An Image Deblocking Method Based on Pre-classified Sparse Representation

Han Zhao #1, Xiaoguang Li #2, Li Liu #3, Li Zhuo #4

Abstract—Block-based Discrete Cosine Transform (BDCT) image compression method inevitably produces annoying blocking artifacts in the case of low-bit-rate compression, as each block is transformed and quantized independently. Blocking artifacts not only seriously affect the subjective image quality, but also affect the performance of automatic analysis. In this paper, we propose a pre-classified sparse representation based deblocking method. We combine the human visual sensitivity based classification and the sparse representation method together. For different contents, the reconstruction threshold can be adaptively adjusted. Experimental results show that the proposed method can improve the deblocking quality of the compressed images effectively.

I. INTRODUCTION

Block-based Discrete Cosine Transform (BDCT) image compression method inevitably produces annoying blocking artifacts, as each block is transformed and quantized independently. Especially in the case of high compression ratio, compressed distortion will be more obvious [1]. Blocking artifacts not only seriously affect the subjective image quality, but also affect the performance of automatic analysis.

In order to alleviate blocking problem of low bit rate compressed images, researchers have made lots of efforts [1-13]. Most of them treat the quantization distortion as a kind of noise, so that it can be removed using smoothing filter at spatial domain or transform domain. There are two drawbacks for the existing works. Firstly, as the filter smoothing image blocks, it may also over smooth the edges and details of the image. That may result from the incomplete separation between the signal and noise. Secondly, the human visual sensitivity is not adequately taken into account. The sensitivity of Human Visual System (HVS) will be different according to image contents. It is necessary to take advantage of human visual sensitivity to improve the image de-blocking quality. Recently, over-complete dictionary based sparse representation as a new signal representation theory received more attention in image denoising [14, 15]. In this paper, we propose a pre-classified sparse representation based deblocking method. In this method, the input image will be divided into different regions based on the HVS sensitivity. Then, an adaptive reconstruction threshold will be estimated for sparse de-noise. Experimental results have shown that the proposed method can achieve better results.

The remainder of this paper is organized as follows: Section II describes the proposed method in details. In Section III, experimental results are discussed. Finally, we conclude the paper in Section IV.

II. THE PROPOSED METHOD

Sparse representation has been introduced in image deblocking. It supports that the sparsity of a kind of high quality images and noises are different in a well-trained sparse space. The sparse space is usually represented by an over-complete dictionary, which is trained using a set of high quality images. Therefore, adjusting the threshold of reconstruction residual will remove the blocking artifacts. Suppose that the artifacts in a DCT based compressed image can be expressed as original image signal and the quantization noise [13,16,17], as shown in (1):

\[ Y = X + W, \]

where, \( X \) is the original image which is uncompressed, \( Y \) is distorted compressed image with blocking artifacts, \( W \) is the quantization noise. We usually take an image block as an unit, so that \( x \) represents a block (size \( n \times n \)) in original image while \( y \) represents the corresponding block in the compressed image. And then the corresponding quantization noise is \( w \). As shown in the following formula:

\[ y = x + w, \]

According to the theory of sparse representation, the process of signal sparse representation can be expressed as follow:

\[ a = \arg \min_a \|a\|_0 \quad \text{Subject to} \quad \|x - Da\|_2^2 \leq t, \]

where, \( a \) is the sparse representing coefficient of signal \( x \), \( t \) is the residual tolerance, \( D \) (\( D = [d_1, d_2, ..., d_K] \)) is an over-complete dictionary, \( d \) (size \( N \times 1 \)) is the atom in dictionary \( D \), \( \| \cdot \|_0 \) is \( l_0 \) norm. The sparse approximation problem can be efficiently solved using several available approximation techniques, including Orthogonal Matching Pursuit (OMP) [18, 19], Basis Pursuit (BP) [20, 21], FOCUSS [22], and so on to get the sparse decomposition coefficients \( a \) (\( a = [a_1, a_2, ..., a_N] \), \( N=n^2 \))
of x on D. Suppose the block x in the original image X can be decomposed sparse by over-complete dictionary D, and also can be relatively complete reconstruction, so we have:

\[ y = x + w = \sum_{m=1}^{M} a_m d_m^n + e + w \]  \hspace{1cm} (4)

