Image/Video Restoration via Multiplanar Autoregressive Model and Low-Rank Optimization

MADING LI and JIAYING LIU, Peking University
XIAOYAN SUN, Microsoft Research Asia
ZHIWEI XIONG, University of Science and Technology of China

In this article, we introduce an image/video restoration approach by utilizing the high-dimensional similarity in images/videos. After grouping similar patches from neighboring frames, we propose to build a multiplanar autoregressive (AR) model to exploit the correlation in cross-dimensional planes of the patch group, which has long been neglected by previous AR models. To further utilize the nonlocal self-similarity in images/videos, a joint multiplanar AR and low-rank based approach is proposed (MARLow) to reconstruct patch groups more effectively. Moreover, for video restoration, the temporal smoothness of the restored video is constrained by the Markov random field (MRF), where MRF encodes *a priori* knowledge about consistency of patches from neighboring frames. Specifically, we treat different restoration results (from different patch groups) of a certain patch as labels of an MRF, and temporal consistency among these restored patches is imposed. The proposed method is also suitable for other restoration applications such as interpolation and text removal. Extensive experimental results demonstrate that the proposed approach obtains encouraging performance comparing with state-of-the-art methods.

CCS Concepts: • **Computing methodologies** → **Image processing**;

Additional Key Words and Phrases: Image/video restoration, multiplanar autoregressive model, low-rank optimization, Markov random field

ACM Reference format:
Mading Li, Jiaying Liu, Xiaoyan Sun, and Zhiwei Xiong. 2019. Image/Video Restoration via Multiplanar Autoregressive Model and Low-Rank Optimization. *ACM Trans. Multimedia Comput. Commun. Appl.* 15, 4, Article 102 (December 2019), 23 pages.
https://doi.org/10.1145/3341728

1 INTRODUCTION

Image/video restoration aims to recover original images/videos 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

This work was supported in part by National Natural Science Foundation of China under contract nos. 61772043 and 61671419, and in part by the Beijing Natural Science Foundation under contract nos. L182002 and 4192025.
Authors' addresses: M. Li and J. Liu (corresponding author), Wangxuan Institute of Computer Technology, Peking University, Zhongguancun North Street 128#, Haidian, Beijing, China; emails: {martinli0822, liujiaying}@pku.edu.cn; X. Sun, Microsoft Research Asia, Danling Street 5#, Haidian, Beijing, China; email: xysun@microsoft.com; Z. Xiong, University of Science and Technology of China, Dept. EEIS, P.O. Box 4, Hefei 230027, Anhui, China; email: zwxiong@ustc.edu.cn.

1551-6857/2019/12-ART102 $15.00
https://doi.org/10.1145/3341728

computer vision. Researchers have been dedicated to such topics in the past few decades. Among different image restoration problems, image restoration from limited samples has attracted many researchers' attention. In particular, we focus on the restoration of images with random-missing pixels, as the problem addressed in several works [7, 15, 16, 19, 21, 27, 30, 31, 36, 47, 49, 51, 54]. Recovery of the random-missing pixels modeled by a random matrix is often required in real applications like compressive sensing [39, 40], inpainting [12, 43], and reconstruction of multispectral images [14, 32]. Besides image restoration, we also focus on video restoration from limited samples, which has been given less attention in the literature [21, 23, 31, 54]. Instead of simply regarding the video sequence as a whole tensor and performing global restoration [31, 54], we propose to process with overlapping clips while utilizing the information redundancy existing not only within a single frame but also between adjacent frames.

The nonlocal prior [3] is one of the most commonly used priors utilizing redundant information in images/videos. Such a prior reflects the fact that there are many similar contents frequently repeating in the whole image or adjacent video frames, which can be well utilized in image/video restoration. 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 restoration. According to the manipulation scheme applied to the patch group, there are generally two kinds of methods in the literature: cube based and mixed base.

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 [9] 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 [8, 33, 34, 51]. These methods perform a global optimization on the data cube, neglecting the local structures inside the cube. In addition, 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 article, 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 optimization, 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 [35, 56]. Zhang et al. [50] presented a group-based sparse representation method, which regards similar patch groups as its basic units. For low-rank optimization, 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 [10, 23]. However, when restoring contents from limited samples, such methods may excessively smooth the result, as they only consider the correlation of pixels at the same location of different patches. In addition, unlike stacking similar patches directly, representing patches by 1D vectors shatters the local information stored in 2D patches.

Upon these analyses, cube-based methods fail to preserve intrinsic contents while maintaining the local information; matrix-based methods shatter the local information while capturing intrinsic contents, which means that they are relatively complementary. Thus, motivated by combining the merits of cube-based and matrix-based methods, a joint multiplanar AR and low-rank approach (MARLow) for single frame restoration was proposed in our previous work [27]. Instead of performing a global optimization on the data cube grouped by similar patches, we proposed the concept of the multiplanar AR model to exploit the local stationarity on different cross-sections of

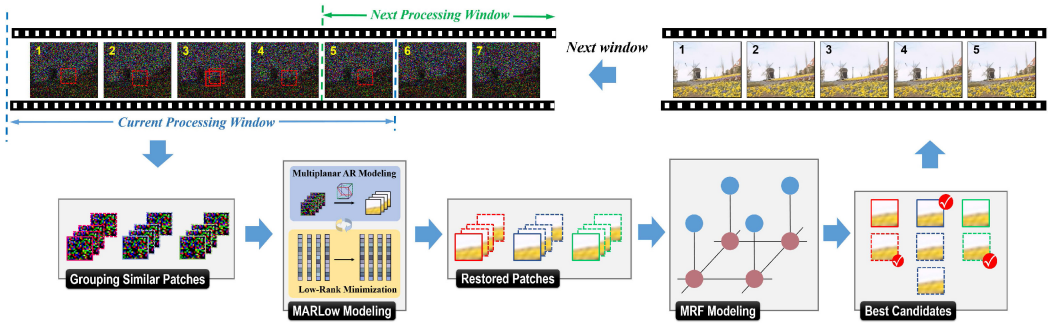


Fig. 1. Framework of the proposed video restoration method. After grouping similar patches from neighboring frames, the joint multiplanar AR and low-rank approach (MARLow) is applied on grouped patches to restore missing pixels. Then, the MRF model is formed, with different restoration results being treated as labels, to select the restoration candidates (patches with red check marks) considering their consistency with neighboring patches. Finally, these candidates are aggregated to restore the input frames.

the data cube. Meanwhile, we jointly considered the matrix grouped by stretched similar patches, in which the intrinsic content of similar patches can be well recovered by low-rank optimization.

Since the temporal continuity of video sequences provides more rich redundant information than a single image, the multiplanar AR model can be used in video restoration to further exploit the similarity in the spatiotemporal domain and help restore missing pixels. However, when extending algorithms from single-image restoration to video restoration, one should always avoid generating flickering artifacts. In other words, even if each single frame is recovered quite well, the visual quality of the overall restored video may not be good enough. This is mainly because the temporal smoothness of video sequences is not taken into consideration when each frame is restored individually. Thus, to produce a satisfying restoration result of the input degraded video, we should not only take into account the local stationarity in neighboring patches (spatial domain), but we must also consider the continuity in the temporal domain simultaneously. In this regard, we propose to use a Markov random field (MRF) model to constrain the temporal smoothness of the output video, ensuring that each restored patch is compatible in its spatiotemporal neighborhood. In summary, we present a multiplanar AR model and low-rank optimization-based video restoration method, with temporal smoothness constrained by an MRF model. The framework of the proposed method is illustrated in Figure 1.

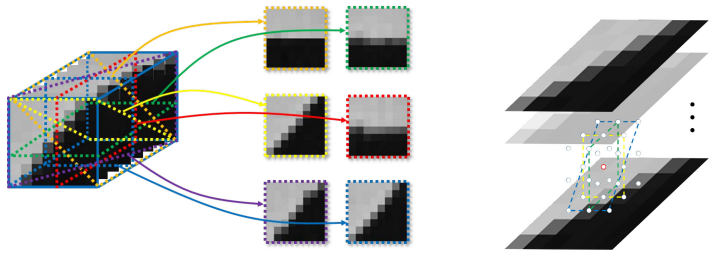
The rest of the article is organized as follows. Section 2 provides a brief review of the relevant literature. Section 3 introduces the proposed image/video restoration method via the multiplanar AR model and low-rank optimization, with a temporal smoothness constraint. Experimental results and analyses are presented in Section 4. Section 5 concludes the article.

2 RELATED WORK

In this section, we briefly review the existing literature that closely relates to the proposed method, including approaches associated with the AR model, low-rank optimization, and tensor completion.

2.1 AR Model

The AR model has been extensively studied in the past decades. The AR model refers to modeling a pixel as the linear combination of its supporting pixels, usually its known neighboring pixels.



(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 patches). Green, blue, and yellow rectangles in dashed lines are cross-dimensional planes passing through the center pixel of a multiplanar AR model (the red circle), containing supporting pixels of the model.

