ADAPTIVE MULTI-DIMENSION SPARSITY BASED COEFFICIENT ESTIMATION FOR COMPRESSION ARTIFACT REDUCTION

Jing Mu[1], Xinfeng Zhang[2], Ruiqin Xiong[1], Siwei Ma[1], and Wen Gao[1]

[1] Institute of Digital Media, Peking University, Beijing 100871, China
[2] Rapid-Rich Object Search (ROSE) Lab, Nanyang Technological University, Singapore
Email: {jmu, rqxiong, swma, wgao}@pku.edu.cn, xfzhang@ntu.edu.sg

ABSTRACT

Sparsity has shown promising results in various image restoration applications. Recent advances have suggested that structured or group sparsity often leads to more powerful results in compression artifact reduction studies. In this paper, we introduce nonlocal multi-dimension sparsity in an adaptive space-transform domain, which performs multi-scale wavelet transform on DCT coefficients of similar patches. The new transform efficiently reduces image redundancies between inner block and inter block simultaneously, thus it can substantially achieve sparse representation for images. Furthermore, a band-based filter is proposed to reduce compression artifacts by shrinking transform coefficients adaptively. Because of the overlapped processing, adaptive aggregation is used to combine different estimates for each block. The proposed algorithm achieves improvement over some methods in terms of both objective and subjective qualities.

Index Terms— Compression artifact reduction, Multi-dimension sparsity, Adaptive shrinkage, Adaptive transform domain

1. INTRODUCTION

Block-based Discrete Cosine Transform (BDCT) has been widely used in the existing compression standards including JPEG, MPEG, H.264 and HEVC. However, due to the fact that the image is independently transformed and quantized by block, the compressed image often suffers from severe degradation. In order to reduce compression artifacts, various researches and studies have been performed in recent years.

One of the most traditional ways is filtering approaches to smooth the block boundaries directly. Iterative image recovery algorithms were proposed using the traditional projection onto convex sets (POCS) [1, 2]. In [3], a Gaussian filter is used to process the pixels around block boundaries to smooth out the artifacts. This kind of method has a low cost and can be easily implemented, but cannot preserve edges and details well. Takeda et al. [4] proposed a denoising method based on a steering kernel regression framework according to the signal information. Also, total variation [5], and Markov random field (MRF) [6] were utilized as image prior models to seek the MAP estimation of the original image. These methods only consider the smoothness in pixel domain, but ignore the compressed DCT coefficients in the bit stream.

The methods in [7] directly processed compressed images in DCT domain to reduce the artifacts. Choy et al. [8] estimates the original DCT coefficients from the quantized ones with the local mean and variance of the coefficients in each sub-band. Recently, Zhang et al. proposed to utilize image block similarity prior model to reduce compression artifacts by the coefficients estimated from non-local blocks [9, 10].

Although these methods can achieve good performances, they exploited a fixed domain (DCT) which is signal independent. Considering the fact that the sparse domain of natural image varies spatially, [11, 12] try to restore images in an adaptive transform domain. BM3D [11] generates remarkable results by applying a 3D transform on a group of similar image patches. However, the signal-independent hard-thresholding of coefficients sometimes produces low-frequency noise and edge ringing. Based on the idea of group sparsity in BM3D, in this paper, we enforce a nonlocal multi-dimension sparsity via DCT and wavelet transform as shown in Fig. 1. DCT is used to reduce the redundancy of inner block. We use wavelet transform to reduce redundancy of inter block in multi-scale. DCT and multi-scale wavelet transform form the adaptive multi-dimension transform domain. The multi-dimension transform offers a powerful mechanism of characterizing the structured sparsities of natural images. To make our approach tractable and robust, a band-based wavelet shrinkage technique that is dependent on the signal is developed. Because of the overlapped processing, different estimates for each pixel need to be combined. Adaptive aggregation is used as a particular averaging procedure. Experimental results show that our approach achieves noticeable improvement in terms of both objective and subjective qualities of the reconstruction images.

