Think Big, Solve Small: Scaling Up Robust PCA with Coupled Dictionaries

Jian Lai   Wee Kheng Leow   Terence Sim   Vaishali Sharma
Dept. of Computer Science, National University of Singapore
{laij, leowwk, tsim, vaishali}@comp.nus.edu.sg

Abstract

Recent advances in robust principle component analysis offers a powerful method for solving a wide variety of low-level vision problems. However, if the input data is very large, especially when high-resolution images are involved, it makes RPCA computationally prohibitive for many real applications. To tackle this problem, we propose a fixed-rank RPCA method that uses coupled dictionaries (FRPCA-CD) to handle high-resolution images. FRPCA-CD downsamples high-resolution images into low-resolution images, performs FRPCA on the low-level images to obtain the low-rank matrix, which is reconstructed at original resolution by coupled dictionaries. Comprehensive tests performed on video background recovery, noise reduction in photometric stereo, and image reflection removal problems show that FRPCA-CD can reduce computation time and memory space drastically without sacrificing accuracy.

1. Introduction

Recent advances in robust PCA (RPCA) offers a powerful method for solving a wide variety of low-level vision problems such as batch alignment of images [10, 12], recovery of video background [6, 7], background modeling for high dynamic range images [10], noise reduction in photometric stereo [10], and removal of shadows, specularity [2, 18], and reflections [7] in images. RPCA solves these problems by arranging images as columns in a data matrix $\mathbf{M}$, and decomposing it into a low-rank matrix $\mathbf{A}$ that contains the desired solutions and a sparse matrix $\mathbf{E}$ that contains noise or errors. The data matrix $\mathbf{M}$ can be very large, especially when high-resolution images or long-duration videos are involved, making RPCA computationally prohibitive for many real applications. Therefore, most existing works on RPCA have restricted to low-resolution images and short-duration videos. To overcome the long-duration problem, we have previously proposed an incremental RPCA method for video background recovery that computes the low-rank and sparse matrices incrementally as new video frames arrive [6]. However, the issue of high-resolution images is not addressed in [6].

This paper proposes an fixed-rank RPCA (FRPCA) method that uses coupled dictionaries (FRPCA-CD) to handle high-resolution images. FRPCA-CD downsamples high-resolution images into low-resolution images, performs FRPCA on the low-level images to obtain the low-rank matrix, which is reconstructed at the original resolution by coupled dictionaries. Comprehensive tests performed on video background recovery, photometric stereo, and image reflection removal problems show that FRPCA-CD can reduce computational time drastically without sacrificing accuracy. In fact, FRPCA-CD is often slightly more accurate than FRPCA, and is significantly more accurate than other methods. Therefore, with coupled dictionaries, FRPCA becomes viable for many vision applications that involve high-resolution images.

2. Related Work

In low-level vision problems such as noise reduction in photometric stereo, video background recovery, and image reflection removal, the rank of the low-rank matrix $\mathbf{A}$ is known. Several methods have exploited the known rank to improve their accuracy. For example, our fixed-rank RPCA (FRPCA) methods [6, 7] constrain the rank of $\mathbf{A}$ to the known rank via exact augmented Lagrange multiplier (ALM) method. The method of [6] speeds up FRPCA by applying incremental SVD, but it is still restricted to low-resolution images. [10] minimizes partial sum of singular values instead of the nuclear norm of $\mathbf{A}$. It is implemented via inexact ALM, and its resultant $\mathbf{A}$ has a rank that is close to but not necessarily equal to the desired rank [10].

In contrast, our FRPCA-CD method adopts the fixed-rank RPCA method of [7] but changes exact ALM to inexact ALM, which is more efficient than and has comparable accuracy as exact ALM [8, 21]. Moreover, FRPCA-CD is designed to handle high-resolution images, although it can be applied to low-resolution images as well.

For video background recovery, we show that our FRPCA method [6] is more robust and accurate than methods based on video frame averaging, PCA, and mixture of Gaussian [15]. For photometric stereo, Oh et al. [10] show that their rank-aware RPCA method is more robust in removing large-amplitude noise than generic RPCA [18] and standard least squares [17]. For reflection removal, we show that our FRPCA method [7] is more robust than RPCA in removing global reflections. These methods are compared with FRPCA-CD in Section 4.

Coupled dictionaries is used by Yang et al. [20] for solving single-image super-resolution problem. They model a feature point as a sparse representation of a dictionary, and couple two corresponding feature points by two dictionaries that share the same dictionary coefficients. They solve this joint sparse coding problem using projected stochastic gradient descent. In contrast, our FRPCA-CD models an image as a linear representation of a dictionary, and couples the coefficients of two dictionaries by another linear model. It solves the coupled dictionaries problem by singular value decomposition (SVD) or FRPCA.

