Discriminative Structured Dictionary Learning on Grassmann Manifolds and Its Application on Image Restoration

Han Pan, Zhongliang Jing, Lingfeng Qiao, and Minzhe Li

***Abstract*—Image restoration is a difficult and challenging problem in various imaging applications. However, despite of the benefits of a single overcomplete dictionary, there are still several challenges for capturing the geometric structure of image of interest. To more accurately represent the local structures of the underlying signals, we propose a new problem formulation for sparse representation with block-orthogonal constraint. There are three contributions. First, a framework for discriminative structured dictionary learning is proposed, which leads to a smooth manifold structure and quotient search spaces. Second, an alternating minimization scheme is proposed after taking both the cost function and the constraints into account. This is achieved by iteratively alternating between updating the block structure of the dictionary defined on Grassmann manifold and sparsifying the dictionary atoms automatically. Third, Riemannian conjugate gradient is considered to track local subspaces efficiently with a convergence guarantee. Extensive experiments on various datasets demonstrate that the proposed method outperforms the state-of-the-art methods on the removal of mixed Gaussian-impulse noise.**

***Index Terms*—Grassmann manifold, image restoration (IR), sparse representation.**

I. Introduction

IMAGE restoration has been widely studied [1]–[4] in the field of computer vision. It aims to recover a clean image from the noise-corrupted observations. The restoration of visual structures can be useful for improving vision tasks, e.g., face hallucination [5], image super-resolution [6], gait recognition [7], [8], visual tracking [9], [10], and face sketch synthesis [11]. The noise on real images can be much more complex empirically. A variety of noise conditions lead to the mixture of multiple sources of noise, such as mixed Gaussian-impulse noise [12] and fixed Gaussian and Poisson noise [13], etc. These happen inevitably and have received less attention. The removal of mixed Gaussian-impulse noise is an important preprocess step in some applications, such as change detection and object recognition in remote sensing.

A variety of noise removal methods have been proposed in past decades [14]–[20]. One method to remove Gaussian noise is sparse representation-based methods [14], [15]. It has been shown that data adaptive dictionaries significantly outperform wavelets-based methods [21]. For the image corrupted by impulse noise, median filters-based methods [16]–[18] were used to detect and remove the irregular pixels. However, these methods may suffer from the loss of important details. Thus, a well-adapted framework is required to take all things into consideration.

In this paper, we focus on structured sparse model selection approach [15] (SSMS). For a fixed width $\sqrt{m}$, SSMS merges the image plane into a data matrix X ($m \times n$) with the patches by $\sqrt{m} \times \sqrt{m}$. And, n denotes the number of the patches. Then, X can be represented by

$$DC \simeq X \tag{1}$$

where D is a dictionary matrix with the size of $m \times k$, and C denotes the corresponding coefficients matrix. The main idea of this scheme is to represent matrix C by a predefined number of blocks. Each block corresponds to a fixed set of columns of D. Subdictionaries of D are selected for stabilizing the encoding procedure, which denoted by $[B_1, \ldots, B_K]$. For a fixed cluster coming from n columns of X, its main idea can be expressed as follows:

$$\begin{aligned} &\min_{D,C} \ \|X - DC\|_F^2 \\ &\text{s.t.} \ \ \|C_i\|_0 \le s, \ B_i^T B_i = I_d, \ 1 \le i \le K \end{aligned} \tag{2}$$

where D is assumed to be composed of K blocks of estimated bases $[B_1, \ldots, B_K]$, corresponding to the columns of X associated to the clusters ω. I_d denotes the identity matrix. And, s stands for the bound of sparsity level.

The success of SSMS depends on the modeling capability of local basis B_i. However, it is not easy to update the block structure of learned dictionary locally because of the computational challenge introduced by $B_i^T B_i = I_d$. To alleviate this problem, we consider some recent progress [22] in the theory of optimization on manifold. The relevant constraints are identified as a smooth manifold structure. The arising problem can be handled by Riemannian optimization framework. This optimization strategy calculates the specific representation error

Manuscript received February 24, 2017; revised July 2, 2017 and August 25, 2017; accepted September 2, 2017. Date of publication September 26, 2017; date of current version September 14, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61603249 and Grant 61673262, and in part by the Key Project of Science and Technology Commission of Shanghai Municipality under Grant 16JC1401100. This paper was recommended by Associate Editor Y. Xia. *(Corresponding author: Han Pan.)*

The authors are with the School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: hanpan@sjtu.edu.cn; zljing@sjtu.edu.cn; qiaolf927@sina.cn; lmznuaa@126.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2017.2751585

by tracking local subspaces. In the context of the directional bases, the improvement of the representation framework leads to the construction of a new dictionary adapted to the image of interest. Then, the most prominent local image content can be approximated.

A. Related Work

There have been a number of studies for the removal of mixed Gaussian-impulse noise recently. These methods can be categorized into three classes, including variational methods [12], [23], nonlocal means filtering methods [24], [25], and sparse representation-based methods [26], [27].

Variational methods have been proposed for restoring sharp discontinuities or texture [12], [23] because of its structure-encouraging norm on the gradient of image. Rodríguez *et al.* [12] presented a composite objective function with total variation regularization term, ℓ_2 and ℓ_1 data fidelity terms, which is solved by an iterative reweighted norm approach. Mumford–Shah functional [23] also was utilized to reconstruct sharp edges. This method was divided into two consecutive stages: 1) identifying the outliers and 2) denoising the resulting image. Then, the resulting problem is solved by preconditioned conjugate gradient method. It can be seen that these methods contain a common procedure for finding the location of impulse noise. However, variational methods may suffer from staircasing effects. In addition, the choice of the regularization parameter is an interesting and complex topic. Xia and Kamel [28] proposed a novel parameter selection method with cooperative neural fusion regularization method.

Among these algorithms, nonlocal means filtering method and its variants are viewed as a state-of-the-art scheme for the removal of Gaussian noise [24], [25]. Its main idea is to utilize the natural redundancy of patterns inside an image. In other words, an internal self-similarity prior is exploited. This has triggered an increasing research interest in the design of restoration methods. Typically, block-matching 3-D (BM3D)-based method [24] and its video case [25] have been proved efficient. The main idea of BM3D is to group the nonlocal similar patches into a 3-D cube and perform a shrinkage operation on the transform coefficients. Inevitably, nonlocal means filtering methods have some limitations. The main one is the oversmooth of low-contrasted areas.

Sparse representation has been viewed as a powerful tool for representing high-dimensional signals. This scheme [29]–[31] is a patch-based approach by constructing over-complete dictionaries and implementing sparse coding. Xiao *et al.* [26] reconstructed the latent image by solving ℓ_1-ℓ_0 constraint-based problem, which turns out to be advantageous in the removal of mixed Gaussian-impulse noise. This method involved three steps, including outliers detection [16] by rank-order absolute difference detector, initially approximating [30] by K-singular value decomposition (K-SVD) and enhancing the latent image by minimizing the resulting energy function. Recently, a mixed soft-hard thresholding-based sparse representation method [27] was presented. It can be seen that the construction of over-completed dictionary is important to the success of these methods. Among them, the update of learned dictionary of these methods, however, tends to ignore high-level structure.

A drawback of the current sparse representation-based methods is that they are restricted to a special kind of image signals. For natural images with rich structures, block-sparse representation-based methods may not provide sufficient representation of local image content across blocks. The restoring methods should address these variations adaptively. Meanwhile, natural image signals have some local structures that go beyond a structured dictionary block-to-block.

B. Contributions

The motivation of this paper is that natural signals cannot be represented by common structured dictionary learning robustly. The main idea in this paper, however, is to learn a discriminative structured dictionary with block and orthogonal constraints simultaneously, i.e., updating independent orthobases block-to-block. Specially, we show this by considering the structures of the cost function and the constraints. Then, the effectiveness of the proposed method is shown in the removal of mixed Gaussian-impulse noise.

Based on the Riemannian optimization framework, we define some smooth cost functions over Grassmann manifold. These smooth functions have an advantage in dealing with ill-conditioned problems. An efficient alternating minimization scheme for block-orthogonal constraint is presented by considering its smooth manifold structure. A Riemannian conjugate gradient method is adopted. Then, a union of locally texture-specific subspaces is constructed for describing common features of natural signals.

The major contributions of this paper can be summarized as follows.

1) A framework for discriminative structured dictionary learning is proposed. This is based on a new problem formulation with block-orthogonal constraint.
2) We propose an alternating minimization scheme that can provide a solution to the resulting optimization problem. This is accomplished by updating the block structure of the dictionary defined on Grassmann manifolds and sparsifying the dictionary atoms automatically.
3) We derive a result that tracks local subspaces based on Riemannian conjugate gradient.

C. Paper Structures

The rest of this paper is organized as follows. Section II presents a new problem formulation for discriminative structured dictionary learning. Section III describes some ingredients about the optimization methods on manifold. Section IV presents an alternating minimization scheme for the resulting problem. Section V summarizes the experimental results. Then, detailed discussion is completed. At last, a comprehensive conclusion is presented in Section VI.

II. Problem Formulations

For the reasons specified in the introduction, a new problem formulation of discriminative structured dictionary learning for finding K-subspaces is presented, which defined on Grassmann

manifold. The proposed scheme can handle the corrupted data effectively. Equation (2) is reformulated as follows:

$$\min_{B_k,C_i,\omega_k} \sum_{k=1}^{K}\left\{\sum_{i\in\omega_k}\|X_i - B_kC_i\|_F^2\right\} \quad \text{s.t. } \|C_i\|_0 \le s_0,\ B_k^TB_k = I_d,\ \{\omega_k\}\in\Omega \tag{3}$$

where $X_i \in \mathbb{R}^{m\times n_k}$ stands for the submatrix of X by n_k data points in the cluster $\omega(j) = k$. And, the set Ω denotes the mapping from data points $(1,\ldots,n)$ to a union of subspaces $(1,\ldots,K)$. Specially, the set $\{\omega_k\}_{k=1}^K$ stands for a clustering of the columns of $\{X_i\}_1^n$. And, the cluster ω_k contains the indexes i corresponding to X_i in the kth cluster. The set Ω can be expressed as follows:

$$\Omega = \left\{\omega_k : \bigcup_{k=1}^{K}\omega_k = [1:n], \omega_j\bigcap\omega_k = \varnothing,\ \forall j\neq k\right\}. \tag{4}$$

The definition of Ω is used to enforce ω_k to be disjoint. Then, a union of local subspaces model B_k is associated to data matrix X_k.

