

Optimization Approaches to Image Enhancement

OPTIMIZATION APPROACHES TO IMAGE ENHANCEMENT

BY

XIAO SHU, M.Sc.

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

© Copyright by Xiao Shu, November 2015

All Rights Reserved

Doctor of Philosophy (2015)
(Electrical & Computer Engineering)

McMaster University
Hamilton, Ontario, Canada

TITLE: Optimization Approaches to Image Enhancement

AUTHOR: Xiao Shu
M.Sc., (Computing and Information Science)
University of Guelph, Canada

SUPERVISOR: Dr. Xiaolin Wu

NUMBER OF PAGES: xiii, 100

To Yao Zhang

Abstract

To human viewers, sharp edges and rich details in an image are often interpreted as high perceptual quality. Due to various types of degradation during the acquisition of an image and the limitations of display devices and human visual system, some information that exists in the acquired image can be difficult to distinguish when the image is displayed directly, affecting its perceptual quality. Many image enhancement techniques have been proposed to fully utilize the dynamic range of the image data and reproduce a visually more appealing and informative image.

In this dissertation, we present two image enhancement techniques. The first is a global approach that utilizes advanced image statistics and finds the best compromise among the factors that affect image quality; the second is a local approach exploiting the fact that the maximum discrimination power of human vision system can only be achieved in a relatively small locality of an image. These two approaches produce visually pleasing results consistently over a wide range of images.

Besides the various types of artifacts, another practical problem affecting the perceptual quality of an enhanced image is the compression noise. Due to the low pass nature of image compression, the high-frequency components of a compressed image with sharp edges often carry large compression error, which can be amplified by image enhancement operator deteriorating the perceptive quality of enhanced image. By

incorporating the non-linear DCT quantization mechanism into the formulation for image enhancement, we propose a new sparsity-based convex programming approach for joint quantization noise removal and enhancement. Experimental results demonstrate significant performance gains of the new approach over existing enhancement methods.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Xiaolin Wu, for his kind assistance, wise advice and great patience. His wide knowledge in image and video processing technologies and sharp vision of future trends provided the foundation stone for this thesis.

I am delighted to have Dr. Sorina Dumitrescu and Dr. Shahin Sirouspour as my thesis committee members and Dr. Ling Guan as the external examiner. I would like to thank them for their constructive comments, and for arranging time for the final oral examination. I would also like to thank Dr. Donald Sprung for chairing the examination.

My warm thanks are due to Guangtao Zhai and Xianming Liu who kindly provided helpful input and experimental resources for this study. I would also like to acknowledge my fellow collaborators Yi Niu, Reza Pournaghi and Jiantao Zhou for generously sharing their advice and knowledge.

Lastly, and most importantly, I wish to thank my wife, Yao Zhang, for her unconditional love, patience and support. I cannot imagine the life without her. To her I dedicate this thesis.

Notation and abbreviations

1D one dimensional.

2D two dimensional.

BBHE brightness preserving bi-histogram equalization.

CLAHE contrast-limited adaptive histogram equalization.

DCT discrete cosine transform.

DSIHE dualistic sub-image histogram equalization.

EME enhancement measure.

HDR high dynamic range.

HE histogram equalization.

HVS human visual system.

i.i.d. independent and identically distributed.

LARCOTM locally adaptive rank-constrained optimal tone mapping.

MSE mean square error.

NNS nonlocal self-similarity.

NQCS narrow quantization constraint set.

OCTM optimal contrast-tone mapping.

PSNR peak signal-to-noise ratio.

QF quality factor.

RMSE root mean square error.

SVT singular value thresholding.

Contents

Abstract	iv
Acknowledgements	vi
Notation and abbreviations	vii
1 Introduction	1
1.1 Image Enhancement Techniques	2
1.2 Original Contributions	7
1.3 Outline of the Dissertation	10
2 Second Order Statistics	11
2.1 Joint Distribution in OTCM	12
2.2 Perceptual Fine Tuning	15
2.3 Entropy Heuristics	17
2.4 Experimental Results	23
2.5 Conclusion	25
3 Preservation of Order Statistics	27
3.1 Order Statistics and Perception	28

3.2	Locally Adaptive Rank-Constrained Optimal Tone Mapping	30
3.3	Experimental Results	36
3.4	Conclusion	46
4	Enhancement of Compressed Image	47
4.1	Quantization Error in DCT Domain	48
4.2	DCT Quantization Error Model	53
4.2.1	Quantization Effects on Linear Signal	55
4.2.2	Quantization Effects on Piecewise Constant Signal	57
4.3	Enhancement Model	62
4.4	Algorithm	66
4.4.1	Nonlocal Self-Similarity	68
4.4.2	Singular Value Thresholding	71
4.4.3	DCT Coefficient Constraint	75
4.5	Experimental Results	79
4.6	Conclusion	89
5	Conclusion and Future Work	90

List of Figures

1.1	A CT image is displayed in four separate windows in four subranges of pixel values, emphasizing on lung, soft tissue, liver and bone, respectively.	5
2.1	The joint distribution of adjacent pixel intensities of an image.	15
2.2	Constructed graph for solving OCTM with entropy heuristics. Only edges that come out from node i are shown.	19
2.3	Constructed graph for solving OCTM with entropy heuristics without self-loop.	20
2.4	Enhancement results of the moon scene.	21
2.5	Enhancement results of the dolphin scene.	22
2.6	Enhancement results of the car scene.	22
2.7	The contrast (EME) of enhanced image as a function of its histogram entropy.	25
3.1	Adelson's checker shadow illusion. In the left image, tile A appears to be much darker than B to a human viewer despite the fact that they have the exact same grey level. However, when a bridge connecting A and B is added as in the right image, the illusion disappears and their shades look identical.	29

3.2	This example shows the effect of different sizes of neighbourhood window D_i . Generally, the output of LARCOTM is of higher contrast for smaller D_i , and the output looks more natural and closer to the result of a global tone mapping operator such as Mantiuk’s method, when D_i is larger.	36
3.3	This example demonstrates the effect of contrast gain upper bound h . As expected, larger h setting results higher contrast of the output images.	37
3.4	12-bit input infrared images and the enhanced images by Mantiuk’s global and local methods, Paris’ method, DDE and the proposed LARCOTM method, respectively.	38
3.5	The order statistics error images for the bulldozer case. The brighter a pixel is, the larger order statistics error at that location is generated by the corresponding tone mapping method.	40
3.6	12-bit input CT images and the enhanced images by Mantiuk’s global and local methods, BACCT and the proposed LARCOTM method, respectively.	42
3.7	Tone mapped natural HDR images by different methods. The second and third rows contain magnified patches of the images in the first row.	44
3.8	Tone mapped natural HDR images by different methods. The second and third rows contain magnified patches of the images in the first row.	45
4.1	JPEG compression artifacts become highly objectionable after image enhancement or image size magnification, or both. The second and third sub-figures from left are regions up-scaled by bi-cubic interpolation and A+ (Timofte <i>et al.</i> , 2014), respectively.	49

4.2	JPEG compression noise is corrected with signal.	50
4.3	DCT quantization noise appears drastically different with small changes in QF, signal phase or amplitude.	52
4.4	Besides QF, JPEG compression noise is sensitive to the angle and phase of the signal.	53
4.5	DCT coefficients of a decreasing linear sequence.	56
4.6	Quantization removes small AC components causing perceivable block- ing artifacts in a simple gradient ramp image.	56
4.7	A sharp edge causes similar quantization artifacts regardless of the phase r	60
4.8	DCT coefficient Y_k as a function of frequency k	62
4.9	Avoid collecting patches of the same phase.	68
4.10	Several widely used test images.	79
4.11	The median PSNR gain as a function of QF.	83
4.12	The median SSIM gain as a function of QF.	83
4.13	Comparison of tested methods in visual quality at QF = 5.	85
4.14	Comparison of tested methods in visual quality at QF = 15.	86
4.15	Comparison of tested methods in visual quality at QF = 25.	87
4.16	Results of image enhancement techniques applied on JPEG compressed images with QF = 90.	88

Chapter 1

Introduction

To human viewers, sharp contrast of edges and subtle tone of smooth surfaces in an image are often interpreted as high perceptual quality. But various conditions, such as foggy weather, poor illumination, camera motion, etc., can make an acquired image look faded and blurry. However, it is not uncommon that the raw image with low perceptual contrast still contains information on the details of the captured scene. Modern image acquisition devices can generate digital images of intensity resolution that exceeds 16 bits (65536 levels of grey). In comparison, the human visual system (HVS) has a quite limited sensitivity in minating different light intensity levels. Under ideal viewing conditions, human viewers can only distinguish up to around 720 levels of grey on a typical medical display (Kimpe and Tuytschaever, 2007). Therefore, there exists a substantial gap between the achievable precision of devices and the discrimination capability of HVS in pixel intensity. This gap presents a serious obstacle in applications where images of high dynamic range (HDR) are scrutinized and acted upon ultimately by human viewers; for instance, in medical imaging it is the doctors not computers that make diagnostics and clinic decisions

relying primarily on visual examination of the input HDR image. Since very early days of image processing many image enhancement techniques have been proposed and used, aiming to fully utilize the dynamic range of the raw sensor data and reproduce a visually more appealing and informative image.

1.1 Image Enhancement Techniques

One of the earliest and most widely used image enhancement techniques is histogram equalization (HE), which remaps pixel values of the input image such that the processed image has as uniform a histogram as possible. Although HE has been taught in almost all image processing and computer vision textbooks for decades, the reasoning why and in what conditions HE can be used as a technique of image enhancement was not well analyzed and understood until very recent (Wu, 2011). As a general purpose image enhancement technique, HE has an intrinsic weakness: its tendency to over-exaggerate contrast at the expense of tone discontinuities in smooth waveform, producing unnatural looking results for many types of images. Visual defects of HE were also observed by many other authors long ago, and they proposed several improved histogram-based contrast enhancement techniques (Gauch, 1992; Stark, 2000; Chen *et al.*, 2006a,b). For instance, brightness preserving bi-histogram equalization (BBHE), which partitions the histogram of the input image and runs HE on each part independently, could prevent the shift of the average intensity in the output image, thus maintaining the brightness outlook of the original image (Kim, 1997). Following the same basic idea, dualistic sub-image histogram equalization (DSIHE) preserves the median intensity of the input image (Wang *et al.*, 1999). These two techniques can bring an improvement visually over HE for very dark or bright images, but still

suffer from the same problem of distorting tone subtlety when the histogram contains spikes, i.e., a large number of pixels share the (or almost) same intensity level.

Contrast-limited adaptive histogram equalization (CLAHE) is one of the first attempts to handle spikes in the input histogram. It applies the standard HE to an intermediate histogram generated from the input histogram with all the spikes being cut down to a threshold (Pisano *et al.*, 1998). Based on the similar procedure, Arici *et al.* proposed to construct the intermediate histogram by balancing between the original input histogram and the uniform histogram (Arici *et al.*, 2009). This idea was further extended using second-order image statistics in (Celik and Tjahjadi, 2011). With proper parameter settings by user, those methods can suppress some undesirable artifacts of HE.

In a more rigorous study of histogram-based image enhancement (Wu, 2011), the author formulated the problem as one of optimal allocation of output dynamic range to maximize contrast gain but with constraint on tone distortion, and proposed an optimal contrast-tone mapping (OCTM) algorithm. The OCTM algorithm generally produces visually more pleasing results compared with the aforementioned techniques. But this work still does not break away from a common shortcoming of all its predecessors: the use of only first-order statistics, namely, the histogram. This shortcoming is very serious, at least theoretically, considering that the very notions of contrast and tone involve the second-order statistics (autocorrelations and signal spectrum of image signals). In fact, the author of (Wu, 2011) realized that spatial structures of pixels should not be ignored in any approach of image enhancement, and accordingly categorized OCTM together with other histogram-based methods into a so-called context-free class of image enhancement methods.

The quality of image sensor has made big strides in the last decade since the popularization of high-quality but inexpensive digital cameras. In recent years, instead of pixel count, the dynamic range of an image sensor is gradually becoming the new focus of digital imaging technology research and development due to the demand of better image quality in low-light conditions. Nowadays, most high-end image sensors on professional grade digital cameras support 14-bit depth for each colour channel. Even mobile devices, like cellphone cameras, start to employ 10-bit image sensors. In contrast, during the same period of time, image display devices receive almost no improvement in terms of bit depth; most latest consumer grade displays still only support 8-bit depth per colour channel. This is because 8-bit depth is sufficient for most applications. The fundamental limiting factor in perceiving high bit-depth image is not technical limitation of display devices but the human viewers themselves; HVS can only distinguish a limited number of light intensity levels in the brightness range of a typical display. The problem of filling the gap between the high precision of image sensors and the low discrimination capability of HVS in pixel intensity gives rise to HDR tone mapping, a new category of image enhancement technologies.

A straightforward way of visualizing subtle details in an HDR image is to display and examine it in different intensity subranges on screen. In medical imaging software, for example, different representations of the HDR image are placed side by side in different windows (Barnes, 1992), as illustrated by Figure 1.1. The drawback of this scheme is that the wholeness and coherence of the image are compromised causing missing or misreading of some vital information by radiologists (John *et al.*, 2004). If in some professional applications (e.g., medical diagnosis) it cannot be helped to display an HDR image in multiple windows of different intensity subranges out of

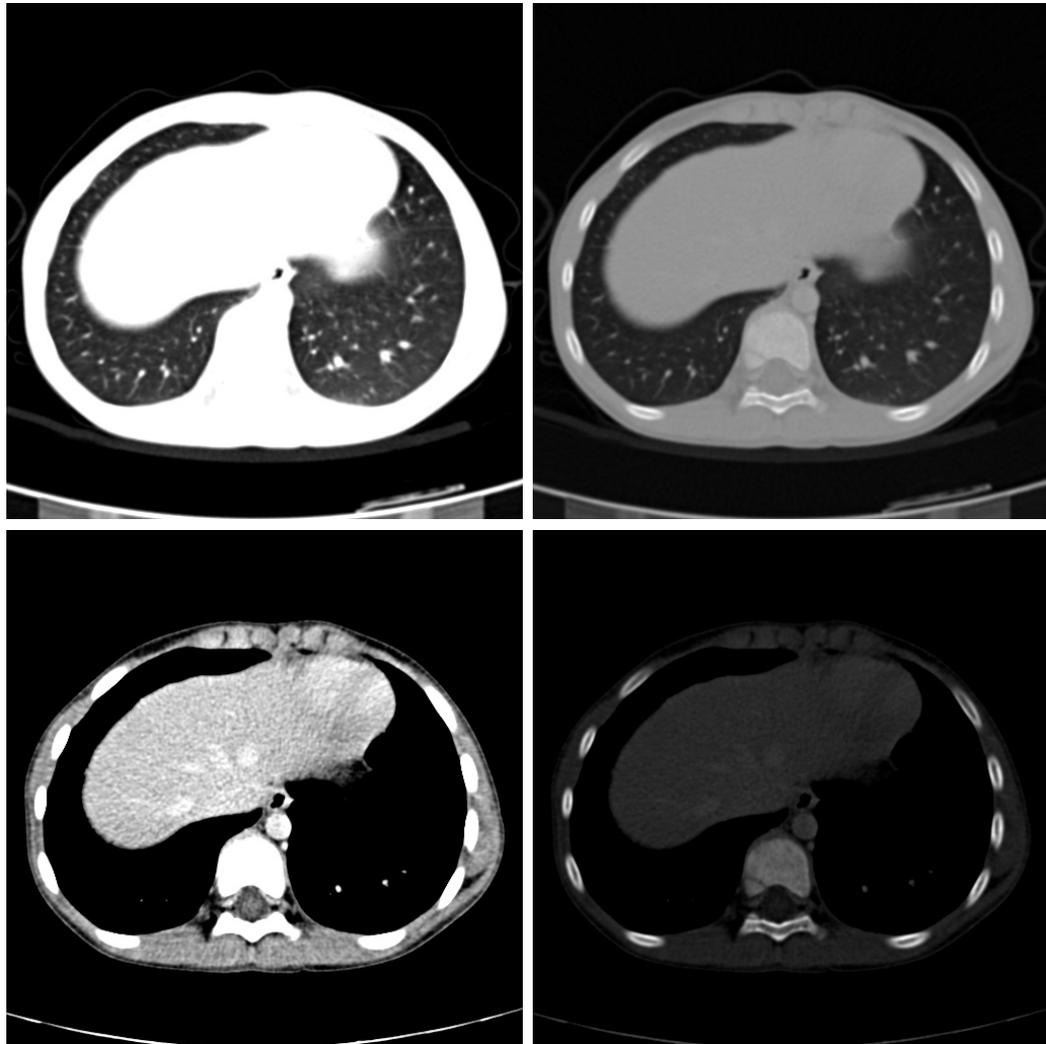


Figure 1.1: A CT image is displayed in four separate windows in four subranges of pixel values, emphasizing on lung, soft tissue, liver and bone, respectively.

the fear of missing any nuances in the raw image data, HDR presentation by intensity windowing is unacceptable in consumer applications. In the latter case, users just want to leisurely appreciate rich details of HDR images in an intuitive, holistic representation; they should not endure the burden and weariness of cross-examining multiple versions of an HDR image. This is the reason for the wide use of display methods of dynamic range compression or commonly referred as HDR tone mapping (Larson *et al.*, 1997; Reinhard *et al.*, 2010). The objective of HDR tone mapping is to augment human vision so that viewers can see details that are otherwise indiscernible to naked eyes. The basic premise of HDR tone mapping is a so-called retinex principle: HVS is more sensitive to relative intensity changes or contrast than to absolute luminance levels (Land and McCann, 1971). This is because as a result of long evolution process, HVS infers intrinsic properties of object surfaces based on contrast and hue, largely factoring out variations of external illumination.

In the retinex principle, all existing HDR tone mapping techniques are designed to squeeze the dynamic range of the low-pass signal and graft the high-pass signal onto the squeezed low-pass signal. Unlike the approach of intensity windowing, the compression of intensity dynamic range is necessarily a many-to-one tone mapping; the process inevitably causes loss or distortion of information for the simple reason that the amplitude quantization precision of pixels exceeds the raw discrimination power of HVS. Therefore, the challenge of HDR tone mapping is how to make as much wanted information conspicuous as possible while preventing or minimizing information loss and visual artifacts, mathematically casting the problem into a framework of constrained optimization.

In the above optimization perspective, HDR tone mapping is a resource allocation

problem: how to allocate a limited discriminating power of HVS and/or display hardware to best visualize subtle details spread in a very large dynamic range of intensity. One common approach to the problem is to first decompose the input image into multiple layers containing coarse to fine scale signals, then reallocate coarse layers with reduced dynamic range so that the recombined image is tone compressed. Early natural HDR image tone mapping techniques, such as homomorphic filtering (Oppenheim *et al.*, 1968), retinex algorithm (Land and McCann, 1971; Rahman *et al.*, 2004) and the like, fall into this class. These techniques often produce visual artifacts, such as halo, double edging, ghosting, etc., near strong image edges. In order to alleviate the problem, many state-of-the-art methods do the decomposition using edge-aware filters to preserve local waveform near edges more faithfully (Aubry *et al.*, 2014; Farbman *et al.*, 2008; Fattal, 2009; Paris *et al.*, 2011). Another HDR tone mapping approach, called gradient-based method, reallocates dynamic range resource in gradient domain and reconstructs a tone compressed image from the modified gradients (Fattal *et al.*, 2002; Mantiuk *et al.*, 2006). Similar to the aforementioned idea of preserving local waveform, this approach tries to retain the same sign of the original derivative wherever in the tone mapped image to reduce visual artifacts. Čadík *et al.* (Čadík *et al.*, 2008) gives an excellent overview and evaluation of many of the HDR tone mapping methods.

1.2 Original Contributions

This dissertation devotes itself to optimization approaches to the field of image enhancement. Several original contributions can be drawn as following.

Our critique of using first-order statistics for image enhancement, histograms in

particular, suggests the necessity of including second-order statistics, i.e., exploiting the spatial information of an image. Following this line of investigation, a natural inquiry is whether the role of histogram in image enhancement can be replaced by the joint distribution of spatially adjacent pixel values. This generalization from first-order to second-order statistics, we find out, can be realized in the existing OCTM framework. Quite surprisingly, the use of joint distribution in the OCTM objective function does not materially increase the complexity of the underlying design algorithm. Besides the above progress, we also propose a new remedy for a common adverse side effect of histogram-manipulation contrast enhancement methods: the distortion of tone subtlety caused by mapping two or more different grey levels into the same value in the output image. Although this side effect is often inevitable for the benefit of increased contrast, a good balance between high contrast of edges and tone subtlety of smooth shades is a critical design issue. In the adoption of joint distribution in the OCTM objective function, we also include an entropy heuristic to control the loss of tone continuity in the enhanced image. The final result is a generalized and significantly improved OCTM algorithm that obtains superior visual quality over a wide range of images consistently.

As discussed above, many current researches on HDR tone mapping do realize the importance of preserving local waveform in maintaining the coherence of the image and reducing visual artifacts, however, none of their solutions are rigorously formulated. As a result, without given carefully chosen parameters, unnatural appearance and visual artifacts due to HDR compression are still plaguing these HDR tone mapping methods. To advance the state of the art, first, we associate common visual

artifacts of HDR compression with changes of order statistics in the image, and propose to constrain tone mapping in terms of rank preservation in an appropriate spatial neighbourhood. Second, we rigorously formulate HDR tone mapping as a constrained optimization problem, with a general objective function to quantify the level of detail conspicuity and a user-tunable parameter to impose a desired degree of consistency in image order statistics. Third, it is shown how the formulated problem of adaptive local rank preserving optimal tone mapping can be solved by linear programming, and the resulting images do achieve superior visual quality to competing HDR methods.

In this research, we notice that, along with the subtle detail, image enhancement techniques often amplify objectionable compression artifacts making them discernible in the enhanced images. But to our surprise, in the very large existing body of research literature on image enhancement, no study has been reported on how compression noises affect the performances of image enhancement. All published works of image enhancement, except few papers explicitly on the topic of combating compression artifacts (a.k.a., soft decoding), assumed the input image data to be uncompressed or mathematically losslessly compressed. This long-time tradition is, unfortunately, an operational convenience in contrary to the real world settings. In this dissertation, we analyze the nature of quantization noises in the discrete cosine transform (DCT) domain, in which most popular JPEG and H.264/HEVC compression standards operate. In particular, we find that the quantization errors of DCT coefficients exhibit complex behaviours after being mapped back into the spatial domain. These behaviours are highly sensitive to quantization precision (QF in JPEG and H.264/HEVC), the

amplitude and phase of the input image signal. Second, we manage to incorporate the non-linear DCT quantization mechanism into the image enhancement problem. Specifically, we propose a new sparsity-based convex programming approach for joint quantization noise removal and enhancement and demonstrate significant performance gains of the new approach over existing image enhancement methods with DCT-domain compressed images.

1.3 Outline of the Dissertation

The rest of the dissertation is organized as follows. Chapter 2 introduces a global approach that utilizes advanced image statistics and finds the best compromise among the factors that affect image quality. Chapter 3 presents a local approach exploiting the fact that the maximum discrimination power of human vision system can only be achieved in a relatively small locality of an image. In Chapter 4, we study the impact of compression noise to image enhancement and propose a new sparsity-based convex programming approach for joint quantization noise removal and enhancement. Finally, Chapter 5 summarizes this dissertation and suggests some interesting future work.

Chapter 2

Second Order Statistics

Image enhancement techniques can be broadly categorized into two classes: local and global. A local approach enhances the detail of an input image by altering the rate of change in intensity between neighbouring pixels on a pixel-by-pixel basis. For example, traditional edge enhancement and high-boost filtering techniques fall into this class. Although the local techniques are effective to enhance certain types of images, they are prone to objectionable artifacts such as ringing and halo, resulting in severely distorted image features. The global image enhancement techniques, such as HE and OCTM, in comparison, do not directly adjust the local waveform. Instead, they manipulate the image histogram using a monotonic tone mapping function that is determined by the histogram of the input image.

