Non-Local Extension of Total Variation Regularization for Image Restoration

Hangfan Liu[1], Ruiqin Xiong[1], Siwei Ma[1], Xiaopeng Fan[2], Wen Gao[1]
[1]Institute of Digital Media, Peking University, Beijing 100871, China
[2]Department of Computer Science, Harbin Institute of Technology, Harbin 150001, China
Email: {liuhf, rqxiong, swma, wgao}@pku.edu.cn, fxp@hit.edu.cn

Abstract—**Total-variation (TV) regularization is widely adopted in image restoration problems to exploit the feature that natural images are smooth with small gradient values at most regions. Basic TV method assumes identical zero-mean Laplacian distribution for the gradients at all pixels. However, for real-world images, the statistics of gradients may not be stationary, and the zero-mean assumption of gradients may not be valid either for a specific pixel. This paper presents a non-local extension of TV regularization for image restoration, called Non-Local Gradient Sparsity Regularization (NGSR). The NGSR model employs a separate gradient value distribution for each pixel. To figure out the distribution parameters, the NGSR method exploits a set of patches which are similar to the patch centered at current pixel and estimates the distribution parameter adaptively. Experimental results demonstrate that the proposed NGSR outperforms traditional TV remarkably for image restoration.**

I. INTRODUCTION

Image restoration is a fundamental problem in image processing. It aims to recover a high quality original image from its corrupted observation, generally formulated as:

$$\boldsymbol{z} = H\boldsymbol{u} + \boldsymbol{n}, \quad (1)$$

where $\boldsymbol{u}$ and $\boldsymbol{z}$ are lexicographically stacked representation of the original image and the noisy observation, respectively, $\boldsymbol{n}$ is additive Gaussian white noise, and the matrix H stands for the degradation operation in the observation process, such as convolution and down-sampling, etc.

Based on Bayesian framework, image restoration is usually transformed into the following minimization problem:

$$\min_{\boldsymbol{u}} \frac{\mu}{2} \|H\boldsymbol{u} - \boldsymbol{z}\|_2^2 + \varphi(\boldsymbol{u}). \quad (2)$$

Here $\|H\boldsymbol{u} - \boldsymbol{z}\|_2^2$ is the l_2-norm data-fidelity term measuring the likelihood that a specific image $\boldsymbol{u}$ produces the observation $\boldsymbol{z}$, $\varphi(\boldsymbol{u})$ is the regularization term or prior knowledge term representing the prior probability of a specific image $\boldsymbol{u}$, μ is a regularization parameter controlling the trade-off between the two competing terms. It is well known that, as a way to describe the characteristic of the original images to restore, the prior image model plays a very important role in image restoration. Typical image prior models assume that the images can be sparsely represented in the transform domain, e.g. in DCT [1] or wavelet [2] domain, or on a trained dictionary [3]. Total-variation (TV) regularization [4, 5, 6, 7] is also widely adopted in image restoration problems, to exploit the feature that natural images are smooth in most regions such that the gradient values at most pixels are very small. The TV image prior model is generally formulated as:

$$\varphi(\boldsymbol{u}) = \mathrm{TV}(\boldsymbol{u}) = \sum_i \|D_i\boldsymbol{u}\| \quad (3)$$

where $D_i\boldsymbol{u} \in \mathbb{R}^2$ is the discrete gradient of $\boldsymbol{u}$ at pixel i, $\|.\|$ is l_1-norm or l_2-norm, corresponding to the anisotropic TV and the isotropic TV, respectively. Obviously, TV favors local smoothness in natural images and it can be regarded as a kind of sparsity prior model in the gradient domain. As a matter of fact, it suggests that the gradient data at all pixels conform to identical zero-mean Laplacian distribution.

The above hypothesis in TV model is correct to some extent. However, it may not be accurate enough, in the sense that the statistics of natural images may not be stationary. In fact, the distributions of gradient data may vary from one region to another so that the gradients at different locations do not necessarily have the same variance. In addition, for any specific pixel, the zero-mean assumption of gradients may not be valid either, especially for the image regions with edges and rich textures. In this paper, we present a non-local extension of traditional TV regularization, allowing each pixel to have a separate gradient distribution. In the proposed model, we manage to estimate the statistics and derive a sparse distribution for gradient data at each pixel adaptively, by exploiting a set of non-locally searched patches which are similar to the patch centered at current pixel. Considering the features of the proposed approach, we call it Non-local Gradient Sparsity Regularization (NGSR).

