An Image Denoising Model via the Reconciliation of the Sparsity and Low-Rankness of the Dictionary Domain Coefficients

YANG Yifan¹², ZHANG Tao²†, WU Di², ZHAO Yu²

1. Key Laboratory of Multidisciplinary Management and Control of Complex Systems of Anhui Higher Education Institutes, Anhui University of Technology, Maanshan 243002, Anhui, China;
2. School of Mathematics and Physics, Anhui University of Technology, Maanshan 243032, Anhui, China

Abstract: Sparse coding has achieved great success in various image restoration tasks. However, if the sparse representation coefficients of the structure (low-frequency information) and texture (high-frequency information) components of the image are under the same penalty constraint, the restoration effect may not be ideal. In this paper, an image denoising model combining mixed norm and weighted nuclear norm as regularization terms is proposed. The proposed model simultaneously exploits the group sparsity of the high frequency and low-rankness of the low frequency in dictionary-domain. The mixed norm is used to constrain the high frequency part and the weighted nuclear norm is used to constrain the low frequency part. In order to make the proposed model easy to solve under the framework of alternative direction multiplier method (ADMM), iterative shrinkage threshold method and weighted nuclear norm minimization method are used to solve the two sub-problems. The validity of the model is verified experimentally through comparison with some state-of-the-art methods.

Key words: image denoising; mixed norm; sparse representation; principal component analysis (PCA) dictionary

CLC number: TP 751

0 Introduction

As a fundamental and important image processing task, image denoising has a wide range of applications [1-3]. Traditional methods are based on local denoising, i.e., all pixels in the vicinity of the current pixel are used to estimate the current pixel. As a result, the estimated pixel can be heavily influenced by the pixels in its neighborhood. In recent years, image denoising methods based on non-local block matching have achieved great success [4-5]. Buades et al. [6] proposed the non-local mean (NLM) filter, which estimates each pixel by a non-local averaging of all the pixels in the image. The most well-known hybrid method for image denoising is the block-matching and 3D filtering (BM3D) reported in Ref. [7], which groups similar patches into 3D arrays and deals...
with these arrays by sparse collaborative filtering, achieving better visual results in several image processing tasks. Jia et al.\cite{41} proposed a joint sparse and low-rank matrix approximation problem by solving the nuclear norm and $l_1$ norm related minimization problem.

Using non-local self-similarity, nuclear norm minimization (NNM) has a wide range of applications in matrix decomposition and low rank representation\cite{42}. Cai et al.\cite{43} proved that the NNM can be easily solved by a soft-thresholding operation on the singular values of observation matrix. Given a matrix $Y$, the aim of NNM is to find a low-rank matrix $X$ by solving the following problem:

$$\hat{X} = \arg \min_{X} \left( \| X - Y \|_F^2 + \lambda \| X \|_0 \right)$$

(1)

where $\| \cdot \|_F$ and $\| \cdot \|_0$ denote the Frobenius norm and nuclear norm, respectively, and $\lambda$ is a trade-off parameter between the fidelity term and the low rank regularization term induced by the nuclear norm. The solution of problem (1) can be obtained by $X = USV^T$, where $Y = USV^T$ is the singular value decomposition (SVD) of $Y$ and $S_V(\Sigma)$ is the soft thresholding operator on diagonal matrix $\Sigma$ with parameter $\lambda$.

In many applications, it is hard for single $\lambda$ to shrink all singular values. Large $\lambda$ will over-penalize the large singular values, and small $\lambda$ preserves large singular values but prevents NNM from obtaining low-rank solutions. In recent year, Gu et al.\cite{44,45} proposed a weighted nuclear norm minimization (WNNM) to balance the rank and thresholding, and achieved competitive results.