Visual defects of HE were well documented by many authors long ago, and many improved histogram-based enhancement techniques have been proposed since (Stark, 2000; Kim, 1997; Chen *et al.*, 2006a,b; Wang *et al.*, 1999; Pisano *et al.*, 1998). However, the use of only first-order statistics, namely, the histogram, does not provide enough information of an image. This fundamental problem prevents these image

enhancement techniques from over or under enhancing certain types of input image. Thus, it is necessary to include second-order statistics, i.e., the spatial information of an image, in the design of a more robust image enhancement technique.

2.1 Joint Distribution in OTCM

In (Wu, 2011) the problem of OCTM is formulated as the following linear program,

$$\begin{aligned}
 & \text{maximize} && \sum_{i=1}^{L-1} p_i s_i \\
 & \text{subject to} && \mathbf{1}^T \mathbf{s} \leq L \\
 & && 0 \leq \mathbf{s} \leq \mathbf{u} \\
 & && \sum_{j=0}^{d-1} s_{i+j} \geq 1, \quad i = 1, \dots, L-d,
 \end{aligned} \tag{2.1}$$

where \mathbf{p} is the histogram vector of the input image of L grey levels; variable s_i is called the context-free contrast at grey level i , which is the unit rate of change from level i to level $i+1$ in the output image independent of pixel locations. Collectively, the $L-1$ elements of vector $\mathbf{s} \in \mathbb{R}^{L-1}$ uniquely determine the integer-valued monotonically non-decreasing transfer function

$$T(i) = \left\lfloor \sum_{t=1}^i s_t \right\rfloor, \quad 0 \leq i \leq L-1 \tag{2.2}$$

that maps grey level i to $T(i)$. It is clear from the definitions of \mathbf{p} , \mathbf{s} and $T(i)$ that the objective function

$$C(T) = \sum_{i=1}^{L-1} p_i s_i = \sum_{i=1}^{L-1} p_i \cdot [T(i) - T(i-1)] \tag{2.3}$$

is a measure of expected context-free contrast achieved by the transfer function $T(i)$, and hence it is the term to be maximized in the linear program. In addition, parameter u is an upper bound on each incremental step s_i and parameter d is for limiting the maximum tone distortion allowed. For the sake of simplicity, we assume that the input and output dynamic ranges are the same, however, the techniques presented in this paper can still work with simple modifications if such assumption is not true.

Like all existing contrast enhancement techniques, OCTM uses histogram as the only image statistics to construct the transfer function $T(i)$ (Kim, 1997; Wang *et al.*, 1999; Pisano *et al.*, 1998). If the population p_i of grey level i is large in the input image, based on its optimization criterion, OCTM allocates a large increment s_i in $T(i)$, increasing the context-free contrast between grey levels i and $i - 1$. This strategy is essentially the same for all existing histogram-based contrast enhancement techniques. For example, HE allocates a piece of dynamic range proportional to the population of each grey level. However, first-order statistics alone cannot characterize perceptual image quality well. In a smooth area consisting of a large number of pixels whose values are close to each other, boosting contrast is counterproductive because this deprives valuable dynamic range resources from some other more interesting or eye-catching areas, and even worse introduces objectionable contour artifacts in the smooth area.

One way of overcoming the limitation of histogram in contrast enhancement is to employ q_{ij} , $i \leq j$, the joint probability that two spatially adjacent pixels take on grey levels i and j respectively. Based on the joint probability q_{ij} , which brings in second-order statistics, we now define the expected context-sensitive contrast $\widehat{C}(T)$

under transfer function $T(i)$ to be

$$\widehat{C}(T) = \sum_{i=0}^{L-1} \sum_{j=i}^{L-1} q_{ij} \cdot [T(j) - T(i)]. \quad (2.4)$$

The fundamental difference between $C(T)$ and $\widehat{C}(T)$ is that the latter accounts for spatial adjacency of two pixels together with their grey levels, but the former totally disregards the spatial relationship between pixels.

For our primary goal of image enhancement, the task becomes to design the transfer function $T(i)$ for a given q_{ij} such that $\widehat{C}(T)$ is maximized. This can easily be integrated into the OCTM framework by replacing the objective in Eq. (2.1) with $\widehat{C}(T)$. Since $T(j) - T(i)$ is the sum of s_k , $i < k \leq j$, Eq. (2.4) is equivalent to,

$$\sum_{i=0}^{L-1} \sum_{j=i}^{L-1} q_{ij} \cdot \left(\sum_{k=i+1}^j s_k \right) = \sum_{k=1}^{L-1} s_k \cdot \left(\sum_{i=0}^{k-1} \sum_{j=k}^{L-1} q_{ij} \right). \quad (2.5)$$

Then, by letting $\hat{\mathbf{p}} \in \mathbb{R}^{L-1}$ be a vector, such that

$$\hat{p}_k = \sum_{i=0}^{k-1} \sum_{j=k}^{L-1} q_{ij}, \quad (2.6)$$

we have

$$\widehat{C}(T) = \hat{\mathbf{p}}^T \mathbf{s}, \quad (2.7)$$

thus, the new context-sensitive contrast measure has the same form as $C(T)$, the objective function of the original OCTM, except for the definition of \mathbf{p} . Therefore, the efficient linear program solvers for OCTM are still applicable to the new problem.

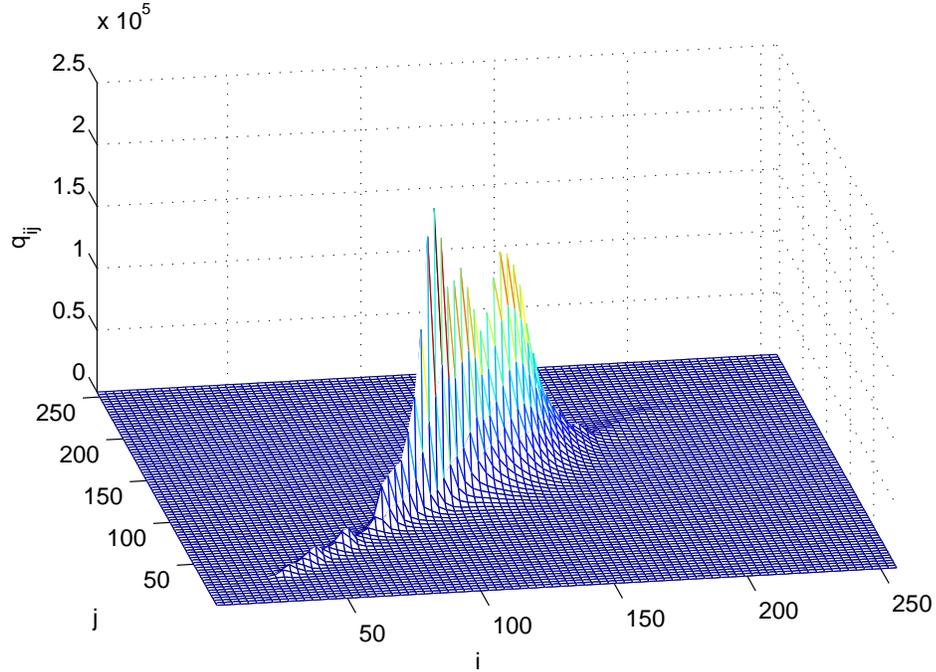


Figure 2.1: The joint distribution of adjacent pixel intensities of an image.

2.2 Perceptual Fine Tuning

The advantages of adopting joint distribution q_{ij} in the framework of OCTM go beyond the inclusion of spatial information in the context-sensitive contrast measure $\hat{C}(T)$ to rectify a serious flaw in the context-free contrast measure $C(T)$. Perhaps, more importantly, q_{ij} makes it possible to factor in many other psychovisual attributes in the design of image enhancement algorithms. For example, most human viewers rate tone subtlety on smooth surfaces and edge sharpness at object boundaries both high in contributing to good image quality for these are vital visual cues in cognition. For this reason, image enhancement cannot be a single minded pursue of high contrast; maximizing $\hat{C}(T)$ directly can still be problematic, risking objectionable contour artifacts in smooth regions as existing histogram-based techniques do.

Consider an image with a large smooth area, in which adjacent pixels have the nearly identical grey levels; the joint probability q_{ij} is very high for pairs of grey levels (i, j) that are in the narrow dynamic range of the smooth area. Figure 2.1 plots q_{ij} for such an example image; the energy of the joint distribution q_{ij} is concentrated along the diagonal where the values of i and j are very close. Simply maximizing $\widehat{C}(T)$ would increase $T(j) - T(i)$ drastically, turning the original smooth shade into contour artifacts. In order to account for the sensitivity of human visual system to tone subtlety in smooth regions, one should reduce the weight of q_{ij} in the objective function when i and j values are very close.

On the other hand, when i and j values differ greatly, i.e., a high contrast can be easily observed between two corresponding adjacent pixels, increasing the contrast further will not noticeably improve image quality. Therefore, the dynamic range of the display should not be wasted on pixels that already exhibit sufficiently high contrast.

In general, different waveforms in pixel localities affect the perceptual image quality differently. In order to account for these effects in image enhancement algorithms, we propose the use of a perceptual weighting function $W(i, j)$ for adjacent pixel pairs having values i and j in the OCTM framework for image enhancement. Combining all the above points, we finally present a new algorithm of linear program for image enhancement:

$$\begin{aligned}
 & \text{maximize} && \sum_{k=1}^{L-1} s_k \cdot \left[\sum_{i=0}^{k-1} \sum_{j=k}^{L-1} q_{ij} W(i, j) \right] \\
 & \text{subject to} && \mathbf{1}^T \mathbf{s} \leq L \\
 & && 0 \leq \mathbf{s} \leq \mathbf{u} \\
 & && \sum_{j=0}^{d-1} s_{i+j} \geq 1, \quad i = 1, \dots, L - d.
 \end{aligned} \tag{2.8}$$

Compared with the original OCTM in Eq. (2.1), this new problem only differs in how the step vector \mathbf{s} is weighted in the objective functions; while OCTM uses histogram \mathbf{p} , the new problem employs the joint probability q_{ij} adjusted by $W(i, j)$. In either case, the weighting vector for \mathbf{s} is constant to the corresponding optimization problem. Therefore, by simply changing the definition of the constants, the original OCTM solvers can also solve the new problem efficiently without much increase in complexity.

2.3 Entropy Heuristics

The transfer function T of a global image enhancement algorithm is a monotonic integer function that preserves rank consistency, i.e., for all i, j such that $i \geq j$, $T(i) \geq T(j)$. However, unless T is an identity function, it is not invertible, which makes it impossible to revert back the enhancement done by T with another transfer function. From another point of view, when multiple intensity levels are merged to the same histogram bin by a transfer function, corresponding pixels become indistinguishable, and some fine detail might be flattened consequently in the enhanced image. This undesired loss of rank information is unavoidable as its a trade-off for higher contrast. But, in order to achieve the overall superior visual quality, it is important to limit the loss and its side effect. To strike a balance between contrast gain and loss of image details, we quantify rank information using the histogram entropy and introduce a new optimization objective function for OCTM as follows,

$$\hat{\mathbf{p}}^T \mathbf{s} - \lambda \sum_{i=0}^{L-1} \bar{p}_i \log \bar{p}_i. \quad (2.9)$$

The first part of the function is a contrast measure as presented before, and the second part, which is prioritized with a Lagrangian multiplier λ , is the entropy of the normalized histogram $\bar{\mathbf{p}}$ of the enhanced output image. If, for example, intensity levels i and $i + 1$ are merged to the same histogram bin by T , i.e., $T(i) = T(i + 1)$, then the histogram entropy decreases after enhancement, as

$$\begin{aligned}
& -p_i \log p_i - p_{i+1} \log p_{i+1} \\
\geq & -(p_i + p_{i+1}) \log(p_i + p_{i+1}) \\
= & -\bar{p}_{T(i)} \log \bar{p}_{T(i)}.
\end{aligned} \tag{2.10}$$

At the same time, the contrast defined by $\hat{\mathbf{p}}^T \mathbf{s}$ might benefit from the extra grey level saved from the dynamic range compression. Overall, the new objective function is to find the optimal combination of the often conflicting goals of rank information preservation and image enhancement.

With the introduction of the entropy heuristics, the proposed optimization problem is now nonlinear, and cannot be solved by the original OCTM solvers, however, we can still efficiently solve the integer version of this problem, where $\mathbf{s} \in \mathbb{Z}^{L-1}$, using dynamic programming. First, let a_1, a_2, \dots, a_n be the indices of the n nonzero components of \mathbf{s} in ascending order, and define $a_0 = 0$ and $a_{n+1} = L$ to simplify the argument. Then, by removing the nonzero components of \mathbf{s} , the contrast measure can be reformulated as

$$\hat{\mathbf{p}}^T \mathbf{s} = \sum_{i=1}^{L-1} \hat{p}_i s_i = \sum_{i=0}^n \hat{p}_{a_i} s_{a_i}. \tag{2.11}$$

Since for each $i \in [0, n]$, $s_{a_i+1}, s_{a_i+2}, \dots, s_{a_{i+1}-1} = 0$ by the definitions, grey levels from a_i to $a_{i+1} - 1$ are all mapped to the same bin $T(a_i)$, i.e., $T(a_i) = T(a_i + 1) = \dots = T(a_{i+1} - 1)$. Thus, similar to the contrast measure, the histogram entropy can

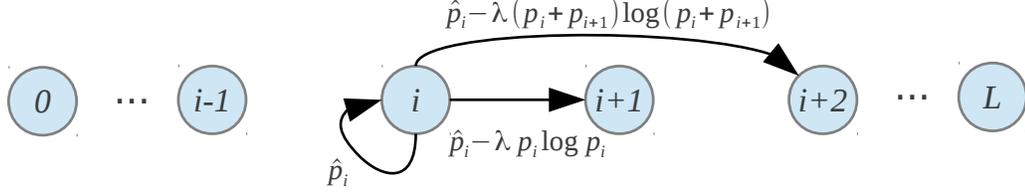


Figure 2.2: Constructed graph for solving OCTM with entropy heuristics. Only edges that come out from node i are shown.

be rewritten as

$$-\sum_{i=0}^{L-1} \bar{p}_i \log \bar{p}_i = -\sum_{i=0}^n \bar{p}_{T(a_i)} \log \bar{p}_{T(a_i)}, \quad (2.12)$$

where $\bar{p}_{T(a_i)} = p_{a_i} + p_{a_i+1} + \dots + p_{a_i+1-1}$. Combining Eq. (2.11) and (2.12), the objective function defined in Eq. (3.5) becomes

$$\sum_{i=0}^n [\hat{p}_{a_i} (s_{a_i} - 1)] + [\hat{p}_{a_i} - \lambda \bar{p}_{T(a_i)} \log \bar{p}_{T(a_i)}]. \quad (2.13)$$

After reformulating the objective function as above, an equivalent problem of finding maximum weight path in a directed graph of $L + 1$ nodes can be constructed accordingly. As illustrated in Figure 2.2, each node in the graph connects to the next d consecutive nodes, where d is the tone distortion constraint defined in Eq. (2.1). The weight of the edge from node i to node $i + k$ for $1 \leq k \leq d$ is set as

$$\hat{p}_i - \lambda \left(\sum_{j=i}^{i+k-1} p_j \right) \log \left(\sum_{j=i}^{i+k-1} p_j \right). \quad (2.14)$$

Except node 0 and L , each node i also has a self-loop with weight \hat{p}_i . Suppose that a path, which takes no more than L steps from node 0 to L , visits different nodes $a'_0, a'_1, \dots, a'_{n'+1}$ in order, and the self-loop of each node a'_i is visited $s'_{a'_i} - 1$ times, then the total weight of the path is exactly the same as the objective function in Eq. (2.13)

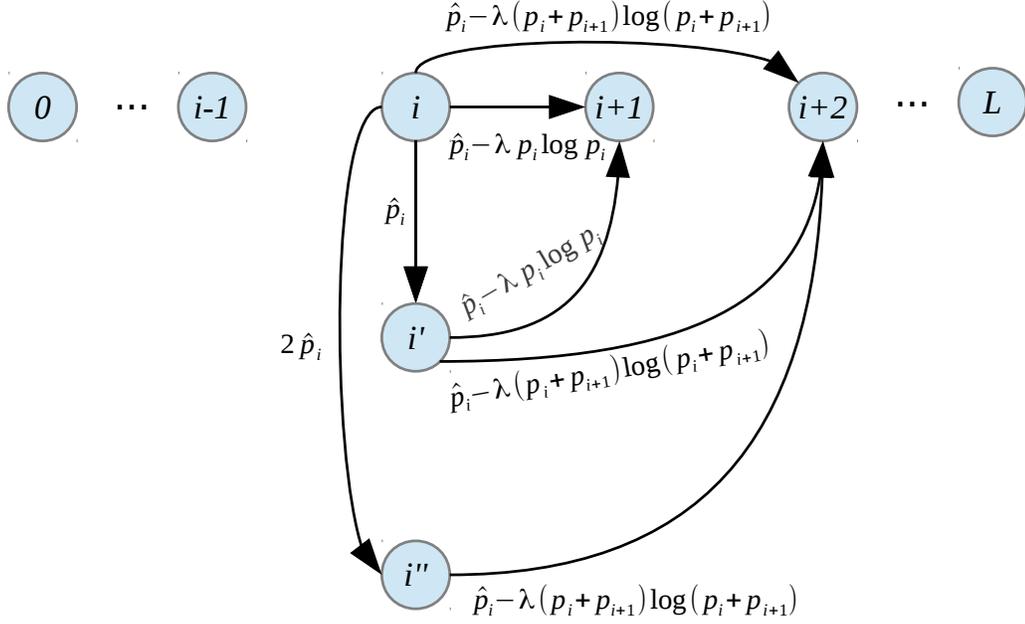


Figure 2.3: Constructed graph for solving OCTM with entropy heuristics without self-loop.

with a, s being replaced by a', s' , respectively. If the number of times of each node i being visited, s'_i , is less than an upper bound u as defined in Eq. (2.1), vector \mathbf{s}' is a feasible solution to the optimization problem. To implement the constraint on the number of times that a node can be visited in a path, we break each self-loop into $u - 1$ dummy nodes $i', i'', \dots, i^{(u-1)}$ for each node i and add an edge to each dummy node $i^{(j)}$ with weight $j\hat{p}_i$ from node i , as shown in Figure 2.3. Each dummy node has the same weighed edges connecting to the other non-dummy nodes as the original node. Obviously, a path from node 0 to L in this reconstructed graph corresponds to a path with no node visited more than u times in the original graph. Thus, a path in the graph also has a corresponding feasible solution to the the optimization problem in Eq. (2.13).

The reverse is also true; each feasible solution has an equivalent valid path in the

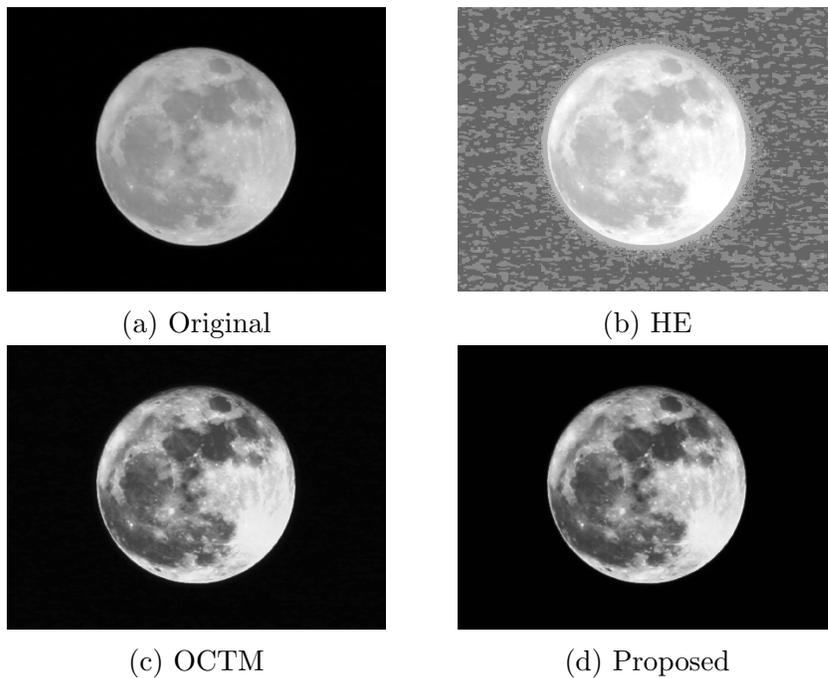


Figure 2.4: Enhancement results of the moon scene.

graph, because the indices a_1, a_2, \dots, a_n of the n nonzero components of a solution \mathbf{s} directly map to the visited nodes and each s_{a_i} indicates the number of times that node a_i is visited in a path by the design of the constructed graph. Hence, \mathbf{s}' of a maximum weight path must be also optimal to the optimization problem, since otherwise, it is possible to find a path with higher weight using the real optimal solution, which contradicts to the assumption. Since the graph is acyclic except for the self-loops, there exist dynamic programming algorithms that can solve the maximum weight path problem in $O(L^2)$ time. Thus, the corresponding optimization problem can also be solved efficiently.

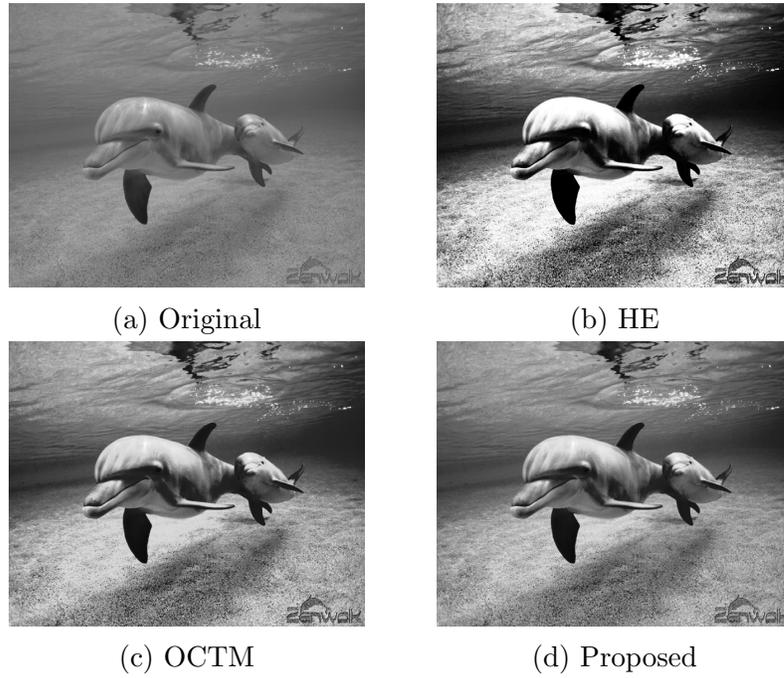


Figure 2.5: Enhancement results of the dolphin scene.