The remainder of this paper is organized as follows. Section II briefly reviews traditional TV restoration. Section III describes the proposed NGSR regularization, showing how the parameters of gradient distribution at different locations are estimated and how the optimization problem is numerically solved efficiently. Experimental results are reported in Section IV and Section V concludes the paper.

II. TRADITIONAL TOTAL VARIATION REGULARIZATION

In this part we review traditional TV regularization. According to the Bayes rule, the maximum a posterior probability (MAP) estimate of $\boldsymbol{u}$ given $\boldsymbol{z}$ is given by

$$\tilde{\boldsymbol{u}} = \arg\max_{\boldsymbol{u}} p(\boldsymbol{u}|\boldsymbol{z}) = \arg\max_{\boldsymbol{u}} \frac{p(\boldsymbol{z}|\boldsymbol{u})p(\boldsymbol{u})}{p(\boldsymbol{z})},$$

which is equivalent to

$$\tilde{\boldsymbol{u}} = \arg\max_{\boldsymbol{u}} \log p(\boldsymbol{z}|\boldsymbol{u}) + \log p(\boldsymbol{u}). \quad (4)$$

978-1-4799-3432-4/14/$31.00 ©2014 IEEE

The term $\log p(\boldsymbol{z}|\boldsymbol{u})$ stands for the likelihood of $\boldsymbol{u}$. Clearly, the conditional probability $p(\boldsymbol{z}|\boldsymbol{u})$ is i.i.d. Gaussian distributed, since $\boldsymbol{z}$ is the sum of $\boldsymbol{u}$ and a white Gaussian noise. The second term $\log p(\boldsymbol{u})$ reflects the prior knowledge of $\boldsymbol{u}$. In TV regularization, the gradient data is supposed to have i.i.d. zero-mean Laplacian distribution. In this setting, the MAP estimate is formulated as

$$\min_{\boldsymbol{u}} \frac{1}{2\sigma_n^2}\|H\boldsymbol{u}-\boldsymbol{z}\|_2^2+\frac{\sqrt{2}}{\sigma_\Delta}\sum_i\|D_i\boldsymbol{u}\| \quad (5)$$

where σ_n^2 and σ_Δ^2 are the variations of the noise and the gradient, respectively. This provides a statistical explanation for the formulation (2) and (3), with $\mu=\frac{\sigma_\Delta}{\sqrt{2}\sigma_n^2}$.

Many algorithms have been developed to solve the TV regularization problems. In recent typical algorithms, such as Split Bregman [8], SALSA [9], etc., the problem is transformed into

$$\min_{\boldsymbol{u}} \frac{\mu}{2}\|H\boldsymbol{u}-\boldsymbol{z}\|_2^2+\sum_i|w_i| \quad s.t.\ w_i=D_i\boldsymbol{u} \quad (6)$$

by variable splitting and then solved by alternating minimization. The same approach will be adopted in our proposed nonlocal extended TV regularization. The details will be discussed in Section III to avoid repetition.

III. Nonlocal Gradient Sparsity Regularization

It is well known that the characteristics of image content might change from location to location. For instance, the gradient values are usually very close to zero in flat area, but can be much larger in regions with rich textures. Therefore, instead of assuming i.i.d. Laplacian distribution as in traditional TV

$$\mathrm{TV}(\boldsymbol{u})=\sum_i(|D_i^{\mathrm{h}}\boldsymbol{u}|+|D_i^{\mathrm{v}}\boldsymbol{u}|), \quad (7)$$

we propose an extended TV regularization term

$$\mathrm{NGSR}(\boldsymbol{u})=\sum_i\left(\frac{\sqrt{2}}{\sigma_{\Delta_i}^{\mathrm{h}}}|D_i^{\mathrm{h}}\boldsymbol{u}-m_{\Delta_i}^{\mathrm{h}}|+\frac{\sqrt{2}}{\sigma_{\Delta_i}^{\mathrm{v}}}|D_i^{\mathrm{v}}\boldsymbol{u}-m_{\Delta_i}^{\mathrm{v}}|\right), \quad (8)$$