Generally, the conventional AR model is defined as

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

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), whereas $\varphi(m,n)$ is the corresponding AR parameter. $\mathcal{N}$ is the set of supporting pixels' offsets, and $\sigma(i,j)$ is the estimation error.

Based on the assumption that natural images/videos have the property of local stationarity, pixels in a local area share the same AR parameters, (i.e., the weight for each neighbor). Thus, pixels in a local patch can be modeled and estimated using the same AR parameters.

AR parameters are often estimated from the low-resolution image [28, 53]. Dong et al. [11] proposed a nonlocal AR model 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 [13, 24]. Since video sequences have the property of temporal smoothness, the AR model can be extended to the temporal domain by combining the local statistics in the single frame. Different from approaches mentioned earlier, we focus on different cross-sections of the data cube grouped by similar nonlocal patches and constrain the local stationarity inside different planes in the data cube simultaneously (Figure 2).

2.2 Low-Rank Optimization

As a commonly used tool in image restoration, low-rank optimization aims to minimize the rank of an input corrupted matrix. It can be used for recovering/completing the intrinsic content of a degraded matrix that is potentially low rank. The motivation is that noises are distributed less regularly than the principal component of the matrix. Given the input matrix P, the original low-rank optimization problem is defined as

$$\min \operatorname{rank}(L), \quad \text{s.t. } L_\Omega = P_\Omega, \tag{2}$$

where Ω represents locations of known elements.

Unfortunately, the original low-rank optimization problem (2) is NP-hard and cannot be solved efficiently. Therefore, we often consider to relax the problem by substituting $\|\cdot\|_*$ for $\operatorname{rank}(\cdot)$ [5]:

$$\min \|L\|_*, \quad \text{s.t. } L_\Omega = P_\Omega, \tag{3}$$

where $\|L\|_*$ is the nuclear norm of matrix L, which is the sum of the singular values. In some sense, this is the tightest convex relaxation of the NP-hard low-rank optimization problem (2). Its

equivalent form

$$\underset{L}{\operatorname{argmin}} \|L - P\|_F^2 + \lambda \|L\|_* \tag{4}$$

has been widely used in low-rank optimization problems. As proposed in other works [10, 23], similar patches in images/videos are collected to form a potentially low-rank matrix. Then, the nuclear norm of the matrix is minimized. Hu et al. [18] and Zhang et al. [49] further presented the truncated nuclear norm, minimizing the sum of small singular values. Ono et al. [38] proposed the block nuclear norm, leading to a suitable characterization of the texture component. In our work, we apply the nuclear norm of matrices, and we use the singular value thresholding (SVT) method [4] to solve the low-rank optimization problem. Jointly combined with our multiplanar model (as elaborated in Section 3), our method produces encouraging image/video restoration results.

2.3 Video Restoration Based on Tensors

Tensor completion methods have been studied to estimate missing values of visual data [6, 17, 25, 26, 31, 48, 54]. Low-rank optimization can also be used on tensor completion. Liu et al. [31] regarded the whole input video or 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. [6] attempted to recover the tensor while simultaneously capturing the underlying structure of it. Zhang et al. [54] proposed a tensor nuclear norm penalized algorithm for video completion from missing entries based on tensor–Singular Value Decomposition (t-SVD). Li et al. [25, 26] presented a multitensor completion that utilized the relationship between different datasets. Hu et al. [17] proposed the twist tensor nuclear norm and made an effective utilization of the temporal redundancy between frames and the spatial relationships between entries.

However, these methods assume that there is a global correlation in video frames, whereas a certain video usually contains multiple scenes and video frames that are not correlated with each other. To tackle the problem, Wang et al. [48] proposed a tensor completion method with spatiotemporal consistency. They introduced a smoothness regularization that ensures the smoothness of successive frames. Nevertheless, the method still neglects the correlation between similar patches in video frames.

3 THE PROPOSED IMAGE/VIDEO RESTORATION

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 multiplanar AR model to utilize information from cross-sections of the data cube grouped by similar patches. Moreover, combined with low-rank optimization and a smoothness constraint built on MRF, we present the joint multiplanar AR and low-rank approach (MARLow) for image/video restoration.

3.1 Multiplanar AR Model

Consider a reference patch of size $n \times n$; a data cube can be constructed by stacking its similar patches of size $n \times n$. Observing different cross-sections (cross-dimensional planes) of the cube (Figure 2(a)), we can see that different cross-sections of the data cube possess local stationarity. Since AR models can measure the local stationarity of 2D signals, we naturally extend the conventional AR model to the multiplanar AR model to measure cross-dimensional planes of 3D cubes. The similar patches are collected from adjacent frames in video restoration, as described in Section 3.3; as for image/single frame restoration, similar patches are grouped in the input image.

The multiplanar AR model consists of supporting pixels from different cross-dimensional planes (as illustrated in Figure 2(b)). For a data cube grouped by similar patches of a 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

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

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 estimation error. Y is the initialization of the input image or video frame X. The reason we use Y here is that it is difficult to find enough known pixels to support the multiplanar AR model under a 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,

$$X_i = T_i(Y) \cdot \varphi_i + \sigma_i, \tag{5}$$

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:

$$\underset{X_i, \varphi_i}{\operatorname{argmin}} \|X_i - T_i(Y) \cdot \varphi_i\|_F^2, \tag{6}$$

where $\| \cdot \|_F$ is the Frobenius norm.

To enhance the stability of the solution, we introduce the Tikhonov regularization to solve this problem. Specifically, a regularization term is included in the optimization problem, forming the following regularized least-square problem

$$\underset{X_i, \varphi_i}{\operatorname{argmin}} \|X_i - T_i(Y) \cdot \varphi_i\|_F^2 + \|\Gamma \cdot \varphi_i\|_F^2, \tag{7}$$

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

In our work, a multiplanar AR model is constructed by pixels on different planes of a tiny $3 \times 3 \times 3$ cube centered at the pixel to be estimated. All planes are 3×3 rectangles, and they all pass through the center pixel of the tiny cube. Figure 3 shows an ablation study that compares the proposed method using different planes. As can be observed in Figure 3(i) through (n), restoration results generated by MARLow using multiple planes are better than using a single plane. Figure 3(n) is the result of the proposed method, which outperforms other results both objectively and subjectively. Table 1 presents more quantitative results for the ablation study. As can be observed, the proposed method (using all planes) presents the best objective results.

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 structures. For instance, assume that there is an edge severely degraded, with only a few pixels on it. After collecting similar patches, low-rank optimization 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, whereas low-rank optimization methods specialize in it.

