Low-rank Modeling and its Applications in Medical Image Analysis

Xiaowei Zhou and Weichuan Yu*

The Hong Kong University of Science and Technology, Hong Kong SAR, China

ABSTRACT

Computer-aided medical image analysis has been widely used in clinics to facilitate objective disease diagnosis. This facilitation, however, is often qualitative instead of quantitative due to the analysis challenges associated with medical images such as low signal-to-noise ratio, signal dropout, and large variations. Consequently, physicians have to rely on their personal experiences to make diagnostic decisions, which in turn is expertise-dependent and prone to individual bias.

Recently, low-rank modeling based approaches have achieved great success in natural image analysis. There is a trend that low-rank modeling will find its applications in medical image analysis. In this review paper, we like to review the recent progresses along this direction. Concretely, we will first explain the mathematical background of low-rank modeling, categorize existing low-rank modeling approaches and their applications in natural image analysis. After that, we will illustrate some application examples of using low-rank modeling in medical image analysis. Finally, we will discuss some possibilities of developing more robust analysis methods to better analyze cardiac images.

Keywords: Review, low-rank, computer vision, medical image analysis

1. INTRODUCTION

In many areas of image analysis, the latent structure underlying image data is assumed to be a low-dimensional subspace. Multiple vectorized images will form a low-rank matrix. Specific examples include background images under different illuminations, dynamic textures with periodicity, a group of similar shapes, and 3-D trajectories of feature points on a rigid object. Therefore, relevant tools such as principal component analysis have been widely used in various problems to explore the low-rank structure of data. In early literatures, the low-rank property of data was often used in preprocessing for dimension reduction or pattern extraction instead of being combined with other features in a unified model to provide a complete solution. A possible reason is the difficulty of rank minimization. Also, it is hard for the conventional approaches to handle outliers or missing values in data. Recent advances in low-rank modeling have proposed powerful tools such as nuclear norm relaxation,[1] robust principal component analysis,[2] and matrix completion[3] for data analysis.

In this paper, we will review the mathematical background of low-rank modeling, recent progresses in this area, and state-of-the-art algorithms to solve related optimization problems with applications in natural image analysis. After that, we will illustrate some examples of using low-rank modeling in medical image analysis. Finally, we will discuss some possibilities of developing more robust algorithms to better analyze cardiac images based on low-rank modeling.

2. MATHEMATICAL BACKGROUND

In many research fields, the high dimensionality of data brings great challenges to data analysis. Typical examples include images in vision problems, documents in natural language processing, users' ratings in recommender systems, and gene expression profiles in bioinformatics. Fortunately, the high-dimensional data usually lie in a subspace with limited dimensions. This fact greatly reduces the complexity of the problems we face.

* Weichuan Yu: E-mail: eeyu@ust.hk; Telephone: +852 2358-7054; Fax: +852 2358-1485

Independent Component Analyses, Compressive Sampling, Wavelets, Neural Net, Biosystems, and Nanoengineering XI,
edited by Harold H. Szu, Proc. of SPIE Vol. 8750, 87500V · © 2013 SPIE
CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2017684

If we represent one image as a column vector and arrange related images as columns of a matrix $\mathbf{D} \in \mathbb{R}^{m \times n}$, the corresponding data matrix $\mathbf{D}$ should read as

$$\mathbf{D} = \mathbf{X} + \mathbf{E}, \tag{1}$$

where $\mathbf{X}$ is a low-rank matrix, *i.e.* $\text{rank}(\mathbf{X}) \ll \min(m,n)$, and $\mathbf{E}$ represents noise in measurements. A conventional approach to finding the low-rank approximation is by minimizing

$$\begin{aligned} &\min_{\mathbf{X}} \ \|\mathbf{D} - \mathbf{X}\|_F^2, \\ &\text{s.t. } \text{rank}(\mathbf{X}) \leq k, \end{aligned} \tag{2}$$

where $\|\mathbf{Y}\|_F = \sqrt{\sum_{i,j} y_{ij}^2}$ denotes the Frobenious norm. Such a minimization is to seek the best rank-k estimate of $\mathbf{D}$ in a least-squares sense. The minimization in (2) can be solved analytically by using singular value decomposition (SVD).[4] According to the matrix approximation lemma (Eckart-Young-Mirsky theorem),[5] the solution to (2) is given by

$$\mathbf{X}^* = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \tag{3}$$

where $\{\mathbf{u}_i\}$, $\{\mathbf{v}_i\}$, and $\{\sigma_i\}$ for $i = 1, \cdots, k$ are the first k left singular vectors, right singular vectors, and singular values of $\mathbf{D}$, respectively. The vectors $\mathbf{u}_1, \cdots, \mathbf{u}_k$ also provide a set of orthonormal bases for the low-dimensional subspace that can best embed the data. This procedure corresponds to Principal Component Analysis (PCA)[6] in statistics.