Dictionary learning has also been applied to PCA and spare representation methods such as [9, 13, 14]. PCA is known to be not robust to large-amplitude noise. Although there are methods that robustify PCA, such as influence function [16], alternating minimization [5], and random sampling [3], they do not guarantee optimal solutions [18]. In contrast, RPCA as defined by Wright et al. [18] has proven performance guarantee. Spare representation methods are related to RPCA but they solve different optimization problems.

3. FRPCA with Coupled Dictionaries

3.1. Robust PCA

Given a data matrix $\mathbf{M}$, robust PCA (RPCA) decomposes $\mathbf{M}$ into a low-rank matrix $\mathbf{A}$ and a sparse error matrix $\mathbf{E}$ by

$$\min_{\mathbf{A},\mathbf{E}} \|\mathbf{A}\|_* + \lambda\|\mathbf{E}\|_1, \text{ subject to } \mathbf{M} = \mathbf{A} + \mathbf{E}, \tag{1}$$

where $\|\cdot\|_*$ denotes the nuclear norm and $\|\cdot\|_1$ denotes the l_1-norm. Wright et al. [18] show that $\mathbf{A}$ can be exactly recovered if $\mathbf{A}$ is sufficiently low-rank and $\mathbf{E}$ is sufficiently sparse. This minimization problem can be solved in several ways. In particular, the augmented Lagrange multiplier (ALM) method, which reformulates Eq. 1 into

$$\min_{\mathbf{A},\mathbf{E}} \|\mathbf{A}\|_* + \lambda\|\mathbf{E}\|_1 + \langle \mathbf{Y}, \mathbf{M}-\mathbf{A}-\mathbf{E}\rangle + \frac{\mu}{2}\|\mathbf{M}-\mathbf{A}-\mathbf{E}\|_F^2, \tag{2}$$

has been shown to be among the most efficient and accurate methods [8]. In Eq. 2, $\mathbf{Y}$ contains the Lagrange multipliers, $\langle \mathbf{U}, \mathbf{V}\rangle$ is the sum of the product of corresponding elements in $\mathbf{U}$ and $\mathbf{V}$, and λ and μ are positive parameters.

Our fixed-rank RPCA (FRPCA) methods [6, 7] revise Eq. 1 to

$$\begin{aligned} &\min_{\mathbf{A},\mathbf{E}} \lambda\|\mathbf{E}\|_1, \\ &\text{subject to } \operatorname{rank}(\mathbf{A}) = \text{known } r, \mathbf{M} = \mathbf{A} + \mathbf{E}. \end{aligned} \tag{3}$$

This problem can be solved by the same ALM method, with the additional restriction of the rank of $\mathbf{A}$ to the fixed known rank r (Section 3.4). In [7], we show that FPRCA offers better performance than RPCA, especially when the noise in the data matrix is not sparse.

3.2. Overview of Proposed Method

Let $\mathbf{M}^*$ denote the data matrix that contains high-resolution images $\mathbf{m}_i^*$ arranged in columns. RPCA decomposes $\mathbf{M}^*$ into a low-rank matrix $\mathbf{A}^*$ and a sparse error matrix $\mathbf{E}^*$. Performing RPCA or fixed-rank RPCA (FRPCA) directly on $\mathbf{M}^*$ requires potentially prohibitive amounts of memory space and computation time. Instead, our fixed-rank RPCA method with coupled dictionaries (FRPCA-CD) downsamples the images $\mathbf{m}_i^*$ into low-resolution versions $\mathbf{m}_i$, forms the low-resolution data matrix $\mathbf{M}$ that contains $\mathbf{m}_i$, and decomposes $\mathbf{M}$ into the corresponding $\mathbf{A}$ and $\mathbf{E}$ by applying FRPCA. Then, the images $\mathbf{a}_i$ in $\mathbf{A}$ are used to reconstruct the high-resolution versions $\widehat{\mathbf{a}}_i$, which are assembled into the matrix $\widehat{\mathbf{A}}$. The goal is to achieve $\widehat{\mathbf{A}} = \mathbf{A}^*$.

Many high-resolution reconstruction methods are available, such as Bilinear and Bicubic interpolation. However, these methods are not ideal because it is widely known that they tend to generate overly smooth images that lose many details. FRPCA-CD, on the other hand, uses coupled dictionaries to achieve more accurate results.

Consider a low-resolution image $\mathbf{m}_i$ that is partitioned into non-overlapping image blocks $\mathbf{b}_{ij}$. The concept of coupled dictionaries is to represent $\mathbf{b}_{ij}$ with respect to dictionary matrix $\mathbf{D}_j$ and coefficient vector $\mathbf{c}_{ij}$:

$$\mathbf{b}_{ij} = \mathbf{D}_j \mathbf{c}_{ij}. \tag{4}$$

The same dictionary $\mathbf{D}_j$ applies to all blocks $\mathbf{b}_{ij}$ of images $\mathbf{m}_i$ at position j. A dictionary $\mathbf{D}_j^*$ is also defined for the corresponding high-resolution block $\mathbf{b}_{ij}^*$:

$$\mathbf{b}_{ij}^* = \mathbf{D}_j^* \mathbf{c}_{ij}^*. \tag{5}$$

As the dictionaries $\mathbf{D}_j$ and $\mathbf{D}_j^*$ are learned separately (Section 3.3), a mapping matrix $\mathbf{W}_j$ is needed to map $\mathbf{c}_{ij}$ to $\mathbf{c}_{ij}^*$:

$$\mathbf{c}_{ij}^* = \mathbf{W}_j \mathbf{c}_{ij}. \tag{6}$$

The matrix $\mathbf{W}_j$ couples dictionaries $\mathbf{D}_j$ and $\mathbf{D}_j^*$.

