The Hong Kong Polytechnic University

Department of Computing

External and Internal Nonlocal Self-Similarity based Models for Image Denoising

Jun Xu

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

August 6, 2017

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

	(Signed)
Jun Xu	(Name of student)

Abstract

The nonlocal self-similarity (NSS) prior of natural images has been extensively studied in many image restoration methods. In this thesis, we exploit the NSS property of external natural images, external guided internal NSS property, and internal NSS property for image denoising tasks.

Acknowledgement

Contents

1	Intr	oduction	1
	1.1	The Camera Imaging Pipeline	1
	1.2	The Image Noise	2
	1.3	The Proposed Methods	3
	1.4	Thesis Structure	4
2	Lite	rature Review	7
	2.1	Synthetic Grayscale Image Denoising	7
	2.2	Realistic Color Image Denoising	7
3	Exte	ernal Non-local Self-Similarity Prior for Additive White Gaus-	
	sian	Noise	9
	3.1	Introduction	9
	3.2	System Design	10
	3.3	Demo System	11
	3.4	Calibration	11
	3.5	Conclusion	11
4	Exte	ernal Prior Guided Internal Prior Learning for Real Noisy Im-	
	age	Denoising	13
	4.1	Learning External Nonlocal Self-Similarity Priors	13
	4.2	System Design	14
	4.3	Demo System	15
	4.4	Calibration	15
	4.5	Conclusion	15
5	Inte	rnal Nonlocal Self-Similarity Prior for Real Color Image De-	
	nois	ing: A Low Rank based Method	17
	5.1	Introduction	17
	5.2	Related Work	18
	5.3	Method	18
	5.4	Experimental Results	18

	5.5	Summary	19
6	Inte	rnal Nonlocal Self-Similarity Prior for Real Color Image De-	
	nois	ing: A Sparse Coding based ethod	21
	6.1	Introduction	21
	6.2	Related Work	22
	6.3	Summary	22
7	A L	arge Real Noisy Image Dataset, with A Comprehensive Eval-	
	uati	on of State-of-the-arts	2 3
	7.1	Introduction	23
	7.2	Related Work	24
	7.3	Summary	24
8	Con	clusions	25
	8.1	Section 1	25
	8.2	Section 2	25
	8.3	Future Work	26

List of Figures

3.1	Figure example: (a) example part one, (c) example part two; (c)	
	example part three	9
3.2	Another Figure example: (a) example part one, (c) example part	
	two; (c) example part three	10
4.1	Figure example: (a) example part one, (c) example part two; (c)	
	example part three	13
4.2	Another Figure example: (a) example part one, (c) example part	
	two; <i>(c)</i> example part three	14

List of Tables

Introduction

Nowadays, CCD or CMOS cameras are becoming more and more important in many aspects of human life such as photography, artificial intelligence, and security system. For each camera product, the camera imaging pipeline in the camera is of particular importance since it is the core part to transform the photons reflected by the real scene being captured in the camera sensor into the pixel values of an image, which can be displayed on a screen. During the camera imaging process, the noise is unavoidablely generated due to many reasons. Two major reasons of noise generation are the discrete nature of light and the thermal agitation, which can cause the photon shot noise and the dark-current noise, respectively. Therefore, image denoising is the problem of recovering the laten clean image from the captured noise version.

1.1 The Camera Imaging Pipeline

The cameras capture the images and store as raw image formats. During the camera imaging pipeline, the photons are transformed into electronics by the photodiode in the camera sensor. The original sensor arrat (also called color filter array, or CFA) contains red, green, and blue channels, and these incomplete channels are transformed into the final RGB files via the raw converter. The camera imaging pipeline includes multiple stages such as reading raw image, black light subtraction, lens correction, demosaicing, noise reduction, white balancing, gamma curve, final color space conversion, etc [browneccv2016]. Basically, a camera imaging pipeline includes demosaicing, white balancing and color space transform, gamut mapping, tone mapping, and JPEG compression [crosschannel]. However, different cameras have varying structures and camera parameters, and hence resulting different imaging effects. Recently, there also exists learning based imaging pipelines which directly learn the natural image priors from the RGB and raw images pairs.

1.2 The Image Noise

In image denoising community, the most commonly studied noise is the additive white Gaussian noise, which is used to model the independent noise in the raw images. The AWGN noise is described as a Gaussian distribution $\mathcal{N}(0,\sigma^2)$, which means that the noise is Gaussian distributed with 0 mean and σ standard deviation. Most of methods are focus on this type of noise since it is a good testing bed for many other image restoration problems such as super-resolution, deblurring, inpainting, etc.

