Multichannel Color Image Denoising via Weighted Schatten p-norm Minimization

Xinjian Huang , **Bo Du*** and **Weiwei Liu***
School of Computer Science, Institute of Artificial Intelligence and
National Engineering Research Center for Multimedia Software, Wuhan University, Wuhan, China
h_xj525279@outlook.com, dubo@whu.edu.cn, liuweiwei863@gmail.com

Abstract

The R, G and B channels of a color image generally have different noise statistical properties or noise strengths. It is thus problematic to apply grayscale image denoising algorithms to color image denoising. In this paper, based on the non-local self-similarity of an image and the different noise strength across each channel, we propose a MultiChannel Weighted Schatten p-Norm Minimization (MCWSNM) model for RGB color image denoising. More specifically, considering a small local RGB patch in a noisy image, we first find its nonlocal similar cubic patches in a search window with an appropriate size. These similar cubic patches are then vectorized and grouped to construct a noisy low-rank matrix, which can be recovered using the Schatten p-norm minimization framework. Moreover, a weight matrix is introduced to balance each channel's contribution to the final denoising results. The proposed MCWSNM can be solved via the alternating direction method of multipliers. Convergence property of the proposed method are also theoretically analyzed . Experiments conducted on both synthetic and real noisy color image datasets demonstrate highly competitive denoising performance, outperforming comparison algorithms, including several methods based on neural networks.

1 Introduction

Noise corruption is inevitable during the image acquisition process and may heavily degrade the visual quality of an acquired image. Image denoising is thus an essential preprocessing step in various image processing and computer vision tasks [Chatterjee and Milanfar, 2010; Ye *et al.*, 2018; Wang *et al.*, 2018b]; moreover, it is also an ideal test platform for evaluating image prior models and optimization methods [Roth and Black, 2005]. As a result, image denoising remains a challenging yet fundamental problem. Early denoising algorithms were mainly devised on the basis of filter and transformation [Dabov *et al.*, 2007b], such as wavelet transform and curvelet transform [Starck *et al.*, 2002]. State-of-the-art denoising methods are mainly based on sparse representation [Dabov *et al.*, 2007b], low-rank approximation [Gu *et al.*, 2017; Xie *et al.*, 2016], dictionary learning [Zhang and Aeron, 2016; Marsousi *et al.*, 2014; Mairal *et al.*, 2012], nonlocal self-similarity [Buades *et al.*, 2005; Dong *et al.*, 2013a; Hu *et al.*, 2019] and neural networks [Liu *et al.*, 2018; Zhang *et al.*, 2017a; Zhang *et al.*, 2017b; Zhang *et al.*, 2018].

Current methods for RGB color image denoising can be categorized into three classes. The first kind involves applying grayscale image denoising algorithms to each channel in a channel-wise manner. However, these methods ignore the correlations between R, G and B channels, meaning that unsatisfactory results may be obtained. The second method is to transform the RGB color image into other color spaces [Dabov *et al.*, 2007a]. This transform, however, may change the noise distribution of the original observation data and introduce artifacts. The third type of method involves making full use of the correlation information across each channel and conduct the denoising task on R, G and B channels simultaneously [Zhang *et al.*, 2017a].

Similar to the case of grayscale images, noise from each channel can be generally, regarded as additive white Gaussian noise. However, the noise levels of each channel are diverse due to camera sensor characteristics and the imagery environment, such as fog, haze, illumination intensity, etc. Moreover, the on-board processing steps in digital camera pipelines may also introduce different noise strength assigned to different channels [Nam *et al.*, 2016]. This reveals the difficulties and challenges faced by RGB color image denoising. Intuitively, if we apply the grayscale image denoising algorithm to color images in a band-wise manner without considering the mutual information and noise difference between each channel, artifacts or false colors could be generated [Mairal *et al.*, 2008]. An example of this is presented in Fig.1. Here, we apply a representative low-rank-based method, WSNM [Xie *et al.*, 2016], and our proposed method, MCWSNM (see Section 2), to the Kodak PhotoCD Dataset[1]. It can be seen from the figure that WSNM retains a large amount of noise and, to some extent, introduces some artifacts. Thus, the issue of how noise differences in each channel should be modeled is key to designing a good RGB color denoising algorithm.

One well-known color image denoising method is color

*Co-corresponding Author.

[1]http://r0k.us/graphics/kodak/

(a) Clean (b) Noisy (c) WSNM (d) MCWSNM

Figure 1: Denoised results of "kodim07" in the Kodak PhotoCD Dataset with $\sigma_r = 50$, $\sigma_g = 5$ and $\sigma_b = 20$. WSNM performs in a band-wise manner; MCWSNM jointly processes three channels jointly and considers their noise differences simultaneously. (It is best to zoom in on these images on-screen).

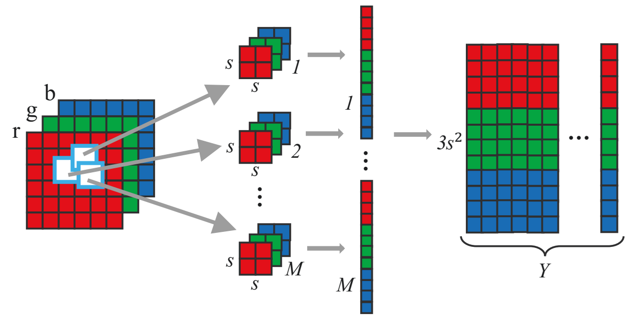


Figure 2: Illusion of the grouped local patches.

block-matching 3D (CBM3D) [Dabov *et al.*, 2007a]. This is a transformation field method that may destroy the noise distribution across each channel and introduce artifacts. The method in [Zhu *et al.*, 2016] concatenates similar patches of RGB channels into a long vector. However, this concatenation operation regards each channel equally and ignores the noise level differences across each channel. For a long time, low-rank theory has been widely used in many fields, such as recommendation system [Wang *et al.*, 2018a], community search [Fang *et al.*, 2020b; Fang *et al.*, 2020a], multi-out task [Liu *et al.*, 2019] and multi-label[Liu *et al.*, 2017], etc.

Recently, based on the non-local self-similarity of an image and SVD, several low-rank-based denoising algorithms have been proposed, such as weighted nuclear norm minimization (WNNM) [Gu *et al.*, 2017] and WSNM [Xie *et al.*, 2016]. The nuclear norm and Schatten p-norm [Xie *et al.*, 2016] are two surrogates of the rank function [Xu *et al.*, 2017a]. In WNNM and WSNM, singular values are assigned different weights using a weight vector. It has been proven that WNNM is a special case of WSNM and that WSNM outperforms WNNM. Multi-channel WNNM (MCWNNM) [Xu *et al.*, 2017b] is an extension of WNNM from grayscale image to color image that operates by introducing a weight matrix to adjust each channel's contribution to the final results based on noise level.