Where \( d_m^n \) & A, \( M = |A| \leq K \), A represents that the corresponding \( a_m \) are non-zero coefficient term, \( e \) is the residual after sparse representation by dictionary D for \( x \). We assume that, if the sparse representation coefficients of \( y \) and \( x \) are the same by using dictionary \( D \), but the residuals of them are different, we will have:

\[ y = \sum_{m=1}^{M} a_m d_m^n + e' = \sum_{m=1}^{M} a_m d_m^n + e + w \]  \hspace{1cm} (5)

From equation (5) we can see that if \( y \) and \( x \) are represented by the same sparse coefficients, the difference between \( x \) and \( y \) can be approximated to be considered as the residuals after sparse representation. If \( x \) can be relatively accurately reconstruct then \( e \) will be very small, the main body of residual error in the right of formula (5) can be regarded as just quantization noise \( w \). Therefore, if we can estimate the residual after sparse representation and find the sparse coefficients of the distortion image on the dictionary \( D \), then, the original image will be restored [9]. But in [9], they treat the entire image as the same, and employ a same residual threshold on the whole image reconstruction. In our opinion, the quantization distortion will depend on the contents of images. Different categories of image blocks will have a certain degree of different sparsity after represented by the same over-complete dictionary.

Therefore, we propose a method to classify the input image first, then, residual thresholds for different image regions will be calculated to process the distorted image. By doing this, we can achieve the purpose of optimizing deblocking quality.

A. The Framework of Our Proposed Method

The framework of our proposed method is shown in Fig.1. Our method is divided into offline stage and online stage. The offline part includes residual thresholds calculation and dictionary training processes. In this stage, we can get 3 residual thresholds and a high quality sparse representation over-complete dictionary. The online part complete the classification and sparse deblocking processes.

B. Offline Dictionary Training

Offline sparse dictionary training steps are shown in Fig. 1 a). Given a set of training samples (blocks), we expect to find a dictionary \( D \) which satisfies that all of the signals in training sample set can be sparse represented by dictionary \( D \). We adopt high quality images without compression distortion to training an over complete sparse dictionary. We firstly randomly extracted \( N_p \) image blocks from these images, with each block’s size is \( n \times n \), then expand every block into column vector (size \( N \) (\( N = n^2 \))). Therefore, we can get a training sample set, an \( N \times N_p \) matrix. Then, the obtained sample set can be used for dictionary iterative training. We use K-SVD algorithm [23] to iteratively solve dictionary. When using K-SVD method to iterate, in addition to the need to enter the training sample set we also need a dictionary as the starting iteration. We choose DCT over-complete dictionary as the initial dictionary \( D_0 \), and the dictionary is obtained by the DCT transformation [24]. Given the sequence \( X(n) = 0, 1...N-1 \), the discrete cosine transform is:

\[ X_c(0) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) \]  \hspace{1cm} (6)

\[ X_c(k) = \frac{2}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) \cos \left( \frac{(2n+1)kr}{2N} \right) \]  \hspace{1cm} (7)

Written in matrix form:

\[ X_c = C_N x \]  \hspace{1cm} (8)

where, \( C_N \) is \( N \times N \) transform matrix whose row vectors are cosine basis. Then use the fractional frequency method to expand the dictionary to which obtained after the DCT transform to over-complete. More specifically, using the DCT transformation dictionary to do more sophisticated traversal and frequency sampling to obtain a new over-complete dictionary. After we get the initial dictionary \( D_0 \) and training set, we can use the K-SVD method to train dictionary.

There are 70 natural images in [9], together with 5 images in Kodak dataset standard image library are used as the sample set. The size of blocks is \( 8 \times 8 \). Randomly extract 100000 blocks to do the sparse dictionary training. The finally training set is \( 64 \times 100000 \), dictionary training iteration times is 40. The obtained training dictionary size is \( 64 \times 512 \). Training resulting dictionary as shown in Fig. 2, where the Fig. 2 a) is the initial dictionary \( D_0 \) and b) is the trained dictionary \( D \). The data in dictionary is normalized, we conduct data expansion processing and block arrangement for the data in dictionary to facilitate the display. Each small block is arranged by an atom of the dictionary, the block is arranged in an inverse manner process to expand the block.

