International Journal of Pattern Recognition
and Artificial Intelligence
Vol. 31, No. 5 (2017) 1756008 (32 pages)

DOI: 10.1142/S0218001417560080

Off-line Text-Independent Writer Recognition: A Survey

Yu-Jie Xiong* and Yue Lu†

Shanghai Key Laboratory of Multidimensional Information Processing
Department of Computer Science and Technology
East China Normal University, Shanghai 200241, P. R. China
**xiong@stu.ecnu.edu.cn*
†ylu@cs.ecnu.edu.cn

Patrick S. P. Wang‡

Northeastern University, Boston, MA 02115, United States
‡patwang@ieee.org

Received 13 July 2016
Accepted 25 October 2016
Published 26 January 2017

Writer recognition is to identify a person on the basis of handwriting, and great progress has been achieved in the past decades. In this paper, we concentrate ourselves on the issue of off-line text-independent writer recognition by summarizing the state of the art methods from the perspectives of feature extraction and classification. We also exhibit some public datasets and compare the performance of the existing prominent methods. The comparison demonstrates that the performance of the methods based on frequency domain features decreases seriously when the number of writers becomes larger, and that spatial distribution features are superior to both frequency domain features and shape features in capturing the individual traits.

Keywords: Off-line; text-independent; writer recognition; feature extraction; classification; datasets; performance evaluation.

1. Introduction

The requirements of personal authentication for information security have placed biometrics at the center of the academic and industrial research.[63] Biometrics refers to the automatic identification or verification of persons using their individual physical or behavioral characteristics. In terms of the traits employed, biometrics can be classified into physiological biometrics and behavioral biometrics. Writer recognition is a branch of behavioral biometrics using handwriting as individual characteristics for authentication. An ideal biometric should be universal, where each person possesses the characteristic; unique, where nobody should share the same

† Corresponding author.

characteristic; permanent, where the characteristics should not change over time; and collectable, where the characteristics should be quantifiable and easy to obtain.[62] It appears that handwriting cannot completely satisfy all the characteristics. However, handwriting is the most widespread carrier of personal behavioral information, and signatures have already been used as the legitimate means to verify an individual's identity for several centuries. People always do some writing in daily life, and the growing popularity of digital cameras and smartphones makes the collection of handwriting become more and more convenient. For these reasons above, handwriting is still an effective way to represent the uniqueness of individual, and plays an essential role in biometric identification.

Writer recognition has attracted lots of researchers who are interested in scientific challenges and potential applications. In the early years, researchers extensively investigated the problem of text-dependent writer recognition. Plamondon and Lorette[80] summarized the progress of writer identification and signature verification. Since the year of 2000, researchers paid continuous attention to reaching the goal of text-independent writer recognition. Successively, numerous approaches for different conditions and languages were proposed. Several surveys[7,14,67,92] were published to cover the progress of writer recognition in the past years. In this paper, we concentrate ourselves on the issue of off-line text-independent writer recognition by summarizing the state of the art methods from the perspectives of feature extraction and classification. We also exhibit some public datasets and compare the performance of some existing prominent approaches.

The remainder of this paper is organized as follows. Section 2 summarizes the problem of off-line text-independent writer recognition and analyzes some related concepts. Feature extraction and classification are presented in Secs. 3 and 4 respectively. Section 5 exhibits some widely used public datasets and Sec. 6 compares the performance of writer recognition approaches. A brief discussion is given in Section 7.