The main challenge of (3) is block-orthogonal constraint because it is not easy to preserve iterate-to-iterate. Specially, a central problem in this context is nonlinear and high-dimensional matrix search spaces. The resulting problem is required to preserve the underlying geometric structure of the problem. This spurs a retraction-based optimization framework that aims to solve (3) on matrix manifold. In this paper, recent advances [22], [32], [33] in optimization algorithms on manifold are considered.

Another challenge of (3) is that it is nondifferentiable while keeping C_i and ω_k fixed. To handle this problem, some smooth techniques are adopted to make it differentiable. There are many formulations of smooth penalty functions. It has been verified that smooth functions [34] are better than ℓ_1 norm. Some nonconvex but smooth objective functions would be solved using a Riemannian conjugate gradient-like minimization process, which can be expressed as follows:

$$h_\mu^{lp} : \mathbb{R}^{m\times d}\to\mathbb{R}^+,\ X\to\sum_{j=1}^{d}\sum_{i=1}^{m}\left(x_{ij}^2+\mu\right)^{\frac{p}{2}},\ 0<p<1 \tag{5}$$

$$h_\mu^{\text{atan}} : \mathbb{R}^{m\times d}\to\mathbb{R}^+,\ X\to\sum_{j=1}^{d}\sum_{i=1}^{m}\text{atan}^2\left(\frac{x_{ij}}{\mu}\right) \tag{6}$$

where $\mu > 0$ is a smoothing parameter.

To handle these difficulties, the geometry of block-orthogonal constraint is exploited. Meanwhile, B and C cannot be determined uniquely [15] via SVD. It should be noted that B is an element of the so-called Stiefel manifold $\text{St}_{d,m} = \{B\in R^{m\times d}|B^TB = I_d\}$. We reformulate the problem (3) into a problem defined on quotient manifold, which also known as Grassmann manifold $G_k \in \text{Gr}_{d,m}$.

In summary, a new problem formulation for discriminative structured dictionary learning by optimizing on the set of symmetric projectors is proposed, which can be expressed as follows:

$$F(G,C,\omega) = \min_{G_k,C_i,\omega_k}\sum_{k=1}^{K}\left\{\sum_{i\in\omega_k}h(X_i - G_kC_i)\right\} \quad \text{s.t. } \{\omega_k\}\in\Omega \tag{7}$$

where $G_k = B_kB_k^T$.

In contrast to other single overcomplete dictionary-based methods [26], [27], [35], the integration of optimization method on Grassmann manifold and smooth function provides a natural way to explore low dimensional structures of learned dictionary, which approximates the significant information of image efficiently. Moreover, the local structures of over-completed dictionary are utilized. The proposed method can learn image features via optimizing on Grassmann manifold. The predefined dictionary is loaded when the denoising process starts.

Remark 1: Compared to other formulations [36]–[38], the proposed method is different in some aspects. First, Zelnik-Manor *et al.* [36] reformulated the problem of dictionary learning problem via introducing the constraint of block sparse. However, this method did not develop an efficient method to update the local subspaces. Second, Bao *et al.* [37] solved the dictionary learning problem with block-orthogonal constraint. But, the nonuniqueness of SVD may lead to the unstable solutions. Third, these methods were just available for Gaussian noise.

Remark 2: In [39] and [40], they derived various kernelized versions of sparse coding and dictionary learning algorithms. The resulting problems were solved via optimizing on Grassmann manifold, which exploit the geometry of symmetric matrices after an isometric mapping. Their methods have been applied to face recognition, action recognition, etc.

III. Ingredients for Optimization on Manifold

To overcome some difficulties of high-dimensional data by block-orthogonality constraints, we provide some ingredients about optimization methods [22] on matrix manifold. These concepts can be viewed as a generalization representations of optimization methods in Euclidean space.

Definition 1 (Retraction): A retraction is a smooth mapping from the tangent bundle $T\mathcal{M}$ onto manifold $\mathcal{M}$ at a point $x\in\mathcal{M}$

$$R : T\mathcal{M}_x\to\mathcal{M}_x$$

which satisfied the following.

1) $R_x(0) = x$, where 0 is the zero element of $T_x\mathcal{M}$.
2) $R(x,0) = x$ for all $x\in\mathcal{M}$.
3) $(d/dt)R(x,t\dot{x})|_{t=0} = \dot{X}$ and $\dot{x}\in T\mathcal{M}$.

It should be noted that retraction locally approximates the geodesics and provides an attractive alternative to the geodesics in the design of optimization algorithms on manifold. Moreover, it provides a natural way to iterate on manifold, and keeps the key properties of convergence.

The rich geometry of Riemannian manifolds allows us to define a flexible projector, i.e., retraction, to move on the manifold along a specified tangent direction at a point. Retraction

is less compensative than logistic mapping. Then, it is taken for pulling back the search direction from the tangent space to manifold.

In the context of optimization methods on manifold, vectors from different tangent spaces cannot be compared directly. Then, vector transport is defined to transport tangent vectors from tangent space to another manifold. It can be viewed as the first-order approximation of parallel transport. More insights about this paper can be referred to [22].

Definition 2 (Vector Transport): A transport on a manifold $\mathcal{M}$ is a smooth mapping

$$\text{Transp} : T\mathcal{M} \otimes T\mathcal{M} \to T\mathcal{M} : (\eta, \xi) \to \text{Transp}_\eta(\xi) \tag{8}$$

where $\xi \in T_x\mathcal{M}$ from $x \in \mathcal{M}$ to a point $R_x(\eta) \in \mathcal{M}$, $\eta \in T_x\mathcal{M}$. Meanwhile, it satisfies the following properties for all $x \in \mathcal{M}$.

1) *Associated Retraction:* There exists a retraction R, called the retraction associated with Transp, such that $\text{Transp}_\eta(\xi) \in T_{R_x(\eta)}\mathcal{M}$.
2) *Consistency:* $\text{Transp}_0(\xi) = \xi$ for all $\xi \in T_x\mathcal{M}$.
3) *Linearity:* $\text{Transp}_\eta(a\xi + b\theta) = a\text{Transp}_\eta(\xi) + b\text{Transp}_\eta(\theta),\ \forall \eta, \xi, \theta \in T_x\mathcal{M},\ a, b \in \mathbb{R}$.

It can be seen that the mapping $\text{Transp}_{y \leftarrow x}(\xi) : \text{T}_x\mathcal{M} \to \text{T}_y\mathcal{M}$ is a vector transport, which depends smoothly on x and y. Meanwhile, it is linear in ξ. And, $\text{Transp}_{y \leftarrow x}$ is the identity map.

The recent advances on the theory of optimization methods on manifold [22] lead to efficient numerical methods on special manifold. In this paper, the combination of retraction and vector transport would provide some basic tools to implement the framework of Riemannian conjugate gradient-based method.

IV. Proposed Method

A. Alternating Minimization Scheme

In this section, a numerical solution to the problem (7) is provided. The proposed method contains three steps: 1) dictionary update; 2) sparse coding; and 3) clusters update. Then, an alternating minimization scheme is presented in Algorithm 1. At last, we would deal with this problem by performing alternating minimization procedure until convergence.

1) Dictionary update G_k via solving

$$\min_{G_k \in \text{Gr}_{d,m}} \sum_{k=1}^{K} \left\{ \sum_{i \in \omega_k} h_\mu(X_i - G_k C_i) \right\}. \tag{9}$$

Given X_i, C_i, and ω_k, the problem (9) is unconstrained optimization problem over Grassmann manifold G_k. It should be noted that the reformulation of (3) defined on nonlinear matrix search spaces appeals for a proper optimization method. Moreover, the resulting problem is separable into K unconstrained subproblems after embedding the constraints into the search spaces. In this paper, we take both the cost function and the constrained search spaces into consideration. Thus, the optimization methods on matrix manifold are adopted. This is a core algorithm in our method. And, this motivates a smooth manifold structure. In particular, a Riemannian conjugate gradient method is implemented for (9). This requires some additional concepts from differential geometry, such as retraction and vector transport [22]. In particular, the details of the minimization process are described in Algorithm 2.

2) Sparse coding C_i via solving

$$\min_{C_i} \sum_{k=1}^{K} \left\{ \sum_{i \in \omega_k} h_\mu(X_i - G_k C_i) \right\}. \tag{10}$$

Given X_i, G_i, and ω_k, (10) is presented for sparse coding. Different from K-SVD-based method [30], some smooth functions are taken for promoting sparsity adaptively. It has been verified that smooth functions described in (5) and (6) are better than ℓ_1-norm. In this paper, this subproblem is solved via conjugate gradient (CG) method in Euclidean space, which is well-established.

3) Update the cluster ω_k via solving the following problem:

$$\omega(j) = \arg \min_{1 \le k \le K} \min_{c \in \mathbb{R}^d} \left(\left\| B_i c - x_j \right\|_2^2 + \lambda \phi_i^2 \right) \tag{11}$$

where λ denotes the regularization parameter, x_j denotes the column of the data matrix X. And, ϕ is considered for stabilizing the solution. Then, $\omega(j)$ ensures a unique clustered index after the minimization process. However, it should be noted that this is a modified version of (7), which contains an additional regularizer. After the update of learned dictionary D, the most appropriate subspace for each representative point $x_{1 \le j \le n} \in \mathbb{R}^m$ can be obtained. Furthermore, kth cluster of underlying subspace from the rows of B_k is completed. Then, the best subspace estimation corresponding to x_j is indexed.