The remainder of the paper is organized as follows. Section 2 utilizes the concept of multi-dimension sparsity and shows how this type of sparsity is incorporated into the framework of artifact reduction. Section 3 gives the implemen-

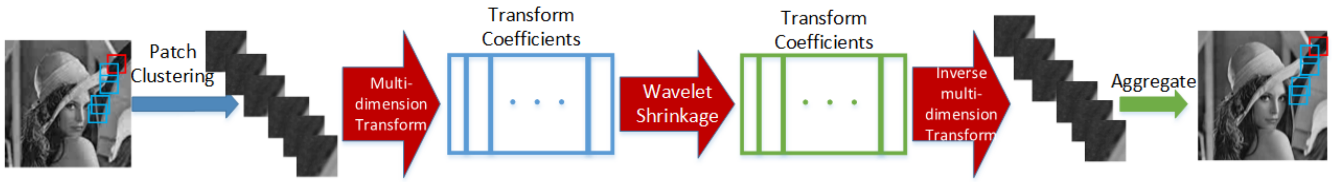


Fig. 1. The flowchart of the proposed multi-dimension sparsity based compression artifact reduction method.

tation details of the framework, and introduces the adaptive wavelet shrinkage to solve the objective problem. Experimental results are reported in Section 4. Section 5 concludes this paper.

2. MULTI-DIMENSION SPARSITY IN TRANSFORM DOMAIN

In this section, we establish an adaptive transform domain, which exploits nonlocal multi-dimension sparsity. This transform can depict the self-similarity of natural images, and retain the sharpness effectively.

2.1. Problem Formulation

We use $\mathbf{x}$ to represent the image, and use $\mathbf{x_i}$ and $\mathbf{X_i}$ to represent pixel intensity and the transform coefficients of a block, which is located in position i of size $B_s \times B_s$. The BDCT compressed image can be modeled as follows,

$$\mathbf{X_i} = \mathbf{T}\left(\mathbf{x_i}\right) \rightarrow \mathbf{Y_i} = \mathbf{Q}\left(\mathbf{X_i}\right) \rightarrow \mathbf{y_i} = \mathbf{T}^{-1}\left(\mathbf{Y_i}\right) \tag{1}$$

where $\mathbf{T}$ is DCT transform, and $\mathbf{Q}$ is the operation of quantization. $\mathbf{Y_i}$ and $\mathbf{y_i}$ are reconstructed coefficients and pixel intensity of block i. Then we have the following equation:

$$\mathbf{y_i} = \mathbf{x_i} + \mathbf{e_i} \tag{2}$$

where $\mathbf{e_i}$ is the compression error for the image patch $\mathbf{x_i}$.

2.2. Nonlocal Multi-dimension Sparsity in Transform Domain

Motivated by the great success of sparse representation [13] and self-similarity [14] in image restoration, we introduce a nonlocal multi-dimension sparsity in transform domain. We divide the image $\mathbf{x}$ into N overlapped patches. For each patch $\mathbf{x_i}$, a group of similar patches $\mathbf{Z_{x_i}}$ is collected. The entire group is formed into a data matrix. According to the self-similarity property of natural images, all the non-local similar patches have similar underlying structures. Then multi-dimension transform $\mathbf{T}^{\mathrm{MD}}$ is used to reduce redundancy of the matrix.

