# A STUDY ON QUANTIZATION EFFECTS OF DCT BASED COMPRESSION

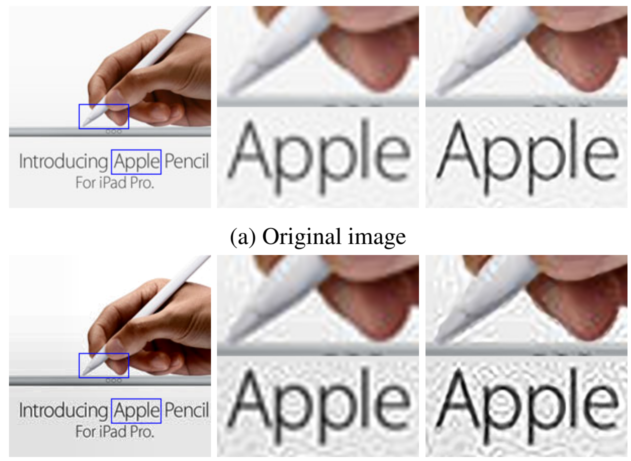
*Xiao Shu*†‡*, Xiaolin Wu*†‡*, Bolin Liu*‡

† Shanghai Jiao Tong University
‡ McMaster University

## ABSTRACT

Quantization in discrete cosine transform (DCT) domain is a widely used lossy compression technique in international multimedia compression standards from audio (e.g., MP3) to image (e.g., JPEG) to video (e.g., H.264). Unlike many degradation sources, like sensor noises, quantization errors of DCT coefficients are signal dependent and difficult to isolate and remove, causing serious artifacts in decompressed and post-processed signals. In this research, quantization errors in the DCT domain are analytically assessed. Our analysis exposes complex behaviors of the DCT quantization errors after being mapped back into the temporal or spatial domain. These behaviors are highly sensitive to quantization precision, the amplitude and phase of the input signal. Based on these observations, we develop a DCT-domain error model to predict and quantify the quantization effects in cases where artifacts are most perceivable to humans, and offer some insights into possible strategies for further suppressing compression noises.

***Index Terms*—** Quantization error, discrete cosine transform, compression standards, perceptual quality.




(a) Original image

(b) Contrast enhanced by CLAHE [2]

**Fig. 1**: JPEG compression artifacts become highly objectionable after image enhancement and/or image size magnification. The second and third sub-figures from left are regions up-scaled by bi-cubic interpolation and A+ [3], respectively.

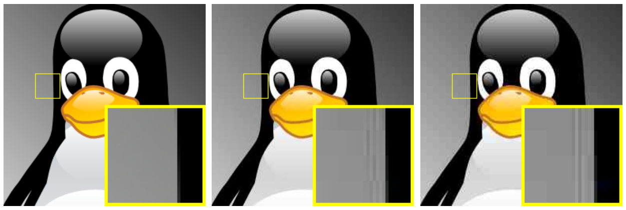
## 1. INTRODUCTION

Discrete cosine transform (DCT) is widely used in the compression of all types of multimedia signals, including audio, image, graphics and video [1]. In the interests of bandwidth economy and cost effectiveness of multimedia products and services, almost all DCT-based international compression standards, such as MP3, MPEG, H.264, etc., operate in the so-called lossy compression mode. That is, the coding process is not strictly invertible; the decompressed multimedia contents are not identical to the original files. Lossy compression is necessary for consumer and industrial applications, because mathematically lossless coding cannot achieve a sufficient level of data volume reduction of most multimedia files that we routinely transmit over the Internet and wireless networks.

In DCT-based lossy compression standards, the reconstruction errors are rooted in quantization of DCT coefficients. Even at a high quality factor (QF) setting, which means high bit-rate or a modest level of compression, quantization artifacts are still perceivable, particularly so on noise-free, sharp synthetic imagery. The compression-induced quality degradation is more pronounced for compound document images, which are characterized by the embedding of graphics arts or texts into an acquired photograph, as exemplified by Fig. 1. This type of quantization artifacts infect graphic arts commonly found in webpages, such as logos, cartons, fonts, etc. The infection becomes highly objectionable after image size magnification or contrast enhancement, which are mundane manipulations in multimedia presentations on large displays.

Unlike other sources of image degradation, for instances, sensor noises and improper illumination, quantization noises of DCT coefficients are signal dependent and hence are far more difficult to be modeled and removed. It is, therefore, important to understand the impacts of DCT quantization on perceptual quality, if one wants to make progress in making lossy compression standards more resistant to compression artifacts [4, 5].

