Image Denoising Algorithm Based on Non Related Dictionary Learning

Yao Nan¹, Wang KaiSheng² and Cai Yue³

1. Department of Jiangsu Electric Power Company Research Institute, NanJing Jiangsu, 211107, China
2. Department of Yangzhou Power Supply Company, YangZhou Jiangsu, 225000, China
3. Department of Nanjing Yinshi Software Co.,Ltd., NanJing Jiangsu, 210037, China
E-mail: yaon_js@aliyun.com

Abstract

In allusion to the partial texture information loss during image denoising process, an image denoising algorithm based on non related dictionary learning is proposed in this article. In this algorithm, the noise image is firstly divided into mutually overlapped image blocks, and a certain quantity of these image blocks are randomly selected for subsequent dictionary learning; then, non related dictionary learning technology is adopted to obtain the redundant dictionary with relatively strong irrelevance; finally, the sparse encoding algorithm is adopted to obtain the sparse representation coefficient of each image block in the redundant dictionary, and such sparse representation coefficients are used to recover the original image. The experiment result shows: since the redundant dictionary obtained through non related dictionary learning technology can strongly represent the image texture information, PSNR (Peak Signal to Noise Ratio) of the algorithm proposed in this article is superior to that of the existing advanced algorithm, and the algorithm can well keep the image detail and texture information, thus to improve visual effect.

Keywords: Dictionary learning; Non related; Redundant dictionary; Sparse encoding; Image denoising

1. Introduction

Digital images are usually polluted by various noise sources during the acquisition or transmission process, and the existence of noises seriously influences the effectiveness and reliability of such subsequent processing as feature extraction, target detection and identification. In order to improve image quality, it is necessary to remove such noise interference in images. Therefore, denoising issue becomes a hot research issue in the fields of computer vision and image processing.

The existing image denoising algorithms can be basically divided into three types: denoising algorithm based on spatial filtering, denoising algorithm based on transform domain filtering and denoising algorithm based on learning. Specifically, the denoising algorithm based on spatial filtering involves Gaussian filtering, bilateral filtering, guide filtering, nonlocal mean filtering, etc. [1,2], wherein the basic thought thereof is to adopt the local self-similarity or nonlocal self-similarity of the image for denoising, and such algorithm usually has the advantages of high computation efficiency, but the denoised images are usually too smooth; the denoising algorithm based on transform domain filtering involves Fourier transform, wavelet transform, BM3D algorithm, etc. [3,4], wherein the basic thought thereof is to adopt the threshold value method and the different
energy distributions of the transformed noise system and the image system to filter noises, and BM3D algorithm is adopted for image block matching in order to convert the similarly structured two-dimension image blocks into three-dimension data through 3D transformation before implementing Wiener filtering; the denoising algorithm based on learning includes K-SVD algorithm [5-7], LSSC algorithm [8] and CSR algorithm [9-13], wherein the basic thought thereof is to adopt the local sparsity of the image for denoising. In allusion to the defects of the existing image denoising algorithms, the irrelevance of the redundant dictionary atoms shall be increased to enable the redundant dictionary obtained by learning to comprehensively describe the image texture information, so an image denoising algorithm based on non related dictionary learning is proposed in this article, wherein the basic thought of this algorithm is to adopt the non related dictionary learning technology to reduce the relevance of the redundant dictionary atoms and improve the image texture information expression ability of the redundant dictionary, thus to remove noises and keep the image detail information. In this algorithm, the input noise image is firstly divided into mutually overlapped image blocks, and a certain quantity of these image blocks are randomly selected as the samples; then, the non related dictionary learning algorithm is adopted to obtain the redundant dictionary of which atoms have small relevance; finally, the efficient ROMP (Regularized Orthogonal Matching Pursuit) algorithm is adopted to find the sparse representation coefficient of each image block for recovering the original image. Compared with similar algorithms, the algorithm proposed in this article can sufficiently keep image texture and detail information to further improve the image denoising effect.