C. Visual Sensitivity Based Image Classification

In recent years, the characteristics of human visual sensitivity have been extensively studied. These researching results prove that, the HVS has different sensitivity to blocking artifacts in different regions [25]. In [26], they have tested the sensitivity of human eyes to blocking artifacts in different bit-rate compressed images. The results prove that, human visual system is very sensitive to the artifacts in low frequency flat areas where the brightness change is small. In addition to the flat areas, eyes are also sensitive to the artifacts at edge texture areas of an image. The visual masking effects come up in high frequency texture areas with complex structures, such as grass, which means human visual system hard to perceive obvious artifacts in these areas. We employ the classification method in [26] to divide image into different kinds of regions. Sobel operator is employed for edge detection. By computing gradient magnitude, an input image is divided into three different kinds of regions: Flat region, Edge region and Texture region. The specific classification steps are list as follows:

Step 1: Calculate the threshold \( TH_1 = \zeta_1 \times g_{max} \) and \( TH_2 = \zeta_2 \times g_{max} \) where \( g_{max} \) is the maximum gradient value. Where \( TH_1, TH_2 \) are the high and low level of the threshold respectively, as a basis for pixel texture classification, \( \zeta_1 \) and \( \zeta_2 \) are 0.12 and 0.06, respectively, they are both recommended values.
respectively. The following rule is employed to determine the category of a block.

\[
\begin{align*}
\text{edge block} & \quad (\text{NumE} > 16) \& (\text{NumE} > \text{NumT}) \\
\text{flat block} & \quad (\text{NumF} > \text{NumE}) \& (\text{NumF} > \text{NumT}) \\
\text{texture block} & \quad \text{others}
\end{align*}
\]  

(10)

D. Calculate Residual Tolerance Parameters

In the block decomposition step described in (3), the optimization problem is a NP-Hard problem. Usually, we can use \( l^1 \)-norm minimization constraint to approximately solve the problem as (11):

\[
\min_{\mathbf{a}} ||\mathbf{a}||_1 \quad \text{Subject to} \quad ||\mathbf{y} - D\mathbf{a}||_2^2 \leq T, \quad (11)
\]

where \( y \) is a column vector of the image block, \( a \) is the sparse representation coefficient. Here the difference between different fault-tolerant parameter \( T \) directly affects deblocking effect. We set a content adaptive threshold \( T \) as (12).

\[
T = T_{\text{base}} + T_{\text{adjust}}, \quad (12)
\]

where \( T_{\text{base}} \) is calculated refer to [9], which is estimated according to the compression parameters \( q \) and the standard deviation of the entire image as shown in (13).

\[
T_{\text{base}} = C \cdot \sigma_N \cdot \left( \frac{a}{q+b} + c \right), \quad (13)
\]

where \( C, a, b, c \) are constants that obtained via data fitting, \( \sigma_N \) is the standard deviation of compressed image.

The \( T_{\text{adjust}} \) is our proposed content adaptive threshold. For different categories, we have estimated different optimal values for \( T_{\text{adjust}} \) as shown in (14):

\[
T_{\text{adjust}}(x_p) = \begin{cases} 
\Delta t_1, x_p & \text{edge block} \\
\Delta t_2, x_p & \text{flat block} \\
\Delta t_3, x_p & \text{texture block}
\end{cases}
\]  

(14)