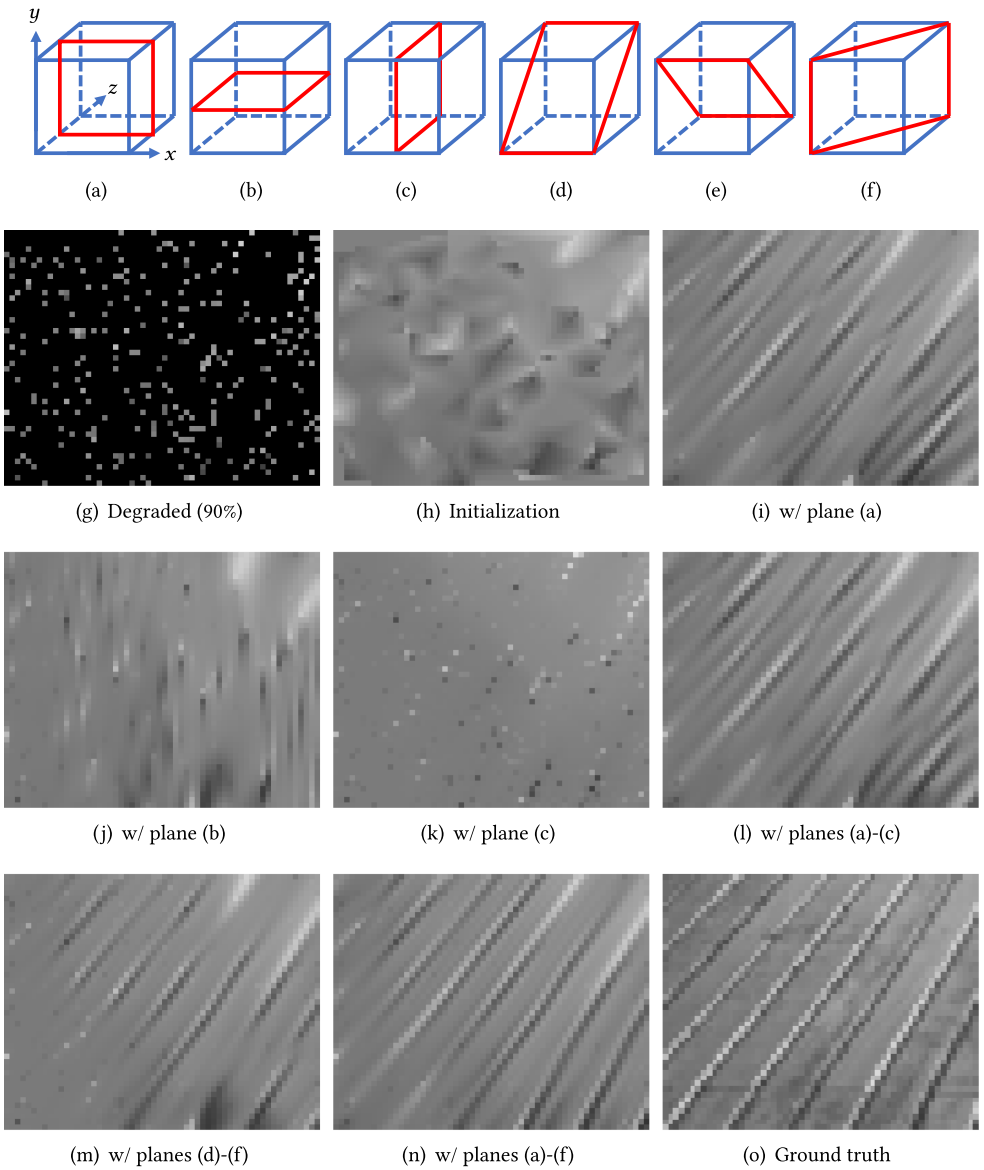


Fig. 3. (a)-(f) Planes used in our experiments. The x-axis and y-axis denote the patch spatial plane, and the z-axis is the dimension that stacking similar patches. (g) Close-up of the input degraded image *img_032*. (h) Initialization by Bilinear. (i)-(n) Restoration results by MARLow with different planes. (o) iGround truth. PSNR results of (i)-(n) are 22.16 dB, 21.60 dB, 22.19 dB, 22.52 dB, 23.32 dB, and **24.33** dB, respectively.

Thus, we propose to combine the multiplanar AR model with low-rank optimization (MARLow) as follows:

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

where the last part is the low-rank optimization 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

Table 1. PSNR (dB) Results of Our Method Using Different Planes

Images	(a)	(b)	(c)	(a)-(c)	(d)-(f)	(a)-(f)
Cathedral	21.18	23.69	21.38	23.65	20.73	**24.91**
Fountain	23.59	23.68	23.13	23.70	23.88	**25.74**
Img_024	13.54	17.56	15.80	17.10	13.65	**20.80**
Img_032	23.32	24.13	22.99	25.52	22.42	**27.09**

The best result in each case is highlighted in bold.

(a) (b) (c) (d)

Fig. 4. Restoration results of *Cathedral* and *Lena* by low rank without our multiplanar AR model, (a, c) and MARLow (b, d). The first row shows the zoom-in regions, and the second row gives full-size images.

patches of the patch located at i. $R_i(X) = [X_{i_1}, X_{i_2}, \ldots, X_{i_N}] \in \mathbb{R}^{n^2 \times N}$ is similar patch group of the reference patch X_{i_1}, and $R_i(Y) = [Y_{i_1}, Y_{i_2}, \ldots, Y_{i_N}] \in \mathbb{R}^{n^2 \times N}$ represents the corresponding patch group extracted from Y.

Figure 4 presents the restoration 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 (please observe the ceiling of the cathedral and the textures on Lena's hat).

For different channels of a frame, 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) from a frame. After that, each patch group is processed by simultaneously considering 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 the multiplanar AR model, the optimization problem in Equation (7) turns into

$$\underset{X_i^k, \varphi_i^k}{\operatorname{argmin}} \sum_{1 \le k \le h} \left(\left\| X_i^k - T_i^k(Y) \cdot \varphi_i^k \right\|_F^2 + \left\| \Gamma \cdot \varphi_i^k \right\|_F^2 \right). \tag{9}$$

(a) 25.77 dB (b) 30.36 dB (c) 26.66 dB (d) 34.38 dB

Fig. 5. (a) and (c) are obtained by the separate procedure. (b) and (d) are obtained by simultaneously processing different channels.

For low-rank optimization, 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 a frame as an example, in patch grouping, we search for similar patches using reference patches with the size $n \times n \times 3$. The single frame restoration problem can be solved by minimizing

$$\underset{X_i^C, \varphi_i^C}{\operatorname{argmin}} \left\| X_i^C - T_i^C(Y^C) \cdot \varphi_i^C \right\|_F^2 + \left\| \Gamma \cdot \varphi_i^C \right\|_F^2 + \mu \left(\left\| R_i^C(X^C) - R_i^C(Y^C) \right\|_F^2 + \left\| R_i^C(X^C) \right\|_* \right), \tag{10}$$

where

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

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

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

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

The notations are given similarly as the preceding definitions. By utilizing the information in different channels, the patch grouping can be more precise. Furthermore, information from different channels compensate for each other and constrain the restoration result. Figure 5 illustrates the difference between processing different channels separately and simultaneously (with 80% of pixels missing). Compared with the separate procedure, the simultaneous approach can significantly improve the performance of our method.

3.3 Temporal Smoothness Constraint

In this section, we introduce a constraint built on MRF to ensure the temporal smoothness of the video. The nonlocal prior can be perfectly applied to video sequences. In other words, instead of only searching for similar patches from a single frame (spatial domain), we expand the searching range to neighboring frames (temporal domain) to find more redundant information. More specifically, we search for similar patches of a reference patch in the current frame and neighboring

frames. After that, the collected similar patches are grouped into a data cube and then processed by MARLow.

Nevertheless, there is another important property that can be utilized in video restoration: the smoothness in the temporal domain. Generally, most video sequences contain few large motion vectors between adjacent frames, which allows us to constrain the restored patches.

Furthermore, we notice that for similar patch sets of different reference patches, there may be several patches located on the same location. Since MARLow reconstructs the whole similar patch set at once, there may be different restoration results for a certain patch. To select the best restoration candidate that satisfies the smoothness constraint, we optimize the following MRF energy function:

$$E(P) = \sum_{p \subset V} E_d\left(P(p)\right) + \sum_{\{p,p'\} \subset V} E_s\left(P(p), P(p')\right). \tag{11}$$

Here, V represents the video sequence. p and p' are temporally neighboring patches—that is, patches with the same spatial coordinates on neighboring frames. $P(p)$ selects one of the restoration candidates of patch p. Notice that there may be no restoration candidate for some patches.

The data term E_d is 0 if there exists at least one restoration candidate for p; otherwise, E_d is $+\infty$. The smoothness term E_s penalizes the incoherent temporal patches, and it is defined as follows:

$$E_s(p_1, p_2) = \|p_1 - p_2\|_F^2. \tag{12}$$

The energy function (11) is optimized using multilabel graph cuts [2]. Although we use MRF in our approach, one can also utilize optical flow to constrain the temporal smoothness.

After minimizing the energy function, the best restoration candidate of each patch can be determined. Patches without restoration candidates are set as the previous restoration result from the last iteration. All restored patches are then aggregated into a whole video sequence V, with overlapped regions averaged. Since it is unpractical to load the whole sequence, we instead process a temporal window with length $(2w + 1)$ containing the kth frame as reference frame and its neighboring frames (from the $(k - w)$th frame to the $(k + w)$th frame) at once, with one frame overlapping with the previous temporal window. If $w = 0$, the problem degenerated to the single frame restoration problem, that is, sequentially process each frame individually. Figure 6 illustrates zoomed area of the first five frames of *Foliage*. As can be observed, processing frames individually cannot restore shed leaves on the road in the first, third, and fourth frames, which can easily cause flickering artifacts. By contrast, the proposed method successfully recovers the information by utilizing redundant information between adjacent frames and maintains temporal consistency of the reconstructed video.

3.4 Optimization

In this section, we present an alternating minimization algorithm to solve the optimization problems in Equations (8) and (10). Take Equation (8) for an example. We address each of the variable X_i and φ_i separately and present an efficient optimization algorithm.

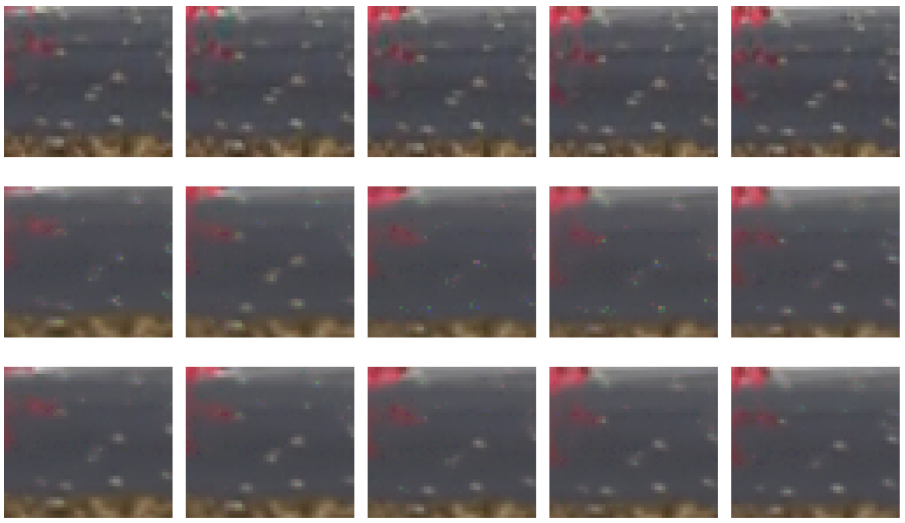
When fixing X_i, the problem turns into

$$\underset{\varphi_i}{\operatorname{argmin}} \|X_i - T_i(Y) \cdot \varphi_i\|_F^2 + \|\Gamma \cdot \varphi_i\|_F^2, \tag{13}$$

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

$$\varphi_i = \left(\hat{Y}^T \hat{Y} + \Gamma^T \Gamma\right)^{-1} \hat{Y} X_i, \tag{14}$$

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



(a) Frame 1 (b) Frame 2 (c) Frame 3 (d) Frame 4 (e) Frame 5

Fig. 6. From top to bottom: Close-ups of ground truth, restoration results of the sequential process, and restoration results of the proposed method.