It should be noted that the problems in (9) and (10) have closed-form solutions. In other words, we can obtain their respective gradients. However, they differ in the domains. Meanwhile, there are no coupling among these three steps. Then, in the setting of manifold, some theoretical results about (9) are provided.

B. Riemannian Conjugate Gradient

In this section, Riemannian conjugate gradient is considered for solving the subproblem (9). And, the combination of the operations of retraction and vector transport gives all the necessary ingredients for optimizing this cost function on Riemannian quotient manifold. There are three basic steps to compute Riemannian conjugate gradient.

1) Computing the Riemannian gradient $\text{grad}\, h_\mu(B)$ defined on Grassmann manifold. After Euclidean gradient $\nabla h_\mu(B)$ for (5) is obtained, an orthogonal projection [41] is provided to project it onto the tangent space $T_x\mathcal{M}$. For $Z = U\Sigma V$, it can be illustrated as follows:

$$\begin{aligned} P_{T_x\mathcal{M}} : \mathbb{R}^{m \times n} \to T_x\mathcal{M} : Z \to P_U Z P_V \\ + P_U^\perp Z P_V + P_U Z P_V^\perp \end{aligned} \tag{12}$$

Algorithm 1 Alternating Minimization Scheme for (7)

Input: $X_0 \in \mathbb{R}^{m\times n}$, K: number of subspaces, d: dimension of subspaces, number of iterations N_1 and N_2, smoothing parameter μ, decaying factor c_μ and regularization parameter $\lambda = 10^{-8}$.
Output: B: learned dictionary, C: learned sparse codes, ω: learned clusters.
for t=1:N_1 **do**
  for i=1:N_2 **do**
    $B = \arg\min_B G(B, C, \omega)$. # see **Algorithm 2**
    $C = \arg\min_C G(B, C, \omega)$. # referred to Eq.(10)
    $\mu^{i+1} = c_\mu \mu^i$
  end for
  for j=1:n **do**
    If $j = 1$, $\omega_k = \varnothing \quad \forall i$
    $j =$ Eq.(11), $1 \le k \le K$.
    $\hat{\omega}_k = \hat{\omega}_k \cup \{j\}$.
  end for
end for

where $P_U = UU^T$, $P_U^\perp = I - UU^T$. And, P_V and $P_V^\perp$ stand for the same definition on V, respectively. This projection provides an alternative way to obtain the tangent space in the context of Grassmann manifold. We obtain the Riemannian gradient

$$\operatorname{grad} h_\mu(B) = P_{T_x\mathcal{M}}\big(\nabla h_\mu(B)\big). \tag{13}$$

Then, a retraction is defined for mapping the tangent space to $\mathrm{Gr}_{d,m}$, which can be expressed as follows:

$$R_B(H) = \mathrm{qf}(B + H) \tag{14}$$

where $\mathrm{qf}(\cdot)$ stands for QR decomposition, which is Gram–Schmidt orthogonalization of columns of A, started from the first column, $B \in \mathcal{M}$, and $H \in T_B\mathcal{M}$.

2) Computing conjugate search direction $d_i \in T_x\mathcal{M}$. Then, a linear combination of the Riemannian gradient with previous search direction d_{i-1} is required to determine. However, d_{i-1} does not lie in the same subspace with $T_x\mathcal{M}$, it is required to be transported to $T_x\mathcal{M}$. In this paper, a simple procedure to transport a tangent vector H at U to the tangent space at V is defined as follows:

$$\mathrm{Transp}_{V\leftarrow U}(H) = \big(I - VV^T\big)H. \tag{15}$$

At last, Hestenes–Stiefel+ formula is taken for updating the new search direction.

3) Computing the subsequent iterate via backtracking line-search. However, a good step-size α has a significant effect on the performance of a nonlinear CG algorithm.

The nonlinear CG method on Grassmann manifold can be outlined as follows. Starting at an initial point $x_0 \in \mathrm{Gr}_{d,m}$, the Riemannian gradient g_0 can be computed. And $d_0 = -g_0$ is selected as initial search direction. In each iteration, suitable step-size α_0 is determined using a backtracking line-search algorithm described in Algorithm 3. The new iterate can be obtained by

$$x_{k+1} = R_{x_k}(\alpha_k d_k). \tag{16}$$

Algorithm 2 Minimizing (9) via Riemannian Conjugate Gradient Method

Require: Given $x_0 \in \mathcal{M}$, $g_0 = \operatorname{grad} h_\mu(x_0)$, $d_0 = -g_0$, $k = 0$.
repeat
  if $\langle g_k, d_k\rangle \ge 0$ **then**
    $d_k = -g_k$
  end if
  $\alpha_k =$ Backtracking(x_k, d_k, g_k) # see **Algorithm 3**
  $x_{k+1} = \mathrm{R}_{x_k}(\alpha_k d_k)$
  $g_{k+1} = \operatorname{grad} h_\mu(x_{k+1})$
  $d_k^+ = \mathrm{Transp}_{x_{k+1}\leftarrow x_k}(d_k)$
  $g_k^+ = \mathrm{Transp}_{x_{k+1}\leftarrow x_k}(g_k)$
  $\beta_k = \max(0, \beta_k^{HS})$
  # update the search direction
  $d_{k+1} = -g_{k+1} + \beta_k d_k^+$
  $k = k + 1$
until Satisfying a stopping criterion

Algorithm 3 Backtracking Line Search

Require: $\alpha_{ini} > 0$, $c, \beta \in (0, 1)$ and $\alpha = \alpha_{ini}$.
repeat
  $\alpha = \beta\alpha$
  # $\mathrm{tr}(\cdot)$ denotes the operation of the trace.
until $h_\mu(\mathrm{R}_x(\alpha d_k)) \le h_\mu(x_k) + c\alpha\mathrm{tr}(g_k^T d_k)$
return step-size $\alpha_k = \alpha$

Finally, the search direction based on vector transport is updated by

$$d_{k+1} = -g_{k+1} + \beta_k \mathrm{Transp}(d_k). \tag{17}$$

Here, we consider two update rules for β_k, namely

$$\beta_k^{FR} = \frac{\langle g_{k+1}, g_{k+1}\rangle}{\langle g_k, g_k\rangle},\ \beta_k^{HS} = \frac{\langle g_{k+1}, g_{k+1} - g_k^+\rangle}{\langle d_{k+1}^+, g_{k+1} - g_k^+\rangle} \tag{18}$$

where $d_k^+ = \mathrm{Transp}_{x_{k+1}\leftarrow x_k}(d_k)$ and $g_k^+ = \mathrm{Transp}_{x_{k+1}\leftarrow x_k}(g_k)$. The former is the well-known Fletcher–Reeves update formula and guarantees convergence of our algorithm. The latter is an adaptation of Hestenes–Stiefel+ formula.

Remark 3: There exist some algorithms for the problems defined on manifolds [42]. The main difference is the formulation of the approximation of Riemannian gradient, e.g., Riemannian metric and connection [22].

C. Convergence Analysis

In this section, theoretical proof of convergence of the proposed method is provided, which involved in the convergence theory of Riemannian optimization framework described in [22] and [41]. The proposed method explores the simplicity of the underlying block-structure of SSMS. It should be noted that the convergence of Algorithm 2 can be guaranteed. Then, a critical point can be calculated. However, the alternating minimization scheme, presented in Algorithm 1, is not guaranteed to be monotonic.

Proposition 1: Let x_i be an infinite sequence of iterates generated by Algorithm 2 with the smooth function h_μ with $0 < \gamma < 1$. Then, we have

$$\lim_{i\to\infty} P_{T_{x_i}\mathcal{M}} \nabla h_\mu(x_i) = 2\gamma^2\big(x_i^* + x_i\big).$$

Proof: First, it can be shown that the iterates stay in a closed and bounded subset of $\mathcal{M}$. Let $L = \{x \in \mathcal{M}\} : h_\mu(x) \le h_\mu(x_0)$ be the level set at x_0. After the application of line search, the sequence of x_i lies in the domain of L. Then, we get

$$h_\mu(x_i) + \gamma^2 \big\|x_i^*\big\|_F^2 + \gamma^2 \|x_i\|_F^2 \le C_0^2,\ i > 0$$

where $h_\mu(x_0) = C_0^2$. This indicates that

$$\gamma^2 \|x_i\|_F^2 \le C_0^2 - h_\mu(x_i) - \gamma^2 \big\|x_i^*\big\|_F^2 \le C_0^2.$$

Then, an upper bound based on the largest singular value can be obtained as follows:

$$\sigma_1(x_i) \le \|x_i\|_F \le C_0/\gamma = C^\sigma.$$

Similarly, a lower bound based on the smallest singular value can be estimated as follows:

$$\gamma^2 \|x_*\|_F^2 = \sum_{j=1}^{k} \frac{\gamma^2}{\sigma_j^2(x_i)} \le C_0^2 - h_\mu(x_i) - \gamma^2 \|x_i\|_F^2 \le C_0^2$$

which indicates that

$$\sigma_k(x_i) \ge \gamma/C_0 = C_0.$$

At last, it can be seen that all x_i lie in the set

$$S = \big\{x \in \mathcal{M} : \sigma_1(x) \le C^\sigma, \sigma_k(x) \ge C_\sigma\big\}.$$

This indicates the set is bounded, and compact.