2. Overview of the Problem

Writer recognition is to identify a person on the basis of handwriting, and have many potential applications. Figure 1 sketches three main modules of a writer recognition

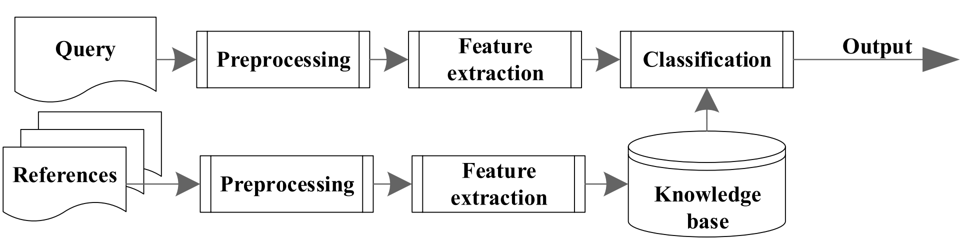


Fig. 1. Framework of a writer recognition system.

system: preprocessing, feature extraction, and classification. The role of the preprocessing module is to clean the handwriting (i.e. remove noise), segment the handwritten image into pieces, normalize the size of pieces, and do some operations which contribute to appropriate feature representation. After that, features of the references are extracted and stored into the knowledge base. Likewise, features of the query are also extracted by the same process. During the classification, trained classifiers assign the unknown query pattern to one of the known patterns under the consideration of the knowledge base.

2.1. *On-line versus off-line*

In terms of data acquisition, writer recognition can be classified into two categories: on-line and off-line. On-line data means that handwriting is collected at the same time when it is produced. The writer usually creates handwriting via a mouse or an electronic pen, and the output contains both sequential and spatial information. Off-line data refers to static image of handwritten document. Handwriting is captured by a scanner or camera, and is stored as an image. Due to the lack of sequential information, off-line writer recognition is considered as a harder task.

2.2. *Writer recognition versus handwriting recognition*

Compared with writer recognition, handwriting recognition is an older and broader research field which has lasted for several decades.[81] Off-line handwriting recognition involves automatic transcription of handwritten image into text. Since different people have different writing ways of the same character, it is necessary to eliminate individual variations of the character among a large number of writers to obtain invariant representation.[16] However, individual variations among different writers are the foundation of writer recognition, in which investigators attempt to discover the specificity of writing style to achieve the goal of authentication. Although writer recognition and handwriting recognition are completely opposite, the information of writer's writing style can be used to reduce ambiguities in general pattern representation of handwriting recognition.

2.3. *Writer recognition versus signature verification*

Writer recognition is to identify a person on the basis of handwriting, while signature verification aims to verify the identity of an individual based on the signature. Handwriting and signature are considered as two basic products of writing, but there are some differences between them. Typically, handwriting is produced with a natural writing attitude. People usually have no legal intention at this time, so they will not try to forge their writing style. Signature often signifies official act and is used for personal authentication. In this paper, we focus on the issue of off-line text-independent writer recognition. Readers who are interested in signature verification

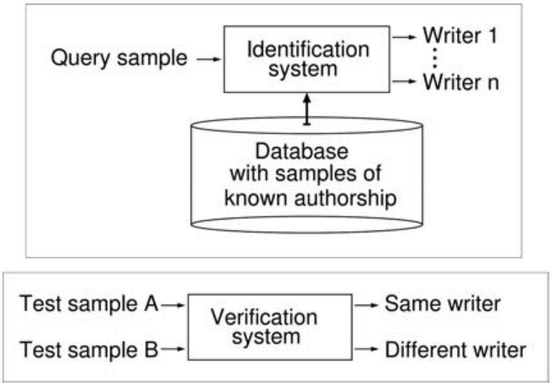


Fig. 2. A writer identification system and a writer verification system.[16]

can find more information in Ref. 59 which provided an impressive overview of signature verification and highlighted the most profitable directions of this field.