In this paper, based on the low-rank property of the non-local self-similar patches and the different noise strength across each channel, we propose a multi-channel weighted Schatten p-norm minimization (MCWSNM) model for RGB color image denoising. More specifically, considering a small local RGB patch in a noisy image, we first find its non-local similar cubic patches in a search window with an appropriate size. These similar cubic patches are then vectorized and concatenated to construct a noisy low-rank matrix; then Schatten p-norm minimization method is performed to recover the noise-free low-rank matrix. Moreover, in order to make full use of the redundant information across each channel, a weight matrix is introduced to balance each channel's contribution to the final denoising results. This model can be solved via alternating direction method of multipliers (ADMM) framework, and each variable can be updated with a closed-form solution. In addition, the convergence property of our proposed algorithm is also provided.

2 Our Proposed Method

2.1 Method Formulation

The image denoising task involves recovering a clean image $\boldsymbol{x}_c$ from its noisy observation data $\boldsymbol{y}_c$, where $c = \{r, g, b\}$ is the index of the R, G and B channels. Here, we assume that the noisy image is corrupted by additive white Gaussian noise $\boldsymbol{n}_c$ with $\sigma_c = (\sigma_r, \sigma_g, \sigma_b)$, where $\sigma_r, \sigma_g, \sigma_b$ denote the noise standard deviation in the R, G, B channels, respectively. Mathematically, $\boldsymbol{y}_c = \boldsymbol{x}_c + \boldsymbol{n}_c$. Low-rank-based denoising methods utilize the low-rank property of the grouped non-local similarity patches; this group procedure is illustrated in Fig.2. Given a noisy RGB color image $\boldsymbol{y}_c$, each local patch of size $s \times s \times 3$ is extracted and stretched into a vector $\boldsymbol{y} = (\boldsymbol{y}_r^T, \boldsymbol{y}_g^T, \boldsymbol{y}_b^T)^T \in \mathbb{R}^{3s^2}$, where $\boldsymbol{y}_r, \boldsymbol{y}_g, \boldsymbol{y}_b \in \mathbb{R}^{s^2}$ are the corresponding patches in the R, G and B channels. For each vectorized local patch $\boldsymbol{y}$, we find its M most similar patches (including $\boldsymbol{y}$ itself) by Euclidean distance in a proper size search window around it. We then stack the M similar patches vector column by column to obtain the noisy patch matrix $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{N} \in \mathbb{R}^{3s^2 \times M}$, where $\boldsymbol{X}$ and $\boldsymbol{N}$ are the corresponding clean and noise patch matrix, respectively. It is clear that $\boldsymbol{Y}$ is a low-rank matrix and $\boldsymbol{X}$ can be recovered using a low-rank approximation algorithm.

As a nonconvex surrogate of the rank function, the weighted Schatten p-norm of a matrix $\boldsymbol{Z} \in \mathbb{R}^{m \times n}$ is defined as $\|\boldsymbol{Z}\|_{\boldsymbol{w},S_p} = (\sum_{i=1}^{min\{m,n\}} w_i \sigma_i^p)^{\frac{1}{p}}, 0 < p < 1$, where $\boldsymbol{w} = (w_1, w_2, \ldots, w_{min\{m,n\}})^T$ is a non-negative weight vector and σ_i is the i-th singular value of $\boldsymbol{Z}$. The weighted nuclear norm [Gu *et al.*, 2017] is a special case of the weighted Schatten p-norm when the power $p = 1$. It has been proven that the Schatten p-norm with $0 < p < 1$ is a better approximation of rank function than the nuclear norm [Zhang *et al.*, 2013].

Intuitively, when denoising a noisy color image, if a channel is heavily polluted, the contribution of this channel to the final denoised results should be less, and vice versa. Thus, a weight matrix $\boldsymbol{W}$ assigned to the noise strength of each channel is introduced. Based on the low-rank property of the non-local self-similarity, we propose the following multi-channel WSNM (MCWSNM) model to recover the local low-rank patch $\boldsymbol{X}$:

$$\min_{\boldsymbol{X}} \|\boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X})\|_F^2 + \lambda \|\boldsymbol{X}\|_{\boldsymbol{w},S_p}^p, \tag{1}$$

where $\lambda > 0$ is a tradeoff parameter used to balance the

Algorithm 1 $\min_{\boldsymbol{X}} \|\boldsymbol{Y}-\boldsymbol{X}\|_F^2 + \lambda\|\boldsymbol{X}\|_{\boldsymbol{w},S_p}^p$ via GST

Input: $\boldsymbol{Y}$, weight $\{w_i\}_{i=1}^r, \lambda, p, K$

Singular value decomposition $\boldsymbol{Y} = \boldsymbol{U\Sigma V}^T$, $\boldsymbol{\Sigma} = diag\{\sigma_1, \sigma_2, \ldots, \sigma_r\}$;
for $i = 1 : r$ **do**
  $\tau_p^{GST}(w_i, \lambda) = (2\lambda w_i(1-p))^{\frac{1}{2-p}} + \lambda w_i p(2\lambda w_i(1-p))^{\frac{p-1}{2-p}}$;
  if $|\sigma_i| \leq \tau_p^{GST}(w_i)$ **then**
    $\delta_i = 0$;
  else
    $\delta_i^{(k)} = |\sigma_i|$;
    for $k = 0, 1, \ldots, K$ **do**
      $\delta_i^{(k+1)} = |\sigma_i| - w_i p(\delta_i^{(k)})^{p-1}$;
      $k = k + 1$;
    end for
    $\delta_i = \mathrm{sgn}(\sigma_i)\delta_i^{(k)}$;
  end if
end for
$\boldsymbol{\Delta} = diag(\delta_1, \delta_2, \ldots, \delta_r)$;

Output: $\boldsymbol{X}^* = \boldsymbol{U\Delta V}^T$

data fidelity term and the regularization term, while $\boldsymbol{W} = \mathrm{diag}(\tau_r\boldsymbol{I}, \tau_g\boldsymbol{I}, \tau_b\boldsymbol{I})$, $\boldsymbol{I} \in \mathbb{R}^{s^2\times s^2}$ is the identity matrix, $(\tau_r, \tau_g, \tau_b) = \min\{\sigma_r, \sigma_g, \sigma_b\}/(\sigma_r, \sigma_g, \sigma_b)$.