Now, given a low-resolution image $\mathbf{a}_i$ in the low-rank matrix $\mathbf{A}$ obtained by applying FRPCA on $\mathbf{M}$, FRPCA-CD computes the coefficients $\mathbf{c}_{ij}$ of image blocks $\mathbf{b}_{ij}$ of $\mathbf{a}_i$

using the low-resolution dictionary $\mathbf{D}_j$ (Eq. 4), and maps $\mathbf{c}_{ij}$ to $\mathbf{c}^*_{ij}$ through mapping matrix $\mathbf{W}_j$ (Eq. 6). Then, it computes the high-resolution blocks $\widehat{\mathbf{b}}^*_{ij}$ using high-resolution dictionary $\mathbf{D}^*_j$ (Eq. 5), and assembles $\widehat{\mathbf{b}}^*_{ij}$ into the resultant high-resolution images $\widehat{\mathbf{a}}_i$. So, FRPCA-CD can be summarized as in Algorithm 1.

Algorithm 1: Fixed-rank RPCA with Coupled Dictionaries (FRPCA-CD)

Input: High-resolution data matrix $\mathbf{M}^*$.
Downsample $\mathbf{M}^*$ to $\mathbf{M}$.
Compute dictionaries $\mathbf{D}_j$ and $\mathbf{D}^*_j$, and mapping matrices $\mathbf{W}_j$ (Section 3.3).
Apply FRPCA to decompose $\mathbf{M}$ to $\mathbf{A}+\mathbf{E}$ (Section 3.4).
Compute coefficients $\mathbf{c}_{ij}$ of column $\mathbf{a}_i$ in $\mathbf{A}$ (Eq. 4).
Map coefficients $\mathbf{c}_{ij}$ to $\mathbf{c}^*_{ij}$ (Eq. 6).
Reconstruct high-resolution data $\widehat{\mathbf{a}}_i$ from $\mathbf{c}^*_{ij}$ (Eq. 5).
Assemble data columns $\widehat{\mathbf{a}}_i$ into $\widehat{\mathbf{A}}$.
Output: High-resolution low-rank matrix $\widehat{\mathbf{A}}$.

3.3. Dictionary Learning

Our coupled dictionary method is inspired by [20], which links the features extracted from a pair of high- and low-resolution images by coupled dictionaries. In contrast, our FRPCA-CD learns a pair of dictionaries and a mapping matrix for each block position of all the images.

Define $\mathbf{B}_j = [\mathbf{b}_{1j} \ \cdots \ \mathbf{b}_{nj}]$ as the matrix that contains the blocks $\mathbf{b}_{ij}$ of images $\mathbf{m}_i$ at position j. A straightforward way to derive the dictionary is to apply singular value decomposition (SVD) on $\mathbf{B}_j$:

$$\mathbf{B}_j = \mathbf{U}_j \mathbf{S}_j \mathbf{V}_j^\top . \tag{7}$$

Then, comparing Eq. 7 and 4 yields

$$\mathbf{D}_j = \mathbf{U}^r_j, \quad [\mathbf{c}_{1j} \ \cdots \ \mathbf{c}_{nj}] = \mathbf{S}^r_j \mathbf{V}^{r\top}_j , \tag{8}$$

where $\mathbf{U}^r_j, \mathbf{V}^r_j$, and $\mathbf{S}^r_j$ contain the first r singular vectors and values. In the implementation, we choose r to retain 95% of the sum of singular values in $\mathbf{S}_j$.

Computing dictionary using SVD is simple and fast, but SVD is not robust to large-amplitude noise in the data. A robust alternative is to apply fixed-rank RPCA (FRPCA) to recover a rank-r matrix $\mathbf{B}^r_j$ from $\mathbf{B}_j$, along with the SVD of $\mathbf{B}^r_j$, which are used to compute the dictionary as in Eq. 8.

Given corresponding coefficients $\mathbf{c}^*_{ij}$ and $\mathbf{c}_{ij}$, the mapping matrix $\mathbf{W}_j$ can be computed by linear least squares method. $\mathbf{W}_j$ plays the role of aligning the orthogonal bases defined by the dictionaries $\mathbf{D}^*_j$ and $\mathbf{D}_j$.