Now, suppose that $\lim_{i\to\infty} \|\operatorname{grad} h_\mu(x_i)\|_F \ne 0$. Then, there is a subsequence in $\{x_i\}_i$. Thus, $\|\operatorname{grad} h_\mu(x_i)\|_F \ge \epsilon > 0, \forall i$. Because x_i has a limit point x_* in the set of S. Based on the property of the continuity of $\operatorname{grad} h_\mu$, this indicates that $\|\operatorname{grad} h_\mu(x_*)\|_F \ge \epsilon$ contradicts in [22, Th. 4.3.1] that every accumulation point is critical point of h_μ. Hence, $\lim_{i\to\infty} \|\operatorname{grad} h_\mu(x_i)\|_F = 0$. ■

The convergence rate of the proposed method is an interesting and complex topic. Some existing works have considered the convergence behavior of Riemannian conjugate gradient methods [22], [41]. It should be noted that our method inherits the convergence guarantees of Riemannian conjugate gradient method in [22]. However, their methods do not discuss the convergence rate. Thus, we studied it experimentally. A detailed illustrative example is provided in Section V-B. Its empirical behavior is presented in Fig. 2. Theoretical analysis on convergence rate is left for future work.

D. Computational Complexity

The proposed method mainly contains three parts: 1) dictionary update; 2) sparse coding; and 3) clusters update. Each iteration at time t computes the values of a point (G_k, C_i, ω_k). The corresponding domain for this point is$\Theta = (\mathbb{R}^{m\times d})^K \times (\mathbb{R}^{d\times n})^K \times \{1, \ldots, K\}^n$. The computational complexity of the proposed method is presented as follows.

1) For the problem of dictionary update in (9), it can be seen that each cluster updates the local basis using Algorithm 2. The complexity of calculating Euclidean gradient for (9) is $\mathcal{O}(mnd)$, which contains multiply add operations. In the step 7 of Algorithm 2, the cost of computing Riemannian gradient by (12) is $\mathcal{O}((2m+n)d^2)$. The complexity of retraction in (14) is $\mathcal{O}((m+n)d^2)$. The cost of backtracking line search in Algorithm 3 is $\mathcal{O}((m+n)d^2)$ because of the introduction of retraction. Similar, the complexity of vector transport in (15) is $\mathcal{O}((m+n)d^2)$. Summed up, the complexity of dictionary update in Algorithm 2 is $\mathcal{O}\{\max(mnd, (2m+n)d^2\}$.
2) Given G_k and ω_k, we can see that the complexity of sparse coding in (10) is $\mathcal{O}(mnd)$. This subproblem is solved by traditional conjugate gradient methods. Thus, the complexity of this procedure is dominated by the calculation of Euclidean gradient with respect to C_i.
3) In (11), The complexity of performing clusters update ω_k for each column is $\mathcal{O}(md^2)$. It can be seen that the complexity of the computation of $B_i c$ is $\mathcal{O}(md)$. The complexity of computing $B_i^T B_i$ is $\mathcal{O}(md^2)$. Finally, to obtain the best cluster for each requires $\mathcal{O}(K-1)$ operations.

In summary, the proposed method for solving problem (3) have a complexity of $\mathcal{O}\{\max(mnd, (2m+n)d^2)\}$ per iteration. It should be noted that the complexity of our method depends on the dimensions of the problem. There are two cases which are relevant to the costs of the proposed method. For simplify, we give formal expressions of these cases as follows.

1) When $d < m < n$, the complexity of the proposed method is $\mathcal{O}((2m+n)d^2)$.
2) When $d < m \ll n$, the complexity of the proposed method is $\mathcal{O}(mnd)$.

E. Implementation Details

The proposed method is directly applied to reduce mixed Gaussian-impulse noise by exploiting a union of local subspaces. Similar to previous work [29], overlapping patches are implemented. Some strategies described in [15] are adopted to implement the denoising procedure. The subspaces are initialized by low frequency discrete cosine transform basis. To deal with the impulse noise, some additional steps are considered to detect the positions of the corruption elements in the current iterate $\hat{X}$. The corrupted pixels can be identified by inspecting the largest entries obtained from $l = \operatorname{abs}(X - \hat{X})$, where $\operatorname{abs}(\cdot)$ calculates the absolute value. This provides a flexible way to determine the corrupted pixels with some additional information.

V. Experimental Results and Discussion

A. Experimental Setting

To validate the effectiveness of the proposed method, we compare it with other related methods. Five restoration methods, including BM3D [24] based method, patch group-based

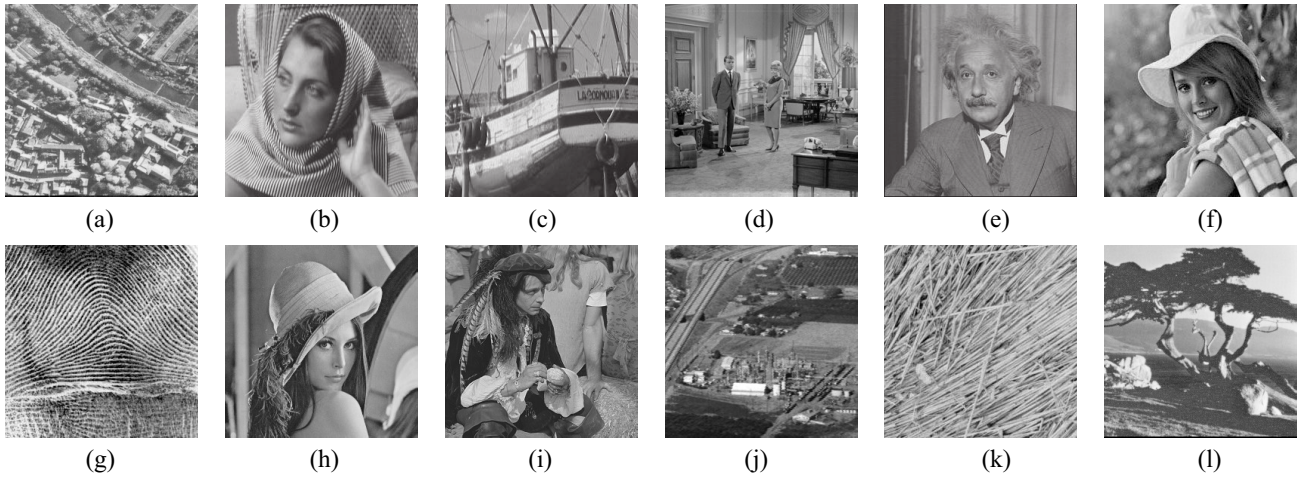


Fig. 1. Twelve images in our experiments. (a) *Aerial*. (b) *Barbara*. (c) *Boat*. (d) *Couple*. (e) *Einstein*. (f) *Elaine*. (g) *Fingerprint*. (h) *Lena*. (i) *Man*. (j) *Plant*. (k) *Straw*. (l) *Tree*.

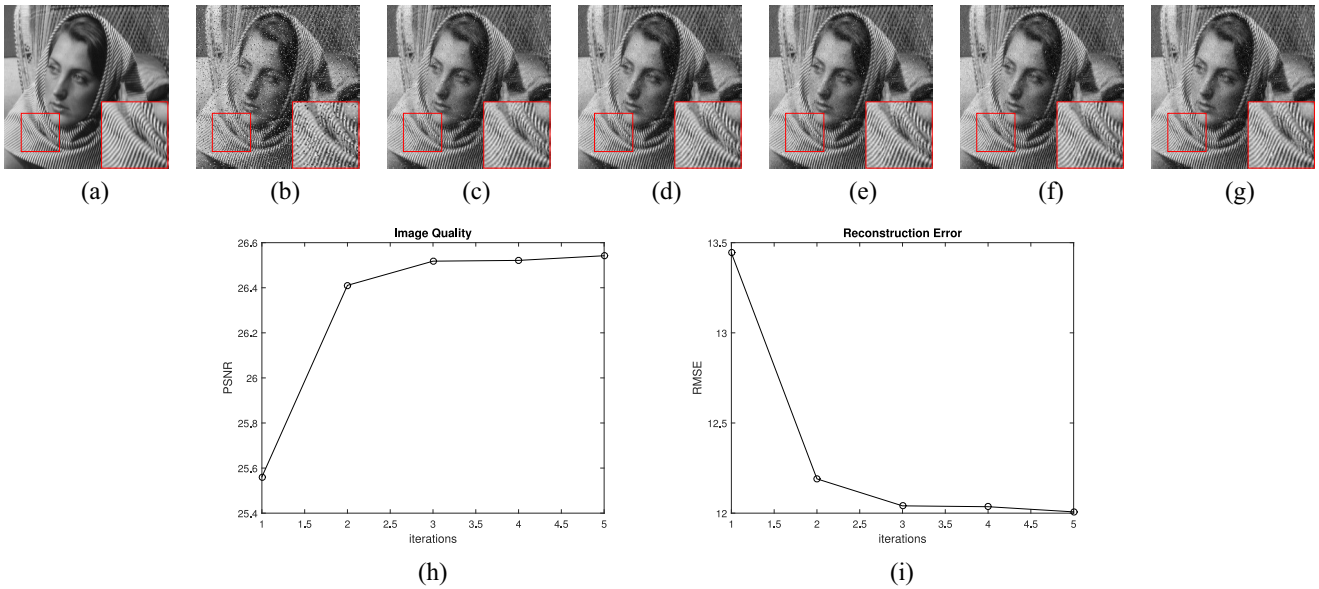


Fig. 2. Illustrative example of the proposed optimization process with five iterations. (a) *barbara*. (b) Noisy image. (c) $t = 1$. (d) $t = 2$. (e) $t = 3$. (f) $t = 4$. (g) $t = 5$. (h) PSNR versus number of iterations. (i) RMSE versus number of iterations.

nonlocal self-similarity prior learning method [43] (PGNP)-based method, ℓ_2-ℓ_1-TV-based method [12], ℓ_1-ℓ_0-based method [26], [27], and patch clustering with low-rank regularization (PCLR) method [44], are taken for a comprehensive comparison in terms of the restoration performance and visual quality. The parameters of these methods are configured defaultly. For BM3D, PCLR, and PGNP-based methods, we first apply adaptive center-weighted median filter [18] (ACWMF) to filter impulse noise. Then, the referred methods are performed for denosing the filtered image individually. In this paper, (σ, ri) stands for the mixture of Gaussian noise with zero mean and different standard deviations, and different levels of random value impulse noise. Accordingly, $(\sigma, \text{sp}, \text{ri})$ stands for mixed Gaussian, salt-and-peppers noise, and random value impulse noise.