2.2 Model Optimization

We utilize the variable splitting method to the MCWSNM model (1). By introducing an augmented variable $\boldsymbol{Z}$, model (1) can be reformulated as a linear constrained optimization problem. Thus, model (1) can be expressed as follows:

$$\min_{\boldsymbol{X},\boldsymbol{Z}} \quad \|\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{X})\|_F^2 + \lambda\|\boldsymbol{Z}\|_{\boldsymbol{w},S_p}^p, \ s.t.\ \boldsymbol{X} = \boldsymbol{Z}. \quad (2)$$

Because the two variables $\boldsymbol{X}$ and $\boldsymbol{Z}$ in problem (2) are separable, this programming problem can be solved using the ADMM framework. The augmented Lagrangian function of (2) is

$$\mathcal{L}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{L}, \rho) = \|\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{X})\|_F^2 + \lambda\|\boldsymbol{Z}\|_{\boldsymbol{w},S_p}^p + \mathrm{Tr}(\boldsymbol{L}^T(\boldsymbol{X}-\boldsymbol{Z})) + \frac{\rho}{2}\|\boldsymbol{X}-\boldsymbol{Z}\|_F^2, \quad (3)$$

where $\boldsymbol{L}$ is the augmented Lagrangian multiplier and $\rho > 0$ is the penalty parameter. By taking derivatives of $\mathcal{L}$ with respect to $\boldsymbol{X}$ and $\boldsymbol{Z}$ and setting the derivative function to be zero, the variables can be alternatively updated as follows:

(1) $\boldsymbol{X}_{k+1}$

$$= (\boldsymbol{W}^T\boldsymbol{W} + \frac{\rho_k}{2}\boldsymbol{I})^{-1}(\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{Y} + \frac{\rho_k}{2}\boldsymbol{Z}_k - \frac{1}{2}\boldsymbol{L}_k). \quad (4)$$

(2) $\boldsymbol{Z}_{k+1}$

$$= \arg\min_{\boldsymbol{Z}} \frac{\rho_k}{2}\|\boldsymbol{Z} - (\boldsymbol{X}_{k+1} + \frac{1}{\rho_k}\boldsymbol{L}_k)\|_F^2 + \lambda\|\boldsymbol{Z}\|_{\boldsymbol{w},S_p}^p. \quad (5)$$

This is a weighted Schatten p-norm minimization problem. It was noted in [Xie *et al.*, 2016] that if the weights follow the non-descending permutation, (5) can be equivalently transformed into independent nonconvex ℓ_p-norm subproblems, the global optima of which can be efficiently solved by the generalized soft-thresholding (GST) algorithm. This is summarized in Algorithm 1.

(3) $\boldsymbol{L}_{k+1} = \boldsymbol{L}_k + \rho_k(\boldsymbol{X}_{k+1} - \boldsymbol{Z}_{k+1})$.

(4) $\rho_{k+1} = \mu\rho_k, \quad (\mu > 1)$.

The above alternative updating steps are repeated until either the convergence condition is satisfied or the number of iterations exceeds a preset K_1. The ADMM algorithm converges when 1) $\|\boldsymbol{X}_{k+1} - \boldsymbol{Z}_{k+1}\|_F \leq tol$, 2) $\|\boldsymbol{X}_{k+1} - \boldsymbol{X}_k\|_F \leq tol$ and 3) $\|\boldsymbol{Z}_{k+1} - \boldsymbol{Z}_k\|_F \leq tol$ are simultaneously satisfied; here, tol is a small tolerance value. The complete updating procedures are summarized in Algorithm 2.

Algorithm 2 Solve MCWSNM via ADMM

Input: data $\boldsymbol{Y}$, weight $\boldsymbol{W}$, p, K_1, $\mu > 1$, $tol > 0$;

Initialization: $\boldsymbol{X}_0 = \boldsymbol{Z}_0 = \boldsymbol{0}$, $\boldsymbol{L}_0 = \boldsymbol{0}$, $\rho_0 > 0$, flag = False, $k = 0$;
while flag == False **do**
  Update $\boldsymbol{X}$ by using (4);
  Update $\boldsymbol{Z}$ by solving (5);
  Update $\boldsymbol{L}$: $\boldsymbol{L}_{k+1} = \boldsymbol{L}_k + \rho_k(\boldsymbol{X}_{k+1} - \boldsymbol{Z}_{k+1})$;
  Update ρ: $\rho_{k+1} = \mu\rho_k$;
  $k = k + 1$;
  if (Convergence condition is satisfied) or ($k \geq K_1$) **then**
    flag = True;
  end if
end while

Output: $\boldsymbol{X}^*$.

2.3 Convergence and Complexity

Although the proposed MCWSNM (1) is nonconvex, the global optimum also can be obtained for the limitations of weights permutation in GST. From Fig.3, it can be seen that $\|\boldsymbol{X}_{k+1} - \boldsymbol{X}_k\|_2$, $\|\boldsymbol{Z}_{k+1} - \boldsymbol{Z}_k\|_2$ and $\|\boldsymbol{X}_{k+1} - \boldsymbol{Z}_{k+1}\|_2$ simultaneously approach 0 during the iteration process. This curve is based on a synthetic experiment on Kodak PhotoCD Dataset and the test image is ”kodim01” (see Experiments Section in detail).

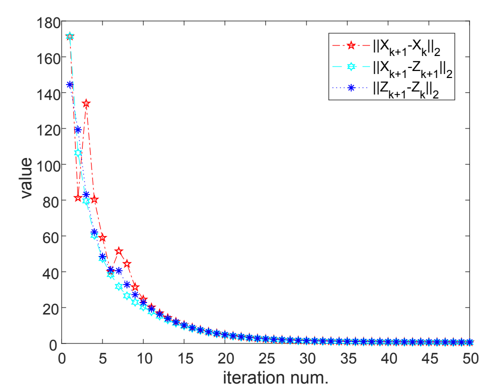


Figure 3: The convergence curves of $\|\boldsymbol{X}_{k+1} - \boldsymbol{X}_k\|_2$, $\|\boldsymbol{Z}_{k+1} - \boldsymbol{Z}_k\|_2$ and $\|\boldsymbol{X}_{k+1} - \boldsymbol{Z}_{k+1}\|_2$ of image ”kodim01” in the Kodak Dataset.

We next provide a theorem that theoretically guarantees the convergence property of Algorithm 2.

Theorem 1. *Assume that the weights $\boldsymbol{w}$ are sorted in a non-descending order and the parameter ρ_k is unbounded; then the sequences $\{\boldsymbol{X}_k\}$, $\{\boldsymbol{Z}_k\}$ and $\{\boldsymbol{L}_k\}$ generated in Algorithm 2 satisfy:* $(a) \lim_{k\to+\infty} \|\boldsymbol{X}_{k+1} - \boldsymbol{Z}_{k+1}\|_F = 0$; $(b) \lim_{k\to+\infty} \|\boldsymbol{X}_{k+1} - \boldsymbol{X}_k\|_F = 0$; $(c) \lim_{k\to+\infty} \|\boldsymbol{Z}_{k+1} - \boldsymbol{Z}_k\|_F = 0$.