PCA has become one of the most popular tools for data analysis because of its analytical solution in computation and the optimality under the assumption of i.i.d. Gaussian noise. However, it is limited in real applications when part of the data are grossly corrupted or missing.[2] Inspired by the recent advances of sparse learning and convex optimization, many new low-rank modeling-based algorithms have been proposed to address the challenges faced by traditional methods. In the following section, we like to introduce two of the most popular models.

2.1 Robust Principal Component Analysis

While PCA is optimal in the case of i.i.d. Gaussian noise, it can be easily corrupted by a few gross noisy points in data.[7] The reason is that it adopts the sum of squared residues to measure the data fidelity, which is not robust to outliers. Previous efforts towards the robust low-rank fitting tried to replace the squared penalty used in (2) with a more robust penalty function such as the Geman-McClure function[7] or the ℓ_1-penalty.[8] The limitation is that these methods use the alternating algorithms to solve the models. Thus, the solution depends on initialization and its optimality cannot be guaranteed.

Recently, people have started to solve the robust low-rank matrix recovery by sparse and low-rank decomposition, where the data matrix $\mathbf{D}$ is decomposed as the sum of a low-rank component $\mathbf{X}$ and a sparse component $\mathbf{E}$ by minimizing the rank of $\mathbf{X}$ and the cardinality of $\mathbf{E}$ simultaneously. The surprise message is that, under some mild assumptions, the low-rank matrix can be exactly recovered by the following convex optimization method named Principal Component Pursuit (PCP):[2]

$$\begin{aligned} &\min_{\mathbf{X},\mathbf{E}} \ \|\mathbf{X}\|_* + \lambda \|\mathbf{E}\|_1, \\ &\text{s.t. } \mathbf{X} + \mathbf{E} = \mathbf{D}. \end{aligned} \tag{4}$$

Here, the nuclear norm $\|\mathbf{X}\|_*$ and the ℓ_1-norm $\|\mathbf{E}\|_1$ are minimized, which are the convex surrogates of rank and cardinality, respectively. Candès *et al.*[2] and Chandrasekaran *et al.*[9] both analyzed the conditions for exact recovery. Briefly speaking, it has been proven in Candès's work[2] that $\mathbf{X}$ and $\mathbf{E}$ can be exactly recovered with high probability if the singular vectors of $\mathbf{X}$ are not sparse and the nonzero entries of $\mathbf{E}$ are sufficiently sparse and randomly distributed. Moreover, a theoretical choice of parameter λ is provided to make the exact recovery possible.

The basic model in (4) has been extended to handle more complicated scenarios. For example, stable PCP[10] relaxes the equality constraint in (4) as $\|\mathbf{X} + \mathbf{E} - \mathbf{D}\|_F \leq \sigma$ to allow the existence of Gaussian noise. In implementation, the following problem is solved

$$\min_{\mathbf{X},\mathbf{E}} \ \|\mathbf{X}\|_* + \lambda\|\mathbf{E}\|_1 + \frac{\mu}{2}\|\mathbf{X} + \mathbf{E} - \mathbf{D}\|_F^2, \tag{5}$$

where μ is a constant determined by the noise level. Other examples include outlier pursuit via group sparsity[11] and the matrix recovery from compressive measurements.[12]

2.2 Matrix Completion

In many applications, we would like to recover a matrix from only a limitedl number of observed entries. A typical example is collaborative filtering for recommender systems, in which we wish to make predictions to users' preference based on the information collected so far. The NetFlix problem[13] is a famous instance. The data is a big matrix $\mathbf{D}$ with each entry $D_{ij} \in \{1, \cdots, 5\}$ recording the rating of user i for movie j. There are around 480K users and 18K movies in the data set, but only 1.2% entries have values since each user only rated about 200 movies on average. The problem is how to predict the ratings that haven't been made yet based on the current observation, or in another word, how to complete the unknown entries in $\mathbf{D}$.

A popular solution is to assume the rating matrix is low-rank based on the fact that a subgroup of users are likely to share similar taste and their ratings are highly correlated. The problem then turns to be recovering a low-rank matrix from partial observation. In turn, the following optimization problem is considered in recent works:[14,15]

$$\begin{aligned} &\min_{\mathbf{X}} \ \|\mathbf{X}\|_*, \\ &\text{s.t. } \mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{D}), \end{aligned} \tag{6}$$

where Ω is the set of observed entries and $\mathcal{P}_\Omega$ denotes the projection of all matrix entries to the set of observed entries constrained by Ω. The equality constraint in (6) means that the entries in the recovered matrix should agree with the users' ratings for the rated entries. Under this constraint, the other unrated entries are predicted by minimizing the nuclear norm of the matrix $\mathbf{X}$, which is the convex relaxation of rank($\mathbf{X}$). Moreover, it can be proven that the solution to (6) will give an exact recovery of the low-rank matrix under certain conditions.[14]

In real applications, the rated entries in Ω might be noisy, and the equality constraint in (6) will be too strict, resulting in over-fitting.[16] Similar to Stable PCP,[10] the following relaxed form of (6) is considered for matrix completion with noise:[3]

$$\min_{\mathbf{X}} \ \frac{1}{2} \ \|\mathcal{P}_\Omega(\mathbf{D} - \mathbf{X})\|_F^2 + \lambda\|\mathbf{X}\|_*, \tag{7}$$

where the parameter λ depends on the noise level.