2. Dictionary Learning Model

Firstly, we simply review the dictionary learning model: for a group of given samples \( Y = (y_1, \ldots, y_K) \in \mathbb{R}^{n \times K} \), the purpose of dictionary learning is to obtain a redundant dictionary \( D = (d_1, d_2, \ldots, d_m) \in \mathbb{R}^{m \times m} \) through learning so that each sample \( y_k (k = 1, \ldots, K) \) can be expressed as sparse vector \( x_k (k = 1, \ldots, K) \) through sparse representation. The dictionary learning problem can be expressed as follows:

\[
\begin{align*}
\min_{D, X} & \quad \|Y - DX\|^2_F \\
\text{s.t.} & \quad \|d_i\|_0, \|x_k\|_0 \leq T_0 \quad \forall i, k
\end{align*}
\]  

In the above formula, \( X = (x_1, x_2, \ldots, x_K) \in \mathbb{R}^{m \times K} \) is coefficient matrix, and \( T_0 \) denotes sparsity. The above dictionary learning problem (1) can be solved by MOD algorithm, K-SVD algorithm or online dictionary learning algorithm [14].

3. Non Related Dictionary Learning Algorithm

Non related dictionary learning algorithm will be introduced in this section. The atom relevancy in the redundant dictionary is an important index for measuring the expression ability of the redundant dictionary, and the weaker atom relevancy indicates stronger dictionary expression ability. The existing dictionary learning models cannot ensure the smaller atom relevancy of the redundant dictionary obtained thereby, thus influencing the performance of the redundant dictionary.
3.1. Non Related Dictionary Learning Model

Non related dictionary learning model will be established in this section. The atom relevancy of the dictionary is defined as follows [15]:

\[
R(d_i, d_j) = \frac{\langle d_i, d_j \rangle}{\|d_i\|_2 \|d_j\|_2} = \frac{d_i^T d_j}{\|d_i\|_2 \|d_j\|_2}
\]

(2)

According to the above definition, \( R(d_i, d_j) \in [0,1] \); when \( d_i \) and \( d_j \) are orthonormal with each other, \( R(d_i, d_j) = 0 \); when \( d_i = \rho d_j \) (\( \rho \) is a nonzero constant), \( R(d_i, d_j) = 1 \).

According to the above relevancy definition (2), the relevancy of the dictionary \( D \) is defined as follows:

\[
R(D) = \sum_{i,j} R^2(d_i, d_j) = \sum_{i,j} \left( \frac{d_i^T d_j}{\|d_i\|_2 \|d_j\|_2} \right)^2
\]

(3)

Since the dictionary atom is a standard vector \( \|d_i\| = 1 \), the relevancy of the dictionary \( D \) can be simplified as follows:

\[
R(D) = \sum_{i,j} R^2(d_i, d_j) = \|D^T D\|_F^2
\]

In the original dictionary learning model (1), the dictionary relevancy constraint is increased so that the dictionary atoms can keep strong irrelevance during the dictionary learning process to improve the dictionary expression ability. On the basis of increasing the relevancy constraint of the dictionary \( D \), non related dictionary learning model can be expressed as follows:

\[
\begin{cases}
\min_{D, X} \|Y - DX\|_F^2 + \lambda \|D^T D\|_F^2 \\
\text{s.t.} \quad \|d_i\| = 1, \|x_{ik}\|_0 \leq T_0 \quad \forall i, k
\end{cases}
\]

(4)

In the above formula, parameter \( \lambda \) is a positive constant, and the second item \( \|D^T D\|_F^2 \) of the objective function can make the dictionary \( D \) learned thereby have relatively strong irrelevance.

3.2. Model Solution

In this section, the basic thought of the alternating direction method [16] is adopted to solve the non related dictionary learning model (4). The model solution process mainly includes two iteration steps: fix the redundant dictionary \( D \) and solve the sparse representation coefficient matrix \( X \); fix the sparse representation coefficient matrix \( X \) and update the redundant dictionary \( D \).

