Compression Noise Estimation and Reduction via Patch Clustering

Xinfeng Zhang[*], Weisi Lin[*], Jiaying Liu[†] and Siwei Ma[‡]

[*]Rapid-Rich Object Search (ROSE) Lab, Nanyang Technological University, Singapore
[†]Institute of Computer Science and Technology, Peking University, Beijing, China
[‡]Institute of Digital Media, Peking University, Beijing, China
E-mail:{xfzhang, wslin}@ntu.edu.sg, {liujiaying, swma}@pku.edu.cn

Abstract—**Images compressed at low bit rates usually suffer from annoying artifacts due to coarse quantization of transform coefficients. In this paper, we propose a soft-thresholding scheme to reduce compression noise with content-based noise level estimation. In the proposed method, a compressed image is divided into multiple similar image patch groups, and the compression noise is estimated from every group respectively based on coefficient distribution in transform domain. For each group of similar patches, soft-thresholding is applied to the singular values in the singular value decomposition (SVD) of every group of similar patches. The threshold is adaptively determined based on the standard deviation of image signals and compression noise. Finally, quantization constraint is applied to estimated images to avoid over-smoothing. Extensive experimental results show that the proposed method improves the quality of compressed images obviously, and outperforms state-of-the-art denoising algorithms significantly.**

I. INTRODUCTION

Block discrete cosine transform (BDCT) is widely used in existing image/video coding standards, e.g., JPEG and MPEG-2, to reduce the spatial correlation among neighboring pixels. In general compression process, an image is first divided into non-overlapped blocks, and then compressed by sequentially transforming with BDCT, quantizing and entropy coding independently for every block. One major problem in BDCT based image compression is that annoying compression noise severely degenerate the image quality especially when it is compressed at very low bit rates. The compression noise not only leads to a poor user experience but also deteriorate the performance of many computer vision algorithms.

In order to improve image quality, lots of image denoising methods are proposed in literatures these years [1]-[8]. Most of existing denoising algorithms focus on dealing with additive white Gaussian noise, which are independent and identical distributed in the whole image. Buades et al. [1] proposed the nonlocal means filter to predict each pixel by a weighted average of its surrounding pixels, where the weights are determined by the similarity of the corresponding image patches located at the source and target coordinates. Takeda et al. [2] proposed a signal-dependent steering kernel regression framework for denoising, which takes the covariance of image local gradients to derive edge directions. However, the filtering strength of these methods is difficult to control, which may lead to blur images due to over-smoothing.

In order to avoid over-smoothing, Zhai et al. [3] utilized quantization intervals to constrain the filtered coefficients in the same range with the original ones. Sun and Cham [4] modeled the original image as a high order Markov random field (MRF) based on the field of experts (FoE) framework and utilized quantization steps to estimate compression noise variance, which is utilized to control filtering strength. Zhang et al. [5] [6] proposed a multi-prediction adaptive fusion framework by modeling transform coefficients with generalized Gaussian distribution (GGD) to remove compression noise. In the latest video coding standards, HEVC, the strength of deblocking filter [7] is adaptively determined according to coding modes and quantization parameters (QP), and the adaptive loop filter (ALF) [8] directly derives filter parameters based on the original images and reconstructed images in encoder side.

In recent years, image sparse prior for group of similar image patches are widely studied in literatures and achieves significant improvement in image restoration [9]-[11]. The well-known denoising filter, BM3D [9], enhanced image sparsity by clustering similar 2-D image patches into 3-D data arrays and collaborative filtering is implemented by shrinking the transform coefficients of these 3-D data. Liu et al. [10] utilized trained over-complete dictionary to get much sparser representation for noisy images, and Ren et al. [11] removed compression noise by soft-thresholding singular values of noisy images.