With φ_i fixed, the problem for updating X_i becomes

$$\underset{X_i}{\operatorname{argmin}} \|X_i - T_i(Y) \cdot \varphi_i\|_F^2 + \mu \left(\|R_i(X) - R_i(Y)\|_F^2 + \|R_i(X)\|_* \right). \tag{15}$$

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. Thus, 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 (15):

$$\underset{M_i}{\operatorname{argmin}} \|M_i - Y_{1_i}\|_F^2 + \mu \left(\|M_i - Y_{2_i}\|_F^2 + \|M_i\|_* \right). \tag{16}$$

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

$$\underset{M_i}{\operatorname{argmin}} \|M_i - Y_i'\|_F^2 + \lambda \|M_i\|_*, \tag{17}$$

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 optimization problem [4]. Its closed-form solution is given as

$$M_i = S_\tau(Y_i'), \tag{18}$$

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

In our work, to utilize the updated data for further patch grouping, we propose a global iteration method. Specifically, after patch grouping, we perform MARLow on every patch group. Then, we solve the energy minimization problem to determine the best restoration patch and aggregate all overlapped patches into an intermediate frame for the next iteration. The iterative procedure continues until it reaches the maximum iteration number.

4 EXPERIMENTAL RESULTS

All of our experiments are implemented on the MATLAB platform. Testing images/videos come from the Sun-Hays dataset [44], the Urban dataset [20], the Berkeley segmentation dataset and benchmark (BSDS) [37], and the VideoSR dataset [29]. 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. The peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index are used to evaluate the objective image quality. To fully assess the proposed method, we first compare our video restoration method with existing methods. Then, we compare our single frame restoration method with state-of-the-art grayscale and color image restoration methods. Finally, we test our method on other interesting applications, such as image interpolation and text removal.

For color images (and videos), we evaluate the performance of the proposed algorithm with random-missing pixels across R, G, and B channels [7, 36, 56]. More specifically, three random binary masks are used to R, G, and B channels, separately, to generate a degraded color image. For an input, we first conduct a simple interpolation-based initialization (say Bilinear interpolation in our experiments) on it to provide enough information for patch grouping.

In our implementation, if not specially stated, the size of each patch is set to 8×8 for grayscale images and $5 \times 5 \times 3$ for color images/videos, with a four-pixel (one-pixel in color images/videos) overlap. The number of similar patches is set to $N = 64$ for grayscale images and $N = 75$ for color images/videos. 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 shown in the following.

4.1 Video Restoration

For video restoration, we have compared our method with a matrix-based denoising method by Ji et al. [23], two state-of-the-art tensor-based video restoration algorithms (t-SVD [54] and LRTC [31]), and a single-frame restoration method GSR [50]. Ji et al. [23] proposed a patch-based video denoising method capable of removing mixed noise, which is a representative matrix-based method. t-SVD is a tensor nuclear norm penalized algorithm for video restoration from missing entries. Generalized from the matrix trace norm, the tensor trace norm is proposed in Liu et al. [31] to deal with tensor completion problems. Both of these methods regard the video with missing pixels as a whole tensor and perform a global restoration process. The experimental settings of video restoration are basically the same with color image restoration. The temporal processing window length is set as 5.

Table 2 shows the average PSNR and SSIM results of different methods for six video sequences. The objective PSNR/SSIM results indicate that our method has a rather large improvement over other methods. Figure 7 and Figure 8 illustrate some of the video restoration results of different methods. They show that although t-SVD and LRTC are able to generate rather good results of static backgrounds, they still produce burrs around edges and boundaries. The method by Ji et al. [22] generates results with better visual qualities, but the method blurs the results to some extent. Our method is able to reconstruct clear and sharp edges, and it generates visually pleasant results.

The second and the third rows of Figure 8 show a particular situation that our method outperforms other tensor-based methods. As can be observed in the results, both LRTC and t-SVD generate severe ghosting artifacts. In fact, this sequence has rather still backgrounds (e.g., the sky region) and moving foregrounds (e.g., the penguin). This kind of sequence is particularly not suitable for methods built on the whole sequence to restore. The reason is that although some parts of the scene remain still, other parts keep changing. For LRTC and t-SVD, as they regard the whole

Table 2. Average PSNR (dB) and SSIM Results of Video Restoration from Different Methods

Sequences	Ratio	LRTC	t-SVD	Ji	GSR	Proposed
Foliage	80%	20.21/0.5651	22.20/0.6159	23.70/0.6918	23.93/0.7729	**30.08/0.9277**
	90%	16.83/0.3697	19.80/0.4278	21.67/0.5539	21.14/0.6039	**25.43/0.8074**
Walk	80%	20.49/0.6488	23.27/0.6820	26.59/0.8497	28.61/0.9194	**35.06/0.9630**
	90%	17.02/0.4786	20.44/0.5363	24.22/0.7769	24.72/0.8314	**29.83/0.9129**
Calendar	80%	17.46/0.4525	19.38/0.5374	20.01/0.6601	21.00/0.8001	**25.47/0.9052**
	90%	14.71/0.2792	17.21/0.3651	18.37/0.5507	18.25/0.6613	**21.43/0.7979**
City	80%	21.93/0.5433	24.20/0.5930	25.47/0.7193	29.55/0.8814	**35.02/0.9490**
	90%	19.47/0.4047	21.98/0.4291	23.53/0.5989	25.20/0.7338	**30.04/0.8756**
Penguin	80%	27.58/0.8570	30.23/0.8951	29.59/0.9255	31.28/0.9465	**37.76/0.9677**
	90%	22.73/0.6962	26.32/0.7969	27.16/0.8867	27.77/0.9046	**32.80/0.9397**
Temple	80%	22.45/0.6746	25.86/0.7600	24.73/0.7792	26.70/0.8928	**32.77/0.9582**
	90%	17.88/0.4598	22.58/0.6089	22.86/0.6910	23.37/0.7791	**27.57/0.8866**
Average		19.90/0.5358	22.79/0.6039	23.99/0.7236	25.13/0.8106	**30.27/0.9076**

The best result in each case is highlighted in bold.

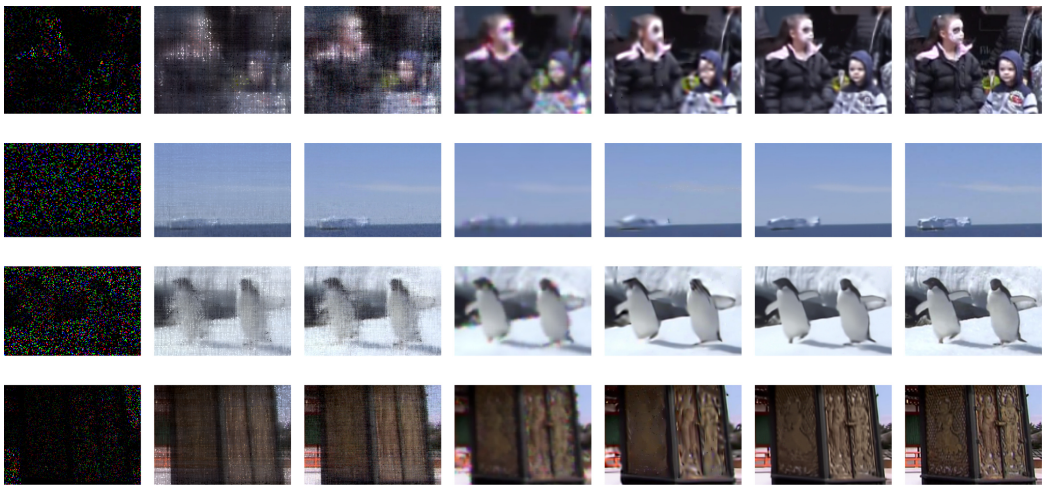
(a) Degraded (b) LRTC (c) t-SVD (d) Ji et al. (e) GSR (f) Proposed (g) Ground truth

Fig. 7. Visual quality comparison of different video restoration methods under an 80% missing rate. From top to bottom: Close-ups of *Foliage*, *Calendar*, *Temple*, and *City*.

sequence as a low-rank tensor, they can be easily affected by the content of input videos and generate noticeable artifacts. However, our method can successfully produce ghost-free results with no flickering artifacts that mainly benefit from MARLow implemented on videos and our temporal smoothness constraint.