To find a suitable fault tolerance threshold for different types of contents in (11), we have conducted several tests in this section. Based on the training images, we have tested the PNSR performance of reconstructed images respect to different \( \Delta t_1, \Delta t_2, \Delta t_3 \) and the compression parameter \( q \). The test results are shown in Fig. 3-5. In the testing, we set a range of \( \Delta t_1, \Delta t_2, \Delta t_3 \) changed in step 2 and fixed \( q \), to find the values that achieve the best average PSNR. Then, for different \( q \), we repeat the testing, to obtain result in Fig. 3-5. From these tests, we find that for a fixed \( q \), the curves are approximate parabola opening down to obtain maximum value at an apex. Fig. 3 a), shows the image deblocking PSNR result respect to the change of \( \Delta t_1 \) when \( q = 15 \). We test the case of \( q = 5-30 \), the optimal parameters have been shown in Fig. 3 b)–Fig. 5 b) . We can observe that \( \Delta t_1 \) fluctuates around the mean respect to \( q \). In Fig. 3-8 b), the mean of the best parameters are shown in the horizontal line. Therefore, we choose the mean curve as the value of \( T_{\text{adjust}} \). In our experiments, we set \( \Delta t_1 = 36.125 \), \( \Delta t_2 = 24.5 \) and \( \Delta t_3 = 25.2 \).

E. Online Classification and Sparse Representation

Deblocking

Firstly, an input compressed image will be tiled into overlapped \( n \times n \) blocks with step 1, and each block is classified into one of the three categories. Then, the block are decompressed using (11) with threshold in (12).
constructed block is obtained via sparse coefficients and the over-complete dictionary $D$.

We select lena and 20 images from Kodak dataset as a testing set. The test images are not included in the training set. In order to obtain the distorted images, the high quality images are JPEG compressed with different quality parameters, we set $q$ to 5, 10, 15 and 20 respectively.

In order to measure the objective performance, the average PSNR of testing images is shown in Fig. 6 and Tab. 1. We can see that the proposed method can achieve better average PSNR than the compared methods.

The subjective performance are shown in Fig. 7. From the left to right, the first column are original images, the second column are JPEG compressed images, Pre-classification based pixel-domain method results are list at the third column, the results of [9] and our method are list in the fourth and fifth column respectively. It can be seen that the visual quality of our method has certain advantage than the compared methods. Our proposed method can adaptively protected the details according to human visual characteristics.

### III. EXPERIMENTS AND DISCUSSION

To solve (11), a fast orthogonal matching pursuit algorithm [23] is employed in our implementation. Then, the resulting block can be reconstructed by using the sparse coefficients and the dictionary. The block is extracted via overlap scheme with the 1 pixel step.

To solve (11), a fast orthogonal matching pursuit algorithm [23] is employed in our implementation. Then, the resulting block can be reconstructed by using the sparse coefficients and the dictionary. The block is extracted via overlap scheme with the 1 pixel step.

To evaluate the performance of our proposed method, we have implemented it with two compare methods on Matlab. The configuration of computer is Intel Core 3.3.GHz CPU with 4GB memory. We use ordinary natural images to train sparse dictionary. We have acquired 70 natural images from literature [9], together with 5 images from Kodak dataset to establish the training sample library. Therefore, a total of 75 images are employed as training set. There are 100,000 blocks which were extracted for sparse dictionary training. The size of dictionary is set to $64 \times 512$.

<table>
<thead>
<tr>
<th>Compression quality factor</th>
<th>Average PSNR(dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>JPG</td>
</tr>
<tr>
<td>$q=5$</td>
<td>24.47</td>
</tr>
<tr>
<td>$q=10$</td>
<td>27.07</td>
</tr>
<tr>
<td>$q=15$</td>
<td>28.44</td>
</tr>
<tr>
<td>$q=20$</td>
<td>29.37</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENT

The work in this paper is supported by the National Natural Science Foundation of China (No.61471013, No.61372149 and No.61370189), the Importation and Development of High-Caliber Talents Project of Beijing Municipal Institutions (No. CIT&TCD201404043, No. CIT&TCD20150311 and No.CIT&TCD201304036.), the Program for New Century Excellent Talents in University (No.NCET-11-0892), Specialized Research Fund for the Doctoral Program of Higher Education (No.20121103110017), the Natural Science Foundation of Beijing (No.4142009), the Science and Technology Development Program of Beijing Education Committee (No.KM201410005002), Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality.

REFERENCES

Fig. 7. The Comparison of deblocking subjective results.