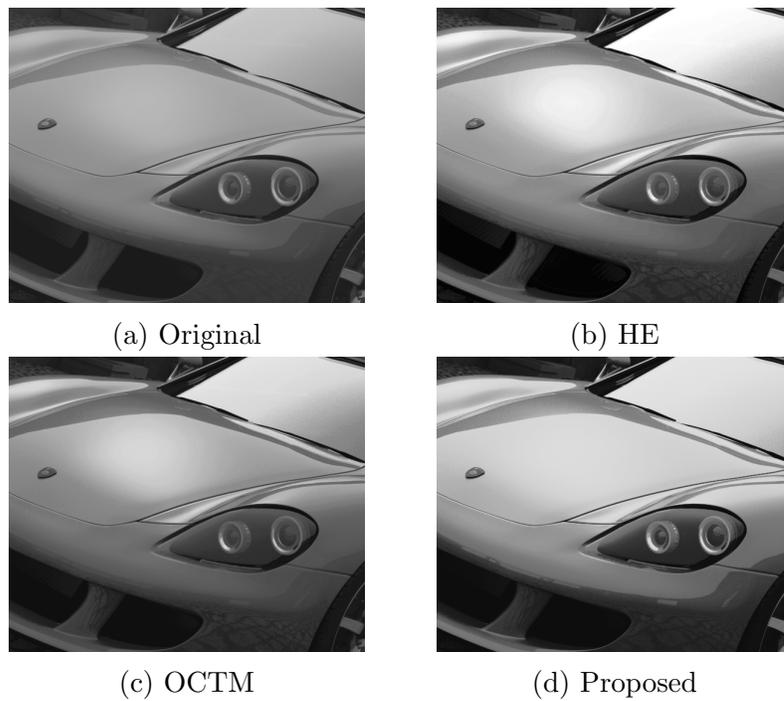


Figure 2.6: Enhancement results of the car scene.

2.4 Experimental Results

Figures 2.4, 2.5 and 2.6 present three sets of images being enhanced by the proposed technique exploiting second-order statistics and entropy heuristics in comparison with those produced by HE and the original OCTM. The transfer functions designed by each technique are also plotted in accompany with the corresponding input histograms to show different behaviours of these techniques in different image statistics.

The test image Moon in Figure 2.4 is a difficult case for traditional histogram-based enhancement methods as we discussed earlier. The input image has a large nearly uniform background region; the corresponding highly skewed histogram misleads HE to disproportionally expand the dynamic range of the background. As a result, HE does not offer the object of interest sufficient dynamic range and ironically reduces its contrast, defeating the very design goal of HE. The quality of HE image is further deteriorated by magnified noises in the background. On the contrary, both OCTM and the proposed method introduce no visible artifact and improve the quality of the input image substantially. Comparatively, the Moon in the enhanced image by the proposed method has higher contrast, because, unlike OCTM which still allocates grey levels to the background pixels, it recognizes the uniformity of the dark region and allocates very few grey levels to the large background and consequently a large dynamic range to the object.

Figures 2.5 and 2.6 compare the performances of the tested enhancement techniques on natural images with narrow but relatively evenly distributed histograms. Despite the good contrast due to the full utilization of the dynamic range, the output of HE flattens subtle smooth shades in some regions and looks unnatural as a whole. The overall impressions of the enhanced images by OCTM and the proposed method,

	Dolphin		Car	
	Entropy	EME	Entropy	EME
Original	6.552	36.8	7.114	25.5
HE	5.893	113.3	5.866	37.4
Proposed ($\lambda = 1.3$)	5.906	117.4	5.878	44.8
GCE	6.487	61.9	6.949	37.3
OCTM	6.487	68.2	6.946	37.5
Proposed ($\lambda = 5.1$)	6.486	71.3	6.944	40.7

Table 2.1: The contrast (EME) and histogram entropy of images enhanced by various techniques.

on the other hand, are visually more pleasing; however, compared with OCTM, the new method preserves details in smooth regions better.

To assess the improvement in the perception of an image by image enhancement, we use image enhancement measure (EME) (Agaian *et al.*, 2007; Panetta *et al.*, 2008) to quantify the perceived contrast of the output. For an input image \mathbf{x} , its EME is defined as,

$$\text{EME}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N 20 \ln \frac{\max(\mathbf{x}_i)}{\min(\mathbf{x}_i)} \quad (2.15)$$

where image block \mathbf{x}_i is one of the N non-overlapping sub-blocks of image \mathbf{x} . As shown in Table 2.1, when the histogram entropy of the enhanced image by each technique is tuned to the same level so that the enhanced images have similar level of detail, the proposed technique provides better quality of enhancement than the competitors, HE, GCE (Arici *et al.*, 2009) and OCTM, measured by EME.

The proposed technique is also more flexible. Demonstrated in Figure 2.7 is the EME of enhanced image as a function of its histogram entropy. As shown in the figure, the contrast of enhanced image, i.e., EME, decreases when we require higher histogram entropy of the output. Most compared techniques only operate in a small range of histogram entropy, but the proposed technique can easily trade off contrast

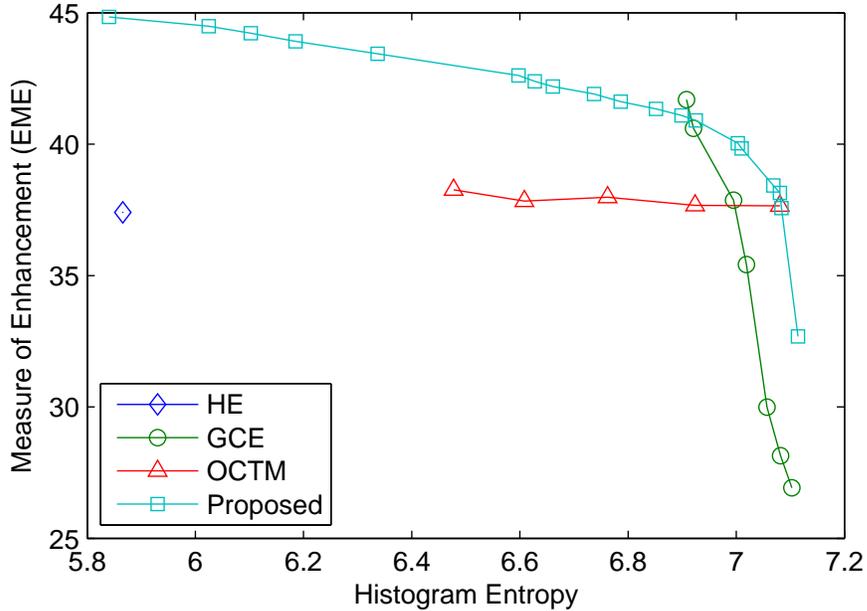


Figure 2.7: The contrast (EME) of enhanced image as a function of its histogram entropy.

gain and histogram entropy for each other based on the preferences of the user.

2.5 Conclusion

This chapter introduces a new image enhancement algorithm in a recently published framework of OCTM. The new algorithm represents a fundamental departure from traditional histogram-based image enhancement techniques (i.e., histogram equalization and all of its variants), in that second-order rather than first-order statistics is used. Perceptual quality attributes, such as contrast and tone, are quantified by joint distribution of the values of spatially adjacent pixels instead of histograms as of today. The problem of image enhancement is then formulated as one of dynamic programming, at the heart of which is a joint distribution-based objective function

that can accommodate various psychovisual properties related to image quality. The new dynamic program algorithm for image enhancement is implemented and its superior performance in visual quality is empirically verified, corroborating with our analytical reasoning.

Chapter 3

Preservation of Order Statistics

We bring the reader into this chapter with the following interesting observation: the type of artifacts commonly found in the results of the retinex HDR image enhancement methods and high-pass filtering methods, such as double edging, ghosting, halo, and ringing are absent in the output images of tone mapping methods based on global histogram transformation (Kim, 1997; Mantiuk *et al.*, 2008; Stark, 2000; Wang *et al.*, 1999; Wu, 2011). The histogram transform type of tone mapping methods (e.g., histogram equalization and its many variants) are classified as global operator because they map an intensity value to another independent of the pixel context (the 2D local waveform), whereas all other methods are classified as local operator as long as the output value of a pixel depends not only on its original value but also on those of its neighbours (Farbman *et al.*, 2008; Fattal *et al.*, 2002; Mantiuk *et al.*, 2006; Paris *et al.*, 2011; Rahman *et al.*, 2004; Reinhard *et al.*, 2002).

3.1 Order Statistics and Perception

The vital difference between the global and local tone mapping operators is in that the former conserves the order statistics of the input image but the latter does not. A tone mapping algorithm is said to preserve order statistics if it maintains the rank of any two pixels' intensity values x_i and x_j , namely $T(x_i) \leq T(x_j)$ if $x_i \leq x_j$ with T being the tone mapping function. Global histogram transform methods, being a point process, necessarily require the non-decreasing monotonicity of $T(\cdot)$ and hence they are order statistics preserving. However, preserving order statistics globally is unnecessarily restrictive in most application scenarios, reducing the effectiveness of an image enhancement method.

On the other hand, local tone mapping do not preserve order statistics because they are local operators in the image domain without imposing relative rank consistency of the processed pixels. Not preserving order statistics lends the existing HDR image enhancement methods greater room for manipulating grey levels to achieve stronger enhancement effects than histogram transform methods, but it also has bad side effects. As relative brightness of object surfaces is an important cue to human visual cognition, HVS is predisposed to distortions of order statistics in an image. Those objectionable artifacts of dynamic range compression, identified at the beginning of this section, are all caused by changed image order statistics within a local proximity in which HVS is sensitive to relative ranking of pixel intensities (Mantiuk *et al.*, 2006).

The strengths and weaknesses of both local and global methods lie in the choice of whether to preserve order statistics globally; while preserving order statistics globally

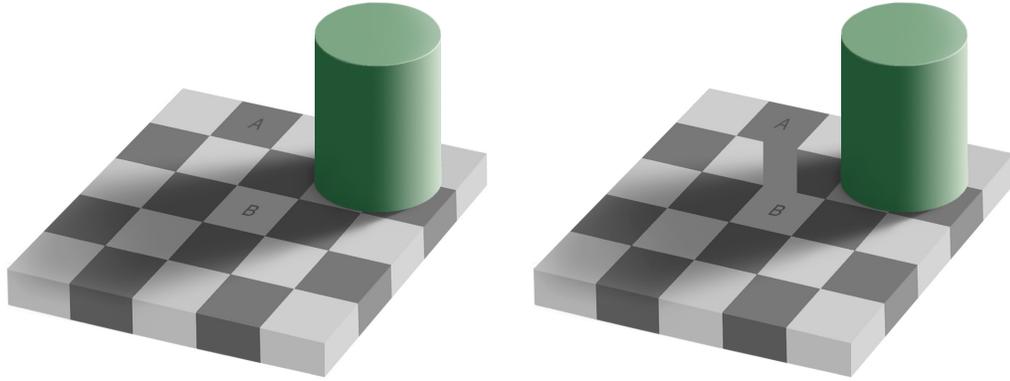


Figure 3.1: Adelson’s checker shadow illusion. In the left image, tile A appears to be much darker than B to a human viewer despite the fact that they have the exact same grey level. However, when a bridge connecting A and B is added as in the right image, the illusion disappears and their shades look identical.

is often too restrictive for tone mapping, ignoring order statistics causes many undesirable artifacts. Conventional tone mapping methods have to sacrifice one merit or another. However, by only preserving order statistics in local regions, a method can be perceptually artifact-free like global methods while not compromising its flexibility like local methods. The reason is that, only if two pixels are adjacent, they can be differentiated easily and accurately, and the artifacts due to rank inconsistency such as those in the existing HDR tone mapping methods become more noticeable. As demonstrated in Figure 3.1, it has been shown that a human viewer cannot reliably tell which one of the two distant pixels has higher intensity, especially when their intensities are similar (Adelson, 2000). Thus, forcing order statistics over two distant pixels provides no or little cognitive value; preserving order statistics locally is functionally the same as preserving order statistics globally in terms of preventing perceptual visual artifacts, while the former is far less restrictive than the latter and offers a greater degree of freedom in tone mapping.

Similar to the global case, preserving order statistics locally can also be defined formally as follows. Suppose \mathbf{x} is the original image and \mathbf{y} is the tone mapped image. If for any pixel i and any pixel j in the neighbourhood set D_i of i , $x_i - x_j$ and $y_i - y_j$ have the same sign, i.e.,

$$\text{sgn}(x_i - x_j) = \text{sgn}(y_i - y_j), \quad (3.1)$$

where $\text{sgn}(\cdot)$ is the sign function, then tone mapped image \mathbf{y} of \mathbf{x} is said to preserve order statistics locally. If for each pixel i , its neighbourhood set D_i contains all pixels in the image, then this definition of preserving local order statistics is equivalent to the global case. Thus, preserving order statistics locally is a more generic case of preserving order statistics globally.

3.2 Locally Adaptive Rank-Constrained Optimal Tone Mapping

The ultimate goal of HDR tone mapping is to enhance an image so that its local details become more visible to human viewers without compromising the wholeness and coherence of the image. Several researches have documented that the detail of an image can be boosted reliably by maximizing the sum of the absolute local contrast (Agaian *et al.*, 2007; Celik and Tjahjadi, 2011; Shu and Wu, 2013); moreover, as the previous discussion suggests, the enhanced image can be perceptually artifact-free by preserving order statistics either globally or locally. Marrying these two ideas together

results a new formulation for HDR image tone mapping as follows,

$$\begin{aligned}
& \text{maximize} && \sum_{i=1}^N \frac{|y_i - s_i|}{s_i} \\
& \text{subject to} && 0 \leq y_i \leq 1, \quad i = 1, \dots, N, \\
& && \text{sgn}(x_i - x_j) = \text{sgn}(y_i - y_j), \\
& && i = 1, \dots, N, \quad j \in D_i,
\end{aligned} \tag{3.2}$$

where the enhanced image \mathbf{y} is the variable, and original image \mathbf{x} and the number of pixels, N , are constant to the problem. Variable s_i , representing the average pixel intensity in the neighbourhood of pixel i , is defined as,

$$s_i = \frac{1}{|D_i|} \sum_{j \in D_i} y_j. \tag{3.3}$$

The objective function is simply the summation of the absolute local Weber contrast of all pixels. The first constraint is to bound the dynamic range of \mathbf{y} to $[0, 1]$ and the second constraint is to preserve order statistics locally as defined in Eq. (3.1).

This optimization problem is non-convex and difficult to solve directly, however, since the enhanced image \mathbf{y} is tone-mapped from the original image \mathbf{x} with the constraint of preserving order statistics, these two images are highly correlated. Thus, the denominator s_i in the objective function can be approximated using the local average pixel intensities of \mathbf{x} ,

$$\tilde{s}_i = \frac{1}{|D_i|} \sum_{j \in D_i} x_j. \tag{3.4}$$

In addition, if the tone mapping method also preserves the order statistics between

each pixel and its local average, or in other words, $y_i - s_i$ always shares the same sign with $x_i - \tilde{s}_i$, then

$$\begin{aligned} |y_i - s_i|/\tilde{s}_i &= (y_i - s_i) \cdot \text{sgn}(y_i - s_i)/\tilde{s}_i \\ &= (y_i - s_i) \cdot \text{sgn}(x_i - \tilde{s}_i)/\tilde{s}_i. \end{aligned} \quad (3.5)$$

Since x_i and \tilde{s}_i are constant to the optimization problem, the approximated objective function is linear. The local order statistics preserving constraints in Eq. (3.2) can also be written as equivalent linear inequalities as follows,

$$\begin{cases} y_i - y_j \leq 0 & \text{if } x_i \leq x_j, \\ y_i - y_j \geq 0 & \text{if } x_i \geq x_j. \end{cases} \quad (3.6)$$

Therefore, this reformulated problem is a linear program. However, the optimal solution of the problem often over-enhances the input image resulting black-white output image with extremely high local contrast and little gradient detail. This is because the order statistics preserving constraint alone cannot prevent the contrast of neighbouring pixels being boosted drastically by the objective function. To alleviate the over-enhancement problem, intuitively, not only does the intensity difference $y_i - y_j$ need to share the same sign as $x_i - x_j$, it should also be relatively proportional to $x_i - x_j$ so that if the original intensity difference of two neighbouring pixels is small, the difference should still be relatively small after enhancement. This idea can be implemented by constraining the ratio of $y_i - y_j$ to $x_i - x_j$ within a certain range, i.e.,

$$l \leq \frac{y_i - y_j}{x_i - x_j} \leq h. \quad (3.7)$$

The parameter h sets the maximum allowed contrast gain in order to maintain the smooth and natural look of the input image; and the purpose of parameter l is to prevent flattening local detail for the sake of global contrast gain. Since in Eq. (3.7), the denominator $x_i - x_j$ is 0 when two adjacent pixels x_i, x_j share the same intensity value, to employ this constraint in a linear program, it must be rewritten in an equivalent linear form as follows,

$$\begin{cases} l \cdot (x_i - x_j) \leq y_i - y_j \leq h \cdot (x_i - x_j) & \text{if } x_i \leq x_j, \\ h \cdot (x_i - x_j) \leq y_i - y_j \leq l \cdot (x_i - x_j) & \text{if } x_i \geq x_j. \end{cases} \quad (3.8)$$

Combining the two cases $x_i \leq x_j$ and $x_i \geq x_j$ together, Eq. (3.8) is equivalent to,

$$L_{i,j} \leq y_i - y_j \leq H_{i,j}, \quad (3.9)$$

where

$$\begin{aligned} L_{i,j} &= \min(l \cdot (x_i - x_j), h \cdot (x_i - x_j)), \\ H_{i,j} &= \max(l \cdot (x_i - x_j), h \cdot (x_i - x_j)). \end{aligned} \quad (3.10)$$

Since x_i, x_j, l, h are all constant to the optimization problem, $L_{i,j}, H_{i,j}$ are also constant. As the lower bound l of the local contrast gain must be greater than or equal to 0, the constraint for local contrast gain in Eq. (3.7) implies the property of preserving order statistics locally in the solution, thus, it is not necessary to use local order statistics preserving constraint as in Eq. (3.6) explicitly in the optimization problem.

Incorporating these constraints, a more practical locally adaptive rank-constrained

optimal tone mapping (LARCOTM) problem can be formulated as,

$$\begin{aligned}
& \text{maximize} && \sum_{i=1}^N \frac{|y_i - s_i|}{\tilde{s}_i} \\
& \text{subject to} && 0 \leq y_i \leq 1, \quad i = 1, \dots, N, \\
& && L_{i,j} \leq y_i - y_j \leq H_{i,j}, \quad (3.11) \\
& && \quad \quad \quad i = 1, \dots, N, \quad j \in D_i, \\
& && \text{sgn}(y_i - s_i) = \text{sgn}(x_i - \tilde{s}_i), \\
& && \quad \quad \quad i = 1, \dots, N,
\end{aligned}$$

Constants l, h are user given parameters based on the requirements of applications. To guarantee the existence of at least one feasible solution, those constants should satisfy the following inequalities,

$$\begin{cases} l \leq 1, \\ h \geq 1. \end{cases} \quad (3.12)$$

Obviously, if these inequalities are all true, the original image \mathbf{x} can satisfy all the constraints in the optimization problem hence a feasible solution.

In some applications, however, it might not be desirable or possible to apply all the restrictions as in Eq. (3.14), as a result, there does not necessarily exist a feasible solution for every input image. Ideally, a robust tone mapping method should always produce some results albeit imperfect. To improve the robustness of the new method, we make the pixel dynamic range constraint a soft constraint and arrive at a more

general formulation LARCOTM as follows,

$$\begin{aligned}
& \text{maximize} && \sum_{i=1}^N \sum_{j \in D_i} |y_i - y_j| - \lambda \sum_{i=1}^N |z_i| \\
& \text{subject to} && a_i \leq y_i + z_i \leq b_i, \quad i = 1, \dots, N, \\
& && L_{i,j} \leq y_i - y_j \leq H_{i,j}, \\
& && i = 1, \dots, N, \quad j \in D_i,
\end{aligned} \tag{3.13}$$

where the enhanced image is $\mathbf{y} + \mathbf{z}$ and \mathbf{z} is a relaxation vector of the pixel dynamic range constraint. Unlike the previous formulation, if the new tone mapping method cannot find a feasible solution y_i within dynamic range $[a_i, b_i]$, it does not give up but tries to clip y_i to a value $y_i + z_i$ within the given dynamic range. The amount of the clipping, $-z_i$, is penalized in the objective function to constrain the compromise. The strength of the penalization is adjustable through setting the Lagrange multiplier λ ; the larger the multiplier, the lower enhancement strength and the lower clipping artifacts in the resulted image as well. With this new formulation in Eq. (3.13), a feasible solution exists regardless the constants $a_i, b_i, l_{i,j}, h_{i,j}$. For example, the following solution must be feasible.

$$\begin{cases} \mathbf{y} = \mathbf{x}, \\ \mathbf{z} = \mathbf{b} - \mathbf{x}. \end{cases} \tag{3.14}$$

Thus, this new method can always yield some results. This LARCOTM problem remains one of linear programming and hence it is still solvable.

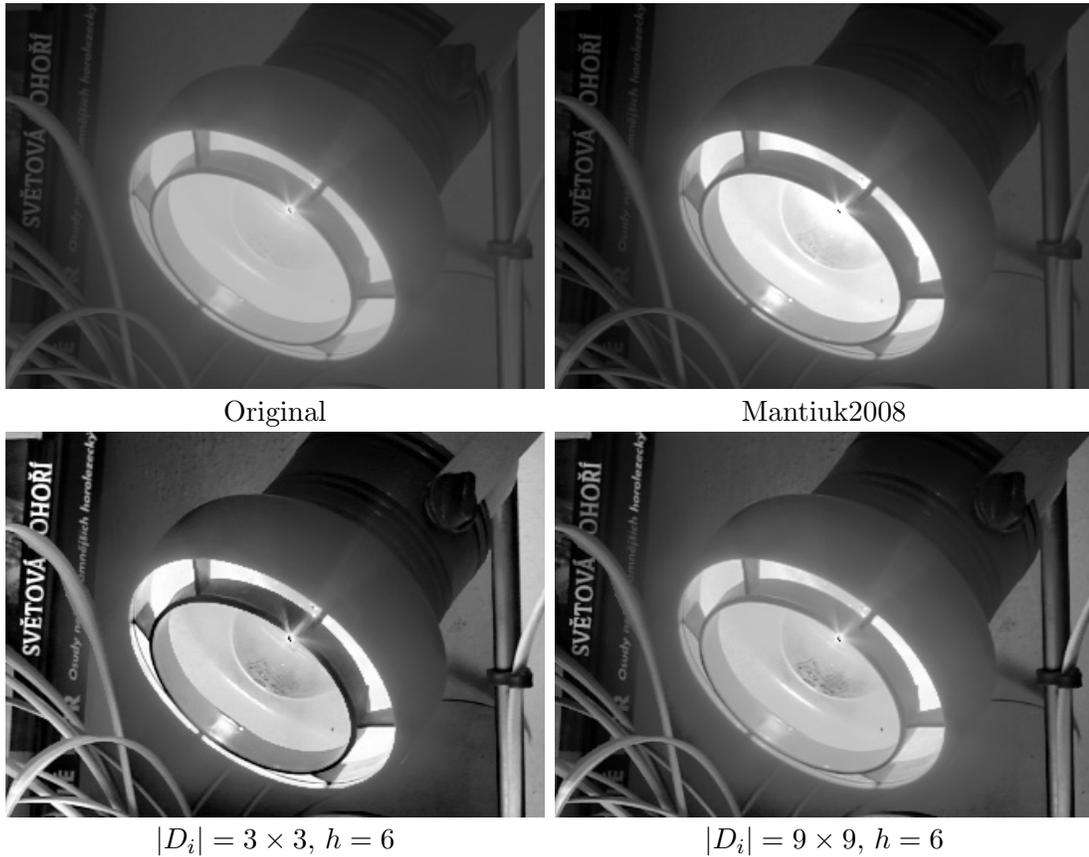


Figure 3.2: This example shows the effect of different sizes of neighbourhood window D_i . Generally, the output of LARCOTM is of higher contrast for smaller D_i , and the output looks more natural and closer to the result of a global tone mapping operator such as Mantiuk’s method, when D_i is larger.