which suggests that the gradient at each pixel has a separate Laplacian distribution. Here $\sigma_{\Delta_i}^{\mathrm{h}}$ and $\sigma_{\Delta_i}^{\mathrm{v}}$ are the standard deviations of the horizontal and the vertical first-order differences at pixel i, respectively. $m_{\Delta_i}^{\mathrm{h}}$ and $m_{\Delta_i}^{\mathrm{v}}$ are the expectations of the horizontal and the vertical differences at pixel i, respectively.

A. Parameter Estimation

A key issue in the proposed NGSR regularization is how to estimate the pixel-wise distribution parameter σ_{Δ_i} and m_{Δ_i}. Of course, the distribution is unknown. To tackle the problem, we follow the idea of non-local estimation. To be specific, to figure out the gradient distribution at a pixel i, we search for a set of patches which are similar to the patch centered at pixel i, and regard the data in these patches as samples of the distribution we would like to learn. Let $\boldsymbol{p}_i$ be the patch of size $N\times N$ centered at pixel i. The similarity between any two patches $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$ is measured by the ℓ_2-distance:

$$d(i,j)=\frac{\|\boldsymbol{p}_i-\boldsymbol{p}_j\|_2^2}{N^2} \quad (9)$$

In our algorithm, K patches that are most similar to $\boldsymbol{p}_i$ are searched via block matching, and we record the center pixel positions of these similar blocks in set $\mathbb{L}_i$. Then the parameter $m_{\Delta_i}^{\mathrm{h}}$ is estimated by

$$m_{\Delta_i}^{h}=\frac{1}{|\mathbb{L}_i|}\sum_{j\in\mathbb{L}_i}D_j^{\mathrm{h}}\boldsymbol{u}, \quad (10)$$

and $\sigma_{\Delta_i}^{\mathrm{h}}$ is estimated by

$$\sigma_{\Delta_i}^{\mathrm{h}}=\sqrt{\frac{1}{|\mathbb{L}_i|}\sum_{j\in\mathbb{L}_i}(D_j^{\mathrm{h}}\boldsymbol{u}-m_{\Delta_i}^{\mathrm{h}})^2}. \quad (11)$$

$m_{\Delta_i}^{\mathrm{v}}$ and $\sigma_{\Delta_i}^{\mathrm{v}}$ can be estimated in the same way. Both the block-matching and the calculation of (10) and (11) require the image $\boldsymbol{u}$, which is unavailable. In our implementation, we generate an initial restoration, denoted by $\boldsymbol{u}^{\mathrm{basic}}$, using existing methods (e.g. traditional TV), and use it for the purpose of block matching and parameter estimation.

B. Algorithm

Based on the proposed NGSR regularization, the MAP estimate of $\boldsymbol{u}$ is:

$$\min_{\boldsymbol{u}}\left\{\begin{array}{c}\frac{1}{2\sigma_n^2}\|H\boldsymbol{u}-\boldsymbol{z}\|_2^2\\ +\sum_i\left(\frac{\sqrt{2}}{\sigma_{\Delta_i}^{\mathrm{h}}}|D_i^{\mathrm{h}}\boldsymbol{u}-m_{\Delta_i}^{\mathrm{h}}|+\frac{\sqrt{2}}{\sigma_{\Delta_i}^{\mathrm{v}}}|D_i^{\mathrm{v}}\boldsymbol{u}-m_{\Delta_i}^{\mathrm{v}}|\right)\end{array}\right\}. \quad (12)$$

Making use of variable splitting technique [9, 10], the problem becomes a constrained optimization:

$$\min_{\boldsymbol{u}}\left\{\begin{array}{c}\frac{1}{2\sigma_n^2}\|H\boldsymbol{u}-\boldsymbol{z}\|_2^2\\ +\sum_i\left(\frac{\sqrt{2}}{\sigma_{\Delta_i}^{\mathrm{h}}}|w_i^{\mathrm{h}}-m_{\Delta_i}^{\mathrm{h}}|+\frac{\sqrt{2}}{\sigma_{\Delta_i}^{\mathrm{v}}}|w_i^{\mathrm{v}}-m_{\Delta_i}^{\mathrm{v}}|\right)\end{array}\right\}$$

$$s.t.\ w_i^{\mathrm{h}}=D_i^{\mathrm{h}}\boldsymbol{u},\ w_i^{\mathrm{v}}=D_i^{\mathrm{v}}\boldsymbol{u} \quad (13)$$