Experiments were conducted on some publicly available datasets [45]–[47]. First, we used the 12 images with size 256×256 from USC-SIPI image database[1] to evaluate the denoised performance. The samples can be seen in Fig. 1. Then, we performed extensive experiments on a subset of three datasets, i.e., categorical image quality (CSIQ) image database [45], Tampere image database [46] (TID2008), and multiply distorted image database [47] (MDID). The subset of CSIQ dataset consists of 30 images with 512×512. The subset of TID2008 dataset has 20 images with 512×384. The subset of MDID dataset is composed of 25 images with 512×384.

To evaluate the quality of reconstructed images, two objective quality metric, including peak signal-to-noise ratio (PSNR) and feature structural similarity (FSIM) index [48], are used in this paper. All the referred methods were tested in a desktop computer shipped with a CPU (2.9 Hz), RAM memory (8 GB).

[1] http://sipi.usc.edu/database/database.php?volume=misc

TABLE I
NUMERICAL RESULTS BY REMOVING MIXED GAUSSIAN AND RANDOM VALUE IMPULSE NOISE

		PSNR(dB)						FSIM					
Image	(σ, ri)	BM3D	ℓ_2-ℓ_1-TV	PCLR	ℓ_1-ℓ_0	PGNP	Proposed	BM3D	ℓ_2-ℓ_1-TV	PCLR	ℓ_1-ℓ_0	PGNP	Proposed
	$(10, 10\%)$	26.70	26.67	25.69	26.25	26.92	**28.36**	0.914	0.911	0.910	0.923	0.920	**0.932**
aerial	$(15, 20\%)$	24.12	24.15	24.39	23.49	24.27	**25.61**	0.866	0.857	0.867	0.874	0.873	**0.896**
	$(20, 30\%)$	22.29	22.24	21.93	22.35	22.77	**23.33**	0.816	0.797	0.822	0.836	0.834	**0.854**
	$(10, 10\%)$	22.77	21.80	24.47	23.78	25.14	**26.53**	0.871	0.842	0.915	0.899	0.928	**0.931**
barbara	$(15, 20\%)$	21.69	21.37	21.88	21.81	22.43	**24.11**	0.822	0.800	0.861	0.838	0.878	**0.885**
	$(20, 30\%)$	20.71	20.47	20.60	21.03	20.75	**22.17**	0.775	0.765	0.806	0.808	0.831	**0.838**
	$(10, 10\%)$	28.10	26.47	28.06	28.27	28.13	**29.38**	0.920	0.909	0.918	0.925	0.923	**0.939**
boat	$(15, 20\%)$	26.10	24.87	25.95	25.54	26.08	**27.68**	0.880	0.868	0.873	0.879	0.885	**0.909**
	$(20, 30\%)$	24.04	23.50	24.03	23.98	24.17	**25.47**	0.841	0.829	0.827	0.838	0.845	**0.876**
	$(10, 10\%)$	28.37	27.17	28.24	29.24	28.40	**30.14**	0.904	0.901	0.899	0.922	0.911	**0.934**
couple	$(15, 20\%)$	26.38	25.44	26.21	26.39	26.35	**26.91**	0.852	0.848	0.838	0.871	0.863	**0.892**
	$(20, 30\%)$	24.55	24.19	24.31	24.73	24.34	**25.39**	0.808	0.806	0.779	0.827	0.812	**0.852**
	$(10, 10\%)$	29.92	29.52	29.86	25.27	29.72	**31.67**	0.911	0.915	0.904	0.892	0.916	**0.934**
einstein	$(15, 20\%)$	27.73	27.78	27.48	24.46	27.10	**28.80**	0.876	0.879	0.857	0.860	0.878	**0.899**
	$(20, 30\%)$	25.82	25.71	25.64	23.92	25.17	**26.52**	0.843	0.837	0.814	0.826	0.823	**0.857**
	$(10, 10\%)$	29.98	29.75	29.49	30.13	31.11	**31.29**	**0.936**	0.923	0.935	0.922	0.932	0.929
elaine	$(15, 20\%)$	27.35	27.88	26.68	27.95	**28.71**	28.66	0.899	0.888	0.885	0.891	**0.902**	0.893
	$(20, 30\%)$	25.22	25.98	24.41	26.33	**27.10**	26.51	0.862	0.859	0.841	0.861	**0.880**	0.856
	$(10, 10\%)$	21.49	18.30	23.04	22.07	23.14	**24.76**	0.908	0.808	0.950	0.921	0.953	**0.957**
fingerprint	$(15, 20\%)$	19.52	17.08	20.97	19.12	21.22	**22.19**	0.843	0.718	0.920	0.837	0.921	**0.926**
	$(20, 30\%)$	17.97	16.23	19.11	17.79	19.32	**21.20**	0.761	0.649	0.877	0.788	0.880	**0.906**
	$(10, 10\%)$	29.03	27.82	29.09	28.94	29.03	**29.95**	0.925	0.907	0.923	0.917	0.930	**0.940**
lena	$(15, 20\%)$	27.37	26.18	27.17	26.78	27.19	**27.76**	0.901	0.869	0.891	0.888	0.902	**0.908**
	$(20, 30\%)$	25.45	24.81	25.44	25.06	25.27	**26.02**	0.868	0.835	0.859	0.843	0.865	**0.873**
	$(10, 10\%)$	26.11	25.57	26.12	26.83	26.16	**27.68**	0.881	0.883	0.887	0.905	0.904	**0.915**
man	$(15, 20\%)$	24.32	24.19	24.60	23.88	24.65	**25.27**	0.832	0.836	0.827	0.844	0.849	**0.872**
	$(20, 30\%)$	23.15	23.24	23.22	21.89	23.28	**23.85**	0.799	0.802	0.778	0.762	0.810	**0.831**
	$(10, 10\%)$	28.41	27.52	28.30	28.81	28.37	**30.15**	0.915	0.902	0.907	0.930	0.919	**0.942**
plant	$(15, 20\%)$	26.15	25.36	25.89	26.06	26.06	**27.26**	0.863	0.852	0.848	0.872	0.874	**0.906**
	$(20, 30\%)$	24.09	23.96	23.93	24.37	24.05	**25.12**	0.812	0.799	0.782	0.840	0.826	**0.865**
	$(10, 10\%)$	25.35	23.63	25.43	26.22	25.40	**27.07**	0.905	0.852	0.931	0.930	0.934	**0.940**
straw	$(15, 20\%)$	23.05	22.33	23.01	22.79	23.16	**24.32**	0.837	0.812	0.880	0.841	0.889	**0.905**
	$(20, 30\%)$	21.22	20.90	21.16	21.51	21.29	**22.51**	0.758	0.745	0.843	0.808	0.845	**0.862**
	$(10, 10\%)$	27.97	27.24	28.10	26.33	27.80	**28.86**	0.906	0.911	0.898	0.916	0.918	**0.929**
tree	$(15, 20\%)$	25.99	25.08	25.68	24.22	25.84	**26.89**	0.878	0.872	0.859	0.868	0.883	**0.898**
	$(20, 30\%)$	23.85	23.43	23.88	23.16	23.98	**24.95**	0.835	0.830	0.822	0.832	0.845	**0.861**
	$(10, 10\%)$	27.01	25.95	27.15	26.84	27.44	**28.82**	0.908	0.889	0.915	0.917	0.924	**0.935**
AVG	$(15, 20\%)$	24.98	24.30	24.99	24.37	25.26	**26.28**	0.862	0.841	0.867	0.864	0.883	**0.899**
	$(20, 30\%)$	23.19	22.89	23.14	23.01	23.46	**24.42**	0.814	0.796	0.820	0.822	0.841	**0.861**
	$(10, 10\%)$	1.81	2.87	1.67	1.98	1.38	/	0.027	0.046	0.020	0.018	0.011	/
Gain	$(15, 20\%)$	1.30	1.98	1.29	1.91	1.02	/	0.037	0.058	0.032	0.035	0.016	/
	$(20, 30\%)$	1.23	1.53	1.28	1.41	0.96	/	0.047	0.065	0.041	0.039	0.020	/

B. Illustrative Example

In this section, we illustrate the proposed method with a simple example. Given a noisy image, we restore the latent image in an iterative manner. Fig. 2 demonstrates the reconstruction process in details. It can be seen that, as the optimization iteration increases, the restored image resembles more and more to the ground truth as indicated by the increase of the PSNR. In other word, the estimated image becomes more accurate as the decrease of root mean square error (RMSE). This implies that the proposed method effectively removes the mixed Gaussian-impulse noise.

Fig. 2(a)–(g) demonstrate the ground-truth image, observed image with noise level $(10, 10\%)$, and the five restored images corresponding to the optimization process, respectively. The selected part is enlarged and displayed in the bottom right corner. Moreover, Fig. 2(h) and (i) are presented to investigate the behavior of the proposed optimization process in terms of PSNR and RMSE. We can see that the proposed method converges in no more than five iterations. Thus, we set the iteration number as 5 in all experiments.

C. Results

1) Evaluation on the Removal of Mixed Gaussian and Random Value Impulse Noise: It is not easy to recover a latent image from the one corrupted by mixed Gaussian and random value impulse noise because of the difficulty of the detection of random value impulse noise. Although some methods are not developed for denoising random value impulse noise naturally, the application of ACWMF can generate a prefilter result.

In Table I, we report the denoising results of all the methods. The strength of Gaussian noise is $\sigma = (10, 15, 20)$. Meanwhile, the ratio of random value impulse noise is $\text{ri} = (10\%, 20\%, 30\%)$, respectively. The best numerical results are highlighted. "AVG" is the symbol used to represent the average scores of all the outcomes. "Gain" denotes the improvement in terms of the average PSNR and FSIM results of the proposed method. The gains of the proposed approach are provided in the bottom of Table I. It can be seen that the proposed method outperforms other related methods.