3.3 Experimental Results

By its formulation, LARCOTM strictly preserves the order statistics between each pixel i and its adjacent pixels in a neighbourhood set D_i . As demonstrated in Figure 3.2, when the neighbourhood D_i is a small 3×3 window centred at pixel i , the result of LARCOTM has high level of contrast and its detail is much more visible than the original. In comparison, using a larger neighbourhood D_i , such as a 9×9 window, the tone-mapped image exhibits relatively lower contrast but looks more

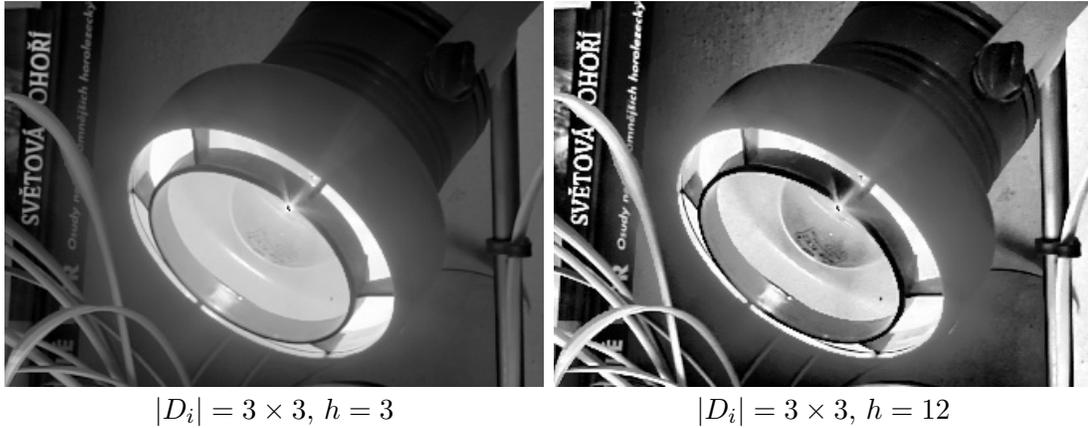


Figure 3.3: This example demonstrates the effect of contrast gain upper bound h . As expected, larger h setting results higher contrast of the output images.

natural. Both settings of neighbourhood sizes yield better detail than a global tone mapping operator such as Mantiuk’s global method (Mantiuk *et al.*, 2008). On the other hand, similar to a global operator, images enhanced by LARCOTM are free of halo and double edge artifacts. As smaller window results higher contrast without causing artifacts, in all of the following presented experiments, the neighbourhood D_i is set to a 3×3 window for all pixels.

Another important parameter of LARCOTM is the upper bound h of contrast gain. In Figure 3.2, h is set to 6, and in Figure 3.3, h is set to 3 and 12, respectively. As expected, with higher upper bound h , the enhanced images have higher contrast grain. However, over exaggerating contrast might distort the relative intensity of different regions resulting unrealistic output images. Thus, the best choice of h depends on the requirements of application, input image and user preferences. Since LARCOTM maximizes local Weber contrast, its results are less sensitive to the choice of the lower bound l of contrast gain. In the presented experiments, the lower bound l is all set to 0.5.

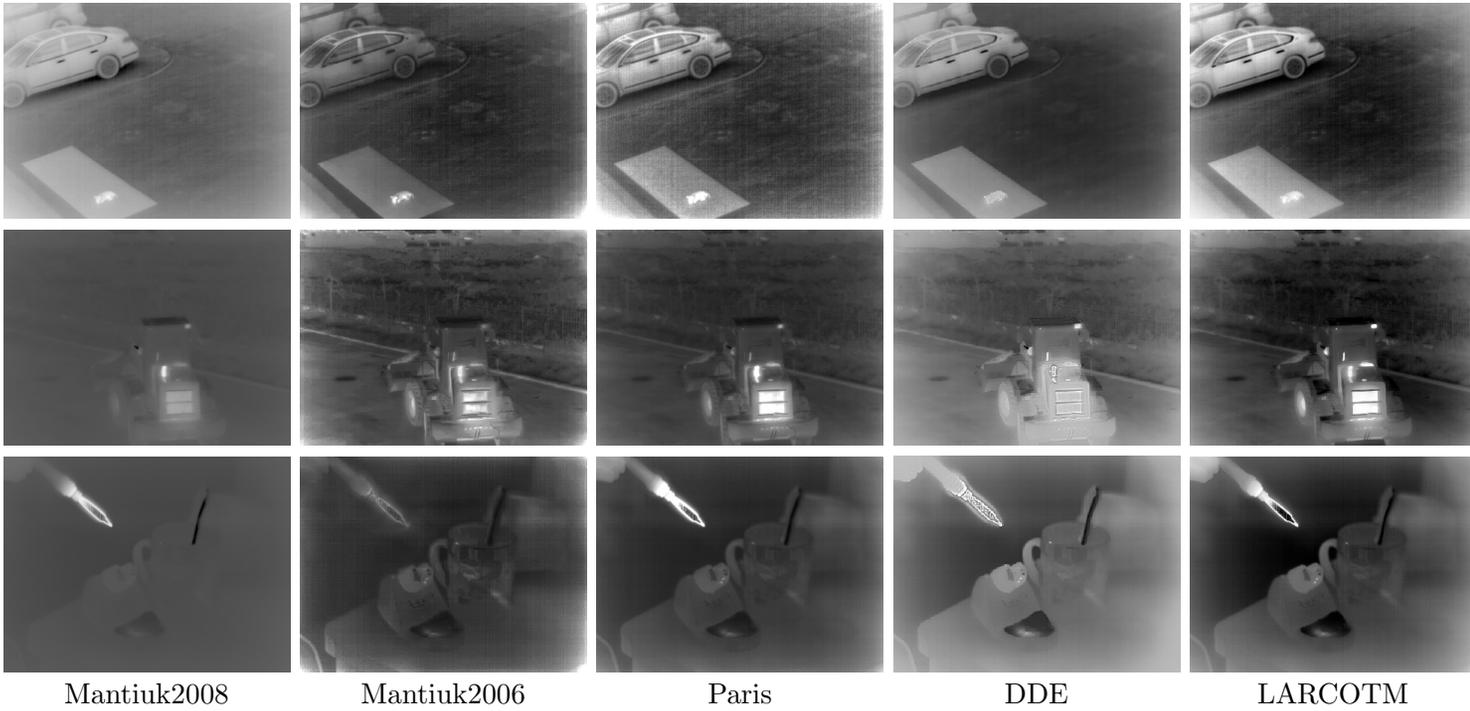


Figure 3.4: 12-bit input infrared images and the enhanced images by Mantiuk's global and local methods, Paris' method, DDE and the proposed LARCOTM method, respectively.

In order to demonstrate the efficacy of the LARCOTM principle, we implemented LARCOTM for various application scenarios. The first test case is the visualization of infrared HDR images as presented in Figure 3.4. The previously mentioned global method by Mantiuk et al. (Mantiuk2008) (Mantiuk *et al.*, 2008), a local approach also by Mantiuk et al. (Mantiuk2006) (Mantiuk *et al.*, 2006), Paris et al.'s local Laplacian filtering method (Paris *et al.*, 2011) and the well-known commercial DDE software (DDE, 2014) in the field of infrared imaging are also tested for comparison purposes. All the four local methods are superior than the global method in terms of exposing image details. Comparatively, while the output images of both Mantiuk's local method and Paris' method show the excellent level of detail, they are overly sensitive to noise and often magnify noise in low quality images. The output images of DDE also have good detail; however, the objects in those images appear flattened and unnatural. By using high-boost filter to enhance edges the DDE method is prone to double-edge artifacts, distorting natural boundaries between objects. The output images of LARCOTM are obtained with the upper bound h of contrast gain set to 8. The results look sharper and more realistic than the competitors, and the relative intensity (temperature) of adjacent objects remains intact in the enhanced images. For example, in the bulldozer case, a human viewer can easily and accurately distinguish the relative temperatures of different parts of the bulldozer in the enhanced infrared image, whereas the other methods can even reverse the ranking of adjacent pixels in temperature (unacceptable in some critical applications). This advantage of LARCOTM can be further visualized using order statistics error image as presented in Figure 3.5. The intensity e_i of each pixel i in these images represents the maximum order statistics error between pixel i in the output image and its neighbours in a 7×7

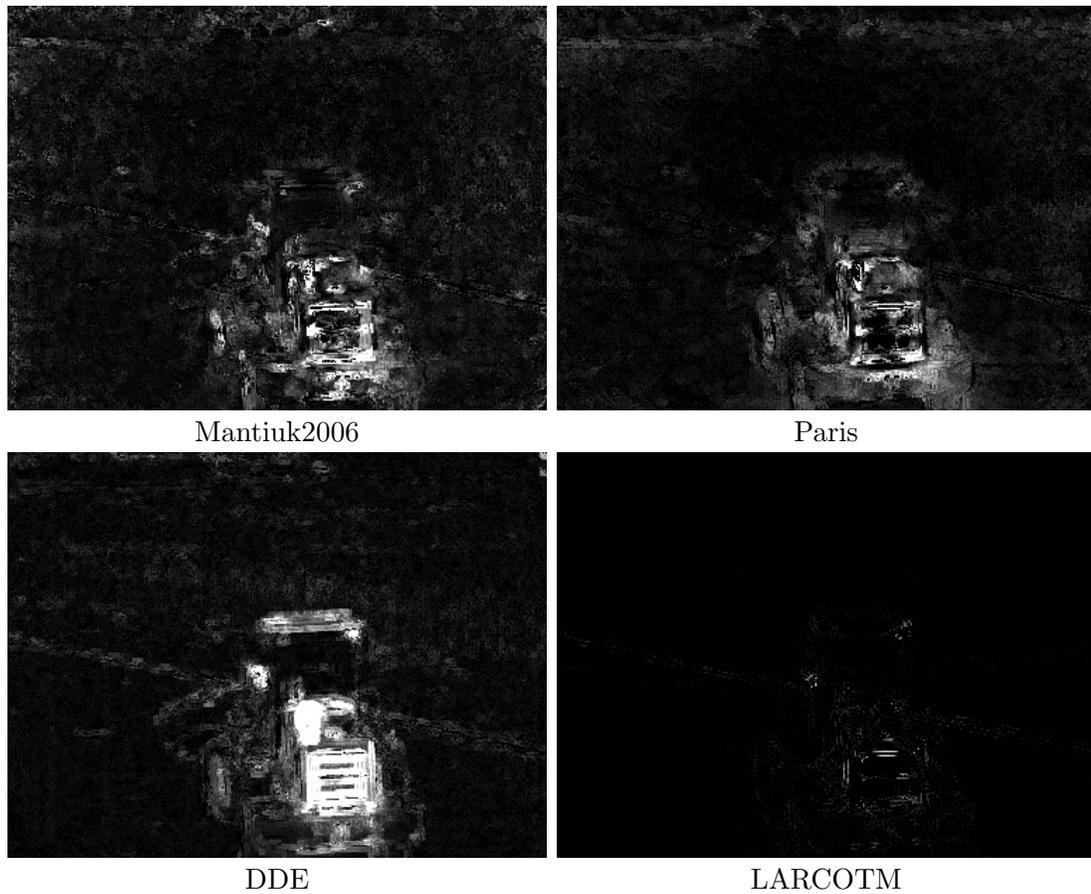


Figure 3.5: The order statistics error images for the bulldozer case. The brighter a pixel is, the larger order statistics error at that location is generated by the corresponding tone mapping method.

window, i.e.,

$$e_i = \max_{j \in D_i} f(i, j), \quad (3.15)$$

where $f(i, j) = 0$ if $\text{sgn}(x_i - x_j) = \text{sgn}(y_i - y_j)$, or $f(i, j) = |x_i - x_j| + |y_i - y_j|$, otherwise. Although in the experiment, LARCOTM is set to preserve the order statistics only in a small 3×3 window, the order statistics in the result is often preserved in a much larger region due to the transitivity of the order relation of pixel intensities. Therefore, as demonstrated by Figure 3.5, the result image of LARCOTM has much smaller order statistics error than its competitors even when we check the error in a large window. Mantiuk's global method is not included in this test because global methods produce no order statistic error by their nature.

The LARCOTM algorithm can also be used for medical imaging. The same CT test image in Figure 1.1 is now enhanced by the proposed LARCOTM algorithm and displayed as an integral image with all pieces of anatomy unified as they physically are (see Figure 3.6). Contrasting the holistic approach of Figure 3.6 with the separation approach of Figure 1.1, radiologists in a small, informal comparative study prefer the new LARCOTM visualization approach.

Compared to the previous infrared image case, two small adjustments are made on the LARCOTM algorithm in retooling it for CT images. First, to ensure that each type of tissue (which can be identified by the intensity) maps to a certain dynamic range as required by radiologist, the first constraint of Eq. (3.11) is replaced with a more restrictive output dynamic range constraint $a_i \leq y_i \leq b_i$, where a_i and b_i are constants determined by the tissue type of each pixel i . Second, the upper bound of contrast gain h is set to 18 for soft tissue pixels and 3 for other pixels, separately. The reason for having a greater enhancement strength for soft tissues is that the intensity

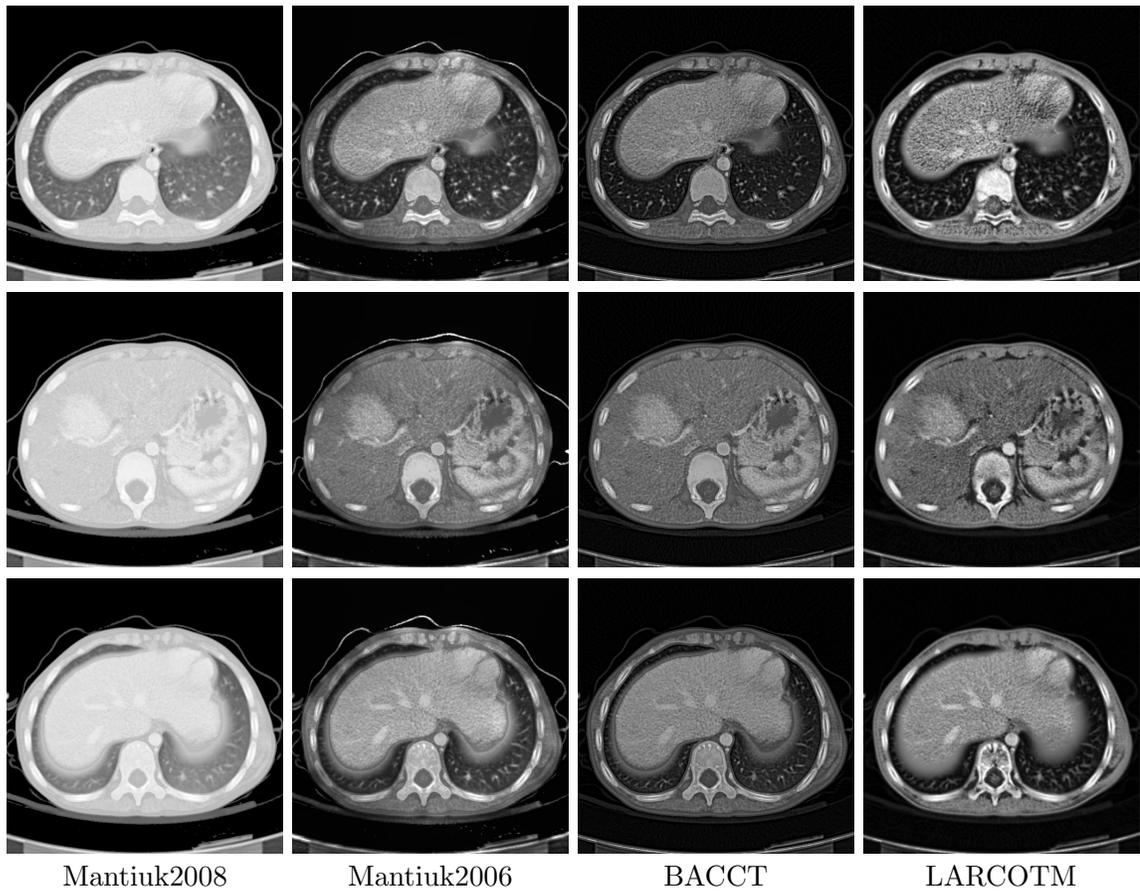


Figure 3.6: 12-bit input CT images and the enhanced images by Mantiuk's global and local methods, BACCT and the proposed LARCOTM method, respectively.

of soft tissue pixels falls in a very small dynamic range, thus, even if the amplitude of soft tissue pixels is allowed to increase by three times more than others, it is still not sufficient to expose all the detail of the soft tissues. To address this difficulty facing all CT image enhancement methods Cohen-Duwek et al. introduced a pre-process phase to increase the dynamic range of soft tissue region with a histogram-based transfer function in their BACCT algorithm (Cohen-Duwek *et al.*, 2011). Although this method does improve the visualization of soft tissues, it often generates false edges around soft tissue regions, hindering the diagnosis. The proposed method, however, can easily allocate more dynamic range resources to soft tissues within the LARCOTM optimization framework without introducing any artifacts. Since the dynamic range of soft tissue is known a priori and it does not overlap with the dynamic ranges of other objects, assigning the soft tissue regions a higher contrast gain h will enhance those regions without much affecting other regions. A general purpose HDR tone mapping technique is also tested for comparison (Mantiuk *et al.*, 2006). Since Mantiuk's global and local methods are not optimized for CT image, the test images have to be pre-processed to boost the contrast of soft tissue using the subroutine of the BACCT algorithm before using the HDR tone mapping techniques, thus these two methods inherit the false edge artifacts from BACCT inevitably. Moreover, as Mantiuk's local method provides no mechanism to preserve the pixel rank consistency and constrain output dynamic range, the same tissue might have noticeable different shades in different positions.

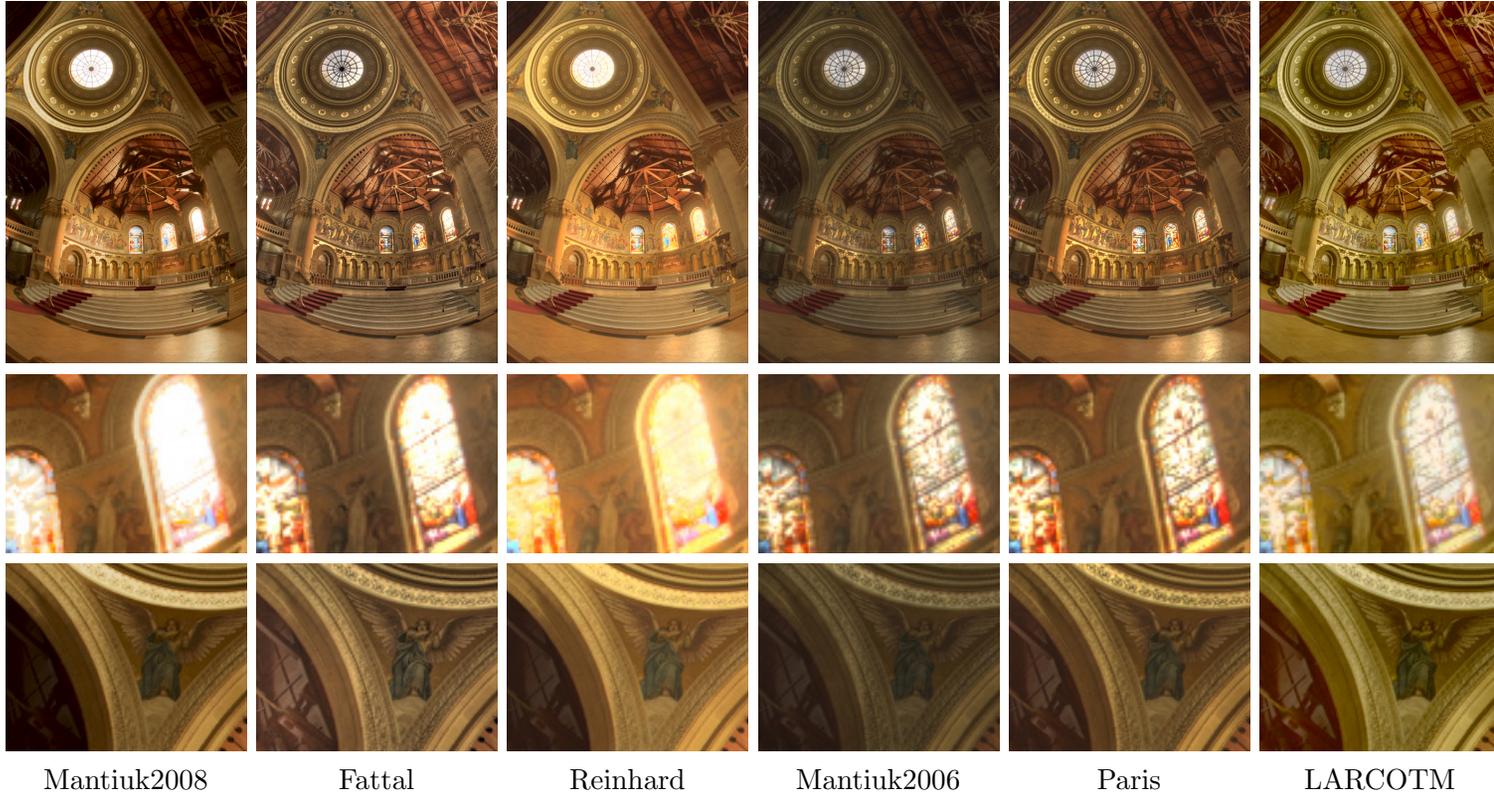


Figure 3.7: Tone mapped natural HDR images by different methods. The second and third rows contain magnified patches of the images in the first row.

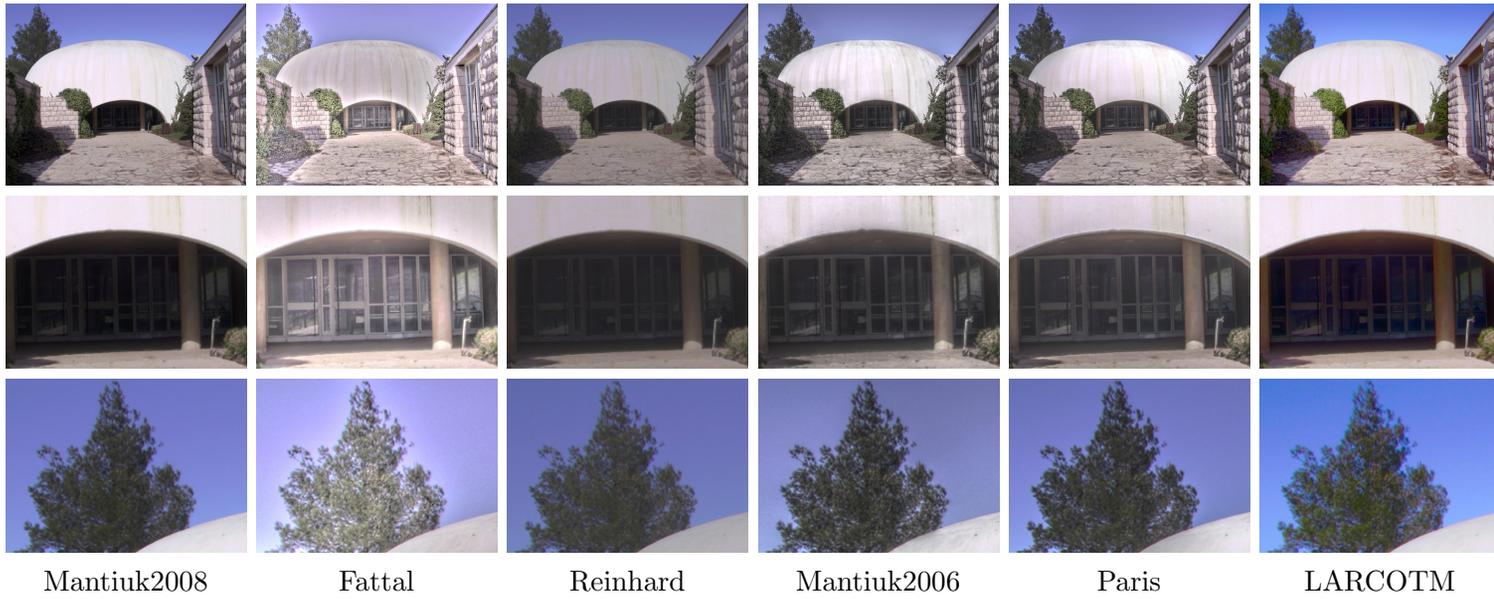


Figure 3.8: Tone mapped natural HDR images by different methods. The second and third rows contain magnified patches of the images in the first row.

