Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

Beyond RPCA: Flattening Complex Noise in the Frequency Domain

Yunhe Wang,[1,3] **Chang Xu,**[2] **Chao Xu,**[1,3] **Dacheng Tao**[2]
[1]Key Laboratory of Machine Perception (MOE), School of EECS, Peking University, China
[2]Centre for Artificial Intelligence, University of Technology Sydney, Australia
[3]Cooperative Medianet Innovation Center, Peking University, China
wangyunhe@pku.edu.cn, Chang.Xu@uts.edu.au,
xuchao@cis.pku.edu.cn, Dacheng.Tao@uts.edu.au

Abstract

Discovering robust low-rank data representations is important in many real-world problems. Traditional robust principal component analysis (RPCA) assumes that the observed data are corrupted by some sparse noise (*e.g.*, Laplacian noise) and utilizes the ℓ_1-norm to separate out the noisy component. Nevertheless, as well as simple Gaussian or Laplacian noise, noise in real-world data is often more complex, and thus the ℓ_1 and ℓ_2-norms are insufficient for noise characterization. This paper presents a more flexible approach to modeling complex noise by investigating their properties in the frequency domain. Although elements of a noise matrix are chaotic in the spatial domain, the absolute values of its alternative coefficients in the frequency domain are constant *w.r.t.* their variance. Based on this observation, a new robust PCA algorithm is formulated by simultaneously discovering the low-rank and noisy components. Extensive experiments on synthetic data and video background subtraction demonstrate that FRPCA is effective for handles complex noise.

Introduction

Natural data such as that derived from video, documents, images, and social networks are often low rank. For example, in a recommendation system, the similarities between movies and between users result in a low rank rating matrix (Babacan et al. 2012; Pan et al. 2010). In another example, a video sequence can be compressed by a representation of much smaller dimensionality than the number of its pixels because of the strong correlation between consecutive video frames (Xin et al. 2015; Nie, Huang, and Ding 2012; Zhao et al. 2014).

In machine learning and data mining (Wold, Esbensen, and Geladi 1987), principal component analysis (PCA) is one of the most popular techniques for exploring the low-rank representation of data. Based on singular value decomposition (SVD), PCA seeks the best low-rank approximation of the data matrix under the ℓ_2-norm, which is exclusive to Gaussian noise and inapplicable to gross noise or outliers (Zhao et al. 2014). Therefore, a number of methods have emerged to enhance PCA's robustness over the last decade (Wright et al. 2009; Ding, He, and Carin 2011; Babacan et al. 2012; Wang et al. 2012; Zhao et al. 2014).

The robust principal component analysis algorithm (RPCA (Wright et al. 2009; Nie, Yuan, and Huang 2014)) is a representative robust improvement of the traditional PCA. Besides the low-rank component of the original data matrix, RPCA additionally employs an ℓ_1-norm to distill a sparse matrix that characterizes large error corruptions, especially Laplacian noise. Since Laplacian noise is common in natural data, RPCA performs well in many real-world problems, *e.g.*, video background subtraction (Xin et al. 2015; Babacan et al. 2012), face modeling (Li et al. 2014; Zhao et al. 2014), and subspace learning (Ding and Fu 2016).

However, the noise seen in practical applications can be more complex than simple Laplacian or Gaussian noise and, furthermore, there might be a mixture of different types of noise with distinct distributions, *e.g.*, a video sequence corrupted by two kinds of Gaussian noise and one Laplacian noise (Zhao et al. 2014). In these challenging cases, traditional PCA and RPCA are ineffective due to their limited capacity to model complex noise.

Recently, efforts have been made to broaden the RPCA algorithm to tackle complex noise (Zhao et al. 2014; Ding, He, and Carin 2011; Babacan et al. 2012). However, these methods tend to be inefficient due to the introduction of more hyper-parameters such as the number of noise distributions, variations of noise and noise level, which largely increase model complexity and redundancy.

Instead of trying to characterize each individual possible noise distribution, we aim to develop a more flexible approach to simultaneously modeling different noise distributions. Based on the observation that the frequency coefficients of *i.i.d.* noise signals in the frequency domain are constant values *w.r.t.* their variance regardless of the type of noise, we propose a new robust PCA algorithm in the frequency domain. The original data matrix can be transformed into the frequency domain and then decomposed into a low-rank component constrained by the trace norm and a noise component that is encouraged to approach some pre-estimated constants. The resulting objective function can be efficiently optimized by the augmented Lagrange multiplier technique. Experimental results demonstrate the power of the proposed algorithm to handle complex noise.

Preliminaries

We first briefly introduce some related works on RPCA. Consider the observed data $\mathbf{y}$ (*e.g.*, an image patch, a video frame, a speech signal) corrupted by some noise, *i.e.*,

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{x}$ is the desired clean data and $\mathbf{n}$ is an additive noise signal. Many methods have been developed to recover $\mathbf{x}$ from $\mathbf{y}$. Robust principal component analysis (RPCA) (Wright et al. 2009), as a representative method, decomposes the corrupted matrix $\mathbf{Y} \in \mathbb{R}^{m\times n}$ into two parts: the desired matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$ and the noise matrix $\mathbf{N} \in \mathbb{R}^{m\times n}$, where m and n are the sample dimensionality and the number, respectively. The desired clean matrix $\mathbf{X}$ is supposed to be low rank, while $\mathbf{N}$ is encouraged to be sparse. The RPCA model can be mathematically described as:

$$\min_{\mathbf{X},\mathbf{N}} ||\mathbf{X}||_* + \lambda||\mathbf{N}||_1 \quad s.t.\ \mathbf{Y} = \mathbf{X} + \mathbf{N}, \tag{2}$$

where $||\mathbf{X}||_*$ is the nuclear norm (Candès and Recht 2009) accumulating all singular values of $\mathbf{X}$, $||\cdot||_1$ denotes the ℓ_1-norm, and λ is the weight to balance two norms. Fcn.2 expects the estimated noise matrix to be sparse, which implies that the noise follows a Laplacian-like distribution. Alternatively, the ℓ_2-norm (Zhao et al. 2014) can be adopted to handle Gaussian noise:

$$\min_{\mathbf{X},\mathbf{N}} ||\mathbf{X}||_* + \lambda||\mathbf{N}||_2^2 \quad s.t.\ \mathbf{Y} = \mathbf{X} + \mathbf{N}. \tag{3}$$

However, the noise hidden in real-world data is often more complex and may be composited of multiple noise distributions, *e.g.*, different Gaussian noise with distinct means and variances.