Fig. 3 provides visual results of all competing methods with noise level $(10, 10\%)$. The noisy image is displayed

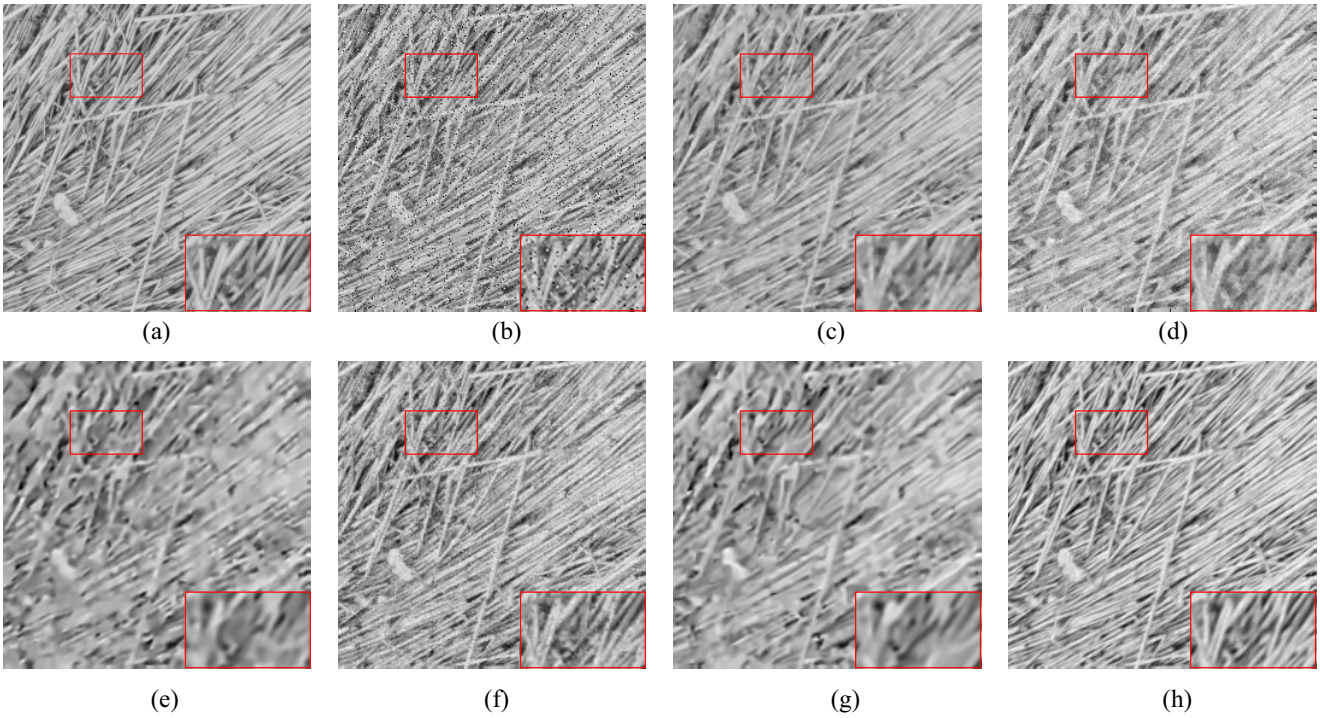


Fig. 3. Restoration results of *straw* image with $(\sigma, \text{ri}) = (10, 10\%)$. (a) For straw, (b) for noisy image, (c) for ACWMF + BM3D (PSNR = 25.35 and FSIM = 0.905), (d) for ℓ_2-ℓ_1-TV-based method (PSNR = 23.63 and FSIM = 0.852), (e) for ACWMF + PCLR-based method (PSNR = 25.43 and FSIM = 0.931), (f) for ℓ_1-ℓ_0-based method (PSNR = 26.22 and FSIM = 0.930), (g) for ACWMF + PGNP (PSNR = 25.40 and FSIM = 0.934), and (h) for proposed method (PSNR = 27.07 and FSIM = 0.940).

in Fig. 3(b). The recovered results of ACWMF+BM3D, ℓ_2-ℓ_1-TV, ACWMF+PCLR, ℓ_1-ℓ_0 and ACWMF+PGNP-based methods are presented in Fig. 3(c)–(g), respectively. It can be noted that some fine details are lost. Visual appearance of the proposed method, provided in Fig. 3(h), validates that the proposed method can reconstruct the geometry of the image fair well. Beyonds these, the continuity of fine details of image is preserved, especially textures and corners. To have a visual comparison about the denoising behavior, the selected part is enlarged.

2) Evaluation on the Removal of Mixed Gaussian, Salt-and-Pepper and Random Value Impulse Noise: To demonstrate the performance over different settings of mixed Gaussian, salt-and-pepper and random value impulse noise, we compare state-of-the-art methods on two noise levels $(\sigma, \text{sp}, \text{ri}) = (10, 10\%, 10\%), (15, 15\%, 05\%)$. Table II provides the numerical results of all competing methods. Similar, the average PSNR and FISM results of the related methods are provided. The gains, in terms of the outcomes of the proposed method, are also presented in the bottom of Table II. It can be seen that our numerical results with the proposed method are quite competitive.

In Fig. 4, we present a representative example of the reconstruction of the *boat* image with noise level $(10, 10\%, 10\%)$. Fig. 4(h) shows visual quality of the proposed method. From the enlarged windows, we can see that our approach is better than other related methods in terms of the local details. It can be observed that ℓ_2-ℓ_1-TV-based method more or less generates artifacts. It can be seen that the visual results of ACWMF+PCLR and ACWMF+PGNP-based methods are plausible. However, these two methods both generate over-smooth edges with details lost to some extent.

3) Experimental Results on Three Datasets: We performed extensive experiments on a subset of three publicly available datasets with three noise levels, i.e., CSIQ [45], TID2008 [46], and MDID [47]. Table III shows the average PSNR and FSIM results of the recovered images of all competing methods. It is clear that the proposed method behaves much better. Due to the limitations of the space, we do not provide the visual results of all competing methods.

4) Evaluation on Running Time: In this section, some experiments are performed for the comparison of the running time of all competing methods. The test image is *barbara* with size 256×256. Table IV presents the running time of the six methods for the removal of mixed Gaussian and random value impulse noise with noise level $(10, 10\%)$. It can be seen that ACWMF+BM3D-based method runs the fastest than the other methods because the BM3D algorithm is written in C with a MATLAB interface.

D. Discussion

To learn a discriminative structured dictionary with block and orthogonal constraints, we studied its minimization strategy in this paper. The numerical results of the proposed method for mixed Gaussian-impulse noise are demonstrated in Tables I–III. Visual results of the proposed method are provided in Figs. 3(h) and 4(h). When the strength of mixed Gaussian-impulse noise-levels increases, the proposed method behaves a significant improvement on both PSNR

TABLE II
NUMERICAL RESULTS BY REMOVING MIXED GAUSSIAN, SALT-AND-PEPPER AND RANDOM VALUE IMPULSE NOISE

		PSNR(dB)						FSIM					
Image	$(\sigma, \text{sp}, \text{ri})$	BM3D	ℓ_2-ℓ_1-TV	PCLR	ℓ_1-ℓ_0	PGNP	Proposed	BM3D	ℓ_2-ℓ_1-TV	PCLR	ℓ_1-ℓ_0	PGNP	Proposed
aerial	$(10, 10\%, 10\%)$	25.41	25.10	25.57	25.62	25.39	**26.68**	0.901	0.898	0.899	**0.914**	0.903	0.912
	$(15, 15\%, 05\%)$	24.00	23.99	24.32	24.36	24.53	**25.60**	0.871	0.878	0.859	0.886	0.875	**0.895**
barbara	$(10, 10\%, 10\%)$	22.75	21.37	22.26	23.32	22.57	**24.36**	0.870	0.831	0.853	0.890	0.880	**0.907**
	$(15, 15\%, 05\%)$	22.42	21.12	21.64	21.79	21.83	**23.65**	0.841	0.811	0.811	0.845	0.828	**0.864**
boat	$(10, 10\%, 10\%)$	27.49	25.57	27.23	27.30	27.20	**28.63**	0.915	0.898	0.910	0.908	0.913	**0.925**
	$(15, 15\%, 05\%)$	26.68	24.49	26.04	26.08	26.04	**27.69**	0.888	0.865	0.875	0.887	0.885	**0.907**
couple	$(10, 10\%, 10\%)$	27.61	26.11	27.53	28.22	27.55	**29.65**	0.892	0.883	0.887	0.907	0.898	**0.926**
	$(15, 15\%, 05\%)$	26.34	25.17	26.28	27.07	26.51	**28.15**	0.857	0.862	0.843	0.879	0.865	**0.902**
einstein	$(10, 10\%, 10\%)$	30.33	28.36	28.29	25.40	28.64	**30.97**	0.912	0.902	0.896	0.885	0.908	**0.915**
	$(15, 15\%, 05\%)$	28.87	26.75	27.80	24.85	27.69	**29.82**	0.889	0.870	0.856	0.863	0.881	**0.893**
elaine	$(10, 10\%, 10\%)$	29.94	28.61	29.72	29.58	**30.63**	30.56	**0.937**	0.915	0.918	0.913	0.927	0.926
	$(15, 15\%, 05\%)$	28.56	27.01	28.68	27.84	**29.26**	28.77	**0.912**	0.886	0.894	0.887	0.907	0.896
fingerprint	$(10, 10\%, 10\%)$	21.39	17.68	20.16	20.95	20.11	**22.89**	0.906	0.779	0.873	0.903	0.874	**0.917**
	$(15, 15\%, 05\%)$	21.35	17.22	19.46	19.16	19.70	**21.94**	0.898	0.758	0.842	0.866	0.854	**0.901**
lena	$(10, 10\%, 10\%)$	28.46	27.09	28.11	28.01	28.11	**29.33**	0.920	0.898	0.914	0.899	0.922	**0.923**
	$(15, 15\%, 05\%)$	27.58	25.68	27.14	27.09	27.39	**28.02**	0.902	0.864	0.891	0.892	0.904	**0.912**
man	$(10, 10\%, 10\%)$	25.57	24.89	25.37	25.94	25.41	**26.38**	0.869	0.866	0.868	0.888	0.879	**0.896**
	$(15, 15\%, 05\%)$	24.49	23.99	24.58	24.90	24.83	**25.98**	0.838	0.840	0.826	0.861	0.853	**0.877**
plant	$(10, 10\%, 10\%)$	27.87	26.47	27.63	28.06	27.45	**29.17**	0.904	0.886	0.899	0.921	0.907	**0.931**
	$(15, 15\%, 05\%)$	26.59	25.14	25.99	26.58	26.21	**27.65**	0.868	0.867	0.849	0.888	0.876	**0.908**
straw	$(10, 10\%, 10\%)$	24.40	22.55	24.38	25.30	24.38	**26.25**	0.883	0.788	0.883	0.916	0.890	**0.927**
	$(15, 15\%, 05\%)$	23.27	22.01	23.14	23.54	23.31	**24.97**	0.839	0.778	0.832	0.867	0.851	**0.907**
tree	$(10, 10\%, 10\%)$	25.78	25.61	26.96	25.77	26.82	**27.52**	0.890	0.894	0.891	0.899	0.901	**0.904**
	$(15, 15\%, 05\%)$	25.04	24.60	26.03	24.91	26.07	**26.53**	0.865	0.870	0.863	0.880	0.883	**0.898**
AVG	$(10, 10\%, 10\%)$	26.41	24.95	26.10	26.12	26.18	**27.69**	0.899	0.869	0.891	0.904	0.900	**0.917**
	$(15, 15\%, 05\%)$	25.43	23.93	25.09	24.84	25.28	**26.56**	0.872	0.845	0.853	0.875	0.872	**0.897**
Gain	$(10, 10\%, 10\%)$	1.28	2.74	1.59	1.57	1.51	/	0.018	0.048	0.026	0.013	0.017	/
	$(15, 15\%, 05\%)$	1.13	2.63	1.47	1.72	1.28	/	0.025	0.052	0.044	0.022	0.025	/