The LARCOTM algorithm is also applicable to natural HDR images; its performance is competitive compared to other HDR tone mapping techniques, as shown in Figures 3.7 and 3.8. In this test case, the upper bound of contrast gain h is set to 4. For colour images, we run the LARCOTM algorithm on intensity channel only and keep the colour unchanged. Beside Mantiuk's and Paris' methods, the comparison group includes other two popular methods by Reinhard et al. (Reinhard *et al.*, 2002) and Fattal et al. (Fattal *et al.*, 2002). Similar to the other tested methods, LARCOTM exhibits rich details in both bright (as in the second row of Figure 3.7) and dark (as in the third row of Figure 3.7 and the second row of Figure 3.8) regions, and it is free of halo artifacts in high-contrast regions (as in the third row of Figure 3.8). As a whole, the results of LARCOTM appear more realistic because of its ability to preserve order statistics.

3.4 Conclusion

The results of this research substantiate the validity of the LARCOTM high dynamic range image display methodology. Operationally, the proposed LARCOTM framework is also simple to use yet flexible enough to be fine tuned for meeting stringent requirements of different applications, in particular for medical and scientific visualization.

Chapter 4

Enhancement of Compressed Image

In most practical image enhancement scenarios, particularly those of consumer applications, the input images are compressed in DCT domain with some loss of fidelity. For practical systems constrained by bandwidth and storage economy, lossy compression is inevitable because mathematically invertible image coding typically achieves only roughly 2:1 compression ratio (Wu and Memon, 1997), still leaving the image file size too large to handle.

Granted, after years of research, development and investment, international compression standards such as JPEG, H.264, HEVC, etc., can offer very high reconstruction quality to the level of perceptual transparency; namely, naked eyes cannot easily discern the differences between the original and the decompressed images. But as demonstrated by this work, small compression noises, despite being transparent to human eyes, can adversely affect the results of many image processing operations,

if left unaccounted for. Especially, compression noises are highly detrimental to inverse operators of high-boosting (sharpening) nature, such as image enhancement, deblurring and superresolution against a convolution kernel.

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 Figure 4.1. 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 enhancement or size magnification, or both, which are mundane manipulations in multimedia presentations on large displays. The problem has troubled many users for a long time but no satisfactory solutions have yet been found. As compound images mix acquired and synthetic contents, they cannot be compressed effectively as vector graphics. Content providers and users have to apply compression standards for natural images/videos on compound images.

4.1 Quantization Error in DCT Domain

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 modelled as random variables following an independent and identically distributed (i.i.d.) Gaussian distribution, while true signal has repetitive patterns hence sparse in some basis (Mairal *et al.*, 2009; Dong *et al.*, 2011). By exploiting this statistical difference between signal and sensor noise, denoising techniques can



Figure 4.1: JPEG compression artifacts become highly objectionable after image enhancement or image size magnification, or both. The second and third sub-figures from left are regions up-scaled by bi-cubic interpolation and A+ (Timofte *et al.*, 2014), respectively.

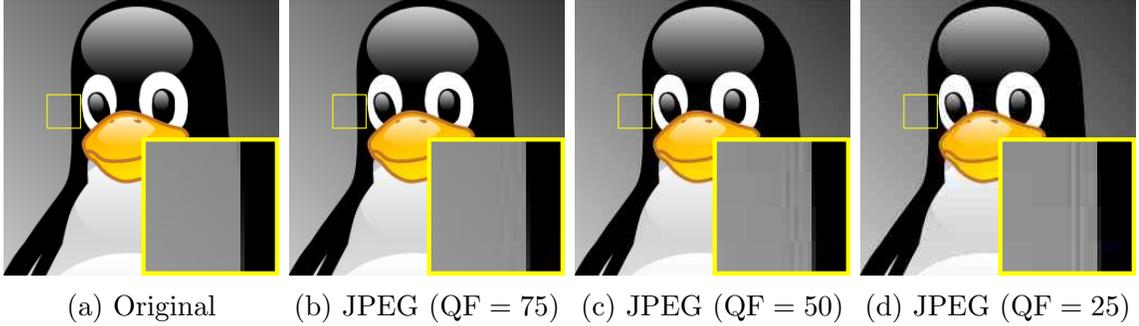


Figure 4.2: JPEG compression noise is corrected with signal.

effectively separate signal and noise in a given noisy observation (Gu *et al.*, 2014). Compression noises, on the other hand, are much more difficult to model than other degradation sources, e.g., motion blur and sensor noises. The non-linearity of quantization operations in image compression systems makes quantization noises image dependent, far from being white and independent.

In main stream DCT-based compression systems, the encoding of signal \mathbf{x} is a three-step process. 1. The discrete cosine transform T is performed on signal \mathbf{x} ; 2. the transformed signal $T(\mathbf{x})$ is subject to quantization Q ; 3. the quantized version $(Q \circ T)(\mathbf{x})$ is coded by an entropy coder C , resulting the code stream $(C \circ Q \circ T)(\mathbf{x})$ for storage or transmission. The decoding process reverses the above three-step encoding process and generates the decompressed signal

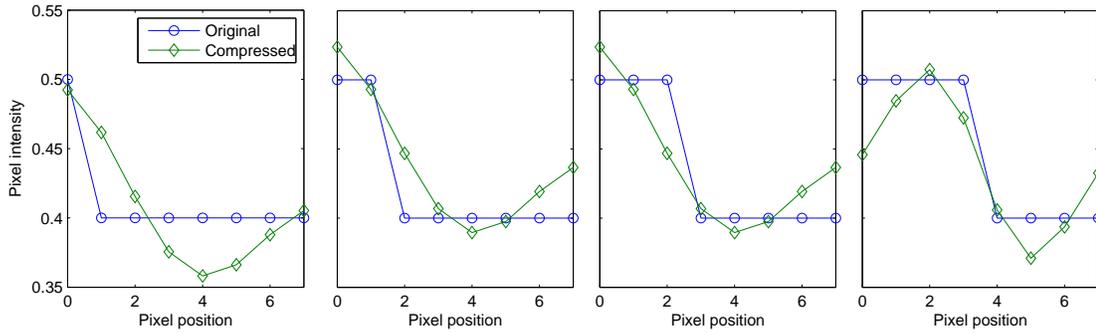
$$\hat{\mathbf{x}} = (T^{-1} \circ Q^{-1} \circ C^{-1})((C \circ Q \circ T)(\mathbf{x})). \quad (4.1)$$

In this closed loop, the entropy decoder C^{-1} and the inverse transform T^{-1} are invertible operators, namely, $C^{-1} \circ C = I$, $T^{-1} \circ T = I$, but the dequantization operator Q^{-1} is not. The approximation error due to $Q^{-1} \circ Q \neq I$ is aggravated and complicated by the non-linearity of the quantization operation Q . In the interest

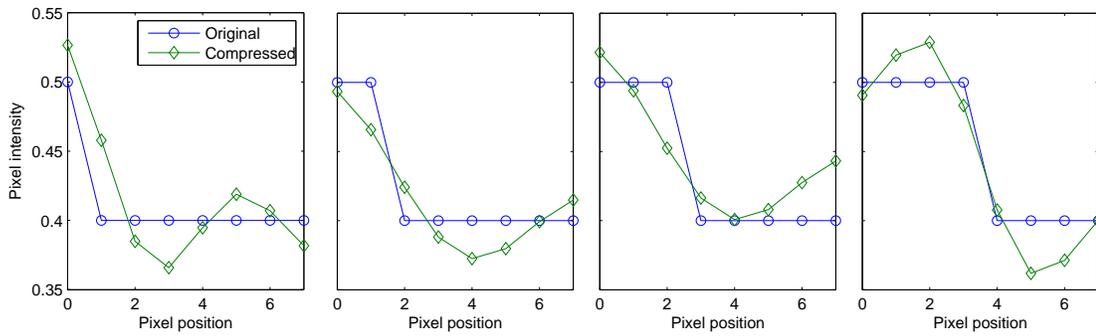
of gaining compression performance, the quantizer Q inclines to demote or outright discard high-frequency DCT coefficients. Setting high frequency components of \mathbf{x} to zero causes periodic ringing artifacts in the reconstructed signal $\hat{\mathbf{x}}$, which are easy to perceive as demonstrated in Figure 4.2. In this set of JPEG-decompressed images, the ringing artifacts not only accompany sharp edges in close proximity and they also agree with the image signal in orientation; in other words, the quantization noises are correlated with the image signal.

Unlike other noise mechanisms in image or video restoration, compression noises are not random in the sense that coding blocks of similar high-frequency contents tend to have similar ringing artifacts. As a result, a particular artifact pattern may occur repetitively in a pixel vicinity. Such signal-dependent noises may resist the treatment of sparsity-based denoising techniques, because the assumption that only the signal as self-similarity is no longer valid.

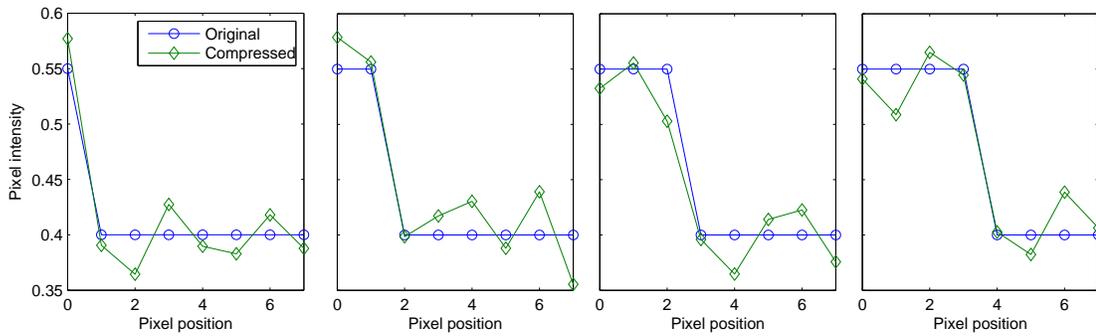
There is yet another complication in modeling and removing DCT quantization errors. That is, the same signal structure can, after through the loop of compression $C \circ Q \circ T$ and decompression $T^{-1} \circ Q^{-1} \circ C^{-1}$, exhibit much varied temporal or spatial patterns, with even immaterial changes in the phase or amplitude of the input signal, and in compression quality factor (QF). This high sensitivity and nonlinearity of error patterns are depicted graphically in Figure 4.3. In this example, the reconstructed versions of the same one dimensional (1D) unit pulse signal of minor linear-type alterations, such as shifting and scaling, behave drastically differently. In two dimensional (2D) image, there are more factors affecting quantization noise. Figure 4.4 shows the quantization effects on image blocks of the same sharp edge but different phases and angles. In each case, there are visible false lines parallel to the edge, however,



(a) Amplitude = 0.1, QF = 45.

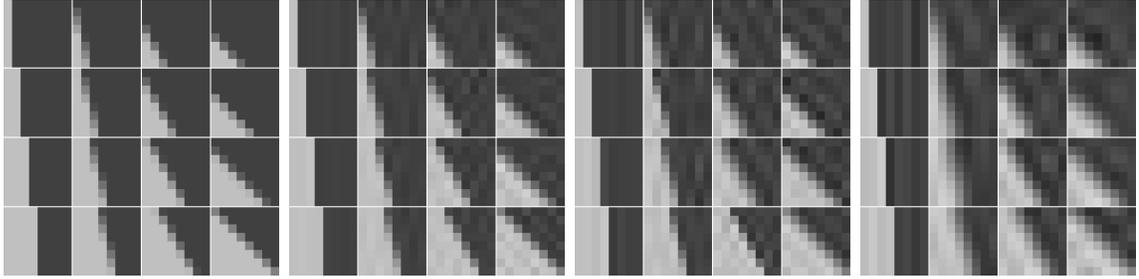


(b) Amplitude = 0.1, QF = 50.



(c) Amplitude = 0.15, QF = 50.

Figure 4.3: DCT quantization noise appears drastically different with small changes in QF, signal phase or amplitude.



(a) Original (b) JPEG (QF = 75) (c) JPEG (QF = 50) (d) JPEG (QF = 25)

Figure 4.4: Besides QF, JPEG compression noise is sensitive to the angle and phase of the signal.

the position, strength and sign of the compression noise vary with a small change in angle or phase. Since there are so many factors affecting the quantization error, using learning based techniques to build a map from noisy observation to true signal for each scenario is impractical.

4.2 DCT Quantization Error Model

DCT based lossy compression techniques realize data volume reduction by trading off the accuracy of the DCT-domain representation of the input signal through quantization. By the definition of DCT, the k -th DCT coefficient of 1D signal x_0, \dots, x_{N-1} is,

$$X_k = \sum_{n=0}^{N-1} x_n \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} \right) k \right]. \quad (4.2)$$

After quantization, the true value of X_k is commonly estimated as,

$$\hat{X}_k = \lfloor X_k / q_k + 0.5 \rfloor \cdot q_k, \quad (4.3)$$

where q_k is quantization interval for the k -th DCT coefficient. In general, q_k is set to decrease with k , due to the fact that most energy of a signal is commonly concentrated in low frequency components.

Using Fourier transform, DCT can be approximated in continuous domain as follows,

$$\begin{aligned}
 X_k &= N \sum_{n=0}^{N-1} \frac{1}{N} x_n \cos \left(2\pi \cdot \frac{n + \frac{1}{2}}{N} \cdot \frac{k}{2} \right) \\
 &\approx N \int_0^1 f(t) \cos \left(2\pi t \frac{k}{2} \right) dt \\
 &= \operatorname{Re} \left[\frac{N}{2} \int_{-\infty}^{\infty} f(t) e^{-2\pi t \frac{k}{2}} dt \right] \\
 &= \operatorname{Re} \left[\frac{N}{2} F \left(\frac{k}{2} \right) \right], \tag{4.4}
 \end{aligned}$$

where f is an integrable function such that

$$\begin{cases} f\left(\frac{n+\frac{1}{2}}{N}\right) = x_n, \\ f(t) = f(-t), \\ f(t) = 0, \quad t > 1, \end{cases} \tag{4.5}$$

and F is the Fourier transform of f . By this equation, if sequence x_0, \dots, x_{N-1} consists of equally spaced samples of function f and f satisfies Eq. (4.5), then a DCT coefficient of the sequence is a sample of f in frequency domain.

4.2.1 Quantization Effects on Linear Signal

Suppose input signal $\mathbf{x}_r = \{x_0, \dots, x_{N-1}\}$ is a decreasing linear sequence, in which the n -th element is,

$$x_n = a \cdot \frac{n + \frac{1}{2}}{N}, \quad 0 \leq n \leq N - 1. \quad (4.6)$$

Then triangular function $a \cdot \text{tri}(t)$, where,

$$\text{tri}(t) = \begin{cases} 1 - |t| & \text{if } |t| < 1 \\ 0 & \text{otherwise,} \end{cases} \quad (4.7)$$

satisfies the conditions in Eq. (4.5), thus by Eq. (4.4), the k -th DCT coefficient of the sequence can be approximated as,

$$\begin{aligned} X_k &\approx \text{Re} \left[\frac{N}{2} \int_{-\infty}^{\infty} a \cdot \text{tri}(t) e^{-i2\pi t \frac{k}{2}} dt \right] \\ &= \frac{aN}{2} \cdot \text{sinc}^2 \left(\frac{k}{2} \right), \end{aligned} \quad (4.8)$$

where $\text{sinc}(\cdot)$ is the normalized sinc function defined as,

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}. \quad (4.9)$$

This linear input signal \mathbf{x}_r is easy to model in temporal domain; its second order derivative is zero everywhere hence sparse. It is comparable to a simple gradient ramp in 2D digital image. However, as visualized in Figure 4.5, this signal is not sparse in DCT domain. By the approximation in Eq. (4.8), the k -th DCT coefficient X_k is only zero for even positive integer k , thus, more than half ($\lfloor N/2 \rfloor + 1$ out of N) of the

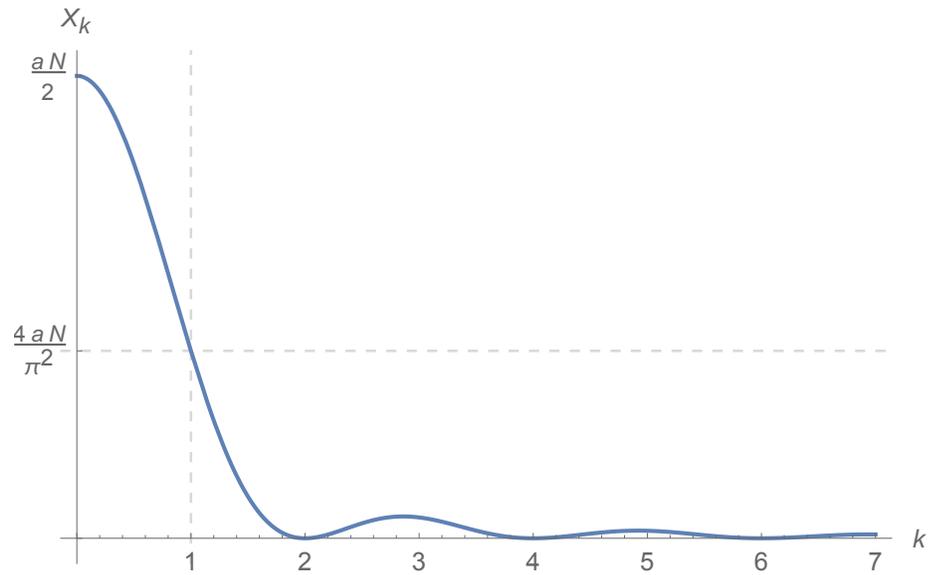


Figure 4.5: DCT coefficients of a decreasing linear sequence.

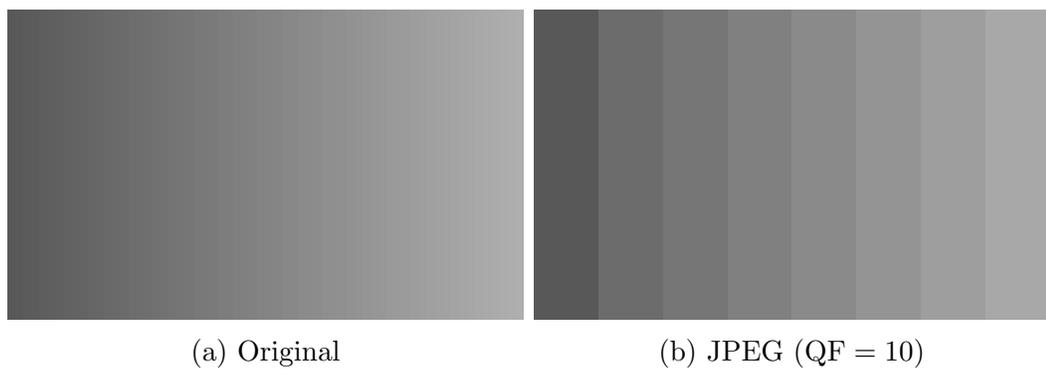


Figure 4.6: Quantization removes small AC components causing perceivable blocking artifacts in a simple gradient ramp image.

coefficients in DCT domain are non-zero. To effectively compress this signal, some of the non-zero DCT coefficients need to be quantized to zero. Suppose for some odd positive integer k_0 , the k_0 -th quantized DCT coefficient is zero, then

$$\begin{aligned}
 X_{k_0} < \frac{q_{k_0}}{2} &\iff \frac{aN}{2} \cdot \text{sinc}^2\left(\frac{k_0}{2}\right) < \frac{q_{k_0}}{2} \\
 &\iff aN \cdot \frac{4}{\pi^2 k_0^2} < q_{k_0} \\
 &\iff a < \frac{\pi^2 k_0^2 q_{k_0}}{4N}.
 \end{aligned} \tag{4.10}$$

Since non-zero DCT coefficient X_k decreases with k while in general, quantization interval q_k increases with k , all the quantized DCT coefficients after the k_0 -th one are zero as well. For example, if we use JPEG with QF = 25 to compress a horizontal gradient ramp, then by Eq. (4.10), quantized DCT coefficient X_1 is zero when a is less than about 4.8. In this case, each coding block becomes uniform after compression as its AC components in DCT domain are all zeros. As shown in Figure 4.6, it is not sufficient to reconstruct the gradient ramp accurately in each block with only the DC component X_0 . More importantly, due to Mach bands illusion, the discontinuity around coding block boundaries is highly perceivable to human, greatly deteriorating the perceptual quality of the compressed image.

4.2.2 Quantization Effects on Piecewise Constant Signal

Similar to linear signal, piecewise constant signal is another case which is simple to model in temporal domain but complex in DCT domain. For instance, let input signal

$\mathbf{x}_s = \{x_0, \dots, x_{N-1}\}$ be a sequence of two steps, i.e.,

$$\underbrace{[a, a, \dots, a]}_m, \underbrace{[0, 0, \dots, 0]}_{N-m}, \quad (4.11)$$

where $a > 0$ and $0 \leq m \leq N$. This sequence is a discrete version of rectangular function $f_s(t)$, where,

$$\begin{aligned} f_s(t) &= a \cdot \text{rect} \left(\frac{t}{2r} \right) \\ &= a \cdot \begin{cases} 0 & \text{if } |t| > r \\ \frac{1}{2} & \text{if } |t| = r \\ 1 & \text{if } |t| < r, \end{cases} \end{aligned} \quad (4.12)$$

and $r = m/N$. As $f_s(t)$ satisfies Eq. (4.5), the DCT of sequence \mathbf{x}_s can be approximated as follows by Eq. (4.4),

$$\begin{aligned} X_k &\approx \text{Re} \left[\frac{N}{2} \int_{-\infty}^{\infty} a \cdot \text{rect} \left(\frac{t}{2r} \right) e^{-i2\pi t \frac{k}{2}} dt \right] \\ &= \frac{N}{2} \cdot 2ar \cdot \text{sinc} \left(2r \cdot \frac{k}{2} \right) \\ &= arN \cdot \text{sinc}(rk). \end{aligned} \quad (4.13)$$

As $\text{sinc}(rk)$ decreases with frequency k in general, if quantization intervals are large enough, quantization effects can be approximated by cutting off high frequency

components. Suppose only the first b DCT coefficients are preserved after quantization, then the restored sequence is

$$\begin{aligned}
 \hat{x}_t &= \int_{-b}^b \frac{2}{N} X_k e^{i2\pi t \frac{k}{2}} d\frac{k}{2} \\
 &= a \int_{-b}^b r \operatorname{sinc}(rk) e^{i2\pi \frac{t}{2} k} dk \\
 &= a \cdot [\operatorname{Si}(br - bt) + \operatorname{Si}(br + bt)],
 \end{aligned} \tag{4.14}$$

where function $\operatorname{Si}(z)$ is sine integral defined as

$$\operatorname{Si}(z) = \int_0^z \operatorname{sinc}(t) dt. \tag{4.15}$$

By aligning the sequence $\{\hat{x}_t | 0 \leq t \leq 1\}$ to the location of the edge, we get the following sequence such that y_0 is the transition of two steps,

$$\hat{y}_t = \hat{x}_{t+r} = a[\operatorname{Si}(-bt) + \operatorname{Si}(2br + bt)]. \tag{4.16}$$

As shown in Figure 4.7, quantization noise in \hat{y}_t has a relatively fixed pattern regardless of the phase. Therefore, if we align the signals by their phases, the noises become aligned as well. This correlation between signal and quantization noise makes them much more difficult to distinguish.