A variety of algorithms have also been proposed to handle noise. The generative denoising model (Gu et al. 2014; Dong, Zhang, and Shi 2011) can be formulated as

$$\hat{\mathbf{X}} = \arg\min ||\mathbf{X} - \mathbf{Y}||_F^2 + \lambda\mathcal{R}(\mathbf{X}), \tag{4}$$

where $\mathcal{R}(\cdot)$ denotes some prior knowledge of $\mathbf{X}$ such as the low rank assumption (Gu et al. 2014; Dong, Shi, and Li 2013) or the sparse assumption (Dong, Zhang, and Shi 2011; Mairal et al. 2009; Xu et al. 2015). In fact, the $||\cdot||_F$ in Fcn.4 is the Frobenius norm of a matrix, equivalent to the ℓ_2-norm for vectors. By replacing $\mathbf{X} - \mathbf{Y}$ with $\mathbf{N}$, these methods act as Gaussian denoisers and can only recover data corrupted by Gaussian noise.

To recover the clean data $\mathbf{X}$, some studies have investigated the complex noise corruption problem. (Ding, He, and Carin 2011; Babacan et al. 2012) introduced Bayesian approaches to RPCA, by assuming that the noise is a sparse noise plus a dense noise, but this approach also lacks applicability to more complex mixed noise. (Zhao et al. 2014) explored a mixture of Gaussians (MoG)-RPCA approach, which had the capability to fit more complex noise via sophisticated assumptions and regularizations. However, it estimates multiple distributions simultaneously by introducing more parameters, making the model more complex.

RPCA in the Frequency Domain

As discussed above, conventional ℓ_1 or ℓ_2-norms are only effective for specific types of noise and cannot be perfectly applied to complex noise. More rigorous assumptions and formulations can be made to address the underlying noise, but more parameters are introduced as a result, and the model becomes difficult to optimize. A more effective approach is required to model complex noise.

The Discrete Cosine Transform

Generally, the energy of natural data in the frequency domain is mainly concentrated in low-frequency positions (Lam and Goodman 2000; Wang et al. 2016). In contrast, the energy of noise signals in the frequency domain is dispersed. Specifically, it is well known that the high-frequency coefficients of noisy images are significantly higher than those of natural images when converted into the frequency domain using spatial-frequency transformations, *e.g.*, FFT (Cooley and Tukey 1965), DCT (Ahmed, Natarajan, and Rao 1974), or PCA (Wold, Esbensen, and Geladi 1987).

Since the low-rank property is widespread in visual data and PCA, RPCA, and their variants are usually evaluated over visual data, we consider the denoising of visual data as an example to illustrate the proposed algorithm. Compared to FFT whose coefficients are imaginary numbers, DCT utilizes real numbers to express frequency coefficients and is an approximate KL-transformation for images. Hence we first convert each corrupted observation into the frequency domain using DCT. Considering that $\mathbf{y} \in \mathbb{R}^{m_1\times m_2}$ is a 2D image or a frame from a video sequence, its DCT frequency representation is:

$$\begin{aligned}\mathcal{C}_{\mathbf{y}_{j_1,j_2}} &= \mathfrak{D}(\mathbf{y}_{i_1,i_2}) \\ &= s_{j_1}s_{j_2}\sum_{i_1=0}^{m_1-1}\sum_{i_2=0}^{m_2-1}\mathbf{C}(i_1,i_2,j_1,j_2)\mathbf{y}_{i_1,i_2} \\ &= S_{j_1,j_2} * \mathbf{y},\end{aligned} \tag{5}$$

where m_1 and m_2 are the height and width of each image, respectively. $m = m_1 \times m_2$ is the dimensionality of each sample. $*$ is the image convolution, and $S_{j_1,j_2}(i_1,i_2) = s_{j_1}s_{j_2}\mathbf{C}(i_1,i_2,j_1,j_2)$ with the same size of $\mathbf{y}$ is generated to calculate the j_1,j_2-th DCT coefficient of $\mathbf{y}$. $\mathcal{C}_{\mathbf{y}} \in \mathbb{R}^{m_1\times m_2}$ is the DCT coefficient matrix of $\mathbf{y}$, and $s_j = \sqrt{1/n}$ if $j = 0$ and $s_j = \sqrt{2/n}$, otherwise. $\mathbf{C}(\cdot,\cdot,\cdot,\cdot)$ represents the cosine basis function:

$$\begin{aligned}&\mathbf{C}(i_1,i_2,j_1,j_2) = \\ &\quad\cos\left(\frac{\pi(2i_1+1)j_1}{2m}\right)\cos\left(\frac{\pi(2i_2+1)j_2}{2n}\right).\end{aligned} \tag{6}$$