Another popular method for collaborative filtering is named Maximum Margin Matrix Factorization (MMM-F).[17] Instead of minimizing the rank of the approximating matrix $\mathbf{X}$, it factorizes $\mathbf{X}$ as a product of two matrices $\mathbf{U} \in \mathbb{R}^{m\times k}$ and $\mathbf{V} \in \mathbb{R}^{n\times k}$, and minimizes the following function

$$\min_{\mathbf{U},\mathbf{V}} \frac{1}{2} \ \|\mathcal{P}_\Omega(\mathbf{D} - \mathbf{U}\mathbf{V}^T)\|_F^2 + \frac{\lambda}{2}(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2). \tag{8}$$

It also fits the data with a low-rank matrix since rank$(\mathbf{U}\mathbf{V}^T) \leq k$. More interestingly, the solution to (8) is equivalent to the solution to (7)[16] since

$$\|\mathbf{X}\|_* = \min_{\mathbf{U},\mathbf{V}:\mathbf{X}=\mathbf{U}\mathbf{V}^T} \frac{1}{2} \ (\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2). \tag{9}$$

According to this equivalence, MMMF can also be considered as a low-rank model. The difference between MMMF and the model in (7) is that (8) is not convex while (7) is.

2.3 Optimization and Algorithms

Low-rank modeling usually requires rank minimization. However, the rank minimization problem (RMP) is combinatorial and known to be NP-hard.[1] A popular approach to solving RMP is replacing rank by the nuclear norm, which is the tight convex surrogate of rank.[1,18] Regarding computation, the nuclear norm has the following property (Theorem 2.1 given in Cai *et al.*[15]):

THEOREM 2.1. *The solution to the following problem*

$$\min_{\mathbf{X}} \ \frac{1}{2} \ \|\mathbf{Z} - \mathbf{X}\|_F^2 + \lambda\|\mathbf{X}\|_* \tag{10}$$

is given by $\mathbf{X}^* = \mathcal{D}_\lambda(\mathbf{Z})$, *where*

$$\mathcal{D}_\lambda(\mathbf{Z}) = \sum_{i=1}^{\min(m,n)} (\sigma_i - \lambda)_+ \mathbf{u}_i \mathbf{v}_i^T, \tag{11}$$

and $\mathbf{u}_i$, $\mathbf{v}_i$ *and* σ_i *correspond the left singular vector, the right singular vector and the singular value of* $\mathbf{Z}$, *respectively.* $\mathcal{D}_\lambda$ refers to the *singular value thresholding* (SVT) operator,[15] which serves as a basic ingredient in many algorithms. For example, the matrix completion problem in (7) is often solved by iteratively performing SVT on the data.

Based on Theorem 2.1, different algorithms can be developed to solve different problems. Two of the most popularly-used methods are the Proximal Gradient (PG) method[19] and the Augmented Lagrangian Method (ALM),[20] which are applicable to a variety of convex problems. The PG is extremely useful to solve the norm-regularized problems such as the model in (7), where the energy is the sum of a differentiable loss function and a nuclear norm regularizer. Moreover, it is often combined with the Nesterov method to accelerate the convergence.[21,22] Examples using the PG method are shown in the works.[16,23,24] The ALM is also named the Alternating Direction Method of Multipliers (ADMM).[25] It provides a powerful framework to solve convex problems with equality constraints such as PCP in (4) and the model in (6). The algorithms used in the works[2,26] belong to this class.

Many other works focus on the factor models,[7,8,17,27,28] where the low-rank matrix $\mathbf{X}$ is written as $\mathbf{U}\mathbf{V}^T$, a product of two matrices with smaller dimensions. Therefore, the rank minimization is not required in these problems. To estimate $\mathbf{U}$ and $\mathbf{V}$, the alternating minimization is often used. In each iteration, only one variable is updated with the other one fixed. Each update aims to decrease the energy function by solving a least-squares problem or using other update rules. In practice, these methods are generally faster than the methods using nuclear minimization. Also, it is more convenient for them to develop parallel or incremental versions of the algorithms. The major drawbacks of these methods are the nonconvex formulations, which means that the optimality in optimization cannot be guaranteed. Hence, the theoretical properties such as the exact recovery of the models in (4) and (6) cannot be proven for the factor models. However, these methods perform generally well in practice and are popularly used in applications.