If sequence \mathbf{x}_s is smoothed with a Gaussian kernel resulting sequence \mathbf{y}_s , then \mathbf{y}_s is a discrete version of $f_g(t) = (f_s * g)(t)$, where,

$$g(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}}. \tag{4.17}$$

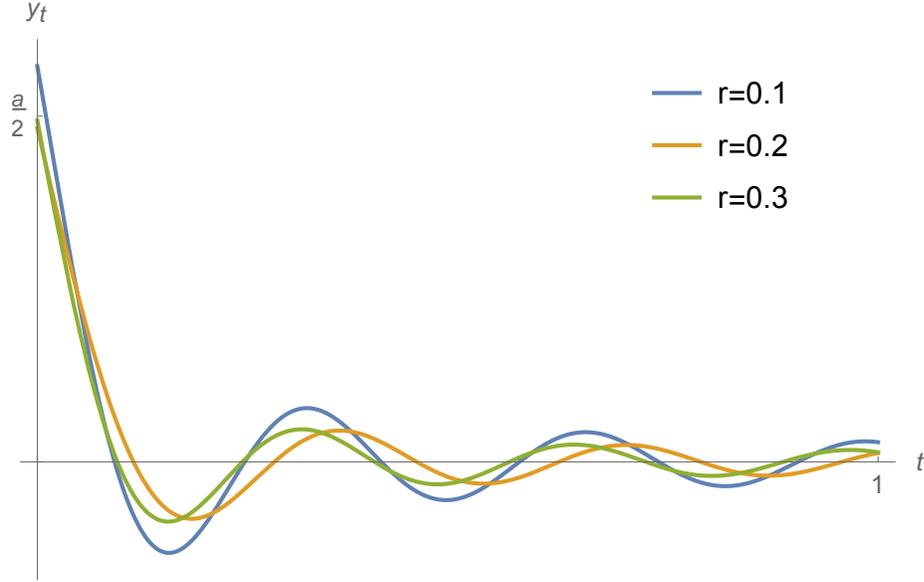


Figure 4.7: A sharp edge causes similar quantization artifacts regardless of the phase r .

By convolution theorem, the Fourier transform of function $f_g(t)$ is,

$$\begin{aligned} F_g\left(\frac{k}{2}\right) &= F_s\left(\frac{k}{2}\right) \cdot G\left(\frac{k}{2}\right) \\ &= 2ar \cdot \text{sinc}(rk) \cdot e^{-2\pi^2\sigma^2k^2/2^2} \end{aligned} \quad (4.18)$$

Thus the k -th DCT coefficient of the blurred sequence \mathbf{y}_s is approximately equal to,

$$Y_k \approx arN \cdot \text{sinc}(rk) \cdot e^{-\pi^2\sigma^2k^2/2} \quad (4.19)$$

For a given frequency k , the absolute quantization error is

$$\epsilon_k = |v_k \cdot q_k - Y_k|. \quad (4.20)$$

where $v_k = \lfloor Y_k/q_k + 0.5 \rfloor$. If the absolute quantization error ϵ_k of DCT coefficient Y_k is sufficiently small, say less than a constant C_ϵ , it has little impact on the quality of the compressed image; if ϵ_k is large, but the relative error $\epsilon_k/|Y_k|$ is small, it still contributes little to the artifacts of the compressed image, as in this case, the quantized DCT coefficient is strong enough to hide the error perceptually. Suppose that to hide the quantization artifacts from frequency k , the relative error $\epsilon_k/|Y_k|$ must be less than $1/3$, i.e.,

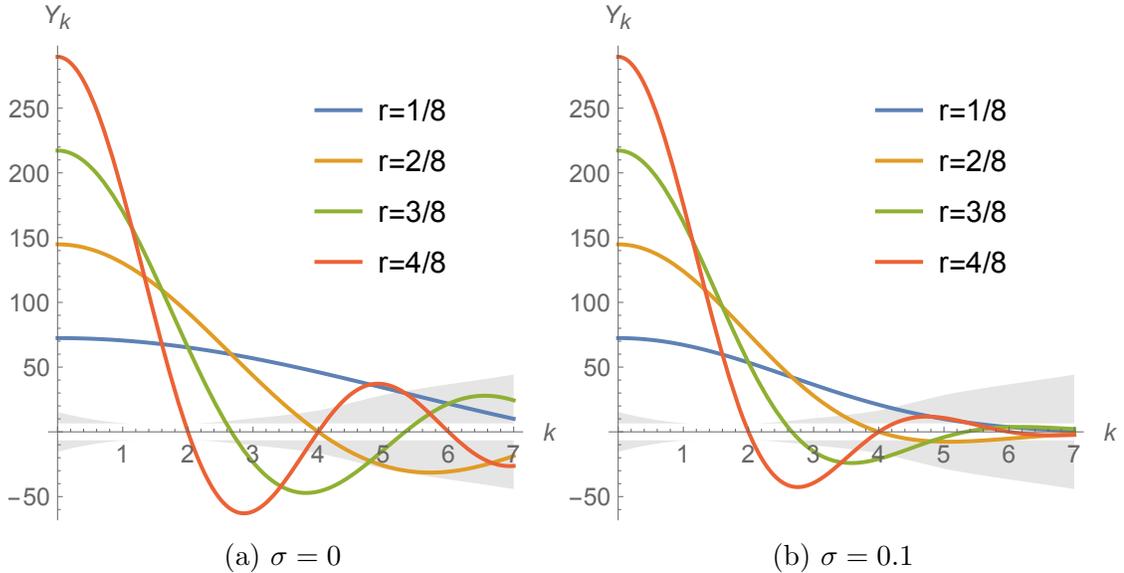
$$\frac{e_k}{|Y_k|} = \left| \frac{v_k \cdot q_k - Y_k}{Y_k} \right| < \frac{1}{3}. \quad (4.21)$$

This inequality is true if and only if $|Y_k| \geq 3q_k/4$. Thus, when,

$$C_\epsilon \leq |Y_k| \leq \frac{3}{4}q_k, \quad (4.22)$$

quantization of DCT coefficient Y_k results both large absolute error and relative error.

Plotted in Figure 4.8 is DCT coefficient Y_k as a function of k . Each curve represents a signal with a different phase r , and amplitude of the first step is $a = 50$; and the quantization intervals are based on the quantization matrix of JPEG with QF = 50. Regions where a coefficient can cause large absolute error and relative quantization error are marked as gray. As demonstrated in Figure 4.8a, the strength and sign of a DCT coefficient and its quantization error depend on various factors, e.g., phase r , smoothness σ and amplitude a . In Figure 4.8b, the sequence is smoothed by a Gaussian kernel with variance σ^2 . As a result, the quantization errors of high frequency coefficients are small (not in the gray regions) compare to the previous case, as those coefficients are close to zero.

Figure 4.8: DCT coefficient Y_k as a function of frequency k .

4.3 Enhancement Model

Recent nonlocal self-similarity (NNS) based image denoising techniques, such as BM3D (Dabov *et al.*, 2006), SAIST (Dong *et al.*, 2013) and WNNM (Gu *et al.*, 2014), have demonstrated their great strength in reconstructing the original image \mathbf{x} from an observation $\mathbf{y} = \mathbf{x} + \mathbf{n}$ contaminated by additive white Gaussian noise \mathbf{n} . NNS refers to the fact that there are many repeated local patterns across a natural image, and those nonlocal similar patches to a given patch can help much the reconstruction of it (Cai *et al.*, 2010). For a local patch \mathbf{y}_i of size m in image \mathbf{y} , we can stack M of its similar patches $\mathbf{y}_{i,j} = \mathbf{R}_{i,j}\mathbf{y}$ across the image together into a matrix $\mathbf{Y}_i \in \mathbb{R}^{m \times M}$, where $\mathbf{R}_{i,j}$ is a matrix extracting the j -th similar patch of the local patch at location i for $1 \leq i \leq N, 1 \leq j \leq M$. Then solving the following nuclear norm minimization (NNM) problem yields a matrix $\hat{\mathbf{X}}_i$ consisting of noise reduced

patches,

$$\hat{\mathbf{X}}_i = \underset{\mathbf{X}_i}{\operatorname{argmin}} \|\mathbf{Y}_i - \mathbf{X}_i\|_F^2 + \lambda \|\mathbf{X}_i\|_*, \quad (4.23)$$

where $\|\cdot\|_F$ and $\|\cdot\|_*$ represent the Frobenius norm and nuclear norm of a matrix, respectively. Although this problem is non-convex, it is tractable by an efficient singular value thresholding (SVT) algorithm (Cai *et al.*, 2010). The whole reconstructed image can be then estimated by aggregating all the denoised patches as,

$$\hat{\mathbf{x}} = \left(\sum_{i=1}^N \sum_{j=1}^M \mathbf{R}_{i,j}^\top \mathbf{R}_{i,j} \right)^{-1} \sum_{i=1}^N \sum_{j=1}^M \mathbf{R}_{i,j}^\top \hat{\mathbf{x}}_{i,j} \quad (4.24)$$

Following this idea, we can formulate the enhancement of DCT-domain compressed image problem as a constrained nuclear norm minimization problem,

$$\begin{aligned} \hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \quad & \sum_{i=1}^N \|\mathbf{X}_i\|_* \\ \text{s. t.} \quad & |\mathbf{Q}\mathbf{T}\tilde{\mathbf{R}}_i\mathbf{H}\mathbf{x} - \gamma_i| \leq 0.5, \\ & i = 1 \dots n \end{aligned} \quad (4.25)$$

where \mathbf{H} is a matrix modelling the degradation of image quality caused by various image capturing conditions, $\tilde{\mathbf{R}}_i$ is a matrix extracting the coding block at location i , \mathbf{T} is the DCT transform matrix, \mathbf{Q} is a diagonal matrix storing quantization table and vector γ_i is the DCT coefficient of the block at location i .

Using hard-decoding technique, each block \mathbf{y}_i of \mathbf{y} is obtained by inverse DCT transform from coefficient vector γ_i as follow,

$$\mathbf{y}_i = \tilde{\mathbf{R}}_i \mathbf{y} = (\mathbf{Q}\mathbf{T})^{-1} \gamma_i, \quad (4.26)$$

and the observed image \mathbf{y} is a degraded version of image $\mathbf{H}\mathbf{x}$ contaminated mainly by DCT-domain quantization noise. Existing sparsity based denoising techniques designed for reducing additive white Gaussian noise generally use $\|\hat{\mathbf{x}} - \mathbf{y}\|_2$ as the fidelity term in their optimization frameworks and leave $\hat{\mathbf{x}}$ unconstrained. In the case of DCT-domain quantization noise, we have more information about the noise: the true value of each DCT coefficient before scalar quantization lies in a known interval,

$$\mathbf{Q}^{-1}(\gamma_i - 0.5) \leq \mathbf{T}\tilde{\mathbf{R}}_i\mathbf{H}\mathbf{x} < \mathbf{Q}^{-1}(\gamma_i + 0.5) \quad (4.27)$$

This constrain confines the solution space of the optimization problem in Eq. (4.25), preventing the sparsity objective function from over-smoothing the output image.

To solve the problem in Eq. (4.25), we split it into two parts. The first part is to find a sparse estimation $\hat{\mathbf{z}}$ of the original image from a given noisy version $\tilde{\mathbf{x}}$, i.e., for each patch group of image $\hat{\mathbf{z}}$,

$$\hat{\mathbf{Z}}_i = \underset{\mathbf{Z}_i}{\operatorname{argmin}} \|\tilde{\mathbf{X}}_i - \mathbf{Z}_i\|_F^2 + \lambda \|\mathbf{Z}_i\|_* \quad (4.28)$$

The noisy version $\tilde{\mathbf{x}}$ of the original image can be estimated directly using $\tilde{\mathbf{x}} = \mathbf{H}^{-1}\mathbf{y}$. Here we use \mathbf{H}^{-1} to represent an inverse operator of \mathbf{H} rather than matrix inverse. Although many types of image degradation can be modelled by a simple product of a degradation matrix \mathbf{H} and the original image \mathbf{x} , the inverse problem is often ill-posed and requires complex non-linear algorithm to find a good solution. Since the observed image \mathbf{y} contains compression noise, if operator \mathbf{H}^{-1} exhibits high-boosting property, which is often the case for unsharp and edge enhancement operators, \mathbf{H}^{-1} could amplify the noise and make an inaccurate estimation of the original image.

Using sparsity prior, the boosted noise can be greatly alleviated by the optimization problem in Eq. (4.28), resulting a better estimation of the original image in vector $\hat{\mathbf{z}}$.

The second part of the problem is to impose the DCT-domain constraint in Eq. (4.25) on the noise reduced estimation $\hat{\mathbf{z}}$ from the first part using the following optimization problem,

$$\begin{aligned} \hat{\mathbf{x}}' = \operatorname{argmin}_{\mathbf{x}} \quad & \|\mathbf{x} - \hat{\mathbf{z}}\|_2 \\ \text{s. t.} \quad & |QT\tilde{\mathbf{R}}_i\mathbf{H}\mathbf{x} - \gamma_i| \leq 0.5, \\ & i = 1 \dots n \end{aligned} \quad (4.29)$$

This is a convex problem solvable by off-the-shelf convex optimization problem solvers. However, if the input image is large, a general purpose solver is too time-consuming for this problem. Instead, we can solve a similar but much simpler problem as follows,

$$\begin{aligned} \hat{\mathbf{x}}' = \operatorname{argmin}_{\mathbf{x}} \quad & \sum_{i=1}^n \|\tilde{\mathbf{R}}_i\mathbf{H}(\mathbf{x} - \hat{\mathbf{z}})\|_2 \\ \text{s. t.} \quad & |QT\tilde{\mathbf{R}}_i\mathbf{H}\mathbf{x} - \gamma_i| \leq 0.5, \\ & i = 1 \dots n \end{aligned} \quad (4.30)$$

Compared with the original problem, the only difference of the reduced problem is that the new problem measures the norm of the error in degraded image domain rather than original image domain. Since the DCT tranform matrix \mathbf{T} is unitary,

$$\begin{aligned} \|\tilde{\mathbf{R}}_i\mathbf{H}(\mathbf{x} - \hat{\mathbf{z}})\|_2 &= \|\mathbf{T}\tilde{\mathbf{R}}_i\mathbf{H}(\mathbf{x} - \hat{\mathbf{z}})\|_2 \\ &= \|\mathbf{T}\mathbf{y}_i - \mathbf{T}\tilde{\mathbf{R}}_i\mathbf{H}\hat{\mathbf{z}}\|_2, \end{aligned} \quad (4.31)$$

where coding block $\mathbf{y}_i = \tilde{\mathbf{R}}_i \mathbf{H} \mathbf{x}$. On the other hand, the DCT coefficient constraint in Eq. (4.27) is applied on each element of vector $\mathbf{T} \mathbf{y}_i$, thus, the optimization problem has a closed-form solution,

$$\hat{\mathbf{y}}_i = \mathcal{C}_{0.5}(\tilde{\mathbf{R}}_i \mathbf{H} \hat{\mathbf{z}}, \mathbf{y}_i), \quad (4.32)$$

where, the DCT-domain clipping operator $\mathcal{C}_\beta(\cdot, \cdot)$ is defined as

$$\mathcal{C}_\beta(\boldsymbol{\alpha}, \boldsymbol{\rho}) = (\mathbf{QT})^{-1} \min(\max(\mathbf{QT} \boldsymbol{\alpha}, \mathbf{QT} \boldsymbol{\rho} - \beta), \mathbf{QT} \boldsymbol{\rho} + \beta). \quad (4.33)$$

Aggregating these DCT blocks together, we get $\hat{\mathbf{y}}$, an estimation in degraded image domain with reduced compression noise, from which an approximate solution $\hat{\mathbf{x}}' = \mathbf{H}^{-1} \hat{\mathbf{y}}$ of the problem in Eq. (4.30) can be easily found. Compared with image $\hat{\mathbf{x}}$, the initial inverse of the observed image \mathbf{y} , $\hat{\mathbf{x}}'$ has lower level of compression noise because of sparsity prior in Eq. (4.28) and still satisfies the DCT domain constraints due to Eq. (4.30).

4.4 Algorithm

Based on the enhancement model discussed in the previous section, our purposed algorithm can be implemented as an iterative process alternatively finding a reconstructed original image and a compression noise reduced observation image as in Algorithm 1. In this section, we address some of the technical issues in the implementation of the algorithm.

Algorithm 1 Image enhancement from compressed image

Input: Compressed image \mathbf{y} , contrast degradation matrix \mathbf{H}

- 1: Estimate compression RMSE ε of \mathbf{y}
- 2: Estimate threshold λ using ε ▷ Eq. (4.43)
- 3: $\beta = 0.2$
- 4: $\mathbf{x}^{(0)} = \mathbf{H}^{-1}\mathbf{y}$
- 5: **for** $k = 1$ to K **do**
- 6: **for each** patch \mathbf{x}_i in $\mathbf{x}^{(k-1)}$ **do**
- 7: Find similar patch group \mathbf{X}_i
- 8: $[\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}] = \text{SVD}(\mathbf{X}_i)$
- 9: $\mathbf{Z}_i = \mathbf{U}\mathcal{T}_\lambda(\mathbf{\Sigma})\mathbf{V}^\top$ ▷ Eq. (4.38)
- 10: **end for**
- 11: Aggregate \mathbf{Z}_i to form image $\mathbf{z}^{(k)}$ ▷ Eq. (4.24)
- 12: **if** $k = K$ **then**
- 13: $\beta = 0.5$
- 14: **end if**
- 15: Clip $\mathbf{H}\mathbf{z}^{(k)}$ using threshold β to get $\mathbf{y}^{(k)}$ ▷ Eq. (4.32)
- 16: $\mathbf{x}^{(k)} = \mathbf{H}^{-1}\mathbf{y}^{(k)}$
- 17: $\lambda = \lambda/2$
- 18: **end for**

Output: Enhanced image $\mathbf{x}^{(K)}$

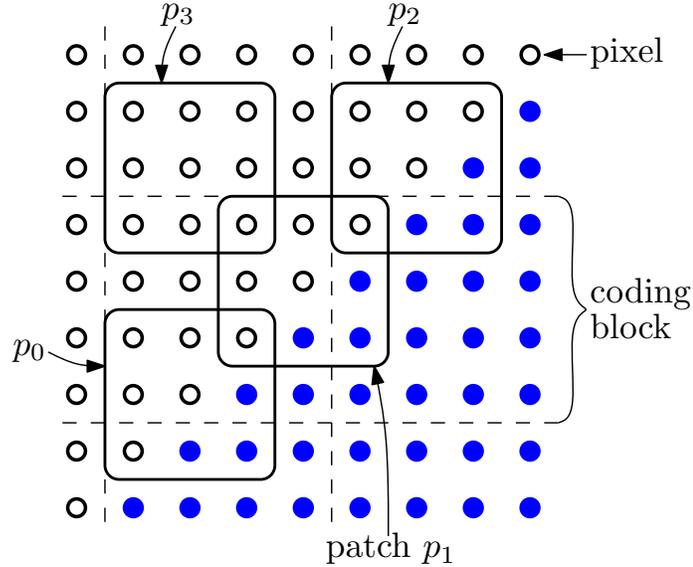


Figure 4.9: Avoid collecting patches of the same phase.

4.4.1 Nonlocal Self-Similarity

NNS based techniques have achieved the state-of-the-art results in removing Gaussian noise (Gu *et al.*, 2014). 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 sections, 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 Figure 4.9, 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 matching 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 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.

In addition to considering patch positions in choosing similar patches, the measurement of patch similarity should be carefully designed to decouple noise from signal. As discussed previously, compression noise in input image can mislead the selection of similar patches. Thus, instead of comparing two patches directly by their squared error, a denoised version of a patch can be used to measure similarity. A simple low-pass filtering can generate a good enough denoised image effectively reducing the impact of compression noise to the measurement of similarity. This technique is only used in the first iteration of a denoising algorithm when the strength of noise is high. In the later iterations, input image becomes less subject to compression noise and it is not necessary to clean up the input image for a more robust measurement of similarity.

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 as demonstrated in Figure 4.6. To prevent 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. For example, in Figure 4.9, patch p_3 is in a smooth area located across two coding blocks; any patch in the same row as p_3 is likely to have the identical blocking artifacts, hence it should not be considered in the patch group of p_3 . Moreover, since natural images are smooth in general, patches in a small windows of smooth area are similar to each other. 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. In practice, we set the search windows to 60×60 for normal patch and shrink the window to 10×10 when the variance of the given patch is less than 3.

4.4.2 Singular Value Thresholding

Ideally, finding a low-rank reconstruction of patch group matrix \mathbf{Y}_i should be formulated as an ℓ_0 -norm minimization problem as follows,

$$\hat{\mathbf{X}}_i = \underset{\mathbf{X}_i}{\operatorname{argmin}} \|\mathbf{Y}_i - \mathbf{X}_i\|_F^2 + \lambda \|\mathbf{X}_i\|_0. \quad (4.34)$$

Since this problem is NP-hard (Cai *et al.*, 2010), in practice, we approximate it with a nuclear norm minimization problem as in Eq. (4.23), which has an efficient closed-form solution,

$$\hat{\mathbf{X}}_i = \mathbf{U} \mathcal{D}_\lambda(\boldsymbol{\Sigma}) \mathbf{V}^\top, \quad (4.35)$$

where \mathbf{U} , $\boldsymbol{\Sigma}$, \mathbf{V} represent the the singular value decomposition (SVD) of \mathbf{Y}_i and $\mathcal{D}_\lambda(\cdot)$ is a soft-thresholding operator,

$$\mathcal{D}_\lambda(\boldsymbol{\Sigma})_{jj} = \begin{cases} \boldsymbol{\Sigma}_{jj} - \lambda & \boldsymbol{\Sigma}_{jj} > \lambda, \\ 0 & \text{otherwise,} \end{cases} \quad (4.36)$$

or simply $\mathcal{D}_\lambda(\boldsymbol{\Sigma})_{jj} = \max(\boldsymbol{\Sigma}_{jj} - \lambda, 0)$.

Although this is a reasonable approximation employed by many applications (Cai *et al.*, 2010; Xie *et al.*, 2014; Dong *et al.*, 2013), it still has some weaknesses. One of its problems is that, in addition to having a lower rank, the Frobenius norm of the optimal solution $\hat{\mathbf{X}}_i$ also decreases with larger threshold λ , since,

$$\|\hat{\mathbf{X}}_i\|_F^2 = \|\mathbf{U} \mathcal{D}_\lambda(\boldsymbol{\Sigma}) \mathbf{V}^\top\|_F^2 = \sum_{j=1}^M \mathcal{D}_\lambda(\boldsymbol{\Sigma})_{jj}^2 \quad (4.37)$$

and $\mathcal{D}_\lambda(\boldsymbol{\Sigma})_{jj}$ is a decreasing function to λ . In the context of image denoising, when we

try to increase the strength of the denoising algorithm by selecting a large threshold λ , it inevitably decreases the second moment of the image reducing the brightness and contrast of the output. Unlike white Gaussian noise, DCT-domain quantization noise could contribute negatively to the second moment of the image, especially when the quality factor is low, hence, image denoising using NNM may pull the result further away from the statistics of the original image.