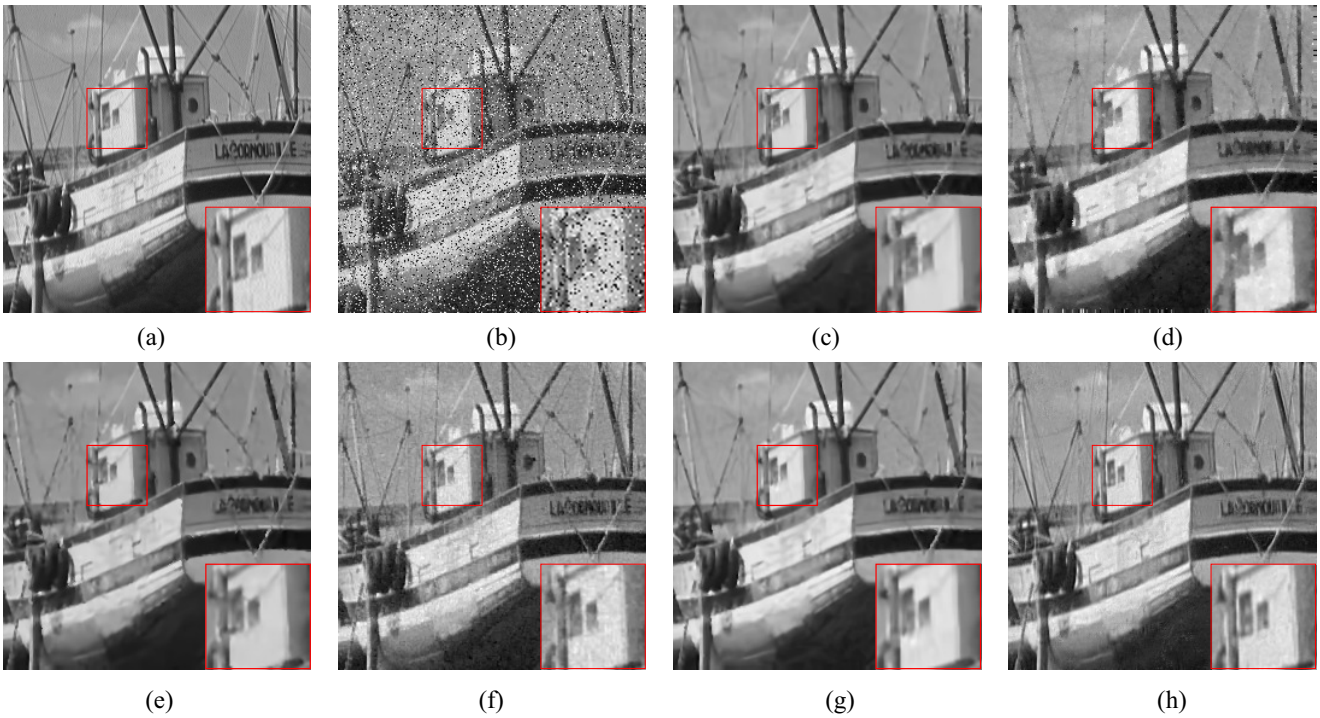


Fig. 4. Restored results of *boat* image with $(\sigma, \text{sp}, \text{ri}) = (10, 10\%, 10\%)$. (a) For *Boat*, (b) for noisy image, (c) for ACWMF + BM3D (PSNR = 27.49 and FSIM = 0.915), (d) for ℓ_2-ℓ_1-TV-based method (PSNR = 25.57 and FSIM = 0.898), (e) for ACWMF + PCLR-based method (PSNR = 27.23 and FSIM = 0.910), (f) for ℓ_1-ℓ_0-based method (PSNR = 27.30 and FSIM = 0.908), (g) for ACWMF + PGNP (PSNR = 27.20 and FSIM = 0.913), and (h) for proposed method (PSNR = 28.63 and FSIM = 0.925).

and FSIM. The visual experiments suggest that the difference may related to the method and optimization method we taken. Moreover, the proposed discriminative structured dictionary learning method, i.e., local subspaces tracking method by optimizing on Grassmann manifold, is critical for the success. The proposed method exploits the natural characteristic of the learned orthogonal bases, which converts the original problem into an optimization problem on Grassman manifold.

TABLE III
AVERAGE PSNR AND FSIM RESULTS FROM THE RECOVERY OF THREE DATASETS WITH THREE NOISE LEVELS

		PSNR(dB)						FSIM					
Image	(σ, ri)	BM3D	ℓ_2-ℓ_1-TV	PCLR	ℓ_1-ℓ_0	PGNP	Proposed	BM3D	ℓ_2-ℓ_1-TV	PCLR	ℓ_1-ℓ_0	PGNP	Proposed
CSIQ	$(10, 10\%)$	27.07	25.25	27.07	26.84	27.12	**28.65**	0.903	0.882	0.901	0.898	0.907	**0.912**
	$(15, 20\%)$	25.15	23.37	25.04	24.77	25.12	**26.13**	0.858	0.826	0.850	0.852	0.862	**0.864**
	$(20, 30\%)$	23.22	21.91	23.17	23.46	23.25	**24.14**	0.812	0.778	0.801	0.810	0.817	**0.823**
MDID	$(10, 10\%)$	28.08	25.74	28.10	27.57	29.09	**30.10**	0.892	0.862	0.885	0.890	0.926	**0.929**
	$(15, 20\%)$	26.35	23.81	26.39	25.78	26.53	**27.64**	0.857	0.807	0.844	0.849	0.881	**0.890**
	$(20, 30\%)$	24.51	22.37	24.52	24.38	24.53	**25.55**	0.820	0.759	0.803	0.807	0.836	**0.852**
TID2008	$(10, 10\%)$	27.81	25.90	27.76	27.93	27.86	**28.74**	0.897	0.874	0.893	0.907	0.904	**0.915**
	$(15, 20\%)$	25.99	23.89	25.95	25.54	26.08	**27.42**	0.855	0.817	0.843	0.852	0.861	**0.883**
	$(20, 30\%)$	24.29	22.42	24.23	24.18	24.30	**25.33**	0.813	0.765	0.793	0.810	0.818	**0.839**

TABLE IV
RUNNING TIME (SECOND) ON IMAGE *Barbara*

	Time(second)					
barbara	BM3D	ℓ_2-ℓ_1-TV	PCLR	ℓ_1-ℓ_0	PGNP	Proposed
$(10, 10\%)$	**0.24**	21.75	48.14	96.55	10.66	137.65

Then, independent orthogonal bases can be updated in each cluster. It can be seen that some drawbacks of the proposed method may occur in some situations. One limitation of the proposed method is that the estimation of the number of local subspaces remains as a challenging problem.

In this paper, an alternative approach to discriminative structured dictionary learning by embedding the constraints into the search spaces is proposed. Specially, the cost function and the constrained search spaces are taken into consideration. The proposed method selectively exploits block-orthogonal constraint. It is remarkable that the combination of block-orthogonal-based spare representation and local subspaces tracking by optimizing on Grassmann manifold leads to good quality reconstructions, when compared to state-of-art methods. The proposed method may provide an alternative way to represent local content of image, and deal with mixed Gaussian-impulse noise. The numerical results in terms of the objective evaluation show the validity and effectiveness of the proposed method.

VI. CONCLUSION

In this paper, we propose a discriminative structured dictionary learning method on Grassmann manifolds for the removal of mixed Gaussian-impulse noise. An alternating minimization scheme via exploiting the particular geometry of block-orthogonal constraint is developed. The proposed approach utilizes the natural orthogonal constraint to improve the representation of complex pattern over the image patches. A Riemannian conjugate gradient-based method, using the framework of retraction-based optimization on manifolds, is developed. This scheme leads to approximate the geometrical structures of image of interest. Experimental results on public datasets show that the proposed method outperforms some state-of-the-art methods in terms of reconstruction errors and visual quality.