Proof. We first prove that the sequence $\{\boldsymbol{L}_k\}$ generated by Algorithm 2 is upper bounded. $\boldsymbol{U}_k\boldsymbol{\Lambda}_k\boldsymbol{V}_k^T$ denotes the SVD of matrix $\frac{1}{\rho_k}\boldsymbol{L}_k + \boldsymbol{X}_{k+1}$ in the $k+1$ iteration, where $\boldsymbol{\Lambda}_k$ is the diagonal singular value matrix. By using the GST algorithm for WSNM, we have $\boldsymbol{Z}_{k+1} = \boldsymbol{U}_k\boldsymbol{\Delta}_k\boldsymbol{V}_k^T$, where $\boldsymbol{\Delta}_k = \{diag(\delta_k^1, \delta_k^2, \ldots, \delta_k^n)\}$ is the diagonal singular value matrix after generalized soft-thresholding operation. Then:

$$
\begin{aligned}
&\|\boldsymbol{L}_{k+1}\|_F^2 = \boldsymbol{L}_k + \rho_k(\boldsymbol{X}_{k+1} - \boldsymbol{Z}_{k+1})\|_F^2 \\
&= \rho_k^2\|\frac{1}{\rho_k}\boldsymbol{L}_k + \boldsymbol{X}_{k+1} - \boldsymbol{Z}_{k+1}\|_F^2 \\
&= \rho_k^2\|\boldsymbol{U}_k\boldsymbol{\Lambda}_k\boldsymbol{V}_k^T - \boldsymbol{U}_k\boldsymbol{\Delta}_k\boldsymbol{V}_k^T\|_F^2 = \rho_k^2\|\boldsymbol{\Lambda}_k - \boldsymbol{\Delta}_k\|_F^2 \\
&= \rho_k^2\|\sum_i rw_i/\rho_k\|_F^2 = \|r\sum_i w_i\|_F^2.
\end{aligned}
$$

Hence, the sequence $\{\boldsymbol{L}_k\}$ is upper bounded.

We then prove that the sequence of the Lagrange function $\{\mathcal{L}(\boldsymbol{X}_{k+1}, \boldsymbol{Z}_{k+1}, \boldsymbol{L}_{k+1}, \rho_{k+1})\}$ is also upper bounded. The inequality $\mathcal{L}(\boldsymbol{X}_{k+1}, \boldsymbol{Z}_{k+1}, \boldsymbol{L}_k, \rho_k) \leq \mathcal{L}(\boldsymbol{X}_k, \boldsymbol{Z}_k, \boldsymbol{L}_k, \rho_k)$ always holds since we have the globally optimal solution of $\boldsymbol{X}$ and $\boldsymbol{Z}$ in their corresponding subproblems (step 3 and step 4 in Algorithm 2. Based on the updating rule of $\boldsymbol{L}$, it yields

$$
\begin{aligned}
&\mathcal{L}(\boldsymbol{X}_{k+1}, \boldsymbol{Z}_{k+1}, \boldsymbol{L}_{k+1}, \rho_{k+1}) \\
&= \|\boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X}_{k+1})\|_F^2 + \|\boldsymbol{Z}_{k+1}\|_{\boldsymbol{w},S_p}^p \\
&+ \langle \boldsymbol{L}_{k+1}, \boldsymbol{X}_{k+1} - \boldsymbol{Z}_{k+1}\rangle + \frac{\rho_{k+1}}{2}\|\boldsymbol{X}_{k+1} - \boldsymbol{Z}_{k+1}\|_F^2 \\
&= \mathcal{L}(\boldsymbol{X}_{k+1}, \boldsymbol{Z}_{k+1}, \boldsymbol{L}_k, \rho_k) + \langle \boldsymbol{L}_{k+1} - \boldsymbol{L}_k, \boldsymbol{X}_{k+1} - \boldsymbol{Z}_{k+1}\rangle \\
&+ \frac{\rho_{k+1} - \rho_k}{2}\|\boldsymbol{X}_{k+1} - \boldsymbol{Z}_{k+1}\|_F^2 \\
&= \mathcal{L}(\boldsymbol{X}_{k+1}, \boldsymbol{Z}_{k+1}, \boldsymbol{L}_k, \rho_k) + \langle \boldsymbol{L}_{k+1} - \boldsymbol{L}_k, \frac{\boldsymbol{L}_{k+1} - \boldsymbol{L}_k}{\rho_k}\rangle \\
&+ \frac{\rho_{k+1} - \rho_k}{2}\|\frac{\boldsymbol{L}_{k+1} - \boldsymbol{L}_k}{\rho_k}\|_F^2 \\
&= \mathcal{L}(\boldsymbol{X}_{k+1}, \boldsymbol{Z}_{k+1}, \boldsymbol{L}_k, \rho_k) + \frac{\rho_{k+1} + \rho_k}{2\rho_k^2}\|\boldsymbol{L}_{k+1} - \boldsymbol{L}_k\|_F^2.
\end{aligned}
$$

Since $\{\boldsymbol{L}_k\}$ is upper bounded, the sequence $\{\boldsymbol{L}_{k+1} - \boldsymbol{L}_k\}$ is also upper bounded. Denote by a the upper bound of $\{\boldsymbol{L}_{k+1} - \boldsymbol{L}_k\}$ for all $k \geq 0$, i.e. $\{\boldsymbol{L}_{k+1} - \boldsymbol{L}_k\} \leq a$, $\forall k \geq 0$. We therefore conclude that

$$
\begin{aligned}
&\mathcal{L}(\boldsymbol{X}_{k+1}, \boldsymbol{Z}_{k+1}, \boldsymbol{L}_{k+1}, \rho_{k+1}) \\
&\leq \mathcal{L}(\boldsymbol{X}_{k+1}, \boldsymbol{Z}_{k+1}, \boldsymbol{L}_k, \rho_k) + \frac{\rho_{k+1} + \rho_k}{2\rho_k^2}a^2 \\
&\leq \mathcal{L}(\boldsymbol{X}_1, \boldsymbol{Z}_1, \boldsymbol{L}_0, \rho_0) + a^2\textstyle\sum_{k=0}^{\infty}\frac{\rho_{k+1}+\rho_k}{2\rho_k^2} \\
&= \mathcal{L}(\boldsymbol{X}_1, \boldsymbol{Z}_1, \boldsymbol{L}_0, \rho_0) + a^2\textstyle\sum_{k=0}^{\infty}\frac{1+\mu}{2\rho_0\mu^k} \\
&\leq \mathcal{L}(\boldsymbol{X}_1, \boldsymbol{Z}_1, \boldsymbol{L}_0, \rho_0) + \frac{a^2}{\rho_0}\textstyle\sum_{k=0}^{\infty}\frac{1}{\mu^{k-1}} \\
&< +\infty.
\end{aligned}
$$

Thus, $\{\mathcal{L}(\boldsymbol{X}_{k+1}, \boldsymbol{Z}_{k+1}, \boldsymbol{L}_{k+1}, \rho_{k+1})\}$ is upper bounded.