3. APPLICATIONS IN COMPUTER VISION

Many objects of interest in vision problems have been modeled as having low-rank, such as images of a convex lambertian surface under various illuminations,[29] dynamic textures changing periodically,[30] a group of active contours with similar shapes,[31] and multiple 3-D trajectories of feature points from a rigid moving object.[32] Intuitively, the low-dimensional subspace models the common pattern underlying the data. Hence, recovering the low-rank structure is critical to high-level tasks such as background subtraction, face recognition, and segmentation. In the following, we review several typical examples based on the aforementioned models in Section 2.

Background subtraction. A direct application of Robust PCA is to model the background in video surveillance, where the task is to detect objects that stand out from the background. As illustrated in papers,[2,33] the background images captured by a static camera can be naturally modeled as a low-rank matrix. Hence, the background can be recovered by PCP as the low-rank component and the foreground objects can be identified as the sparse component. Zhou *et al.*[34] combined the Markov Random Field with matrix completion to achieve better accuracy for moving object detection. He *et al.*[35] proposed an online algorithm for background subtraction, where the low-rank model is updated incrementally.

Face recognition. Face recognition algorithms generally require a high-quality training set to build a classifier. However, face images in real data sets are usually corrupted by various defects such as shadows, specularities and occlusions.[2,36] It has been shown that face images of a person under various illuminations can be modeled by a low-rank matrix.[2,7,36] If PCP is applied, the local defects could be removed as the sparse component, while a correct description of the person's face could be obtained from the low-rank component. This procedure improves the quality of training data and boosts the performance of current algorithms for face recognition.[36]

Image alignment and rectification. Image alignment is to transform different images into the same coordinate system. Peng *et al.*[37] proposed to solve the problem by rank minimization based on the assumption that a batch of aligned images should form a low-rank matrix. The parameters of transformation τ were estimated by solving

$$\begin{aligned} \min_{\tau,\mathbf{X},\mathbf{E}} \ & \|\mathbf{X}\|_* + \lambda\|\mathbf{E}\|_1, \\ \text{s.t. } & \mathbf{X} + \mathbf{E} = \mathbf{D} \circ \tau, \end{aligned} \tag{12}$$

where each column of $\mathbf{D}$ corresponds to an image to be aligned and $\mathbf{D}\circ\tau$ indicates the images after transformation. Zhang *et al.*[38] applied the model (12) to generate transform-invariant low-rank textures (TILT), where $\mathbf{D}$ denotes images to be rectified. The TILT can be further used in various problems such as camera calibration, 3D reconstruction, character recognition, *etc.*

Subspace clustering. Another extended model of PCP is the Low-Rank Representation (LLR) for subspace clustering.[39] The task is to partition data points into multiple subspaces. A popular method is spectral clustering, where the segmentation is achieved by cutting a graph. In the graph, each node represents one pixel in the original image and a edge represents represents the affinity between two neighboring pixels. In LLR, each data point is represented by a linear combination of its neighbors within the same subspace, and the coefficients $\mathbf{X}$ is estimated by

$$\begin{aligned} \min_{\mathbf{X},\mathbf{E}} \ & \|\mathbf{X}\|_* + \lambda\|\mathbf{E}\|_{2,1}, \\ \text{s.t. } & \mathbf{D} = \mathbf{D}\mathbf{X} + \mathbf{E}. \end{aligned} \tag{13}$$

It is shown that,[39] if the data points in $\mathbf{D}$ are from several orthogonal subspaces, $\mathbf{X}$ derived from (13) will be block-diagonal. Intrinsically, $\mathbf{X}$ identifies the affinity among data points, and its block-diagonal structure indicates clusters in the data. Thus, $\mathbf{X}$ provides a preferred affinity matrix to perform spectral clustering.

Image restoration. A popular application of matrix completion is image or video restoration. In some cases, it is desired to reconstruct the lost or corrupted parts of an image, which might be caused by texts or logos superposed on the image. This process is named image inpainting.[40] As a natural image is approximately low-rank,[41] the missing pixels can be filled back by matrix completion. Another application is video denoising.[42] To remove the defects in a video, unreliable pixels in the video are first detected and labeled as missing. Then, the image patches are grouped such that the patches in each group share similar underlying structure and form a low-rank matrix approximately. Finally, the matrix completion is carried out on each patch group to restore the image.

Other applications of low-rank modeling include saliency detection by low-rank background modeling,[43] object tracking via structured Robust PCA,[44,45] block-wise partition for parsing façade via rank-one approximation,[46] fusion of model scores by rank-two matrix reconstruction,[47] and so on.

4. APPLICATIONS IN MEDICAL IMAGE ANALYSIS

While low-rank modeling has been popular in the machine learning and computer vision community, the applications of low-rank modeling in medical image analysis are relatively limited. Here we give a brief summary of these applications.