In contrast, [13] and [20] assume that the high-resolution and low-resolution blocks share the same coefficients, i.e. $c_{ij} = c^*_{ij}$. This assumption can be easily violated when the numbers of columns in $\mathbf{D}_j$ and $\mathbf{D}^*_j$ are different. FRPCA-CD introduces a mapping matrix $\mathbf{W}$ to map c_{ij} to c^*_{ij} and it produces a more accurate estimation.

3.4. Fixed-Rank RPCA

When the rank r of the low-rank matrix $\mathbf{A}$ is known, fixed-rank RPCA (FRPCA) [7] can be used in place of RPCA. [7] implements FRPCA via exact augmented Lagrange multiplier (ALM) method. We adapt it and change the implementation to inexact ALM, which is more efficient than and has comparable accuracy as exact ALM. Our FRPCA method can be summarized as in Algorithm 2.

Algorithm 2: Fixed-rank RPCA (FRPCA)

Input: $\mathbf{M}, r$.
1 $\mathbf{A} = \mathbf{0}, \mathbf{E} = \mathbf{0}$.
2 $\mathbf{Y} = \mathrm{sgn}(\mathbf{M})/J(\mathrm{sgn}(\mathbf{M}))$, $\mu > 0$, $\rho > 1$, $\lambda > 0$.
3 **repeat**
4 $\quad \mathbf{U}, \mathbf{S}, \mathbf{V} = \mathrm{SVD}(\mathbf{M} - \mathbf{E} + \mathbf{Y}/\mu)$.
5 $\quad$ **if** *rank*$(T_{1/\mu}(\mathbf{S})) < r$ **then**
6 $\quad\quad \mathbf{A} = \mathbf{U}\, T_{1/\mu}(\mathbf{S}) \mathbf{V}^\top$,
7 $\quad$ **else**
8 $\quad\quad \mathbf{A} = \mathbf{U}^r\, \mathbf{S}^r\, \mathbf{V}^{r\top}$.
9 $\quad$ **end**
10 $\quad \mathbf{E} = T_{\lambda/\mu}(\mathbf{M} - \mathbf{A} + \mathbf{Y}/\mu)$.
11 $\quad \mathbf{Y} = \mathbf{Y} + \mu(\mathbf{M} - \mathbf{A} - \mathbf{E}),\ \mu = \rho\mu$.
12 **until** *convergence*;
Output: $\mathbf{A}, \mathbf{E}$.

In Line 2 of Algorithm 2, $\mathrm{sgn}(\cdot)$ computes the sign of each matrix element, and $J(\cdot)$ computes a scaling factor

$$J(\mathbf{X}) = \max\left(\|\mathbf{X}\|_2, \lambda^{-1}\|\mathbf{X}\|_\infty\right) \tag{9}$$

as recommended in [8]. The function T_ϵ in Line 5, 6 and 10 is a soft thresholding or shrinkage function [1]. The parameters μ and ρ are set to their recommended default values of $1.25/\sigma_1$ and 1.5, respectively, where σ_1 is the largest singular value of the initial $\mathbf{Y}$, whereas λ is set to the theoretical optimal value $1/\sqrt{\max(m,n)}$ [2].

4. Experiment

This section describes the applications of FRPCA-CD to three vision tasks: video background recovery, noise reduction in photometric stereo, and image reflection removal.

4.1. Data Preparation

For background recovery, we downloaded a high resolution video from Youtube captured by a stationary camera. The video was recorded in a park with people walking. 100 frames of the video were used and resized to 3840×2176. The ground truth was generated by replacing the moving human region in the first frame with the background region from the other frames.

For photometric stereo, we used the Bunny dataset in [4], which consisted of 40 images generated by Cook-Torrance reflectance model. Ground truth surface normal and lighting condition of 40 images were available. Each image had a resolution of 256×256. The average ratio of specularity and shadow were 8.4% and 24%, respectively.

For reflection removal, 46 images of size 1600×896 were captured. One image without reflection was used as the ground truth and the other 45 images with real global reflections were used as test images.

4.2. Procedure

For background recovery and reflection removal, images were arranged as columns in the data matrix $\mathbf{M}^*$. The size of $\mathbf{M}^*$ was $(3840{\times}2176){\times}100$ and $(1600{\times}896){\times}45$ for background recovery and reflection removal, respectively. Since the background in the video was stationary and images containing reflections were aligned, their corresponding matrix $\mathbf{M}^*$ had a rank of 1. The recovered rank-1 matrix $\widehat{\mathbf{A}}$ contained the stationary background and reflection-free image, respectively, for these two problems. Mean-squared error (MSE) was measured between the recovered images in $\widehat{\mathbf{A}}$ and the ground truth. For background recovery, two MSEs were calculated: one was the MSE of the whole image across all images (denoted as MSE-W); the other was the MSE of the manually marked region containing moving human of 10 regular sampling frames (denoted as MSE-R).

