Multi-channel weighted nuclear norm minimization for color image denoising

Anonymous ICCV submission

Paper ID 572

Abstract

Most of existing image denoising methods focus on grayscale images which contain a single channel, while little attention has been paid to color images containing multiple channels. In order to apply the existing denoising methods to color images, commonly used strategies are: 1) to process separately the patches in R, G, B channels; or 2) to process jointly the patch vectors concatenated by corresponding patches in the R, G, B channels. However, both processing strategies would generate false colors or artifacts of the color images, which will definately lose a certain of information in the color images. In this paper, we propose a new multi-channel model for color image denoising. We also concatenate the patches in the same positions of R, G ,B channels to form joint patch vectors. In the data term, we employ the weighted Frobinous norm, in which we add a weighting matrix to adaptively deal with different channels with different weights. For each concatenated patch, we search its non-local similar patches to it and form the patch matrix. To fully exploit the low rank property of the non-local similar patches, we employ the well-defined weighted nuclear norm as the regularization term. The resulting multi-channel weighted nuclear norm minimization model is much more complex than the original weighted nuclear norm minimization (WNNM) model. We transform the proposed model into a linear constrained non-convex optimization problem, which can be solved by the alternating direction method of multipliers (ADMM) algorithm. For each alternative updating step, the problem has closed-form solution. We also proof that the proposed algorithm has convergence results. The proposed model has been evaluated on color image denosing problem. Experiments on benchmark datasets demonstrate that the proposed model outperforms state-of-the-art denoising methods on synthetic as well as real-world noisy images.

1. Introduction

Image denoising is an important problem in enhancing the image quality in computer vision systems. It aims to recover the clean image $\mathbf{x}$ from the noisy observation $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where $\mathbf{n}$ is often assumed to be additive white Gaussian noise (AWGN). Though most image denoising methods [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12] have achieved excellant denoising performance on grayscale images, little attention has been paid to the color counterparts. Directly extending these methods to color images by common strategies, e.g., processing each channel of color images seperately, would generate false colors or artifacts [13].

There are several work [13, 14, 15, 16, 17, 18] proposed specifically for color image denoising. These methods either treat equally the R, G, B channels in color images [13, 14, 15] or ignore the non-local self similarity property of natural images [16, 17, 18]. On one hand, since the noise structures and levels among the R, G, B channels in color images are quite different due to the processing steps, such as demosaicing, white balance, etc., of in-camera imaging pipelines [19], the methods treating each channel equally would generate false colors or artifacts [13]. In order to deal with the R, G, B channels in color images more effectively, different noise properties of these channels should be considered in denoising color images, especially real noisy images. On the other hand, the performance of existing denoising methods would be largely depressed if the non-local self similarity property of natural images is ignored [2, 9]. This can be validated by the success of the state-of-the-art denoising methods [2, 6, 8, 9] which all exployed the non-local self similarity (NSS) property of natural images in different ways.

Recently, the weighted nuclear norm minimization method (WNNM) [9] is among the most successful methods on image denoising task which employ the NSS property o f natural images. In this paper, we proposed a noval multi-channel WNNM model for color image denoising by intorducing a weighting matrix to the WNNM method. However, if we add a weighting matrix $\mathbf{W}$ to the WNNM model, the resulting multi-channel WNNM model no longer has closed-form solutions as WNNM. This makes the proposed model more challenging to solve. To solve this problem, we reformulate the proposed multi-channel WNNM problem into a linear constrained non-convex program with

two variables. And the relaxed problem can be solved by employing the alternating direction method of multipliers (ADMM) [20]. Each variable can be updated with closed-form solution [9, 21]. The convergency resutls with detailed proof are given to guarantee a rational termination of the proposed model.

2. Related Work

2.1. Nuclear Norm Minimization

As the tightest convex surrogate function of the matrix rank minimization [22, 23], the nuclear norm minimization (NNM) problem has been extensive studied in low rank matrix approximation (LRMA) [24, 25, 26, 27]. A standard nuclear norm minimization problem is as follows:

$$\hat{\mathbf{X}} = \arg\min_{X} \|\mathbf{Y} - \mathbf{X}\|_F^2 + \lambda\|\mathbf{X}\|_*. \tag{1}$$

This NNM problem has closed-form solution by soft-thresholding the singular values of the matrix $\mathbf{Y}$ as

$$\hat{\mathbf{X}} = \mathbf{U}\mathcal{S}_{\frac{\lambda}{2}}(\mathbf{\Sigma})\mathbf{V}^\top \tag{2}$$

where $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ is the singular value decomposition (a.k.a. Eckart-Young Decomposition [28]) of $\mathbf{Y}$ and $\mathcal{S}_\tau(\bullet)$ is the soft-thresholding function with parameter $\tau > 0$:

$$\mathcal{S}_\tau(\mathbf{\Sigma}_{ii}) = \max(\mathbf{\Sigma}_{ii} - \tau, 0) \tag{3}$$

One limitation of the original NNM model is that it treats all the singular values equally but ignore the different importance of them. To make the NNM mmodel more flexible at processing sigular values, it has been extended to the truncated nuclear norm minimization model [29], the partial sum minimization of singular values [30], and the weighted nuclear norm minimization (WNNM) model [9], etc. Among these models, the WNNM model has been applied on grayscale image denoising problem with highly effective performance. This model adds weights to each singular values and the problem is:

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{X}\|_F^2 + \|\mathbf{X}\|_{\boldsymbol{w},*} \tag{4}$$

is firstly proposed for grayscale image denoising problem, where $\|\mathbf{X}\|_{\boldsymbol{w},*} = \sum_i w_i\sigma_i(\mathbf{X})$ is the weighted nuclear norm of matrix $\mathbf{X}$ and $\boldsymbol{w} = [w_1, ..., w_n]^\top, w_i \geq 0$ is the weight vector. According to the Remark 1 of [31], the problem (4) has closed-form solution if the weights are in a non-decreasing order

Though having achieved excellent performance on grayscale image denoising task, the WNNM method could not be applied on color image denoising in a direct manner. Of course we can apply the WNNM on each channel seperately, but it has been studied that this manner would get inferior performance when compared to the power of this model on grayscale images. In this paper, we would add a weighting matrix to the WNNM model and naturally extend it to deal with color images and maintain its powerful ability on exploring the non-local self similarity property of the natural images. The proposed multi-channel WNNM model can be solved via the famous ADMM [20] algorithm. For each variable, we can derive a closed-form solution and hence the overall model can be solved in an efficient way. Besides, we analysis the convergence results of the proposed model.

2.2. Color Image Denoising

Numerous methods [1, 2, 4, 5, 6, 8, 9, 10, 11, 12] has been proposed for grayscale image denoising task. These methods could be trivially employed to denoise each channel in color images separately. For example, the CBM3D [14] first transform the RGB image into luminance-chrominance color space (e.g., YCbCr) and perform BM3D [2] for each channels separately with the patches only being grouped in luminance channel. However, as pointed out in [13], the performance would be largely reduced when compared to the performance of these methods on grayscale images. This method [13] performs color image denoising by concatenating the patches in R, G, B channels into a single vector. However, the concatenation would generate false colors and artifacts [13]. The major reason is that the concatenation methods treat multiple channels equally and ignore the different properties among these channels. To better explore the difference and correlation among the three channels in color images, several methods [15, 16, 17, 18, 32, 33, 34] have been proposed to deal with real noisy image denoising task. These methods often firstly estimate the parameters of the assumed noise model (usually Gaussian), and then perform denoising with the estimated noise model. In this paper, we propose a weighting matrix (diagonal) which add differents noise levels on different channels for color image denoising. The proposed multi-channel method are able to deal with the distinct noise property among these channels. Experiments will demonstrate that the proposed model achieves better results than other competing methods for color image denoising.