We next prove the sequences of $\{\boldsymbol{X}_k\}$ and $\{\boldsymbol{Z}_k\}$ are upper bounded. Since $\{\mathcal{L}(\boldsymbol{X}_k, \boldsymbol{Z}_k, \boldsymbol{L}_k, \rho_k)\}$ and $\{\boldsymbol{L}_k\}$ are upper bounded and

$$
\begin{aligned}
&\|\boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X}_k)\|_F^2 + \|\boldsymbol{Z}_k\|_{\boldsymbol{w},S_p}^p \\
&= \mathcal{L}(\boldsymbol{X}_k, \boldsymbol{Z}_k, \boldsymbol{L}_{k-1}, \rho_{k-1}) - \langle \boldsymbol{L}_{k-1}, \boldsymbol{X}_k - \boldsymbol{Z}_k\rangle \\
&- \frac{\rho_{k-1}}{2}\|\boldsymbol{X}_k - \boldsymbol{Z}_k\|_F^2 \\
&= \mathcal{L}(\boldsymbol{X}_k, \boldsymbol{Z}_k, \boldsymbol{L}_{k-1}, \rho_{k-1}) - \langle \boldsymbol{L}_{k-1}, (\boldsymbol{L}_k - \boldsymbol{L}_{k-1})/\rho_{k-1}\rangle \\
&- \frac{\rho_{k-1}}{2}\|\frac{\boldsymbol{L}_k - \boldsymbol{L}_{k-1}}{\rho_{k-1}}\|_F^2 \\
&= \mathcal{L}(\boldsymbol{X}_k, \boldsymbol{Z}_k, \boldsymbol{L}_{k-1}, \rho_{k-1}) + \frac{\|\boldsymbol{L}_{k-1}\|_F^2 - \|\boldsymbol{L}_k\|_F^2}{2\rho_{k-1}},
\end{aligned}
$$

and the sequence $\{\boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X}_k)\}$ and $\{\boldsymbol{Z}_k\}$ are upper bounded. Since $\boldsymbol{L}_{k+1} = \boldsymbol{L}_k + \rho_k(\boldsymbol{X}_{k+1} - \boldsymbol{Z}_{k+1})$, $\{\boldsymbol{X}_k\}$ is also upper bounded. Thus, there exists at least one accumulation point for $\{\boldsymbol{X}_k, \boldsymbol{Z}_k\}$. More specifically, we obtain that

$$
\lim_{k\to+\infty}\|\boldsymbol{X}_{k+1} - \boldsymbol{Z}_{k+1}\|_F = \lim_{k\to+\infty}\frac{1}{\rho_k}\|\boldsymbol{L}_{k+1} - \boldsymbol{L}_k\|_F = 0
$$

and that the accumulation point is a feasible solution to the objective function. Thus, equation (a) is proved.

Finally, we prove that the change of sequence $\{\boldsymbol{X}_k\}$ and $\{\boldsymbol{Z}_k\}$ in adjacent iterations tends to be 0. For $\boldsymbol{X}_{k+1}$ and $\boldsymbol{X}_k = \frac{1}{\rho_{k-1}}(\boldsymbol{L}_k - \boldsymbol{L}_{k-1}) + \boldsymbol{Z}_k$, we have

$$
\begin{aligned}
&\lim_{k\to\infty}\|\boldsymbol{X}_{k+1} - \boldsymbol{X}_k\|_F \\
&= \lim_{k\to\infty}\|(\boldsymbol{W}^T\boldsymbol{W} + \frac{\rho_k}{2}\boldsymbol{I})^{-1}(\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{Y} + \frac{\rho_k}{2}\boldsymbol{Z}_k - \frac{1}{2}\boldsymbol{L}_k) \\
&- \frac{1}{\rho_{k-1}}(\boldsymbol{L}_k - \boldsymbol{L}_{k-1}) - \boldsymbol{Z}_k\|_F \\
&= \lim_{k\to\infty}\|(\boldsymbol{W}^T\boldsymbol{W} + \frac{\rho_k}{2}\boldsymbol{I})^{-1}(\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{Y} - \boldsymbol{W}^T\boldsymbol{W}\boldsymbol{Z}_k \\
&- \frac{1}{2}\boldsymbol{L}_k) - \frac{1}{\rho_{k-1}}(\boldsymbol{L}_k - \boldsymbol{L}_{k-1})\|_F \\
&\leq \lim_{k\to\infty}\|(\boldsymbol{W}^T\boldsymbol{W} + \frac{\rho_k}{2}\boldsymbol{I})^{-1}(\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{Y} - \boldsymbol{W}^T\boldsymbol{W}\boldsymbol{Z}_k \\
&- \frac{1}{2}\boldsymbol{L}_k)\|_F + \frac{1}{\rho_{k-1}}\|(\boldsymbol{L}_k - \boldsymbol{L}_{k-1})\|_F \\
&= 0.
\end{aligned}
$$

Since $\boldsymbol{Z}_{k+1} = \frac{1}{\rho_k}\boldsymbol{L}_k - \frac{1}{\rho_k}\boldsymbol{L}_{k+1} + \boldsymbol{X}_{k+1}$, we conclude that

$$\begin{aligned}
&\lim_{k\to\infty} \|\boldsymbol{Z}_{k+1} - \boldsymbol{Z}_k\|_F \\
&= \lim_{k\to\infty} \|\frac{1}{\rho_k}\boldsymbol{L}_k - \frac{1}{\rho_k}\boldsymbol{L}_{k+1} + \boldsymbol{X}_{k+1} - \boldsymbol{Z}_k\|_F \\
&= \lim_{k\to\infty} \|\boldsymbol{X}_k + \frac{1}{\rho_{k-1}}\boldsymbol{L}_{k-1} - \boldsymbol{Z}_k + \boldsymbol{X}_{k+1} \\
&- \boldsymbol{X}_k - \frac{1}{\rho_{k-1}}\boldsymbol{L}_{k-1} + \frac{1}{\rho_k}\boldsymbol{L}_k - \frac{1}{\rho_k}\boldsymbol{L}_{k+1}\|_F \\
&\leq \lim_{k\to\infty} \|\sum_i \frac{rw_i}{\rho_{k-1}}\|_F + \|\boldsymbol{X}_{k+1} - \boldsymbol{X}_k\| \\
&+ \|\frac{1}{\rho_{k-1}}\boldsymbol{L}_{k-1} - \frac{1}{\rho_k}\boldsymbol{L}_k + \frac{1}{\rho_k}\boldsymbol{L}_{k+1}\|_F = 0.
\end{aligned}$$

This completes the proof. □

The main computational cost in a single iteration of Algorithm 2 consists of two parts. The first part involves updating $\boldsymbol{X}$, in which the time complexity is $\mathcal{O}(\max\{s^4M, M^3\})$. The second part involves updating $\boldsymbol{Z}$. The predominant cost of this updating is the SVD and the calculation of $\boldsymbol{X}^*$. Its complexity is $\mathcal{O}(\min\{s^4M, s^2M^2\} + s^2r^4M)$. The costs for updating $\boldsymbol{L}$ and ρ can be ignored. Therefore, the total time complexity of our MCWSNM for solving problem (1) is about $\mathcal{O}(K_1M^3)$.

3 Experiments

To illustrate the performance of the proposed MCWSNM, we implement MCWSNM on synthetic and real color noisy images. We also compare the proposed method with several recent proposed methods, including NC [Lebrun *et al.*, 2015], NCSR [Dong *et al.*, 2013b], PGPD [Xu *et al.*, 2015], MCWNNM [Xu *et al.*, 2017b], DnCNN [Zhang *et al.*, 2017a], FFDNet [Zhang *et al.*, 2018] and IRCNN [Zhang *et al.*, 2017b]. In more detail, NC is an online blind image denoising platform, NCSR and PGPD are designed for grayscale images and MCWNNM is a color image denoising method while DnCNN, FFDNet, and IRCNN are three methods that operate on the basis of convolutional neural networks. All the parameters of the comparison algorithms are either optimally assigned, or chosen as described in the reference papers.