2.4. *Text-dependent versus text-independent*

On the basis of the content of handwriting, writer recognition is divided into two categories: text-dependent and text-independent.[80] The former assumes that the references and the query must be the same text content. This survey focuses on off-line text-independent writer recognition, which eliminates the restriction of text content. Text-independent methods usually require sufficient handwritten text of each writer to extract robust statistical features for pattern representation. Thus, the minimal amount of enough handwritten text is of crucial importance.[15,36,37]

2.5. *Writer identification and writer verification*

Writer identification and writer verification are two different tasks of writer recognition.[111] As shown in Fig. 2, writer identification is to find the author of the query, similar to a one-to-many search. The output of an identification system is a sorted list of candidates of writers. Writer verification is a one-to-one comparison of two handwriting samples. The goal of verification is to determine whether they are produced by the same person or not. It is obvious that writer identification techniques can be applied to writer verification favorably.[1,13,20,30,88]

3. Feature Extraction

Features widely used for off-line text-independent writer recognition can be generally categorized into two types: texture and shape features. When texture features are used, handwriting is characterized as a series of texture properties. On the other hand, when shape features are used, handwriting is characterized as a group of

Table 1. Feature categories.

Categories		References	Advantages	Disadvantages
	Frequency domain features	Gabor features[46,48,108,119] Fourier spectral features[5] Moment-based Gabor features[96,97] Extended Gabor features[51,52,55] Gabor wavelet features[49,50,99,118]	Describes the global texture of handwriting by the frequency content.	The performance decreases seriously when the number of writers becomes larger; time-consuming.
Texture features	Spatial distribution features	Run-length histogram[27] Gray level co-occurrence matrix[86] Text line-based feature[65,75,82] Gradient feature[8,23,83] Edge-hinge features[21,109] Edge structure coding[110] MPP contours[2–4], LBP[39,77] Geometrical features[43] Gray level run matrix[29,30] SIFT[33,34,57,106,113–115] SURF,[98] oBIF columns[76]	Retains the local structure information; can be applied to different languages; keeps high identification accuracy when the number of writers becomes larger.	Needs abundant handwritten text of each writer.
Shape features		Allographs[17,18,93–95] Graphemes[1,10–12,64,105] Fragments[30,37,48,60,61,100–102,107]	Similar to forensic handwriting analysis; obtains the codebook by learning algorithms.	The query and the references should be written by the same language.
Deep learning		CNN[26,35,117]	Features are learned from the data directly.	Needs enormous training data.

segmented shapes. Furthermore, texture features can be classified into two subcategories: frequency domain features and spatial distribution features. Frequency domain features concern the global traits of handwriting in frequency domain. Spatial distribution features concern local spatial structures of handwriting. With the emergence of deep learning, deep neural networks (DNN) are also used as popular tools of feature extraction for writer recognition. A brief description of different categories of features is shown in Table 1.

3.1. *Texture features*

Texture features reflect the traits of handwriting from two different aspects. The first ones are frequency domain features, which treat handwriting as a whole texture, and utilize frequency transform techniques to extract features. The second ones are spatial distribution features, which treat handwriting as a series of edges, contours, and strokes, and employ the spatial distribution of specific parts to describe the characteristics of handwriting.

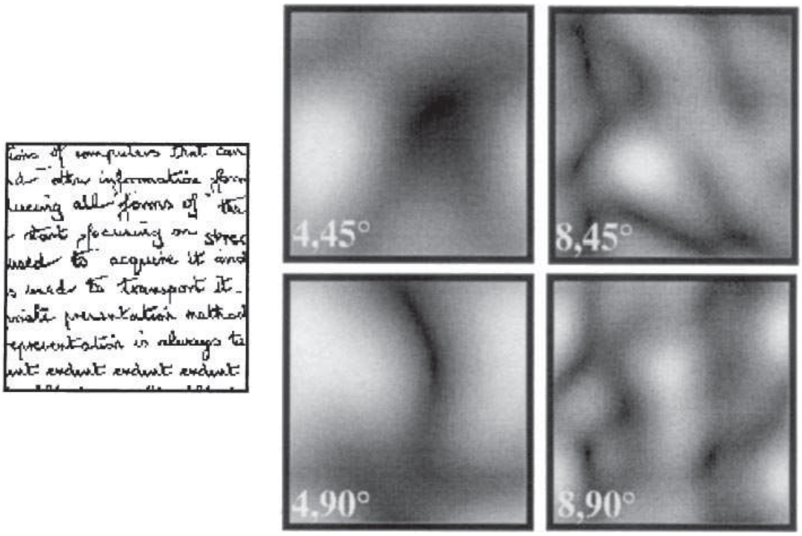


Fig. 3. The original handwriting and Gabor filters outputs.[86]