3. Multi-channel Weighted Nuclear Norm Minimization

3.1. The Problem

Given In order to treat differently each channel in color images, we propose to add a weighting matrix to the original WNNM model and the resulting problem becomes

$$\min_{\mathbf{X}} \|\mathbf{W}(\mathbf{Y} - \mathbf{X})\|_F^2 + \|\mathbf{X}\|_{\boldsymbol{w},*}. \tag{5}$$

where $\mathbf{W}$ is the weighting matrix. For simplisity, we assume $\mathbf{W}$ to be a diagonal matrix.

Unfortunately, the proposed multi-channel WNNM problem cannot be solved in an analytical form. In [31], when the weights on singular values are non-descending, the weighted nuclear norm proximal operator can have global optimum with closed-form solution. However, such property is not valid for the multi-channel WNNM model. The reason is that the weighting matrix $\mathbf{W}$ is added to the matrix $\mathbf{X}$ instead of its singular values. Besides, the elements in $\mathbf{W}$ is not in a non-descending order with respect to the singular value of $\mathbf{X}$. This makes the proposed model more difficult to optimize than the original WNNM model.

This can be solved by introducing an augmented variable $\mathbf{Z}$, and the above multi-channel WNNM problem is equivalent to a linearly constrained non-convex problem with two variables.

$$\min_{\mathbf{X},\mathbf{Z}} \|\mathbf{W}(\mathbf{Y}-\mathbf{X})\|_F^2 + \|\mathbf{Z}\|_{\boldsymbol{w},*} \quad \text{s.t.} \quad \mathbf{X}=\mathbf{Z}. \tag{6}$$

This is an optimization problem with functions of two variables $\mathbf{X}$ and $\mathbf{Z}$ with linearly constrained condition of $\mathbf{X}=\mathbf{Z}$. In fact, this proplem can be solved by the alternating direction method of multipliers (ADMM) algorithm, which will be introduced in the next subsection.

3.2. Optimization

To solve the above optimization problem, we first derive its augmented Lagrangian function as

$$\begin{aligned}\mathcal{L}(\mathbf{X},\mathbf{Z},\mathbf{A},\rho) =& \|\mathbf{W}(\mathbf{Y}-\mathbf{X})\|_F^2 + \|\mathbf{Z}\|_{\boldsymbol{w},*} \\ &+ \langle \mathbf{A}, \mathbf{X}-\mathbf{Z}\rangle + \frac{\rho}{2}\|\mathbf{X}-\mathbf{Z}\|_F^2\end{aligned} \tag{7}$$

where $\mathbf{A}$ is the augmented Lagrangian multiplier and $\rho>0$ is the penalty parameter. After some simple calculations, we can obtain the following equivalent form of the Lagrangian function

$$\begin{aligned}\mathcal{L}(\mathbf{X},\mathbf{Z},\mathbf{A},\rho) =& \|\mathbf{W}(\mathbf{Y}-\mathbf{X})\|_F^2 + \|\mathbf{Z}\|_{\boldsymbol{w},*} \\ &+ \frac{\rho}{2}\|\mathbf{X}-\mathbf{Z}+\rho^{-1}\mathbf{A}\|_F^2\end{aligned} \tag{8}$$

We initialize the matrix variables $\mathbf{X}_0$, $\mathbf{Z}_0$, and $\mathbf{A}_0$ to be zero matrix of suitable size. Taking derivative of the Lagrangian function $\mathcal{L}$ with respect to the variables $\mathbf{X}$ and $\mathbf{Z}$ and setting the derivative function to be zero, we can alternatively update the iterations of the ADMM algorithm as follows:

(1) Update $\mathbf{X}$ while fixing $\mathbf{Z}$ and $\mathbf{A}$:

$$\mathbf{X}_{k+1} = \arg\min_{\mathbf{X}} \|\mathbf{W}\mathbf{Y}-\mathbf{W}\mathbf{X}\|_F^2 + \frac{\rho_k}{2}\|\mathbf{X}-\mathbf{Z}_k+\rho_k^{-1}\mathbf{A}_k\|_F^2 \tag{9}$$

This is a mixed weighted least square and standard least square problem and we could derive its closed-form solution:

$$\mathbf{X}_{k+1} = (\mathbf{W}^\top\mathbf{W}+\frac{\rho_k}{2}\mathbf{I})^{-1}(\mathbf{W}^\top\mathbf{W}\mathbf{Y}+\frac{\rho_k}{2}\mathbf{Z}_k-\frac{1}{2}\mathbf{A}_k) \tag{10}$$

(2) Update $\mathbf{Z}$ while fixing $\mathbf{X}$ and $\mathbf{A}$:

$$\mathbf{Z}_{k+1} = \arg\min_{\mathbf{Z}} \frac{\rho_k}{2}\|\mathbf{Z}-(\mathbf{X}_{k+1}+\rho_k^{-1}\mathbf{A}_k)\|_F^2 + \|\mathbf{Z}\|_{\boldsymbol{w},*} \tag{11}$$

According to the Theorem 1 in [31], given the $\mathbf{X}_{k+1}+\rho_k^{-1}\mathbf{A}_k = \mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^\top$ be the SVD of $\mathbf{X}_{k+1}+\rho_k^{-1}\mathbf{A}_k$, where $\mathbf{\Sigma}_k = \begin{pmatrix}\text{diag}(\sigma_1,\sigma_2,...,\sigma_n)\\ \mathbf{0}\end{pmatrix} \in \mathbb{R}^{m\times n}$, then the global optimum of the above problem is $\hat{\mathbf{Z}} = \mathbf{U}_k\hat{\mathbf{\Sigma}}_k\mathbf{V}_k^\top$, where $\hat{\mathbf{\Sigma}}_k = \begin{pmatrix}\text{diag}(\hat{\sigma}_1,\hat{\sigma}_2,...,\hat{\sigma}_n)\\ \mathbf{0}\end{pmatrix} \in \mathbb{R}^{m\times n}$ and $(\hat{\sigma}_1,\hat{\sigma}_2,...,\hat{\sigma}_n)$ is the solution to the following convex optimization problem:

$$\begin{aligned}&\min_{\hat{\sigma}_1,\hat{\sigma}_2,...,\hat{\sigma}_n} \sum_{i=1}^{n}(\sigma_i-\hat{\sigma}_i)^2 + \frac{2w_i}{\rho_k}\hat{\sigma}_i \\ &\text{s.t.} \quad \hat{\sigma}_1 \geq \hat{\sigma}_2 \geq ... \geq \hat{\sigma}_n \geq 0.\end{aligned} \tag{12}$$

According to the Remark 1 in [31], the problem above has closed-form solution

$$\hat{\sigma}_i = \begin{cases} 0 & \text{if } c_2<0 \\ \frac{c_1+\sqrt{c_2}}{2} & \text{if } c_2 \geq 0 \end{cases} \tag{13}$$

where $c_1=\sigma_i-\epsilon$, $c_2=(\sigma_i-\epsilon)^2-\frac{8C}{\rho_k}$ and C is set as $\sqrt{2n}$ by experience in image denoising.

