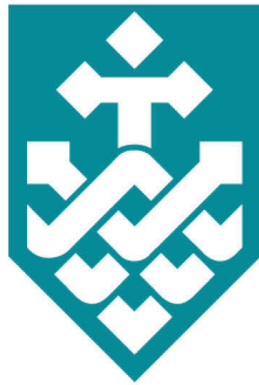


# Facial Image Restoration and Retrieval Through Orthogonality



Dayong Tian

Faculty of Engineering and Information Technology

University of Technology Sydney

A thesis submitted for the degree of

*Doctor of Philosophy*

2017



To my loving parents  
*Jubao Tian* and *Shuxia Song*  
and my wife  
*Yiwen Wei*



## **Certificate of Original Authorship**

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Dayong Tian



## Acknowledgements

I would like to express my special appreciation and thanks to my supervisor Professor Dacheng Tao. He accepted my application three years ago and gave me the chance to change my life. In last three years, he consistently taught me how to research. From reading literatures to writing my own papers, from implementing others' methods to devising my own methods, I learned a lot under his instructions.

I also wish to give special thanks to Chaoyue Wang, Guoliang Kang and Jiang Bian. Without their helps, my life in Sydney would not be so easy. I am grateful to Baosheng Yu and Zhe Chen. Our work stations are neighboring. Their opinions deeply impressed me. I would like to give my gratitude to Changxing Ding, Kede Ma, Maoying Qiao and Tongliang Liu for their helps on my research.

I would like to give thanks to my friends I met in Conversation@UTS, especially to Sakada Ko. Ko is an exchange student from Kyoto University of Foreign Studies. He shared lots of his experiences on learning English to me.

It is my fortune to met my friends, Archer, Berry, Beryl, Billy, Carrie, Celine, Chain, CK, Danny, David, Emma, Peyton, Rhonda, Vivi, Vivian, William, Wilton and Zoe. I cannot remember how many times Berry and Chain mixed drinks for us, how many times I was drunk with Rhonda and how many times Zoe drove me home. It is them who make my life diverse and give me completely different experiences. I will miss every minute spent with them.

Finally, I would like to express my gratitude to my family, my parents and my wife, for their encouragement and support.

# Abstract

Orthogonality has different definitions in geometry, statistics and calculus. This thesis studies how to incorporate orthogonality to facial image restoration and retrieval tasks. A facial image restoration method and three retrieval methods were proposed.

Blur in facial images significantly impedes the efficiency of recognition approaches. However, most existing blind deconvolution methods cannot generate satisfactory results, due to their dependence on strong edges which are sufficient in natural images but not in facial images. A novel method is proposed in this report. Point spread functions (PSF) are represented by the linear combination of a set of pre-defined orthogonal PSFs and similarly, an estimated intrinsic sharp face image (EI) is represented by the linear combination of a set of pre-defined orthogonal face images. In doing so, PSF and EI estimation is simplified to discovering two sets of linear combination coefficients which are simultaneously found by the proposed coupled learning algorithm. To make the method robust to different kinds of blurry face images, several candidate PSFs and EIs are generated for a test image, and then a non-blind deconvolution method is adopted to generate more EIs by those candidate PSFs. Finally, a blind image quality assessment metric is deployed to automatically select the optimal EI.

On the other hand, the orthogonality is incorporated into the proposed Unimodal image retrieval method. Hashing methods have been widely investigated for fast approximate nearest neighbor searching in large datasets. Most existing methods use binary vectors in lower dimensional spaces to represent data points that are usually real vectors of higher dimensionality. The proposed method divides the hashing



process into two steps. Data points are first embedded in a low-dimensional space, and the Global Positioning System (GPS) method is subsequently introduced but modified for binary embedding. Data-independent and data-dependent methods are devised to distribute the satellites at appropriate locations. The proposed methods are based on finding the tradeoff between the information losses in these two steps. Experiments show that the data-dependent method outperforms other methods in different-sized datasets from 100K to 10M. By incorporating the orthogonality of the code matrix, both data-independent and data-dependent methods are particularly impressive in experiments on longer bits.