$\mathbf{T}^{\mathrm{MD}}$ is the combination of a 2D DCT and a multi-scale wavelet transform, which is illustrated in Fig. 2. DCT reduces the redundancy of inner block, but ignores the correlation of DCT coefficients between similar patches, i.e. the redundancy of inter block. 2D DCT of each block encourages the alignment of sparse coefficients along the column direction only. In other words, it does not treat the row and column spaces equally. We introduce multi-scale wavelet transform to reduce the redundancy between similar patches and get more sparse coefficients. Section 3 will show how to use the prior into decoded image reconstruction. Therefore, the nonlocal multi-dimension sparsity in transform domain is written as,

$$\left\| \mathbf{\Theta_x} \right\|_1 = \sum_{i=1}^{n} \left\| \mathbf{T}^{\mathrm{MD}}\left(\mathbf{Z_{x_i}}\right) \right\|_1 \tag{3}$$

3. COMPRESSION ARTIFACT REDUCTION VIA MULTI-DIMENSION SPARSITY IN TRANSFORM DOMAIN

Using Eq. (3) as a prior of image, the proposed optimization problem for our method is formulated as:

$$\mathbf{x} = \arg\min_{\mathbf{x}} \left\| \mathbf{x} - \mathbf{y} \right\|_2^2 + \alpha \left\| \mathbf{\Theta_x} \right\|_1 \tag{4}$$

where $\mathbf{x}$ is original image, and $\mathbf{y}$ is the compressed image.

The optimization problem in Eq. (4) is non-convex and quite difficult to solve directly because of the non-differentiability and non-linearity of the sparsity term. Section 3.2 shows an efficient solution based on adaptive wavelet shrinkage, which is one of the main contributions of this paper.

3.1. Wavelet Shrinkage for l^1-Minimization

The frequencies of pixels appearing in the overlapping patches are roughly equal, hence

$$\left\| \mathbf{x} - \mathbf{y} \right\|_2^2 \approx c \left\| \mathbf{x_i} - \mathbf{y_i} \right\|_2^2 \tag{5}$$

where c is a positive constant. Considering the unitary property of $\mathbf{T}^{\mathrm{MD}}$, then we rewrite Eq. (4) as Eq. (6).

$$\mathbf{x} = \arg\min_{\mathbf{x}} \sum_i \left(\left\| \mathbf{W}\left(\mathbf{T}^{2D}\left(\mathbf{Z}_{\mathbf{x}_i}\right)\right) - \mathbf{W}\left(\mathbf{T}^{2D}\left(\mathbf{Z}_{\mathbf{y}_i}\right)\right) \right\|_2^2 \right) + \alpha \sum_i \left\| \mathbf{W}\left(\mathbf{T}^{2D}\left(\mathbf{Z}_{\mathbf{x}_i}\right)\right) \right\|_1 \quad (6)$$

where $\mathbf{T}^{2D}$ is DCT matrix and $\mathbf{W}$ is multi-scale wavelet transform matrix. Using mentioned above can Eq. (6) as follows,

$$\mathbf{x} = \arg\min_{\mathbf{x}} \| \mathbf{\Theta_x} - \mathbf{\Theta_y} \|_2^2 + \alpha \| \mathbf{\Theta_x} \|_1 \quad (7)$$

Eq. (7) is a standard $l^1 - l^2$ problem, which can be solved by soft-thresholding [15]. Based on Eq. (1), we rewrite Eq. (6) as Eq. (8)

$$\mathbf{x} = \arg\min_{\mathbf{x}} \sum_i \left(\| \mathbf{W}\left(\mathbf{Z}_{\mathbf{X}_i}\right) - \mathbf{W}\left(\mathbf{Z}_{\mathbf{Y}_i}\right) \|_2^2 \right) + \alpha \sum_i \left(\| \mathbf{W}\left(\mathbf{Z}_{\mathbf{X}_i}\right) \|_1 \right) \quad (8)$$

For each block i, we can get Eq. (9):

$$\hat{\mathbf{Z}}_{\mathbf{X}_i} = \max\left(\left|\mathbf{Z}_{\mathbf{X}_i}\right| - \tau, 0\right) \cdot \operatorname{sgn}\left(\mathbf{Z}_{\mathbf{X}_i}\right) \quad (9)$$

where τ is threshold of different band for block i. Then $\mathbf{x}$ is gained by inverse wavelet transform and 2D DCT.