Noise level of MCWSNM and most comparative methods should be provided as parameters. In the synthetic experimental case, the noise $(\sigma_r, \sigma_g, \sigma_b)$ in the R, G and B channels are assumed to be known. In the case involving real noise, we utilize the noise estimation method outlined in [Chen *et al.*, 2015] to estimate the noise level of the noisy image for each channel. We implement the grayscale denoising methods, NCSR and PGPD, on the color noisy images in a bandwise manner with the corresponding channel noise level.

3.1 Synthetic Noisy Color Image Experiments

The Kodak PhotoCD Dataset is first utilized in our synthetic experiments. It includes 24 color images, each of which is either 768×512 or 512×768 in size. The noisy image is generated by adding zero mean Gaussian noise with $\sigma_r = 40$, $\sigma_g = 20$ and $\sigma_b = 30$. In MCWSNM, we set the local search window size of each patch as 20, the similar patch number $M = 70$, each patch size $s = 6$, $K_1 = 10$, $K = 4$, $c = 2\sqrt{2}$, $\lambda = 0.6$, $p = 0.999$ and $\rho = 3$.

The line graph of the PSNR results for MCWSNM and the comparison methods are presented in Fig.4. It can be seen from this graph that our proposed method outperforms the other competing methods in most cases. Moreover, Fig.5 and Fig.6 show the denoised results of "kodim1" and "kodim3"

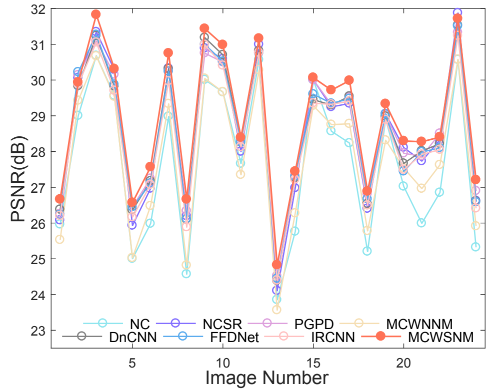


Figure 4: Line graph of the denoising results in Kodak Dataset.

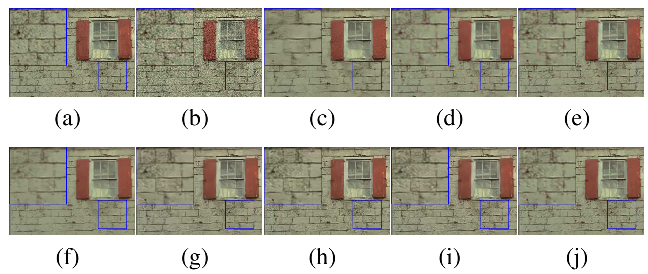


Figure 5: Denoised images of "kodim1" in Kodak PhotoCD Dataset with $\sigma_r = 40$, $\sigma_g = 20$ and $\sigma_b = 30$. ((a): Original, (b): Noisy, (c): NC, (d): NCSR, (e): PGPD, (f): MCWNNM, (g): DnCNN, (h): FFDNet, (i): IRCNN, (j): MCWSNM).

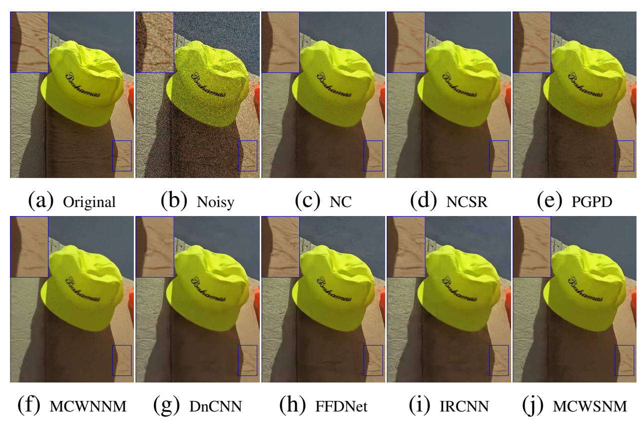
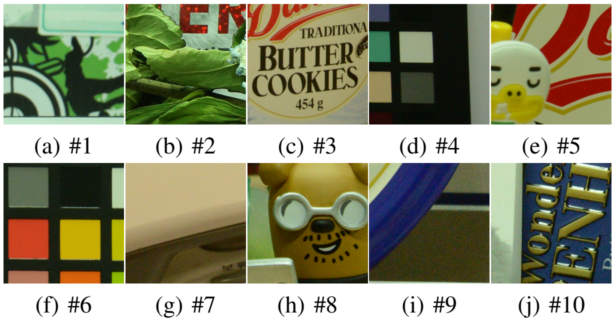
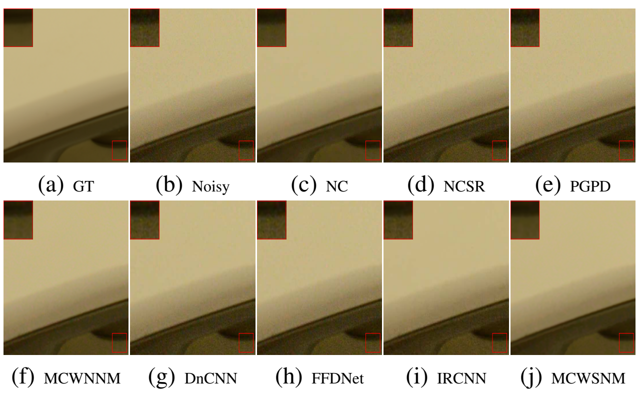


Figure 6: Denoised images of cropped "kodim3" in the Kodak PhotoCD Dataset with $\sigma_r = 40$, $\sigma_g = 20$ and $\sigma_b = 30$.



(a) #1 (b) #2 (c) #3 (d) #4 (e) #5
(f) #6 (g) #7 (h) #8 (i) #9 (j) #10

Figure 7: Ten noisy images in the CC Dataset.



(a) GT (b) Noisy (c) NC (d) NCSR (e) PGPD
(f) MCWNNM (g) DnCNN (h) FFDNet (i) IRCNN (j) MCWSNM

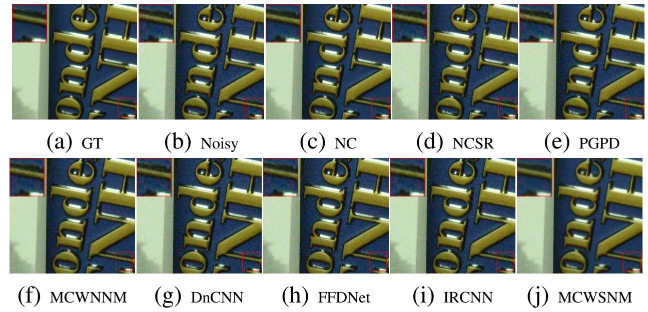
Figure 8: Denoised images of cropped #7 in real Dataset CC.

in Kodak PhotoCD Dataset, respectively. Visual inspection reveals that MCWSNM can obtain the best visual result and clear texture information. Color artifacts are generated by NCSR and PGPD, while NC over-smooths the image. While MCWNNM, DnCNN, FFDNet and IRCNN can get a clean image, some of the detailed information is lost compared with the original image.