4.2 Grayscale Image Restoration

For grayscale images, we compare our method with state-of-the-art grayscale image restoration methods BPFA [56], BNN [38], SAIST [10], JSM [51], and DIP [46]. BPFA considers a nonparametric Bayesian method using learned dictionaries to recover incomplete images, and it can be also



(a) Degraded (b) LRTC (c) t-SVD (d) Ji et al. (e) GSR (f) Proposed (g) Ground truth

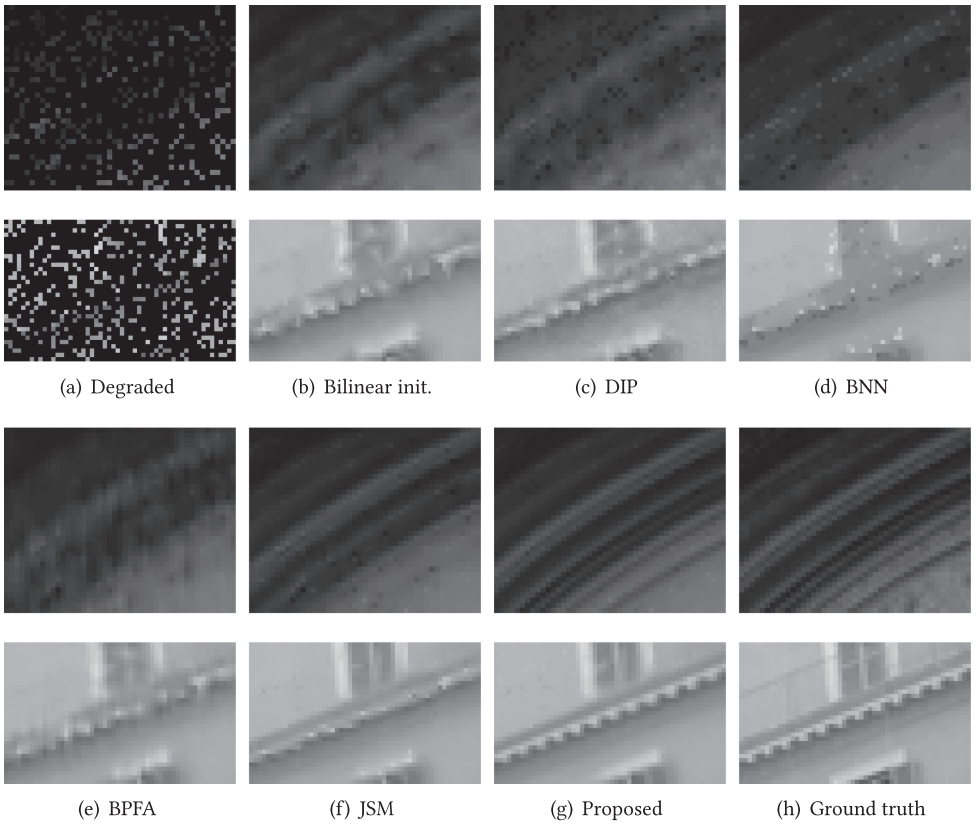
Fig. 8. Visual quality comparison of different video restoration methods under a 90% missing rate. From top to bottom: Close-ups of *Walk*, *Penguin*, *Penguin*, and *Temple*.

Table 3. PSNR (dB) and SSIM Results of Grayscale Image Restoration from Different Methods Under 80% and 90% Missing Rates

Images	Ratio	DIP	BNN	BPFA	JSM	SAIST	Proposed
Canyon	80%	27.15/0.7873	26.19/0.7622	28.01/0.8279	27.79/0.8188	28.28/0.8279	**28.48/0.8411**
	90%	24.86/0.6886	23.43/0.6370	24.94/0.6909	24.95/0.6959	25.39/0.7143	**25.93/0.7437**
Cathedral	80%	25.37/0.7397	23.71/0.6854	26.45/0.8081	26.82/0.8316	27.44/0.8443	**27.81/0.8675**
	90%	22.32/0.6097	21.50/0.5432	23.76/0.6849	23.82/0.6566	24.26/0.6797	**25.07/0.7588**
Chalet	80%	20.38/0.7199	19.72/0.6416	20.69/0.7139	20.89/0.7189	**21.24**/0.7517	21.09/**0.7544**
	90%	18.08/0.5858	17.56/0.4796	18.63/0.5560	18.76/0.5630	**19.01**/0.6083	18.90/**0.6190**
Cockpit	80%	23.55/0.6968	23.28/0.7315	25.11/0.8011	26.23/0.8287	26.65/0.8354	**26.86/0.8514**
	90%	21.89/0.6070	20.56/0.5849	22.17/0.6556	23.05/0.6913	23.63/0.7112	**23.91/0.7428**
Fountain	80%	26.86/0.7809	25.51/0.7254	27.25/0.8008	27.91/0.8181	28.49/0.8308	**28.70/0.8447**
	90%	24.62/0.6992	22.82/0.5765	24.30/0.6565	24.75/0.6739	25.37/0.7009	**25.89/0.7375**
Ruin	80%	27.92/0.8439	26.41/0.8081	27.81/0.8530	28.41/0.8590	28.70/0.8604	**29.02/0.8707**
	90%	25.68/0.7860	23.93/0.7139	25.30/0.7582	25.73/0.7685	26.00/0.7716	**26.54/0.7983**
Skyscraper	80%	23.94/0.7765	22.37/0.7830	24.67/0.8460	24.77/0.8445	25.81/0.8658	**26.16/0.8779**
	90%	21.41/0.5955	20.14/0.6798	21.95/0.7277	22.07/0.7402	22.45/0.7582	**22.86/0.7842**
Village	80%	23.59/0.7145	22.86/0.7001	23.97/0.7575	23.16/0.7497	24.10/0.7817	**25.01/0.7976**
	90%	21.83/0.6551	20.56/0.5604	21.54/0.6140	20.53/0.6014	21.80/0.6540	**22.45/0.6783**
Average		23.72/0.7054	22.53/0.6633	24.16/0.7345	24.35/0.7413	24.91/0.7623	**25.29/0.7855**

The best result in each case is highlighted in bold.

applied to color images. BNN introduces a new convex prior block nuclear norm to characterize texture components. SAIST takes a low-rank approach toward simultaneous sparse coding, developing the spatially adaptive iterative SVT for image restoration. JSM presents a joint statistical model in the hybrid space-transform domain. DIP is a latest deep learning–based method, and it utilizes neural networks to capture low-level image statistics prior.



(a) Degraded (b) Bilinear init. (c) DIP (d) BNN

(e) BPFA (f) JSM (g) Proposed (h) Ground truth

Fig. 9. Comparison of grayscale image restoration results of *Cathedral* and *Fountain* by different methods with 80% of pixels missing.

Table 3 shows PSNR/SSIM results of different methods on test images with 80% and 90% of pixels missing. From Table 3, the proposed method achieves the highest PSNR and SSIM in most cases, which fully demonstrates the effectiveness of our method. Specifically, the improvement on PSNR is 0.38 dB on average compared with the second best algorithm (i.e., SAIST).

Figure 9 compares the visual quality of restoration results for test images (with 80% of pixels missing). From Figure 9, it can be observed that BNNs fail to restore image details. BPFA and DIP show better performance and recover more details. Nonetheless, there are plenty of noises along edges recovered by BPFA, and DIP produces scratchy details. As for JSM and SAIST, they both produce noises on image details in addition to not generating fine details. Our method presents the best visual quality by preserving image details and edges.

4.3 Color Image Restoration

We compare our method with state-of-the-art color image restoration methods BPFA [56], GSR [50], KSVD [36], ST-NLTV [7], and DIP [46]. GSR represents group-based sparse representation, which enforces the intrinsic local sparsity and nonlocal self-similarity of images simultaneously in a unified sparse representation framework. KSVD learns the correlation between different R, G, and B channels based on learning models for sparse color image representation. ST-NLTV extends the nonlocal total variation regularization by taking advantage of the gradient of a multicomponent image.