(3) Update $\mathbf{A}$ while fixing $\mathbf{X}$ and $\mathbf{Z}$:

$$\mathbf{A}_{k+1} = \mathbf{A}_k + \rho_k(\mathbf{X}_{k+1}-\mathbf{Z}_{k+1}) \tag{14}$$

(4) Update ρ_k as $\rho_{k+1}=\mu*\rho_k$, where $\mu>1$ is a .

The above 4 alternative updating steps are repeated until the convergence conditions are satisfied or the number of iterations exceeds a preset maximum number, e.g., K_1. The overall algorithm will achieve its convergence conditions when $\|\mathbf{X}_{k+1}-\mathbf{Z}_{k+1}\|_F \leq \text{Tol}$, $\|\mathbf{X}_{k+1}-\mathbf{X}_k\|_F \leq \text{Tol}$, and $\|\mathbf{Z}_{k+1}-\mathbf{Z}_k\|_F \leq \text{Tol}$ are simutaniously satisfied, where $\text{Tol}>0$ is a small tolerance. We summerize the optimization steps in Algorithm 1 (A1). We give a theorem, i.e., Theorem 1, to guarantee the convergence of the proposed Algorithm 1. Note that since the weighted nuclear norm is non-convex in general, we employ an unbounded sequence of $\{\rho_k\}$ here to make sure that the Algorithm 1 is convergent.

Theorem 1. *Assume the weights in $\boldsymbol{w}$ are in a non-descending order, the sequence $\{\mathbf{X}_k\}$, $\{\mathbf{Z}_k\}$, and $\{\mathbf{A}_k\}$ generated in Algorithm 1 satisfy:*

$$(1)\ \lim_{k\to\infty}\|\mathbf{X}_{k+1}-\mathbf{Z}_{k+1}\|_F = 0; \tag{15}$$

$$(2)\ \lim_{k\to\infty}\|\mathbf{X}_{k+1}-\mathbf{X}_k\|_F = 0; \tag{16}$$

$$(3)\ \lim_{k\to\infty}\|\mathbf{Z}_{k+1}-\mathbf{Z}_k\|_F = 0. \tag{17}$$

A1: Solve Multi-channel WNNM via ADMM

Input: Matrices $\mathbf{Y}$ and $\mathbf{W}$, $\mu > 1$, Tol > 0, $K_1 > 0$;
Initialization: $\mathbf{X}_0 = \mathbf{Z}_0 = \mathbf{A}_0 = \mathbf{0}$, $\rho_0 > 0$, T = False, $k = 0$;
While (T == false) **do**
1. Update $\mathbf{X}_{k+1}$ as
$\mathbf{X}_{k+1} = (\mathbf{W}^\top\mathbf{W} + \frac{\rho_k}{2}\mathbf{I})^{-1}(\mathbf{W}^\top\mathbf{W}\mathbf{Y} + \frac{\rho_k}{2}\mathbf{Z}_k - \frac{1}{2}\mathbf{A}_k)$
2. Update $\mathbf{Z}_{k+1}$ by solving the WNNM problem
$\min_{\mathbf{Z}} \frac{\rho_k}{2}\|\mathbf{Z} - (\mathbf{X}_{k+1} + \rho_k^{-1}\mathbf{A}_k)\|_F^2 + \|\mathbf{Z}\|_{\boldsymbol{w},*}$
3. Update $\mathbf{A}_{k+1}$ as $\mathbf{A}_{k+1} = \mathbf{A}_k + \rho_k(\mathbf{X}_{k+1} - \mathbf{Z}_{k+1})$
4. Update $\rho_{k+1} = \mu * \rho_k$;
5. $k \leftarrow k + 1$;
if ($\|\mathbf{X}_{k+1} - \mathbf{Z}_{k+1}\|_F / \|\mathbf{Z}_{k+1}\|_F <$ Tol) or ($k \geq K_1$)
5. T $\leftarrow$ True
end if
end while
Output: Matrices $\mathbf{X}$ and $\mathbf{Z}$.

Proof. We give proof sketch here and detailed proof of this theorem can be found in Appendix. We can first proof that the sequence $\{\mathbf{A}_k\}$ generated by Algorithm 1 is upper bounded. Since $\{\rho_k\}$ is unbounded, that is $\lim_{k\to\infty}\rho_k = +\infty$, we can proof that the sequence of Lagrangian function $\{\mathcal{L}(\mathbf{X}_{k+1}, \mathbf{Z}_{k+1}, \mathbf{A}_k, \rho_k)\}$ is also upper bounded. Hence, both $\{\mathbf{WY} - \mathbf{WX}_k\}$ and $\{\mathbf{Z}_k\}$ are upper bounded. According to Eq. (14), we can proof that $\lim_{k\to\infty}\|\mathbf{X}_{k+1} - \mathbf{Z}_{k+1}\|_F = \lim_{k\to\infty}\rho_k^{-1}\|\mathbf{A}_{k+1} - \mathbf{A}_k\|_F = 0$, and (1) is proofed. Then we can proof that $\lim_{k\to\infty}\|\mathbf{X}_{k+1} - \mathbf{X}_k\|_F \leq \lim_{k\to\infty}\|(\mathbf{W}^\top\mathbf{W} + \frac{\rho_k}{2}\mathbf{I})^{-1}(\mathbf{W}^\top\mathbf{W}\mathbf{Y} - \mathbf{W}^\top\mathbf{W}\mathbf{Z}_k - \frac{1}{2}\mathbf{A}_k)\|_F + \rho_k^{-1}\|\mathbf{A}_k - \mathbf{A}_{k-1}\|_F = 0$ and hence (2) is proofed. Then (3) can be proofed by checking that $\lim_{k\to\infty}\|\mathbf{Z}_{k+1} - \mathbf{Z}_k\| \leq \lim_{k\to\infty}\|\boldsymbol{\Sigma}_{k-1} - \mathcal{S}_{\boldsymbol{w}/\rho_{k-1}}(\boldsymbol{\Sigma}_{k-1})\|_F + \|\mathbf{X}_{k+1} - \mathbf{X}_k\|_F + \rho_k^{-1}\|\mathbf{A}_{k-1} + \mathbf{A}_{k+1} - \mathbf{A}_k\|_F = 0$, where $\mathbf{U}_{k-1}\boldsymbol{\Sigma}_{k-1}\mathbf{V}_{k-1}^\top$ is the SVD of the matrix $\mathbf{X}_k + \rho_{k-1}\mathbf{A}_{k-1}$. The proof skecth of Theorem 1 is end. □

4. Multi-channel WNNM for Color Image Denoising

In this section, we apply the proposed multi-channel WNNM model on color image denoising problem. The multi-channel WNNM model can make use of the non-local self similarity property of natural images while treating each channel adaptively. In real-world noisy images, the noise are first emerged in the RAW data, i.e., color filter array (CFA). The major noise generated in real noisy images are due to the discrete nature of light and thermal agitation [], which can be modeled as Poisson and Gaussian distribution, respectively. Since the Poisson distribution can be approximately modeled by Gaussian distribution, the overall noise model in each channel of the color image could be Gaussian. Some methods would first transform the color images into other color spaces such as luminance/chrominance space such as YCbCr or Lab, but this would change the noise distribution in each channel. Hence, in this work we still choose to deal with the RGB channels in color images. Besides, even though if the demosaicing of RAW image generate similar distribution in noise in different channels, the channel-wise scaling in white balance would definitely change the noise distribution in each channel. Thus, the noise in R, G, B channels are definitely different which can be described by different noise levels and structures. According to above analysis, color image denoising is to recover the latent clean image $\mathbf{x}$ from the observed noisy version $\mathbf{y}_c = \mathbf{x}_c + \mathbf{n}_c$, where $c \in \{R, G, B\}$ represent the R, G, B channels in color images and $\mathbf{n}_c$ is the noise in the c channel (assumed to be additive white Gaussian noise).