If the redundant dictionary \( D \) is fixed, then the dictionary learning model (4) can be simplified as follows:
The above optimization problem (5) can be converted into a series of sparse encoding problems:

\[ \forall k \begin{cases} 
\min_{x_k} & \|y_k - Dx_k\|_F^2 \\
\text{s.t.} & \|x_k\|_0 \leq T_0 
\end{cases} \tag{6} \]

There are many mature sparse encoding algorithms, such as OMP (Orthogonal Matching Pursuit) algorithm [17], ROMP (Regularized Orthogonal Matching Pursuit) algorithm [18] and SL0 algorithm [19]. In this article, the simple and efficient ROMP algorithm is adopted to solve the above sparse encoding problem (6) to obtain the sparse representation coefficient matrix \( X \).

If the sparse representation coefficient matrix \( X \) is fixed, then the dictionary learning model (4) can be simplified as follows:

\[ \begin{cases} 
\min_{d_i} & \|Y - DX\|_F^2 + \lambda \|D^T D\|_F^2 \\
\text{s.t.} & \|d_i\|_2 = 1, \forall i 
\end{cases} \tag{7} \]

After standardization treatment, the dictionary atoms can meet the condition \( \|d_i\|_2 = 1 \), and the above optimization problem (7) can be converted into an unrestricted optimization problem:

\[ \dot{D} = \min_D \left\{ L(D) = \|Y - DX\|_F^2 + \lambda \|D^T D\|_F^2 \right\} \tag{8} \]

The steepest descent method can be used to solve the unrestricted optimization problem. The gradient of the function \( L(D) \) is calculated as follows:

\[ \frac{\partial \|Y - DX\|_F^2}{\partial D} = \frac{\partial tr\left((Y - DX)^T (Y - DX)\right)}{\partial D} = 2(DX - Y)X^T \tag{9} \]

\[ \frac{\partial \|D^T D\|_F^2}{\partial D} = \frac{\partial tr\left(DD^T DD^T\right)}{\partial D} = 4DD^T \tag{10} \]

According to formulae (9) and (10), the gradient of \( L(D) \) is as follows:
According to above formula (11), the iteration formula for updating the dictionary $D$ is as follows:

$$D^{(n+1)} = D^{(n)} - t \frac{\partial L(D^{(n)})}{\partial D^{(n)}}$$

$$= D^{(n)} - t \left( 4D^{(n)}(D^{(n)})^T D^{(n)} + 2D^{(n)}XX^T - YX^T \right)$$

(12)

In the above formula, $t > 0$ denotes iteration step size. Formula (12) is repeatedly used for iteration till convergence in order to obtain the numerical solution $\hat{D}$ for updating the redundant dictionary $D$.

The non related redundant dictionary $D$ can be obtained through repeating the above two iteration steps, and the detailed algorithm steps are as shown in Algorithm 1.

**Algorithm 1: Non Related Dictionary Learning Algorithm**

- **Algorithm Input:** a group of samples $Y = (y_1, \cdots, y_K) \in R^{n \times K}$
- **Initialization:** standardized random dictionary $D^{(0)}$, iteration times $L$, integer $l = 1$
  - **Steps 1** Sparse Encoding: adopt ROMP algorithm to solve the sparse representation coefficient matrix $X^{(l)}$ obtained according to formula (5);
  - **Step 2** Dictionary Updating: solve the optimization problem (7) according to the iteration formula (12) to obtain $\hat{D}$, and update the redundant dictionary $D^{(l)} = \hat{D}$;
  - **Step 3** Atom Standardization: standardize the dictionary atom vector $d_{i}^{(l)} = d_{i}^{(l)}/\|d_{i}^{(l)}\|_2$;
  - **Step 4** If $l < L$ is true, then repeat Step 1~Step 3;
- **Output the Result:** redundant dictionary $D^{(L)}$.

**Algorithm Annotation:**

(a) Selection of step size $t > 0$: the step size shall meet the condition $L(D^{(n+1)}) < L(D^{(n)})$;

(b) Selection of parameter $\lambda$: the optimization problem (4) can be regarded as the multi-objective programming result obtained through linear weighted sum method, so the traditional $\alpha$—method can be used to determine parameter $\lambda$.

**4. Image Denoising Algorithm based on Non Related Dictionary Learning**

The denoising algorithm proposed in this article includes two processes: firstly, the non related dictionary learning algorithm is adopted to obtain the redundant dictionary of which atoms have relatively strong irrelevance; then, the efficient ROMP algorithm is
adopted to solve the sparse representation coefficient of each image block in the dictionary, and meanwhile such coefficient is adopted to recover the original image.