Inspired by the above approach, group sparse coding (GSC) has shown great potential in a variety of image processing tasks in recent years. Instead of using a single patch as the basic unit of sparse coding, image processing is accomplished using groups of similar patches. GSC has great potential for solving many image problems that takes full advantage of local sparsity and non-local self-similarity\cite{46,47,48,49}. Specifically, an image $x \in \mathbb{R}^n$ is divided into $n$ overlapped patches of size $\sqrt{d} \times \sqrt{d}$, $d \in \mathbb{R}$. For each patch $x_i$, $m$ similar matching patches are selected from a $W \times W$ sized search window to form a set $K_i$. Note that the k-nearest neighbor (KNN) algorithm\cite{50} is commonly used to search for similarly matched patches. Then all patches in $K_i$ are stacked into a matrix $X_i \in \mathbb{R}^{d \times m}$, i.e., $X_i = \{x_{i,1}, x_{i,2}, \cdots, x_{i,m}\}$. The matrix $X$, consisting of patches with similar structure is called a patch group, where $\{x_{i,j}\}_{j=1}^m$ denotes the $j$-th patch in the $i$-th patch group. Similar to GSC, given a dictionary $D$, each patch group $X_i$ can be sparsely represented by solving the following $l_1$-norm minimization problem,

$$\hat{X}_i = \arg \min_{X} \left( \frac{1}{2} \| Y - D X_i \|_F^2 + \lambda \| X_i \|_0 \right)$$

(2)

where $A_i$ represents the group sparse coefficient of each patch group $X_i$. However, since $l_1$-norm minimization is an NP-hard optimization problem, it is often relaxed to the convex $l_1$-norm minimization. Unfortunately, for some practical image processing tasks, such as image restoration, $l_1$-norm minimization is just an estimation to the $l_0$-norm minimization and cannot obtain the desirable sparse solution. Various norm minimization methods, like the weighted $l_1$-norm minimization\cite{51}, the $l_{1,2}$-norm minimization\cite{52} and the weighted Schatten norm minimization\cite{53}, are proposed to solve the $l_0$-norm minimization problem.

Inspired by GSC prior and low rank prior mentioned above, in this paper we will propose a denoising model in which mixed norm and nuclear norm are employed to regularize the coefficients of the high and low frequency of the image. The proposed denoising model can be efficiently solved by the ADMM algorithm\cite{54}, which provides a closed solution for each subproblem of the model.

1 Preliminaries

1.1 Nuclear Norm and Mixed Norm

Here, we briefly introduce nuclear norm and mixed norm\cite{55}.

Definition 1 The weighted nuclear norm of a matrix $X$ is defined as

$$\| X \|_{w,\|,\|_F} = \sum_{i=1}^{\min\{d,n\}} w_i \sigma_i = \text{Tr}(W \Sigma)$$

(3)

where $w = [w_1, w_2, \cdots, w_{\min\{d,n\}}]$ and $\sigma$ means the $i$-th singular value of $X$. The diagonal matrix of $W$ and $\Sigma$ consisting of elements $w_i$ and $\sigma_i$, respectively.

Definition 2 The mixed $(1,1)$ norm of a matrix $X$ is defined as

$$\| X \|_{1,1} = \sum_{i=1}^{d} \sum_{j=1}^{n} |x_{i,j}|$$

(4)

The mixed $(2,1)$ norm of a matrix $X$ is defined as


\[ \| X \|_{2,1} = \sum_{i=1}^{d} \sqrt{\sum_{j=1}^{\min\{d,n\}} x_{ij}^2} \]  

(5)

1.2 Adaptive Dictionary Learning

In Ref. [24], an adaptive dictionary learning approach is designed. For each patch group \( X_i \), its adaptive dictionary can be learned from its observation \( Y_i \in \mathbb{R}^{n \times n} \). Specifically, SVD is applied to \( Y_i \):

\[ Y_i = U_i V_i^T = \sum_{j=1}^{\min\{d,m\}} \delta_{ij} u_j v_j^T \]  

(6)

where \( \Delta_i = \text{diag}(\delta_{i,1}, \delta_{i,2}, \ldots, \delta_{i,\min\{d,m\}}) \) is a diagonal matrix, and \( u_j, v_j \) are the columns of \( U_i \) and \( V_i \), respectively. Then each dictionary atom \( d_i \) of the adaptive dictionary \( D \), for every patch group \( Y_i \), is defined as:

\[ d_{ij} = \begin{cases} u_j, & \text{if } j = 1, 2, \ldots, \min\{d,m\} \\ v_j, & \text{otherwise} \end{cases} \]  

(7)

Thus, an adaptive dictionary has been learned, i.e.,

\[ D_i = [d_{i,1}, d_{i,2}, \ldots, d_{i,\min\{d,m\}}] \]  

(8)

It can be seen that the designed dictionary learning approach only requires one SVD operation for each patch group.