In this paper, we investigate the compression noise estimation and reduction problems for block discrete cosine transform (BDCT) based image compression methods. In our proposed method, we remove compression noise by soft-thresholding singular values of every group of image patches with similar structures. Then, image patches are reconstructed with the shrunken singular values and a noise-alleviative image is further reconstructed by weighted average of these image patches. For each group, we first derive the standard deviation of image signals and compression noise adaptively from the noisy image, which jointly determine the threshold for singular values. Then, the image noise is reduced by thresholding singular values of similar image patches. This process can be performed iteratively to improve the image quality with updated image signals. Finally, in order to avoid the over-smoothing, narrow quantization constraint (NQC) is applied to restored images.

The remainder of this paper is organized as follows. Section II introduces the proposed compression noise estimation and reduction method. Extensive experimental results and analysis are reported in Section III, and Section IV concludes the paper.

II. COMPRESSION NOISE REDUCTION AND ESTIMATION

A. Framework of soft-thresholding denoising

Soft-thresholding is widely used in image denoising problem based on image sparse prior model in transform domain. In order to enhance image sparsity, we construct image sparse representation with similar image patches. For an image, we denote $G_i = \{y_j \mid \|y_j - y_i\|_2^2 < \varepsilon\}$ as a group of image patches with size of $N \times N$. Since the image patches are very similar within groups, they can be sparsely represented with a few basis vectors in certain vector space. In this paper we utilize Singular Value Decomposition (SVD) of similar image patches to construct their sparse representation space. We organize a group of image patches into a matrix, denoted as $\mathbf{Y}_i$, each column of which corresponds to an image patch. Then, a group of image patches can be decomposed as,

$$\mathbf{Y}_i = \mathbf{U}_i \mathbf{S}_i \mathbf{V}_i \tag{1}$$

where $\mathbf{U}_i$ and $\mathbf{V}_i$ are unitary matrices, the columns of each of which form a set of orthonormal vectors. $\mathbf{S}_i$ is a diagonal matrix with non-negative real numbers on the diagonal, which is referred to as singular values.
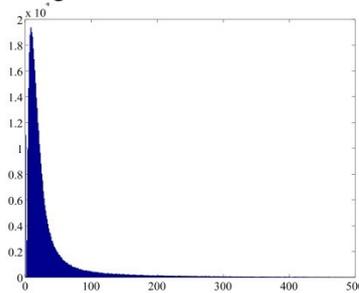

Fig.1 Histogram of singular values of groups of image patches.

Since the image patches in one group are very similar, only a small number of non-zero singular values are needed to represent them. This can be verified by Fig.1, which shows that singular value approaches Laplace distribution. Then, we can reduce compression noise by applying soft-thresholding operation to singular values of groups of image patches,

$$D_\tau(s_k) = \begin{cases} s_k - \tau sign(s_k), & if \ |s_k| > \tau \\ 0, & if \ |s_k| \le \tau \end{cases} \tag{2}$$

where s_k is the k-th singular value of matrix $\mathbf{Y}_i$. Based on the discussion in [12], when signal follows Laplace distribution, the optimal threshold value for the k-th singular value, τ_k, is,

$$\tau_k = 2\sqrt{2}\sigma_{n,G_i}^2/\sigma_{x,k} \tag{3}$$

where σ_{n,G_i} and $\sigma_{x,k}$ are the standard deviation of compression noise and image signals corresponding to the k-th singular value in group G_i, respectively. The image patches are restored as,

$$\hat{\mathbf{X}}_i = \mathbf{U}_i \mathbf{S}_i^{(\tau)} \mathbf{V}_i \tag{4}$$

where $\mathbf{S}_i^{(\tau)}$ is the matrix with shrunked singular values. Then, the high quality image is reconstructed with weighted average of all the overlapped image patches after thresholding. The weights are designed with the number of non-zero singular values,

$$w_i = \max(1 - \frac{r_i}{M_i}, \frac{1}{M_i}) \tag{5}$$

where r_i and M_i are the number of non-zero singular values and the number of the diagonal elements of singular matrix S_i for group G_i. Therefore, the most important factor for our soft-thresholding method is to estimate the standard deviation of compression noise and original image signals.