3.1.1. *Frequency domain features*

Handwriting can be treated as a particular texture, which is defined as the variations of gray level that form certain repeated patterns. Therefore, texture analysis techniques can be used for feature extraction.[80] Frequency transform techniques are typical representatives of them. Said *et al.*[86] proposed a texture analysis approach using multichannel 2-d Gabor filtering technique (see Fig. 3). The multichannel 2-d Gabor filters is a series of Gabor filters with different spatial frequencies and orientations. The mean value and standard deviation of filtered images are used as global features. If a orientations and b frequencies for each orientation are selected to create Gabor filters, a feature vector with $2 \times a \times b$ elements will be extracted. The mathematical expressions of the 2-d Gabor model[45] are

$$h_e(x, y, f, \theta) = g(x, y) \cdot \cos[2\pi f(x \cos\theta + y \sin\theta)], \tag{1}$$

$$h_o(x, y, f, \theta) = g(x, y) \cdot \sin[2\pi f(x \cos\theta + y \sin\theta)], \tag{2}$$

where f and θ are the spatial frequency and orientation of the Gabor filter. The even and odd symmetric Gabor filters are denoted as h_e and h_o, respectively. $g(x, y)$ is a 2-d Gaussian function

$$g(x, y) = \frac{1}{2\pi\sigma^2} \cdot \exp[-(x^2 + y^2)/2\sigma^2], \tag{3}$$

where σ is the space constant. Zhu *et al.*[119] adopted the spatial frequency responses of Gabor filters as features for Chinese writer identification. He and Tang[46] used both Gabor features and autocorrelation function to extract features. Ubul *et al.*[108] utilized the Gabor features combined with feature selection technique for Uyghur writer identification. Shahabi and Rahmati[96,97] improved 2-d Gabor filters with moment-based features. Instead of mean value and standard deviation of filtered images, the moments are extracted by sliding windows as features for Farsi writer identification. Helli and Moghadam[51,52,55] presented the extended Gabor model called as XGabor, and defined the ratio of the strength of the filtered image versus the strength of the original handwriting image as the feature for Persian writer identification, where the strength of image is the sum of all image pixel values.

Shen *et al.*[99] introduced the Gabor wavelet technique to extract features owing to the advantage that wavelet representation is able to describe spatial structure of the images while preserving information of spatial relations. The handwriting image is decomposed into a series of wavelet subbands and wavelet coefficients (e.g. the energies of wavelet sub-bands) are used as the features. Wavelet coefficients are employed to build different models for the handwriting of each writer. He *et al.*[44,45] presented a wavelet-based generalized Gaussian density model which can reduce the excessive calculation cost to replace the traditional Gabor model. He *et al.*[49] also proposed a method using wavelet based hidden Markov tree model (HMTM) for Chinese writer identification. The wavelet transform of the image $f(x, y)$ is defined as

