MARLow: A Joint Multiplanar Autoregressive and Low-Rank Approach for Image Completion

Mading Li[1], Jiaying Liu[1], Zhiwei Xiong[2], Xiaoyan Sun[2], Zongming Guo[1]

[1]Institute of Computer Science and Technology, Peking University
[2]Microsoft Research Asia

Abstract. In this paper, we propose a novel multiplanar autoregressive (AR) model to exploit the correlation in cross-dimensional planes of a similar patch group collected in an image, which has long been neglected by previous AR models. On that basis, we then present a joint multiplanar AR and low-rank based approach (**MARLow**) for image completion from random sampling, which exploits the nonlocal self-similarity within natural images more effectively. Specifically, the multiplanar AR model constraints the local stationarity in different cross-sections of the patch group, while the low-rank minimization captures the intrinsic coherence of nonlocal patches. The proposed approach can be readily extended to multichannel images (e.g. color images), by simultaneously considering the correlation in different channels. Experimental results demonstrate that the proposed approach significantly outperforms state-of-the-art methods, even if the pixel missing rate is as high as 90%.

Keywords: Image completion, multiplanar autoregressive model, low-rank minimization

1   Introduction

Image restoration aims to recover original images from their low-quality observations, whose degradations are mostly generated by defects of capturing devices or error prone channels. It is one of the most important techniques in image/video processing, and low-level computer vision. In our work, we mainly focus on an interesting problem: image completion from random sampling, which has attracted many researchers' attention [1,2,3,4,5,6,7,8]. The problem is to recover the original image from its degraded observation, which has missing pixels randomly distributed. Such problem is a typical ill-posed problem, and different kinds of image priors have been employed.

One of the most commonly used image priors is the nonlocal prior [9], also known as the self-similarity property of natural images. Such prior reflects the fact that there are many similar contents frequently repeated in the whole image, which can be well utilized in image completion. A classic way is to process the collected similar patch groups. The reason is that similar degraded patches contain complementary information for each other, which contributes to the
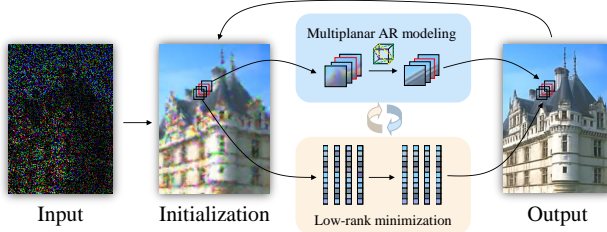
Fig. 1. The framework of the proposed image completion method MARLow. After obtaining an initialization of the input image, similar patches are collected. Then, the joint multiplanar autoregressive and low-rank approach is applied on grouped patches. After all patches are processed, overlapped patches are aggregated into a new intermediate image, which can be used as the input for the next iteration.

completion. According to the manipulation scheme applied to the patch group, there are generally two kinds of methods in the literature:

Cube-based methods stack similar patches directly, and then manipulate the data cube. The well-known denoising method Block-Matching and 3D filtering (BM3D) algorithm [10] is one of the most representative cube-based methods, which performs a 1D transform on each dimension of the data cube. The idea has been widely studied, and many extensions have been presented [11,12,7]. These methods perform a global optimization on the data cube, neglecting the local structures inside the cube. Also, they process the data cube along each dimension, failing to consider the correlation that exists in cross-dimensional planes of the data cube. In this paper, we propose a multiplanar autoregressive (AR) model to address these problems. Specifically, the multiplanar AR model is to constrain the local stationarity in different sections of the data cube. Nonetheless, the multiplanar AR model is not good at smoothing the intrinsic structure of similar patches.

Matrix-based methods stretch similar patches into vectors, which are spliced to form a data matrix. Two popular approaches, sparse coding and low-rank minimization, can be applied to such matrices. For sparse coding, the sparse coefficients of each vector in the matrix should be similar. This amounts to restricting the number of nonzero rows of the sparse coefficient matrix [13,14]. Zhang *et al.* [15] presented a group-based sparse representation method, which regards similar patch groups as its basic units. For low-rank minimization, since the data matrix is constructed by similar vectors, the rank of its underlying clean matrix to be recovered should be low. By minimizing the rank of the matrix, inessential contents (*e.g.* the noise) of the matrix can be eliminated [16,4]. However, in image completion from random samples, such methods may excessively smooth the result, since they only consider the correlation of pixels at the same location of different patches. Also, unlike stacking similar patches directly, representing image patches by vectors shatters the local information stored in image patches.

Upon these analyses, these two kinds of methods seem to be relatively complemented. Thus, motivated by combining the merits of cube-based and matrix-based methods, we present a joint multiplanar autoregressive and low-rank ap-

proach (**MARLow**) for image completion (Figure 1). Instead of performing a global optimization on the data cube grouped by similar patches, we propose the concept of the multiplanar autoregressive model to exploit the local stationarity on different cross-sections of the data cube. Meanwhile, we jointly consider the matrix grouped by stretched similar patches, in which the intrinsic content of similar patches can be well recovered by low-rank minimization. In summary, our contributions lie in three aspects:

- We propose the concept of multiplanar autoregressive model, to characterize the local stationarity of cross-dimensional planes in the patch group.
- We present a joint multiplanar autoregressive and low-rank approach (MAR-Low) for image completion from random sampling, along with an efficient alternating optimization method.
- We extend our method to multichannel images by simultaneously considering the correlation in different channels, presenting encouraging performance.

2   Related Work

In this section, we briefly review and discuss the existing literature that closely relates to the proposed method, including approaches associated with the autoregressive model and low-rank minimization.