However, the realistic noise in real-world natural images captured by cameras are much more complex than the synthetic AWGN noise being widely studied. The major reason is that, during the imaging pipeline, the noise will be generated. The key reason of noise generation is unstable measurement from the discrete nature of light and the thermal agitation. The major sources of noise generated during the imaging pipeline are the random noise, the spatial non-uniformity, and quantization noise. The random noise includes photon shot noise, dark current, and readout noise. The spatial non-uniformity noise includes the fixed pattern noise (PRNU, DCNU), CCD/CMOS specific noise.

A simplified model including various noise sources (for each pixel) can be approximately defined as follows:

$$P = f((q_{cv}(C+D) + N_{reset})q_{out} + N_{out}) + Q.$$

$$(1.1)$$

Now the above equation is explained in details. P is the raw pixel value, C is the number of absorbed electrons (charges) transformed from the photons via the photon-diodes in the camera sensor, D is the number of absorbed electrons generated by dark current, g_{cv} is the equivalent capacitance (EC) of the photo-diode, N_{reset} is the thermal noise generated by the readout circuitry (or reset noise related to reset voltage), g_{out} is the gain factor during voltage to pixel value conversion (readout), N_{out} is the readout noise, f is the camera response function, usually a linear function before attaining a saturation threshold, Q is the quantization error happened during rounding to interger values. The quantization noise is normally negligible compared to the readout noise.

Though can be approximated as Gaussian or Poisson distribution, these noise sources will be largely changed to be more complex during the in-camera imaging pipeline, which has been analyzed in [crosschannel]. Hence, the real-world noise is much more complex than the traditional additive white Gaussian noise, and should be paid more attention.

In order to obtain good images while still maintaining the structures and details of the captured scene, image denoising is a good choice. In this thesis, I will present my work on image denoising, including synthetic noisy images and real noisy images. Note that the realistic noise in real color noisy images are much more complex than the usually synthetic additive white Gaussian noise (AWGN). This is because that during the camera imaging pipeline, the noise generated will be modified to be more complex than the noise in the raw sensor data. Hence, in this thesis, we pay more attention to the realistic noise generated during the camera imaging pipeline.

1.3 The Proposed Methods

This thesis is made up of several work I have done during my study, in which I focus on designing new and better image denoising algorithms.

Firstly, we propose a method for denoising synthetic AWGN noise, from which we can study the performance of the nonlocal self-similarity priors of natural images. In fact, we propose to learn the external NSS priors and apply the learned model on denoising AWGN noise. The proposed method achieves state-of-the-art performance on AWGN denoising on both effectiveness and efficiency.

Basing on the success on the synthetic noise removal, I propose to exploit the power of the NSS priors in natural images to deal with the complex realistic noise in real-world noisy images. Specifically, we propose three methods exploiting the NSS priors of natural images for real noisy image denoising, which can be introduced as follows.

In the first method, I propose to learn the NSS prior from the external natural images, and then apply the learned external prior to guide the learning of the internal NSS prior of the input real noisy image. The experiments on two commonly used datasets and a new one we constructed to implement the shortage of existing datasets, demonstrate that the proposed method can achieve better performance than existing color image denoising methods such as CBM3D [cbm3d], the state-of-the-art Gaussian noise removal methods

[bm3d; mlp; csr], and the real noisy image denoising methdos [] including a commercial software Neat Image [neatimage], which is embedde into the famous PhotoShop CS for image processing tasks.

In the second method, I propose to employ the low rank model describe fully the the internal NSS prior, basing on the observed fact that the similar image patches can be contanated as a matrix of low rank. Different from the previous work, I extend the WNNM model and apply it to multi-channel version to make it feasible for color image denoising.

In the third method, is to use the sparse coding based method with additional weighting scheme to regard the local noise in real noisy images as a Gaussian and the prior is used to deal with the real noisy image.

Finally, to make my thesis more comprehensive, I construct a large benchmark of real noisy images captured by different types of famous commercial cameras, on which I also evaluate the image denoising methods mentioned above and the proposed methods in this thesis.

The structure of this thesis is organized as follows: in the chapter 2, we review the literatures in the image denoising area; in the chapter 3, we introduce the fully external method; in the chapter 4, we introduce the external prior guided internal method; in the chapter 5, we introduce the internal method based on low ran model; in the chapter 6, we introduce the internal method based on sparse coding model; in the chapter 7, we introduce the real noisy image dataset we construct, and finally evaluate the proposed methods with the compared competing methods, both for synthetic AWGN or Poisson noise and real noise, including the commercial software designed especially for real noise.

1 4 Thesis Structure

Chapter 2: Literature Review

In this chapter, we review the related work and give a detailed introduction of the literature. We will first review the most representative work on additive white Gaussian noise removal. I review the detailed work on camera imaging pipeline and realistic noise generated in the camera sensors. I will also review the work on real noisy image denoising.

