avg.	**29.42**	29.33	29.23	29.25	**28.33**	28.27	28.16	28.16	**27.49**	27.43	27.34	27.33	**26.81**	26.77	26.61	26.69	**26.25**	26.22	26.07	26.12
	0.833	0.832	0.823	0.830	**0.799**	0.798	0.789	0.795	**0.768**	0.767	0.760	0.766	**0.740**	**0.740**	0.724	**0.740**	0.716	0.716	0.701	0.716
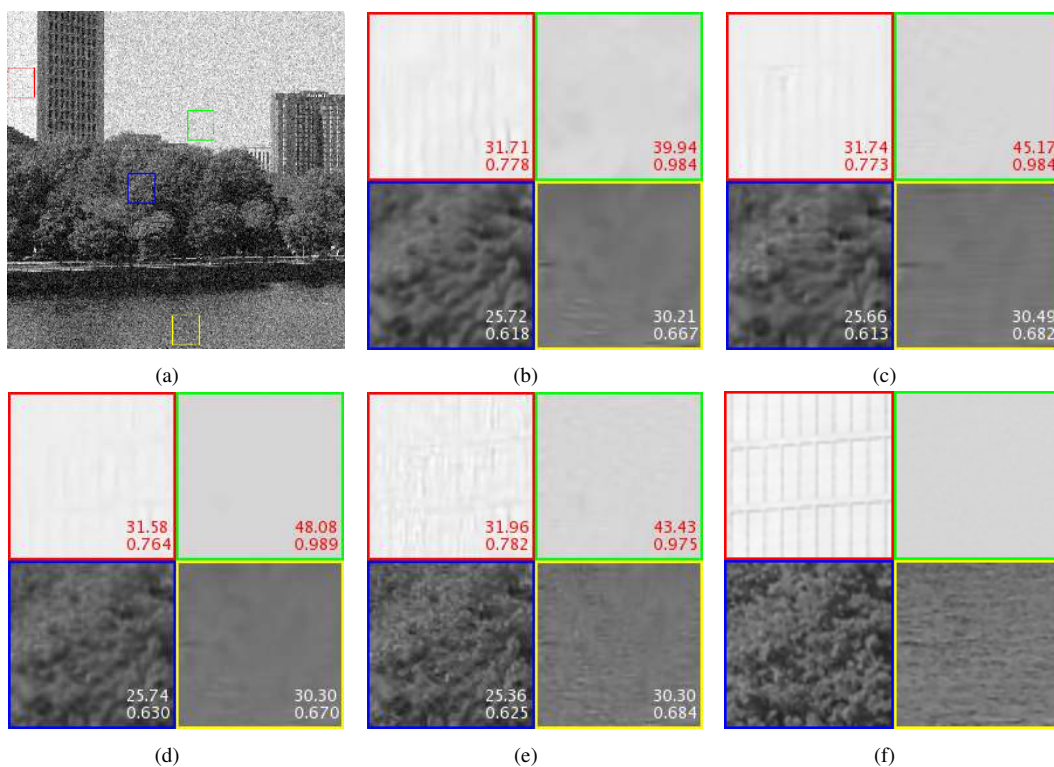
Fig. 8: Denoising results on image 7. (a) Noisy image with AWGN of standard deviation 30; cropped and zoom-in denoised