For photometric stereo, we adopted Lambertian model

$$\mathbf{M}^* = \mathbf{N}^\top \mathbf{L}, \tag{10}$$

where $\mathbf{M}^*$ was an $m{\times}n$ observation matrix, $\mathbf{N}$ was the $3{\times}m$ surface normal matrix, and $\mathbf{L}$ was the $3 \times n$ light direction matrix. The observations in each image were arranged as a column in $\mathbf{M}^*$, which had a rank of 3 for Lambertian model [10]. Noise was injected by randomly replacing a fraction of the pixels in each image in $\mathbf{M}^*$ by i.i.d. uniformly distributed noise. Noise ratio ranged from 10% to 50%. Therefore, photometric stereo was modeled as the RPCA problem $\mathbf{M}^* = \mathbf{A}^* + \mathbf{E}^*$, where $\mathbf{A}^* = \mathbf{N}^\top \mathbf{L}$ and $\mathbf{E}^*$ contains specularity, shadow and noise. After recovering $\widehat{\mathbf{A}}$, an approximation of $\mathbf{A}^*$, surface normals in $\mathbf{N}$ were computed by least squares method given the known lighting conditions. Mean error was measured in degree between computed surface normals and ground truth.

Test programs were implemented in Matlab and ran in a Windows 7 PC with Intel i7-5930 CPU and 32GB RAM.

Besides comparing the performance of FRPCA-CD with existing methods for the three tasks, the experiments also aimed to elucidate three aspects of FRPCA-CD:

- Downsampling rate vs. block size
  Intuitively, larger downsampling rate and larger block size result in the fewest number of blocks and thus the least amount of computation. This aspect was tested in all three tasks. Downsampling rate was set at 2, 4, 8, and 16 for background recovery and reflection removal, and 2, 4, and 8 for photometric stereo. Downsampling rate of 16 was omitted for photometric stereo because the resolution of the images was just 256×256. Block size was set to 16, 32, and 64 for all tests.

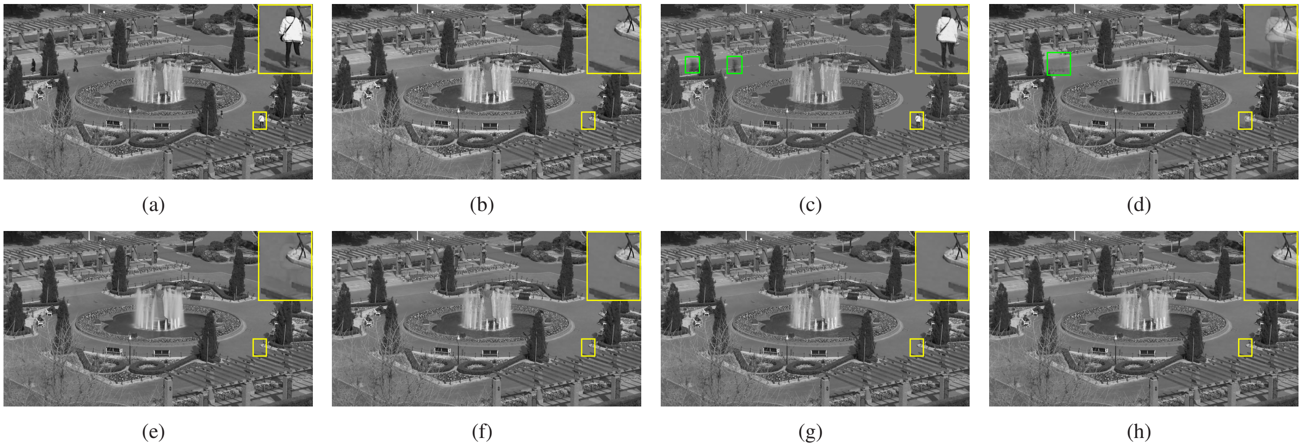
- Dictionary learning method
  FRPCA-CD has two methods of learning coupled dictionaries: by SVD (FRPCA-CD-S) and by FRPCA (FRPCA-CD-F). In principle, FRPCA is more robust than SVD in handling large-amplitude noise. This aspect was tested in the photometric stereo task at noise ratio of 10%. For background recovery and reflection removal, FRPCA-CD-S was sufficiently accurate and FRPCA-CD-F was omitted.

- High-resolution reconstruction
  Bicubic interpolation is a standard method for high-resolution reconstruction of an image. So, we compared FRPCA-CD (specifically, FRPCA-CD-S) with a variation of our method whose coupled dictionaries were replaced by bicubic interpolation (FRPCA-BI). This aspect was tested in the reflection remvoal task.

4.3. Video Background Recovery Results

Table 1 (a) and (b) illustrate the effect of block size and downsampling rate on FRPCA-CD's MSE. When both block size and downsampling rate are 16, the low-resolution block size is $16/16 = 1$, which contains too little information to learn the low-resolution dictionaries. Therefore, this test case is omitted. In Table 1 (a), all MSE-W values are similar for different combinations of parameter values. A possible reason is that most of the regions in the images are background, which overwhelm the moving human region. Compared to MSE-W, MSE-R of moving human region indeed is larger than MSE-W on the whole image, which is verified in Table 1 (b). MSE-R values vary more with different combinations of parameter value. When block size is fixed, the smallest downsampling rate always yields the least error. The best result is achieved by block size 64 and downsampling rate 2.