The patches in color image $\mathbf{y}$ are of size $p \times p \times 3$. For each patch $\mathbf{y}_j$, we search its non-local similar patches in a large area and stack the similar patches column by column. The resulting matrix $\mathbf{Y}_j \in \mathbb{R}^{3p^2\times n}$, where n is the number of similar patches. The corresponding matrices containing the clean patches and the channel-wise noise are defined as $\mathbf{X}_j$ and $\mathbf{N}_j$, respectively. Since $\mathbf{X}_j$ is made of similar patches, it should be a low rank matrix. And hence the multi-channel WNNM model proposed in this paper can be used here. Compared to the original single channel WNNM model [9] proposed for grayscale image denoising

$$\min_{\mathbf{X}_j}\|\mathbf{W}_j(\mathbf{Y}_j - \mathbf{X}_j)\|_F^2 + \|\mathbf{X}_j\|_{w,*}. \tag{18}$$

When the weighting matrix $\mathbf{W}_j = \frac{1}{\sigma_n^2}\mathbf{I}$, where $\mathbf{I} \in \mathbb{R}^{3p^2\times 3p^2}$ is the identity matrix, the multi-channel WNNM model will reduce to the WNNM model as a special case. The design of WNNM model also motivate us to consider a similar design of the weighting matrix. In order to deal with color image denoising task, the weighting matrix $\mathbf{W}_j$ should be modified to be suitable for multi-channel cases. In fact, for the color image denoising task, a holistic try of the weighting matrix $\mathbf{W}_j$ could be

$$\mathbf{W}_j = \begin{pmatrix} \sigma_{rj}^{-2}\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_{gj}^{-2}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_{bj}^{-2}\mathbf{I} \end{pmatrix}. \tag{19}$$

Here, for simplicity, we assume that the noise in different channels are independent to each other. The experimental results have already demonstrated that this simple assumption could already generate the best denoising performance on benchmark real noisy image dataset. In this paper, we did not consider the correlations of noise among different channels, which is the future work of our research line. The determination of the weight vector in weighted nuclear norm is the same as in the WNNM model [31]. We set the weight vector as $w_i^{k+1} = \frac{C}{|\sigma_i(\mathbf{X}_k)|+\epsilon}$.

A2: Color Image Denoising by Multi-channel WNNM

Input: Noisy image $\mathbf{y}$, noise levels $\{\sigma_r, \sigma_g, \sigma_b\}$;
Initialization: $\hat{\mathbf{x}}^{(0)} = \mathbf{y}$, $\mathbf{y}^{(0)} = \mathbf{y}$;
for $k = 1 : K_2$ **do**
1. Set $\mathbf{y}^{(k)} = \hat{\mathbf{x}}^{(k-1)}$;
2. Extracte local patches $\{\mathbf{y}_j\}_{j=1}^N$ from $\mathbf{y}^{(k)}$;
   for each patch $\mathbf{y}_j$ **do**
3. Search non-local similar patches $\mathbf{Y}_j$;
4. Estimate $\mathbf{X}_j$ by applying the Algorithm 1 to $\mathbf{Y}_j$;
   end for
5. Aggregate $\{\mathbf{X}_j\}_{j=1}^N$ to form the image $\hat{\mathbf{x}}^{(k)}$;
end for
Output: Denoised image $\hat{\mathbf{x}}^{K_2}$.

The multi-channel WNNM is applied to the non-local similar patches of each local patch in the noisy image $\mathbf{y}$. And then all the patches are aggregated together to form the final recovered image $\hat{\mathbf{y}}$. We also perform the denoising procedure for several (K_2) iterations to obtain better denoising results. In Algorithm 2 (A2), we summarizes the denoising steps of multi-channel WNNM model on color image denoising.

5. Experiments

We evaluate the proposed method on synthetic noisy images as well as real noisy images. The synthetic noisy images are generated by adding additive white Gaussian noise with known noise standard derivations $\sigma_r, \sigma_g, \sigma_b$ for the R, G, B channels, respectively. We compare the proposed method with other state-of-the-art denoising algorithms including CBM3D [2, 14], MLP [10], CSF [11], WNNM [9], TNRD [12], "Noise Clinic" [16, 33], and the commercial software Neat Image [34].

In order to take fully comparison with the original WNNM method, we extended the WNNM method [31] in three directions. The first is to apply the WNNM method on each channel separately and we still call this method "WNNM". The second is to concatenate the corresponding patches in the R, G, B channels into a joint patch and perform denoising in a joint manner. We call this method "WNNM1". Note that both "WNNM" and "WNNM1" have closed form solutions since they are directly extended from the original WNNM. The third is to set the weighting matrix $\mathbf{W}$ in the proposed multi-channel WNNM model as $\mathbf{W} = \sigma_n^2\mathbf{I}$. This is to more clearly validate the effectiveness of the weighting matrix by reducing the multi-channel WNNM model to its special case: the WNNM model solved by ADMM algorithm. We call this method "WNNM2". We set the same parameters settings for the "WNNM2" method and the proposed multi-channel WNNM method (called "Proposed"). For fair comparison, for "WNNM", the corresponding noise levels σ_c of the c ($c = r, g, b$) channel is input as known parameter; for "WNNM1", we input the noise level as $\sigma = \sqrt{(\sigma_r^2 + \sigma_g^2 + \sigma_b^2)/3}$ and tune the other parameters to achieve its best denoising performance (i.e., highest average PSNR results); for "WNNM2", we employ the same parameter settings as the proposed multi-channel WNNM method, which will be introduced in details in the following sections.

5.1. Experiments on Synthetic Noisy Images

In this section, we compare the proposed method with other competing method [9, 10, 11, 12, 14, 16, 34] on 24 high quality color images from the Kodak PhotoCD Dataset (http://r0k.us/graphics/kodak/), which are shown in Fig. 1. Then we add additive white Gaussian noise with different standard deviations to different channels of the color images. The standard deviations of noise we add to the R, G, B channels of the 24 clean images are 40, 20, 30, respectively. We set the patch size as $p = 6$, the number of non-local similar patches as $n = 70$, the window size for searching similar patches as $W = 20$. For the proposed multi-channel WNNM model, we set the regularization parameter as $\lambda = 4$, the penalty parameter as $\rho = 6$, the $\mu = 1.1$, the number of iterations in Algorithm 1 as $K_1 = 10$, the number of iterations in Algorithm 2 as $K_2 = 8$.

We perform quantitative comparison on the 24 high quality images from the Kodak PhotoCD Dataset, which are widely used for color image denoising task. The PSNR results of CBM3D [2], MLP [10], TNRD [12], NC [16, 33], NI [34], "WNNMCW" [31], "WNNMJ", "WNNMJadmm" and the proposed multi-channel WNNM methods are listed in Table 1. The best PSNR results of each image are highlighted in bold. One can see that on all the 24 images, our method achieves the best PSNR values. On average, our proposed method has 0.45dB PSNR improvements over the second best method, i.e., "WNNMJ" and much higher PSNR gains over other competing methods. Fig. 4 shows the denoised images of a scene in the Kodak PhotoCD Dataset . We can see that CBM3D, NC, and NI would either remain noise or generate artifacts, while MLP, TNRD "WNNMCW", "WNNMJ", "WNNMJadmm" over-smooth much the image. By using the proposed multi-channel WNNM model, our method preserves the structures (e.g., edges and textures) better across the R, G, B channels and generate less artifacts than other denoising methods, leading to visually pleasant outputs. More visual comparisons can be found in the supplementary file.