B. Compression noise estimation via patch clustering

Compression noise is mainly caused by quantization, and is related with both quantization steps and image signal distribution. Considering variations of image signals, we estimate standard deviation of compression noise and image signals for every group individually, based on the assumption that image patches in the same group approximately follow the same distribution. For every group of image patches, we first take the average of all image patches, $\bar{y}_i$, as an initial estimation of original signals.

In order to estimate the compression noise level, we take the following image spatial model,

$$r(x_a, x_b) = \sigma_x \rho_H^{|m_A - m_B|} \rho_V^{|n_A - n_B|} \tag{6}$$

where $r(\cdot)$ is the correlation coefficient function for any two image pixels x_a and x_b with their coordinates (m_A, n_A) and (m_B, n_B), and ρ_H and ρ_V are the correlation coefficient of neighboring pixels in the horizontal direction and vertical direction, respectively. Then, the variance of DCT coefficients can be calculated as,

$$\sigma_X^2(u, v) = \sigma_x^2 \sum_{(j1, j2) \in C} t_{i,j1} t_{i,j2} \rho_H^{|j1-j2|_H} \rho_V^{|j1-j2|_V} \tag{7}$$

where $t_{i,j1}$ and $t_{i,j2}$ are the elements of DCT matrix and $|j1 - j2|_H$ and $|j1 - j2|_V$ represent the distance of pixel $j1$ and $j2$ in horizontal direction and vertical direction, respectively. Based on the derived standard deviation of DCT coefficients in Eqn.(7), the compression noise variance in every band is able to be formulated as follows.

$$\sigma_n^2(u, v) = \int_{-\frac{q}{2}}^{\frac{q}{2}} (n - \mu_n(u, v))^2 f(n|X(u, v)) dn \tag{8}$$

where $q = Q(u, v)$ is the quantization step for band (u, v) and $f(\cdot)$ is the probability density function of compression noise, which is determined by the distribution of DCT coefficients. In this paper, we assume that the DCT coefficients follow Gaussian distribution. Then, the standard deviation of noise in group G_i, σ_{n,G_i}, is estimated by averaging $\sigma_n(u, v)$ in all the bands of $N \times N$ image patch. The standard deviation of image patch corresponding to k-th singular value is calculated as,

$$\sigma_{x,k} = \sqrt{max\left(\frac{s_k^2}{M_k} - \sigma_{n,Gi}^2, 0\right)} \tag{9}$$

C. Iteration improvement and quantization constraint

In noise estimation, we take the average of similar patches as an initial estimation to derive noise variance, which may lead to bias to compression noise level. Therefore, we

perform the soft-thresholding operation to every group iteratively by updating the standard deviation of compression noise as,

$$\sigma_{n,G_i}^{(k+1)} = max\left(\sigma_{n,G_i}^{(k)} - \nabla^{(k)}, 0\right) \qquad (10)$$

where $\nabla^{(k)}$ is the standard deviation of the difference between the reconstructed images in the k-th iteration and the decoded image. In every iteration, the reconstructed image is divided into blocks as that in compression process, and each block is transformed with DCT. The estimated DCT coefficients are projected to a narrow quantization interval to avoid over-smoothing [3].

III. EXPERIMENTAL RESULTS

In this section, we evaluate our proposed method by comparing to representative image denoising methods, NLM[1], BM3D[9], PSW[3] and FoE[4]. Herein, NLM and BM3D are two general denoising methods for different noise types, while PSW and FoE are two well-known denoising methods specially designed for compression noise. PSW and FoE can estimate their parameters from compressed images automatically, but NLM and BM3D need the standard deviation of compression noise to determine their parameters. In order to compare with their best performance, we assign the real standard deviation of compression noise to NLM and BM3D algorithms, which is not available for denoising problem in practice. Six test images, which are widely used in image processing, are firstly compressed by JPEG, and then restored by the compared methods and our proposed method.