Table 2 shows running time for different block size and downsampling rate. It is not surprising that the least running time is achieved by the largest block size (i.e., 64) and downsampling rate (i.e., 16), which leads to the smallest



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 1. Sample test results for background recovery. (a) Sample image, (b) ground truth background, and the results of (c) MoG, (d) PCA, (e) RPCA, (f) FRPCA, (g) FRPCA-CD 64-2, and (h) FRPCA-CD 64-16. The top-right box shows the zoomed-in view in the little yellow box. The green boxes highlight shadows that are not removed.

Table 1. Background recovery results. (a) MSE-W of whole images with respect to block size and downsampling rate; (b) MSE-R of moving region with respect to block size and downsampling rate.

(a) MSE-W of whole image

block size	downsampling rate			
	2	4	8	16
16	19.48	19.4	19.52	-
32	19.35	19.52	19.53	19.4
64	**19.22**	19.34	19.73	19.33

(b) MSE-R of moving region

block size	downsampling rate			
	2	4	8	16
16	31.88	31.89	41.29	-
32	31.01	33.89	41.57	36.66
64	**28.23**	31.14	44.49	29.67

Table 2. Background recovery results. This table shows running time with respect to block size and downsampling rate.

Running time (sec.)

block size	downsampling rate			
	2	4	8	16
16	616	384	337	-
32	489	258	198	176
64	424	276	157	**140**

Table 3. Comparison of background recovery methods. This table shows MSE-W, MSE-R and running time.

MSE-W, MSE-R, and Running time (sec.)

methods	MoG	PCA	RPCA	FRPCA	FRPCA-CD	
					64-2	64-16
MSE-W	31.23	232.54	22.17	19.5	**19.22**	19.33
MSE-R	739	745	38.99	31.76	**28.23**	29.67
Time (s)	10829	**12.9**	12149	10499	423.7	140.4

number of dictionaries to be built and the lowest resolution images that are processed by FRPCA (Algorithm 1, Step 3).

Table 3 compares PCA [11], mixture of Gaussian (MoG) [15], RPCA, FRPCA [7] and FRPCA-CD. For FRPCA-CD, the parameter values that give the least MSE-R and running time are selected. They are denoted as FRPCA-CD 64-2 and FRPCA-CD 64-16. It can be seen that MoG and PCA have the largest MSE-W and MSE-R, which is consistent with the test results in [6]. Using the known rank, FRPCA outperforms RPCA. The results of FRPCA-CD methods are slightly better than those of FRPCA, even though FRPCA-CD is much faster than FRPCA.

As PCA requires only one SVD step, it has the least running time. MoG is very slow due to iterative updating of the mixture of Gaussian. RPCA and FRPCA have similar long running time because they run many SVD in their iteration steps for low rank optimization. On the other hand, FRPCA-CD is much faster than FRPCA and RPCA. It spends less than 5% of the computation time of FRPCA.

Figure 1 displays sample results for background recovery. In the results of MoG and PCA, it can be clearly seen that the moving person highlighted by yellow box has not been removed and there is shadow indicated by green boxes. RPCA, though achieves a better result, cannot completely recover the background. FRPCA and both versions of FRPCA-CD achieve the best results.

4.4. Photometric Stereo Results

Table 4 (a) and (b) compare two different methods of learning dictionary with different combination of parameters. It shows that the result of FRPCA-CD-F is

Table 4. Comparison of coupled dictionaries learning methods for photometric stereo. (a) Mean error in degree of FRCPA-CA-F and FRPCA-CD-S, (b) running time in second of FRCPA-CA-F and FRPCA-CD-S.

(a) MSE

block size	downsampling rate		
	2	4	8
16	7.47 / 9.97	8.24 / 10.77	13.11 / 14.63
32	6.31 / 10.06	6.63 / 10.95	8.42 / 12.42
64	**5.25 / 10.4**	5.27 / 10.64	6.01 / 12.07

(b) Running time (sec.)

block size	downsampling rate		
	2	4	8
16	8.05 / 1.67	6.09 / 1.03	5.15 / 0.89
32	4.01 / 1.01	3.14 / 0.55	2.66 / 0.47
64	3.07 / 0.85	2.4 / 0.42	**2.2 / 0.35**

Table 5. Noise reduction for photometric stereo. The table shows mean error and standard deviation with respect to noise ratio.