2 Proposed Model

In existing group sparse models, sparse coding is conducted the same on different components of the image such as structure and texture. In fact, an image can be decomposed into the low frequency (structure) and high frequency (texture) components, and it is difficult for single regularization term to enforce sparsity on different components. So in this subsection, we aim to recover the sparse coefficients of high and low frequency components simultaneously by solving the following problem:

\[ L(A_i, W_i; Z_i) = \frac{1}{2} \left( \| D_i A_i - Y_i \|_F^2 + \mu_1 \| E_i \|_{2,1} \right) + \frac{\mu_2}{2} \left( \| W_i \|_{*,*} + \left( Z_i A_i - (W_i + E_i) \right) \right) \]  

\[ + \frac{\beta}{2} \left( \| A_i - (W_i + E_i) \|_F^2 \right) \]  

(9)

where \( A_i \) has the same meaning as that of group sparse model (2). \( \mu_1, \mu_2 \) are regularization parameters which control the balance between the fidelity term and the regularization terms. \( E_i \) and \( W_i \) represent the high frequency coefficient and the low frequency coefficient corresponding to the texture and structure components of the image, respectively.

The corresponding augmented Lagrange function of Eq. (11) is defined by

\[ \frac{L(A, W, E; Z)}{\Delta} = \frac{1}{2} \left( \| D_i A_i - Y_i \|_F^2 + \mu_1 \| E_i \|_{2,1} \right) + \frac{\mu_2}{2} \left( \| W_i \|_{*,*} + \left( Z_i A_i - (W_i + E_i) \right) \right) + \frac{\beta}{2} \left( \| A_i - (W_i + E_i) \|_F^2 \right) \]  

(10)

where \( \beta \) is a positive parameter. In the following we use the ADMM algorithm to solve these proposed problem. Eq. (10) is decomposed into the following subproblems.

1) \( A_i \)-subproblem

For fixed \( W_i, E_i \), the minimization problem for \( A_i \) becomes

\[ A_i^{k+1} = \arg \min_{A_i} \left\{ \frac{1}{2} \left( \| D_i A_i - Y_i \|_F^2 \right) + \frac{\beta}{2} \left( \| A_i - (W_i - E_i) \|_F^2 \right) \right\} \]  

(11)

This is a quadratic problem and has a closed-form solution:

\[ A_i^{k+1} = (\beta(W_i - E_i) - Z_i^k + D_i^k Y_i)/(1 + \beta) \]  

2) \( W_i \)-subproblem

The subproblem of \( W_i \) is a weighted nuclear norm minimization (WNMM) problem and is computed as follows:

\[ W_i^{k+1} = \arg \min_{W_i} \left\{ \frac{\mu_2}{2} \| W_i \|_{*,*} + \frac{\beta}{2} \left( \| A_i - (W_i + E_i) \|_F^2 \right) \right\} \]  

where \( A_i^{k+1} + Z_i^k - E_i^k = U\Sigma V^T \) is a singular value decomposition of \( A_i^{k+1} + Z_i^k - E_i^k \), and \( S_{\alpha, \beta} \) is the soft-thresholding operator with weight vector \( \alpha \).

\[ S_{\alpha, \beta} = \max(\Sigma_{\alpha, \beta} - \frac{\mu_2}{\beta}, 0) \]  

(13)

3) \( E_i \)-subproblem

\[ E_i^{k+1} = \arg \min_{E_i} \left\{ \frac{\mu_1}{2} \| E_i \|_{2,1} + \frac{\beta}{2} \left( \| A_i - (W_i + E_i) \|_F^2 \right) \right\} \]  

\[ \implies \arg \min_{E_i} \left\{ \frac{\mu_1}{2} \| E_i \|_{2,1} + \frac{1}{2} \left( \| A_i - (W_i + E_i) \|_F^2 \right) \right\} \]  

The subproblem of \( E_i \) is given explicitly by the two-dimensional shrinkage
\[ E_{i+1}(j) = \max \left\{ \| E_i \|_1 - \frac{\mu_i}{\beta}, 0 \right\} \odot \frac{E_i(j)}{\| E_i \|_2} \]  
\[ \| x_i^{k+1} - x_i^k \|_2 \leq \varepsilon \]  

where \( E_i = A_i^{k+1} + \frac{Z_i^{k+1}}{\beta} - W_i^{k+1} \) and \( \odot \) represents point-wise product.