3.2. Bandwise Threshold Estimation

In many methods, the threshold is an empirical and universal rule, which is independent on the signal. However, this threshold is not suitable for the multi-dimension transform coefficients which are highly sparse. We will provide an adaptively chosen threshold dependent on signal.

As demonstrated in [15], under the assumption of Laplacian prior, according to Bayes theory, the rule for choosing the threshold τ is

$$\tau = 2\sqrt{2}\sigma_e^2 \Big/ \sigma_{\mathbf{W}\left(\mathbf{Z}_{\mathbf{X}_i}\right)} \quad (10)$$

σ_e^2 is the mean square error of compression coefficients. The compression error, which can be approximated as the uniform distribution, is caused by quantizing the coefficients of a sensed image to a number of discrete levels. According to the expectation of uniform distribution, we can get σ_e^2

$$\sigma_e^2 = \frac{1}{12} Q^2 \quad (11)$$

where $Q = [q_{i,j}]_{i,j=1,\ldots,8}$ is the quantization table.

In addition, $\sigma_{\mathbf{W}\left(\mathbf{Z}_{\mathbf{X}_i}\right)}$ denotes the variance of signal coefficients. For simplicity, we exploit 2D wavelet transform. For band k of DCT coefficients, we can get four levels, HH^k, HL^k, LH^k, LL^k, so we estimate $\sigma_{\mathbf{W}\left(\mathbf{Z}_{\mathbf{X}_i}\right)}$ separately, i.e. for

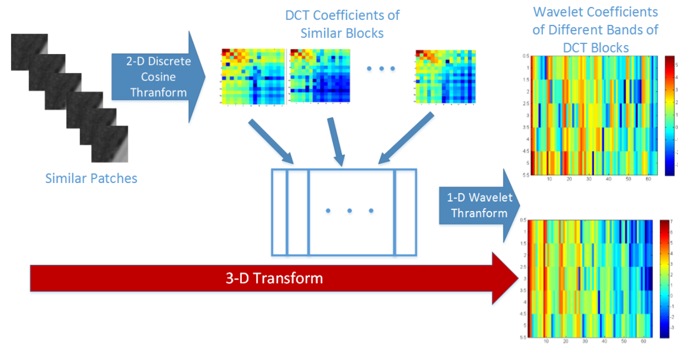


Fig. 2. Diagram of multi-dimension transform. Inverse transform is the mirror flow of this process. (We take 1-D wavelet as an example).

each band k, $\sigma^k_{\mathbf{W}\left(\mathbf{Z}_{\mathbf{X}_i}\right)} = \left\{\sigma^k_{\mathrm{HH}}, \sigma^k_{\mathrm{HL}}, \sigma^k_{\mathrm{LH}}, \sigma^k_{\mathrm{LL}}\right\}$. Because the number of the samples of each wavelet level is limited, the variance of coefficients cannot be calculated accurately. We approximate the covariance using the maximum absolute coefficients of each band. Taking coefficients of HH level of band k as an example, we can get Eq. (12)

$$\sigma^k_{\mathrm{HH}} = \max\left(abs\left(\mathbf{W}_{\mathrm{HH}}\left(\mathbf{Z}^k_{\mathbf{x}_i}\right)\right)\right) \quad (12)$$

According to Eq. (10), we can get adaptive of different levels of wavelet transform of each DCT band.