As for the corrupted matrix $\mathbf{Y} = [\text{vec}(\mathbf{y}_1),...,\text{vec}(\mathbf{y}_n)]$, in which each column is an observed sample, we calculate its frequency coefficient matrix $\mathcal{C}_{\mathbf{Y}}$ as

$$\mathcal{C}_{\mathbf{Y}} = \mathbf{S}\mathbf{Y} = \mathbf{S}\mathbf{X} + \mathbf{S}\mathbf{N} = \mathcal{C}_{\mathbf{X}} + \mathcal{C}_{\mathbf{N}}, \tag{7}$$

where $\mathcal{C}_{\mathbf{X}}$ and $\mathcal{C}_{\mathbf{Y}}$ are the coefficient matrices of $\mathbf{X}$ and $\mathbf{Y}$, respectively, and

$$\mathbf{S} = [\text{vec}(S_{1,1}),...,\text{vec}(S_{m_1,m_2})], \tag{8}$$

and the inverse DCT can be formulated as $\mathbf{Y} = \mathbf{S}^T\mathcal{C}_\mathbf{Y} = \mathbf{S}^T\mathcal{C}_\mathbf{X} + \mathbf{S}^T\mathcal{C}_\mathbf{N}$. Note that $\mathbf{S}$ is an orthogonal matrix, *i.e.*, $\mathbf{S}^T\mathbf{S} = \mathbf{I}$, where $\mathbf{I} \in \mathbb{R}^{m\times m}$ is an identity matrix, $\mathcal{C}_\mathbf{y} = \mathcal{C}_\mathbf{x} + \mathcal{C}_\mathbf{n}$, and $\mathcal{C}_\mathbf{Y} = \mathcal{C}_\mathbf{X} + \mathcal{C}_\mathbf{N}$.

Complex Noise in the Frequency Domain

Accurately modeling each individual noise component in complex noise is difficult (Zhao et al. 2014), even if the complex noise is the mixture of two kinds of Gaussian noise. In the spatial domain, different noise distributions will produce various shapes and values. In contrast, we observe that noises following different distributions have similar characteristics in the frequency domain.

All elements in the noise matrix $\mathbf{N}$ are supposed to follow an independent and identical distribution. Corruptions can happen at any location, *e.g.*, a pedestrian or a light can appear at any location in a video (Laplacian noise) or whispers can happen in any part of audio data(Gaussian nose). We consider a generative form of complex noise

$$\mathbf{N} \sim \sum_{p=1}^{P} \pi_p \mathcal{D}_p(\mu_p, \sigma_p^2), \tag{9}$$

where P is the number of distributions and $\mathcal{D}_p$ denotes the p-th distribution; for example, Gaussian distribution, Laplacian distribution, Exponential distribution *etc*. μ_p and σ_p^2 are the expectation and the variance of $\mathcal{D}_p$, respectively, and π_p is the proportion of $\mathcal{D}_p$ in the noise signal.

It is common to fix the expectation of the noise to 0, *i.e.*, $\mu = \sum_{p=1}^{P} \pi_p\mu_p = 0$. This is because adding a constant to all elements of an arbitrary matrix does not influence its properties, and a small constant additive illumination change will not obviously change the visual quality. Thus, the variance of the mixed noise can be estimated as $\sigma^2 = \sum_{p=1}^{P} \pi_p\sigma_p^2$ given independent noise components. We omit the subscript and represent the expectation and variance of the complex noise as μ and variance σ^2.

After transfering the noise matrix $\mathbf{N}$ into the frequency domain using an orthogonal transform, its alternative coefficients (AC) are irrelevant to its expectation but only relevant to its variance. By exploiting the property of the DCT and its corresponding transformation matrix $\mathbf{S}$, we have the following theorem.

Theorem 1. *Given an arbitrary matrix $\mathbf{N} \in \mathbb{R}^{m\times n}$, assume that all of its elements are* i.i.d. *with expectation μ and variance σ^2,* i.e.*, $E[\mathbf{N}_{i,j}] = \mu$, $D[\mathbf{N}_{i,j}] = \sigma^2$. The expectation of any of its squared AC coefficients $\mathcal{C}_{\mathbf{N}}{}^2_{i,j}$, $i = 2, ..., m, j = 1, ..., n$, in the frequency domain via an orthogonal matrix $\mathbf{S}$ is equal to σ^2.*

Proof. The DCT coefficient matrix of the given matrix $\mathbf{N}$ is $\mathcal{C}_\mathbf{N} = \mathbf{SN}$, and its direct coefficients (DC) are in the first row with $\mathcal{C}_{\mathbf{N}1,j} = \frac{1}{m}\sum_{j=1}^{n} \mathbf{N}_{i,j}$, which is irrelevant to the variance of $\mathbf{N}$. Thus, the DC should not be considered in the proposed approach.

For any AC $\mathcal{C}_{\mathbf{N}i,j}$, $\mathcal{C}_{\mathbf{N}i,j} = \mathbf{S}_i\mathbf{N}(:, j)$ $(i > 1)$, where $\mathbf{N}(:, j)$ is the j-th column of $\mathbf{N}$ and $\mathbf{S}_i$ denotes the i-th row of $\mathbf{S}$ in Fcn.8. We can calculate the expectation of its squared value $\mathcal{C}_{\mathbf{N}}{}^2_{i,j}$ as

$$\begin{aligned} E\left[\mathcal{C}_{\mathbf{N}}{}^2_{i,j}\right] &= \left(E\left[\mathcal{C}_{\mathbf{N}i,j}\right]\right)^2 + D\left[\mathcal{C}_{\mathbf{N}i,j}\right] \\ &= \left(\mathbf{S}_i E\left[\mathbf{N}(:, j)\right]\right)^2 + \mathbf{S}_i D\left[\mathbf{N}(:, j)\right]\mathbf{S}_i^T \\ &= \left(\mathbf{S}_i\mu\right)^2 + \mathbf{S}_i\Sigma\mathbf{S}_i^T \\ &= \mu^T\mathbf{S}_i^T\mathbf{S}_i\mu + \mathbf{S}_i\Sigma\mathbf{S}_i^T \\ &=< \mathbf{S}_i^T\mathbf{S}_i, \mu\mu^T > + < \mathbf{S}_i^T\mathbf{S}_i, \Sigma >, \end{aligned} \tag{10}$$