Autoregressive model   The autoregressive (AR) model has been extensively studied in the last decades. AR model refers to modeling a pixel as the linear combination of its supporting pixels, usually its known neighboring pixels. Based on the assumption that natural images have the property of local stationarity, pixels in a local area share the same AR parameters, *i.e.* the weight for each neighbor. AR parameters are often estimated from the low-resolution image [17,18]. Dong *et al.* proposed a nonlocal AR model [19] using nonlocal pixels as supporting pixels, which is taken as a data fidelity constraint. The 3DAR model has been proposed to detect and interpolate the missing data in video sequences [20,21]. Since video sequences have the property of temporal smoothness, AR model can be extended to temporal space by combining the local statistics in the single frame. Different from approaches mentioned above, we focus on different cross-sections of the data cube grouped by similar nonlocal patches in a single image and constrain the local stationarity inside different planes in the data cube simultaneously (Figure 2).

Low-rank minimization   As a commonly used tool in image completion, low-rank minimization aims to minimize the rank of an input corrupted matrix. It can be used for recovering/completing the intrinsic content of a degraded potentially low-rank matrix. The original low-rank minimization problem is NP-hard, and cannot be solved efficiently. Candès and Recht [22] proposed to relax the problem by using nuclear norm of the matrix, *i.e.* the sum of singular values, which has been widely used in low-rank minimization problems since then. As proposed in [16,4], similar patches in images/videos are collected to form a potentially low-rank matrix. Then, the nuclear norm of the matrix is minimized. Zhang *et al.* further presented the truncated nuclear norm [6], minimizing the sum of

small singular values. Ono *et al.* [23] proposed the block nuclear norm, leading to a suitable characterization of the texture component. Low-rank minimization can also be used on tensor completion. Liu *et al.* [5] regarded the whole input degraded color image as a potentially low-rank tensor, and defined the trace norm of tensors by extending the nuclear norm of matrices. However, most general natural images are not potentially low-rank. Thus, Chen *et al.* [24] attempted to recover the tensor while simultaneously capturing the underlying structure of it. In our work, we apply the nuclear norm of matrices, and we use singular value thresholding (SVT) method [25] to solve the low-rank minimization problem. Jointly combined with our multiplanar model (as elaborated in Section 3), our method produces encouraging image completion results.

3 The Proposed Image Completion Method

As discussed in previous sections, cube-based methods and matrix-based methods have their drawbacks, and they complement each other in some sense. In this section, we introduce the proposed multiplanar AR model to utilize information from cross-sections of the data cube grouped by similar patches. Moreover, combined with low-rank minimization, we present the joint multiplanar autoregressive and low-rank approach (MARLow) for image completion. At the end of this section, we extend the proposed method to multichannel images. For an input degraded image, we first conduct a simple interpolation-based initialization on it (see Figure 1), to provide enough information for patch grouping.

3.1 Multiplanar AR Model

Considering a reference patch of size $n \times n$, we collect its similar nonlocal patches. For a data cube grouped by similar patches, we observe its different cross-sections (cross-dimensional planes). As shown in Figure 2 (a), different cross-sections of the data cube possess local stationarity. Since AR models can measure the local stationarity of image signals, we naturally extend the conventional AR model to the multiplanar AR model to measure cross-dimensional planes.
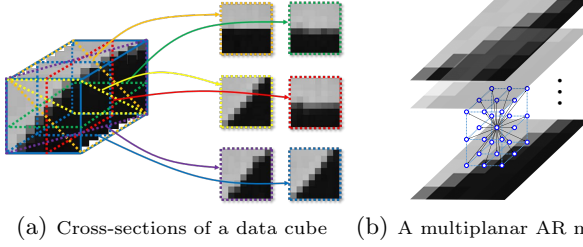
Generally, the conventional AR model is defined as

$$X(i, j) = \sum_{(m,n) \in \mathcal{N}} X(i + m, j + n) \cdot \varphi(m, n) + \sigma(i, j), \qquad (1)$$

where $X(i, j)$ represents the pixel located at (i, j). $X(i + m, j + n)$ is the supporting pixel with spatial offset (m, n), while $\varphi(m, n)$ is the corresponding AR parameter. $\mathcal{N}$ is the set of supporting pixels' offsets and $\sigma(i, j)$ is the noise.

In our work, the multiplanar AR model consists supporting pixels from different cross-dimensional planes (as illustrated in Figure 2 (b)). For a data cube grouped by similar patches of an image patch located at i, the multiplanar AR model of pixel $X_i(j, k, l)$ with offset (j, k, l) in the data cube is defined as

$$X_i(j, k, l) = \sum_{m \in \mathcal{N}_1} \sum_{\substack{(p, q) \\ \in \mathcal{N}_2}} Y_i(j + m, k + p, l + q) \cdot \varphi_i(m, p, q) + \sigma_i(j, k, l), \qquad (2)$$

(a) Cross-sections of a data cube   (b) A multiplanar AR model

Fig. 2. (a) Different cross-sections of a data cube grouped by similar patches also possess local stationarity, which can be well processed by AR models. (b) White dots represent pixels (*i.e.* small rectangles in 6×6 image patches). Black arrows connect the center pixel of a multiplanar AR model with its supporting pixels.

where $\mathcal{N}_1$ represents the set of supporting pixels' planar offsets and $\mathcal{N}_2$ represents the set of supporting pixels' spatial offsets (assuming the order of the multiplanar AR model $N_{order} = |\mathcal{N}_1| \times |\mathcal{N}_2|$). $Y_i(j + m, k + p, l + q)$ is the supporting pixel with offset (m, p, q) in the data cube and $\varphi_i(m, p, q)$ is the corresponding AR parameter. $\sigma_i(j, k, l)$ is the noise. Y is the initialization of the input image X. The reason we use Y here is that it is difficult to find enough known pixels to support the multiplanar AR model under high pixel missing rate.