Chapter 3: External Nonlocal Self-Similarity Prior Learning for Synthetic Gaussian Noise Removal

In this chapter, I will introduce our work on external nonlocal self-similarity (NSS) prior learning for synthetic Gaussian noise removal. As far as we know, this work is the first to learn the NSS priors of natural clean images, while previous work only utilize the NSS priors of input noisy image for online denoising. The advantages of this offline learning is that it can preserve the details of natural images while being much faster then most online denoising methods.

Chapter 4: External Prior Guided Internal Prior Learning for Real Noisy Image Denoising

In this chapter, I will introduce our work on external prior guided internal prior learning method for real noisy image denoising. This work can maintain the advantages of both sides: from the external perspective, the method can preserve the structures of natural images better than the internal methods, while from the perspective of internal method, the proposed method can recover the details of the input noisy image better than the external methods.

Chapter 5: Multi-channel Weighted Nuclear Norm Minimization for Real Color Image Denoising

In this chapter, we introduce a multi-channel weighted nuclear norm minimization (MC-WNNM) method. This method regards different channels in RGB images differently to adaptively process the real color noisy images. Besides, this work also propose a new strategy for color image denoising. Experiments demonstrate that the proposed method can achieve better performance on real color image denoising than existing state-of-the-art methods, including some commercial software.

Chapter 6: A Triple Weighted Sparse Coding Scheme for Realistic Noisy Image Denoising

In this chapter, I introduce a novel sparse coding based method for real color image denoising. In this method, I regard the noise in each of the local region in the real noisy image as a Gaussian, and propose a triplely weighted scheme to deal with the complex realistic noise in real color noisy images. Experiments show that the proposed method performs better and faster than the nuclear norm based method mentioned in previous chapter.

Chapter 7: A Benchmark on Real Color Noisy Image, with Comprehensive Evaluation of State-of-the-art

To fully boost the research of real color noisy image denoising, we construct a large benchmark on real color noisy images. This dataset is collected from several representative cameras with comprehensive settings on contents, lighting, ISO, shutter, and aperture, etc. Based on this newly established dataset, we fully evaluated existing denoising methods, including the methods designed for synthetic Gaussian noise and the methods designed especially for real color noise. We believe that this new dataset will largely boost the research of the image denoising especially the realistic image denoising problems.

Literature Review

"Mens cujusque is est Quisque" – "Mind Makes the Man"

— Samuel Pepys

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

2.1 Synthetic Grayscale Image Denoising

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

2.2 Realistic Color Image Denoising

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A

blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

External Non-local Self-Similar ity Prior for Additive White Gaussian Noise

Innovation distinguishes between a leader and a follower.

— Steve Jobs
(CEO Apple Inc.)

3.1 Introduction



Fig. 3.1: Figure example: (*a*) example part one, (*c*) example part two; (*c*) example part three

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A

blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.



Fig. 3.2: Another Figure example: (*a*) example part one, (*c*) example part two; (*c*) example part three

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

3.2 System Design

3.3 Demo System

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

3.4 Calibration

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

3.5 Conclusion

External Prior Guided Internal Prior Learning for Real Noisy Image Denoising

Innovation distinguishes between a leader and a follower.

— Steve Jobs
(CEO Apple Inc.)

4.1 Learning External Nonlocal Self-Similarity Priors



Fig. 4.1: Figure example: (*a*) example part one, (*c*) example part two; (*c*) example part three

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between

this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.



Fig. 4.2: Another Figure example: (*a*) example part one, (*c*) example part two; (*c*) example part three

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

4.2 System Design

4.3 Demo System

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

4.4 Calibration

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

4.5 Conclusion

Internal Nonlocal Self-Similarity Prior for Real Color Image Denoising: A Low Rank based Method

Users do not care about what is inside the box, as long as the box does what they need done.

— Jef Raskin about Human Computer Interfaces

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

5.1 Introduction

There is no need for special content, but the length of words should match the language.

5.2 Related Work

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

5.3 Method

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

5.4 Experimental Results

There is no need for special content, but the length of words should match the language.

5.5 Summary

Internal Nonlocal Self-Similarity Prior for Real Color Image Denoising: A Sparse Coding based ethod

box, as long as the box does what they need done.

— **Jef Raskin** about Human Computer Interfaces

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

6.1 Introduction

There is no need for special content, but the length of words should match the language.

6.2 Related Work

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

6.3 Summary

A Large Real Noisy Image Dataset, with A Comprehensive Evaluation of State-of-the-arts

Users do not care about what is inside the box, as long as the box does what they need done.

— **Jef Raskin** about Human Computer Interfaces

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

7.1 Introduction

7.2 Related Work

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

7.3 Summary

Conclusions

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

8.1 Section 1

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

8.2 Section 2

There is no need for special content, but the length of words should match the language.

8.3 Future Work