

## DOCTORAL THESIS

### Writer identification using wavelet, contourlet and statistical models

He, Zhenyu

*Date of Award:*  
2006

[Link to publication](#)

#### General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

# Writer Identification Using Wavelet, Contourlet and Statistical Models

**HE Zhenyu**

A thesis submitted in partial fulfillment of the requirements  
for the degree of  
Doctor of Philosophy

Principal Supervisor: Prof. TANG Yuan Yan

Hong Kong Baptist University

December 2006

# Abstract

Even in such a highly developed time, handwriting still plays a very important role in human society. And handwriting generally is regarded as a sign of the writer. A long history before, people have realized the importance of finding the true writer of one unknown handwriting document. In fact, writer identification based on the handwriting (there are many forms of handwritings, such as signatures, letters, notes, etc.) has a wide applicable field: to confirm the document authenticity in the financial sphere, to solve the expert problems in criminology, etc.

With the coming of computer era, most manual works can be well carried out by computer automatically. As a result, a great deal of time, efforts and money are saved. Naturally, the automatic writer identification comes into scientists' views. Nowadays, the automatic writer identification of handwriting is receiving growing interests from both academia and industry. And more and more researchers and firms put money and energies on it. However, automatic writer identification is a complex problem involved in many science disciplines, including computer vision, pattern recognition, etc., and has many challenging items. Because of that, in spite of continuous efforts, writer identification is still a world challenging issue and far from being well solved.

Generally speaking, writer identification is a typical pattern recognition problem which consists of several steps: pre-processing, feature extraction, similarity measurement, performance evaluation. Among these steps, the feature extraction is the core one.

The following works are done in this research.

- We give a detailed introduction of writer identification, including concept, function and importance in human society, and other relevant knowledge. In addition, by consulting several tens of papers in this field, especially the most recent ones, we discuss the main types of writer identification and systematically recall the development on off-line, text-independent writer identification.

- We describe a whole procedure for off-line, text-independent writer identification, and introduce the function of each step in this procedure. In particular, we design a method for pre-processing of handwriting images. Till now, the pre-processing of the handwriting images is the most challenging task for writer identification. Most pre-processing algorithms proposed aim to the handwriting written in regular form, and nearly all researches in writer identification parry the handwriting written in irregular form. But our method can help us deal with any handwriting manually to generate high-quality pre-processed handwriting images for next processing step. Moreover, we also establish our own handwriting database for writer identification which consists of hundreds of handwriting images stored in gray-scale or binary image format.
- Finding and extracting the radical feature from handwriting image for writing identification is the main research task in this dissertation. We introduce wavelet techniques to the off-line, text-independent writer identification. From our research works, we find that the histogram of the wavelet coefficients of normalized handwriting image can be well characterized by the Generalized Gaussian Model (GGD). As a consequence, we propose a wavelet-based GGD model for writer identification, which achieves a better identification result than those of the existing methods. In addition, our method also greatly reduces the computational cost. To better describe the mutual dependence relations between wavelet coefficients, we propose a wavelet-based Hidden Markov Tree (HMT) model for writer identification. This method also outperforms the existing methods, and saves the computation time as well. Furthermore, to capture more orientation information of handwriting images, we design a method combining the GGD model and the contourlet transform, which is lately developed for efficient image representation and owns a more powerful ability in reflecting directional information than that of the separable wavelet. These three methods aim to serve the gray handwritings. While for the binary handwriting, our experiments tell us that these statistical models are unavailable. Therefore, for the writer identification based on the binary handwriting, we develop a hybrid method combining the Gabor transform and fractal dimension. This

method utilizes the Gabor's advantage in multi-directional decomposition and the fractal dimension's advantage in precise measurement.

- Finally, we discuss the possibility and realization of a hybrid writer identification combining the text-independent and text-dependent features. In practice, multi-features based systems provably improve accuracy, robustness and applicability of the identification system. We propose a simple but whole procedure for the hybrid system. Besides, we also introduce our current attempt in the text-dependent writer identification.

# Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	v
Table of Contents	vi
List of Figures	ix
List of Tables	xv
Chapter 1 Introduction	1
1.1 Why We Need Writer Identification . . . . .	1
1.2 The Importance of Handwriting and the Concept of Writer Identification	4
1.3 The Theoretical Foundation of Writer Identification . . . . .	6
1.4 On-line and Off-line Writer Identification . . . . .	7
1.5 Text-dependent and Text-independent Writer Identification . . . . .	10
1.6 Signature Verification and Writer Identification . . . . .	12
1.7 Writer Identification and Handwriting Recognition . . . . .	14
1.8 Handwriting Experts and Features Used by Them for Writer Identifi- cation . . . . .	15
1.9 Previous Works . . . . .	16
1.10 Original Contributions . . . . .	19
1.11 Outline of the Thesis . . . . .	20