Table 4. PSNR (dB) and SSIM Results of Color Image Restoration from Different Methods Under 80% and 90% Missing Rates

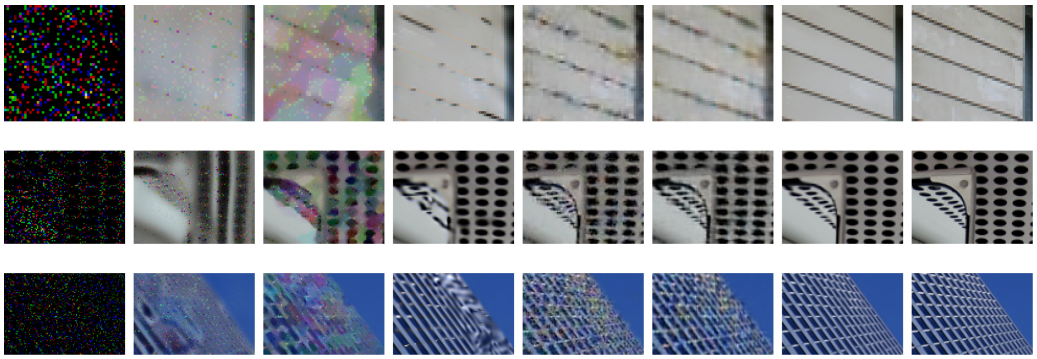
Images	Ratio	DIP	ST-NLTV	GSR	KSVD	BPFA	Proposed
Abbey	80%	23.94/0.7159	26.67/0.8268	25.81/0.8351	28.90/0.9067	28.58/0.8963	**29.77/0.9103**
	90%	22.90/0.6604	22.97/0.6869	23.42/0.7399	25.29/0.7999	25.21/0.7947	**26.51/0.8299**
Boardwalk	80%	22.22/0.5043	26.07/0.7261	25.17/0.7220	**28.92/0.8695**	28.86/0.8544	28.83/0.8430
	90%	20.34/0.3831	19.29/0.4920	23.10/0.5797	25.39/**0.7238**	25.68/0.7213	**25.90**/0.7056
Burial	80%	25.50/0.6249	27.96/0.7479	27.88/0.7608	28.91/0.8260	29.23/0.8117	**30.40/0.8377**
	90%	24.83/0.5659	23.65/0.5624	25.49/0.6409	25.44/0.6746	26.45/0.6860	**27.68/0.7203**
Inn	80%	24.88/0.6980	27.53/0.8080	27.14/0.8238	30.09/**0.9079**	30.00/0.8996	**30.71**/0.9038
	90%	23.83/0.6400	23.79/0.6590	24.48/0.7154	26.36/0.7989	26.53/0.8012	**27.39/0.8101**
Phone Booth	80%	25.24/0.7709	27.32/0.8439	30.11/0.9197	31.19/0.9449	32.53/0.9469	**34.39/0.9494**
	90%	24.80/0.7483	23.05/0.7454	25.87/0.8310	25.86/0.8461	27.46/0.8801	**29.75/0.8971**
Img_001	80%	18.90/0.5329	27.89/0.8461	28.10/0.8742	29.49/0.9147	30.57/0.9183	**32.68/0.9235**
	90%	21.08/0.5566	23.51/0.7354	23.92/0.7630	25.34/0.8056	26.19/0.8282	**28.65/0.8496**
Img_002	80%	23.08/0.7232	26.17/0.8474	27.79/0.9276	27.92/0.9136	28.53/0.9240	**32.28/0.9644**
	90%	21.13/0.6172	21.69/0.6731	27.85/**0.9280**	24.14/0.8043	24.89/0.8295	**28.11**/0.9172
Img_003	80%	19.88/0.5111	24.07/0.7604	24.16/0.7997	26.04/0.8593	26.29/0.8587	**28.12/0.8964**
	90%	19.22/0.4387	18.38/0.4910	21.01/0.6469	22.75/0.7257	22.96/0.7271	**24.66/0.7927**
Img_004	80%	16.73/0.4518	21.09/0.7643	30.54/0.9651	25.66/0.9267	26.30/0.9285	**34.51/0.9809**
	90%	16.03/0.3381	16.98/0.4951	20.42/0.8027	20.99/0.7911	21.91/0.8180	**28.32/0.9468**
Img_005	80%	21.44/0.8260	24.01/0.8981	28.14/0.9676	26.49/0.9399	26.83/0.9409	**36.41/0.9874**
	90%	20.14/0.7737	19.45/0.7967	20.98/0.8738	22.18/0.8397	22.19/0.8478	**27.81/0.9615**
Average		21.81/0.6041	23.58/0.7203	25.57/0.8058	26.37/0.8410	26.86/0.8457	**29.64/0.8814**

The best result in each case is highlighted in bold.

(a) Degraded (b) DIP (c) ST-NLTV (d) GSR (e) KSVD (f) BPFA (g) Proposed (h) Ground truth

Fig. 10. Comparison of color image restoration results of different methods with 80% of pixels missing. From top to bottom: Close-ups of *Burial*, *Phone Booth*, and *img_002*.

Table 4 lists PSNR/SSIM results of different methods on color images from Sun-Hays and Urban datasets with 80% and 90% of pixels missing. It is clear that the proposed method achieves the highest PSNR/SSIM in most of the cases. Compared with grayscale images, our image restoration 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 and 0.0357 in terms



(a) Degraded (b) DIP (c) ST-NLTV (d) GSR (e) KSVD (f) BPFA (g) Proposed (h) Ground truth

Fig. 11. Comparison of color image restoration results of different methods with 90% of pixels missing. From top to bottom: Close-ups of *img_003*, *img_004*, and *img_005*.

of PSNR and SSIM on average, respectively. Note that the highest PSNR and SSIM improvements over the runner-ups are 8.27 dB (over GSR on *img_005* under an 80% missing rate) and 0.1289 (over BPFA on *img_004* under a 90% missing rate), respectively.

Figure 10 and Figure 11 compare the visual quality of color image restoration results for test images (with 80% and 90% of pixels missing, respectively). Apparently, all comparing methods perform well on flat regions. However, DIP fails to recover clear edges, whereas ST-NLTV cannot restore fine details. KSVD and BPFA are better on recovering details, but they generate noticeable artifacts around edges. GSR produces sharper edges, but its results are somewhat blurred. The results of our method are of the best visual quality, especially under a higher missing rate.

4.4 Text Removal

Text removal is one of the classic cases 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 several state-of-the-art algorithms: the Content-Aware Fill feature in Photoshop CS6 [1], FoE [41], JSM [51], BPFA [56], and DIP [46]. Our experimental settings of text removal are the same with those in color image restoration. Table 5 shows the PSNR and SSIM results of different methods. Figure 12 presents visual comparison of different approaches, which further illustrates the effectiveness of our method.

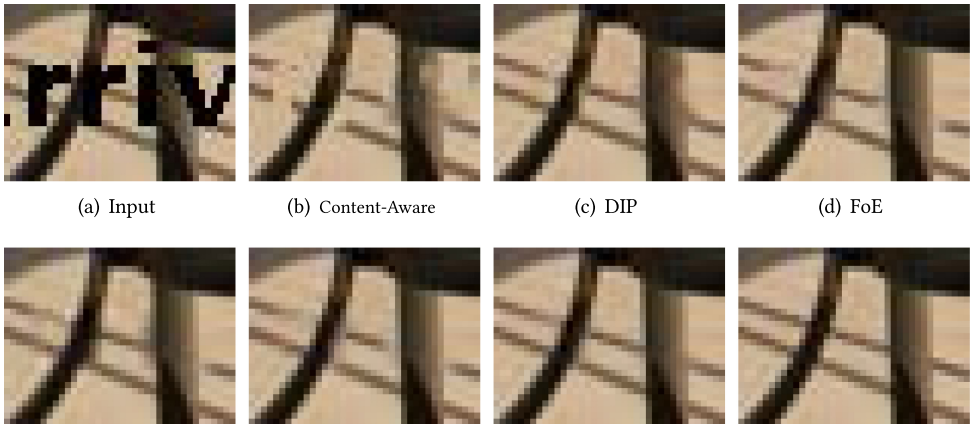
4.5 Image Interpolation

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 limited 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 limited 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 existing interpolation methods. The compared methods include AR model–based interpolation algorithms NEDI [28] and SAI [53], as well as a directional cubic convolution interpolation DCC [55]. Objective results are given in Table 6 and subjective comparisons are demonstrated in Figure 13, which reveal the competitiveness of our method in image interpolation.

Table 5. PSNR (dB) and SSIM Results of Text Removal from Different Methods

Images	Content-Aware	FoE	BPFA	JSM	DIP	Proposed
Aqueduct	32.14/0.9307	37.21/0.9653	35.84/0.9524	36.87/0.9636	35.41/0.9497	**37.98/0.9678**
Badlands	28.37/0.9224	31.92/0.9606	31.76/0.9515	32.19/0.9608	30.57/0.9423	**33.33/0.9682**
Barn	28.51/0.9254	32.38/0.9659	31.82/0.9551	34.01/0.9683	30.26/0.9336	**34.36/0.9712**
Balcony	27.50/0.9305	34.23/0.9743	33.56/0.9633	35.78/0.9799	33.99/0.9685	**36.79/0.9835**
Average	29.13/0.9273	33.93/0.9665	33.25/0.9556	34.71/0.9681	32.56/0.9485	**35.61/0.9727**

The best result in each case is highlighted in bold.



(a) Input (b) Content-Aware (c) DIP (d) FoE

(e) BPFA (f) JSM (g) Proposed (h) Ground truth

Fig. 12. Visual quality comparison of text removal for the *Balcony* image.