5.2. Experiments on Real Noisy Images

In the second part, we compare the proposed method with other competing methods on the 15 real noisy images, , which are shown in Fig. 2, with "ground truth" clean images [17]. The noisy images were collected under controlled in-

Table 1. PSNR(dB) results of different denoising algorithms on 20 natural images.

	$\sigma_r = 40, \sigma_g = 20, \sigma_b = 30$								
Image#	**CBM3D**	**MLP**	**TNRD**	**Noise Clinic**	**Neat Image**	**WNNM**	**WNNM1**	**WNNM2**	**Proposed**
1	25.24	25.70	25.74	24.90	23.85	26.01	25.95		26.66
2	28.27	30.12	30.21	25.87	25.90	30.08	30.11		30.20
3	28.81	31.19	31.49	28.58	26.00	31.58	31.61		32.25
4	27.95	29.88	29.86	25.67	25.82	30.13	30.16		30.49
5	25.03	26.00	26.18	25.15	24.38	26.44	26.39		26.82
6	26.24	26.84	26.90	24.74	24.65	27.39	27.30		27.98
7	27.88	30.28	30.40	27.69	25.63	30.47	30.54		30.98
8	25.05	25.59	25.83	25.30	24.02	26.71	26.75		26.90
9	28.44	30.75	30.81	27.44	25.94	30.86	30.92		31.49
10	28.27	30.38	30.57	28.42	25.87	30.65	30.68		31.26
11	26.95	28.00	28.14	24.67	25.32	28.19	28.16		28.63
12	28.76	30.87	31.05	28.37	26.01	30.97	31.06		31.48
13	23.76	23.95	23.99	22.76	23.53	24.27	24.15		24.89
14	26.02	26.97	27.11	25.68	24.94	27.20	27.15		27.57
15	28.38	30.15	30.44	28.21	26.06	30.52	30.60		30.81
16	27.75	28.82	28.87	26.66	25.69	29.27	29.21		29.96
17	27.90	29.57	29.80	28.32	25.85	29.78	29.79		30.40
18	25.77	26.40	26.41	25.70	24.74	26.63	26.56		27.22
19	27.30	28.67	28.81	26.52	25.40	29.19	29.22		29.57
20	28.96	30.40	30.76	25.90	24.95	30.79	30.83		31.07
21	26.54	27.53	27.60	26.48	25.06	27.80	27.75		28.34
22	27.05	28.17	28.27	26.60	25.36	28.21	28.16		28.64
23	29.14	32.31	32.51	23.24	26.13	31.89	31.97		32.34
24	25.75	26.41	26.53	25.73	24.55	27.10	27.03		27.59
Average	27.13	28.54	28.68	26.19	25.24	28.84	28.83		29.31

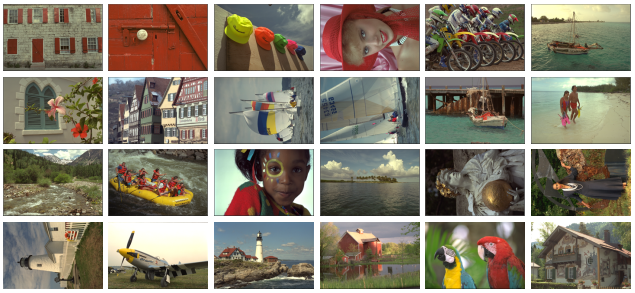


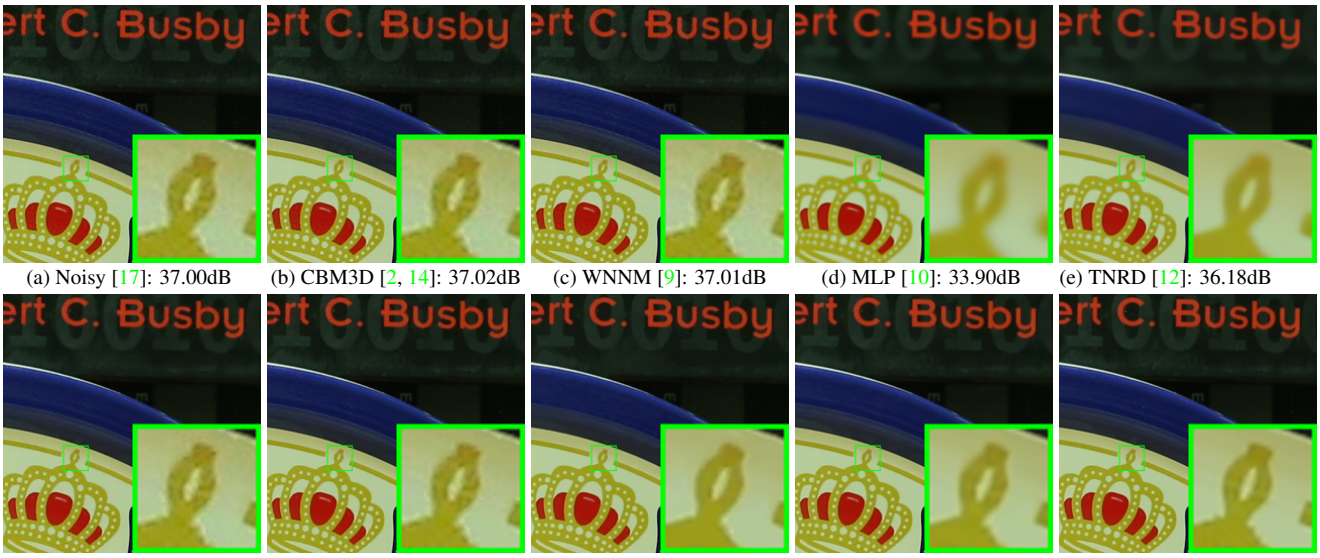
Figure 1. The 24 high quality color images from the Kodak PhotoCD Dataset.

door environment. Each scene was shot 500 times under the same camera and camera setting. The mean image of the 500 shots is roughly taken as the "ground truth", with which the PSNR can be computed. Since the image size is very large (about 7000×5000) and the scenes of this dataset share repetitive contents, the authors of [17] cropped 15 smaller images (of size 512×512) to perform experiments. In this section, we do not compare with the "WNNMCW" method due to its inferior performance.

We firstly perform quantitative comparison on the 15 cropped images used in [17]. The PSNR results of CBM3D [2], WNNM [9], MLP [10], TNRD [12], NC [16, 33], NI [34] and CC [17] are listed in Table 2 (The results of CC are copied from the original paper [17]). The best and second best PSNR results of each image are highlighted in red and blue, respectively. One can see that on 9 out of the 15 images, our method achieves the best PSNR values. CC achieves the best PSNR on 3 of the 15 images. It should be noted that in the CC method, a specific model is trained for each camera and camera setting, while our method uses the same model for all images. On average, our proposed method has 0.28dB PSNR improvements over [17] and much higher PSNR gains over other competing methods. Fig. 4 shows the denoised images of a scene captured by Canon 5D Mark 3 at ISO = 3200. We can see that CBM3D, WNNM, NC, NI and CC would either remain noise or generate artifacts, while MLP, TNRD over-smooth much the image. By using the proposed multi-channel WNNM model, our method preserves the structures (e.g., edges and textures) better across the R, G, B channels and generate less artifacts than other denoising methods, leading to visually pleasant outputs. More visual comparisons can be found in the supplementary file.