3.2 Real Noisy Color Image Experiments

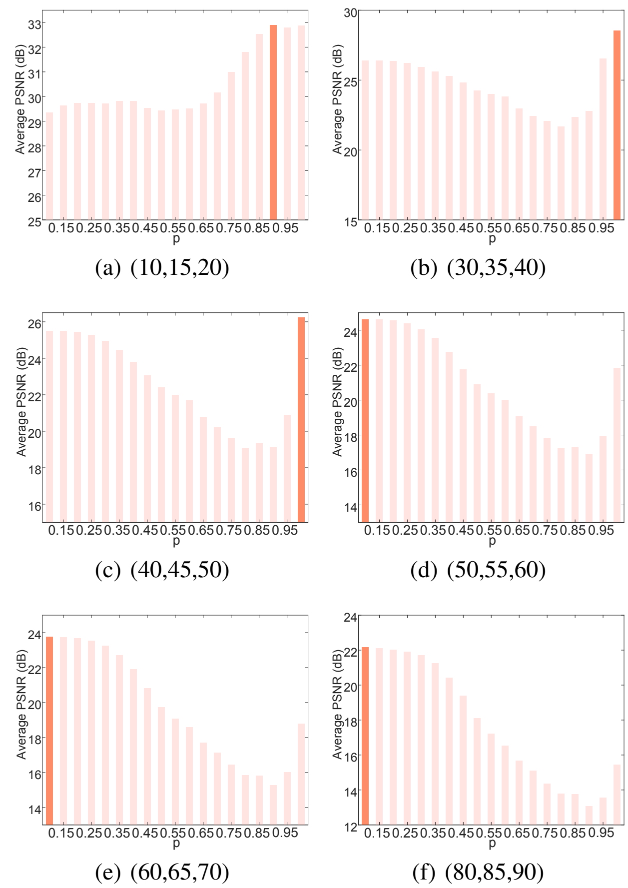
We implement the WCWSNM on a real noisy color image dataset, CC [Nam *et al.*, 2016], to evaluate the method's performance. This dataset includes 11 static scenes. The noisy images were captured in an indoor environment with different cameras and camera settings. For each scene, 500 images were taken using the same camera and camera settings. The mean image of each scene was then calculated to generate the noisy-free images, which were roughly regarded as the ground truth (GT). As the size of the original image is very large, 60 cropped smaller images with size 512×512 were provided in [Nam *et al.*, 2016]. In this paper, we select 10 cropped images (see Fig.7) to conduct the experiment.

For MCWSNM, we set the local search window size for each patch as 20, the similar patch number $M = 60$, each patch size $s = 5$, $K_1 = 10$, $K = 4$, $c = 2\sqrt{2}$, $\lambda = 0.6$, $p = 0.999$ and $\rho = 3$. Because the GT are provided, a quantitative assessment of each method is accessible. PSNR results for different cameras and camera settings are provided in Table 1. It can be seen from the table that the proposed MCWSNM obtains the highest PSNR value in most noisy images. Fig. 8 and Fig. 9 show the visual results of #7 and #10 images in CC dataset, respectively. Compared with other methods, MCWSNM gets the best visual effect. It not only remove the noise completely, but also preserve the detailed information effectively.



(a) GT (b) Noisy (c) NC (d) NCSR (e) PGPD
(f) MCWNNM (g) DnCNN (h) FFDNet (i) IRCNN (j) MCWSNM

Figure 9: Denoised images of cropped image #10 in the real Dataset CC.



(a) (10,15,20) (b) (30,35,40)
(c) (40,45,50) (d) (50,55,60)
(e) (60,65,70) (f) (80,85,90)

Figure 10: The influence of changing p on denoised results under different noise levels $(\sigma_r, \sigma_g, \sigma_b)$ on 24 images in the Kodak PhotoCD Dataset.

Camera Settings	#	NC	NCSR	PGPD	MCWNNM	DnCNN	FFDNET	IRCNN	MCWSNM
Canon 5D, ISO=3200	1	38.5689	37.8693	36.8979	40.8160	40.8445	40.6137	41.3212	**42.1242**
	2	35.0985	34.7144	33.8614	**36.5158**	35.7199	35.6380	36.0826	35.8209
Nikon D600, ISO=3200	3	36.8310	36.4562	35.8916	38.9327	38.9287	38.7748	**39.3098**	37.9467
	4	38.6322	36.3762	35.9296	40.1002	39.3217	38.7881	40.4201	**40.4434**
Nikon D800, ISO=1600	5	38.2259	36.8040	36.2972	39.4494	39.8094	39.5719	39.9118	**40.0471**
	6	39.0023	37.3847	36.4443	42.3815	40.6930	40.2738	**42.8365**	41.9094
Nikon D800, ISO=3200	7	36.5622	34.2501	33.9318	39.5804	36.7582	36.2557	38.9257	**41.5596**
	8	35.9500	33.8988	33.5135	36.7548	36.0972	35.8297	38.2784	**37.8922**
Nikon D600, ISO=6400	9	33.2295	31.1022	30.7797	36.5241	32.1810	31.9962	33.4272	**38.2520**
	10	32.2609	30.6858	30.4058	32.2105	31.6635	31.4748	32.7171	**32.4324**

Table 1: PSNR results (dB) of real color image CC Dataset.

3.3 Analysis of Power p

It is necessary to analyze the most appropriate setting of power p for different noise levels $(\sigma_r, \sigma_g, \sigma_b)$. We utilize 24 images in Kodak PhotoCD Dataset to test the proposed MCWSNM with different p under different noise levels added to the R, G and B channels. The results are presented in Fig.10. In each subfigure, the vertical coordinate represents the average PSNR value under a certain noise level, while the horizontal coordinate denotes the values of p changing from 0.1 to 1 with interval 0.05. We use six levels in this test: $\sigma_c = \{(10,15,20),(30,35,40),\ldots,(60,65,70),(80,85,90)\}$. It is clear from the histogram that at low and medium noise levels, the best value for p is 1, while at a high noise level, the best value of p is 0.1. This is mainly because, in the strong noise case (which means more rank components of the data are contaminated) highly ranked parts should penalized heavily, while lower-ranked parts should be penalized less.

4 Conclusion

In this paper, based on the low-rank property of the non-local self-similarity, we propose a MCWSNM method for color image denoising. For noisy color images, which generally hold different noise strength in each band, a weight matrix assigned to the noise level of each channel is introduced in order balance each channel's the contribution to the final estimation result. MCWSNM can be efficiently solved via ADMM optimization framework. Theorem 1 theoretically analyzes the convergence property of our proposed algorithm. Experiments on synthetic and real datasets demonstrate that the proposed method can obtain satisfactory results on the color image denoising task.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 61822113, 61976161, 62041105, the Science and Technology Major Project of Hubei Province (Next-Generation AI Technologies) under Grant 2019AEA170, the Natural Science Foundation of Hubei Province under Grants 2018CFA050, the Fundamental Research Funds for the Central Universities under Grant 413000092 and 413000082.