In social networks, heterogeneous multimedia data correlates to each other, such as videos and their corresponding tags in YouTube and image-text pairs in Facebook. Nearest neighbor retrieval across multiple modalities on large data sets becomes a hot yet challenging problem. Hashing is expected to be an efficient solution, since it represents data as binary codes. As the bit-wise XOR operations can be fast handled, the retrieval time is greatly reduced. Few existing multi-modal hashing methods consider the correlation among hashing bits. The correlation has negative impact on hashing codes. When the hashing code length becomes longer, the retrieval performance improvement becomes slower. The proposed method incorporates a so-called minimum correlation constraint which can be treated as a generalization of orthogonality constraint. Experiments show the superiority of the proposed method becomes greater as the code length increases.

Deep neural network is expected to be an efficient way for multi-modal hashing. We propose a hybrid neural network which consists of a convolutional neural network for facial images and a full-connected neural network for tags or labels. The minimum correlation regularization is imposed on the parameters of output layers. Experiments validates the superiority of the proposed hybrid neural network.

# Contents

<b>Contents</b>	<b>ix</b>
<b>List of Figures</b>	<b>xii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.1.1 Orthogonality in Geometry . . . . .	1
1.1.2 Orthogonality in Statistics . . . . .	2
1.1.3 Orthogonality in Calculus . . . . .	3
1.1.4 Facial Image Restoration . . . . .	4
1.1.5 Image Retrieval . . . . .	4
1.1.6 Multi-modal Retrieval . . . . .	5
1.2 Summary of Contributions . . . . .	5
<b>2 Coupled Learning for Facial Deblur</b>	<b>7</b>
2.1 Introduction . . . . .	8
2.2 On Combining Coupled Learning and BIQA for Facial Deblur . .	12
2.2.1 Construct $\mathbf{A}$ . . . . .	12
2.2.2 Calculate $\mathbf{x}$ . . . . .	13
2.2.3 Calculate $\alpha$ and $\beta$ . . . . .	14
2.2.4 Generate candidate results . . . . .	16
2.2.5 Assessing candidate results . . . . .	16
2.2.6 Simultaneous restoration and recognition . . . . .	19
2.3 Efficient Implementation for Symmetric PSF . . . . .	21
2.3.1 PSF direction estimation . . . . .	22

## CONTENTS

2.4	Experiments . . . . .	24
2.4.1	PSF direction estimation . . . . .	25
2.4.2	Facial deblur . . . . .	26
2.4.3	Simultaneous facial deblur and recognition . . . . .	27
2.4.4	Experiments on camera-shaking blur . . . . .	29
2.4.5	Experiments on real blur . . . . .	31
2.4.6	Computational efficiency . . . . .	32
2.5	Conclusions . . . . .	33
<b>3</b>	<b>Global Hashing System for Fast Image Search</b>	<b>36</b>
3.1	Introduction . . . . .	36
3.2	Related Work . . . . .	38
3.3	Methodology . . . . .	39
3.3.1	Global Positioning/Hashing System . . . . .	39
3.3.2	Data-dependent method (GHS-DD) . . . . .	40
3.3.3	Optimization . . . . .	42
3.3.4	Data-independent method (GHS-DI) . . . . .	44
3.3.5	Parameters $r_s$ and $\rho$ . . . . .	46
3.4	Relations to Existing Methods . . . . .	48
3.4.1	Iterative Quantization (ITQ) . . . . .	49
3.4.2	Inductive Hashing on Manifolds (IMH) . . . . .	49
3.4.3	Spectral Hashing (SH) . . . . .	50
3.4.4	Spherical Hashing (SpH) . . . . .	50
3.5	Experiments . . . . .	51
3.5.1	Protocols and baselines . . . . .	51
3.5.2	Quantitative evaluation . . . . .	53
3.5.3	Computational efficiency . . . . .	54
3.5.4	Incorporating label information . . . . .	57
3.5.5	Classification with hashing codes . . . . .	58
3.5.6	Performance on facial images . . . . .	59
3.6	Conclusion . . . . .	61