In this paper, we analyze the effects of quantizing DCT coefficients in the image domain after the inverse DCT. Our analysis reveals that quantization errors in the DCT domain



(a) Original (b) JPEG (QF = 50) (c) JPEG (QF = 25)

**Fig. 2**: JPEG compression noise is corrected with signal.

exhibit complex behaviors after being mapped back into the spatial domain. Moreover, the quantization effects in the spatial domain can be easily mistaken as image features. As such, main stream image denoising algorithms, if directly applied on quantization noises, can even strengthen the compression artifacts. However, we find that this commonly overlooked problem can be greatly alleviated by applying some simple strategies in collecting similar patches.

## 2. DCT COMPRESSION ARTIFACTS

The process of capturing, storing and displaying a digital image is far from perfect; it often introduces objectionable errors, such as motion blur, lens distortion, moiré pattern, sensor noise, compression noise, etc., into the final reproduction of a scene. Some errors are independent to and statistically distinct from signal. For example, sensor noise can be modeled as random variables following an independent and identically distributed Gaussian distribution, while true signal has repetitive patterns hence sparse in some basis [6, 7]. By exploiting this statistical difference between signal and sensor noise, denoising techniques can effectively separate signal and noise in a given noisy observation [8]. Compression noises, on the other hand, are much more difficult to model. The non-linearity of quantization operations in image compression systems makes quantization noises image dependent, far from being white and independent.

For example, as shown in Figure 2, JPEG compressed images often have visible boundaries between adjacent DCT coding blocks and discernible ringing artifacts accompanying sharp edges. Blocking artifacts are mainly the results of the quantization error of low-frequency components in the DCT domain. Since we know where the blocking artifacts could happen, they are relatively easy to remove [9, 10, 11, 12]. In comparison, ringing artifacts caused by quantization error of high-frequency components are more difficult to isolate and reduce, because they are not independent from signal and appear to have reoccurring patterns just like signal [13, 14, 15]. In the following, we analyze this problem of reoccurring noise patterns for 1D signal.

By the definition of DCT, the $k$-th DCT coefficient of 1D signal $\boldsymbol{x} = \{x_0, \dots, x_{N-1}\}$ is,

$$X_k = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}\left(n+\frac{1}{2}\right)k\right] = N\sum_{n=0}^{N-1}\frac{1}{N}x_n \cos\left(2\pi\cdot\frac{n+\frac{1}{2}}{N}\cdot\frac{k}{2}\right). \quad (1)$$

By using Fourier transform, coefficient $X_k$ in Eq. (1) can be approximated with a continuous function $\tilde{F}(k)$,

$$\tilde{F}(k) = N\int_0^1 f(t)\cos\left(2\pi t\frac{k}{2}\right)\mathrm{d}t = \frac{N}{2}\int_{-\infty}^{\infty} f(t)e^{-i2\pi t\frac{k}{2}}\,\mathrm{d}t \quad (2)$$

where $f$ is an integrable even function such that

$$\begin{cases} f(\frac{n+\frac{1}{2}}{N}) = x_n, \\ f(t) = f(-t), \\ f(t) = 0, \quad t > 1. \end{cases} \quad (3)$$

By this equation, if sequence $x_0, \dots, x_{N-1}$ consists of equally spaced samples of function $f$ and $f$ satisfies Eq. (3), then a DCT coefficient of the sequence is a sample of $f$ in frequency domain.

With function $\tilde{F}_k$, the inverse DCT transform of $X_k$ can be approximated using the inverse Fourier transform of $\tilde{F}_k$,

$$\hat{x}_n = \tilde{f}(t) = \frac{2}{N}\int_{-\infty}^{\infty}\tilde{F}(k)\,e^{i2\pi t\frac{k}{2}}\,\mathrm{d}\frac{k}{2} = f(t), \quad (4)$$

where $t = (n+\frac{1}{2})/N$.

Now, suppose input signal $\boldsymbol{x}_s$ is a piecewise constant signal with two steps as follows,

$$\boldsymbol{x}_s = \{\underbrace{1, 1, \dots, 1}_{m}, \underbrace{0, 0, \dots, 0}_{N-m}\}, \quad (5)$$

where $0 \leq m \leq N$. Sequence $\boldsymbol{x}_s$ is a discrete version of rectangular function $f_s(t)$,

$$f_s(t) = \mathrm{rect}\left(\frac{t}{2r}\right) = \begin{cases} 0 & \text{if } |t| \geq r \\ 1 & \text{if } |t| < r, \end{cases} \quad (6)$$

where $r = m/N$. As $f_s(t)$ satisfies Eq. (3), by Eq. (2), the DCT coefficient $X_k$ of sequence $\boldsymbol{x}_s$ is approximately,

$$\tilde{F}_s(k) = \frac{N}{2}\int_{-\infty}^{\infty}\mathrm{rect}\left(\frac{t}{2r}\right)e^{-i2\pi t\frac{k}{2}}\,\mathrm{d}t = rN\cdot\mathrm{sinc}(rk). \quad (7)$$

where $\mathrm{sinc}(x) = \sin(\pi x)/(\pi x)$ is the normalized sinc function.