Fig.2 plots the PSNR curves of the different image restored methods for JPEG compressed images at different quality factors (QF), which range from 1 to 100 (1 means lowest compression quality and 100 corresponds to highest compression quality) to scale quantization matrix. These results show that our proposed method significantly improve the quality of JPEG images in a large quality range. Especially for image *Barbara*, our method achieves more than 3.5+ dB compared with decoded JPEG image, even compared with state-of-the-art denoising methods, it also achieves up to 2+dB. Although the true values of compression noise variance are utilized to determine parameters in NLM and BM3D, their performances are still obvious inferior to our methods. In addition, the performance of NLM decreases along with QF increase, which implicitly shows the global noise variance cannot well reflect content-dependent compression noise level.

Besides objective quality improvement of our proposed method, the perceptual quality of the restored images is also significantly improved. Fig.3 shows the restored image, *Barbara*, and our method achieves most visual pleasing results, especially in the scarf area.

IV. CONCLUSIONS

In this paper, we propose a soft-thresholding scheme to improve the quality of compressed images. The main contribution of this work is the data-driven estimation method for standard deviation of compression noise from groups of image patches with similar structures. Experimental results demonstrate that the proposed scheme significantly improves the quality of JPEG images. In future work, we will further optimize the accuracy of compression noise estimation and utilize it to improve other image processing applications.

REFERENCES

[1] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in Proc. IEEE Int. Conf. CVPR, vol. 2. Jun. 2005, pp. 60–65.

[2] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," IEEE Trans. Image Process., vol. 16, no. 2, pp. 349–366, Feb. 2007.

[3] G. Zhai, W. Zhang, X. Yang, W. Lin, and Y. Xu, "Efficient Image Deblocking Based on Postfiltering in Shifted Windows," IEEE Trans. on Circuits and Systems for Video Technology, vol. 18, no. 1, pp. 122–126, Jan. 2008.

[4] D. Sun and W.-K. Cham, "Postprocessing of Low Bit-Rate Block DCT Coded Images Based on a Fields of Experts Prior," IEEE Transactions on Image Processing, vol. 16, no. 11, pp. 2743–2751, Nov. 2007.

[5] X. Zhang, R. Xiong, X. Fan, S. Ma, and W. Gao, "Compression Artifact Reduction by Overlapped-Block Transform Coefficient Estimation With Block Similarity," IEEE Trans. on Image Processing, vol. 22, no. 12, pp. 4613–4626, Dec. 2013.

[6] X. Zhang, R. Xiong, S. Ma, and W. Gao, "Reducing Blocking Artifacts in Compressed Images via Transform-Domain Non-local Coefficients Estimation," in IEEE International Conference on Multimedia and Expo (ICME), 2012, pp. 836–841.

[7] A. Norkin, G. Bjontegaard, A. Fuldseth, M. Narroschke, M. Ikeda, K. Andersson, M. Zhou, and G. Van der Auwera, "HEVC Deblocking Filter," IEEE Trans. on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1746–1754, Dec. 2012.

[8] X. Zhang, R. Xiong, S. Ma, and W. Gao, "Adaptive loop filter with temporal prediction," in Picture Coding Symposium (PCS), 2012, 2012, pp. 437–440.

[9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering," IEEE Trans. on Image Processing, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[10] X. Liu, X. Wu, J. Zhou, and D. Zhao, "Data-driven Sparsity-based Restoration of JPEG-compressed Images in Dual Transform-Pixel Domain," in Proc. IEEE Int. Conf. CVPR, vol. 2. Jun. 2015, pp. 5171–5178.

[11] J. Ren, J. Liu, M. Li, W. Bai, and Z. Guo, "Image Blocking Artifacts Reduction via Patch Clustering and Low-Rank Minimization," in Data Compression Conference (DCC), 2013, 2013, pp. 516–516.

[12] S.G. Chang, Bin Yu, M. Vetterli. Adaptive wavelet thresholding for image denoising and compression. IEEE Trans. on Image Processing. Sep. 2000, 9(9):1532–1546.