$$W_{s,f}(x, y) = f(x, y) * \psi_s(x, y), \tag{4}$$

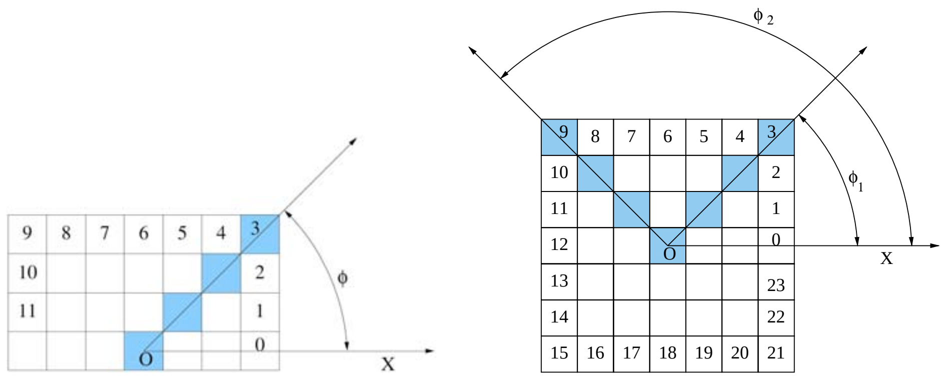
where $*$ is the 2-d convolution operator, s is the scale, and $\psi_s(x, y)$ is the multi-scale 2-d wavelet function. $\psi_s(x, y)$ is defined as

$$\psi_s(x, y) = (1/s^2)\psi(x/s, y/s). \tag{5}$$

Zhang *et al.*[118] proposed a hybrid method combining Gabor wavelet and mesh fractal dimension for feature extraction. The first step of the hybrid method is to decompose the original handwriting by the 2-d Gabor filters. Then, decomposed images are reshaped to 1-d sequences. After that, the sequences are decomposed using the wavelet filters. Finally, the mesh fractal dimension of each sequence is the Gabor wavelet-fractal feature vector. Al-Dmour and Abu Zitar[5] presented a hybrid Fourier spectral statistical approach for Arabic writer identification. Bertolini *et al.*[13] utilized local phase quantization (LPQ) extracted by the 2-d short-term Fourier transform for writer identification.

3.1.2. *Spatial distribution features*

Different from frequency domain features, spatial distribution features represent the characteristics of handwriting by statistical information of spatial structures extracted from the edges, keypoints and text lines. Dinstein and Shapira[27] employed black pixel run-length histograms for writer recognition. Another type of texture feature is gray level co-occurrence matrix, which is often used as a benchmark feature.[86] Djeddi *et al.*[29,30] extracted the run-length features from the gray level run matrix (GLRM) to describe the characteristics of handwriting, which takes four orientations (0°, 45°, 90°, 135°) into account. It is well known that the gradient



(a) Edge-direction feature

(b) Edge-hinge feature

Fig. 4. Edge-direction feature and edge-hinge feature.[21]

contains both magnitude and direction information, and can be used to describe the property of texture. Ram and Moghaddam[83] used the direction information of gradient features for writer identification. Their experiments showed that the direction intervals $[0°, 60°]$ and $[300°, 360°]$ are the most important and essential discriminative ranges in Persian handwriting. Chanda *et al.*[23] used both chain-code-based direction features (quantified into four directions) and gradient-based direction features (quantified into 16 directions) for Bengali writer identification, and similar method[8] was also proposed for Arabic writer identification.

Bulacu *et al.*[21] proposed two kinds of edge-based features: edge-direction feature and edge-hinge feature (see Fig. 4). Edge-direction feature is extracted from a binary image which is composed of edge pixels. Each edge pixel emerging from the central edge pixel is checked, and the direction of each verified instance is accumulated into the histograms of probability distribution. The primary innovation of the edge-hinge feature is to consider both edges emerging from the central edge pixel. Subsequently, joint probability distribution of directions of two edges is computed. Van der Maaten and Postma[109] improved the edge-hinge feature by multi-scale analysis. Li and Ding[68] extended the edge-hinge feature into grid microstructure feature (GMF, see Fig. 5) for Chinese writer identification. The edge-hinge feature inspects the directions of the two nearest edge pixels emerging from the central edge pixel, while GMF focuses on recording the positions of each edge pixel pairs in the grid. Xu *et al.*[116] adopted GMF combined with weighted feature matching method to solve the problem of sensitivity of pen-width. Wen *et al.*[110] characterized the frequent structures distribution of edge fragments on multiple scales using the edge structure coding (ESC), which can be considered as a generalization of GMF. He and Schomaker[47] extended edge-hinge feature to delta-n hinge features, which are rotation-invariant. Besides the distribution of edge pixels, structural information extracted from contours are also commonly used. Abdi *et al.*[2–4] extracted the lengths, directions, angles and curvatures from the minimum-perimeter polygon (MPP) contours