The image noise is considered as additive noise, and the observation model thereof is as follows:

\[ Y = X + \nu \]  

(13)

In the above formula, \( Y \) is the image polluted by noise, \( X \) is the original image, and \( \nu \) is the random noise.

4.1. Non Related Redundant Dictionary Learning

Firstly, the noise image \( Y \) is divided into \( K \) mutually overlapped image blocks with the size of \( \sqrt{n} \times \sqrt{n} \), and the \( k \) th image block is expressed as column vector \( y_k \in \mathbb{R}^n \); then, \( L \) image blocks are randomly selected from \( K \) image blocks as the sample set \( \{ p_l \} (l = 1, \ldots, L) \); then, the samples are trained to obtain the non related redundant dictionary \( D \in \mathbb{R}^{m \times n} \). The non related dictionary learning problem is expressed as follows:

\[
\begin{aligned}
& \min_{D, \alpha} \| P - DQ \|_F^2 + \lambda \|D^T D\|_F^2 \\
& \text{s.t.} \quad \|d_j\|_1 \|q_l\|_0 \leq T_0 \quad \forall i, l
\end{aligned}
\]  

(14)

In the above formula, \( P = \{ p_1, p_2, \ldots, p_L \} \), \( Q = \{ q_1, q_2, \ldots, q_L \} \). The non related dictionary learning algorithm is adopted to solve the optimization problem (14) to obtain the non related redundant dictionary \( D \).

4.2. Image Denoising

In this section, the non related redundant dictionary \( D \) obtained through learning is adopted to recover the original image. Specifically, ROMP algorithm is adopted to solve the sparse encoding problem to obtain the sparse representation coefficient which is used to recover the original image. In order to adopt image sparsity for denoising, it is necessary to obtain the sparse representation coefficient of each image block \( y_k \in \mathbb{R}^n \) in the dictionary \( D \). Through the non related redundant dictionary \( D \) obtained in section 3.1, the sparse encoding problem used for solving the sparse representation coefficient \( \alpha_k \) can be expressed as follows:

\[
\begin{aligned}
& \min_{\alpha_k} \| y_k - Da_k \|_2^2 \\
& \text{s.t.} \quad \|a_k\|_0 \leq T_0
\end{aligned}
\]  

(15)

In the above formula, \( D = [d_1, d_2, \ldots, d_m] \), and \( d_j \) is the \( j \) th column of the dictionary \( D \) (or the \( j \) th atom). ROMP algorithm is adopted to obtain the sparse representation coefficient \( a_k \) corresponding to the \( k \) th image block.

Through the non related redundant dictionary \( D \) obtained by learning and the sparse representation coefficient \( \alpha_k \), the \( k \) th denoised image block \( x_k = Da_k \) can be obtained.
Image block $x_k$ is jointed according to position and the overlapped parts of the image blocks are averaged to obtain the denoised integral image as follows:

$$\hat{X} = \left( \sum_{k} R_k^T R_k \right)^{-1} \left( \sum_{k} R_k^T x_k \right)$$

(16)

In the above formula, $R_k$ is used to extract the matrix of the $k$th image block. For example, $R_k Y$ denotes the $k$th image block of image $Y$.

5. Experiment Result Analysis

In this section, the image denoising experiment is adopted to verify the algorithm performance and compare the algorithm proposed in this article with K-SVD algorithm [6] and BM3D algorithm [3]. Experiment 1 is used to compare the denoising visual effects of different algorithms; Experiment 2 aims at presenting the influence of noise on algorithm performance and different reconstruction effects; Experiment 3 aims at presenting the influence of dictionary atom number on algorithm performance; Experiment 4 aims at presenting the influence of the overlapped pixels of the image block on algorithm performance.