An intuitive solution to this problem is to completely preserve all the singular values that are above the threshold λ , i.e., to replace the soft-thresholding operator $\mathcal{D}_\lambda(\cdot)$ with a hard-thresholding operator

$$\mathcal{T}_\lambda(\boldsymbol{\Sigma})_{jj} = \begin{cases} \boldsymbol{\Sigma}_{jj} & \boldsymbol{\Sigma}_{jj} > \lambda, \\ 0 & \text{otherwise.} \end{cases} \quad (4.38)$$

Since $\mathcal{T}_\lambda(\boldsymbol{\Sigma})_{jj} > 0$ if and only if $\mathcal{D}_\lambda(\boldsymbol{\Sigma})_{jj} > 0$, the resulting matrices $\hat{\mathbf{X}}_i$ by the two threshold operators have the exact same rank. Thus, the new operator $\mathcal{T}_\lambda(\cdot)$ does not change the low rank property of the solution, however, in this case, the solution is closer to \mathbf{Y}_i statistically in terms of the second moment.

This method coincides with the idea of reweighted nuclear norm minimization where large singular values are given smaller weight to achieve better low rank approximation (Li *et al.*, 2014). It can also be interpreted as a special case of weighted nuclear norm minimization (WNNM) (Gu *et al.*, 2014) as follows. If for $\sigma_j(\mathbf{X}_i)$, the j -th singular value of \mathbf{X}_i , we assign a weight w_j ,

$$w_j = \begin{cases} 0 & \sigma_j(\mathbf{Y}_i) > \lambda, \\ \lambda & \text{otherwise.} \end{cases} \quad (4.39)$$

Since the weights $w_{1\dots M}$ are in a non-descending order, by the theory of WNNM, applying the hard-thresholding operator $\mathcal{T}_\lambda(\cdot)$ on Σ yields an optimal solution for the following optimization problem,

$$\hat{\mathbf{X}}_i = \underset{\mathbf{X}_i}{\operatorname{argmin}} \|\mathbf{Y}_i - \mathbf{X}_i\|_F^2 + \|\mathbf{X}_i\|_{w,*} \quad (4.40)$$

where $\|\mathbf{X}_i\|_{w,*}$ is the weighted sum of the singular values of matrix \mathbf{X}_i .

Now, the question is how to set the parameter λ of the NNM problem making it more effective against DCT-domain quantization noise. In the formulation of Eq. (4.23), λ is a weight balancing the sparse and fidelity regularization terms. If sparsity is given too much weight, it tends to over-smooth the image and cause degradation in brightness and contrast as discussed previously; if the weight is too small, noise remains visible. From the perspective of the solution to the problem in Eq. (4.35), λ is a threshold eliminating small singular values of matrix $\mathbf{Y}_i = \mathbf{U}\Sigma\mathbf{V}^\top$, where row vector $\sigma_j\mathbf{v}_j^\top$ in matrix $\Sigma\mathbf{V}^\top$ consists of the coefficient of each patch in \mathbf{Y}_i with respect to the j -th basis vector in sparse dictionary \mathbf{U} (Dong *et al.*, 2013). Image denoising by sparse optimization is based on the fact that signal is likely sparse under some basis while noise is i.i.d. under the same basis. Furthermore, the energy of compression error generally is small in comparison with the strength of signal, especially when the quantization factor is set to a practical range. Thus, removing small coefficients, which originated most likely from noise than signal, results an output closer to the true signal. The mean square error (MSE) ε^2 of compression can then

be approximated by,

$$\begin{aligned}
\varepsilon^2 &= \frac{1}{mM} \|\mathbf{Y}_i - \mathbf{X}_i\|_F^2 \\
&\approx \frac{1}{mM} \|\mathbf{Y}_i - \hat{\mathbf{X}}_i\|_F^2 \\
&= \frac{1}{mM} \|\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top - \mathbf{U}\mathcal{T}_\lambda(\boldsymbol{\Sigma})\mathbf{V}^\top\|_F^2 \\
&= \frac{1}{mM} \|\boldsymbol{\Sigma} - \mathcal{T}_\lambda(\boldsymbol{\Sigma})\|_F^2 \\
&= \frac{1}{mM} \sum_{j \in L} \sigma_j^2
\end{aligned} \tag{4.41}$$

where set L contains indices of singular values that are less than λ . For the same input image, using a lower quality setting increases the compression noise (i.e., ε^2), which in turn increases the small singular values according to Eq. (4.41). To compensate this, threshold λ must increase as well to keep the size of set L unchanged so that the sparsity of the signal is preserved. Considering that the compression error is i.i.d. on each basis vector, implying that small singular values are of similar strength, threshold λ should be proportional to root mean square error (RMSE) ε as,

$$\varepsilon^2 \propto \frac{1}{mM} \sum_{j \in L} \lambda^2 \tag{4.42}$$

Therefore, with an empirical constant C_λ , threshold λ can be set as

$$\lambda = C_\lambda \varepsilon \sqrt{\frac{mM}{|L|}} \approx C_\lambda \varepsilon \sqrt{\max(M, m)}, \tag{4.43}$$

where we assume $|L| \approx \min(M, m)$ due to the sparsity of the true signal. This threshold selecting method requires the knowledge of the strength of the compression

error ε , which is commonly unknown to the decoder. If error ε is indeed not provided by the encoder, various no-reference peak signal-to-noise ratio (PSNR) estimation techniques (Turaga *et al.*, 2004; Ichigaya *et al.*, 2006; Brandão and Queluz, 2008) can be used to estimate ε with sufficient accuracy for finding an appropriate threshold λ .

4.4.3 DCT Coefficient Constraint

Most iterative denoising techniques, such as (Osher *et al.*, 2005; Dong *et al.*, 2013; Gu *et al.*, 2014), employ some regularization mechanisms to add a portion of filtered noise back to the denoised image in each iteration in order to reduce the loss of high frequency information as the result of multiple rounds of smoothing operators. The idea of adding noise back enables denoising techniques to remove large noise aggressively by over-smoothing the image in the first few iterations without completely removing the detail in the process. Then, during the following iterations, the image can be refined gradually using smoothing operators of lower strength. Mainly designed to deal with Gaussian noise, many of these above mentioned denoising techniques implement this iterative regularization by simply adding the difference between the observed noisy image and smoothed image back to the smoothed image, and use the result as the input noisy image for the next iteration.

For our DCT quantization noise reduction algorithm, this noise feedback method can be written as,

$$\begin{aligned}
 \mathbf{y}_i^{(k)} &= \mathbf{z}_i^{(k)} + \delta(\mathbf{y}_i - \mathbf{z}_i^{(k)}) \\
 &= \delta\mathbf{y}_i + (1 - \delta)\mathbf{z}_i^{(k)} \\
 &= \mathbf{T}^{-1}[\delta\mathbf{T}\mathbf{y}_i + (1 - \delta)\mathbf{T}\mathbf{z}_i^{(k)}]
 \end{aligned} \tag{4.44}$$

where coding block \mathbf{y}_i is at location i in the observed image \mathbf{y} as defined in Eq. (4.32), coding block $\mathbf{y}_i^{(k)}$ is the denoised version of \mathbf{y}_i from the k -th iteration of the algorithm, coding block $\mathbf{z}_i^{(k)} = \tilde{\mathbf{R}}_i \mathbf{H} \mathbf{z}^{(k)}$ is the smoothed block at location i as in Eq. (4.30) and δ is a weight parameter adjusting the strength of noise feedback. As shown in Eq. (4.44), in DCT domain, this noise feedback process finds a weighted average between DCT coefficients $\mathbf{T} \mathbf{z}_i^{(k)}$ of the smoothed image and coefficients $\mathbf{T} \mathbf{y}_i$ of the noisy observation, adding image detail along with some reduced noise back to the result.

Similarly, the clipping operator $\mathcal{C}_\beta(\boldsymbol{\alpha}, \boldsymbol{\rho})$ introduced in the previous section in Eq. (4.33) has the effect of blending the smoothed image $\boldsymbol{\alpha}$ with DCT-domain quantization error tainted observation image $\boldsymbol{\rho}$ as well. By design, the clipping operator finds an image that is close to image $\boldsymbol{\alpha}$ and has all of its DCT coefficients lying within the given quantization intervals $[\mathbf{Q} \mathbf{T} \boldsymbol{\rho} - \beta, \mathbf{Q} \mathbf{T} \boldsymbol{\rho} + \beta]$. The output image $\mathcal{C}_\beta(\boldsymbol{\alpha}, \boldsymbol{\rho})$ is closer to $\boldsymbol{\alpha}$ if threshold β is large, and it is closer to $\boldsymbol{\rho}$ if β is small. Therefore, with adjustable strength using parameter β , the clipping operator is also a suitable noise feedback function for our algorithm as follows,

$$\mathbf{y}_i^{(k)} = \mathcal{C}_\beta(\mathbf{z}_i^{(k)}, \mathbf{y}_i). \quad (4.45)$$

This formulation is the same as the solution to the DCT coefficient constraint problem in Eq. (4.32) except for the threshold β . Since applying the clipping operator multiple times is equivalent to applying it once with the smallest threshold β , i.e.,

$$\mathcal{C}_\beta(\mathcal{C}_{0.5}(\mathbf{z}_i^{(k)}, \mathbf{y}_i), \mathbf{y}_i) = \mathcal{C}_{\min(\beta, 0.5)}(\mathbf{z}_i^{(k)}, \mathbf{y}_i), \quad (4.46)$$

we only need to use the clipping operator once with threshold $\beta \leq 0.5$ to solve the

DCT coefficient constraint problem and add filtered noise back to the result.

By the theory of narrow quantization constraint set (NQCS), the DCT coefficient clipping threshold β should be sufficiently small in order to achieve the optimal results in terms of PSNR (Park and Kim, 1999). For example, the authors of NQCS demonstrated that fixing threshold $\beta = 0.1$ is good enough for various images; several research papers on JPEG image deblocking and denoising reported that setting $\beta = 0.3$ often yields best results (Zhai *et al.*, 2008b; Liew and Yan, 2004; Sun and Cham, 2007). The best choice of β depends on the distributions of the DCT coefficients of the original image, quantization factors and characteristics of the smoothing technique. Although smaller threshold β generates PSNR-plausible results, it often brings blocking and ringing artifacts back to the result, deteriorating its perceptual visual quality.

To alleviate this problem, in the last iteration of our algorithm, instead of solving the optimization problem in Eq. (4.30), whose solution is given in Eq. (4.32) using the clipping operator, we solve a modified problem as follows,

$$\begin{aligned}
 \mathbf{y}^{(k)} = \underset{\mathbf{y}}{\operatorname{argmin}} \quad & \|\nabla^2(\mathbf{y} - \mathbf{H}\mathbf{z}^{(k)})\|_2^2 \\
 & + \alpha \sum_{i=1}^n \|\tilde{\mathbf{R}}_i(\mathbf{y} - \mathbf{H}\mathbf{z}^{(k)})\|_2^2 \\
 \text{s. t.} \quad & |\mathbf{Q}\mathbf{T}\tilde{\mathbf{R}}_i\mathbf{y} - \gamma_i| \leq \beta, \\
 & i = 1 \dots n.
 \end{aligned} \tag{4.47}$$

In addition to minimizing the difference between the smoothed image and output image, the objective function of this modified problem also minimizes the difference between their second order derivatives. This new regularization term encourages

adding filtered noise back to locations that are discontinuous in the smoothed image, so that, artifacts are less noticeable in the output image perceptually. The modified problem in Eq. (4.47) is solvable using augmented Lagrangian method, which is more expensive than the clipping operator in Eq. (4.32) in terms of computational complexity. However, since our algorithm only solves this problem once during the last iteration, this technique can improve the visual quality of the output image without significantly increasing the overall cost.

Alternatively, we can obtain the goal of eliminating the visual artifacts by adjusting DCT clipping threshold β and singular value threshold λ in the last two iterations. The idea is that, if in the last iteration K , most of the DCT coefficients of smoothed block $\mathbf{z}_i^{(K)}$ are already within the quantization intervals $[\mathbf{QT}\mathbf{y}_i - 0.5, \mathbf{QT}\mathbf{y}_i + 0.5]$, then artifacts cannot be reintroduced to the results by the clipping operator in Eq. (4.32) with threshold $\beta^{(K)} = 0.5$. To insure the condition that most DCT coefficients satisfy the quantization interval constraints, the strength of the smoothing operator must be reduced in the last iteration by using a smaller singular value threshold $\lambda^{(K)}$. By Eqs. (4.41) and (4.43), the standard deviation of the difference between the noisy input image \mathbf{y} and smoothed image $\mathbf{z}^{(K)}$ is,

$$\frac{\|\mathbf{y} - \mathbf{z}^{(K)}\|_2}{\sqrt{n}} \approx \frac{\lambda^{(K)}}{C_\lambda \sqrt{\max(M, m)}}, \quad (4.48)$$

which is roughly proportional to threshold $\lambda^{(K)}$, thus, decreasing threshold $\lambda^{(K)}$ also reduces the variance of $\mathbf{y} - \mathbf{z}^{(K)}$ in DCT domain and makes DCT coefficients of image $\mathbf{z}^{(K)}$ more likely stay within quantization interval. On the other hand, the clipping threshold $\beta^{(K-1)}$ in the second last iteration should also be small in order to make each DCT coefficient of the clipped image close to the centre of quantization interval,



Figure 4.10: Several widely used test images.

limiting DCT coefficient overflow caused by the next smoothing operator. However, if both thresholds β and λ are too small, it weakens the effect of noise reduction. In practice, we find that setting clipping threshold $\beta^{(K-1)} = 0.2$ and singular value threshold $\lambda^{(K)} = \lambda^{(1)}/4$ works well for most input images.

4.5 Experimental Results

To demonstrate the performance of the proposed technique, we first turn off the image enhancement part by setting degradation matrix \mathbf{H} as an identity matrix, and compare the results with the state-of-the-art denoising and JPEG artifact removal techniques. The comparison group is composed of the following methods: one JPEG deblocking method: the ACR algorithm (Zhai *et al.*, 2008a); two denoising methods: the BM3D algorithm (Dabov *et al.*, 2006) and WNNM algorithm (Gu *et al.*, 2014); and three JPEG soft-decoding methods: the TV algorithm (Bredies and Holler, 2012), DicTV (Chang *et al.*, 2014) algorithm and DTPD algorithm (Liu *et al.*, 2015). As the denoising approaches BM3D and WNNM are not designed specifically for dealing

Image	JPEG	ACR	BM3D	WNNM	TV	DicTV	DTPD	Proposed
Lenna	30.64	+0.23	+0.92	+0.66	+0.12	+0.48	+1.45	+1.58
Parrot	32.37	+0.28	+0.75	-2.85	-0.17	+0.41	+1.36	+1.68
Hat	31.47	+0.22	+0.89	-2.56	+0.21	+0.45	+1.39	+1.53
Flower	30.10	+0.16	+0.99	-0.48	+0.12	+0.32	+1.49	+1.78
Monarch	28.32	+0.07	+1.32	+1.94	+1.56	+1.39	+2.64	+2.90
Leaves	28.90	+0.26	+1.70	+1.17	+0.70	+1.69	+3.14	+3.44
Barbara	30.41	+0.15	+1.36	+1.43	-1.19	+1.20	+2.81	+3.25
Boat	31.65	+0.34	+1.17	-0.94	-0.27	+0.60	+1.55	+1.84
House	33.72	+0.39	+0.95	-13.98	+0.10	+0.14	+1.70	+1.60
Bike	27.22	+0.08	+1.03	+1.21	+0.06	+0.87	+1.98	+2.27
Median	30.53	+0.22	+1.01	+0.09	+0.11	+0.54	+1.62	+1.81

Table 4.1: PSNR gains (dB) of different denoising algorithms at QF = 25.

Image	JPEG	ACR	BM3D	WNNM	TV	DicTV	DTPD	Proposed
Lenna	32.96	+0.05	+0.80	+0.85	-0.05	+0.06	+1.44	+1.70
Parrot	34.80	+0.10	+0.64	-0.49	-0.39	-0.11	+1.39	+1.63
Hat	33.64	+0.05	+0.89	+0.01	+0.01	-0.01	+1.60	+1.83
Flower	32.43	+0.02	+0.98	+0.46	+0.12	-0.20	+1.68	+2.08
Monarch	30.73	+0.00	+1.32	+1.99	+1.75	+1.33	+2.62	+3.12
Leaves	31.64	+0.04	+1.78	+0.45	+0.64	+1.58	+3.29	+3.86
Barbara	33.56	+0.05	+1.28	+1.68	-1.75	+0.56	+2.51	+2.98
Boat	34.42	+0.09	+1.20	+0.89	-0.64	-0.11	+1.52	+2.04
House	35.79	+0.13	+0.81	-10.87	-0.11	-0.69	+1.48	+1.52
Bike	29.95	+0.01	+1.10	+1.45	-0.15	+0.84	+2.21	+2.64
Median	33.26	+0.05	+1.04	+0.66	-0.08	+0.03	+1.64	+2.06

Table 4.2: PSNR gains (dB) of different denoising algorithms at QF = 50.

with JPEG compression noise, they cannot estimate the compression error from the input JPEG image but require an estimation of the error variance as a user input. To make a fair comparison, we provide the true variance of the compression error to these methods as a known parameter, so their performances should reflect their best results in removing JPEG compression noise.

We select several widely used images in the literature as test images (thumbnailed in Figure 4.10). All images are 256×256 in size. Tables 4.1, 4.2 and 4.3 list the PSNR results of the compared algorithms on the test images compressed using JPEG with QF set to 25, 50 and 80, respectively. As shown in the tables, the proposed

Image	JPEG	ACR	BM3D	WNNM	TV	DicTV	DTPD	Proposed
Lenna	36.59	+0.00	+0.57	+0.53	-0.69	-1.25	+1.22	+1.57
Parrot	38.31	+0.01	+0.46	+0.49	-0.99	-1.45	+1.07	+1.47
Hat	37.27	+0.02	+0.77	+0.96	-0.71	-1.41	+1.65	+2.14
Flower	36.20	+0.00	+0.98	+1.29	-0.57	-1.56	+1.71	+2.32
Monarch	34.73	-0.00	+1.22	+1.87	+1.01	+0.11	+2.33	+3.10
Leaves	35.92	-0.01	+1.74	+2.63	-0.30	-0.02	+3.10	+3.97
Barbara	37.61	+0.00	+0.89	+1.20	-2.59	-1.56	+1.69	+2.09
Boat	38.38	+0.01	+1.04	+1.21	-1.55	-2.29	+1.04	+1.77
House	39.11	+0.01	+0.87	+0.90	-0.88	-2.52	+1.70	+1.98
Bike	34.53	+0.00	+1.12	+1.63	-1.19	-0.52	+2.03	+2.70
Median	36.93	+0.00	+0.94	+1.21	-0.80	-1.43	+1.69	+2.12

Table 4.3: PSNR gains (dB) of different denoising algorithms at QF = 80.

Image	JPEG	ACR	BM3D	WNNM	TV	DicTV	DTPD	Proposed
Lenna	0.8835	+0.0097	+0.0231	+0.0177	+0.0041	+0.0050	+0.0260	+0.0310
Parrot	0.9060	+0.0114	+0.0186	-0.0141	+0.0063	+0.0032	+0.0193	+0.0243
Hat	0.8752	+0.0094	+0.0221	+0.0021	+0.0058	+0.0058	+0.0272	+0.0315
Flower	0.8816	+0.0080	+0.0271	+0.0123	+0.0040	+0.0045	+0.0306	+0.0382
Monarch	0.8969	+0.0069	+0.0425	+0.0453	+0.0396	+0.0371	+0.0514	+0.0536
Leaves	0.9234	+0.0112	+0.0386	+0.0431	+0.0302	+0.0362	+0.0472	+0.0487
Barbara	0.9033	+0.0079	+0.0245	+0.0210	-0.0255	+0.0078	+0.0348	+0.0395
Boat	0.8905	+0.0102	+0.0254	+0.0188	-0.0004	+0.0045	+0.0293	+0.0340
House	0.8741	+0.0060	+0.0110	-0.0431	+0.0006	+0.0024	+0.0168	+0.0166
Bike	0.8798	+0.0055	+0.0259	+0.0180	+0.0019	+0.0134	+0.0401	+0.0458
Median	0.8870	+0.0087	+0.0250	+0.0179	+0.0040	+0.0054	+0.0299	+0.0361

Table 4.4: SSIM gains of different denoising algorithms at QF = 25.

Image	JPEG	ACR	BM3D	WNNM	TV	DicTV	DTPD	Proposed
Lenna	0.9221	+0.0021	+0.0107	+0.0071	-0.0011	-0.0094	+0.0144	+0.0186
Parrot	0.9374	+0.0033	+0.0084	-0.0031	-0.0017	-0.0108	+0.0096	+0.0134
Hat	0.9182	+0.0032	+0.0142	+0.0070	-0.0025	-0.0095	+0.0200	+0.0236
Flower	0.9253	+0.0014	+0.0174	+0.0107	+0.0029	-0.0076	+0.0218	+0.0277
Monarch	0.9300	+0.0009	+0.0282	+0.0294	+0.0278	+0.0229	+0.0350	+0.0377
Leaves	0.9533	+0.0013	+0.0248	+0.0240	+0.0203	+0.0211	+0.0299	+0.0322
Barbara	0.9456	+0.0019	+0.0130	+0.0132	-0.0174	-0.0056	+0.0181	+0.0215
Boat	0.9319	+0.0029	+0.0165	+0.0145	-0.0017	-0.0069	+0.0189	+0.0237
House	0.9103	+0.0015	+0.0040	-0.0369	-0.0085	-0.0166	+0.0100	+0.0124
Bike	0.9280	+0.0008	+0.0162	+0.0130	+0.0022	+0.0005	+0.0276	+0.0321
Median	0.9290	+0.0017	+0.0152	+0.0118	-0.0014	-0.0072	+0.0194	+0.0237

Table 4.5: SSIM gains of different denoising algorithms at QF = 50.

Image	JPEG	ACR	BM3D	WNNM	TV	DicTV	DTPD	Proposed
Lenna	0.9559	+0.0001	+0.0015	-0.0018	-0.0045	-0.0198	+0.0061	+0.0079
Parrot	0.9629	+0.0003	+0.0018	-0.0001	-0.0041	-0.0192	+0.0041	+0.0061
Hat	0.9580	+0.0010	+0.0053	+0.0031	-0.0060	-0.0214	+0.0109	+0.0131
Flower	0.9623	+0.0001	+0.0090	+0.0089	-0.0004	-0.0156	+0.0119	+0.0150
Monarch	0.9626	+0.0000	+0.0121	+0.0120	+0.0131	+0.0050	+0.0174	+0.0191
Leaves	0.9789	-0.0003	+0.0114	+0.0128	+0.0081	+0.0046	+0.0131	+0.0147
Barbara	0.9738	+0.0001	+0.0034	+0.0028	-0.0116	-0.0169	+0.0056	+0.0073
Boat	0.9659	+0.0002	+0.0066	+0.0065	-0.0053	-0.0204	+0.0069	+0.0101
House	0.9530	+0.0002	+0.0040	+0.0023	-0.0133	-0.0346	+0.0111	+0.0121
Bike	0.9679	+0.0000	+0.0087	+0.0086	-0.0011	-0.0118	+0.0130	+0.0156
Median	0.9627	+0.0001	+0.0060	+0.0048	-0.0043	-0.0180	+0.0110	+0.0126

Table 4.6: SSIM gains of different denoising algorithms at QF = 80.

technique improves over the hard-decoded JPEG by around 2dB in PSNR. It leads in PSNR gain in almost every test case and has more than 0.2dB advantage over the second best method. As a reference, we also list objective fidelity assessment results by more sophisticated image quality metric SSIM (Wang *et al.*, 2004) in Tables 4.4, 4.5 and 4.6 for different QF settings. As shown in the tables, the SSIM results also confirm the superiority of the proposed algorithm over the tested technologies.