6. Conclusion and Future Work

Most existing color image denoising methods treat the R, G, B channels equally and ignore the different noise structures in different channels. Common strategies that processing each channel seperately and concatenating the RGB values into joint vectors would generate false color or artifacts. In this paper, we proposed a noval model for color image denoising which can explore the different noise structures among the R, G, B channels and exploit the nonlocal self similarity property of natural images. Specifically, we introduced a weighting matrix, which are employed to describe the noise levels of different channels, to the original weighted nuclear norm minimization (WNNM) model. Though the proposed model no longer has closed-form solution, we successfully solved the proposed model



(a) Noisy [17]: 37.00dB (b) CBM3D [2, 14]: 37.02dB (c) WNNM [9]: 37.01dB (d) MLP [10]: 33.90dB (e) TNRD [12]: 36.18dB

(f) NI [34]: 37.68dB (g) NC [16, 33]: 38.76dB (h) CC [17]: 38.37dB (i) Ours: **40.50**dB (j) Mean Image [17]

Figure 2. Denoised images of a region cropped from the real noisy image "Canon 5D Mark 3 ISO 3200 1" [17] by different methods. The images are better to be zoomed in on screen.

Table 2. PSNR(dB) results of different methods on 15 cropped real noisy images used in [17].

Camera Settings	**CBM3D**	**MLP**	**TNRD**	**NI**	**NC**	**CC**	**WNNM1**	**WNNM2**	**Proposed**
Canon 5D Mark III ISO = 3200	39.76	39.00	39.51	35.68	36.20	38.37	39.74	39.98	**41.13**
	36.40	36.34	36.47	34.03	34.35	35.37	35.12	36.65	**37.28**
	36.37	36.33	36.45	32.63	33.10	34.91	33.14	34.63	**36.52**
Nikon D600 ISO = 3200	34.18	34.70	34.79	31.78	32.28	34.98	35.08	35.08	**35.53**
	35.07	36.20	36.37	35.16	35.34	35.95	36.42	36.84	**37.02**
	37.13	39.33	39.49	39.98	40.51	**41.15**	40.78	39.24	39.56
Nikon D800 ISO = 1600	36.81	37.95	38.11	34.84	35.09	37.99	38.28	38.61	**39.26**
	37.76	40.23	40.52	38.42	38.65	40.36	41.24	40.81	**41.43**
	37.51	37.94	38.17	35.79	35.85	38.30	38.04	38.96	**39.55**
Nikon D800 ISO = 3200	35.05	37.55	37.69	38.36	38.56	**39.01**	39.93	37.97	38.91
	34.07	35.91	35.90	35.53	35.76	36.75	37.32	37.30	**37.41**
	34.42	38.15	38.21	40.05	40.59	39.06	**41.52**	38.68	39.39
Nikon D800 ISO = 6400	31.13	32.69	32.81	34.08	34.25	34.61	**35.20**	34.57	34.80
	31.22	32.33	32.33	32.13	32.38	33.21	33.61	33.43	**33.95**
	30.97	32.29	32.29	31.52	31.76	33.22	33.62	34.02	**33.94**
Average	35.19	36.46	36.61	35.33	35.65	36.88	37.27	37.12	**37.71**

via the famous ADMM algorithm by introducing an additional variable with a linear constraint. The transformed problem has convergence property and can be solved in an alternative updating manner and both variables can be updated with closed-form solutions. We applied the proposed multi-channel WNNM model on color image denoising problem. Extensive experiments on benchmark datasets demonstrate that the proposed model outperforms the other competing denoising methods on both synthetic color noisy images as well as real-world noisy images. The introduce of the weighting matrix can indeed boost the performance of the original WNNM model on color image denoising. We believe that this work can be extended in at least three directions. Firstly, the proposed weighting matrix can be introduced into other methods designed for denoising grayscale images. Secondly, the weighting matrix beyond the diagonal form, such as correlation form [35], may bring better performance on color image denoising. Thirdly, the proposed multi-channel WNNM model can be further extended to deal with images with more channels, such as the hyperspectral images in remote sensing applications. We will focus our futhre work on these three directions.

Figure 3. The 15 cropped real noisy images used in [17].

7. A. Proof of Theorem 1.

Proof. 1. Firstly, we proof that the sequence $\{\mathbf{A}_k\}$ generated by Algorithm 1 is upper bounded. Let $\mathbf{X}_{k+1} + \rho_k^{-1}\mathbf{A}_k = \mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^\top$ be its SVD in the $(k+1)$-th iteration. According to Corollary 1 of [31], we can have the SVD of $\mathbf{Z}_{k+1}$ as $\mathbf{Z}_{k+1} = \mathbf{U}_k\hat{\mathbf{\Sigma}}_k\mathbf{V}_k^\top = \mathbf{U}_k\mathcal{S}_{\frac{\boldsymbol{w}}{\rho_k}}(\mathbf{\Sigma}_k)\mathbf{V}_k^\top$. Then we have

$$\begin{aligned}
\|\mathbf{A}_{k+1}\|_F &= \|\mathbf{A}_k + \rho_k(\mathbf{X}_{k+1} - \mathbf{Z}_{k+1})\|_F && (20)\\
&= \rho_k\|\rho_k^{-1}\mathbf{A}_k + \mathbf{X}_{k+1} - \mathbf{Z}_{k+1}\|_F && (21)\\
&= \rho_k\|\mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^\top - \mathbf{U}_k\mathcal{S}_{\frac{\boldsymbol{w}}{\rho_k}}(\mathbf{\Sigma}_k)\mathbf{V}_k^\top\|_F && (22)\\
&= \rho_k\|\mathbf{\Sigma}_k - \mathcal{S}_{\frac{\boldsymbol{w}}{\rho_k}}(\mathbf{\Sigma}_k)\|_F && (23)\\
&= \rho_k\sqrt{\sum_i(\mathbf{\Sigma}_k^{ii} - \mathcal{S}_{\frac{w_i}{\rho_k}}(\mathbf{\Sigma}_k^{ii}))^2} && (24)\\
&\leq \rho_k\sqrt{\sum_i(\frac{w_i}{\rho_k})^2} = \sqrt{\sum_i w_i^2}. && (25)
\end{aligned}$$

The inequality above can be proofed as follows: given the diagonal matrix $\mathbf{\Sigma}_k$, we define $\mathbf{\Sigma}_k^{ii}$ as the i-th element of $\mathbf{\Sigma}_k^{ii}$. If $\mathbf{\Sigma}_k^{ii} \geq \frac{w_i}{\rho_k}$, we have $\mathcal{S}_{\frac{w_i}{\rho_k}}(\mathbf{\Sigma}_k^{ii}) = \mathbf{\Sigma}_k^{ii} - \frac{w_i}{\rho_k} \geq 0$. If $\mathbf{\Sigma}_k^{ii} < \frac{w_i}{\rho_k}$, we have $\mathcal{S}_{\frac{w_i}{\rho_k}}(\mathbf{\Sigma}_k^{ii}) = 0 < \mathbf{\Sigma}_k^{ii} + \frac{w_i}{\rho_k}$. After all, we have $|\mathbf{\Sigma}_k^{ii} - \mathcal{S}_{\frac{w_i}{\rho_k}}(\mathbf{\Sigma}_k^{ii})| \leq \frac{w_i}{\rho_k}$ and hence the inequality holds true. Hence, the sequence $\{\mathbf{A}_k\}$ is upper bounded.