References

[Buades *et al.*, 2005] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A non-local algorithm for image denoising. In *CVPR*, pages 60–65, 2005.

[Chatterjee and Milanfar, 2010] Priyam Chatterjee and Peyman Milanfar. Is denoising dead? *IEEE Transactions on Image Processing*, 19(4):895–911, 2010.

[Chen *et al.*, 2015] Guangyong Chen, Fengyuan Zhu, and Pheng-Ann Heng. An efficient statistical method for image noise level estimation. In *ICCV*, pages 447–485, 2015.

[Dabov *et al.*, 2007a] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Color image denoising via sparse 3D collaborative filter with grouping constraint in Luminance-Chrominance space. In *ICIP*, pages 313–316, 2007.

[Dabov *et al.*, 2007b] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen O. Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.

[Dong *et al.*, 2013a] Weisheng Dong, Guangming Shi, and Xin Li. Nonlocal image restoration with bilateral variance estimation: A low-rank approach. *IEEE Transactions on Image Processing*, 22(2):700–711, 2013.

[Dong *et al.*, 2013b] Weisheng Dong, Lei Zhang, Guangming Shi, and Xin Li. Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing*, 22(4):1620–1630, 2013.

[Fang *et al.*, 2020a] Yixiang Fang, Xin Huang, Lu Qin, Ying Zhang, Wenjie Zhang, Reynold Cheng, and Xuemin Lin. A survey of community search over big graphs. *The VLDB Journal*, 29:353–392, 2020.

[Fang *et al.*, 2020b] Yixiang Fang, Yixing Yang, Wenjie Zhang, Xuemin Lin, and Xin Gao. Effective and efficient community search over large heterogeneous information networks. *Proceedings of the VLDB Endowment*, 13(6):854–867, 2020.

[Gu *et al.*, 2017] Shuhang Gu, Qi Xie, Deyu Meng, Wangmeng Zuo, Xiangchu Feng, and Lei Zhang. Weighted nuclear norm minimization and its applications to low level vision. *International Journal of Computer Vision*, 121(2):183–208, 2017.

[Hu *et al.*, 2019] Zhentao Hu, Zhiqiang Huang, Xinjian Huang, Fulin Luo, and Renzhen Ye. An adaptive nonlocal gaussian prior for hyperspectral image denoising. *IEEE Geoscience and Remote Sensing Letters*, 16(9):1487–1491, 2019.

[Lebrun *et al.*, 2015] Marc Lebrun, Miguel Colom, and Jean-Michel Morel. The noise clinic: A blind image denoising algorithm. *IPOL Journal*, 5:1–54, 2015.

[Liu *et al.*, 2017] Weiwei Liu, Ivor W. Tsang, and Klaus-Robert Müller. An easy-to-hard learning paradigm for multiple classes and multiple labels. *Journal of Machine Learning Research*, 18(94):1–38, 2017.

[Liu *et al.*, 2018] Ding Liu, BihanWen, Xianming Liu, Zhangyang Wang, and Thomas S. Huang. When image denoising meets high-level vision tasks: A deep learning approach. In *IJCAI*, pages 842–848, 2018.

[Liu *et al.*, 2019] Weiwei Liu, Donna Xu, Ivor W. Tsang, and Wenjie Zhang. Metric learning for multi-output tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):408–422, 2019.

[Mairal *et al.*, 2008] Julien Mairal, Michael Elad, and Guillermo Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, 2008.

[Mairal *et al.*, 2012] Julien Mairal, Francis R. Bach, and Jean Ponce. Task-driven dictionary learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):791–804, 2012.

[Marsousi *et al.*, 2014] Mahdi Marsousi, Kaveh Abhari, Paul Babyn, and Javad Alirezaie. An adaptive approach to learn overcomplete dictionaries with efficient numbers of elements. *IEEE Transactions on Signal Processing*, 62(12):3272–3283, 2014.

[Nam *et al.*, 2016] Seonghyeon Nam, Youngbae Hwang, Yasuyuki Matsushita, and Seon Joo Kim. A holistic approach to cross-channel image noise modeling and its application to image denoising. In *CVPR*, pages 1683–1691, 2016.

[Roth and Black, 2005] Stefan Roth and Michael J. Black. Fields of experts: A framework for learning image priors. In *CVPR*, volume 2, pages 860–867, 2005.

[Starck *et al.*, 2002] Jean-Luc Starck, Emmanuel J. Candès, and David L. Donoho. The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 11(6):670–684, 2002.

[Wang *et al.*, 2018a] Zengmao Wang, Yuhong Guo, and Bo Du. Matrix completion with preference ranking for top-N recommendation. In *IJCAI*, pages 3585–3591, 2018.

[Wang *et al.*, 2018b] Zheng Wang, Xiang Bai, Mang Ye, and Shin'ichi Satoh. Incremental deep hidden attribute learning. In *ACM MM*, pages 72–80, 2018.

[Xie *et al.*, 2016] Yuan Xie, Shuhang Gu, Yan Liu, Wangmeng Zuo, Wensheng Zhang, and Lei Zhang. Weighted Schatten p-norm minimization for image denoising and background subtraction. *IEEE Transactions on Image Processing*, 25(10):4842–4857, 2016.

[Xu *et al.*, 2015] Jun Xu, Lei Zhang, Wangmeng Zuo, David Zhang, and Xiangchu Feng. Patch group based nonlocal self-similarity prior learning for image denoising. In *ICCV*, pages 244–252, 2015.

[Xu *et al.*, 2017a] Chen Xu, Zhouchen Lin, and Hongbin Zha. A unified convex surrogate for the Schatten-p norm. In *AAAI*, pages 926–932, 2017.

[Xu *et al.*, 2017b] Jun Xu, Lei Zhang, David Zhang, and Xiangchu Feng. Multi-channel weighted nuclear norm minimization for real color image denoising. In *ICCV*, pages 1096–1104, 2017.

[Ye *et al.*, 2018] Mang Ye, Zheng Wang, Xiangyuan Lan, and Pong C. Yuen. Visible thermal person re-identification via dual-constrained top-ranking. In *IJCAI*, pages 1092–1099, 2018.

[Zhang and Aeron, 2016] Zemin Zhang and Shuchin Aeron. Denoising and completion of 3D data via multidimensional dictionary learning. In *IJCAI*, pages 2371–2377, 2016.

[Zhang *et al.*, 2013] Min Zhang, Zheng-Hai Huang, and Ying Zhang. Restricted p-isometry properties of nonconvex matrix recovery. *IEEE Transactions on Information Theory*, 59(7):4316–4323, 2013.

[Zhang *et al.*, 2017a] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.

[Zhang *et al.*, 2017b] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep CNN denoiser prior for image restoration. In *CVPR*, pages 2808–2817, 2017.

[Zhang *et al.*, 2018] Kai Zhang, Wangmeng Zuo, and Lei Zhang. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018.

[Zhu *et al.*, 2016] Fengyuan Zhu, Guangyong Chen, and Pheng-Ann Heng. From noise modeling to blind image denoising. In *CVPR*, pages 420–429, 2016.