Image compression. Image compression using PCA has a long history.[48] The basic idea is that most of image contents can be described by only a few principal components, which require less storage space compared to the original image. The number of required components depends on the compression quality required in specific applications. For example,[48] an image is partitioned into blocks, which are further classified into different region types such as tissue or background. Then, different types of blocks are coded separately with different numbers of components to achieve the best compression ratio. A higher compression ratio in experiments was achieved compared to the discrete cosine transform (DCT).[48] The disadvantage is the requirement to store the principal components while the DCT doesn't need to. However, Taur and Tao[48] also stated that it was unnecessary to store the principal components for each image since the images from the same modality have similar statistics.

Image denoising. Low-rank based image denoising shares the similar idea as image compression. A signal can be represented by a limited number of principal components and the remaining components correspond to noise and thus can be removed. The difference is that denoising is often carried out on a sequence of images with each image being a column of the matrix. In MR image denoising, for example, an image sequence consists of multiple echo images,[49] a dynamic image sequence[50] or multiple diffusion-weighted images.[51] Nguyen *et al.*[50] integrated a low-rank model, a Rician noise model and an edge-preserving smoothness model in a MAP framework. In a more recent work,[52] Candes *et al.* used the SVT operator in (11) instead of the classical PCA to achieve more robust results. In these methods, a theoretical framework is proposed to select the optimal thresholding parameter, which is very convenience in practical applications. The drawback of these methods is the possibility of removing local image details that cannot be modeled as low-rank.

Image segmentation. Active shape model[53] has been widely used in medical image segmentation.[54] It increases the robustness of deformable models for image segmentation. It constructs a statistical shape space from a large set of annotated images and constrains the candidate shape in this shape space. Concretely, each shape (*e.g.* the endocardial surface of the left ventricle) is represented by a vector $\mathbf{s} = [x_1, \cdots, x_p, y_1, \cdots, y_p, z_1, \cdots, z_p]^T$ where (x_i, y_i, z_i) denotes a landmark on the parametric surface. In the shape space, the candidate shape is expressed as

$$\mathbf{s}(\mathbf{w}) = \bar{\mathbf{s}} + \mathbf{\Phi}\mathbf{w}, \tag{14}$$

where $\bar{\mathbf{s}}$ is the mean shape; $\mathbf{\Phi} \in \mathbb{R}^{3p\times r}$ is a matrix consisting of vectors describing different modes of shape variations in the training data; $\mathbf{w}$ is a vector of coefficients to represent the candidate shape in the shape space, which is determined by matching the shape to the features in the image. The number of columns in $\mathbf{\Phi}$ is often small. Thus, the candidate shape is limited in a low-dimensional space. In other words, the active shape model intrinsically admits a low-rank assumption on the population of shapes. Moreover, $\bar{\mathbf{s}}$ and $\mathbf{\Phi}$ are often obtained by applying PCA to the shapes in the training set. The limitation of this method is the requirement of annotated data sets for training. Also, it is often doubted that the existing shapes in the training set is sufficient to model the object shape in a new image.

Image reconstruction. Recently, image reconstruction based on low-rank modeling draws more and more attention. The idea is to make use of the temporal coherence in dynamic imaging to reduce the number of sampling. In MR imaging, for example, Liang *et al.*[55] proposed the concept of partial separability (PS) to model a spatial-temporal MR image $\rho(\mathbf{x}, t)$ as

$$\rho(\mathbf{x}, t) = \sum_{\ell=1}^{L} \phi_\ell(\mathbf{x}) v_\ell(t), \tag{15}$$

where $\{\phi_\ell(\mathbf{x})\}$ and $\{v_\ell(t)\}$ for $\ell = 1, \cdots, L$ are sets of spatial and temporal components to represent the image and L is the order of the model. Correspondingly, any sample in the $(\mathbf{k}, t)$-space can be expressed as $c(\mathbf{k}, t) = \sum_{\ell=1}^{L} u_\ell(\mathbf{k}) v_\ell(t)$, where $u_\ell(\mathbf{k})$ is the Fourier transform of $\phi_\ell(\mathbf{x})$. Using matrix notations, we have

$$\mathbf{C} = \mathbf{U}\mathbf{V}, \tag{16}$$

where $C_{ij} = c(\mathbf{k}_i, t_j)$, $U_{i\ell} = u_\ell(\mathbf{k}_i)$ and $V_{\ell j} = v_\ell(t_j)$. Since the images are temporally coherent, L can be very small, which gives a low-rank model of the coefficients $\mathbf{C}$ in the $(\mathbf{k}, t)$-space. Correspondingly, a small number of samples are sufficient to estimate $\mathbf{C}$ and reconstruct the image sequence. For example, $\{u_\ell(\mathbf{k})\}$ and $\{v_\ell(t)\}$ can be obtained by fully sampling L columns and rows of the $(\mathbf{k}, t)$-space.[55] Alternatively, $\mathbf{C}$ can be estimated by random sampling the $(\mathbf{k}, t)$-space followed by solving the matrix completion problem

$$\min_{\mathbf{U},\mathbf{V}} \|\mathcal{P}_\Omega(\mathbf{U}\mathbf{V}) - \mathcal{P}_\Omega(\mathbf{D})\|_F^2, \tag{17}$$

where $\mathcal{P}_\Omega$ indicates the sampling in a random set Ω and $\mathcal{P}_\Omega(\mathbf{D})$ corresponds to the measurement in Ω.