where Σ is the covariance matrix of $\mathbf{N}(:, j)$. Since the expectation of $\mathbf{N}(:, j)$ is only reflected by its DC, when we only tackle ACs, the expectation of $\mathbf{N}(:, j)$ is equal to 0. Thus, the expectation of $\mathcal{C}_{\mathbf{N}}{}^2_{i,j}$ is:

$$\begin{aligned} E\left[\mathcal{C}_{\mathbf{N}}{}^2_{i,j}\right] &= 0 + < \mathbf{S}_i^T\mathbf{S}_i, \Sigma > \\ &= \sigma^2 < \mathbf{S}_i^T\mathbf{S}_i, \mathbf{I} > \\ &= \sigma^2||\mathbf{S}_i^T||_2^2 = \sigma^2, \end{aligned} \tag{11}$$

which means that the expectation of $\mathcal{C}_{\mathbf{N}}{}^2_{i,j}$ is only relevant to the variance of $\mathbf{N}$. Thus, we have $\mathcal{C}_{\mathbf{N}}{}^2_{i,j} = \sigma^2$ and $|\mathcal{C}_{\mathbf{N}i,j}| = \sigma\sqrt{2/\pi}$, $i = 2, ..., m, j = 1, ..., n$. □

Remark: $\mathbf{S}$ in Theorem 1 is an orthogonal matrix, which can take various forms such as those in PCA, FFT, or some learned orthogonal transform. Additionally, examples can be either one dimensional (audio, documents) or two-dimensional (images, videos).

According to Theorem 1, we have obtained an insightful observation for modeling noise signal $\mathbf{N}$, namely that its absolute ACs in the frequency domain should approach a constant value *w.r.t.* the noise variance. Thus, we propose the following model for estimating the desired clean data:

$$\begin{aligned} \min_{\mathbf{X},\mathbf{N}} &||\mathbf{X}||_* + \lambda||\,|\mathbf{SN}| - \xi||_F \\ s.t.\ \ &\mathbf{Y} = \mathbf{X} + \mathbf{N}, \end{aligned} \tag{12}$$

where $\mathbf{SN}$ converts the noise matrix $\mathbf{N}$ into the frequency domain. ξ is a synthetic matrix with the same size of $\mathbf{Y}$. Elements in the first row of ξ are equal to 0, which discards the DCs of $\mathbf{N}$, and other elements are equal to $\sigma\sqrt{2/\pi}$.

Optimization

$||\cdot||_*$ and $||\cdot||_F$ of Fcn.12 are calculated in the spatial and frequency domains respectively, which is inconvenient for minimization. Fortunately, the DCT is an orthogonal transform whose transformation matrix $\mathbf{S}$ is a full-rank matrix, *i.e.*, rank$(\mathbf{S}) = m$. The rank of $\mathcal{C}_\mathbf{X}$ in the frequency domain is therefore equal to the rank of $\mathbf{X}$ in the spatial domain, *i.e.*, rank$(\mathcal{C}_\mathbf{X})$ = rank$(\mathbf{SX})$ = rank$(\mathbf{X})$. Fcn.12 can then be rewritten as:

$$\begin{aligned} \min_{\mathcal{C}_\mathbf{X},\mathcal{C}_\mathbf{N}} &||\mathcal{C}_\mathbf{X}||_* + \lambda||\,|\mathcal{C}_\mathbf{N}| - \xi||_F^2 \\ s.t.\ \ &\mathcal{C}_\mathbf{Y} = \mathcal{C}_\mathbf{X} + \mathcal{C}_\mathbf{N}. \end{aligned} \tag{13}$$

Compared to the objective function of RPCA (Fcn.2), the proposed Fcn.13 escapes the optimization burden of the ℓ_1-norm by adopting an F-norm to constrain the frequency coefficients of the noise signal in the frequency domain. The

inexact augmented Lagrange multiplier (IALM (Lin, Chen, and Ma 2010)) for solving RPCA can also be applied to effectively optimize Fcn.13. As a result, RPCA complexity and convergency analyses can be simply adapted to the proposed algorithm.

By introducing the augmented Lagrange multiplier, the loss function can be defined as:

$$\begin{aligned}\mathcal{L}\left(\mathcal{C}_{\mathbf{X}},\mathcal{C}_{\mathbf{N}},E,\gamma\right)=&||\mathcal{C}_{\mathbf{X}}||_*+\lambda||\left|\mathcal{C}_{\mathbf{N}}\right|-\xi||_F^2+\\&<E,\mathcal{C}_{\mathbf{Y}}-\mathcal{C}_{\mathbf{X}}-\mathcal{C}_{\mathbf{N}}>+\frac{\gamma}{2}||\mathcal{C}_{\mathbf{Y}}-\mathcal{C}_{\mathbf{X}}-\mathcal{C}_{\mathbf{N}}||_F^2,\end{aligned}\tag{14}$$

where E is the Lagrange multiplier and γ is a weight parameter to balance the original objective function and the equality constraint. Fcn.14 can be divided into two parts and solved separately.