## CONTENTS

<b>4</b>	<b>Learning decorrelated hashing codes for multi-modal Retrieval</b>	<b>62</b>
4.1	Introduction . . . . .	62
4.2	Methodology . . . . .	64
4.2.1	Problem Formulation . . . . .	65
4.2.2	Optimization . . . . .	67
4.2.3	Implementation details . . . . .	69
4.3	Experimental Results . . . . .	70
4.3.1	Data sets . . . . .	71
4.3.2	Evaluation Metrics . . . . .	72
4.3.3	Baselines . . . . .	72
4.3.4	Results . . . . .	74
4.3.5	Convergence Study . . . . .	76
4.3.6	Computation Efficiency . . . . .	76
4.3.7	Performance on Facial Images . . . . .	77
4.4	Conclusion . . . . .	77
<b>5</b>	<b>A Hybrid Neural Network for Multimodal Hashing</b>	<b>79</b>
5.1	Introduction . . . . .	79
5.2	Methodology . . . . .	80
5.3	Experimental Results . . . . .	81
<b>6</b>	<b>Algorithmic stability and sharp generalization error bounds</b>	<b>83</b>
6.1	Introduction . . . . .	83
<b>7</b>	<b>Conclusions</b>	<b>85</b>
<b>A</b>	<b>Results of Image Hashing Method</b>	<b>87</b>
<b>B</b>	<b>Results of Multimodal Retrieval Method</b>	<b>98</b>
<b>C</b>	<b>Results of Hybrid Deep Neural Network</b>	<b>109</b>
	<b>References</b>	<b>120</b>

# List of Figures

1.1	Illustration of three coordinate systems. (a) Cartesian coordinate system. (b) Polar coordinate system. . . . .	2
1.2	Illustration of principal component analysis on a 2-dimensional data set . . . . .	3
2.1	Illustration of the drawbacks of the sparse prior. The image is blurred by a Gaussian kernel of standard deviation 0-8. The $l_1$ -norm of sparse coefficients monotonously decreases as the standard deviation increases. Hence, the sparse prior may lead to a blurry result. . . . .	9
2.2	The framework of our method. . . . .	10
2.3	The difference between projections of an image and its blurred counterpart. 90% images in FERET dataset are used to construct $\mathbf{D}$ and the first 10 projections of a test image in the remaining 10% images and its blurred counterpart which is blurred by a Gaussian PSF with standard deviation 2 are shown here. It can be found that projections on the first left singular vector of these two images are very close to each other. The difference is approximately 0.1%. . . . .	15
2.4	Typical results and their corresponding scores. -2 denotes a failure, 0 is a blurry result, and 2 is a sharp image. . . . .	17