For an $n \times n$ patch, assuming N patches are collected, the aforementioned multiplanar AR model can be transformed into a matrix form, that is,

$$X_i = T_i(Y) \cdot \varphi_i + \sigma_i, \tag{3}$$

where $X_i \in \mathbb{R}^{(n^2 \times N) \times 1}$ is a vector containing all modeled pixels. $T_i(\cdot)$ represents the operation that extract supporting pixels for X_i. Each row of $T_i(Y) \in \mathbb{R}^{(n^2 \times N) \times N_{order}}$ contains values of supporting pixels of each pixel and $\varphi_i \in \mathbb{R}^{N_{order} \times 1}$ is the multiplanar AR parameter vector.

Thus, the optimization problem for X_i and φ_i can be formulated as follows,

$$\underset{X_i, \varphi_i}{\arg\min} \|X_i - T_i(Y) \cdot \varphi_i\|_F^2. \tag{4}$$

where $\| \cdot \|_F$ is the Frobenius norm. In order to enhance the stability of the solution, we introduce the Tikhonov regularization to solve this problem. Specifically, a regularization term is included in the minimization problem, forming the following regularized least-square problem

$$\underset{X_i, \varphi_i}{\arg\min} \|X_i - T_i(Y) \cdot \varphi_i\|_F^2 + \|\Gamma \cdot \varphi_i\|_F^2, \tag{5}$$

where $\Gamma = \alpha I$ and I is an identity matrix.
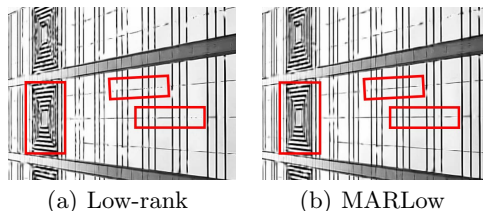
3.2 MARLow

Since the multiplanar AR model is designed to constrain a pixel with its supporting pixels on different cross-sections of the patch group, it can deal more

efficiently with local image structures. For instance, assume there is an edge on an image that is severely degraded, with only a few pixels on it. After collecting similar patches, low-rank minimization or other matrix-based methods may regard the remaining pixels as noises and remove them. However, with the multiplanar AR model, these pixels can be used to constrain each other and strengthen the underlying edge. Nevertheless, AR models are not suitable for smoothing the intrinsic structure, while low-rank minimization methods specialize in it. So we propose to combine the multiplanar AR model with low-rank minimization (MARLow) as follows,

$$
\begin{aligned}
\operatorname*{argmin}_{X_i,\varphi_i} &\|X_i - T_i(Y)\cdot\varphi_i\|_F^2 + \|\varGamma\cdot\varphi_i\|_F^2 \\
&+ \mu\left(\|R_i(X) - R_i(Y)\|_F^2 + \|R_i(X)\|_*\right),
\end{aligned}
\tag{6}
$$

where the last part is the low-rank minimization term restricting the fidelity while minimizing the nuclear norm (*i.e.* $\|\cdot\|_*$) of the data matrix. $R_i(\cdot)$ is an extraction operation that extracts similar patches of the patch located at i. $R_i(X) = [X_{i_1}, X_{i_2}, ..., X_{i_N}] \in \mathbb{R}^{n^2\times N}$ is similar patch group of the reference patch X_{i_1}, and $R_i(Y) = [Y_{i_1}, Y_{i_2}, ..., Y_{i_N}] \in \mathbb{R}^{n^2\times N}$ represents the corresponding patch group extracted from Y.

Figure 3 presents the completion results by using only low-rank without the multiplanar AR model, and by MARLow. From the figure, we can see that MARLow can effectively connect fractured edges.

(a) Low-rank   (b) MARLow

Fig. 3. (a) Completion result by low-rank without our multiplanar AR model (19.48 dB/0.8933); (b) Completion result generated by MARLow (**19.95 dB/0.8980**). Clearly, the result obtained by MARLow has higher visual quality, which can be observed in marked regions.

3.3 Multichannel Image Completion

For multichannel images, instead of applying the straightforward idea, that is, the separate procedure (*i.e.* processing different channels separately and combining the results afterward), we present an alternative scheme to simultaneously process different channels. At first, we collect similar patches of size $n \times n \times h$ (where h represents the number of channels) in a multichannel image. After that, each patch group is processed by simultaneously considers all channels. Specifically, the collected patches can be formed into h data cubes by stacking slices (of size $n \times n \times 1$) in the corresponding channel of different patches. For multiplanar

AR model, the minimization problem in Equation (5) turns into

$$\operatorname*{argmin}_{X_i^k, \varphi_i^k} \sum_{1 \leq k \leq h} \left(\left\| X_i^k - T_i^k(Y) \cdot \varphi_i^k \right\|_F^2 + \left\| \Gamma \cdot \varphi_i^k \right\|_F^2 \right) \tag{7}$$

For low-rank minimization, N collected patches are formed into a potentially low-rank data matrix of size $(n^2 \times h) \times N$ by representing each patch as a vector.