The loss function of the desired data $\mathcal{C}_{\mathbf{X}}$ is

$$\begin{aligned}\mathcal{L}\left(\mathcal{C}_{\mathbf{X}},E,\gamma\right)=&||\mathcal{C}_{\mathbf{X}}||_*-<E,\mathcal{C}_{\mathbf{X}}>\\&+\frac{\gamma}{2}||\mathcal{C}_{\mathbf{Y}}-\mathcal{C}_{\mathbf{X}}-\mathcal{C}_{\mathbf{N}}||_F^2,\end{aligned}\tag{15}$$

which can be simplified as

$$\begin{aligned}\mathcal{L}\left(\mathcal{C}_{\mathbf{X}},E,\gamma\right)=&\frac{1}{\gamma}||\mathcal{C}_{\mathbf{X}}||_*+\frac{1}{2}||\mathcal{C}_{\mathbf{X}}-(\mathcal{C}_{\mathbf{Y}}-\mathcal{C}_{\mathbf{N}}+\frac{1}{\gamma}E)||_F^2\\=&\frac{1}{\gamma}||\mathcal{C}_{\mathbf{X}}||_*+\frac{1}{2}||\mathcal{C}_{\mathbf{X}}-W||_F^2,\end{aligned}\tag{16}$$

where $W=\mathcal{C}_{\mathbf{Y}}-\mathcal{C}_{\mathbf{N}}+\frac{1}{\gamma}E$. The above function is a traditional low-rank minimization problem that can be directly solved using the soft-thresholding scheme (Cai, Candès, and Shen 2010):

$$\hat{\mathcal{C}}_{\mathbf{X}}=U\mathcal{S}_{\gamma^{-1}}(\mathbf{\Sigma})V^T,\tag{17}$$

where γ^{-1} is the parameter for shrinking singular values, and $W=U\mathbf{\Sigma}V^T$ is the singular value decomposition of W. $\mathbf{\Sigma}$ is a diagonal matrix whose diagonal elements $\mathbf{\Sigma}_{ii}$ are singular values of W, and

$$\mathcal{S}_{\gamma^{-1}}(\mathbf{\Sigma})_{ii}=\max(\mathbf{\Sigma}_{ii}-\gamma^{-1},0).\tag{18}$$

On the other hand, the objective function *w.r.t.* the noisy data $\mathcal{C}_{\mathbf{N}}$ is

$$\begin{aligned}\mathcal{L}\left(\mathcal{C}_{\mathbf{N}},E,\gamma\right)=&\lambda||\left|\mathcal{C}_{\mathbf{N}}\right|-\xi||_F^2-<E,\mathcal{C}_{\mathbf{N}}>+\\&\frac{\gamma}{2}||\mathcal{C}_{\mathbf{Y}}-\mathcal{C}_{\mathbf{X}}-\mathcal{C}_{\mathbf{N}}||_F^2,\end{aligned}\tag{19}$$

which can be simplified as

$$\begin{aligned}\mathcal{L}\left(\mathcal{C}_{\mathbf{N}},E,\gamma\right)=&\frac{\lambda}{\gamma}||\left|\mathcal{C}_{\mathbf{N}}\right|-\xi||_F^2+\\&\frac{1}{2}||\mathcal{C}_{\mathbf{N}}-(\mathcal{C}_{\mathbf{Y}}-\mathcal{C}_{\mathbf{X}}+\frac{1}{\gamma}E)||_F^2.\end{aligned}\tag{20}$$

Generally, the problem involving two F-norms can be minimized directly, but the absolute value function $|\cdot|$ in Fcn.20 makes the optimization extremely difficult. An alternative is to estimate signs in advance (Lee et al. 2006). Since $\mathcal{C}_{\mathbf{N}}$ in Fcn.20 is supposed to be consistent with $\mathcal{C}_{\mathbf{Y}}-\mathcal{C}_{\mathbf{X}}+\frac{1}{\gamma}E$, we define two auxiliary variables

$$Q=\text{sign}(P)\circ\xi,\tag{21}$$

Algorithm 1 Frequency RPCA for data recovery.

Input: Observed data $\mathbf{Y}$, weight parameter λ;
1: Convert $\mathbf{Y}$ into the frequency domain $\mathcal{C}_{\mathbf{Y}}\leftarrow\mathbf{SY}$;
2: Initialization: $E^0=\mathcal{C}_{\mathbf{Y}}/J(\mathcal{C}_{\mathbf{Y}})$, $\gamma_0>0$, $\rho>1$, $k=0$;
3: **repeat**
4: **Update** $\mathcal{C}_{\mathbf{X}}$ (Fcn.15):
5: $(U,\mathbf{\Sigma},V)=\text{svd}(\mathcal{C}_{\mathbf{Y}}-\mathcal{C}_{\mathbf{N}}^k+\gamma_k^{-1}E^k)$;
6: $\mathcal{C}_{\mathbf{X}}^{k+1}\leftarrow U\mathcal{S}_{\gamma_k^{-1}}(\mathbf{\Sigma})V^T$
7: **Update** $\mathcal{C}_{\mathbf{N}}$ (Fcn.19):
8: $P\leftarrow\mathcal{C}_{\mathbf{Y}}-\mathcal{C}_{\mathbf{X}}^{k+1}+\gamma_k^{-1}E^k$;
9: $Q\leftarrow\text{sign}(P)\circ\xi$;
10: $\mathcal{C}_{\mathbf{N}}\leftarrow(2\lambda Q+\gamma_k P)/(2\lambda+\gamma_k)$;
11: $E^{k+1}\leftarrow E^k+\gamma_k(\mathcal{C}_{\mathbf{Y}}-\mathcal{C}_{\mathbf{X}}^{k+1}-\mathcal{C}_{\mathbf{N}}^{k+1})$;
12: $\gamma_{k+1}\leftarrow\rho\gamma_k$, $k\leftarrow k+1$;
13: **until** convergence
14: Convert $\mathcal{C}_{\mathbf{X}}^k$ and $\mathcal{C}_{\mathbf{N}}^k$ into the spatial domain:
15: $\mathbf{X}\leftarrow\mathbf{S}^T\mathcal{C}_{\mathbf{X}}^k$, $\mathbf{N}\leftarrow\mathbf{S}^T\mathcal{C}_{\mathbf{N}}^k$
Output: Estimated data $\mathbf{X}$ and $\mathbf{N}$.