Rectangular function $f_s(t)$ is sparse in temporal domain; its first order derivative is a pulse at point $\pm r$ and zero elsewhere. However, in DCT domain, this function has infinite

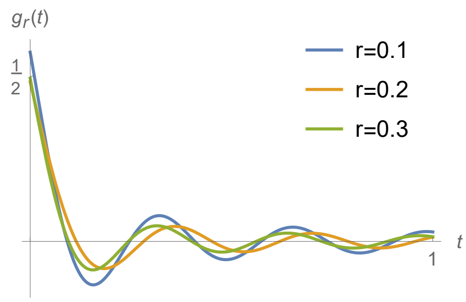



**Fig. 3**: A sharp edge causes similar quantization artifacts regardless of its position $r$ in a DCT block. In this plot, the number of preserved DCT coefficient $b = 6$.

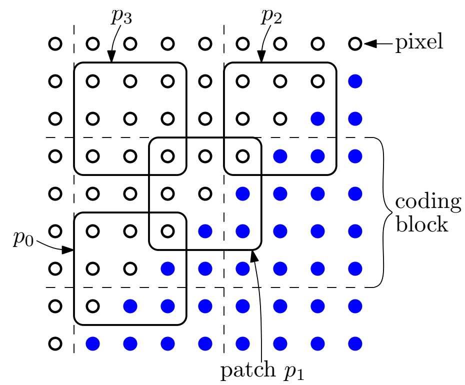



**Fig. 4**: Avoid collecting patches of the same phase.

number of nonzero components as in Eq. (7). To effectively compress this signal, some of the nonzero DCT coefficients need to be quantized to zero. As the magnitude of $\mathrm{sinc}(rk)$ decreases with frequency $k$ in general, if quantization intervals are large enough, quantization effects on $\tilde{F}_s(k)$ can be approximated by cutting off high frequency components. Suppose only the first $b$ DCT coefficients are preserved after quantization and for any $k > b$, $\tilde{F}_s(k)$ is quantized to zero, then by Eq. (4), the reconstructed function $\tilde{f}_s(t)$ is

$$\begin{aligned} \tilde{f}_s(t) &= \int_{-b}^{b} r\,\mathrm{sinc}(rk)\, e^{i2\pi \frac{t}{2} k}\, \mathrm{d}k \\ &= \mathrm{Si}(br + bt) + \mathrm{Si}(br - bt), \end{aligned} \tag{8}$$

where function $\mathrm{Si}(z) = \int_0^z \mathrm{sinc}(t)\,\mathrm{d}t$ is the sine integral.

By left shifting $\tilde{f}_s(t)$ and aligning the edge between the two steps to point zero, we arrive at a new function $g_r(t)$,

$$g_r(t) = \tilde{f}_s(t + r) = \mathrm{Si}(2br + bt) - \mathrm{Si}(bt). \tag{9}$$

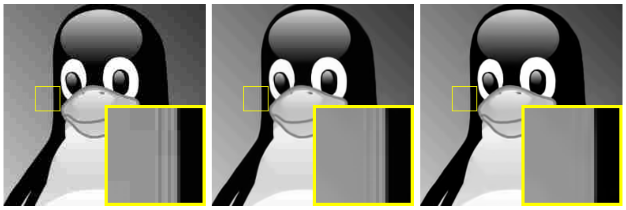
As sine integral function $\mathrm{Si}(z)$ oscillates with a period of 2, the oscillation period of function $g_r(t)$ is $2/b$, which is independent to the edge position $r$ of the original signal. Furthermore, it can be shown numerically that, for any $r_1, r_2$ such that $br_1, br_2 > 0.1$, the difference between the oscillation phases of $g_{r_1}(t)$ and $g_{r_2}(t)$ is less than $1/5$ of the oscillation period $2/b$. Therefore, different edge positions $r$ do not significantly change the shape of function $g_r(t)$. As shown in Fig. 3, quantization noise in $g_r(t)$ has a relatively fixed pattern regardless of the edge position $r$. Thus, if we align the different signals by their edges, the noises become aligned as well. This correlation between signal and quantization noise makes them much more difficult to distinguish statistically.