Taking an RGB image for an example, in patch grouping, we search for similar patches using reference patches with the size $n \times n \times 3$. The multichannel image completion problem can be solved by minimizing

$$\operatorname*{argmin}_{X_i^C, \varphi_i^C} \left\| X_i^C - T_i^C(Y^C) \cdot \varphi_i^C \right\|_F^2 + \left\| \Gamma \cdot \varphi_i^C \right\|_F^2$$
$$+ \mu \left(\left\| R_i^C(X^C) - R_i^C(Y^C) \right\|_F^2 + \left\| R_i^C(X^C) \right\|_* \right), \tag{8}$$

where

$$X_i^C = \begin{bmatrix} X_i^R \\ X_i^G \\ X_i^B \end{bmatrix} \in \mathbb{R}^{(n^2 \times N \times 3) \times 1},$$

$$\varphi_i^C = \begin{bmatrix} \varphi_i^R \\ \varphi_i^G \\ \varphi_i^B \end{bmatrix} \in \mathbb{R}^{(N_{order} \times 3) \times 1},$$

$$T_i^C(Y^C) = \begin{bmatrix} T_i(Y^R) & 0 & 0 \\ 0 & T_i(Y^G) & 0 \\ 0 & 0 & T_i(Y^B) \end{bmatrix} \in \mathbb{R}^{(n^2 \times N \times 3) \times (N_{order} \times 3)},$$

$$R_i^C(X^C) = \begin{bmatrix} X_{i_1}^R & X_{i_2}^R & \cdots & X_{i_N}^R \\ X_{i_1}^G & X_{i_2}^G & \cdots & X_{i_N}^G \\ X_{i_1}^B & X_{i_2}^B & \cdots & X_{i_N}^B \end{bmatrix} \in \mathbb{R}^{(n^2 \times 3) \times N},$$

$$R_i^C(Y^C) = \begin{bmatrix} Y_{i_1}^R & Y_{i_2}^R & \cdots & Y_{i_N}^R \\ Y_{i_1}^G & Y_{i_2}^G & \cdots & Y_{i_N}^G \\ Y_{i_1}^B & Y_{i_2}^B & \cdots & Y_{i_N}^B \end{bmatrix} \in \mathbb{R}^{(n^2 \times 3) \times N}.$$

The notations are given similarly as the preceding definitions. By utilizing the information in multichannel images, the patch grouping can be more precise. Furthermore, rich information in different channels can compensate for each other and constrain the completion result. Figure 4 illustrates the difference between processing different channels separately and simultaneously (with 80% pixels missing). Compared with the separate procedure, the multichannel image completion approach can significantly improve the performance of our method. Thus, in Section 5, for those methods dedicated to gray-scale image completion, we do not apply the separate procedure to them to obtain color image completion results since it may be unfair. Instead, we compare our multichannel image completion method with other state-of-the-art color image completion methods.

(a) 25.77 dB   (b) 30.36 dB   (c) 26.66 dB   (d) 34.38 dB
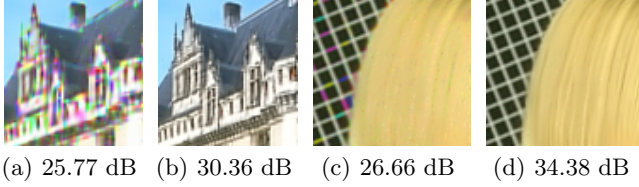
Fig. 4. (a) and (c) are obtained by the separate procedure. (b) and (d) are obtained by our multichannel image completion method.

4 Optimization

In this section, we present an alternating minimization algorithm to solve the minimization problems in Equation (6) and (8). Take Equation (6) for an example. We address each of the variable X_i and φ_i separately and present an efficient optimization algorithm.

When fixing X_i, the problem turns into

$$\underset{\varphi_i}{\operatorname{argmin}} \|X_i - T_i(Y) \cdot \varphi_i\|_F^2 + \|\Gamma \cdot \varphi_i\|_F^2, \tag{9}$$

which is a standard regularized linear least square problem, and can be solved by ridge regression. The closed-form solution is given by

$$\varphi_i = \left(\hat{Y}^T \hat{Y} + \Gamma^T \Gamma\right)^{-1} \hat{Y} X_i, \tag{10}$$

where $\hat{Y} = T_i(Y)$.

With φ_i fixed, the problem for updating X_i becomes

$$\underset{X_i}{\operatorname{argmin}} \|X_i - T_i(Y) \cdot \varphi_i\|_F^2 \\ + \mu \left(\|R_i(X) - R_i(Y)\|_F^2 + \|R_i(X)\|_* \right). \tag{11}$$

Here we notice that X_i and $R_i(X)$ contain the same elements. Their only difference is the formation: X_i is a vector and $R_i(X)$ is a matrix. Since we use Frobenius norm here, the value of the norm does not change if we reform the vector into a matrix form. So we reform X_i into a matrix M_i corresponding to $R_i(X)$ (in this way, $R_i(X)$ does not need to be reformed, and it can be represented by M_i directly). The vector $T_i(Y) \cdot \varphi_i$ is also reformed into a matrix form, represented by Y_{1_i}. By denoting $Y_{2_i} = R_i(Y)$, we can get the simplified version of Equation (11):

$$\underset{M_i}{\operatorname{argmin}} \|M_i - Y_{1_i}\|_F^2 + \mu \left(\|M_i - Y_{2_i}\|_F^2 + \|M_i\|_* \right). \tag{12}$$

It is a modified low-rank minimization problem and can be transformed into the following formation

$$\underset{M_i}{\operatorname{argmin}} \|M_i - Y_i'\|_F^2 + \lambda \|M_i\|_*, \tag{13}$$

where $Y_i' = (1-\lambda)Y_{1_i} + \lambda Y_{2_i}$ and $\lambda = \mu/(\mu+1)$. The problem now turns into a standard low-rank minimization problem [25]. Its closed-form solution is given as

$$M_i = S_\tau(Y_i'), \tag{14}$$

where $S_\tau(\cdot)$ represents the soft shrinkage process.

