

# Chapter 15

## Statistical Features for Text-Independent Writer Identification

**Zhenan Sun**  
*NLPR, CAS, China*

**Bangyu Li**  
*NLPR, CAS, China*

**Tieniu Tan**  
*NLPR, CAS, China*

### ABSTRACT

*Automatic writer identification is desirable in many important applications including banks, forensics, archeology, and so forth. A key and still open issue in writer identification is how to represent the distinctive and robust features of individual handwriting. This chapter presents three statistical feature models of handwritings in paragraph-level, stroke-level, and point-level, respectively, for text-independent writer identification. The three methods evolve from coarse to fine, showing the technology roadmap of handwriting biometrics. The proposed methods are evaluated on CASIA handwriting databases and perform well in both Chinese and English handwriting datasets. And the experimental results show that fine scale handwriting primitives are advantageous in text-independent writer identification. The best performing method adopts the probability distribution function and the statistical dynamic features of tripoint primitives for handwriting feature representation, achieving 95% writer identification accuracy on CASIA-HandwritingV2 with 1,500 handwritings from more than 250 subjects. And a demo system of online writer identification is developed to demonstrate the potential of current algorithms for real world applications.*

DOI: 10.4018/978-1-60566-725-6.ch015

## INTRODUCTION

Handwriting has been widely accepted as a unique characteristic of individuals for a long history. So the identity of a subject has been associated with his handwriting in both forensic and civilian applications. Traditional writer identification relies on manual efforts, which greatly limits the applicability of handwriting biometrics. With fast development of computer, sensor and pattern recognition, automatic writer identification has become possible and attracted great interests from researchers. However, accurate writer authentication is still a very challenging task because handwriting as a behavioral biometrics is less stable and distinctive than biological biometric traits such as iris and fingerprint.

The approaches for automatic writer authentication can be classified into two broad categories based on their requirements on the textual contents of handwriting. One category of approaches are named as *text-dependent* because they require the characters or words of the handwriting samples used for match should be similar or even identical, e.g. signature verification. In contrary, the other category of approaches are named as *text-independent* because they need no prior knowledge on the text contents of handwriting and the probe and gallery handwritings may be totally different in contents. Text-dependent approaches may align the corresponding text elements between two handwritings for precise one-to-one match, which is advantageous in pattern recognition. However, text-dependent approaches are limited in specific texts and easier to be cracked with forgery. In contrary, **text-independent writer identification** (TIWI) is insensitive to text content but it needs to find the similarity of two intra-class handwritten documents with different contents. So TIWI is a more challenging task but it is more resistant to forgery and has much broader applications.

Digital handwriting can be acquired into computer through off-line or on-line methods. Off-line methods scan the text blocks on handwritten

documents to digital images. So image processing and analysis are the main technical tools of off-line writer identification. With fast development of pen-enabled electronic devices such as smart phone, PDA and Tablet PCs, temporal handwriting signals including pen-position, pen-down or pen-up, pen-pressure, pen-altitude, pen-azimuth at each sampling time can be recorded online. So online handwriting is in a higher dimensional signal format (6D or even higher), which contains much more information than two-dimensional image of off-line handwriting.

Automatic writer identification has a long history in pattern recognition and references of traditional methods can be found in the review paper of (Plamondon, & Lorette, 1989). This paper addresses the problem of text-independent writer identification (TIWI) from both off-line and on-line handwritings. Firstly the recent research works in text-independent writer identification are surveyed as follows.

Typical off-line TIWI methods proposed in literature are based on texture analysis (Said, Tan, & Baker, 2000; Said, Peake, Tan, & Baker, 1998), graphemes (Bensefia, Paquet, & Heutte, 2005a & 2005b), text line (Marti, Messerli, & Bunke, 2001; Hertel, & Bunke, 2003), Hidden Markov Model (Schlapbach, & Bunke, 2004), edge-based directional probability distributions (Bulacu, Schomaker, & Vuurpijl, 2003), hinge transform (Bulacu, & Schomaker, 2007), allo-graphic fraglet codebooks (Bulacu, & Schomaker, 2007; Schomaker, & Bulacu, 2004). Said et al. (1998; 2000) globally regarded the handwriting image as a texture rather than a collection of text elements (Said, Tan, & Baker, 2000; Said, Peake, Tan, & Baker, 1998). So well-developed texture analysis methods such as Gabor filtering and gray-scale co-occurrence matrix can be directly used for handwriting feature representation. Texture-based methods have the advantage of robustness to noisy handwriting documents but they lose the local features for accurate identification. In a database of 40 writers, the texture methods (Said,

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/statistical-features-text-independent-writer/36926](http://www.igi-global.com/chapter/statistical-features-text-independent-writer/36926)

## Related Content

---

### Watermark Embedding for Multiscale Error Diffused Halftone Images by Adopting Visual Cryptography

Yuanfang Guo, Oscar C. Auand Ketan Tang (2017). *Biometrics: Concepts, Methodologies, Tools, and Applications* (pp. 137-156).

[www.irma-international.org/chapter/watermark-embedding-for-multiscale-error-diffused-halftone-images-by-adopting-visual-cryptography/164602](http://www.irma-international.org/chapter/watermark-embedding-for-multiscale-error-diffused-halftone-images-by-adopting-visual-cryptography/164602)

### Feature Selection in High Dimension

Sébastien Gadatand Sébastien Gadat (2011). *Advances in Face Image Analysis: Techniques and Technologies* (pp. 97-119).

[www.irma-international.org/chapter/feature-selection-high-dimension/43823](http://www.irma-international.org/chapter/feature-selection-high-dimension/43823)

### A Critical Overview of Net Zero Energy Buildings and Fuzzy Cognitive Maps

Eleni S. Verginiand Peter P. Groumpos (2015). *International Journal of Monitoring and Surveillance Technologies Research* (pp. 20-43).

[www.irma-international.org/article/a-critical-overview-of-net-zero-energy-buildings-and-fuzzy-cognitive-maps/146153](http://www.irma-international.org/article/a-critical-overview-of-net-zero-energy-buildings-and-fuzzy-cognitive-maps/146153)

### Extended Automata for Temporal Planning of Interacting Agents

Christine Largouët, Omar Krichenand Yulong Zhao (2017). *International Journal of Monitoring and Surveillance Technologies Research* (pp. 30-48).

[www.irma-international.org/article/extended-automata-for-temporal-planning-of-interacting-agents/182505](http://www.irma-international.org/article/extended-automata-for-temporal-planning-of-interacting-agents/182505)

### Application of Artificial Neural Networks to Reliable Nuclear Data for Nonproliferation Modeling and Simulation

Pola Lydia Lagari, Vladimir Sobes, Miltiadis Alamaniotisand Lefteri H. Tsoukalas (2016). *International Journal of Monitoring and Surveillance Technologies Research* (pp. 54-64).

[www.irma-international.org/article/application-of-artificial-neural-networks-to-reliable-nuclear-data-for-nonproliferation-modeling-and-simulation/180666](http://www.irma-international.org/article/application-of-artificial-neural-networks-to-reliable-nuclear-data-for-nonproliferation-modeling-and-simulation/180666)