Image Restoration and Background Separation
Using Sparse Representation Framework

Liu, Shikun

Abstract

In this paper, we introduce patch-based PCA denoising and k-SVD dictionary learning method for the tasks of image restoration and background separation. We carry out an empirical evaluation of the performance of the algorithm in terms of quantity and quality. The results show that both methods appear to be competitive with the state-of-the-art image restoration and background subtraction algorithms, despite its simplicity, especially for high resolution images.

Keywords: Dictionary learning, Background separation, Sparse representation, Principal component analysis, k-SVD algorithm
Contents

1 Introduction 3

2 Dictionary Learning 3
   2.1 Predefined Dictionaries ............................................. 4
   2.2 Learned Dictionaries ................................................ 4
      2.2.1 Principal Component Analysis .................................. 4
      2.2.2 K-SVD Algorithm ................................................. 6

3 Experiments and Performance Evaluation 6
   3.1 Image Restoration .................................................. 6
   3.2 Background Subtraction ............................................ 9

4 Conclusions 10

Acknowledgements 10

References 11
1 Introduction

Olshausen & Field [1] first introduced the concept of sparse representations in 1997. After that, it becomes the focus of much research in the area of machine learning and signal processing, leading some of the famous state-of-the-art face recognition [2], handwriting recognition [3] algorithms.

Since it has been proven that images can be sparsely represented by structural primitives, sparse coding has been widely studied to solve the inverse problems in restoration, denoising, and background foreground detection applications [4] using $\mathcal{L}_0$ norm and $\mathcal{L}_1$ norm minimization technique.

The linear decomposition of a signal using a few atoms of a dictionary for either predefined or learned one has can be over-complete. Unlike decomposition based on principal component analysis and its variants, these model do not impose to be orthogonal, which offers more flexibility to adapt the natural representation of the data.

It directly follows that dictionary learning becomes essential and critical to achieve state-of-the-art results. Both predefined and learned dictionaries are introduced in the paper, however we mainly focus on learned dictionaries in this paper. Specifically, we used K-SVD and PCA method to build the dictionaries and conduct image restoration and background separation tasks. The performance achieve superior result in both quantitative and empirical analysis.

2 Dictionary Learning

Dictionary learning problem is to find a dictionary that approximates elements of a signal class using as few atoms as possible. Let a dictionary $\Phi = [\phi_1, \phi_2, \cdots, \phi_N]$ defined a $P \times N$ matrix whose columns are $\mathcal{L}_2$ norm atoms $\phi_i$, and a sparse coefficient vector $\alpha$ which most its elements are negligible. Suppose that $P$—sampled signal $x_k$ is the linear combination of $N$ components with possible Gaussian noise, it gives

$$Y = \sum_{k=1}^{N} x_k + \epsilon, \quad \sigma^2 \epsilon = \text{Var}[\epsilon] < \infty.$$ 

We need to solve the following constrained optimization problem,

$$\min_{\alpha_1, \alpha_2, \cdots, \alpha_N} \sum_{k=1}^{N} \|\alpha_k\|_0 \text{ such that } \|Y - \sum_{k=1}^{N} \Phi \alpha_k\| < \tau$$

where $\|\alpha\|_0$ is a pseudo-norm representing the number of nonzero elements in the vector. The constraint in the optimization problem accounts for the presence of noise and model
imperfection. Therefore, $\tau$ is predefined sufficiently small constant for noise compensation. If there is no noise, then the linear combination model would give a exact $\tau = 0$, then the inequality constraint can be substituted by a equality constraint.

2.1 Predefined Dictionaries

The signal transformation has been around as long as half a century ago. It became tremendously popular with the introduction of Fast Fourier Transform. Discrete Cosine/Sine transform is the result of assuming an anti-symmetric extension of the signal. Hence, it produces a smooth and continuous representation on image recovery. We include visual representations of two well-studied dictionaries below.

![Top 100 8 x 8 Patch Predefined Dictionaries](image)

Figure 1: Top 100 8 x 8 Patch Predefined Dictionaries

2.2 Learned Dictionaries

However, learning the dictionary instead of using predefined ones have shown to dramatically improve the performance of signal reconstruction [4]. Some recent results [5] on dictionary learning accessing the whole training set to minimize the cost function under some constraint. This paper uses the similar idea with low computational cost.