REFERENCES

[1] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Process. Mag.*, vol. 14, no. 2, pp. 24–41, Mar. 1997.

[2] B. K. Gunturk and X. Li, *Image Restoration: Fundamentals and Advances*. Boca Raton, FL, USA: CRC Press, 2012.

[3] Y. Peng, J. Suo, Q. Dai, and W. Xu, "Reweighted low-rank matrix recovery and its application in image restoration," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2418–2430, Dec. 2014.

[4] Q. Liu, M. Chen, and D. Zhou, "Single image haze removal via depth-based contrast stretching transform," *Sci. China Inf. Sci.*, vol. 58, no. 1, pp. 1–17, 2015.

[5] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "A comprehensive survey to face hallucination," *Int. J. Comput. Vis.*, vol. 106, no. 1, pp. 9–30, 2014.

[6] Y. Hu, N. Wang, D. Tao, X. Gao, and X. Li, "SERF: A simple, effective, robust, and fast image super-resolver from cascaded linear regression," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4091–4102, Sep. 2016.

[7] H. El-Alfy, I. Mitsugami, and Y. Yagi, "Gait recognition based on normal distance maps," *IEEE Trans. Cybern.*, to be published.

[8] Q. Zou, L. Ni, Q. Wang, Q. Li, and S. Wang, "Robust gait recognition by integrating inertial and RGBD sensors," *IEEE Trans. Cybern.*, to be published.

[9] Y. Sui, G. Wang, and L. Zhang, "Correlation filter learning toward peak strength for visual tracking," *IEEE Trans. Cybern.*, to be published.

[10] L. Wang, L. Zhang, and Z. Yi, "Trajectory predictor by using recurrent neural networks in visual tracking," *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3172–3183, Oct. 2017.

[11] N. Wang, X. Gao, L. Sun, and J. Li, "Bayesian face sketch synthesis," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1264–1274, Mar. 2017.

[12] P. Rodríguez, R. Rojas, and B. Wohlberg, "Mixed Gaussian-impulse noise image restoration via total variation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Kyoto, Japan, 2012, pp. 1077–1080.

[13] C. Zou and Y. Xia, "Poissonian hyperspectral image superresolution using alternating direction optimization," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 9, pp. 4464–4479, Sep. 2016.

[14] M. Aharon, M. Elad, and A. Bruckstein, "$\mathbb{K}$-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[15] G. Yu, G. Sapiro, and S. Mallat, "Image modeling and enhancement via structured sparse model selection," in *Proc. 17th IEEE Int. Conf. Image Process. (ICIP)*, Hong Kong, Sep. 2010, pp. 1641–1644.

[16] R. Garnett, T. Huegerich, C. Chui, and W. He, "A universal noise removal algorithm with an impulse detector," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1747–1754, Nov. 2005.

[17] H. Hwang and R. A. Haddad, "Adaptive median filters: New algorithms and results," *IEEE Trans. Image Process.*, vol. 4, no. 4, pp. 499–502, Apr. 1995.

[18] T. Chen and H. R. Wu, "Adaptive impulse detection using center-weighted median filters," *IEEE Signal Process. Lett.*, vol. 8, no. 1, pp. 1–3, Jan. 2001.

[19] Y. Xia, H. Leung, and M. S. Kamel, "A discrete-time learning algorithm for image restoration using a novel L_2-norm noise constrained estimation," *Neurocomputing*, vol. 198, pp. 155–170, Jul. 2016.

[20] W. Feng, P. Qiao, and Y. Chen, "Fast and accurate poisson denoising with trainable nonlinear diffusion," *IEEE Trans. Cybern.*, to be published.

[21] J. Wright *et al.*, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.
[22] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2009.
[23] J.-F. Cai, R. H. Chan, and M. Nikolova, "Two-phase approach for deblurring images corrupted by impulse plus Gaussian noise," *Inverse Problems Imag.*, vol. 2, no. 2, pp. 187–204, 2008.
[24] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
[25] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3D transform-domain collaborative filtering," in *Proc. 15th Eur. Signal Process. Conf.*, vol. 1. Poznań, Poland, pp. 145–149.
[26] Y. Xiao, T. Zeng, J. Yu, and M. K. Ng, "Restoration of images corrupted by mixed Gaussian-impulse noise via ℓ_1–ℓ_0 minimization," *Pattern Recognit.*, vol. 44, no. 8, pp. 1708–1720, 2011.
[27] M. Filipovic and A. Jukic, "Restoration of images corrupted by mixed Gaussian-impulse noise by iterative soft-hard thresholding," in *Proc. 22nd Eur. Signal Process. Conf. (EUSIPCO)*, Lisbon, Portugal, 2014, pp. 1637–1641.
[28] Y. Xia and M. S. Kamel, "Novel cooperative neural fusion algorithms for image restoration and image fusion," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 367–381, Feb. 2007.
[29] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
[30] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process.*, vol. 17, no. 1, pp. 53–69, Jan. 2008.
[31] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.
[32] N. Boumal and P.-A. Absil, "RTRMC: A Riemannian trust-region method for low-rank matrix completion," in *Proc. Adv. Neural Inf. Process. Syst.*, Granada, Spain, 2011, pp. 406–414.
[33] L. Simonsson and L. Eldén, "Grassmann algorithms for low rank approximation of matrices with missing values," *BIT Numer. Math.*, vol. 50, no. 1, pp. 173–191, 2010.
[34] G. Gasso, A. Rakotomamonjy, and S. Canu, "Recovering sparse signals with a certain family of nonconvex penalties and DC programming," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4686–4698, Dec. 2009.
[35] J. Ho, M.-H. Yang, J. Lim, K.-C. Lee, and D. J. Kriegman, "Clustering appearances of objects under varying illumination conditions," in *Proc. IEEE Comput. Soc. Conf. Proc. Comput. Vis. Pattern Recognit.*, vol. 1. 2003, pp. 11–18.
[36] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Dictionary optimization for block-sparse representations," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2386–2395, May 2012.
[37] C. Bao, J. Cai, and H. Ji, "Fast sparsity-based orthogonal dictionary learning for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Sydney, NSW, Australia, 2013, pp. 3384–3391.
[38] C. Rusu and B. Dumitrescu, "Block orthonormal overcomplete dictionary learning," in *Proc. 21st Eur. Signal Process. Conf. (EUSIPCO)*, Marrakesh, Morocco, 2013, pp. 1–5.
[39] M. Harandi, C. Sanderson, C. Shen, and B. Lovell, "Dictionary learning and sparse coding on Grassmann manifolds: An extrinsic solution," in *Proc. Int. Conf. Comput.Vis.*, Sydney, NSW, Australia, 2013, pp. 3120–3127.
[40] M. Harandi, R. Hartley, C. Shen, B. Lovell, and C. Sanderson, "Extrinsic methods for coding and dictionary learning on grassmann manifolds," *Int. J. Comput.Vis.*, vol. 114, nos. 2–3, pp. 113–136, 2015.
[41] B. Vandereycken, "Low-rank matrix completion by Riemannian optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1214–1236, 2013.
[42] S. Mittal and P. Meer, "Conjugate gradient on Grassmann manifolds for robust subspace estimation," *Image Vis. Comput.*, vol. 30, nos. 6–7, pp. 417–427, 2012.
[43] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng, "Patch group based nonlocal self-similarity prior learning for image denoising," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 244–252.
[44] F. Chen, L. Zhang, and H. Yu, "External patch prior guided internal clustering for image denoising," in *Proc. IEEE Int. Conf. Comput. Vis.*, Santiago, Chile, 2015, pp. 603–611.
[45] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, 2010, Art. no. 011006.
[46] N. Ponomarenko *et al.*, "TID2008-a database for evaluation of full-reference visual quality assessment metrics," *Adv. Modern Radioelectron.*, vol. 10, no. 4, pp. 30–45, 2009.
[47] W. Sun, F. Zhou, and Q. Liao, "MDID: A multiply distorted image database for image quality assessment," *Pattern Recognit.*, vol. 61, pp. 153–168, Jan. 2017.
[48] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

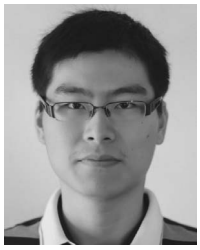
Han Pan received the B.S. and M.S. degree from Xidian University, Xi'an, China, in 2006 and 2009, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2014.

He is currently a Post-Doctoral Fellow with Shanghai Jiao Tong University. His current research interests include information fusion, optimization algorithms on Riemannian manifold, and space robotics.

Zhongliang Jing was born in Sichuan, China, in 1960. He received the B.S., M.S., and Ph.D. degrees in electronics and information technology from Northwestern Polytechnical University, Xi'an, China, in 1983, 1988, and 1994, respectively.

He is currently a Cheung Kong Professor, and the Executive Dean with the School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai, China. He has led many research projects (e.g., National Natural Science Foundation, 863 National High Tech. Plan), had published six books, and authored over 170 journal papers, and over 30 national invention patents. His current research interests include multisource information acquisition, processing and fusion, target tracking, optimal estimation and stochastic control, and aerospace control, and information processing.

Prof. Jing is an Editorial Board Member of *Science China-Information Sciences* as well as the *International Journal of Space Science and Engineering*. He has held concurrently panel of the 863 national high technology plan.



Lingfeng Qiao received the B.S. degree in aerospace engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014, where he is currently pursuing the Ph.D. degree in control science and control engineering.

His current research interests include pattern recognition, signal processing, and robotics.

Minzhe Li was born in Jiangsu, China. He received the B.S. and M.S. degrees from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2009 and 2012. He is currently pursuing the Ph.D. degree with Shanghai Jiao Tong University, Shanghai, China.

His current research interests include tracking and information fusion.