images by (b) SAPCA-BM3D [11], (c) LSSC [12], (d) NCSR [14], and (e) S-GHP; (f) ground truth. The PSNR and SSIM values are shown in the close-ups.
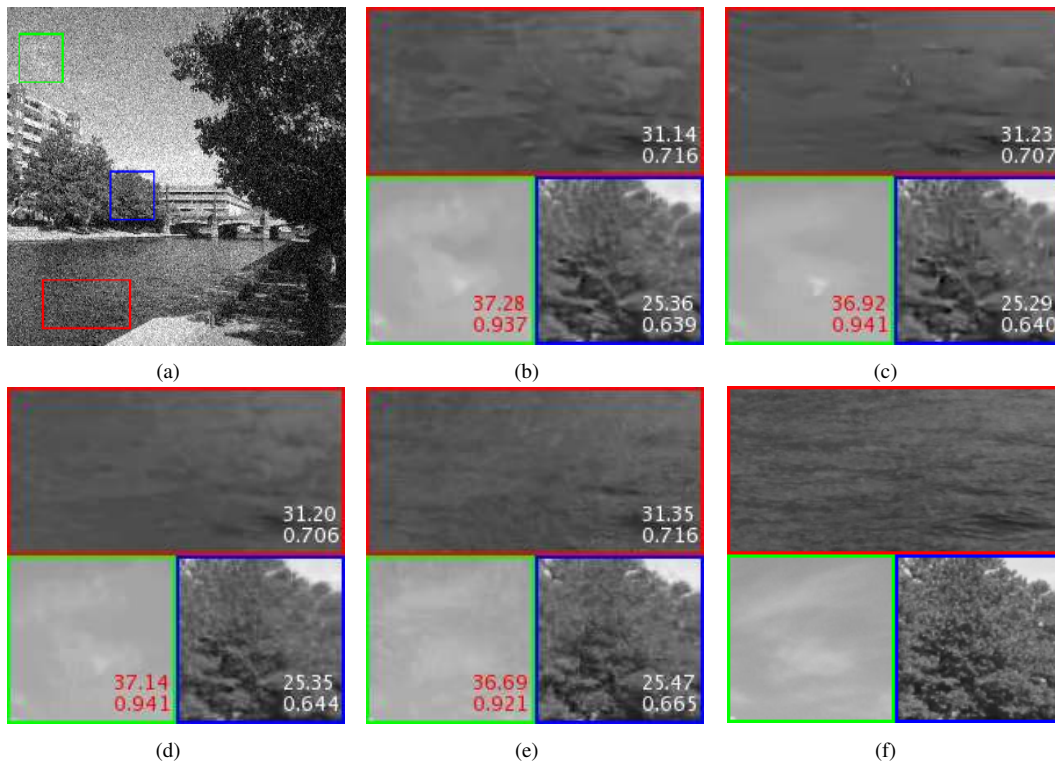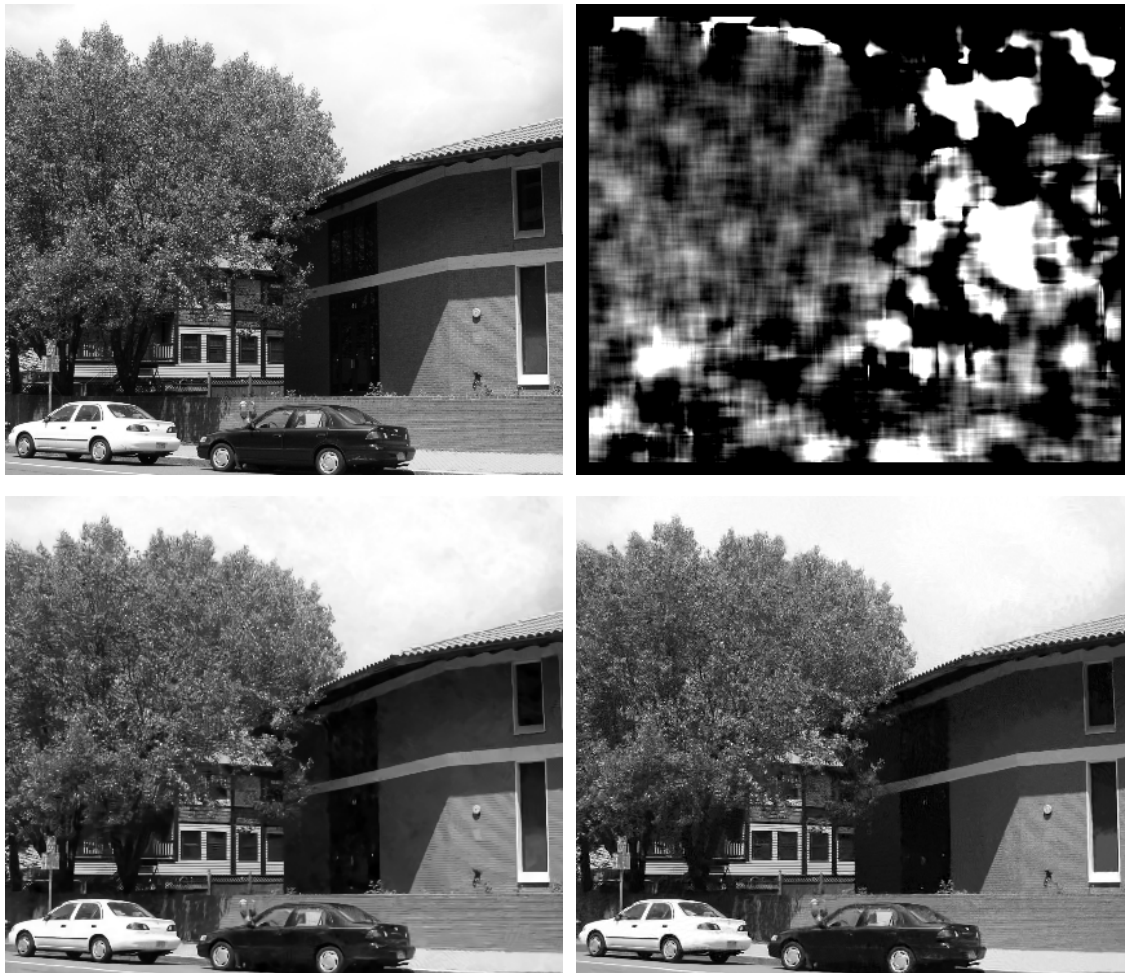
Fig. 10: From left to right and from top to bottom: original image, the difference PSNR map ($\mathbf{p}_{BM3D} - \mathbf{p}_{GHP}$), the denoised images by SAPCA-BM3D and S-GHP.
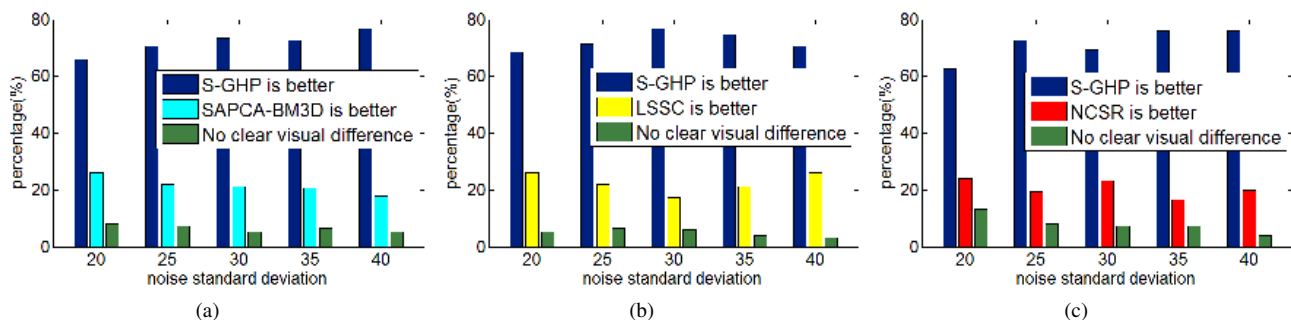