Table 6. PSNR (dB) and SSIM Results of Interpolation from Different Methods

Images	NEDI	SAI	DCC	Proposed
img_017	25.00/0.8764	**25.25/0.8847**	25.10/0.8821	24.64/0.8745
img_022	27.22/0.8539	**27.55**/0.8641	27.39/0.8621	27.48/**0.8655**
img_024	18.57/0.6860	18.54/0.6926	19.27/0.7110	**19.64/0.7284**
img_032	27.89/0.8714	28.38/0.8787	28.47/0.8786	**28.65/0.9004**
Average	24.67/0.8219	24.93/0.8300	25.06/0.8334	**25.10/0.8422**

The best result in each case is highlighted in bold.

4.6 Removal of Salt and Pepper Noise

Salt and pepper noise (S&P noise), which corrupts images/videos with either maximum or minimum values, can be removed by the proposed method. The mask for an S&P noise degraded image can be automatically generated by regarding pixels with maximum or minimum values (i.e., 0 or 255) as 0 and others as 1. Intuitively, the quality of the mask is affected by the number of pixels with values of 0 or 255 in the uncorrupted image. The more pixels with values of 0 or 255, the more inaccurate the generated mask is. We have already demonstrated the effectiveness of our method given accurate masks. Therefore, for the input of our method, we choose images with lots of pixels of values 0 and 255 (in one of the test images, the percentage of such pixels is more than 25%) and randomly add 80% to 90% S&P noise on them. Figure 14 gives some denoising results. State-of-the-art methods Noise Adaptive Fuzzy Switching Median filter (NAFSM) [45] and Adaptive Weighted Mean Filter (AWMF) [52] are compared in the experiment.

(a) Bicubic (b) NEDI (c) SAI

(d) DCC (e) Proposed (f) Ground truth

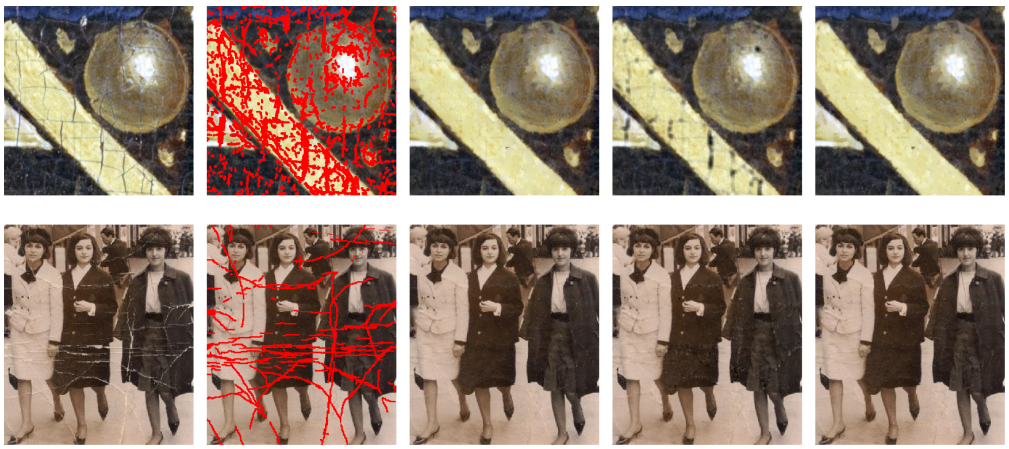
Fig. 13. Subjective comparison of interpolation for *img_017* from the Urban dataset.

(a) Noisy (b) NAFSM (c) AWMF (d) Proposed (e) Ground truth

Fig. 14. Comparisons of S&P noise removal for images from BSDS300. The noise density is 80% in the first row and 90% in the second row.

4.7 Other Applications

Our method can also be applied to other applications, such as inpainting old photos and restoring artworks virtually. Figure 15 shows some examples. The first row is the restoration of a small part of the jewels from the God the Father panel in the Ghent Altarpiece. The mask is generated by crack detection [42]. The second row is an old photo, and its map is user specified. From the figure, it can be observed that our method successfully repair the cracks without introducing many artifacts.



(a) Input (b) Mask (c) GSR (d) BPFA (e) Proposed

Fig. 15. Results for virtual restoration of artwork and an old photo.

5 CONCLUSION

In this work, we introduce the new concept of the multiplanar model, which exploits the cross-dimensional correlation in similar patches collected in images/videos. Moreover, a joint multiplanar AR and low-rank approach for video restoration from limited samples is presented, along with an alternating optimization algorithm. To ensure the temporal smoothness in restored videos, we also introduce a temporal constraint built on an MRF model. Extensive experimental results have demonstrated the effectiveness of our method on image/video restoration. Our method also generates comparable results when facing tasks as interpolation, text removal, and S&P noise removal.

REFERENCES

[1] Adobe Research. n.d. Content-Aware Fill. Retrieved October 23, 2019 from https://research.adobe.com/project/content-aware-fill/.

[2] Y. Boykov, O. Veksler, and R. Zabih. 2001. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 11 (Nov. 2001), 1222–1239. DOI: https://doi.org/10.1109/34.969114

[3] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. 2005. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation* 4, 2 (2005), 490–530.

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

[5] Emmanuel J. Candès and Benjamin Recht. 2009. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics* 9, 6 (2009), 717–772. DOI: https://doi.org/10.1007/s10208-009-9045-5

[6] Y.-L. Chen, C.-T. Hsu, and H.-Y. M. Liao. 2014. Simultaneous tensor decomposition and completion using factor priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 3 (March 2014), 577–591. DOI: https://doi.org/10.1109/TPAMI.2013.164

[7] G. Chierchia, N. Pustelnik, B. Pesquet-Popescu, and J.-C. Pesquet. 2014. A nonlocal structure tensor-based approach for multicomponent image recovery problems. *IEEE Transactions on Image Processing* 23, 12 (Dec. 2014), 5531–5544. DOI: https://doi.org/10.1109/TIP.2014.2364141

[8] K. Dabov, A. Foi, and K. Egiazarian. 2007. Video denoising by sparse 3D transform-domain collaborative filtering. In *Proceedings of the European Signal Processing Conference*. 145–149.

[9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. 2007. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing* 16, 8 (Aug. 2007), 2080–2095. DOI: https://doi.org/10.1109/TIP.2007.901238

[10] Weisheng Dong, Guangming Shi, and Xin Li. 2013. Nonlocal image restoration with bilateral variance estimation: A low-rank approach. *IEEE Transactions on Image Processing* 22, 2 (Feb. 2013), 700–711. DOI: https://doi.org/10.1109/TIP.2012.2221729

[11] W. Dong, L. Zhang, R. Lukac, and G. Shi. 2013. Sparse representation based image interpolation with nonlocal autoregressive modeling. *IEEE Transactions on Image Processing* 22, 4 (April 2013), 1382–1394. DOI: https://doi.org/10.1109/TIP.2012.2231086

[12] J.-J. Fadili, J.-L. Starck, and F. Murtagh. 2009. Inpainting and zooming using sparse representations. *Computer Journal* 52, 1 (2009), 64–79.

[13] W. B. Goh, M. N. Chong, S. Kalra, and D. Krishnan. 1996. Bi-directional 3D auto-regressive model approach to motion picture restoration. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '96)*, Vol. 4. 2275–2278. DOI: https://doi.org/10.1109/ICASSP.1996.545876

[14] Mohammad Golbabaee and Pierre Vandergheynst. 2012. Hyperspectral image compressed sensing via low-rank and joint-sparse matrix recovery. In *Proceedings of the 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'12)*. IEEE, Los Alamitos, CA, 2741–2744.

[15] Liangtian He and Yilun Wang. 2014. Iterative support detection-based split Bregman method for wavelet frame-based image inpainting. *IEEE Transactions on Image Processing* 23, 12 (Dec. 2014), 5470–5485. DOI: https://doi.org/10.1109/TIP.2014.2362051

[16] F. Heide, W. Heidrich, and G. Wetzstein. 2015. Fast and flexible convolutional sparse coding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15)*. 5135–5143. DOI: https://doi.org/10.1109/CVPR.2015.7299149

[17] Wenrui Hu, Dacheng Tao, Wensheng Zhang, Yuan Xie, and Yehui Yang. 2015. A new low-rank tensor model for video completion. arXiv:1509.02027.

[18] Yao Hu, Debing Zhang, Jieping Ye, Xuelong Li, and Xiaofei He. 2013. Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 9 (2013), 2117–2130. DOI: https://doi.org/10.1109/TPAMI.2012.271

[19] Yao Hu, Chen Zhao, Deng Cai, Xiaofei He, and Xuelong Li. 2016. Atom decomposition with adaptive basis selection strategy for matrix completion. *ACM Transactions on Multimedia Computing, Communications, and Applications* 12, 3 (June 2016), Article 43, 25 pages. DOI: https://doi.org/10.1145/2903716

[20] J. B. Huang, A. Singh, and N. Ahuja. 2015. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15)*. 5197–5206. DOI: https://doi.org/10.1109/CVPR.2015.7299156

[21] V. Jakhetiya, W. Lin, S. P. Jaiswal, S. C. Guntuku, and O. C. Au. 2017. Maximum a posterior and perceptually motivated reconstruction algorithm: A generic framework. *IEEE Transactions on Multimedia* 19, 1 (Jan. 2017), 93–106.