The basic PS model can be further extended to integrate other sparse properties in specific domains. The spatial component $\phi_\ell(\mathbf{x})$ (image pattern) often has a sparse representation in certain domains such as wavelets and total variation.[56–58] The temporal component $v_\ell(t)$ is usually periodic or bandlimited, which results in sparsity in the Fourier domain.[59,60] Also, the low-rank property can be modeled to be regionally dependent.[61] Instead of using the PS model, some works[57,58] impose the low-rank property by nuclear norm minimization, which gives convex formulations. Low-rank modeling methods have also been applied to other modalities such as CT[62,63] and PET.[64]

5. DISCUSSIONS

The low-rank modeling is based on the coherence among multiple images. Such prior knowledge can be used to compress images, to remove random noise, and to reduce the sampling rate in image acquisition. Recent progresses in sparse learning and optimization provide powerful tools to model the low-rank property of data.

We noticed that the low-rank modeling was rarely used in ultrasound image analysis. Previous methods for real-time 3-D echocardiography (RT3DE) analysis often focused on a single image or an image pair. In fact, many objects in RT3DE analysis also form a low-dimensional space, such as appearances of myocardium in a sequence, shapes of left ventricle throughout a cardiac cycle, and multi-view volumes in 3-D imaging.[65] Such a low dimensionality can largely reduce the complexity of analysis. Zhou *et al.*[66] proposed an automatic algorithm to track the mitral leaflet in echocardiography by modeling the smoothly-moving myocardium as the low-rank background and detecting the fast-moving mitral leaflet as outliers in the low-rank representation. We believe that the global low-rank modeling would largely improve the robustness of algorithms for cardiac image analysis.

While the low-rank modeling has shown great potentials in solving problems in computer vision and medical image analysis, there are two main limitations for the low-rank modeling based methods.

The first limitation is the heavy computational cost. Estimation of a low-rank matrix usually needs to solve some optimization problems. Although the problems are often convex, solving them still requires many iterations of complicated matrix computations such as SVD or solving large linear systems. These computation becomes more expensive when processing a sequence of 3-D volumetric images. Moreover, many applications in medical image analysis require online processing integrated in the imaging system. Fortunately, more and more efforts have been carried out to develop incremental or distributed algorithms to make the low-rank models more practical.[67,68]

The second limitation is that the low-rank assumption might be violated. For example, the image variation due to irregular tissue motion or abnormal respiration cannot be described by low-rank models. It is important to validate the low-rank assumption for specific problems before applying the corresponding tools.

REFERENCES

[1] Fazel, M., *Matrix rank minimization with applications*, PhD thesis, Stanford University (2002).

[2] Candès, E., Li, X., Ma, Y., and Wright, J., "Robust principal component analysis?," *Journal of the ACM* **58**(3), 11 (2011).

[3] Candes, E. and Plan, Y., "Matrix completion with noise," *Proceedings of the IEEE* **98**(6), 925–936 (2010).

[4] Golub, G. and van Van Loan, C., [*Matrix Computations*], The Johns Hopkins University Press (1996).

[5] Eckart, C. and Young, G., "The approximation of one matrix by another of lower rank," *Psychometrika* **1**(3), 211–218 (1936).

[6] Jolliffe, I. and MyiLibrary, [*Principal component analysis*], Wiley Online Library (2002).

[7] De La Torre, F. and Black, M., "A framework for robust subspace learning," *International Journal of Computer Vision* **54**(1), 117–142 (2003).

[8] Ke, Q. and Kanade, T., "Robust l1 norm factorization in the presence of outliers and missing data by alternative convex programming," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], (2005).

[9] Chandrasekaran, V., Sanghavi, S., Parrilo, P., and Willsky, A., "Rank-sparsity incoherence for matrix decomposition," *SIAM Journal on Optimization* **21**(2), 572–596 (2011).

[10] Zhou, Z., Li, X., Wright, J., Candes, E., and Ma, Y., "Stable principal component pursuit," in [*Proceedings of the IEEE International Symposium on Information Theory*], 1518–1522 (2010).

[11] Xu, H., Caramanis, C., and Sanghavi, S., "Robust pca via outlier pursuit," *IEEE Transactions on Information Theory* **58**, 3047 –3064 (may 2012).

[12] Wright, J., Ganesh, A., Min, K., and Ma, Y., "Compressive principal component pursuit," in [*Proceedings of the IEEE International Symposium on Information Theory*], (2012).