With the input random sampled image Y and the mask matrix M_{mask} indicating known pixels (0s for missing pixels and 1s for known pixels), our alternating minimization algorithm for image completion from random sampling can be summarized in Algorithm 1.

Algorithm 1 A Joint Multiplanar Autoregressive and Low-Rank Approach for Image Completion

---

Input: Y and M_{mask}.
$X^0 = Bilinear(Y, M_{mask})$.
for $i = 1$ to $maxIter$ **do**
   Patch grouping.
   for each image patch group X_k in $X^{(i-1)}$ **do**
      Estimate φ_k according to Equation (9).
      Estimate X_k according to Equation (11).
   end for
   Estimate $X^{(i)}$ by aggregating all overlapped patches.
   $X^{(i)}(M_{mask}) = Y(M_{mask})$.
   $Y = X^{(i)}$.
end for
Output: The restored image $X^{(i)}$.

---

5 Experimental Results

Experimental results of compared methods are all generated by the original authors' codes, with the parameters manually optimized. Both objective and subjective comparisons are provided for a comprehensive evaluation of our work. Peak Signal to Noise Ratio (PSNR) and structural similarity (SSIM) index are used to evaluate the objective image quality. In our implementation, if not specially stated, the size of each image patch is set to 8×8 ($5 \times 5 \times 3$ in color images) with four-pixel (one-pixel in color images) overlap. The number of similar patches is set to $N = 64$ for gray-scale image and $N = 75$ for color image. Other parameters in our algorithm are empirically set to $\alpha = \sqrt{10}$, $\mu = 10$. Please see the electronic version for better visualization of the subjective comparisons. More results can be found in the supplementary materials.

5.1 Gray-Scale Image Completion

For gray-scale images, we compare our method with state-of-the-art gray-scale image completion methods BPFA [14], BNN [23], ISDSB [2], and JSM [7]. Table 1

Table 1. PSNR (dB) and SSIM results of gray-scale image completion from different methods under 80% missing rate. The best result in each case is highlighted in bold.

Image	ISDSB	BNN	BPFA	JSM	Proposed
House	25.61/0.8052	27.76/0.8381	30.19/0.8717	33.00/0.8944	**34.70/0.9070**
Lena	27.31/0.8142	28.58/0.8416	30.95/0.8794	31.49/0.8836	**32.84/0.9043**
Cameraman	21.72/0.7584	22.65/0.7867	24.04/0.8082	25.18/0.8439	**25.49/0.8581**
Pepper	27.06/0.8205	27.87/0.8389	29.85/0.8529	31.75/0.8664	**32.59/0.8781**
Average	25.42/0.7996	26.72/0.8263	28.76/0.8530	30.35/0.8721	**31.41/0.8869**

Table 2. PSNR (dB) and SSIM results of color image completion from different methods under 80% and 90% missing rates. The best result in each case is highlighted in bold.

Image	Ratio	FoE	ST-NLTV	GSR	BPFA	Proposed
Castle	80%	25.71/0.8424	26.60/0.8414	25.66/0.8588	29.22/0.9099	**30.36/0.9124**
	90%	23.38/0.7655	23.39/0.7507	22.99/0.7771	25.09/0.8189	**26.55/0.8509**
Woman	80%	19.83/0.7948	21.37/0.7992	31.71/0.9460	30.33/0.9363	**34.38/0.9561**
	90%	17.16/0.6520	16.79/0.5973	20.02/0.8016	24.20/0.8640	**30.12/0.9179**
Soldier	80%	23.95/0.8434	25.10/0.8320	25.54/0.8963	28.92/0.9352	**30.78/0.9473**
	90%	21.03/0.7237	20.44/0.6911	21.82/0.7922	24.09/0.8368	**26.34/0.8893**
Average		21.84/0.7703	22.28/0.7520	24.62/0.8453	26.98/0.8835	**29.76/0.9123**

shows PSNR/SSIM results of different methods on test images with 80% pixels missing. From Table 1, the proposed method achieves the highest PSNR and SSIM in all cases, which fully demonstrates the effectiveness of our method. Specifically, the improvement on PSNR is 1.06 dB and that on SSIM is 0.0148 on average compared with the second best algorithm (*i.e.* JSM).

Figure 5 compares the visual quality of completion results for test images (with 80% pixels missing). From Figure 5, ISDSB and BNN successfully recover the boundaries of the image, but fail to restore rich details. BPFA performs better completion on image details. Nonetheless, there are plenty of noises along edges recovered by BPFA. At the first glance, the completion results of JSM and our method are both of high quality. However, if we get a closer look, it can be observed that there are isolated noises on image details (such as structures on Lena's hair and her hat) in the result generated by JSM. JSM also cannot recover tiny structures. Our method presents the best visual quality, especially on image details and edges.