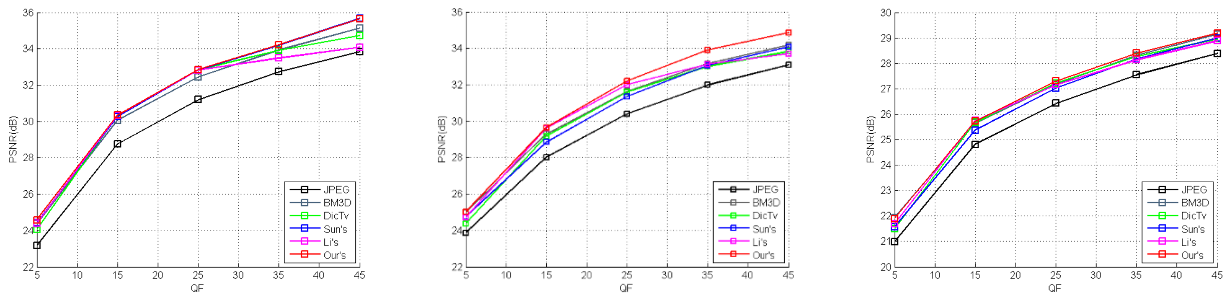
3.3. Global Estimation by Aggregation and Quantization Constraint

Due to overlapping of blockwise estimation, more than one block-estimate can be located at exactly the same coordinate. Hence, aggregation is performed by a weighted averaging at those pixel positions where there are overlapping blockwise estimates.

In general, the blockwise estimates have different variance for each pixel. However, it is quite demanding to take into consideration all these effects. Aggregation weights are inversely proportional to the total sample variance of the corresponding blockwise estimates. That is, noisier block-wise estimates have smaller weights. The mean square error of transform coefficients is as follows,

$$\begin{aligned} E\left[\left(\hat{\mathbf{X}} - \mathbf{X}\right)^2\right] &= E\left\{\left[\max\left(\left|\mathbf{Z_x}\right| - \tau, 0\right) \cdot \operatorname{sgn}\left(\mathbf{Z_x}\right) - \mathbf{X}\right]^2\right\} \\ &= \begin{cases} E\left(e^2 + \mathbf{X}^2\right) & -\tau \le \mathbf{X} \le \tau \\ E\left(e^2 + \tau^2\right) & \text{else} \end{cases} \\ &\le \sigma_e^2 + E\left(\tau^2\right) \end{aligned} \quad (13)$$

where $\hat{\mathbf{X}}$ is the estimation of original DCT coefficients $\mathbf{X}$,



(a) Vessels (b) Barbara (c) Straw

Fig. 3. PSNR vs. JPEG QF of Different Methods.

which is get from Eq. (9). So we assign the weight as

$$w = \frac{1}{\left(\sigma_e^2 + \|\tau\|^2\right)} \tag{14}$$

Finally, for each block, we can get a data matrix reconstructed by the local patch clustering. Then the processed data matrix is reformed into image patches and relocated to their original positions by w to average all of the estimates. Each of the estimated block coefficients should be in the quantization interval $\left[\mathbf{Y_i}^{\min}, \mathbf{Y_i}^{\max}\right]$

$$\mathbf{Y_i}^{\min} = \mathbf{Y_i} - \frac{Q}{2} \qquad \mathbf{Y_i}^{\max} = \mathbf{Y_i} + \frac{Q}{2} \tag{15}$$

Table 1. PSNR Results of Six Compression Artifact Reduction Algorithms When QF = 35 (UNIT: DB)

Image	JPEG	Li's	BM3D	DicTv	Sun's	Ours
Vessels	32.73	33.49	34.00	33.91	34.20	**34.21**
Barbara	31.98	33.10	33.25	33.01	33.06	**33.92**
Straw	27.54	28.15	28.31	28.26	28.17	**28.39**
Avg	30.75	31.58	31.85	31.72	31.82	**32.16**

Table 2. SSIM Results of Six Compression Artifact Reduction Algorithms when QF = 35

Image	JPEG	Li's	BM3D	DicTv	Sun's	Ours
Vessels	0.954	0.963	0.965	0.962	0.970	**0.971**
Barbara	0.927	0.932	0.941	0.928	0.940	**0.947**
Straw	0.908	0.914	0.911	0.914	0.906	**0.921**
Avg	0.930	0.936	0.939	0.935	0.939	**0.946**

4. EXPERIMENTAL RESULTS

In this section, we evaluate the efficiency of the proposed approach and compare it with some recently presented methods. To make our comparative study more thoroughly, we exploit both conventional images (256×256) and biomedical images (96×96).