We employ augmented Lagrangian method to solve the problem. The augmented Lagrange function of (13) is

$$\mathcal{L}_A(\boldsymbol{u},\boldsymbol{w}^{\mathrm{h}},\boldsymbol{w}^{\mathrm{v}})=\frac{1}{2\sigma_n^2}\|H\boldsymbol{u}-\boldsymbol{z}\|_2^2+\sum_i\left(\frac{\sqrt{2}}{\sigma_{\Delta_i}^{\mathrm{h}}}|w_i^{\mathrm{h}}-m_{\Delta_i}^{\mathrm{h}}|+\frac{\sqrt{2}}{\sigma_{\Delta_i}^{\mathrm{v}}}|w_i^{\mathrm{v}}-m_{\Delta_i}^{\mathrm{v}}|\right)+\frac{\beta}{2}\left(\|\boldsymbol{w}^{\mathrm{h}}-D^{\mathrm{h}}\boldsymbol{u}\|_2^2+\|\boldsymbol{w}^{\mathrm{v}}-D^{\mathrm{v}}\boldsymbol{u}\|_2^2\right)-(\boldsymbol{w}^{\mathrm{h}}-D^{\mathrm{h}}\boldsymbol{u})^T\boldsymbol{\lambda}^{\mathrm{h}}-(\boldsymbol{w}^{\mathrm{v}}-D^{\mathrm{v}}\boldsymbol{u})^T\boldsymbol{\lambda}^{\mathrm{v}} \quad (14)$$

where $\boldsymbol{w}^{\mathrm{h}}$ and $\boldsymbol{w}^{\mathrm{v}}$ are lexicographically stacked version of w_i^{h} and w_i^{v}, β is regularization parameter associated with penalty terms $\|\boldsymbol{w}^{\mathrm{h}}-D^{\mathrm{h}}\boldsymbol{u}\|_2^2$ and $\|\boldsymbol{w}^{\mathrm{v}}-D^{\mathrm{v}}\boldsymbol{u}\|_2^2$, $\boldsymbol{\lambda}^{\mathrm{h}}$ and $\boldsymbol{\lambda}^{\mathrm{v}}$ are the Lagrange multipliers associated with the constraints $\boldsymbol{w}^{\mathrm{h}}=D^{\mathrm{h}}\boldsymbol{u}$ and $\boldsymbol{w}^{\mathrm{v}}=D^{\mathrm{v}}\boldsymbol{u}$, respectively. The problem can be solved by solving (15) and (16) iteratively:

$$(\boldsymbol{u}_{(k+1)},\boldsymbol{w}_{(k+1)}^{\mathrm{h}},\boldsymbol{w}_{(k+1)}^{\mathrm{v}})=\min_{\boldsymbol{u},\boldsymbol{w}^{\mathrm{h}},\boldsymbol{w}^{\mathrm{v}}}\mathcal{L}_A(\boldsymbol{u}_{(k)},\boldsymbol{w}_{(k)}^{\mathrm{h}},\boldsymbol{w}_{(k)}^{\mathrm{v}}) \quad (15)$$

$$\begin{cases}\boldsymbol{\lambda}_{(k+1)}^{\mathrm{h}}=\boldsymbol{\lambda}_{(k)}^{\mathrm{h}}-\beta_{(k)}(\boldsymbol{w}_{(k+1)}^{\mathrm{h}}-D^{\mathrm{h}}\boldsymbol{u}_{(k+1)})\\ \boldsymbol{\lambda}_{(k+1)}^{\mathrm{v}}=\boldsymbol{\lambda}_{(k)}^{\mathrm{v}}-\beta_{(k)}(\boldsymbol{w}_{(k+1)}^{\mathrm{h}}-D^{\mathrm{v}}\boldsymbol{u}_{(k+1)})\end{cases} \quad (16)$$

where k is iteration number. We use alternating direction technique [11, 12] to decompose (14) into three sub-problems, each of which can be solved efficiently.