[22] Hui Ji, Sibin Huang, Zuowei Shen, and Yuhong Xu. 2011. Robust video restoration by joint sparse and low rank matrix approximation. *SIAM Journal on Imaging Sciences* 4, 4 (2011), 1122–1142.

[23] Hui Ji, Chaoqiang Liu, Zuowei Shen, and Yuhong Xu. 2010. Robust video denoising using low rank matrix completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'10)*. 1791–1798. DOI: https://doi.org/10.1109/CVPR.2010.5539849

[24] A. Kokaram and P. Rayner. 1994. Detection and interpolation of replacement noise in motion picture sequences using 3D autoregressive modelling. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS'94)*, Vol. 3. 21–24. DOI: https://doi.org/10.1109/ISCAS.1994.409091

[25] Chao Li, Lili Guo, and Andrzej Cichocki. 2014. Multi-tensor completion for estimating missing values in video data. In *Proceedings of the 2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS'14) and the 15th International Symposium on Advanced Intelligent Systems (ISIS'14)*. IEEE, Los Alamitos, CA, 1339–1342.

[26] Chao Li, Qibin Zhao, Junhua Li, Andrzej Cichocki, and Lili Guo. 2015. Multi-Tensor Completion with Common Structures. Retrieved October 23, 2019 from http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9437.

[27] Mading Li, Jiaying Liu, Zhiwei Xiong, Xiaoyan Sun, and Zongming Guo. 2016. MARLow: A joint multiplanar autoregressive and low-rank approach for image completion. In *Proceedings of the European Conference on Computer Vision (ECCV'16)*. 819–834. DOI: https://doi.org/10.1007/978-3-319-46478-7_50

[28] X. Li and M. T. Orchard. 2001. New edge-directed interpolation. *IEEE Transactions on Image Processing* 10, 10 (Oct. 2001), 1521–1527. DOI: https://doi.org/10.1109/83.951537

[29] R. Liao, X. Tao, R. Li, Z. Ma, and J. Jia. 2015. Video super-resolution via deep draft-ensemble learning. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV'15)*. 531–539. DOI: https://doi.org/10.1109/ICCV.2015.68

[30] Ji Liu, P. Musialski, P. Wonka, and J. Ye. 2009. Tensor completion for estimating missing values in visual data. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV'09)*. 2114–2121. DOI: https://doi.org/10.1109/ICCV.2009.5459463

[31] J. Liu, P. Musialski, P. Wonka, and J. Ye. 2013. Tensor completion for estimating missing values in visual data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 1 (2013), 208–220. DOI: https://doi.org/10.1109/TPAMI.2012.39

[32] Luca Lorenzi, Farid Melgani, and Grégoire Mercier. 2013. Missing-area reconstruction in multispectral images under a compressive sensing perspective. *IEEE Transactions on Geoscience and Remote Sensing* 51, 7 (2013), 3998–4008.

[33] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian. 2012. Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms. *IEEE Transactions on Image Processing* 21, 9 (Sept. 2012), 3952–3966. DOI: https://doi.org/10.1109/TIP.2012.2199324

[34] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi. 2013. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Transactions on Image Processing* 22, 1 (Jan. 2013), 119–133. DOI: https://doi.org/10.1109/TIP.2012.2210725

[35] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. 2009. Non-local sparse models for image restoration. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV'09)*. 2272–2279. DOI: https://doi.org/10.1109/ICCV.2009.5459452

[36] J. Mairal, M. Elad, and G. Sapiro. 2008. Sparse representation for color image restoration. *IEEE Transactions on Image Processing* 17, 1 (Jan. 2008), 53–69.

[37] D. Martin, C. Fowlkes, D. Tal, and J. Malik. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the International Conference on Computer Vision (ICCV'01)*, Vol. 2. 416–423.

[38] S. Ono, T. Miyata, and I. Yamada. 2014. Cartoon-texture image decomposition using blockwise low-rank texture characterization. *IEEE Transactions on Image Processing* 23, 3 (March 2014), 1128–1142. DOI: https://doi.org/10.1109/TIP.2014.2299067

[39] Holger Rauhut, Karin Schnass, and Pierre Vandergheynst. 2008. Compressed sensing and redundant dictionaries. *IEEE Transactions on Information Theory* 54, 5 (2008), 2210–2219.

[40] Justin Romberg. 2009. Compressive sensing by random convolution. *SIAM Journal on Imaging Sciences* 2, 4 (2009), 1098–1128.

[41] Stefan Roth and Michael J. Black. 2009. Fields of experts. *International Journal of Computer Vision* 82, 2 (2009), 205–229.

[42] Tijana Ružić, Bruno Cornelis, Ljiljana Platiša, Aleksandra Pižurica, Ann Dooms, Wilfried Philips, Maximiliaan Martens, Marc De Mey, and Ingrid Daubechies. 2011. Virtual restoration of the Ghent Altarpiece using crack detection and inpainting. In *Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems*. 417–428.

[43] Huanfeng Shen, Xinghua Li, Liangpei Zhang, Dacheng Tao, and Chao Zeng. 2014. Compressed sensing-based inpainting of aqua moderate resolution imaging spectroradiometer band 6 using adaptive spectrum-weighted sparse Bayesian dictionary learning. *IEEE Transactions on Geoscience and Remote Sensing* 52, 2 (2014), 894–906.

[44] L. Sun and J. Hays. 2012. Super-resolution from Internet-scale scene matching. In *Proceedings of the 2012 IEEE International Conference on Computational Photography (ICCP'12)*. 1–12. DOI: https://doi.org/10.1109/ICCPhot.2012.6215221

[45] Kenny Kal Vin Toh and Nor Ashidi Mat Isa. 2010. Noise adaptive fuzzy switching median filter for salt-and-pepper noise reduction. *IEEE Signal Processing Letters* 17, 3 (2010), 281–284.

[46] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. 2018. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18)*. 9446–9454.

[47] H. Wang, Y. Cen, Z. He, R. Zhao, Y. Cen, and F. Zhang. 2017. Robust generalized low-rank decomposition of multimatrices for image recovery. *IEEE Transactions on Multimedia* 19, 5 (May 2017), 969–983.

[48] Hua Wang, Feiping Nie, and Heng Huang. 2014. Low-Rank Tensor Completion with Spatio-Temporal Consistency. Retrieved October 23, 2019 from http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8580.

[49] Debing Zhang, Yao Hu, Jieping Ye, Xuelong Li, and Xiaofei He. 2012. Matrix completion by truncated nuclear norm regularization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'12)*. 2192–2199. DOI: https://doi.org/10.1109/CVPR.2012.6247927

[50] Jian Zhang, Debin Zhao, and Wen Gao. 2014. Group-based sparse representation for image restoration. *IEEE Transactions on Image Processing* 23, 8 (Aug. 2014), 3336–3351. DOI: https://doi.org/10.1109/TIP.2014.2323127

[51] Jian Zhang, Debin Zhao, Ruiqin Xiong, Siwei Ma, and Wen Gao. 2014. Image restoration using joint statistical modeling in a space-transform domain. *IEEE Transactions on Circuits and Systems for Video Technology* 24, 6 (June 2014), 915–928. DOI: https://doi.org/10.1109/TCSVT.2014.2302380

[52] Peixuan Zhang and Fang Li. 2014. A new adaptive weighted mean filter for removing salt-and-pepper noise. *IEEE Signal Processing Letters* 21, 10 (2014), 1280–1283.

[53] Xiangjun Zhang and Xiaolin Wu. 2008. Image interpolation by adaptive 2-D autoregressive modeling and soft-decision estimation. *IEEE Transactions on Image Processing* 17, 6 (June 2008), 887–896. DOI: https://doi.org/10.1109/TIP.2008.924279

[54] Z. Zhang, G. Ely, S. Aeron, N, Hao, and M. Kilmer. 2014. Novel methods for multilinear data completion and de-noising based on tensor-SVD. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'14)*. 3842–3849. DOI: https://doi.org/10.1109/CVPR.2014.485

[55] Dizhi Zhou, Xinyue Shen, and Wenjie Dong. 2012. Image zooming using directional cubic convolution interpolation. *IET Image Processing* 6, 6 (2012), 627–634.

[56] M. Zhou, H. Chen, J. Paisley, L. Ren, L. Li, Z. Xing, D. Dunson, G. Sapiro, and L. Carin. 2012. Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images. *IEEE Transactions on Image Processing* 21, 1 (Jan. 2012), 130–144. DOI: https://doi.org/10.1109/TIP.2011.2160072

Received August 2018; revised May 2019; accepted June 2019