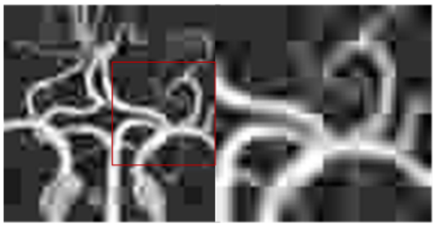
4.1. Parameter Setting

The experiments are implemented on MATLAB platform. In our implementation, parameters of the proposed method including L, B_s, and c are set empirically, and are set fixed for all test images. Concretely, the size of each block, i.e., B_s, is set to be 8×8, the size of training window for searching similar patches, i.e. , $L \times L$ is set to be 41×41, and the number of best matched blocks, i.e.$L \times L$, N is set to be 16 when $QF<25$, and N is set to be 32 when $QF \geq 25$. Since multi-dimension transform coefficients may not strictly conform to Laplacian distribution, the threshold τ is adjusted by a factor δ so as to get the best possible performance of shrinkage. δ is empirically chosen, only depending on the value of σ_e^2.

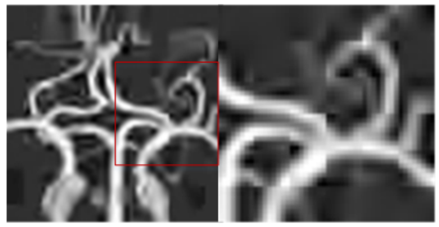
4.2. Experimental Comparisons

The test images, Barbara, Straw and Vessels, are compressed by JPEG at different QFs. Our approach is compared with four representative methods in literature, i.e., deblocking methods, i.e. Sun's [16] and Li's method [17], and denoising methods, i.e. the classic BM3D [11] and DicTv [18]. It is worth emphasizing that Suns method is known as one of the state-of-the-art algorithms for compression artifact reduction, and BM3D is regarded as the state-of-the-art algorithms for denoising.

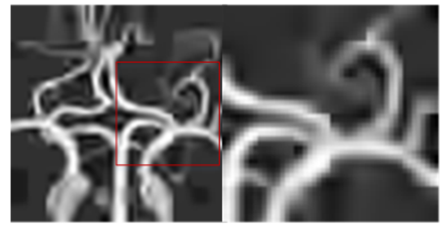
Fig. 3 plots the PSNR curves of compared algorithms on the test images, which are coded by JPEG compression standard with five quality factors (QF): 5, 15, 25, 35 and 45 respectively. Quality factors, which range from 1 to 100, are indexes of a set of quantization matrixes. The larger QF values, the less compression error. The PSNR curves clearly show that the proposed technique achieves the best reconstruction performance for all test images on all quality factors. Take $QF = 35$ as an example, PSNR and SSIM are listed in Table 1 and Table 2, where we have highlighted the best method.



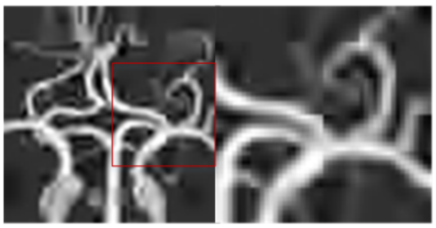
(a) JPEG (PSNR=24.30dB)



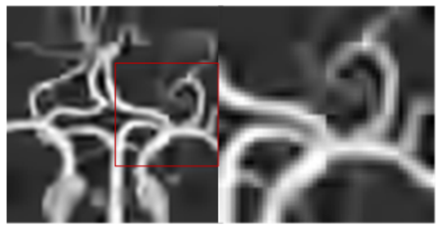
(b) Li's (PSNR=23.08dB)



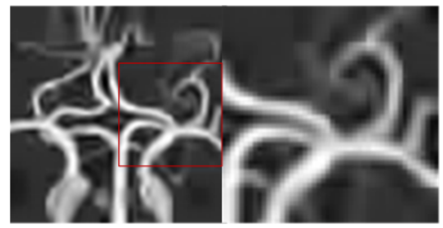
(c) BM3D(PSNR=24.37dB)



(d) DicTv (PSNR=24.05dB)



(e) Sun's (PSNR=24.36dB)



(f) Ours (PSNR=24.59dB)

Fig. 4. Visual Comparison of Reconstructed Images (QF=5).