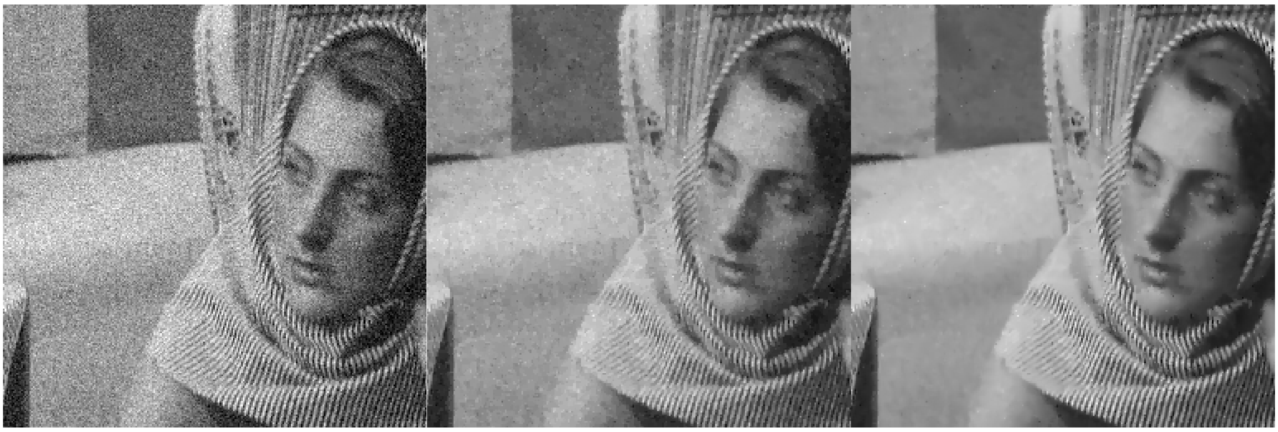
1) $\boldsymbol{w}^h$-problem and $\boldsymbol{w}^v$-problem

Given $\boldsymbol{u}$, $\boldsymbol{w}^{\mathrm{v}}$, the optimization problem associated with $\boldsymbol{w}^{\mathrm{h}}$ can be simplified as

$$\mathcal{L}_A(\boldsymbol{w}^{\mathrm{h}})=\sum_i\frac{\sqrt{2}}{\sigma_{\Delta_i}^{\mathrm{h}}}|w_i^{\mathrm{h}}-m_{\Delta_i}^{\mathrm{h}}|+\frac{\beta}{2}\left\|\boldsymbol{w}^{\mathrm{h}}-D^h\boldsymbol{u}-\frac{\lambda^{\mathrm{h}}}{\beta}\right\|_2^2 \quad (17)$$

The solution is a simple shrinkage operation:

(a) Noisy image ($\sigma_{\mathrm{n}} = 20$) (b) TV (PSNR=28.28 dB) (c) NGSR (PSNR=29.45 dB)



(a) Noisy image ($\sigma_{\mathrm{n}} = 20$) (b) TV (PSNR=26.93 dB) (c) NGSR (PSNR=28.63 dB)

Fig.1: Visual comparison. (a) Noisy image (b) TV denoised (c) NGSR denoised. Please enlarge the pictures for better comparison.

$$\boldsymbol{w}^{\mathrm{h}} = \boldsymbol{m}_{\Delta}^{\mathrm{h}} + \max\left(\left|\widetilde{\boldsymbol{w}^{\mathrm{h}}} - \boldsymbol{m}_{\Delta}^{\mathrm{h}}\right| - \frac{\sqrt{2}}{\beta\cdot\boldsymbol{\sigma}_{\Delta}^{\mathrm{h}}}, 0\right)\cdot \mathrm{sgn}(\widetilde{\boldsymbol{w}^{\mathrm{h}}}) \quad (18)$$