## LIST OF FIGURES

2.5	The symmetry of PSFs. The top row illustrates three PSFs: a Gaussian kernel, a linear motion kernel, and the combination of both. The bottom row illustrates five function bases, i.e., $m = n = 1$ , $m = 1 \& n = 2$ , $m = 1 \& n = 3$ , $m = 2 \& n = 3$ , and $m = n = 3$ . The function bases of odd $m$ and $n$ are symmetric about the two axes and can be used to represent the three symmetric PSFs. . . .	22
2.6	Illustration of proposed PSF direction estimation algorithm. A combination of linear motion and Gaussian PSF locates at the center of the graph. $\phi$ 's with different $\sigma$ locate around the PSF according to their direction. The numbers are the maximum values of $\int_{\Omega}  \mathcal{F}(\phi')   \mathcal{F}(\phi)  d\Omega$ on each direction. . . . .	23
2.7	Results of experiments on the subset of FERET dataset. (a) Test images and their PSFs; (b) results of the proposed method; (c) results using the method proposed by Krishnan <i>et al.</i> [55]; (d) results of FADEIN [83] and (e) results using the method proposed by Pan <i>et al.</i> [86] . . . . .	26
2.8	Results of experiments on the subset of CMU-PIE dataset. (a) Test images and their PSFs; (b) results of the proposed method; (c) results using the method proposed by Krishnan <i>et al.</i> [55]; (d) results of FADEIN [83] and (e) results using the method proposed by Pan <i>et al.</i> [86] . . . . .	28
2.9	Results of experiments on the subset of extended Yale B dataset. (a) Test images and their PSFs; (b) results of the proposed method; (c) results using the method proposed by Krishnan <i>et al.</i> [55]; (d) results of FADEIN [83] and (e) results using the method proposed by Pan <i>et al.</i> [86] . . . . .	29
2.10	Results of experiments on the subset of CMU-PIE dataset. (a) Test images and their PSFs; (b) results of the proposed method; (c) results using the method proposed by Krishnan <i>et al.</i> [55] and (d) results using the method proposed by Pan <i>et al.</i> [86] . . . . .	32

## LIST OF FIGURES

2.11	Results of experiments on the subset of extended Yale B dataset. (a) Test images and their PSFs; (b) results of the proposed method; (c) results using the method proposed by Krishnan <i>et al.</i> [55] and (d) results using the method proposed by Pan <i>et al.</i> [86] . . . . .	33
2.12	The restoration results on FRGC 2.0. From top to bottom, they are OBs, and results of our method, Krishnan <i>et al.</i> [55] and Pan <i>et al.</i> [86], respectively. The results marked by red rectangles are generated by the linear combinations of the function bases.	34
3.1	Illustration of a GPS. A satellite broadcasts its current time to the receiver (red spot). The distance is calculated by multiplying travel velocity of electromagnetic waves with the difference of the receivers' current time and the received satellite time. (a) The distances of a receiver to three satellites can uniquely determine its location on the Earth surface. (b) Such distributed satellites fail to uniquely determine the receiver's location. . . . .	40
3.2	MAP on CIFAR-10 dataset for GHS-DI and GHS-DD. When $r_s$ approximates 0, both methods fail to get satisfactory results. The performance of both methods become stable after $r_s$ is larger than 1. On the other hand, GHS-DI gets its best results when $\rho$ is in the interval $[0.5, 1]$ , while it is $[0.7, 1]$ for GHS-DD. For $c < 16$ , the best results appear when $\rho$ approximates 1 because enough amounts of principal components should be selected. . . . .	47
3.3	Mean F-measure of hash lookup with Hamming radius 2 for different methods on SUN397, GIST1M and SIFT10M. . . . .	51
3.4	Mean F-measure of hash lookup with Hamming radius 2 and MAP for different methods on CIFAR-10. . . . .	55
3.5	The query images and the query results returned by compared methods with 32 hash bits. . . . .	56
3.6	Quantization loss of each iteration on CIFAR-10. (a) CCA-GHS-DD; (b) CCA-GHS-DI. . . . .	58
3.7	Classification accuracy (%) on MNIST. . . . .	59

## LIST OF FIGURES

4.1	Illustration of <b>Proposition 2</b> and <b>Proposition 3</b> . If $\mathbf{W}^i \in \mathbb{R}^{2 \times 3}$ , its column vectors will align with the centerlines of an equilateral triangle. If $\mathbf{W}^i \in \mathbb{R}^{2 \times 4}$ , its column vectors will align with the diagonals of a square. The affine transformation will change the relative positions among vectors but the overall structure is kept. For example, in the equilateral triangle, point B is transformed to the clockwise direction of point A. . . . .	68
4.2	F1-score on MIRFlickr and NUS-WIDE data sets . . . . .	74
4.3	Convergence curves on MIRFlickr and NUS-WIDE data sets . . . . .	75
5.1	The network structure for double-view data . . . . .	80