Fig. 11: The distributions of subjective evaluation outputs. (a) S-GHP vs. SAPCA-BM3D; (b) S-GHP vs. LSSC; (c) S-GHP vs. NCSR.

ACKNOWLEDGEMENT

REFERENCES

[1] R. Fergus, B. Singh, A. Hertzmann, S. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," in *Proc. ACM SIGGRAPH*, pp. 787-794, 2006.

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

[3] D. Krishnan, R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," in *Proc. Neural Inf. Process. Syst.*, pp. 1033-1041, 2009.

[4] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1-4, pp. 259-268, Nov. 1992.

[5] M. Wainwright and S. Simoncelli, "Scale mixtures of gaussians and the statistics of natural images," in *Proc. Neural Inf. Process. Syst.*, vol. 12, pp.855-861, 1999.

[6] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using a scale mixture of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338-1351, Nov. 2003.

[7] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736-3745, Dec. 2006.

[8] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838-1857, Jul. 2011.

[9] A. Buades, B. Coll, and J. Morel, "A review of image denoising methods, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490-530, 2005.

[10] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080-2095, Aug. 2007.

[11] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "From local kernel to nonlocal multiple-model image denoising," *Int. J. Computer Vision*, vol. 86, no. 1, pp. 1-32, Jan. 2010.

[12] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. Int. Conf. Comput. Vis.*, pp. 2272-2279, Sept. 29 2009-Oct. 2 2009.

[13] W. Dong, L. Zhang, and G. Shi, "Centralized sparse representation for image restoration," in *Proc. Int. Conf. Comput. Vis.*, pp. 1259-1266, 6-13 Nov. 2011.

[14] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620-1630, Apr. 2013.

[15] E. Hadjidemetriou, M. D. Grossberg, and S. K. Nayar, "Multiresolution histograms and their use for recognition," *IEEE. Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 7, pp. 831-847, Jul. 2004.

[16] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *Int. J. Computer Vision*, vol. 62, no. 1-2, pp. 61-81, Apr. 2005.

[17] M. Varma and A. Zisserman, "Classifying images of materials: achieving viewpoint and illumination independence," in *Proc. Eur. Conf. Comput. Vis.*, pp. 255-271, 2002.

[18] W. Zuo, L. Zhang, C. Song, and D. Zhang, "Texture Enhanced Image Denoising via Gradient Histogram Preservation," in *Proc. Int. Conf. Compu. Vis. Pattern Recognit.*, 2013.

[19] K. Suzuki, I. Horiba, and N. Sugie, "Efficient approximation of neural filters for removing quantum noise from images," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1787-1799, Jul. 2002.

[20] V. Jain and H. Seung, "Natural image denoising with convolutional networks," in *Proc. Neural Inf. Process. Syst.*, pp. 769-776, 2008.

[21] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: can plain neural networks compete with BM3D?," in *Proc. Int. Conf. Compu. Vis. Pattern Recognit.*, pp. 2392-2399, 16-21 June 2012.