with $\widetilde{\boldsymbol{w}^{\mathrm{h}}} = D^{\mathrm{h}}\boldsymbol{u} + \boldsymbol{\lambda}^{\mathrm{h}}/\beta$. Here $\boldsymbol{m}_{\Delta}^{\mathrm{h}}$ and $\boldsymbol{\sigma}_{\Delta}^{\mathrm{h}}$ are lexicographically stacked version of $m_{\Delta_i}^{\mathrm{h}}$ and $\sigma_{\Delta_i}^{\mathrm{h}}$ for the whole image. The shrinkage operation is performed component by component. Similarly, the $\boldsymbol{w}^{\mathrm{v}}$-problem can be solve by

$$\boldsymbol{w}^{\mathrm{v}} = \boldsymbol{m}_{\Delta}^{\mathrm{v}} + \max\left(\left|\widetilde{\boldsymbol{w}^{\mathrm{v}}} - \boldsymbol{m}_{\Delta}^{\mathrm{v}}\right| - \frac{\sqrt{2}}{\beta\cdot\boldsymbol{\sigma}_{\Delta}^{\mathrm{v}}}, 0\right)\cdot \mathrm{sgn}(\widetilde{\boldsymbol{w}^{\mathrm{v}}}) \quad (19)$$

with $\widetilde{\boldsymbol{w}^{\mathrm{v}}} = D^{\mathrm{v}}\boldsymbol{u} + \boldsymbol{\lambda}^{\mathrm{v}}/\beta$.

2) $\boldsymbol{u}$-problem

With $\boldsymbol{w}^{\mathrm{h}}$ and $\mathbf{w}^{\mathrm{v}}$ fixed, the $\boldsymbol{u}$-problem can be rewritten as:

$$\mathcal{L}_A(\boldsymbol{u}) = \frac{\mu}{2}\|H\boldsymbol{u} - \boldsymbol{z}\|_2^2 + \frac{\beta}{2}\left(\left\|D^{\mathrm{h}}\boldsymbol{u} - (\boldsymbol{w}^{\mathrm{h}} - \frac{\boldsymbol{\lambda}^{\mathrm{h}}}{\beta})\right\|_2^2 + \left\|D^{\mathrm{v}}\boldsymbol{u} - (\boldsymbol{w}^{\mathrm{v}} - \frac{\boldsymbol{\lambda}^{\mathrm{v}}}{\beta})\right\|_2^2\right). \quad (20)$$

This is evidently a quadratic function which can be solved easily. When the H matrix represents a convolution operation, (20) can be solved very efficiently in the Fourier transform domain:

$$\widetilde{\boldsymbol{u}} = \boldsymbol{\mathcal{F}}^{-1}\left(\frac{\boldsymbol{\mathcal{F}}^*(D^{\mathrm{h}})\circ\boldsymbol{\mathcal{F}}\left(\boldsymbol{w}^{\mathrm{h}} - \frac{\boldsymbol{\lambda}^{\mathrm{h}}}{\beta}\right) + \boldsymbol{\mathcal{F}}^*(D^{\mathrm{v}})\circ\boldsymbol{\mathcal{F}}\left(\boldsymbol{w}^{\mathrm{v}} - \frac{\boldsymbol{\lambda}^{\mathrm{v}}}{\beta}\right) + \frac{\mu}{\beta}\boldsymbol{\mathcal{F}}^*(\mathrm{H})\circ\boldsymbol{\mathcal{F}}(\boldsymbol{z})}{\boldsymbol{\mathcal{F}}^*(D^{\mathrm{h}})\circ\boldsymbol{\mathcal{F}}(D^{\mathrm{h}}) + \boldsymbol{\mathcal{F}}^*(D^{\mathrm{v}})\circ\boldsymbol{\mathcal{F}}(D^{\mathrm{v}}) + \frac{\mu}{\beta}\boldsymbol{\mathcal{F}}^*(\mathrm{H})\circ\boldsymbol{\mathcal{F}}(\mathrm{H})}\right), \quad (21)$$

where $\boldsymbol{\mathcal{F}}$ is two-dimensional discrete Fourier transform, "*" denotes complex conjugacy, " ∘ " denotes component-wise multiplication, and the division is performed component-wisely too.