and

$$P=\mathcal{C}_{\mathbf{Y}}-\mathcal{C}_{\mathbf{X}}+\frac{1}{\gamma}E,\tag{22}$$

to reformulate Fcn.20 as:

$$\mathcal{L}\left(\mathcal{C}_{\mathbf{N}},E,\gamma\right)=\frac{\lambda}{\gamma}||\mathcal{C}_{\mathbf{N}}-Q||_F^2+\frac{1}{2}||\mathcal{C}_{\mathbf{N}}-P||_F^2,\tag{23}$$

where $\circ$ is the element-wise product. When $\lambda>0$ and $\gamma>0$, it is obvious that the solution of Fcn.23 is

$$\hat{\mathcal{C}}_{\mathbf{N}}=\frac{2\lambda Q+\gamma P}{2\lambda+\gamma}.\tag{24}$$

The variance of the noise matrix σ^2 in ξ can be regarded as a hyper parameter, or can be approximately estimated from the corrupted matrix $\mathbf{Y}$. Since the highest frequency coefficients of the desired clean matrix $\mathbf{X}$ are almost zeros, the corresponding coefficients of the corrupted data $\mathbf{Y}$ are mainly dominated by the noise. According to Theorem 1, we can estimate σ as

$$\sigma\sqrt{2/\pi}\approx\frac{1}{n}\sum_{i=1}^{n}|\mathcal{C}_{\mathbf{Y}\,m,i}|.\tag{25}$$

Finally, the multipliers are updated in Fcn.26, where $\rho>1$ is a constant.

$$E=E+\gamma(\mathcal{C}_{\mathbf{Y}}-\mathcal{C}_{\mathbf{X}}-\mathcal{C}_{\mathbf{N}}),\quad\gamma=\rho\gamma.\tag{26}$$

By solving $\hat{\mathcal{C}}_{\mathbf{X}}$ and $\hat{\mathcal{C}}_{\mathbf{N}}$ iteratively, we can optimize Fcn.12 and obtain the desired data and noisy data simultaneously. Alg. 1 summarizes the proposed approach, wherein, parameter settings and the initialization method refer to those in IALM (Lin, Chen, and Ma 2010).

Experiments

We next evaluated the performance of the proposed method and compared it with the state-of-the-art methods on synthetic data and real video data. RPCA (Wright et al. 2009),

Table 1: A comparison on the performance of FRPCA and other state-of-the-art methods.

Noise type		RPCA	VBRPCA	RegL1ALM	MoG-RPCA	FRPCA	TFRPCA
Sparse noise	PSNR(*dB*)	27.91	20.63	20.51	16.82	30.38	**30.48**
	RRE	0.0553	0.1155	0.1290	0.2249	0.0515	**0.0510**
	Time(s)	1120.29	476.53	**192.91**	769.29	707.27	728.56
Gaussian noise	PSNR(*dB*)	26.29	20.45	20.63	16.76	30.54	**30.64**
	RRE	0.0708	0.1198	0.1276	0.2333	0.0506	**0.0501**
	Time(s)	1027.78	780.86	**209.16**	955.73	740.82	743.62
Mixed Gaussian	PSNR(*dB*)	25.31	20.15	17.25	21.63	27.53	**27.58**
	RRE	0.0806	0.1263	0.2026	0.1239	0.0730	**0.0727**
	Time(s)	910.16	858.22	**140.30**	1158.12	685.78	652.97
Mixed complex	PSNR(*dB*)	24.67	19.88	17.40	21.65	26.67	**26.72**
	RRE	0.0872	0.1315	0.2032	0.1259	0.0792	**0.0786**
	Time(s)	909.30	901.84	**133.36**	951.52	416.84	400.92

VBRPCA (Babacan et al. 2012), PRMF (Wang et al. 2012), RegL1ALM (Zheng et al. 2012), and MoG-RPCA (Zhao et al. 2014) were used as comparison methods. Codes for these methods were obtained from their corresponding authors' webpages. The proposed FRPCA only has one hyper parameter λ, which was empirically set as $10 \times m^{-1/2}$ for the synthetic data and $m^{-1/2}$ for the video data, where m is the dimensionality of each sample.

Synthetic Data Simulation

Dataset generation. Several synthetic datasets were first generated to evaluate FRPCAs performance. Five famous and commonly used pictures were selected as examples: *Peppers*, *Lena*, *Butterfly*, *Baboon*, and *Cameraman*. These pictures were pre-processed to 128×128 pixels and grayscale images, as shown in Figure 1. Then, 400 109×109 images were cropped from each picture and copied 4 times, *i.e.*, each picture was converted into a $\mathbf{X} \in \mathbb{R}^{11881 \times 1600}$ matrix. Since there is strong conjunct information between several neighboring image areas, $\mathbf{X}$ has an intrinsic low-rank structure.

Figure 1: The test images.

Noise generation. We further synthetically added certain types of noise to the generated ground truth matrices: 1) sparse noise: 20% of elements were corrupted by uniform noise within $[-0.2, 0.2]$; 2) Gaussian noise: all elements were corrupted by Gaussian noise with $\mu = 0$ and $\sigma = 0.05$; 3) mixed Gaussian noise: 40% of elements were corrupted by Gaussian noise with $\mu = 0$ and $\sigma = 0.1$ and the others by Gaussian noise with $\mu = 0$ and $\sigma = 0.05$; and 4) mixed complex noise: the ground truth data were corrupted by the sparse noise and then corrupted by the mixed Gaussian noise as above.