The best average PSNR of the proposed algorithm are found to outperform all competitors. In addition to its objective performance, the new JPEG reconstruction technique also appears to obtain better perceptual quality of the test images. The processed results of the image reconstruction algorithms for the test images Vessels and Barbara with $QF = 5$ and $QF = 15$ are shown in Fig. 4 and Fig. 5. In order to observe the visual quality, we compare whole image and details of Vessels. Obviously, our proposed method preserves both sharp edges and smooth areas, and can show much clearer and better visual results than the other competing methods.

5. CONCLUSIONS

In this paper, a multi-dimension sparsity is introduced, which efficiently characterizes the intrinsic sparsity of natural images in an adaptive space-transform domain. In order to reduce redundancy of DCT coefficients gained from non-local similar patches, we use multi-scale wavelet transform to reduce the redundancy of inter DCT blocks. In addition, wavelet shrinkage is used to solve the problem. And the threshold is adaptively chosen based on the variance of different band coefficients. Last but not least, because of the overlap of pathces, adaptive aggregation is used to combine different estimations for each patch. Experimental results show that the proposed method achieves a remarkably better reconstruction quality than recently presented methods in both the subjective and the objective quality.

6. REFERENCES

[1] Avideh Zakhor, "Iterative procedures for reduction of blocking effects in transform image coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 2, no. 1, pp. 91–95, 1992.

[2] Shigenobu Minami and Avideh Zakhor, "An optimization approach for removing blocking effects in transform coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 5, no. 2, pp. 74–82, 1995.

[3] Howard C Reeve III and Jae S Lim, "Reduction of blocking effects in image coding," *Optical Engineering*, vol. 23, no. 1, pp. 230134–230134, 1984.

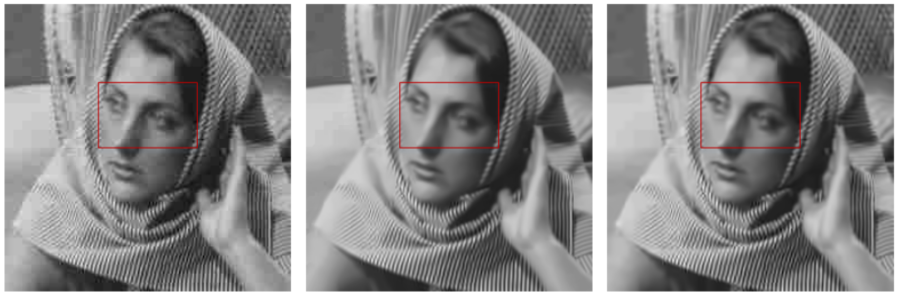
[4] Hiroyuki Takeda, Sina Farsiu, and Peyman Milanfar, "Kernel regression for image processing and reconstruction," *Image Processing, IEEE Transactions on*, vol. 16, no. 2, pp. 349–366, 2007.

[5] H. Liu, R. Xiong, S. Ma, X. Fan, and W. Gao, "Non-local extension of total variation regularization for image restoration," in *IEEE International Symposium on Circuits and Systems*, Jun. 2014, pp. 1102–1105.