[13] ACM SIGKDD and Netflix, "Proceedings of KDD cup and workshop," (2007). Proceedings available online at `http://www.cs.uic.edu/~liub/KDD-cup-2007/proceedings.html`.

[14] Candès, E. and Recht, B., "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics* **9**(6), 717–772 (2009).

[15] Cai, J., Candès, E., and Shen, Z., "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization* **20**, 1956 (2010).

[16] Mazumder, R., Hastie, T., and Tibshirani, R., "Spectral regularization algorithms for learning large incomplete matrices," *The Journal of Machine Learning Research* **11**, 2287–2322 (2010).

[17] Srebro, N., Rennie, J., and Jaakkola, T., "Maximum-margin matrix factorization," *Advances in neural information processing systems* **17**(5), 1329–1336 (2005).

[18] RECHT, B., FAZEL, M., and PARRILO, P., "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review* **52**(3), 471–501 (2010).

[19] Moreau, J., "Proximité et dualité dans un espace hilbertien," *Bulletin de la Société Mathématique de France* **93**(2), 273–299 (1965).

[20] Bertsekas, D., [*Nonlinear programming*], Athena Scientific (1999).

[21] Nesterov, Y., "Gradient methods for minimizing composite objective function." CORE Discussion Papers (2007).

[22] Beck, A. and Teboulle, M., "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences* **2**(1), 183–202 (2009).

[23] Ji, S. and Ye, J., "An accelerated gradient method for trace norm minimization," in [*Proceedings of the 26th Annual International Conference on Machine Learning*], 457–464, ACM (2009).

[24] Toh, K. and Yun, S., "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pacific Journal of Optimization* **6**(615-640), 15 (2010).

[25] Boyd, S., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning* **3**(1), 1–122 (2010).

[26] Lin, Z., Chen, M., and Ma, Y., "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *Arxiv preprint arXiv:1009.5055* (2010).

[27] Okatani, T. and Deguchi, K., "On the wiberg algorithm for matrix factorization in the presence of missing components," *International Journal of Computer Vision* **72**(3), 329–337 (2007).
[28] Wen, Z., Yin, W., and Zhang, Y., "Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm," *Mathematical Programming Computation* , 1–29 (2010).
[29] Basri, R. and Jacobs, D., "Lambertian reflectance and linear subspaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(2), 218–233 (2003).
[30] Doretto, G., Chiuso, A., Wu, Y., and Soatto, S., "Dynamic textures," *International Journal of Computer Vision* **51**(2), 91–109 (2003).
[31] Blake, A. and Isard, M., [*Active contours*], Springer (2000).
[32] Vidal, R. and Hartley, R., "Motion segmentation with missing data using powerfactorization and gpca," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], (2004).
[33] Oliver, N., Rosario, B., and Pentland, A., "A bayesian computer vision system for modeling human interactions," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(8), 831–843 (2000).
[34] Zhou, X., Yang, C., and Yu, W., "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**(3), 597–610 (2013).
[35] He, J., Balzano, L., and Lui, J., "Online robust subspace tracking from partial information," *Arxiv preprint arXiv:1109.3827* (2011).
[36] Chen, C., Wei, C., and Wang, Y., "Low-rank matrix recovery with structural incoherence for robust face recognition," in [*IEEE Conference on Computer Vision and Pattern Recognition*], (2012).
[37] Peng, Y., Ganesh, A., Wright, J., Xu, W., and Ma, Y., "Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(11), 2233–2246 (2012).
[38] Zhang, Z., Ganesh, A., Liang, X., and Ma, Y., "Tilt: Transform invariant low-rank textures," *International Journal of Computer Vision* **99**, 1–24 (2012).
[39] Liu, G., Lin, Z., and Yu, Y., "Robust subspace segmentation by low-rank representation," in [*Proceedings of the 27th International Conference on Machine Learning*], (2010).
[40] Bertalmio, M., Sapiro, G., Caselles, V., and Ballester, C., "Image inpainting," in [*Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*], 417–424, ACM Press/Addison-Wesley Publishing Co. (2000).
[41] Zhang, D., Hangzhou, C., Hu, Y., Ye, J., Li, X., and He, X., "Matrix completion by truncated nuclear norm regularization," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], (2012).
[42] Ji, H., Liu, C., Shen, Z., and Xu, Y., "Robust video denoising using low rank matrix completion," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 1791–1798 (2010).
[43] Shen, X. and Wu, Y., "A unified approach to salient object detection via low rank matrix recovery," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], (2012).
[44] Ding, T., Sznaier, M., and Camps, O., "Fast track matching and event detection," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], (2008).
[45] Ayazoglu, M., Sznaier, M., and Camps, O., "Fast algorithms for structured robust principal component analysis," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], (2012).
[46] Yang, C., Han, T., Quan, L., and Tai, C., "Parsing façade with rank-one approximation," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], (2012).
[47] Ye, G., Liu, D., Jhuo, I., and Chang, S., "Robust late fusion with rank minimization," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], (2012).
[48] Taur, J. and Tao, C., "Medical image compression using principal component analysis," in [*Proceedings of the International Conference on Image Processing*], **1**, 903–906, IEEE (1996).
[49] Bydder, M. and Du, J., "Noise reduction in multiple-echo data sets using singular value decomposition," *Magnetic resonance imaging* **24**(7), 849–856 (2006).
[50] Nguyen, H., Peng, X., Do, M., and Liang, Z., "Spatiotemporal denoising of mr spectroscopic imaging data by low-rank approximations," in [*Proceedings of IEEE International Symposium on Biomedical Imaging*], 857–860, IEEE (2011).