2.2.1 Principal Component Analysis

In the past few years, it has been shown preferable to denoise the image patch-wise instead of each pixel individually using a over-complete dictionaries which learned from noisy image.
Performing PCA [6] may achieve a good performance due to the patch redundancy property of natural images. By selecting the highest variance, PCA retrieves the most dominant pattern in the image. However, this method will have serious limitations. The first is PCA cannot represent rare patches in the images, since they contribute weakly in the total variance and they usually not sparse in the basis. Second, the natural properties of PCA will have few dominant feature spaces which truly represent the original image while the remaining feature spaces resembling noises.

![Noisy Barbara Test Image](image1)

**Figure 2:** Feature Spaces Visualization From a Noisy Test Image

In order to overcome those problems, we propose a sliding square window technique with size $W \times W$ which conduct PCA on patches with a step $\Delta = \frac{W}{2}$. Since the sliding windows can be over-lapped, it gives a higher chance to provide true noise-free candidates. The final results are the uniform average of those candidates for each pixel. We adopt this idea into color image which apply the PCA to each color channel. Then, we build our desired dictionary as follows.

**Algorithm 1** Patch-based PCA on Multichannel data

1: **Input:** $y$ : input image, $W$ : size of sliding window, $k$ : top $k$ features space
2: $[D_1,D_2,D_3] \leftarrow \text{size}(y)$
3: **for** each channel in $y$ **do**
4: \hspace{1em} **for** $i = 1 : W : D_1$ **do**
5: \hspace{2em} **for** $j = 1 : W : D_2$ **do**
6: \hspace{3em} $P_i \leftarrow \text{im}(i : i + W - 1, j : j + W - 1)$
7: $D \leftarrow P - \text{mean}(P)$
8: $\lambda_i \leftarrow \text{eig}(\text{cov}(D))$
9: **return** Top $k$ features space corresponding to $k$ largest $\lambda_i$’s
2.2.2 K-SVD Algorithm

K-SVD algorithm [7] which uses the core idea of K-Means is an iterative method which constantly update the atoms in the predefined dictionary. Since K-SVD is flexible and is able to work with any pursuit method, a lot of works [8, 9] have applied involving natural data and prove its superior performance. In this paper, we use orthogonal matching pursuit for updating atoms in each iteration.

Algorithm 2 K-SVD Method

1: Input: Y : input data, Φ : a predefined dictionary
2: Initialize a predefined dictionary Φ
3: Use OMP to solve $\Phi = \arg\min_{\alpha_1, \ldots, \alpha_k} ||Y - \sum_{k=1}^{N} \Phi \alpha_k|| < \tau$
4: for Update Φ do
5: Remove an atom $\phi_i$
6: Compute the approximation error $E_i = ||Y - \sum_{j \neq i} \Phi_i \alpha_i||$
7: Restrict $E_k$ to represent only data points which were actually using $\phi_i$
8: Compute the SVD of $E_k$
9: Replace $\phi_i$ with the first column of SVD.
10: Return: The dictionary Φ which gives the sparsest representation of data Y

Even though K-SVD performs well for in application, it is rather computationally expensive when we have a very large dimension of the dictionary.

3 Experiments and Performance Evaluation

3.1 Image Restoration

For the image restoration task, both gray-scale and color images are evaluated. We provide two qualitative evaluation measurements, Peak Signal to Noise Ratio (PSNR), and Structural SIMilarity (SSIM) [10] which are defined as

$$PSNR = 10 \log_{10} \left( \frac{255^2}{MSE} \right), \quad SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy}C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}.$$

where $MSE$ is mean square estimation of the input data, $c_1 = (0.01L)^2, c_2 = (0.03L)^2, L = 255$ to stabilize the division with weak denominator, $\mu_x, \mu_y, \sigma_x, \sigma_y$ and $\sigma_{xy}$ are local means, standard deviations and cross-covariance for input data.

The SSIM will be range in 0 to 1 which the estimated image has a similar structure to the original image if the SSIM is close to 1; the PSNR is used in quality measurement with the higher value the better quality of reconstructed data.
We apply Wiener filter, along with 2 different window-size PCA, and k-SVD with block-size 16 dictionary method to compare our results. The following gives the most corrupted images we test with Gaussian noise $\sigma = 50$.