2. Secondly, we proof that the sequence of Lagrangian function $\{\mathcal{L}(\mathbf{X}_{k+1}, \mathbf{Z}_{k+1}, \mathbf{A}_k, \rho_k)\}$ is also upper bounded. Since the global optimal solution of $\mathbf{X}$ and $\mathbf{Z}$ in corresponding subproblems, we always have $\mathcal{L}(\mathbf{X}_{k+1}, \mathbf{Z}_{k+1}, \mathbf{A}_k, \rho_k) \leq \mathcal{L}(\mathbf{X}_k, \mathbf{Z}_k, \mathbf{A}_k, \rho_k)$. Based on the updating rule that $\mathbf{A}_{k+1} = \mathbf{A}_k + \rho_k(\mathbf{X}_{k+1} - \mathbf{Z}_{k+1})$, we have $\mathcal{L}(\mathbf{X}_{k+1}, \mathbf{Z}_{k+1}, \mathbf{A}_{k+1}, \rho_{k+1}) = \mathcal{L}(\mathbf{X}_{k+1}, \mathbf{Z}_{k+1}, \mathbf{A}_k, \rho_k) + \langle\mathbf{A}_{k+1} - \mathbf{A}_k, \mathbf{X}_{k+1} - \mathbf{Z}_{k+1}\rangle + \frac{\rho_{k+1}-\rho_k}{2}\|\mathbf{X}_{k+1} - \mathbf{Z}_{k+1}\|_F^2 = \mathcal{L}(\mathbf{X}_{k+1}, \mathbf{Z}_{k+1}, \mathbf{A}_k, \rho_k) + \frac{\rho_{k+1}+\rho_k}{2\rho_k^2}\|\mathbf{A}_{k+1} - \mathbf{A}_k\|_F^2$. Since the sequence $\{\|\mathbf{A}_k\}$ is upper bounded, the sequence $\{\|\mathbf{A}_{k+1} - \mathbf{A}_k\|_F\}$ is also upper bounded. Denote by a the upper bound of $\{\|\mathbf{A}_{k+1} - \mathbf{A}_k\|_F\}$, we have $\mathcal{L}(\mathbf{X}_{k+1}, \mathbf{Z}_{k+1}, \mathbf{A}_{k+1}, \rho_{k+1}) \leq \mathcal{L}(\mathbf{X}_1, \mathbf{Z}_1, \mathbf{A}_0, \rho_0) + a\sum_{k=0}^{\infty}\frac{\rho_{k+1}+\rho_k}{2\rho_k^2} = \mathcal{L}(\mathbf{X}_1, \mathbf{Z}_1, \mathbf{A}_0, \rho_0) + a\sum_{k=0}^{\infty}\frac{\mu+1}{2\mu^k\rho_0} \leq \mathcal{L}(\mathbf{X}_1, \mathbf{Z}_1, \mathbf{A}_0, \rho_0) + \frac{a}{\rho_0}\sum_{k=0}^{\infty}\frac{1}{\mu^{k-1}}$. The last inequality holds since $\mu + 1 < 2\mu$. Since $\sum_{k=0}^{\infty}\frac{1}{\mu^{k-1}} < \infty$, the sequence of Lagrangian function $\mathcal{L}(\mathbf{X}_{k+1}, \mathbf{Z}_{k+1}, \mathbf{A}_{k+1}, \rho_{k+1})$ is upper bound.

3. Thirdly, we proof that the sequences of $\{\mathbf{X}_k\}$ and $\{\mathbf{Z}_k\}$ are upper bounded. Since $\|\mathbf{W}(\mathbf{Y} - \mathbf{X})\|_F^2 + \|\mathbf{Z}\|_{\boldsymbol{w},*} = \mathcal{L}(\mathbf{X}_k, \mathbf{Z}_k, \mathbf{A}_{k-1}, \rho_{k-1}) - \langle\mathbf{A}_k, \mathbf{X}_k - \mathbf{Z}_k\rangle - \frac{\rho_k}{2}\|\mathbf{X}_k - \mathbf{Z}_k\|_F^2 = \mathcal{L}(\mathbf{X}_k, \mathbf{Z}_k, \mathbf{A}_{k-1}, \rho_{k-1}) + \frac{1}{2\rho_k}(\|\mathbf{A}_{k-1}\|_F^2 - \|\mathbf{A}_k\|_F^2)$. Thus $\{\mathbf{W}(\mathbf{Y} - \mathbf{X}_k)\}$ and $\{\mathbf{Z}_k\}$ are upper bounded, and hence the sequence $\{\mathbf{X}_k\}$ is bounded by Cauchy-Schwarz inequality and triangle inequality. We can obtain that $\lim_{k\to\infty}\|\mathbf{X}_{k+1} - \mathbf{Z}_{k+1}\|_F = \lim_{k\to\infty}\rho_k^{-1}\|\mathbf{A}_{k+1} - \mathbf{A}_k\|_F = 0$ and the equation (1) is proofed.

4. Then we can proof that $\lim_{k\to\infty}\|\mathbf{X}_{k+1} - \mathbf{X}_k\|_F = \lim_{k\to\infty}\|(\mathbf{W}^\top\mathbf{W} + \frac{\rho_k}{2}\mathbf{I})^{-1}(\mathbf{W}^\top\mathbf{W}\mathbf{Y} - \mathbf{W}^\top\mathbf{W}\mathbf{Z}_k - \frac{1}{2}\mathbf{A}_k) - \rho_k^{-1}(\mathbf{A}_k - \mathbf{A}_{k-1})\|_F \leq \lim_{k\to\infty}\|(\mathbf{W}^\top\mathbf{W} + \frac{\rho_k}{2}\mathbf{I})^{-1}(\mathbf{W}^\top\mathbf{W}\mathbf{Y} - \mathbf{W}^\top\mathbf{W}\mathbf{Z}_k - \frac{1}{2}\mathbf{A}_k)\|_F + \rho_k^{-1}\|\mathbf{A}_k - \mathbf{A}_{k-1}\|_F = 0$ and hence (2) is proofed.

5. Then (3) can be proofed by checking that $\lim_{k\to\infty}\|\mathbf{Z}_{k+1} - \mathbf{Z}_k\|_F = \lim_{k\to\infty}\|\mathbf{X}_k + \rho_k^{-1}\mathbf{A}_{k-1} - \mathbf{Z}_k + \mathbf{X}_{k+1} - \mathbf{X}_k + \rho_k^{-1}\mathbf{A}_{k-1} + \rho_k^{-1}\mathbf{A}_k - \rho_k^{-1}\mathbf{A}_{k+1}\|_F \leq \lim_{k\to\infty}\|\mathbf{\Sigma}_{k-1} - \mathcal{S}_{\boldsymbol{w}/\rho_{k-1}}(\mathbf{\Sigma}_{k-1})\|_F + \|\mathbf{X}_{k+1} - \mathbf{X}_k\|_F + \rho_k^{-1}\|\mathbf{A}_{k-1} + \mathbf{A}_{k+1} - \mathbf{A}_k\|_F = 0$, where $\mathbf{U}_{k-1}\mathbf{\Sigma}_{k-1}\mathbf{V}_{k-1}^\top$ is the SVD of the matrix $\mathbf{X}_k + \rho_{k-1}\mathbf{A}_{k-1}$. □