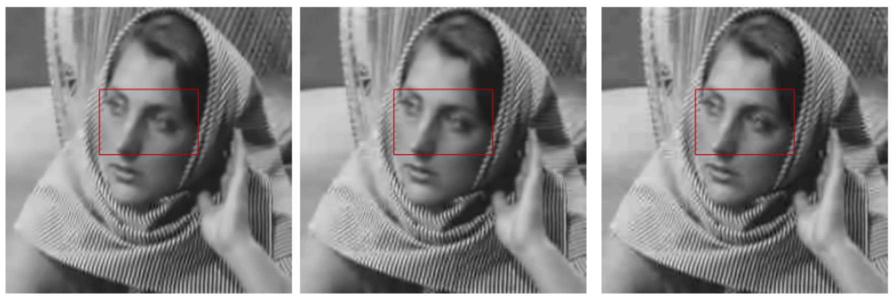
[6] Thomas Meier, King N Ngan, and Gregory Crebbin, "Reduction of blocking artifacts in image and video coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 9, no. 3, pp. 490–500, 1999.

[7] Kiryung Lee, Dong Sik Kim, and Taejeong Kim, "Regression-based prediction for blocking artifact reduction in jpeg-compressed images," *Image Processing, IEEE Transactions on*, vol. 14, no. 1, pp. 36–48, 2005.

[8] Steven SO Choy, Yuk-Hee Chan, and Wan-Chi Siu, "Reduction of block-transform image coding artifacts by using local statistics of transform coefficients," *IEEE Signal Processing Letters*, vol. 4, no. 1, pp. 5–7, 1997.



(a) JPEG (PSNR=28.01dB) (b) Li's (PSNR=29.50dB) (c) BM3D (PSNR=29.30dB)



(d) DicTv (PSNR=29.17dB) (e) Sun's (PSNR=28.85dB) (f) Ours (PSNR=29.65dB)

Fig. 5. Visual Comparison of Reconstructed Images (QF=15).

[9] Xinfeng Zhang, Ruiqin Xiong, Siwei Ma, and Wen Gao, "Reducing blocking artifacts in compressed images via transform-domain non-local coefficients estimation," in *Multimedia and Expo (ICME), 2012 IEEE International Conference on*. IEEE, 2012, pp. 836–841.

[10] Xinfeng Zhang, Ruiqin Xiong, Xiaopeng Fan, Siwei Ma, and Wen Gao, "Compression artifact reduction by overlapped-block transform coefficient estimation with block similarity," *Image Processing, IEEE Transactions on*, vol. 22, no. 12, pp. 4613–4626, 2013.

[11] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *Image Processing, IEEE Transactions on*, vol. 16, no. 8, pp. 2080–2095, 2007.

[12] H. Liu, R. Xiong, J. Zhang, and W. Gao, "Image denoising via adaptive soft-thresholding based on non-local samples," June 2015.

[13] Kun Huang, Allen Y Yang, and Yi Ma, "Sparse representation of images with hybrid linear models," in *Image Processing, 2004. ICIP'04. 2004 International Conference on*. IEEE, 2004, vol. 2, pp. 1281–1284.

[14] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel, "A non-local algorithm for image denoising," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 2, pp. 60–65.

[15] Ivan Selesnick, "A derivation of the soft-thresholding function," *Polytechnic Institute of New York University*, 2009.

[16] Deqing Sun and Wai-Kuen Cham, "Postprocessing of low bit-rate block dct coded images based on a fields of experts prior," *Image Processing, IEEE Transactions on*, vol. 16, no. 11, pp. 2743–2751, 2007.

[17] Mading Li, Jiaying Liu, Jie Ren, and Zongming Guo, "Patch-based image deblocking using geodesic distance weighted low-rank approximation," in *Visual Communications and Image Processing Conference, 2014 IEEE*. IEEE, 2014, pp. 101–104.

[18] Huibin Chang, Michael K Ng, and Tieyong Zeng, "Reducing artifacts in jpeg decompression via a learned dictionary," *Signal Processing, IEEE Transactions on*, vol. 62, no. 3, pp. 718–728, 2014.