Chapter 2	Overview of the Writer Identification System	22
2.1	The Whole Procedure . . . . .	22
2.2	Data Acquisition . . . . .	24
2.3	Pre-processing . . . . .	25
2.4	Feature Extraction . . . . .	31
2.5	Similarity Measurement . . . . .	32
2.6	Performance Evaluation . . . . .	33
Chapter 3	Wavelets and Wavelet-based GGD Model for Writer Identification	36
3.1	Important Knowledge of Wavelets and Wavelet Transform . . . . .	36
3.1.1	The Fourier Transform and Windowed Fourier Transform . . . . .	37
3.1.2	Continuous Wavelet Transform . . . . .	39
3.1.3	Discrete Wavelet Transform . . . . .	41
3.1.4	Multiresolution Analysis (MRA) of $L^2(\mathbf{R})$ . . . . .	42
3.1.5	Mallat Decomposition and Reconstruction Algorithms of Functions in $L^2(\mathbf{R})$ . . . . .	44
3.2	Statistical Models in Wavelet Domain . . . . .	47
3.2.1	The Energy Signature in Wavelet Domain . . . . .	48
3.2.2	The Marginal and Joint Distribution of Wavelet Coefficients . . . . .	49
3.3	A Novel Wavelet-based GGD Method for Writer Identification . . . . .	54
3.3.1	Our Research Discovery . . . . .	54
3.3.2	The Whole Procedure . . . . .	56
3.3.3	Feature Extraction . . . . .	56
3.3.4	Similarity Measurement . . . . .	58
3.3.5	Performance Evaluation . . . . .	59
3.4	Experiments and Relevant Discussions . . . . .	59
3.4.1	Database Used in Our Experiments . . . . .	59
3.4.2	Identification Performance Evaluation 1 . . . . .	59
3.4.3	Identification Performance Evaluation 2 . . . . .	64
3.5	Summary . . . . .	68

Chapter 4	Writer Identification Using Hidden Markov Model in Wavelet Domain	70
4.1	Hidden Markov Model and Relative Statistical Knowledge	70
4.1.1	Expectation Maximization (EM) Algorithm	70
4.1.2	GMM Model and EM for GMM Model	72
4.1.3	Hidden Markov Model	75
4.2	Hidden Markov Models in Wavelet Domain	79
4.2.1	GMM Model for a Single Wavelet Coefficient	79
4.2.2	Independence Mixture Model	79
4.2.3	WD-HMM and EM for WD-HMM	80
4.3	Writer Identification Using WD-HMM	83
4.3.1	The Whole Procedure	83
4.3.2	Feature Extraction	83
4.3.3	Similarity Measurement	85
4.3.4	Performance Evaluation	87
4.4	Experiments and Relevant Discussions	88
4.4.1	Database Used in Our Experiments	88
4.4.2	Identification Performance Evaluation 1	88
4.4.3	Identification Performance Evaluation 2	90
4.5	Summary	91
Chapter 5	Writer Identification Using Contourlet	95
5.1	Motivation of the New Wavelet-like Frame	95
5.2	Ridgelet	98
5.3	Curvelet	100
5.4	Contourlet-based Method for Writer Identification	102
5.4.1	Multi-dimensional Sampling	102
5.4.2	Contourlet Transform	105
5.4.3	Why We Choose the Contourlet for Writer Identification	111
5.4.4	The Whole Procedure	113
5.5	Experiments and Relevant Discussions	113
5.5.1	Identification Performance Evaluation 1	113



5.5.2	Identification Performance Evaluation 2 . . . . .	117
5.5.3	Experimental Observations . . . . .	118
5.6	Summary . . . . .	119
Chapter 6 Writer Identification Using a Hybrid Method Combining Gabor Wavelet and Fractal Theory		124
6.1	Gabor Function . . . . .	124
6.1.1	Gabor Filter Design . . . . .	125
6.1.2	Gabor Feature Representation . . . . .	126
6.2	Fractal Theory . . . . .	129
6.2.1	Hausdorff Dimensions . . . . .	130
6.2.2	Box Counting Dimensions and Mesh Fractal Dimensions . . . . .	131
6.2.3	Calculation of Mesh Fractal Dimension . . . . .	133
6.3	The Whole Procedure . . . . .	134
6.3.1	Feature Extraction . . . . .	134
6.3.2	Similarity Measurement . . . . .	136
6.3.3	Performance Evaluation . . . . .	136
6.4	Experiments . . . . .	137
6.5	Summary . . . . .	140
Chapter 7 Future Works		144
Curriculum Vitae		160
Publications		161