References

[1] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 60–65, 2005. 1, 2

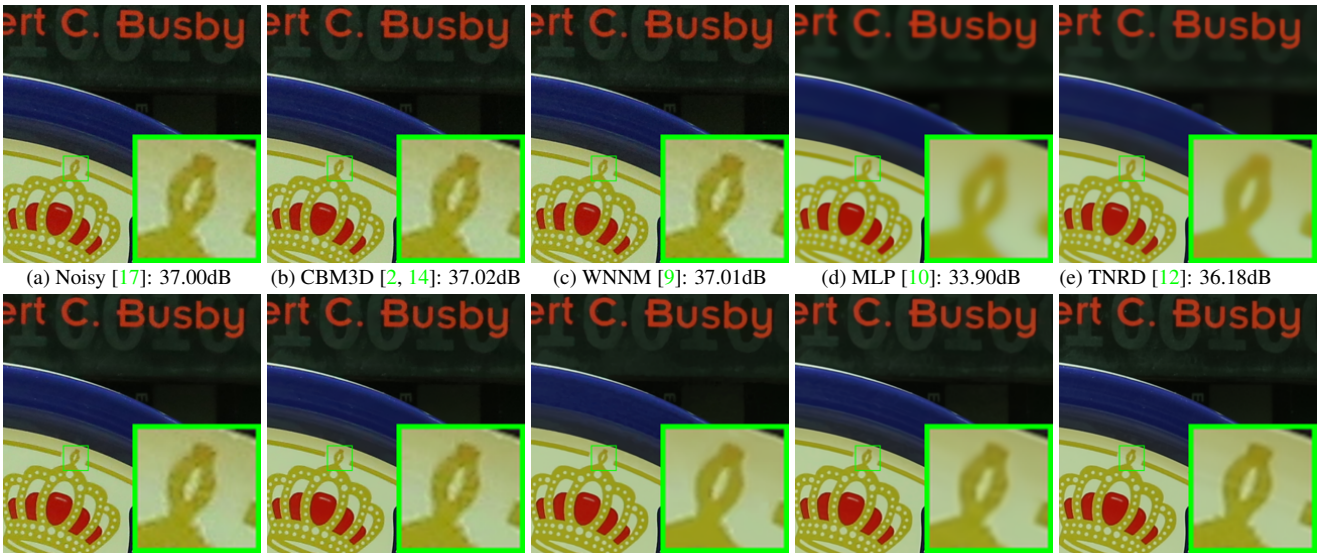
[2] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. 1, 2, 5, 6, 7, 9

[3] S. Roth and M. J. Black. Fields of experts. *International Journal of Computer Vision*, 82(2):205–229, 2009. 1

[4] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006. 1, 2

[5] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. *IEEE International Conference on Computer Vision (ICCV)*, pages 479–486, 2011. 1, 2

[6] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. *IEEE International Conference on Computer Vision (ICCV)*, pages 2272–2279, 2009. 1, 2



(a) Noisy [17]: 37.00dB (b) CBM3D [2, 14]: 37.02dB (c) WNNM [9]: 37.01dB (d) MLP [10]: 33.90dB (e) TNRD [12]: 36.18dB

(f) NI [34]: 37.68dB (g) NC [16, 33]: 38.76dB (h) CC [17]: 38.37dB (i) Ours: **40.50**dB (j) Mean Image [17]

Figure 4. Denoised images of a region cropped from the real noisy image "Canon 5D Mark 3 ISO 3200 1" [17] by different methods. The images are better to be zoomed in on screen.

[7] W. Dong, L. Zhang, G. Shi, and X. Li. Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing*, 22(4):1620–1630, 2013. 1

[8] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng. Patch group based nonlocal self-similarity prior learning for image denoising. *IEEE International Conference on Computer Vision (ICCV)*, pages 244–252, 2015. 1, 2

[9] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2862–2869, 2014. 1, 2, 4, 5, 6, 7, 9

[10] H. C. Burger, C. J. Schuler, and S. Harmeling. Image denoising: Can plain neural networks compete with BM3D? *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2392–2399, 2012. 1, 2, 5, 6, 7, 9

[11] U. Schmidt and S. Roth. Shrinkage fields for effective image restoration. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2774–2781, June 2014. 1, 2, 5

[12] Y. Chen, W. Yu, and T. Pock. On learning optimized reaction diffusion processes for effective image restoration. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5261–5269, 2015. 1, 2, 5, 6, 7, 9

[13] Julien Mairal, Michael Elad, and Guillermo Sapiro. Sparse representation for color image restoration. *IEEE Transactions on image processing*, 17(1):53–69, 2008. 1, 2

[14] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space. *IEEE International Conference on Image Processing (ICIP)*, pages 313–316, 2007. 1, 2, 5, 7, 9

[15] Z. Gong, Z. Shen, and K.-C. Toh. Image restoration with mixed or unknown noises. *Multiscale Modeling & Simulation*, 12(2):458–487, 2014. 1, 2

[16] M. Lebrun, M. Colom, and J.-M. Morel. Multiscale image blind denoising. *IEEE Transactions on Image Processing*, 24(10):3149–3161, 2015. 1, 2, 5, 6, 7, 9

[17] S. Nam, Y. Hwang, Y. Matsushita, and S. J. Kim. A holistic approach to cross-channel image noise modeling and its application to image denoising. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1683–1691, 2016. 1, 2, 6, 7, 8, 9

[18] F. Zhu, G. Chen, and P.-A. Heng. From noise modeling to blind image denoising. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 1, 2

[19] Hakki Can Karaimer and Michael S. Brown. A software platform for manipulating the camera imaging pipeline. *European Conference on Computer Vision (ECCV)*, October 2016. 1

[20] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, January 2011. 2

[21] Canyi Lu, Changbo Zhu, Chunyan Xu, Shuicheng Yan, and Zhouchen Lin. Generalized singular value thresholding. *AAAI*, 2015. 2

[22] Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010. 2

[23] Maryam Fazel. Matrix rank minimization with applications. *PhD thesis, Stanford University*, 2002. 2

[24] Nathan Srebro, Tommi Jaakkola, et al. Weighted low-rank approximations. *ICML*, 3(2003):720–727, 2003. 2

[25] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010. 2

[26] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011. 2

[27] Zhouchen Lin, Risheng Liu, and Zhixun Su. Linearized alternating direction method with adaptive penalty for low-rank representation. *Advances in neural information processing systems*, pages 612–620, 2011. 2

[28] Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936. 2

[29] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He. Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(9):2117–2130, Sept 2013. 2

[30] T. H. Oh, Y. W. Tai, J. C. Bazin, H. Kim, and I. S. Kweon. Partial sum minimization of singular values in robust pca: Algorithm and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(4):744–758, April 2016. 2

[31] 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*, pages 1–26, 2016. 2, 3, 5, 7

[32] C. Liu, R. Szeliski, S. Bing Kang, C. L. Zitnick, and W. T. Freeman. Automatic estimation and removal of noise from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):299–314, 2008. 2

[33] M. Lebrun, M. Colom, and J. M. Morel. The noise clinic: a blind image denoising algorithm. http://www.ipol.im/pub/art/2015/125/. Accessed 01 28, 2015. 2, 5, 6, 7, 9

[34] Neatlab ABSoft. Neat Image. https://ni.neatvideo.com/home. 2, 5, 6, 7, 9

[35] Nicholas J. Higham. Computing the nearest correlation matrixa problem from finance. *IMA Journal of Numerical Analysis*, 22(3):329, 2002. 6