Figure 3: Gray-Scale Test Image Restoration Visualization
We further test two sets of image, gray-scaled Barbara and color Lena using three levels of Gaussian noise with $\sigma = 10, 20, 50$ to conclude our evaluation using PSNR and SSIM measurements we mentioned above. The following table gives the complete results in the form of $[\text{PSNR} | \text{SSIM}]$ in each block.

<table>
<thead>
<tr>
<th></th>
<th>Barbera</th>
<th>Lena</th>
<th>Barbera</th>
<th>Lena</th>
<th>Barbera</th>
<th>Lena</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noise</td>
<td>28.15</td>
<td>0.7147</td>
<td>28.14</td>
<td>0.9611</td>
<td>27.00</td>
<td>0.4753</td>
</tr>
<tr>
<td>Wiener</td>
<td>28.05</td>
<td>0.7935</td>
<td>29.41</td>
<td>0.9733</td>
<td>26.23</td>
<td>0.7336</td>
</tr>
<tr>
<td>PCA16</td>
<td>28.77</td>
<td>0.8186</td>
<td>31.77</td>
<td>0.9828</td>
<td>25.85</td>
<td>0.6312</td>
</tr>
<tr>
<td>PCA32</td>
<td>27.92</td>
<td>0.7841</td>
<td>29.69</td>
<td>0.9727</td>
<td>26.14</td>
<td>0.6820</td>
</tr>
<tr>
<td>KSVD</td>
<td>33.91</td>
<td>0.9275</td>
<td>33.96</td>
<td>0.9895</td>
<td>30.16</td>
<td>0.8670</td>
</tr>
</tbody>
</table>

Table 1: Image Restoration Quantitative Evaluation In Terms Of PSNR and SSIM

From the table, we can see K-SVD gives the best result in all methods we tested above, especially for highly corrupted images. Also, PCA does a decent performance for low-corrupted images and show to be similar on different window sizes. However, if the images were over-damaged, the Wiener filter surpass the PCA method even though it washes out the images.
3.2 Background Subtraction

For background subtraction task, we use K-SVD (which shown to be best from the last task) to build our desired over-complete dictionary for background image. After that, we use orthogonal matching pursuit to find coefficient matrix which represents input foreground image. Since foreground has objects which background does not have, it is reasonable to have the high-sparsity coefficients representing the foreground objects. We use this idea to apply an unsupervised classifier K-means method to classify the coefficient data into two groups. We then call the smaller size of group foreground objects.

We use PETS 2001 datasets to create our background and foreground dataset. We crop the dataset into $200 \times 200$ color image, and we use a overlapped $10 \times 10$ and $5 \times 5$ patch to build the dictionary. We apply 30 dictionary size with 20 sparsity for patch size 10 and 20 dictionary size with 10 sparsity for patch size 5. The followings are our results.

![Background Image](image1.png) ![Foreground Image 1](image2.png) ![Detected Objects 1 (10 \times 10)](image3.png) ![Detected Objects 1 (5 \times 5)](image4.png)

![Background Image](image5.png) ![Foreground Image 2](image6.png) ![Detected Objects 2 (10 \times 10)](image7.png) ![Detected Objects 2 (5 \times 5)](image8.png)

Figure 5: Visualization of Background and Foreground Objects Using K-Means Clustering

From above, we can observe the human has been detected perfectly in patch size 10 and have some noises in patch size 5. The car object loses some details probably because the similar color between the car window and road.
4 Conclusions

In this paper, we applied PCA and K-SVD dictionary learning methods in a sparse representation framework to conduct tasks of image restoration and background subtraction. In image restoration task, we successfully recovered both gray-scale and color damaged images in the level of structure and accuracy. We have shown that K-SVD gives the best performance especially for a highly-damaged image with Gaussian noise level of $\sigma \geq 20$.

In background subtraction task, we used an unsupervised learning K-Means clustering to classify coefficient vector based on OMP algorithm. The algorithm has given superior performance when applied a large patch size but remained some noises dealing with a smaller one.

In order to solve and improve these problems, we will continue to discover and design a more powerful algorithm which would achieve better results in the future.

Acknowledgements

I would love to express my great gratitude to my project advisor Prof. Constantino Lagoa who introduced and guided me through the concept of convex and sparse optimization. Without his endless support and persistent help, this project would not have been completed.
References