## 3. STRATEGIES FOR SUPPRESSING COMPRESSION NOISE

Nonlocal self-similarity (NNS) based techniques have achieved the state-of-the-art results in removing Gaussian noise [8]. The NNS prior assumes that noise is independent to signal, hence by comparing a group of similar patches, noise can be isolated from signal. However, as we argued in previous section, compression noise is not random but correlated with the signal; similar patches have similar compression noise, especially when they also have the same relative position to DCT coding blocks. Moreover, patches with matched artifacts can be easily mistaken as being similar using square error metric. Thus, collecting similar patches without taking their contents or positions into consideration inevitably puts multiple instances of the same quantization artifacts into a sample patch group; consequently, such reoccurring noises cannot be separated from the true signal by the NNS prior alone.

For example, as shown in Fig. 4, patches $p_0$ and $p_2$ are both located on a $45°$ high-contrast edge, and their positions relative to coding blocks are the same, hence they have exactly the same ringing artifacts caused by the quantization of the edge in DCT domain. In contrast, patch $p_1$ on the same edge also suffers from ringing artifacts but with a slightly different pattern than those of $p_0$ and $p_2$ as a result of being aligned differently to coding blocks than the other two. Due to its distinct noise patterns, patch $p_1$ is ranked lower in terms of the similarity to $p_0$, however, it is a better candidate for the sample patch group in combating reoccurring artifacts. Therefore, when compiling the sample patch group for patch $p_0$, other patches of the same position in relative to coding blocks, like $p_2$, should be avoided if $p_0$ is around a high-contrast edge. A special case is that, when the edge is horizontal or vertical, patches in the same row or column are potentially distorted by the same artifacts, thus their similarity rating must be reduced accordingly as well.

The above discussed techniques are designed to deal with patches with ringing artifacts around strong edges. Patches in smooth areas are generally free of ringing artifacts since their high frequency coefficients are near zero and the corresponding quantization errors are negligible. However, these patches are not immune to blocking artifacts. Due to the lack of other textures, the boundaries of coding blocks are actually more discernible perceptually in those smooth areas. To prevent



(a) JPEG (QF $= 15$) (b) DTPD (c) Proposed

**Fig. 5**: By applying the discussed strategies in collecting similar patches, JPEG soft-decoding method DTPD can be more effective against ringing artifacts.

these blocking artifacts being matched as similar patch features causing reoccurring artifacts in sample patch group, the same strategy of choosing only unaligned patches as previous case can be employed. Furthermore, since natural images are smooth in general, patches in a small windows of smooth area are similar to each other. Therefore, for a patch from a smooth area, patches in close proximity are sufficient to build a good sample patch group. If the search window is small enough, there are few patches perfectly aligned with the given patch, hence reducing the risk of collecting too many patches with repeated artifacts.

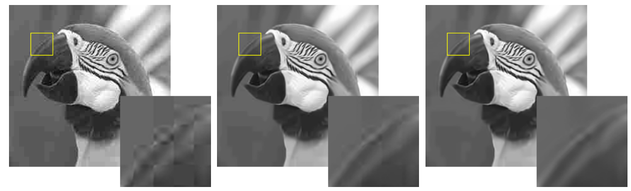
## 4. EXPERIMENTAL RESULTS

To validate these ideas, we implemented the aforementioned strategies in a state-of-the-art NNS-based JPEG soft-decoding method DTPD [15]. As shown in Fig. 5, the original algorithm, although proven highly effective to restore JPEG-compressed images, still leaves some objectionable ringing artifacts. By simply choosing similar patches more strategically, the algorithm can remove residual ringing artifacts without extra computational cost. In fact, as the proposed method avoids patch candidates with possible repeated artifacts, it needs to examine fewer patches hence 10% faster than the original DTPD algorithm on average.

The visual quality improvement of the proposed method over DTPD is also apparent for natural images as shown in Fig. 6. With low QF settings, the hard-decoded JPEG image show some severe ringing artifacts around sharp edges, as predicted by the previous analysis. Some of the repetitive noise patterns are still visible in the output of DTPD. In comparison, by avoiding aligned patches in relative to coding blocks, the improved method does not easily mistake these artifacts as true image features and can remove these artifacts effectively.

In objective evaluations, the improved method has about 0.4dB advantage over DTPD in PSNR on average using LIVE image dataset. For reference, a learning-based JPEG artifact-removal technique ARCNN [14] is also tested. As shown in Table 1, the improved method leads in both average PSNR and SSIM for every QF setting. The PSNR advantage of the improved method is even more evident for large QFs as shown in Fig. 7.

