university of groningen

University Medical Center Groningen

University of Groningen

Statistical pattern recognition for automatic writer identification and verification

Bulacu, Marius Lucian

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2007

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Bulacu, M. L. (2007). Statistical pattern recognition for automatic writer identification and verification s.n.

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 10-02-2018

Statistical Pattern Recognition for Automatic Writer Identification and Verification

Marius Lucian Bulacu

Front cover: A schematic description of the extraction of the ″hinge″ feature.
Back cover: A 20x20 Kohonen self-organized map of handwritten graphemes.

RIJKSUNIVERSITEIT GRONINGEN

Statistical Pattern Recognition for Automatic Writer Identification and Verification

Proefschrift

ter verkrijging van het doctoraat in de
Gedrags- en Maatschappijwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
donderdag 15 maart 2007
om 16.15 uur

door

Marius Lucian Bulacu

geboren op 8 oktober 1973
te Boekarest (Roemenië)

Promotor:	Prof. dr. L. Schomaker
Beoordelingscommissie:	Prof. dr. J. Daugman
	Prof. dr. F. Groen
	Prof. dr. J. Roerdink

ISBN: 90–367–2912–2

To my parents

Contents

1 Introduction **1**
1.1 Writer identification as a behavioral biometric . . . 2
1.2 Writer identification in forensics . . . 2
1.3 Writer identification vs. Handwriting recognition . . . 4
1.4 Writer identification vs. Writer verification . . . 4
1.5 Text-dependent vs. Text-independent methods . . . 5
1.6 Within-writer variance vs. Between-writer variation . . . 6
1.7 Factors causing variability in handwriting . . . 7
1.8 Factors determining individuality of handwriting . . . 9
1.9 A survey of recent research in the field . . . 12
1.10 Main assumptions underlying the methods proposed in the thesis . . . 15
1.11 Overview of the thesis . . . 16

I Texture-Level Approach **19**

2 Writer Identification Using Edge-Based Directional Features **21**
2.1 Introduction . . . 21
2.2 Experimental data . . . 22
2.3 Features . . . 23
2.3.1 Edge-direction distribution . . . 23
2.3.2 A new feature: edge-hinge distribution . . . 25
2.3.3 Run-length distributions . . . 27
2.3.4 Autocorrelation . . . 28
2.3.5 Entropy . . . 28
2.4 Results . . . 28

2.4.1 Evaluation method . . . 28
2.4.2 Analysis of performances . . . 30
2.4.3 Stability test . . . 32
2.5 Conclusions . . . 32

3 Characterizing Handwriting Individuality Using Localized Oriented Edge Fragments 35
3.1 Introduction . . . 35
3.2 Experimental data . . . 36
3.3 Feature extraction . . . 37
3.3.1 Edge-direction distribution . . . 39
3.3.2 Edge-hinge distribution . . . 40
3.3.3 Run-length distributions . . . 41
3.3.4 A new feature: horizontal co-occurrence of edge angles . . . 41
3.3.5 Brush function: ink density distribution . . . 42
3.4 Results . . . 42
3.4.1 Comparison lower- vs. upper-case and entire- vs. split-line . . . 43
3.4.2 Voting feature combination . . . 48
3.5 Conclusions . . . 48

II Allograph-Level Approach 51

4 Grapheme Clustering for Writer Identification and Verification 53
4.1 Introduction . . . 53
4.2 Theoretical model . . . 55
4.3 Datasets . . . 57
4.4 Segmentation method . . . 59
4.5 Grapheme codebook generation . . . 60
4.6 Computing writer-specific grapheme-emission PDFs . . . 61
4.7 Results . . . 62
4.7.1 Writer identification . . . 62
4.7.2 Writer verification . . . 64
4.8 Conclusions . . . 67

5 Feature Fusion for Text-Independent Writer Identification and Verification 69
5.1 Introduction . . . 69
5.2 Experimental datasets . . . 73
5.3 Textural features . . . 75
5.3.1 Contour-direction PDF (*f1*) . . . 78

5.3.2 Contour-hinge PDF (*f2*) . . . 79
5.3.3 Direction co-occurrence PDFs (*f3h*, *f3v*) . . . 81
5.3.4 Other texture-level features: run-length PDFs (*f5h*, *f5v*), autocorrelation (*f6*) . . . 81
5.4 Allographic features . . . 83
5.5 Feature matching and fusion for writer identification and verification . . . 85
5.6 Results . . . 88
5.6.1 Performances of individual features . . . 88
5.6.2 Performances of feature combinations . . . 91
5.7 Discussion . . . 95
5.8 Conclusions . . . 100

6 Concluding the thesis **101**
6.1 Summary and Contributions . . . 101
6.2 Further research directions . . . 105

A GRAWIS: Groningen Automatic Writer Identification System **107**
A.1 Introduction . . . 107
A.2 Visualization tool . . . 107
A.3 Examples of writer identification hit lists . . . 108
A.4 Examples of writer verification errors . . . 108

Bibliography **117**

Index **127**

Publications **131**

Samenvatting **133**

Acknowledgements **139**