Mean Error and Standard Deviation

noise ratio	mean error (in degree)					standard deviation (in degree)				
	LS	RPCA	PSSV	FRPCA	FRPCA-CD	LS	RPCA	PSSV	FRPCA	FRPCA-CD
10%	13.39	11.42	8.66	7.77	**5.27**	8.73	7.02	5.52	4.94	**3.47**
20%	17.21	12.42	9.39	9.12	**6.46**	10.81	7.74	6.11	6.30	**4.80**
30%	21.01	13.98	10.52	10.74	**8.53**	12.60	8.87	7.26	8.02	**6.53**
40%	24.64	16.69	12.58	13.26	**11.82**	14.22	10.52	9.33	10.31	**9.06**
50%	28.15	21.34	16.55	17.41	**15.94**	15.61	12.99	12.39	13.08	**11.85**

much better than that of FRPCA-CD-S due to the large-amplitude noise. However, as it requires multiple SVD steps for dictionary building, FRPCA-CD-F is not as fast as FRPCA-CD-S.

Table 5 compares the standard LS [17], RPCA [19], partial sum of singular values (PSSV) [10], FRPCA and FRPCA-CD. For FPRCA-CD, parameter values that achieve the smallest mean error are selected. As LS is sensitive to the large amount of corruption, it performs poorly. PSSV and FRPCA significantly outperform RPCA because the sparse assumption of noise is violated. It also shows that FRPCA has better results than RPCA when noise ratio is less than 30%. However, with the facilitation of the coupled dictionaries, FRPCA-CD is more accurate than PSSV under all amounts of corruption. Moreover, FRPCA-CD achieves the smallest standard deviation.

4.5. Reflection Removal Results

Table 6 (a) and (b) show the effect of block size and downsampling rate on FRPCA-CD's MSE and running time. The MSE values of FRPCA-CD are similar for different combinations of block size and downsampling rate. Again, the least running time is achieved by the largest block size and downsamplng rate.

Table 7 compares FRPCA-BI with downsampling rates 2 and 16, FRPCA and FRPCA-CD in reflection removal problem. For FRPCA-CD, we choose block size 64 downsampling rate 2 (FRPCA-CD 64-2) and block size 64 downsampling rate 16 (FRPCA-CD 64-16). It can be seen that FRPCA-CD achieves similar result to FRPCA. Although image resolution is not high, FRPCA-CD still significantly accelerates the procedure of FRPCA. It takes only $1/10$ the execution time of FRPCA. Although FRPCA-CD is slightly slower than FRPCA-BI with the same downmpling rate, FRPCA-CD performs much more accurately than FRPCA-BI.

Figure 2 shows some test results for reflection removal. As downsampling rate is 16, we can see that FRPCA-BI recovers an overly smooth image, whereas FRPCA-CD achieves a sharp and accurate image.

4.6. Summary

For video background recovery and photometric stereo noise removal, block size of 64 and downsampling rate of 2 produce the smallest error. On the other hand, for reflection removal, block size of 64 and downsampling rate of 16 produce the lowest error. Nevertheless, other downsampling rates with block size of 64 achieve comparable accuracy. Larger block size contains more information and smaller downsampling rate loses less information. So, we can conclude that larger block size and smaller downsampling rate produce highest accuracy. On the other hand, larger block size and larger downsampling rate produces the fewest number of blocks, and achieves the least running time. This observation is consistent for all three test problems.

5. Conclusion

This paper presented a fixed-rank RPCA method with coupled dictionaries (FRPCA-CD). Instead of directly applying FRPCA to the high-resolution image data, FRPCA-CD first downsamples the original images to low resolution, which is processed by FRPCA. The output low-rank data is then used to reconstruct images at the original resolution using the proposed coupled dictionaries. Two different ways of building coupled dictionaries are

Table 6. Reflection removal results. (a) and (b) show the MSE and running time with respect to block size and downsampling rate, respetively.

(a) MSE

block size	downsampling rate			
	2	4	8	16
16	109.8	109.8	110.3	-
32	109.7	109.6	110.1	109.8
64	109.6	109.5	109.9	**109.2**

(b) Running time (sec.)

block size	downsampling rate			
	2	4	8	16
16	48.1	24.6	20.7	-
32	31.2	15.2	11.2	10.0
64	26.6	12.2	8.0	**7.2**