5.2   Color Image Completion

We compare our method with state-of-the-art color image completion methods FoE [26], BPFA [14], GSR [15] and ST-NLTV [1]. Table 2 lists PSNR/SSIM results of different methods on color images with 80% and 90% pixels missing. It is clear that the proposed method achieves the highest PSNR/SSIM in all cases. Compared with gray-scale images, our image completion method performs even better on color images judging from the average PSNR and SSIM. The proposed method outperforms the second best method (*i.e.* BPFA) by 2.78 dB on PSNR
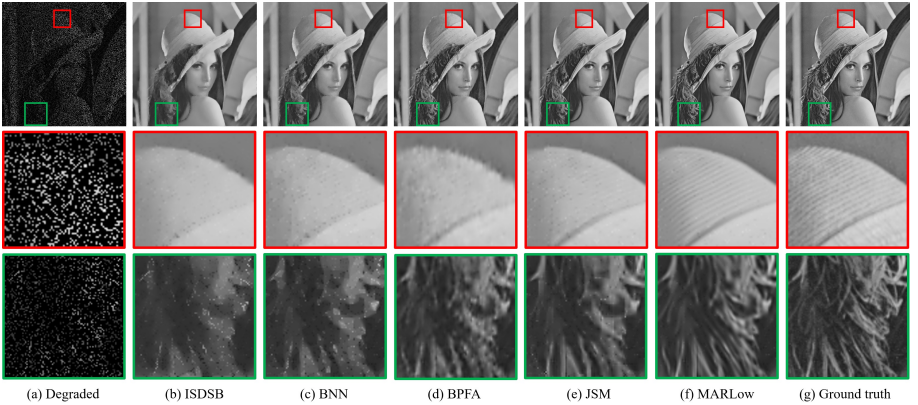
Fig. 5. Comparison of completion results of different methods with 80% pixels missing. From left to right: the degraded image, results of ISDSB, BNN, BPFA, JSM, our method, and the ground truth. The second and third rows show the corresponding close-ups.

and 0.0288 on SSIM. Note that, when tested on image *Woman* with 90% pixels missing, the PSNR and SSIM improvements achieved by our method over BPFA are 5.92 dB and 0.0539, respectively.

Figure 6 shows the visual quality of color image completion results for test images (with 90% pixels missing). Apparently, all the comparing methods are doing great on flat regions. However, FoE and ST-NLTV cannot restore fine details. GSR is better on recovering details, but it generates noticeable artifacts around edges and fails to connect fractured edges. BPFA produces sharper edges, but its performance under higher missing rate is not satisfying. The result of our method is of the best visual quality, especially under higher missing rate.

We also compare our method with state-of-the-art low-rank matrix/tensor completion based methods TNNR [6], LRTC [5] and STDC [24]. Since these methods regard the whole image as a potentially low-rank matrix, the input image should have strong correlations between its columns or rows. Thus, to be fair, we also test this kind of images to evaluate the performance of our method. From Figure 7, TNNR and LRTC tends to erase tiny objects of the image, such as the colorful items (see the close-ups in Figure 7). STDC imports noticeable noises into the whole image. The proposed method presents not only accurate completion on sharp edges, but also high-quality textures, exhibiting the best visual quality.

5.3 Text Removal

Text removal is one of the classic case of image restoration. The purpose of text removal is to recover the original image from a degraded version by removing the text mask. We have compared our method with four state-of-the-art algorithms: KR [27], FoE [28], JSM [7] and BPFA [14]. Our experimental settings of text removal are the same with those in color image restoration. Tab. 3 shows the PSNR and SSIM results of different methods. Fig. 8 presents visual comparison of different approaches, which further illustrates the effectiveness of our method.
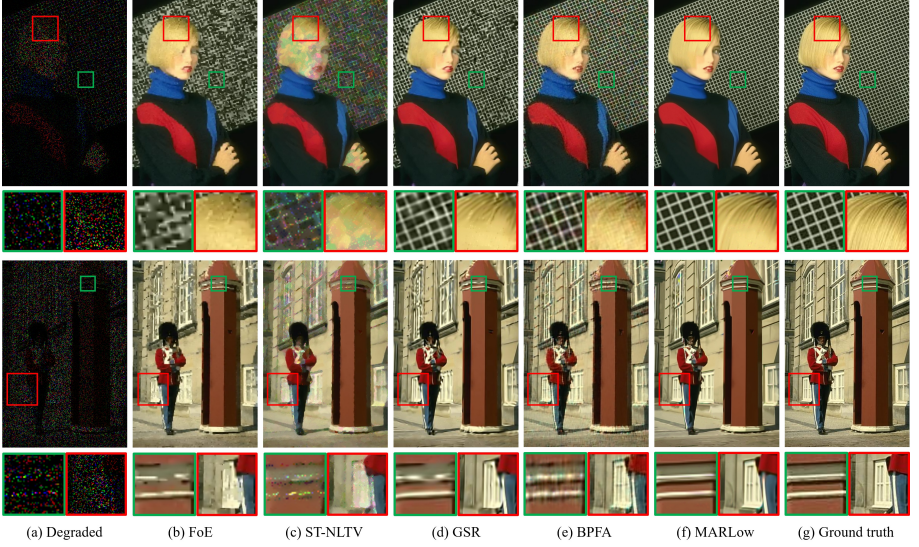
Fig. 6. Comparison of color image completion results of different methods with 90% pixels missing. From left to right: the degraded image, results of FoE, ST-NLTV, GSR, BPFA and our method, the ground truth.
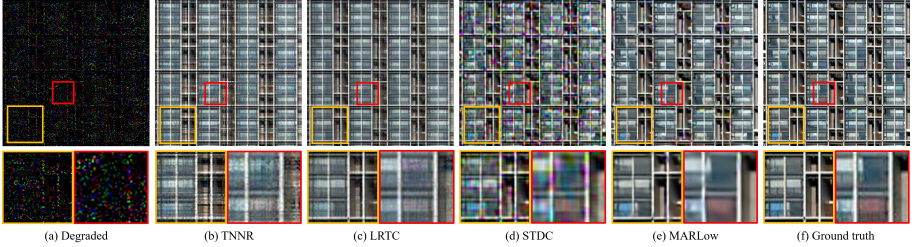
Fig. 7. Visual quality comparison of low-rank based methods. (a) The degraded image (with 90% pixels missing, 6.12 dB/0.0292); (b) - (e) Completion results by TNNR (17.58 dB/0.7276), LRTC (19.78 dB/0.8226), STDC (17.65 dB/0.7839), and the proposed method (**21.71 dB/0.9021**). (f) The ground truth image.