TABLE I. A SUMMARY OF THE NGSR ALGORITHM

Input: The observed noisy image $\boldsymbol{z}$, σ_n, basic estimate $\boldsymbol{u}^{\mathrm{basic}}$, and β, μ.

Initialization: $\boldsymbol{u} = \boldsymbol{z}$, $\boldsymbol{\lambda}^{\mathrm{h}} = \boldsymbol{\lambda}^{\mathrm{v}} = \mathbf{0}$.

Use block matching within $\boldsymbol{u}^{\mathrm{basic}}$, calculate $\boldsymbol{m}_{\Delta}^{\mathrm{h}}$, $\boldsymbol{m}_{\Delta}^{\mathrm{v}}$, $\boldsymbol{\sigma}_{\Delta}^{\mathrm{h}}$, $\boldsymbol{\sigma}_{\Delta}^{\mathrm{v}}$ according to Eq. (10) and (11);

while Outer stopping criteria unsatisfied ***do***

while Inner stopping criteria unsatisfied ***do***

solve $\boldsymbol{w}^{\mathrm{h}}$-problem by computing Eq. (18);

solve $\boldsymbol{w}^{\mathrm{v}}$-problem by computing Eq. (19);

solve $\boldsymbol{u}$-problem by computing Eq. (21);

end while

Update multipliers $\boldsymbol{\lambda}^{\mathrm{h}}$, $\boldsymbol{\lambda}^{\mathrm{v}}$ by Eq. (8);

Choose $\beta_{(k+1)} \geq \beta_{(k)}$;

end while

Output: Denoised image $\boldsymbol{u}^{\mathrm{final}}$.

IV. EXPERIMENTAL RESULTS

In this section, we evaluate the efficiency of the proposed NGSR regularization and compare it with the traditional TV method. Due to limited space, we only consider a special case of image restoration in which the matrix H is identity matrix I so that the problem becomes image denoising. However, it is clear that the proposed NGSR regularization can also be applied to other image restoration problems, because almost all image restoration problems can be decomposed into a denoising problem and a Least Square problem, via variable splitting techniques.

Since gradient data may not strictly conform to Laplacian distribution, the μ in TV is adjusted by a factor δ_μ so as to get the best possible performance of TV. For the same reason, the estimated $\boldsymbol{\sigma}_\Delta^{\mathrm{h}}$ and $\boldsymbol{\sigma}_\Delta^{\mathrm{v}}$ are adjusted by a factor δ_Δ. Both δ_μ and δ_Δ are empirically chosen, according to the value of σ_n. The number of similar patches K is set to 50 and the patch size N is set to 7.

The proposed NGSR regularization is compared with traditional TV, which is also the first step of NGSR to generate $\boldsymbol{u}^{\text{basic}}$. As can be seen in Table II and Table III, NGSR outperforms TV evidently in all cases as respect to PSNR and SSIM, with σ_n varies from 10 to 60. The PSNR gain achieved by proposed NGSR over traditional TV can be as much as 1.3dB and the SSIM gain can be up to 0.1727. As to the average improvement, the PSNR gain is 0.983dB, and the SSIM gain is 0.0736.The reconstruction results of Cameraman and Barbara (a 256×256 fragment) in the case $\sigma_n = 20$ are shown in Fig. 1. Obviously, the results of NGSR are perceptually better. Other pictures are not listed here due to limited space. We also tested NGSR on images like Lena (512×512), Hill (512×512) and Baboon (512×512), etc. All the results suggest that NGSR outperforms TV in restoration performance.

V. CONCLUSIONS

This paper introduces a non-local extension of total variation regularization, which models the sparsity of the image gradient data with pixel-wise distributions, reflecting the non-stationary nature of image statistics. Taking advantage of the self-similarity of natural images, the proposed approach estimates the statistics of image gradient at a particular pixel from a group of non-locally searched patches which are similar to the patch at current pixel. The gradient data in these non-local similar patches are regarded as samples of the gradient distribution to learn. Experiments show that the proposed approach, called Nonlocal Gradient Sparse Regularization (NGSR), can model image gradients more accurately and outperform traditional TV remarkably, producing better restored images with higher PSNR and SSIM values and better perceptual qualities.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (61370114, 61121002, 61073083), Beijing Natural Science Foundation (4132039, 4112026) and Research Fund for the Doctoral Program of Higher Education (20120001110090, 20100001120027).