Fig. 2. Image PSNR values with different restoration methods at different JPEG QF

(a) JPEG (25.6 dB)　　　　　(b) NLM (26.46 dB)　　　　　(c) BM3D (26.45 dB)

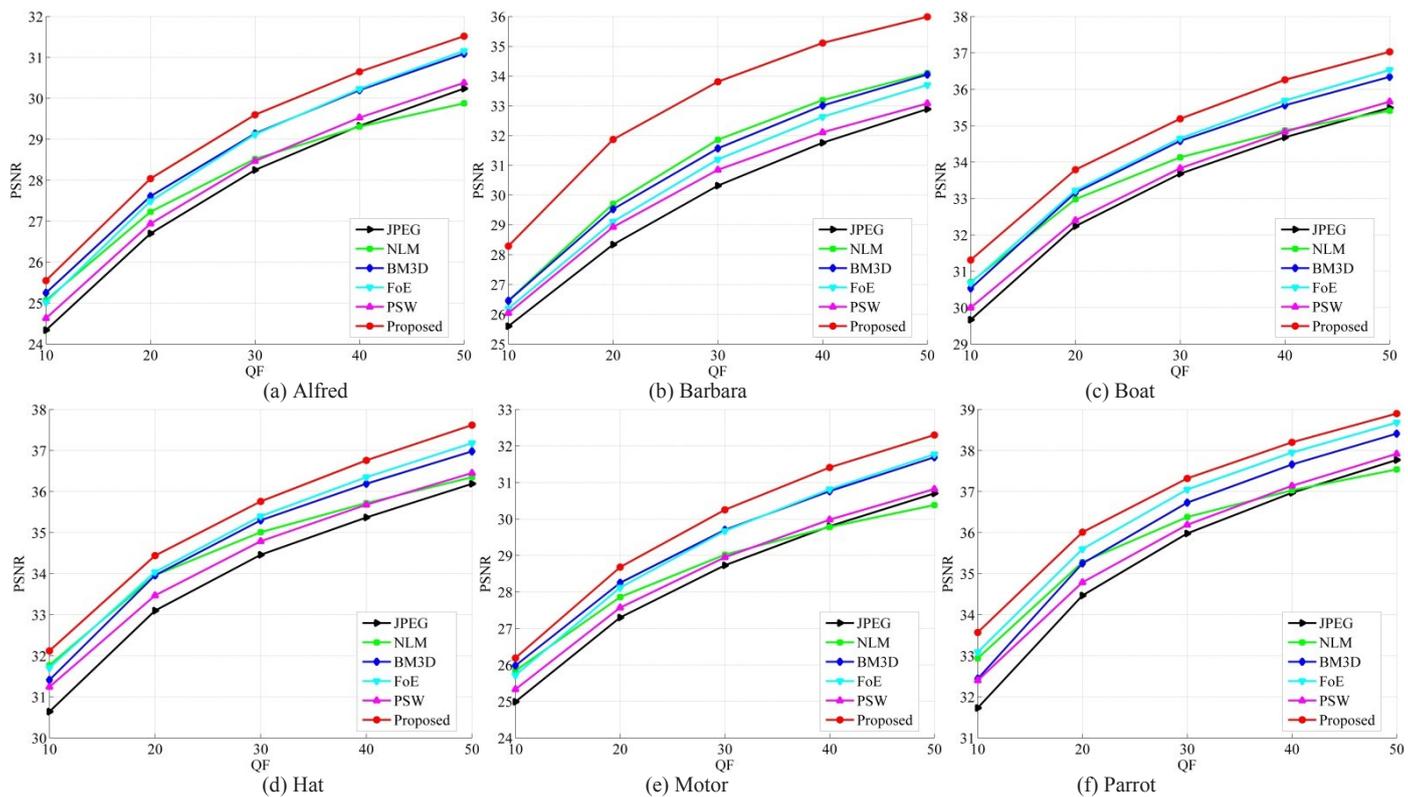(b) FoE (26.19 dB)　　　　　(c) PSW (26.04 dB)　　　　　(d) Proposed (28.29 dB)

Fig. 3. Image visual quality comparison of different restoration methods for *Barbara* at QF=10.