4) \( Z \)-update

We update \( Z \) by

\[ Z^{k+1} = Z^k + \left[ A_i^{k+1} - (W_i^{k+1} + E_i^{k+1}) \right] \]  

Based on the above discussion, we summarize the complete algorithm of the proposed model (9) for image decomposition and denoising in Algorithm 1.

**Algorithm 1**

Initialization: Input \( x^0 = y^0 = y, A_i^0 = 0, W_i^0 = 0, E_i^0 = 0, Z_i^0 = 0 \).

Iteration:
1. While \( k = 1 \) to Max-Iter or Eq. (16) is satisfied.
2. Iterative regularization \( y^k = x^{k-1} + \mu (y - y^{k-1}) \)
3. For each patch \( y \), in \( y^k \) do
4. Find nonlocal similar patches from group \( Y \).
5. Compute \( A_i^{k+1} \) by Eq. (11).
6. Compute \( W_i^{k+1} \) by Eq. (12).
7. Compute \( E_i^{k+1} \) by Eq. (14).
8. Compute \( Z_i^{k+1} \) by Eq. (15).
9. Get the estimation: \( X = D_i A_i \).
10. End for
11. \( k = k + 1 \)
12. Aggregate \( X \) to form the denoised image \( x^{k+1} \).
13. End while.
14. Output: The final denoised image \( \hat{x} \).

For simplicity, we terminate the Algorithm 1 with the relative change of \( X \) in all experiments, i.e.,

\[ \left\| x^{k+1} - x^k \right\|_2 \leq \varepsilon \]

The recovered high and low frequency components of the image can be obtained by reconstructing the PCA dictionary from \( \hat{X}_H = D_i \hat{E}_i, \hat{X}_L = D_i \hat{W}_i \), and then aggregating all the estimated patch groups \( \hat{X}_H, \hat{X}_L \).

### 3 Experiments

In this section, we present various numerical results to illustrate the performance of the proposed model for image decomposition and denoising. Figure 1 shows the test images used in this paper.

#### 3.1 Image Decomposition Results

For image decomposition, the goal of the proposed model is to separate the high frequency information from the low frequency information in the image. Generally speaking, the structure and contours in an image belong to the low frequency components, while small details, textures and noise in an image belong to the high frequency components. We use the image Starfish shown in Fig. 2(a) to give a clear insight into the behavior of the proposed decomposition model. Figure 2(b) and 2(c) show the "cartoon + texture" decomposition result on noiseless Starfish image. For noisy Starfish image in Fig. 2(d), the proposed model also produces reasonable high-low frequency decomposition. The Starfish image is decomposed into simple cartoon component (low frequency) and noisy component (high frequency).

### 3.2 Image Denoising Results and Comparison with Other Methods

For image denoising, we compare the proposed method with other advanced methods including non-local means (NLM [6]), BM3D [7] and EPLL [23] in terms of objective and perceptual metrics. Both NLM and BM3D are based on the prior of non-local self similarity, and EPLL is a patch-prior based denoising method which ensures denoising performance in ensemble approximation by local patch denoising. In the proposed method, the size of patch \( \sqrt{b} \times \sqrt{b} \) is set to \( 6 \times 6 \) for \( 5 \leq \sigma \leq 20 \). The parameters \((\mu, \rho_1, \rho_2)\) are set to \((7, 10, 40), (7, 9, 80)\).
and (6, 8, 180) for noise level $\sigma_n = 5$, $\sigma_n = 10$ and $\sigma_n = 15$, respectively. In Fig. 3, we show the visual comparison of the denoising results with noise level $\sigma_n = 15$ obtained by different methods. From visual comparison, we can see that the images restored by NLM and BM3D seem over-smooth and EPLL produces some bad artifacts. Images restored by our algorithm has better denoising performance and less staircase artifacts compared with the other three methods.