TABLE II. PSNR COMPARISON (UNIT: dB)

Image	*Cameraman* 256×256		*Barbara* 512×512		*Peppers* 512×512		*House* 256×256	
σ_n	*TV*	*NGSR*	*TV*	*NGSR*	*TV*	*NGSR*	*TV*	*NGSR*
10	32.32	**33.45**	31.13	**33.09**	33.90	**34.42**	34.39	**34.97**
20	28.28	**29.45**	26.93	**28.63**	31.00	**31.78**	31.28	**32.22**
30	26.19	**27.47**	24.95	**26.16**	29.27	**30.03**	29.51	**30.44**
40	24.78	**26.08**	23.78	**24.74**	28.01	**28.72**	28.20	**29.09**
50	23.87	**25.05**	23.08	**23.85**	27.01	**27.66**	27.14	**27.99**
60	23.05	**24.24**	22.51	**23.29**	26.25	**26.83**	26.27	**27.05**

TABLE III. SSIM COMPARISON

Image	*Cameraman* 256×256		*Barbara* 512×512		*Peppers* 512×512		*House* 256×256	
σ_n	*TV*	*NGSR*	*TV*	*NGSR*	*TV*	*NGSR*	*TV*	*NGSR*
10	0.8454	**0.9206**	0.8608	**0.9220**	0.8630	**0.8677**	0.8817	**0.8828**
20	0.7172	**0.8449**	0.7247	**0.8608**	0.8096	**0.8312**	0.8342	**0.8510**
30	0.6398	**0.7922**	0.6302	**0.7504**	0.7789	**0.8054**	0.8089	**0.8285**
40	0.5840	**0.7511**	0.5638	**0.6837**	0.7556	**0.7852**	0.7885	**0.8099**
50	0.5578	**0.7244**	0.5258	**0.6370**	0.7379	**0.7672**	0.7724	**0.7929**
60	0.5254	**0.6981**	0.4911	**0.6044**	0.7222	**0.7527**	0.7565	**0.7770**

REFERENCES

[1] A. Foi, V. Katkovnik, and K. Egiazarian, “Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images,” *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1395–1411, May 2007.

[2] D.L. Donoho, “De-noising by soft-thresholding”, *IEEE Trans. on Information Theory*, vol. 41, no. 3, pp. 613-627, 1995.

[3] M. Elad, M. Aharon, “Image denoising via sparse and redundant repretantions over learned dictionaries for sparse representation”, *IEEE Trans. on Signal Processing*, vol. 54, no. 11, pp. 4311-4322, 2006.

[4] A. Chambolle, “An algorithm for total variation minimization and applications”, *J. Math. Imaging Vision.*, vol. 20, no. 1-2, pp. 89–97, 2004.

[5] A. Chambolle and P. L. Lions, “Image recovery via total variation minimization and related problems”, *Numer. Math.*, vol. 76, no. 2, pp. 167–188, 1997.

[6] T. F. Chan, S. Esedoglu, F. Park, and A. Yip, “Recent developments in total variation image restoration”, *CAM Report 05-01*, Department of Mathematics, UCLA, 2004.

[7] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction", *SIAM J. Imaging Sci.*, vol. 1, no. 3, pp. 248–272, 2008.

[8] T. Goldstein, S. Osher, “The Split Bregman Algorithm for L1 Regularized Problems,” *SIAM J. Imaging Sci.*, vol. 2, pp. 323–343, 2009.

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

[10] J. F. Cai, S. Osher, and Z. W. Shen, “Split Bregman methods and frame based image restoration,” *Multiscale Model. Simul.*, vol. 8, pp.5057-5071, 2009.

[11] C. Li, W. Yin, and Y. Zhang, “TVAL3: TV Minimization by Augmented Lagrangian and Alternating Direction Algorithm,” 2009, [online] available at http://www.caam.rice.edu/~optimization/L1/TVAL3/.

[12] S. Chrétien, “An alternating ℓ_1 approach to the compressed sensing problem,” *IEEE Signal Processing Letters*, vol. 17, no. 2, pp. 181–184, 2010.