In the following simulation, the gray level image with the size of $256 \times 256$ pixels is selected, and then 20,000 image blocks are randomly selected for dictionary learning, wherein the size of the image block is $8 \times 8$, the number of the overlapped pixels of the neighboring image blocks is 7 and the atom number of the dictionary $D$ is 1024, namely $D \in \mathbb{R}^{64 \times 1024}$. Specifically, PSNR (Peak Signal-to-Noise) is adopted to measure the algorithm reconstruction performance.

Experiment 1: Comparison of denoising visual effects of algorithms

Four images are denoised for the comparison of the denoising visual effects of three algorithms in this experiment. Specifically, Gaussian white noise with the mean variance of $\sigma = 50$ is added to the images, and the visual effects of the four denoised image are as shown in Figure 1. According to the figure, compared with K-SVD algorithm and BM3D algorithm, the image denoised by the algorithm proposed in this article has clearer image texture and is more visually approximate to the original image.
Figure 1. Comparison of Denoising Visual Effects of Three Algorithms

Experiment 2: Influence of noise on algorithm performance
This experiment aims at comparing the algorithm performances under different noise mean variances, wherein the noise is Gaussian white noise. PSNR values (the first three images in Experiment 1 are selected) of three algorithms under different noise mean variances are as shown in Table 1. According to Table 1, compared with K-SVD algorithm and BM3D algorithm, the algorithm proposed in this article has higher PSNR value and stronger noise adaptability.

Table 1. Comparison of PSNR Values of Three Algorithms

<table>
<thead>
<tr>
<th>σ</th>
<th>barbara Algorithm</th>
<th>cameraman Algorithm</th>
<th>westconcordorthophoto Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K-SVD</td>
<td>BM3D</td>
<td>in this Article</td>
</tr>
</tbody>
</table>

Experiment 3: Influence of overlapped pixel number on algorithm performance
This experiment aims at presenting the influence of the overlapped pixel number on algorithm performance, wherein the noise is Gaussian white noise with mean
variance of $\sigma = 30$, the pixel number is from 0 to 7, and other simulation conditions are not changed. PSNR values (mean value obtained from the four images in Experiment 1) of the two algorithms under the condition of different overlapped pixel numbers are as shown in Figure 2. According to Figure 2, the performances of the algorithm proposed in this article and K-SVD algorithm are both improved along with the increment of the overlapped pixel number, and when the overlapped pixel number is more than or equal to 6, the performance of the algorithm proposed in this article becomes stable.

![Figure 2. Influence of Overlapped Pixel Number](image1)

Experiment 4: Influence of Dictionary Atom Number on Algorithm Performance

This experiment aims at presenting the influence of dictionary atom number on algorithm performance, wherein the noise is Gaussian white noise with mean variance of $\sigma = 30$, the dictionary atom number is from 200 to 600, and other simulation conditions are not changed. PSNR values (mean value obtained from the four images in Experiment 1) under the condition of different dictionary atom numbers are as shown in Figure 3. According to Figure 3, the minimum atom number for the algorithm proposed in this article to reach a stable state is about 800 while that for K-SVD algorithm is 1000. Obviously, the non related dictionary obtained in this article has better expression ability.

![Figure 3. Influence of Dictionary Atom Number](image2)
6. Conclusion

In allusion to the problem that it is difficult for the existing denoising algorithms to keep image details, an image denoising algorithm based on non related dictionary learning is proposed in this article. Firstly, the non related dictionary learning algorithm is adopted to obtain the redundant dictionary; then, ROMP algorithm is adopted to find the sparse representation coefficient of each image block in the redundant dictionary, and the sparse representation coefficient is used for image denoising. Since the non related redundant dictionary can comprehensively describe the image texture and detail information, the denoising quality of the image is improved. Relevant experiment results show; compared with similar algorithms, PSNR value of the algorithm proposed in this article is superior to that of the existing advanced algorithm, and the algorithm can well keep the image detail and texture information, thus to improve visual effect.

Reference


Authors

Yao Nan. He was born in Jiangsu, China, in 1976. He Received M.S. Degrees in Computer software engineering from NanJing University, JiangSu, China, in 2005, respectively. He is currently work in Jiangsu Electric Power Company Research Institute. His research direction includes the power image intelligent technology and information technology. He has received a number of scientific and Technological Progress Award.