The peak signal-to-noise (PSNR) ratio is employed to measure the quality of the recovered images which is defined as:

$$\text{PSNR}(u, f) = 10 \log \left( \frac{255^2}{\frac{1}{mn} \left\| u - f \right\|_2^2} \right)$$

where $u$ is the restored image, $f$ is the true image, $m$ and $n$ denote the size of the image. However, the PSNR only considers the global statistics of image pixel values with neglect of the local visual factor of human eyes, which make it is impossible to access local image quality. To this end, an objective image quality evaluation index conforming to the characteristics of human visual system, structural similarity (SSIM)\(^26\), was proposed. The advantage of SSIM is that it assesses the reconstructed image from the perspective of luminance, contrast and structure. Thus, we select the PSNR and SSIM as quality evaluation criteria to quantitatively assess the reconstruction performance.

The PSNR and SSIM results of all methods are shown in Table 1. The proposed method has an average PSNR gain of up to 1.5dB, 0.31dB and 0.46dB compared with NLM, BM3D and EPLL, respectively. Overall, the proposed method achieves better numerical results.
Table 1 Results with different methods for PSNR and SSIM

<table>
<thead>
<tr>
<th>Images</th>
<th>Noise level</th>
<th>NLM</th>
<th>BM3D</th>
<th>EPLL</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR/dB</td>
<td>SSIM</td>
<td>PSNR/dB</td>
<td>SSIM</td>
<td>PSNR/dB</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barbara</td>
<td>5</td>
<td>35.62</td>
<td>0.963 3</td>
<td>37.96</td>
<td>0.974 2</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>33.31</td>
<td>0.933 5</td>
<td>34.46</td>
<td>0.953 7</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>31.48</td>
<td>0.906 1</td>
<td>32.54</td>
<td>0.932 8</td>
</tr>
<tr>
<td>Starfish</td>
<td>5</td>
<td>37.32</td>
<td>0.970 0</td>
<td>37.24</td>
<td>0.969 7</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>33.28</td>
<td>0.935 7</td>
<td>33.27</td>
<td>0.932 5</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>31.09</td>
<td>0.904 2</td>
<td>31.06</td>
<td>0.901 9</td>
</tr>
<tr>
<td>Light</td>
<td>5</td>
<td>32.07</td>
<td>0.975 8</td>
<td>35.94</td>
<td>0.986 7</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>29.71</td>
<td>0.950 0</td>
<td>31.33</td>
<td>0.968 3</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>27.46</td>
<td>0.917 8</td>
<td>28.88</td>
<td>0.949 2</td>
</tr>
<tr>
<td>Parrot</td>
<td>5</td>
<td>37.31</td>
<td>0.953 7</td>
<td>38.62</td>
<td>0.962 7</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>34.04</td>
<td>0.906 6</td>
<td>34.95</td>
<td>0.936 6</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>31.99</td>
<td>0.854 6</td>
<td>32.97</td>
<td>0.917 5</td>
</tr>
<tr>
<td>Kodim</td>
<td>5</td>
<td>36.98</td>
<td>0.958 6</td>
<td>38.59</td>
<td>0.970 2</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>33.27</td>
<td>0.911 8</td>
<td>34.85</td>
<td>0.942 0</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>31.03</td>
<td>0.859 2</td>
<td>32.73</td>
<td>0.915 9</td>
</tr>
<tr>
<td>Lena</td>
<td>5</td>
<td>36.99</td>
<td>0.958 7</td>
<td>38.55</td>
<td>0.970 0</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>33.26</td>
<td>0.907 7</td>
<td>34.72</td>
<td>0.945 2</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>31.28</td>
<td>0.858 8</td>
<td>32.59</td>
<td>0.921 4</td>
</tr>
</tbody>
</table>

4 Conclusion

This paper proposed a new group sparse denoising model, in which the weighted nuclear norm and $l_2$ mixed norm are employed as regularization terms to simultaneously enforce the priors on low and high frequency of the image. The proposed model can be solved efficiently based on the ADMM framework. Experimental results have demonstrated that the proposed method achieves comparable performance to some state-of-the-art methods.

References


[18] Donoho D L. For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution[J]. Communications on Pure and Applied Mathematics, 2006, 59(6): 797-829.