We use the peak signal-to-noise ratios (PSNR) and the relative reconstruction error (RRE) for performance assessment. PSNR is a standard criterion for evaluating recovery, and RRE directly calculates the accumulated data error, which is calculated as:

$$\text{RRE} = \frac{||\hat{\mathbf{X}} - \mathbf{X}||_F}{||\mathbf{X}||_F}, \tag{27}$$

where $\mathbf{X}$ is the ground truth matrix, $\hat{\mathbf{X}}$ is the estimated matrix, and a smaller RRE means better performance. Moreover, we also tested the performance of replacing the traditional nuclear norm with the truncated nuclear norm, *i.e.*, $||\mathbf{X}||_{*,r} = \sum_{i=r+1}^{\min(m,n)} \sigma_i(\mathbf{X})$, where $\sigma_i(\mathbf{X})$ denotes the i-th maximum singular value of $\mathbf{X}$. The truncated nuclear norm accumulates the sum of $\min(m, n) - r$ minimum singular values, which effectively preserves the main structure of the input data. We set $r = 10$ and denote the modified FRPCA as TFRPCA. Here, RPCA is implemented by the IALM. Additionally, some of the comparison methods (VBPCA, RegL1ALM, and Mog-RPCA) must estimate the rank of the desired matrix $\mathbf{X}$ for satisfactory initialization, and each has more than two hyper-parameters. We set $\text{rank}(\mathbf{X}) = 40$ for these methods, and tuned their parameters to make results comparable. Except for the original RPCA, the other methods are sensitive to initialization and parameter settings.

Detailed results are presented in Table 1. Results were calculated by averaging PSNR and RRE values over five pictures, and the highest evaluation result in each case is highlighted in bold. Traditional RPCA still shows the best performance of the four compared methods. Mog-RPCA was designed to fit multiple noise distributions simultaneously, thus its performance on mixed complex noise is better than that on simple sparse noise and Gaussian noise. Overall, it can clearly be seen that the performance of the proposed FRPCA is superior to that of the original RPCA in the spatial domain, whether assessed by PSNR or RRE. The PSNR of FRPCA is about 2*dB* higher than those of RPCA for every noise type. TFRPCA obtains the highest accuracy of all the methods by embedding the truncated nuclear norm. It is worth noting that when we set $r = 20$, the performance of TFRPCA deteriorates since larger singular values can also be corrupted when the noise level is relatively high.

We also report the MATLAB running times for each algorithm. Although RegL1RPCA is fastest, its accuracy is poor. FRPCA and TFRPCA are clearly more efficient than the original RPCA and Mog-RPCA, since the processing of

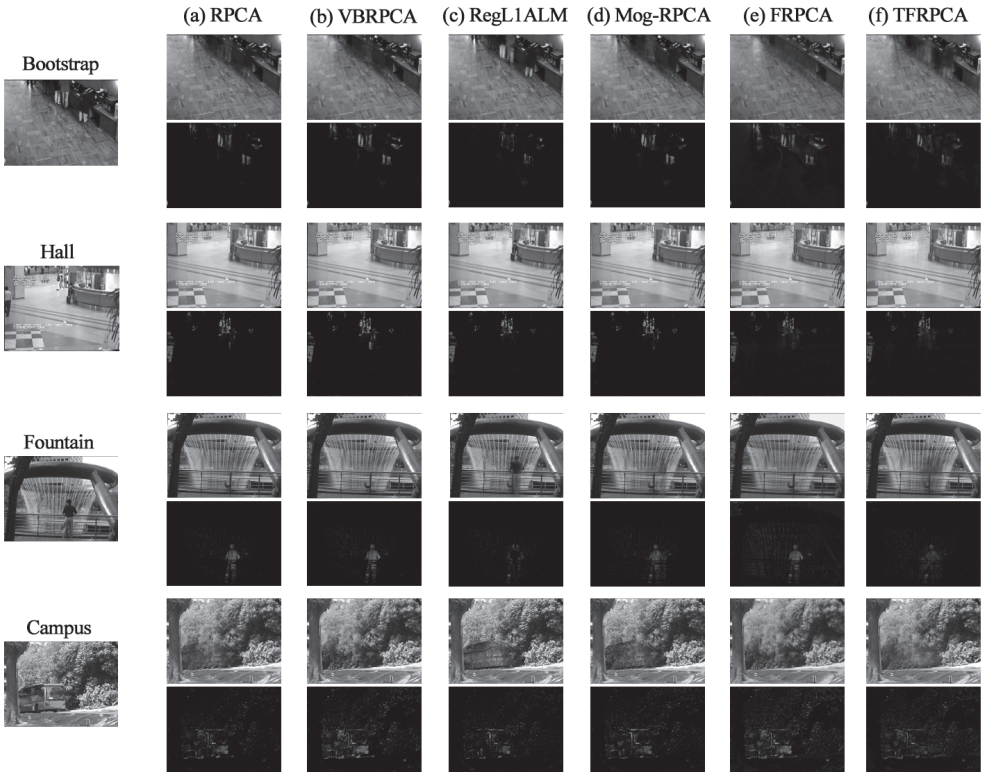

Figure 2: Results of video background subtraction using the proposed FRPCA and other state-of-the-art methods.

complex noise in the frequency domain is better at removing complex noise.

Additionally, we performed a qualitative comparison of the different recovery approaches. The images estimated by FRPCA are much more visually pleasing than those produced by the other methods. Detailed results can be found in the supplementary materials.

Video Background Subtraction

Video background subtraction is one successful application of RPCA in which a static camera captures a video sequence and a clean background is the output. Generally, RPCA assumes that the backgrounds of a given video can construct a low-rank matrix and thus burst pixels can be regarded as sparse noise such as a pedestrian, a dog, or a light. Although RPCA can effectively divide the input video into backgrounds and noisy images, some abnormal pixels are not separated. Thus, we further apply the proposed FRPCA to this task to illustrate the superiority of the proposed approach.