Table 3. PSNR (dB) and SSIM results of text removal from different methods. The best result in each case is highlighted in bold.

Image	KR	FoE	BPFA	JSM	Proposed
Barbara	29.59/0.9578	30.18/0.9585	32.91/0.9647	36.56/0.9839	**37.81/0.9862**
Parthenon	29.69/0.9374	31.87/0.9535	31.90/0.9506	33.07/0.9631	**33.27/0.9656**
Butterfly	30.22/0.9717	30.04/0.9713	30.07/0.9595	31.85/0.9797	**33.18/0.9844**
Foreman	40.51/0.9848	38.81/0.9863	38.53/0.9733	39.70/0.9870	**43.30/0.9887**
Average	32.50/0.9629	32.73/0.9674	33.35/0.9620	35.30/0.9784	**36.89/0.9812**

(a) Degraded   (b) FoE   (c) BPFA   (d) MARLow   (e) Ground truth
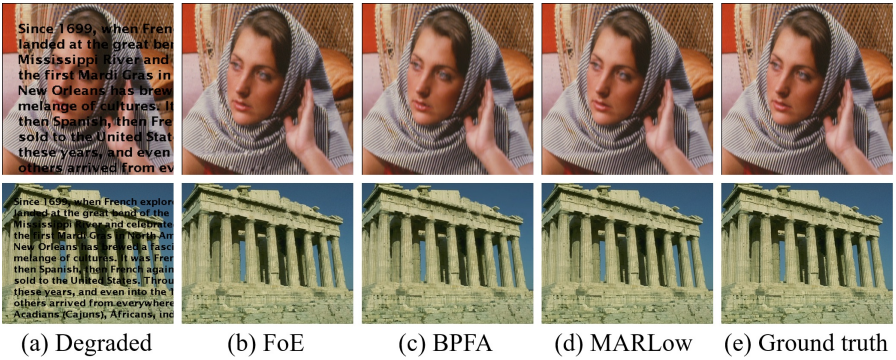
Fig. 8. Visual quality comparison of text removal for image *Barbara* and *Parthenon*. (a) The degraded image with text mask; (b) - (d) Restoration results by FoE, BPFA, and the proposed method. (e) The ground truth image.

5.4 Image Interpolation

The proposed method can also be applied on basic image processing problems, such as image interpolation. In fact, image interpolation can be regarded as a special circumstance of image restoration from random samples. To be more specific, locations of the known/missing pixels in image interpolation are fixed. Since our method is designed to deal with image restoration from random samples, we do not utilize this feature in our current implementation. Even so, we evaluate the performance of the proposed method with respect to image interpolation by comparing with other state-of-the-art interpolation methods. The compared methods including AR model based interpolation algorithms NEDI [17] and SAI [18], and a directional cubic convolution interpolation DCC [29]. Objective results are given in Tab. 4 and subjective comparisons are demonstrated in Fig. 9, showing that proposed method is competitive with other methods.

Table 4. PSNR (dB) and SSIM results of interpolation from different methods. The best result in each case is highlighted in bold.

Images	BC	NEDI	SAI	DCC	Proposed
Cameraman	25.37/0.8629	25.42/0.8626	25.88/0.8709	**25.92/0.8731**	25.36/0.8664
Lena	33.94/0.9149	33.78/0.9142	34.71/0.9193	34.50/0.9197	**34.90/0.9225**
Lighthouse	26.93/0.8436	26.37/0.8386	26.65/0.8445	27.19/**0.8483**	**27.22**/0.8462
Monarch	31.86/0.9561	31.78/0.9555	33.02/0.9623	32.92/0.9623	**33.16/0.9640**
Average	29.52/0.8944	29.33/0.8927	30.07/0.8992	30.13/**0.9009**	**30.16**/0.8998

6 Conclusion

In this work, we introduce the new concept of the multiplanar model, which exploits the cross-dimensional correlation in similar patches collected in a single image. Moreover, a joint multiplanar autoregressive and low-rank approach for

(a) Bicubic   (b) NEDI   (c) SAI   (d) DCC   (e) MARLow   (f) Ground truth
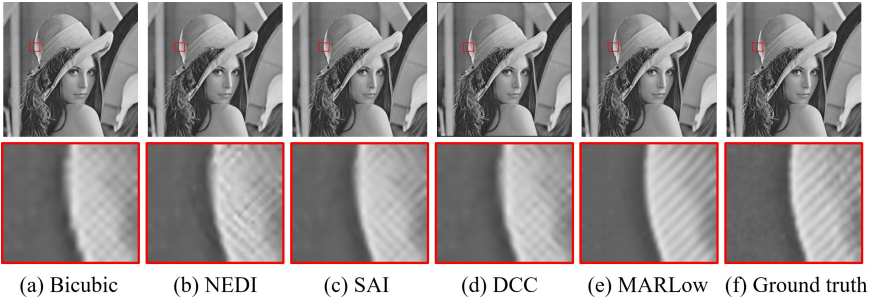
Fig. 9. Subjective comparison of interpolation for image *Lena*. Results of (a) Bicubic, (b) NEDI, (c) SAI, (d) DCC, (e) the proposed method, and (e) the ground truth image.

image completion from random sampling is presented, along with an alternating optimization algorithm. Our image completion method can be extended to multichannel images by utilizing the correlation in different channels. Extensive experiments on different applications have demonstrated the effectiveness of our method. Future works include the extensions on more other applications, such as video completion and hyperspectral imaging. We are also interested in adaptively choosing the size of the processing image patch since it might improve the completion result.