Compared with other techniques, the proposed technique works consistently well at vastly different QF settings. As demonstrated in Figures. 4.11 and 4.12, the proposed technique is ahead of the competitions at all QF settings except when QF = 5. Only in that case, the proposed technique does not perform as well as DTPD in terms of median PSNR and SSIM gain. Although QF = 5 is often used in JPEG denoising research to showcase the capability of a technique, it has no practical value as compressing a down-scaled version of the input image with slightly larger QF could easily yield better output image than using QF = 5 directly. Furthermore, if we trade off time by increasing the number of iterations K , the proposed technique can outperform DTPD in both PSNR and SSIM while still being faster than DTPD at

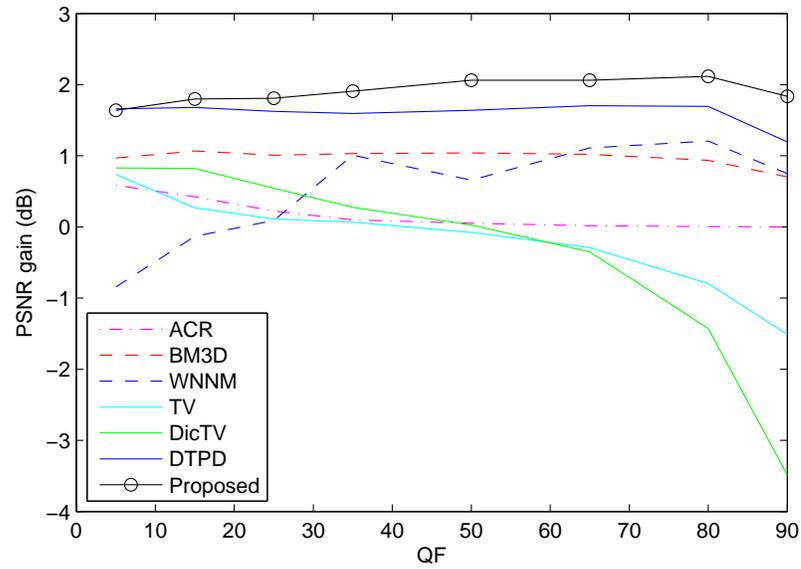


Figure 4.11: The median PSNR gain as a function of QF.

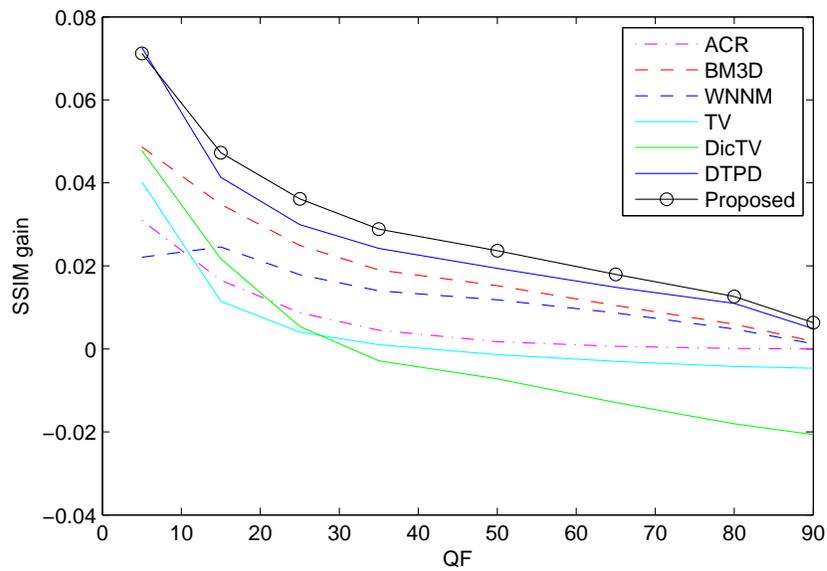


Figure 4.12: The median SSIM gain as a function of QF.

QF = 5.

In addition to its superior performance in objective fidelity metric, the proposed approach also obtains better perceptual quality of the denoised images. As shown in Figures 4.13, 4.14 and 4.15 are some samples of the results from the tested algorithms. The output images of the proposed approach shows no discernible blocking and ringing artifacts even at low QF settings. The proposed approach preserves detail and edge structure visibly better than most of other techniques.

Like all the tested techniques except BM3D whose main functions are implemented in C++ and compiled to native code, the reference implementation of the proposed technique is written in pure MATLAB language, rendering it unfavourable in comparison of time cost with BM3D. Besides BM3D, the only other method faster than the proposed technique in the comparison group is ACR, which only reduces blocking artifacts and does not perform as well as most of the compared techniques in terms of either PSNR or SSIM.

As shown earlier in this chapter in Figure 4.1, image enhancement routines often exaggerate small compression error in the JPEG compressed input image, deteriorating the perceptive qualities of enhancement results in real-world applications. By using the tone-mapping function introduction in Chapter 2 as the inverse degradation matrix \mathbf{H}^{-1} , we can jointly optimize for reduction of compression artifact and output contrast of tone mapping with the proposed image enhancement framework in order to combat artifacts caused by image enhancement. In addition to the images fetched directly from popular websites as demonstrated in Figure 4.1, we also tested our approach using JPEG compressed satellite and natural images and present the results in Figure 4.16. In these test cases, the JPEG compression QF of input image

Figure 4.13: Comparison of tested methods in visual quality at $QF = 5$.

Figure 4.14: Comparison of tested methods in visual quality at $QF = 15$.

Figure 4.15: Comparison of tested methods in visual quality at $QF = 25$.

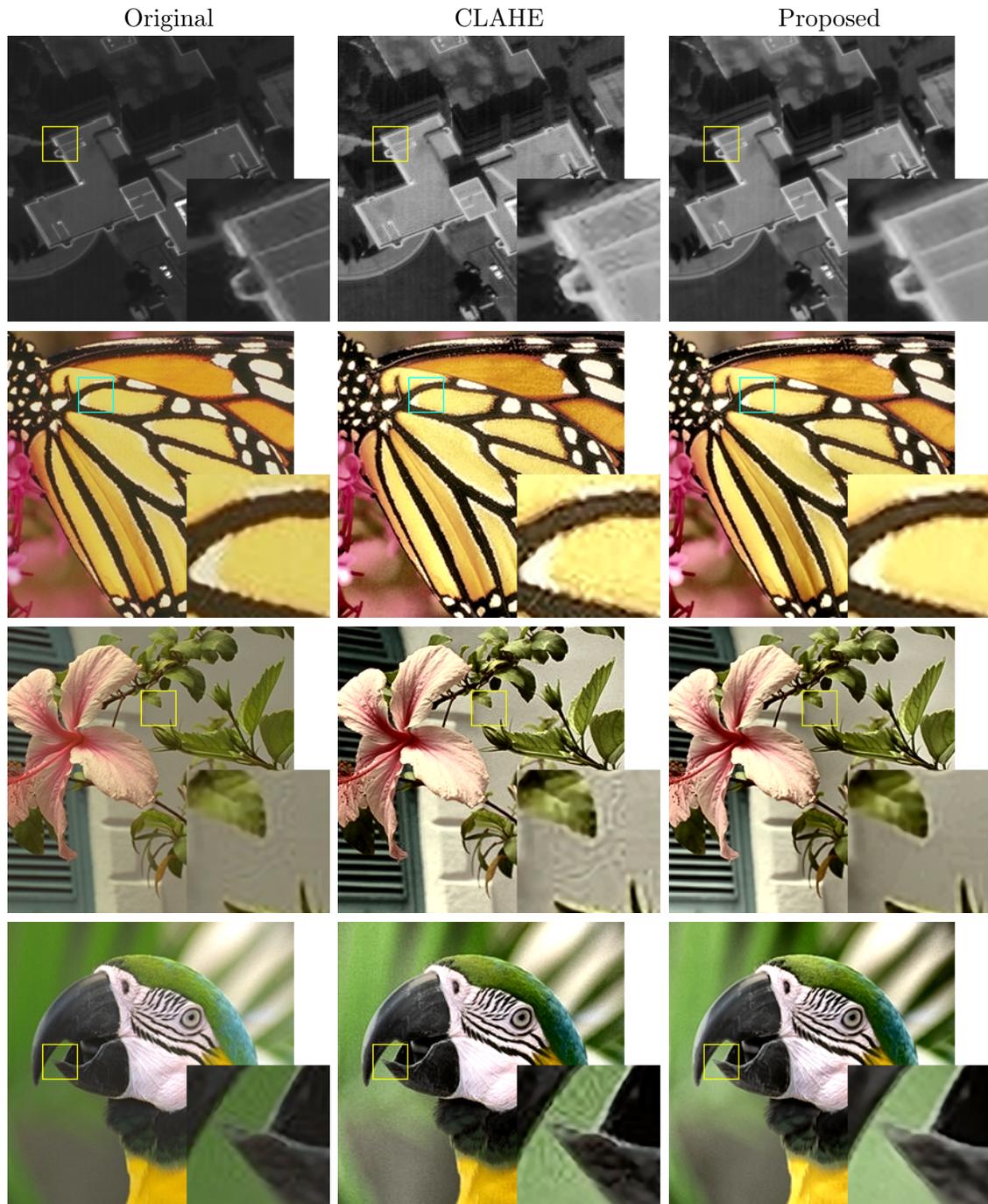


Figure 4.16: Results of image enhancement techniques applied on JPEG compressed images with $QF = 90$.

is set to 90, which is considered as a high-quality compression setting in most applications. As shown in Figure 4.16, the compression artifacts of the original images are indeed hardly perceivable in normal scale. However, conventional image enhancement routines mistake the small compression artifacts as image detail and boost their visibility along with the contrast of true signal. Our approach, on the other hand, successfully separates true signal from compression noise and exhibits little artifacts in the presented examples.

4.6 Conclusion

Due to the low pass nature of image compression, the high-frequency components of a compressed image with sharp edges often carry large compression error. While high-frequency compression noise is relatively indiscernible in the original image as HVS is more sensitive to low-frequency noise, image enhancement operator with high-boosting property can amplify the problem deteriorating the perceptive quality of enhanced image. By incorporating the non-linear DCT quantization mechanism into the formulation for image enhancement, we propose new sparsity-based convex programming approach for joint quantization noise removal and enhancement. Experimental results demonstrate significant performance gains of the new approach over existing enhancement methods.

Chapter 5

Conclusion and Future Work

Human viewers perceive images with sharp edges and rich details as of high visual quality. Because of various types of degradation during the capture of an image and the limitations of display devices and HVS, some information that exists in the acquired image can be difficult to distinguish when the image is displayed directly, resulting poor quality of the image to human viewers. Aiming to fully utilize the dynamic range of the image data and reproduce a visually more appealing and informative image, many image enhancement techniques have been proposed and used since very early days of image processing.

Existing image enhancement techniques can be broadly categorized into two classes: global and local. The global image enhancement techniques, such as HE, manipulate the image histogram using a monotonic tone mapping function that is determined by the histogram of the input image. Based on the OCTM framework, the technique proposed in Chapter 2 is one of such global techniques. Different than existing techniques in the same class, our proposed image enhancement technique utilizes advanced image statistics and finds the best compromise among the factors that affect

image quality, and produces visually pleasing results consistently over a wide range of images. Despite the complexity introduced by the advanced image statistics, the problem can still be solved efficiently.

LARCOTM, the technique introduced in Chapter 3, belongs to the second class, local image enhancement techniques. A local technique enhances the detail of an input image by altering the rate of change in intensity between neighbouring pixels on a pixel-by-pixel basis. Although the local techniques have a larger degree of freedom in manipulating images hence more effective against images with low contrast, they are prone to objectionable artifacts such as ringing and halo, resulting in severely distorted image features. To alleviate these common problems of local techniques, LARCOTM utilizes the fact that the maximum discrimination power of human vision system can only be achieved in a relatively small locality of an image. LARCOTM is fundamentally different from existing image enhancement techniques in that the former can preserve pixel value order statistics within localities in which human foveal vision retains maximum sensitivity, while the latter cannot. As a result, images enhanced by LARCOTM are free of artifacts that plague other local methods.

Although image enhancement can reveal subtle details and enhance the contrast of the original image, it also inevitably magnifies noise due to its high-boosting nature. In Chapter 4, we investigate the problem of enhancing image distorted by compression noise. To suppress noises, some techniques add a noise reduction phase; some common denoising algorithms, such as median filter or Wiener filter, are used before or after the enhancement phase. However, since DCT quantization noise is not independent from signal, these techniques are ineffective against compression artifacts. By incorporating the non-linear DCT quantization mechanism into the formulation for

image enhancement, we propose new sparsity-based convex programming approach for joint quantization noise removal and enhancement. Experimental results demonstrate significant performance gains of the new approach over existing enhancement methods.

For future studies, we find that the current implementation of LARCOTM requires to solve a large scale linear programming problem, which is difficult to do in real time on a computer commonly used in hospitals and clinics. A possible strategy of improving computational efficiency is to find the optimal solution to a downscaled version of the original image first, then upscale the solution and use it as the starting point for a larger version, so on and so forth, until the solution to the original image is found. Another possible improvement worth investigating is to find approximate solution using greedy algorithm. A simple example of such approach is to find the local optimal solutions to all small blocks and then fuse them into an approximate solution for the whole image.

Bibliography

- Adelson, E. H. (2000). Lightness perception and lightness illusions. In *The New Cognitive Neurosciences*, pages 339–351. MIT Press, 2nd edition.
- Agaian, S. S., Silver, B., Panetta, K., *et al.* (2007). Transform coefficient histogram-based image enhancement algorithms using contrast entropy. *Image Processing, IEEE Transactions on*, **16**(3), 741–758.
- Arici, T., Dikbas, S., and Altunbasak, Y. (2009). A histogram modification framework and its application for image contrast enhancement. *Image processing, IEEE Transactions on*, **18**(9), 1921–1935.
- Aubry, M., Paris, S., Hasinoff, S. W., Kautz, J., and Durand, F. (2014). Fast local laplacian filters: Theory and applications. *ACM Transactions on Graphics (TOG)*, **33**(5), 167.
- Barnes, J. (1992). Characteristics and control of contrast in CT. *Radiographics*, **12**(4), 825–837.
- Brandão, T. and Queluz, M. P. (2008). No-reference image quality assessment based on DCT domain statistics. *Signal Processing*, **88**(4), 822–833.

- Bredies, K. and Holler, M. (2012). A total variation-based JPEG decompression model. *SIAM Journal on Imaging Sciences*, **5**(1), 366–393.
- Čadík, M., Wimmer, M., Neumann, L., and Artusi, A. (2008). Evaluation of HDR tone mapping methods using essential perceptual attributes. *Computers & Graphics*, **32**(3), 330–349.
- Cai, J.-F., Candès, E. J., and Shen, Z. (2010). A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, **20**(4), 1956–1982.
- Celik, T. and Tjahjadi, T. (2011). Contextual and variational contrast enhancement. *Image Processing, IEEE Transactions on*, **20**(12), 3431–3441.
- Chang, H., Ng, M. K., and Zeng, T. (2014). Reducing artifacts in JPEG decompression via a learned dictionary. *Signal Processing, IEEE Transactions on*, **62**(3), 718–728.
- Chen, Z., Abidi, B. R., Page, D. L., Abidi, M., *et al.* (2006a). Gray-level grouping (GLG): an automatic method for optimized image contrast enhancement-part I: the basic method. *Image Processing, IEEE Transactions on*, **15**(8), 2290–2302.
- Chen, Z., Abidi, B. R., Page, D. L., Abidi, M., *et al.* (2006b). Gray-level grouping (GLG): an automatic method for optimized image contrast enhancement-part II: the variations. *Image Processing, IEEE Transactions on*, **15**(8), 2303–2314.
- Cohen-Duwek, H., Spitzer, H., Weitzen, R., and Apter, S. (2011). A biologically-based algorithm for companding computerized tomography (CT) images. *Computers in biology and medicine*, **41**(6), 367–379.

- Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2006). Image denoising with block-matching and 3D filtering. In *Electronic Imaging 2006*, pages 606414–606414. International Society for Optics and Photonics.
- DDE (2014). Digital detail enhancement (DDE) – technical note. http://www.flir.com/uploadedfiles/Eurasia/MMC/Tech_Notes/TN_0003_EN.pdf. [Online; accessed 13-Nov-2014].
- Dong, W., Li, X., Zhang, L., and Shi, G. (2011). Sparsity-based image denoising via dictionary learning and structural clustering. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 457–464. IEEE.
- Dong, W., Shi, G., and Li, X. (2013). Nonlocal image restoration with bilateral variance estimation: a low-rank approach. *Image Processing, IEEE Transactions on*, **22**(2), 700–711.
- Farbman, Z., Fattal, R., Lischinski, D., and Szeliski, R. (2008). Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM Transactions on Graphics (TOG)*, volume 27, page 67. ACM.
- Fattal, R. (2009). Edge-avoiding wavelets and their applications. *ACM Transactions on Graphics (TOG)*, **28**(3), 22.
- Fattal, R., Lischinski, D., and Werman, M. (2002). Gradient domain high dynamic range compression. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 249–256. ACM.
- Gauch, J. M. (1992). Investigations of image contrast space defined by variations on

- histogram equalization. *CVGIP: Graphical Models and Image Processing*, **54**(4), 269–280.
- Gu, S., Zhang, L., Zuo, W., and Feng, X. (2014). Weighted nuclear norm minimization with application to image denoising. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2862–2869. IEEE.
- Ichigaya, A., Kurozumi, M., Hara, N., Nishida, Y., and Nakasu, E. (2006). A method of estimating coding PSNR using quantized DCT coefficients. *Circuits and Systems for Video Technology, IEEE Transactions on*, **16**(2), 251–259.
- John, A., Huda, W., Scalzetti, E. M., Ogden, K. M., and Roskopf, M. L. (2004). Performance of a single lookup table (LUT) for displaying chest CT images. *Academic radiology*, **11**(6), 609–616.
- Kim, Y.-T. (1997). Contrast enhancement using brightness preserving bi-histogram equalization. *Consumer Electronics, IEEE Transactions on*, **43**(1), 1–8.
- Kimpe, T. and Tuytschaever, T. (2007). Increasing the number of gray shades in medical display systems how much is enough? *Journal of digital imaging*, **20**(4), 422–432.
- Land, E. H. and McCann, J. (1971). Lightness and retinex theory. *JOSA*, **61**(1), 1–11.
- Larson, G. W., Rushmeier, H., and Piatko, C. (1997). A visibility matching tone reproduction operator for high dynamic range scenes. *Visualization and Computer Graphics, IEEE Transactions on*, **3**(4), 291–306.

- Li, Y.-F., Zhang, Y.-J., and Huang, Z.-H. (2014). A reweighted nuclear norm minimization algorithm for low rank matrix recovery. *Journal of Computational and Applied Mathematics*, **263**, 338–350.
- Liew, A. W. and Yan, H. (2004). Blocking artifacts suppression in block-coded images using overcomplete wavelet representation. *Circuits and Systems for Video Technology, IEEE Transactions on*, **14**(4), 450–461.
- Liu, X., Wu, X., Zhou, J., and Zhao, D. (2015). Data-driven sparsity-based restoration of JPEG-compressed images in dual transform-pixel domain. In *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pages 5171–5178.
- Mairal, J., Bach, F., Ponce, J., Sapiro, G., and Zisserman, A. (2009). Non-local sparse models for image restoration. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2272–2279. IEEE.
- Mantiuk, R., Myszkowski, K., and Seidel, H.-P. (2006). A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception (TAP)*, **3**(3), 286–308.
- Mantiuk, R., Daly, S., and Kerofsky, L. (2008). Display adaptive tone mapping. In *ACM Transactions on Graphics (TOG)*, volume 27, page 68. ACM.
- Oppenheim, A. V., Schaffer, R. W., and Stockham Jr, T. G. (1968). Nonlinear filtering of multiplied and convolved signals. *Audio and Electroacoustics, IEEE Transactions on*, **16**(3), 437–466.

- Osher, S., Burger, M., Goldfarb, D., Xu, J., and Yin, W. (2005). An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, **4**(2), 460–489.
- Panetta, K., Wharton, E. J., Agaian, S. S., *et al.* (2008). Human visual system-based image enhancement and logarithmic contrast measure. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, **38**(1), 174–188.
- Paris, S., Hasinoff, S. W., and Kautz, J. (2011). Local laplacian filters: edge-aware image processing with a laplacian pyramid. *ACM Transactions on Graphics (TOG)*, **30**(4), 68.
- Park, S. H. and Kim, D. S. (1999). Theory of projection onto the narrow quantization constraint set and its application. *Image Processing, IEEE Transactions on*, **8**(10), 1361–1373.
- Pisano, E. D., Zong, S., Hemminger, B. M., DeLuca, M., Johnston, R. E., Muller, K., Braeuning, M. P., and Pizer, S. M. (1998). Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital imaging*, **11**(4), 193–200.
- Rahman, Z.-u., Jobson, D. J., and Woodell, G. A. (2004). Retinex processing for automatic image enhancement. *Journal of Electronic Imaging*, **13**(1), 100–110.
- Reinhard, E., Stark, M., Shirley, P., and Ferwerda, J. (2002). Photographic tone reproduction for digital images. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 267–276. ACM.

- Reinhard, E., Heidrich, W., Debevec, P., Pattanaik, S., Ward, G., and Myszkowski, K. (2010). *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann.
- Shu, X. and Wu, X. (2013). Image enhancement revisited: From first order to second order statistics. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 886–890. IEEE.
- Stark, J. A. (2000). Adaptive image contrast enhancement using generalizations of histogram equalization. *Image Processing, IEEE Transactions on*, **9**(5), 889–896.
- Sun, D. and Cham, W.-K. (2007). Postprocessing of low bit-rate block DCT coded images based on a fields of experts prior. *Image Processing, IEEE Transactions on*, **16**(11), 2743–2751.
- Timofte, R., De Smet, V., and Van Gool, L. (2014). A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Computer Vision—ACCV 2014*, pages 111–126. Springer.
- Turaga, D. S., Chen, Y., and Caviedes, J. (2004). No reference PSNR estimation for compressed pictures. *Signal Processing: Image Communication*, **19**(2), 173–184.
- Wang, Y., Chen, Q., and Zhang, B. (1999). Image enhancement based on equal area dualistic sub-image histogram equalization method. *Consumer Electronics, IEEE Transactions on*, **45**(1), 68–75.
- Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, **13**(4), 600–612.

- Wu, X. (2011). A linear programming approach for optimal contrast-tone mapping. *Image Processing, IEEE Transactions on*, **20**(5), 1262–1272.
- Wu, X. and Memon, N. (1997). Context-based, adaptive, lossless image coding. *Communications, IEEE Transactions on*, **45**(4), 437–444.
- Xie, Q., Meng, D., Gu, S., Zhang, L., Zuo, W., Feng, X., and Xu, Z. (2014). On the optimal solution of weighted nuclear norm minimization. *arXiv preprint arXiv:1405.6012*.
- Zhai, G., Zhang, W., Yang, X., Lin, W., and Xu, Y. (2008a). Efficient deblocking with coefficient regularization, shape-adaptive filtering, and quantization constraint. *Multimedia, IEEE Transactions on*, **10**(5), 735–745.
- Zhai, G., Zhang, W., Yang, X., Lin, W., and Xu, Y. (2008b). Efficient image deblocking based on postfiltering in shifted windows. *Circuits and Systems for Video Technology, IEEE Transactions on*, **18**(1), 122–126.