[22] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," in *Proc. Int. Conf. Compu. Vis. Pattern Recognit.*, pp. 2657-2664, 20-25 June 2011.

[23] T. S. Cho, C. L. Zitnick, N. Joshi, S. B. Kang, R. Szeliski, and W. T. Freeman, "Image restoration by matching gradient distributions," *IEEE. Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 683-694, Apr. 2012.

[24] T. S. Cho, N. Joshi, C. L. Zitnick, S. B. Kang, R. Szeliski, and W. T. Freeman, "A content-aware image prior," in *Proc. Int. Conf. Compu. Vis. Pattern Recognit.*, pp. 169-176, 13-18 June 2010.

[25] N. Joshi, C. L. Zitnick, R. Szeliski, and D. Kriegman, "Image deblurring and denoising using color priors," in *Proc. Int. Conf. Compu. Vis. Pattern Recognit.*, pp. 1550-1557, 20-25 June 2009.

[26] M. Welling, G. Hinton, and S. Osindero, "Learning sparse topographic representations with products of Student-t distributions," in *Proc. Neural Inf. Process. Syst.*, 9-14 Dec. 2002.

[27] S. Roth and M. J. Black, "Fields of experts: a framework for learning image priors," in *Proc. Int. Conf. Compu. Vis. Pattern Recognit.*, vol. 2, pp. 860-867, 20-25 June 2005.

[28] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: a dictionary-learning algorithm for the analysis sparse model," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 661-677, Feb. 2013.

[29] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE. Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 4, pp. 791-804, Apr. 2012.

[30] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. Int. Conf. Comput. Vis.*, pp. 479-486, 6-13 Nov. 2011.

[31] A. Levin, B. Nadler, F. Durand, and W. T. Freeman, "Patch complexity, finite pixel correlations and optimal denoising," in *Proc. Eur. Conf. Comput. Vis.*, 2012.

[32] J. Jancsary, S. Nowozin, and C. Rother, "Loss-specific training of non-parametric image restoration models: a new state of the art," in *Proc. Eur. Conf. Comput. Vis.*, 2012.

[33] I. Daubechies, M. Defriese, and C. DeMol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413-1457, Nov. 2004.

[34] R. C. Gonzalez and R. E. Woods, "Digital Image Processing," *Publishing House of Electronics Industry, Beijing*, 2005.

[35] L. Lin, X. Liu, and S.-C. Zhu, "Layered graph matching with composite cluster sampling," *IEEE. Trans. Pattern Anal. Mach. Intell.*, vol 32, no. 8, pp. 1426-1442, Aug. 2010.

[36] L. Lin, X. Liu, S. Peng, H. Chao, Y. Wang, and B. Jiang, "Object categorization with sketch representation and generalized samples," *Pattern Recogn.*, vol. 45, no. 10, pp. 3648-3660, 2012.

[37] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. H. Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Comp. Vis. Graph. Image Process.*, vol. 39, no. 3, pp. 355-368, Sep. 1987.

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

[39] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[40] J. Bioucas-Dias, and M. Figueiredo, "Multiplicative noise removal using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1720-1730, July 2010.

[41] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345 C 2356, Sept. 2010.

[42] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681-695, Mar. 2011.

[43] B. Julesz, "Textons, the elements of texture perception, and their interactions," *Nature*, vol. 290, pp. 91-97, Mar. 1981.

[44] X. Liu and D. Wang, "A spectral histogram model for texton modeling and texture discrimination," *Vision Research*, vol. 42, no. 23, pp. 2617C2634, Oct. 2002.

[45] J. J. Peissig, M. E. Young, E. A. Wasserman, and I. Biederman, "The role of edges in object recognition by pigeons," *Perception*, vol. 34, pp. 1353 - 1374, 2005.

[46] C. L. Zitnick and D. Parikh, "The role of image understanding in contour detection," *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 622 - 629, 16-21 June 2012.

[47] B. Zhang, J. M. Fadili, and J. L. Starck, "Wavelets, ridgelets, and curvelets for Poisson noise removal," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1093 - 1108, July 2008.

[48] F. J. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 34, pp. 246C254, Dec. 1948.

[49] M. Mäkitalo and A. Foi, "Optimal inversion of the Anscombe transformation in low-count Poisson image denoising," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 99 - 109, Jan. 2011.

[50] L. Zhang, W. Dong, D. Zhang, and G. Shi, "Two-stage Image Denoising by Principal Component Analysis with Local Pixel Grouping," *Pattern Recogn.*, vol. 43, no. 4, pp. 1531-1549, Apr. 2010.

[51] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378C2386, Aug. 2011.

[52] M. Zontak, I. Mosseri, and M. Irani, "Separating signal from noise using patch recurrence across scales," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[53] J. A. Snyman, *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*, Springer, 2005.