Four commonly utilized video sequences were selected for this experiment: *Bootstrap*, *Hall*, *Fountain*, and *Campus* (Li et al. 2004), containing 3055, 3584, 523, and 1439 frames, respectively. *Bootstrap* and *Hall* are indoor scenes, *Fountain* and *Campus* are outdoor scenes. We extracted 400 frames from Fountain and 600 frames from the other observations and then conducted background subtraction experiments individually. All the comparison methods described above were implemented, and the results are shown in Figure 2.

The background estimated by FRPCA is much clearer and visually pleasing than those produced by the other methods; only FRPCA removes all the people in the *Bootstrap* video. Moreover, although each of these methods extracts a relatively clean background, the proposed FRPCA can remove various types of noise simultaneously, *e.g.*, a moving bus, pedestrians, and the background illumination. Additionally, TFRPCAs performance with $r = 10$ is inferior to that of FRPCA because the largest singular values can also be corrupted by real-world complex noise.

Conclusions

Here we examine the data recovery problem with complex noise with complex distribution and structure. Instead of independently characterizing each individual possible noise distribution, we investigate the properties of complex noise in the frequency domain. We find that the squared values of all AC components of a noise matrix are similar, based on which a new robust PCA in the frequency domain is developed. The complexity of the proposed model will not be significantly influenced by the complexity of the noise, and IALM optimization technique can be applied for efficient solution. Experiments on synthetic data and video background subtraction show that the proposed method can perform good recovery even when the noise is extremely complex. The proposed FRPCA shows obvious advantages

over state-of-the-art approaches in terms of recovery accuracy and time cost.

Acknowledgements

We thank supports of NSFC 61375026 and 2015BAF15B00, and ARC Projects: FT-130101457, DP-140102164 and LE-140100061.

References

Ahmed, N.; Natarajan, T.; and Rao, K. R. 1974. Discrete cosine transform. *Computers, IEEE Transactions on* 100(1):90–93.

Babacan, S. D.; Luessi, M.; Molina, R.; and Katsaggelos, A. K. 2012. Sparse bayesian methods for low-rank matrix estimation. *IEEE TIP* 60(8):3964–3977.

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

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

Cooley, J. W., and Tukey, J. W. 1965. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation* 19(90):297–301.

Ding, Z., and Fu, Y. 2016. Robust multi-view subspace learning through dual low-rank decompositions. In *AAAI*.

Ding, X.; He, L.; and Carin, L. 2011. Bayesian robust principal component analysis. *IEEE TIP* 20(12):3419–3430.

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

Dong, W.; Zhang, L.; and Shi, G. 2011. Centralized sparse representation for image restoration. In *ICCV*, 1259–1266.

Gu, S.; Zhang, L.; Zuo, W.; and Feng, X. 2014. Weighted nuclear norm minimization with application to image denoising. In *CVPR*.

Lam, E. Y., and Goodman, J. W. 2000. A mathematical analysis of the dct coefficient distributions for images. *IEEE TIP* 9(10):1661–1666.

Lee, H.; Battle, A.; Raina, R.; and Ng, A. Y. 2006. Efficient sparse coding algorithms. In *NIPS*.

Li, L.; Huang, W.; Gu, I. Y.-H.; and Tian, Q. 2004. Statistical modeling of complex backgrounds for foreground object detection. *IEEE TIP* 13(11):1459–1472.

Li, Y.; Liu, J.; Li, Z.; Zhang, Y.; Lu, H.; Ma, S.; et al. 2014. Learning low-rank representations with classwise block-diagonal structure for robust face recognition. In *AAAI*.

Lin, Z.; Chen, M.; and Ma, Y. 2010. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*.

Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G.; and Zisserman, A. 2009. Non-local sparse models for image restoration. In *CVPR*.

Nie, F.; Huang, H.; and Ding, C. H. 2012. Low-rank matrix recovery via efficient schatten p-norm minimization. In *AAAI*.

Nie, F.; Yuan, J.; and Huang, H. 2014. Optimal mean robust principal component analysis. In *ICML*.

Pan, W.; Xiang, E. W.; Liu, N. N.; and Yang, Q. 2010. Transfer learning in collaborative filtering for sparsity reduction. In *AAAI*.

Wang, N.; Yao, T.; Wang, J.; and Yeung, D.-Y. 2012. A probabilistic approach to robust matrix factorization. In *ECCV*.

Wang, Y.; Shi, M.; You, S.; and Xu, C. 2016. Dct inspired feature transform for image retrieval and reconstruction. *IEEE TIP* 25(9):4406–4420.

Wold, S.; Esbensen, K.; and Geladi, P. 1987. Principal component analysis. *Chemometrics and intelligent laboratory systems* 2(1-3):37–52.

Wright, J.; Ganesh, A.; Rao, S.; Peng, Y.; and Ma, Y. 2009. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *NIPS*.

Xin, B.; Tian, Y.; Wang, Y.; and Gao, W. 2015. Background subtraction via generalized fused lasso foreground modeling. In *CVPR*.

Xu, J.; Zhang, L.; Zuo, W.; Zhang, D.; and Feng, X. 2015. Patch group based nonlocal self-similarity prior learning for image denoising. In *CVPR*.

Zhao, Q.; Meng, D.; Xu, Z.; Zuo, W.; and Zhang, L. 2014. Robust principal component analysis with complex noise. In *ICML*.

Zheng, Y.; Liu, G.; Sugimoto, S.; Yan, S.; and Okutomi, M. 2012. Practical low-rank matrix approximation under robust l 1-norm. In *CVPR*.