References

1. Chierchia, G., Pustelnik, N., Pesquet-Popescu, B., Pesquet, J.C.: A nonlocal structure tensor-based approach for multicomponent image recovery problems. Image Processing, IEEE Transactions on **23**(12) (Dec 2014) 5531–5544

2. He, L., Wang, Y.: Iterative support detection-based split bregman method for wavelet frame-based image inpainting. Image Processing, IEEE Transactions on **23**(12) (Dec 2014) 5470–5485

3. Heide, F., Heidrich, W., Wetzstein, G.: Fast and flexible convolutional sparse coding. In: Computer Vision and Pattern Recognition (CVPR), IEEE Conference on. (June 2015) 5135–5143

4. Ji, H., Liu, C., Shen, Z., Xu, Y.: Robust video denoising using low rank matrix completion. In: Computer Vision and Pattern Recognition (CVPR), IEEE Conference on. (June 2010) 1791–1798

5. Liu, J., Musialski, P., Wonka, P., Ye, J.: Tensor completion for estimating missing values in visual data. In: Computer Vision, IEEE International Conference on. (Sept 2009) 2114–2121

6. Zhang, D., Hu, Y., Ye, J., Li, X., He, X.: Matrix completion by truncated nuclear norm regularization. In: Computer Vision and Pattern Recognition (CVPR), IEEE Conference on. (June 2012) 2192–2199

7. Zhang, J., Zhao, D., Xiong, R., Ma, S., Gao, W.: Image restoration using joint statistical modeling in a space-transform domain. Circuits and Systems for Video Technology, IEEE Transactions on **24**(6) (June 2014) 915–928

8. Zhang, Z., Ely, G., Aeron, S., Hao, N., Kilmer, M.: Novel methods for multilinear data completion and de-noising based on tensor-svd. In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. (June 2014) 3842–3849

9. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. Multiscale Modeling & Simulation **4**(2) (2005) 490–530

10. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. Image Processing, IEEE Transactions on **16**(8) (Aug 2007) 2080–2095

11. Dabov, K., Foi, A., Egiazarian, K.: Video denoising by sparse 3D transform-domain collaborative filtering. In: Signal Processing Conference, European. (Sept 2007) 145–149

12. Maggioni, M., Boracchi, G., Foi, A., Egiazarian, K.: Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms. Image Processing, IEEE Transactions on **21**(9) (Sept 2012) 3952–3966

13. Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Non-local sparse models for image restoration. In: Computer Vision, IEEE International Conference on. (Sept 2009) 2272–2279

14. Zhou, M., Chen, H., Paisley, J., Ren, L., Li, L., Xing, Z., Dunson, D., Sapiro, G., Carin, L.: Nonparametric bayesian dictionary learning for analysis of noisy and incomplete images. Image Processing, IEEE Transactions on **21**(1) (Jan 2012) 130–144

15. Zhang, J., Zhao, D., Gao, W.: Group-based sparse representation for image restoration. Image Processing, IEEE Transactions on **23**(8) (Aug 2014) 3336–3351

16. Dong, W., Shi, G., Li, X.: Nonlocal image restoration with bilateral variance estimation: A low-rank approach. Image Processing, IEEE Transactions on **22**(2) (Feb 2013) 700–711

17. Li, X., Orchard, M.: New edge-directed interpolation. Image Processing, IEEE Transactions on **10**(10) (Oct 2001) 1521–1527

18. Zhang, X., Wu, X.: Image interpolation by adaptive 2-d autoregressive modeling and soft-decision estimation. Image Processing, IEEE Transactions on **17**(6) (June 2008) 887–896

19. Dong, W., Zhang, L., Lukac, R., Shi, G.: Sparse representation based image interpolation with nonlocal autoregressive modeling. Image Processing, IEEE Transactions on **22**(4) (April 2013) 1382–1394

20. Goh, W., Chong, M., Kalra, S., Krishnan, D.: Bi-directional 3D auto-regressive model approach to motion picture restoration. In: Acoustics, Speech, and Signal Processing, IEEE International Conference on. Volume 4. (May 1996) 2275–2278 vol. 4

21. Kokaram, A., Rayner, P.: Detection and interpolation of replacement noise in motion picture sequences using 3D autoregressive modelling. In: Circuits and Systems, IEEE International Symposium on. Volume 3. (May 1994) 21–24 vol.3

22. Candès, E., Recht, B.: Exact matrix completion via convex optimization. Foundations of Computational Mathematics **9**(6) (2009) 717–772

23. Ono, S., Miyata, T., Yamada, I.: Cartoon-texture image decomposition using blockwise low-rank texture characterization. Image Processing, IEEE Transactions on **23**(3) (March 2014) 1128–1142

24. Chen, Y.L., Hsu, C.T., Liao, H.Y.: Simultaneous tensor decomposition and completion using factor priors. Pattern Analysis and Machine Intelligence, IEEE Transactions on **36**(3) (March 2014) 577–591

25. Cai, J.F., Cands, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM Journal on Optimization **20**(4) (2010) 1956–1982

26. Roth, S., Black, M.: Fields of experts: a framework for learning image priors. In: Computer Vision and Pattern Recognition (CVPR), IEEE Conference on. Volume 2. (June 2005) 860–867 vol. 2

27. Takeda, H., Farsiu, S., Milanfar, P.: Robust kernel regression for restoration and reconstruction of images from sparse noisy data. In: IEEE International Conference on Image Processing. (2006) 1257–1260

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

29. Zhou, D., Shen, X., Dong, W.: Image zooming using directional cubic convolution interpolation. Image Processing, IET **6**(6) (2012) 627–634