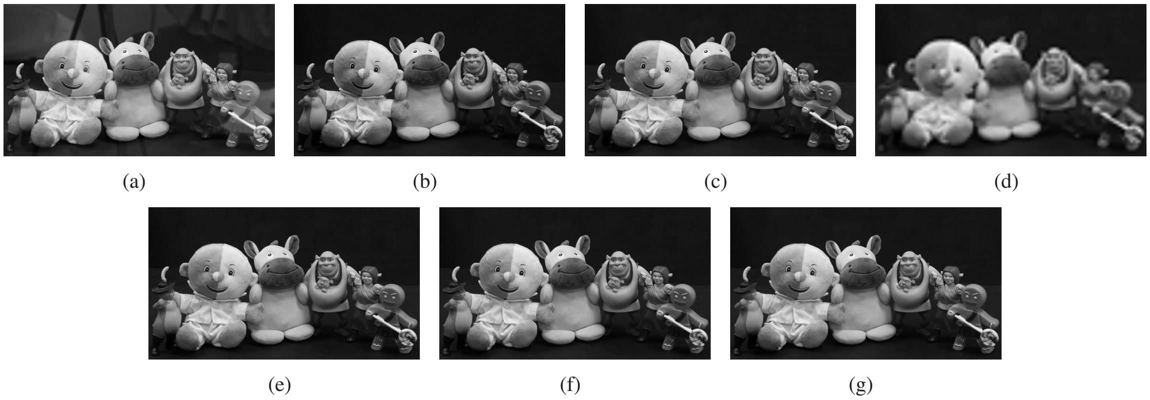


Figure 2. Sample test results for reflection removal. (a) Sample image, (b) ground truth image, and the results of (c) FRPCA-BI 2, (d) FRPCA-BI 16, (e) FRPCA, (f) FRPCA-CD 64-2, and (g) FRPCA-CD 64-16.

Table 7. Comparison of high-resolution reconstruct methods for reflection removal. This table shows MSE and running time.

MSE and Running time (sec.)

methods	FRPCA	FRPCA-BI		FRPCA-CD	
		2	16	64-2	64-16
MSE	109.7	140.62	326.43	109.59	**109.20**
Time (s)	75.17	17.53	**1.24**	26.6	7.21

proposed: by SVD and by FRPCA. FRPCA version is more robust than SVD version in dealing with large-amplitude noise.

Comprehensive tests were performed on three low-level computer vision problems: background recovery, photometric stereo and reflection removal. Test results show that FRPCA-CD uses less computation time than FRPCA without sacrificing accuracy. When the original data is high dimensional and contain redundant information, a large downsampling rate will lead to a much more efficient and sufficiently accurate result. Building dictionary with FRPCA significantly outperforms that with SVD when data is severely corrupted. Test results also show that coupled dictionaries are superior to other high-resolution reconstruction methods, such as bicubic interpolation. Therefore, with the facilitation of coupled dictionaries, FRPCA becomes viable for many vision applications that involve high dimensional data.

Acknowledgement

This research is supported by MOE grant MOE2014-T2-1-062.

References

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

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

[3] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6), June 1981.

[4] S. Ikehata, D. Wipf, Y. Matsushita, and K. Aizawa. Robust photometric stereo using sparse regression. In *Proc. CVPR*, 2012.

[5] Q. Ke and T. Kanade. Robust l1 norm factorization in the presence of outliers and missing data by alternative convex programming. In *Proc. CVPR*.

[6] J. Lai, W. K. Leow, and T. Sim. Incremental fixed-rank robust PCA for video background recovery. In *Proc. Computer Analysis of Images and Patterns*, 2015.

[7] W. K. Leow, Y. Cheng, L. Zhang, T. Sim, and L. Foo. Background recovery by fixed-rank robust principal component analysis. In *Proc. Computer Analysis of Images and Patterns*, 2013.

[8] Z. Lin, M. Chen, L. Wu, and Y. Ma. The augmented Lagrange multiplier method for exact recovery of corrupted

low-rank matrices. Technical Report UILU-ENG-09-2215, Univ. Illinois at Urbana-Champaign, 2009.

[9] M. Mohammadi, E. Fatemizadeh, and M. Mahoor. Pca-based dictionary building for accurate facial expression recognition via sparse representation. *Journal of Visual Communication and Image Representation*, 25(5):1082 – 1092, 2014.

[10] T.-H. Oh, H. Kim, Y.-W. Tai, J.-C. Bazin, and I. S. Kweon. Partial sum minimization of singular values in rpca for low-level vision. In *Proc. ICCV*, 2013.

[11] N. Oliver, B. Rosario, and A. Pentland. A bayesian computer vision system for modeling human interactions. *IEEE Trans. on PAMI*, 22(8):831–843, 2000.

[12] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Trans. on PAMI*, 34(11):2233–2246, 2012.

[13] J. G. Shi and C. Qi. Face hallucination based on PCA dictionary pairs. In *Proc. ICIP*, 2013.

[14] P. Sprechmann, A. M. Bronstein, and G. Sapiro. Learning robust low-rank representations. *CoRR*, abs/1209.6393, 2012.

[15] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. CVPR*, 1999.

[16] F. D. L. Torre and M. Black. A framework for robust subspace learning. *International Journal of Computer Vision*, 54(1-3), 2003.

[17] R. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.

[18] J. Wright, Y. Peng, Y. Ma, A. Ganesh, and S. Rao. Robust principal component analysis: Exact recovery of corrupted low-rank matrices by convex optimization. In *Proc. NIPS*, pages 2080–2088, 2009.

[19] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Proc. ACCV*, 2010.

[20] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang. Coupled dictionary training for image super-resolution. *IEEE Trans. on Image Processing*, 21(8):3467–3478, 2012.

[21] Z. Zhou, X. Li, J. Wright, E. J. Candès, and Y. Ma. Stable principal component pursuit. In *Proc. Int. Symp. Information Theory*, pages 1518–1522, 2010.