(a) JPEG (QF $= 5$) (b) DTPD (c) Proposed



(d) JPEG (QF $= 15$) (e) DTPD (f) Proposed

**Fig. 6**: Visual quality comparison

| QF | Measure | JPEG | ARCNN | DTPD | Proposed |
|---|---|---|---|---|---|
| 10 | PSNR | 26.371 | 27.585 | 27.467 | **27.640** |
| | SSIM | 0.7681 | 0.8043 | 0.8017 | **0.8074** |
| 20 | PSNR | 28.643 | 29.878 | 29.714 | **30.035** |
| | SSIM | 0.8463 | 0.8704 | 0.8682 | **0.8759** |
| 30 | PSNR | 29.985 | 31.263 | 31.080 | **31.475** |
| | SSIM | 0.8806 | 0.9014 | 0.8981 | **0.9049** |
| 40 | PSNR | 30.975 | 32.232 | 32.091 | **32.540** |
| | SSIM | 0.9005 | 0.9175 | 0.9155 | **0.9218** |

**Table 1**: PSNR and SSIM results of tested techniques.

## 5. CONCLUSION

DCT quantization errors exhibit complex behaviors after being mapped back into the spatial domain, and they can be easily mistaken as regular structures. Based on these observations, we develop a DCT-domain error model to predict and quantify the quantization effects in cases where artifacts are most perceivable to humans. This model can further improve the effectiveness of denoising techniques in combating compression artifacts.

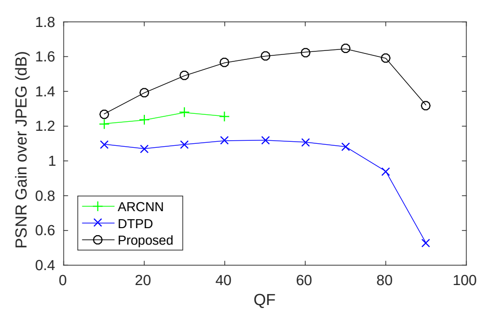



**Fig. 7**: The PSNR gains of tested techniques over JPEG.

## 6. REFERENCES


[1] Gregory K Wallace, "The JPEG still picture compression standard," *IEEE transactions on consumer electronics*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.

[2] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987.

[3] Radu Timofte, Vincent De Smet, and Luc Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Computer Vision–ACCV 2014*, pp. 111–126. Springer, 2014.

[4] Mark A Robertson and Robert L Stevenson, "DCT quantization noise in compressed images," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, no. 1, pp. 27–38, 2005.

[5] Jun Xie and Liang-Tien Chia, "Study on the distribution of dct residues and its application to rd analysis of video coding," *Journal of Visual Communication and Image Representation*, vol. 19, no. 7, pp. 411–425, 2008.

[6] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman, "Non-local sparse models for image restoration," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 2272–2279.

[7] Weisheng Dong, Xin Li, Lei Zhang, and Guangming Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 457–464.

[8] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng, "Weighted nuclear norm minimization with application to image denoising," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 2862–2869.

[9] AS Al-Fohoum and Ali M Reza, "Combined edge crispiness and statistical differencing for deblocking jpeg compressed images," *Image Processing, IEEE Transactions on*, vol. 10, no. 9, pp. 1288–1298, 2001.

[10] Zixiang Xiong, Michael T Orchard, and Ya-Qin Zhang, "A deblocking algorithm for jpeg compressed images using overcomplete wavelet representations," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 7, no. 2, pp. 433–437, 1997.

[11] Jim Chou, Matthew Crouse, and Kannan Ramchandran, "A simple algorithm for removing blocking artifacts in block-transform coded images," *Signal Processing Letters, IEEE*, vol. 5, no. 2, pp. 33–35, 1998.

[12] Guangtao Zhai, Wenjun Zhang, Xiaokang Yang, Weisi Lin, and Yi Xu, "Efficient deblocking with coefficient regularization, shape-adaptive filtering, and quantization constraint," *Multimedia, IEEE Transactions on*, vol. 10, no. 5, pp. 735–745, 2008.

[13] 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.

[14] Chao Dong, Yubin Deng, Chen Change Loy, and Xiaoou Tang, "Compression artifacts reduction by a deep convolutional network," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 576–584.

[15] Xianming Liu, Xiaolin Wu, Jiantao Zhou, and Debin Zhao, "Data-driven sparsity-based restoration of JPEG-compressed images in dual transform-pixel domain," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5171–5178.