[51] Lam, F., Babacan, S., Haldar, J., Schuff, N., and Liang, Z., "Denoising diffusion-weighted mr magnitude image sequences using low rank and edge constraints," in [*Proceedings of the IEEE International Symposium on Biomedical Imaging*], 1401–1404, IEEE (2012).

[52] Candes, E., Sing-Long, C., and Trzasko, J., "Unbiased risk estimates for singular value thresholding and spectral estimators," *arXiv preprint arXiv:1210.4139* (2012).

[53] Cootes, T., Taylor, C., Cooper, D., and Graham, J., "Active shape models – their training and application," *Computer Vision and Image Understanding* **61**(1), 38–59 (1995).

[54] Zhu, Y., Papademetris, X., Sinusas, A., and Duncan, J., "Segmentation of the left ventricle from cardiac mr images using a subject-specific dynamical model," *IEEE Transactions on Medical Imaging* **29**(3), 669–687 (2010).

[55] Liang, Z., "Spatiotemporal imagingwith partially separable functions," in [*Proceedings of the IEEE International Symposium on Biomedical Imaging*], 988–991, IEEE (2007).

[56] Lustig, M., Donoho, D., Santos, J., and Pauly, J., "Compressed sensing mri," *IEE Signal Processing Magazine* **25**(2), 72–82 (2008).

[57] Lingala, S., Hu, Y., DiBella, E., and Jacob, M., "Accelerated dynamic mri exploiting sparsity and low-rank structure: kt slr," *IEEE Transactions on Medical Imaging* **30**(5), 1042–1054 (2011).

[58] Majumdar, A. and Ward, R., "Exploiting rank deficiency and transform domain sparsity for mr image reconstruction," *Magnetic resonance imaging* **30**(1), 9–18 (2012).

[59] Zhao, B., Haldar, J., Brinegar, C., and Liang, Z., "Low rank matrix recovery for real-time cardiac mri," in [*Proceedings of the IEEE International Symposium on Biomedical Imaging*], 996–999, IEEE (2010).

[60] Zhao, B., Haldar, J., Christodoulou, A., and Liang, Z.-P., "Image reconstruction from highly undersampled (k,t)-space data with joint partial separability and sparsity constraints," *IEEE Transactions on Medical Imaging* **31**(9), 1809 –1820 (2012).

[61] Christodoulou, A., Babacan, S., and Liang, Z., "Accelerating cardiovascular imaging by exploiting regional low-rank structure via group sparsity," in [*Proceedings of the IEEE International Symposium on Biomedical Imaging*], 330–333, IEEE (2012).

[62] Cai, J., Jia, X., Gao, H., Jiang, S., Shen, Z., and Zhao, H., "Cine cone beam ct reconstruction using low-rank matrix factorization: algorithm and a proof-of-princple study," *arXiv preprint arXiv:1204.3595* (2012).

[63] Gao, H., Cai, J., Shen, Z., and Zhao, H., "Robust principal component analysis-based four-dimensional computed tomography," *Physics in Medicine and Biology* **56**(11), 3181 (2011).

[64] Rahmim, A., Tang, J., and Zaidi, H., "Four-dimensional (4d) image reconstruction strategies in dynamic pet: beyond conventional independent frame reconstruction," *Medical physics* **36**, 3654 (2009).

[65] Grau, V., Becher, H., and Noble, J., "Registration of multiview real-time 3-d echocardiographic sequences," *IEEE Transactions on Medical Imaging* **26**(9), 1154–1165 (2007).

[66] Zhou, X., Yang, C., and Yu, W., "Automatic mitral leaflet tracking in echocardiography by outlier detection in the low-rank representation," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], (2012).

[67] Balzano, L., Nowak, R., and Recht, B., "Online identification and tracking of subspaces from highly incomplete information," in [*Proceedings of The 48th Annual Allerton Conference on Communication, Control, and Computing*], 704–711 (2010).

[68] Mairal, J., Bach, F., Ponce, J., and Sapiro, G., "Online learning for matrix factorization and sparse coding," *The Journal of Machine Learning Research* **11**, 19–60 (2010).