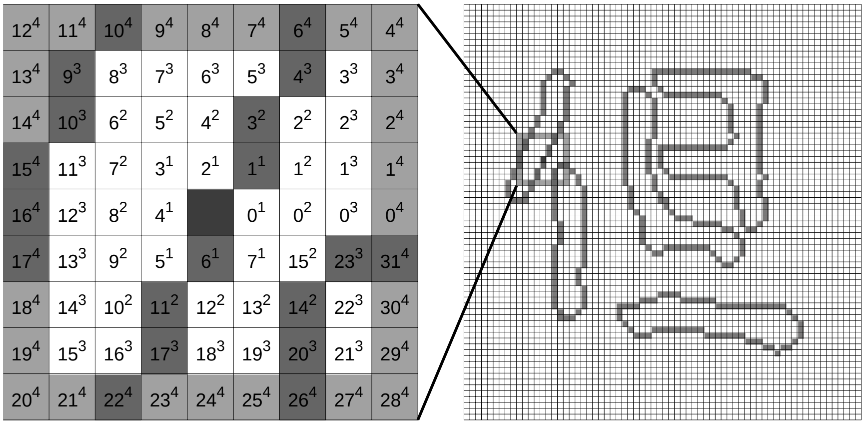


Fig. 5. Extraction of the GMF.[68]

of Arabic handwriting to construct the feature vectors. Hassaine *et al.*[43] proposed a set of geometrical features to characterize English writing style. Newell and Griffin[76] adopted oriented basic image feature (oBIF) columns for writer identification, and enhanced the performance by encoding the writer's style as the deviation from the mean encoding for the population of writers. Some well-known textural descriptors for keypoints are applicable to writer recognition, such as scale invariant feature transform (SIFT)[72] and speeded-up robust feature (SURF).[9] Woodard *et al.*[113] used the quantized SIFT as the local feature for Arabic writer recognition. Fiel and Sablatnig[33,34] utilized the SIFT descriptors for both writer identification and writer retrieval. Fecker *et al.*[32] proposed an SIFT-based method for historical Arabic writer identification. Christlein *et al.*[25] used the RootSIFT-based GMM supervectors to encode the features for each writer. Hu *et al.*[57] presented two coding strategies as improved fisher kernels coding and locality-constrained linear coding to encode the SIFT descriptors to describe the writing style. Wu *et al.*[114] employed the SIFT descriptors with both scale and orientation information extracted from word regions for multilingual writer identification. Tang *et al.*[106] presented an identification system using the SIFT descriptors combined with triangular features. Xiong *et al.*[115] employed the modified SIFT descriptors with contour-directional feature to create a two-stage system for writer identification. Sharma and Dhaka[98] proposed language-free writer identification based on SURF. local binary Patterns (LBP)[78] are also introduced for writer recognition. The LBP maps each pixel to an integer code representing the relationship between the center pixel and its neighborhoods. It encapsulates the local geometry at each pixel by encoding binarized differences with neighbor pixels as

$$\mathrm{LBP} = \sum_{n=0}^{N-1} s(p_n, p_c) * 2^n, \tag{6}$$

where p_c is the central pixel being encoded, p_n are N symmetrically and uniformly sampled points on the periphery of a circular area of p_c, and $s(p_n, p_c)$ is a binarization function. A widely used binarization function $s(p_n, p_c)$ is defined as

$$s(p_n, p_c) = \begin{cases} 1\colon p_n \geq p_c, \\ 0\colon p_n < p_c. \end{cases} \tag{7}$$

Nicolaou *et al.*[77] presented an oriented texture feature set based on LBP. Bertolini *et al.*[13] assessed the performance of LBP and LPQ, and demonstrated that both LBP and LPQ surpass GLCM features by a considerable margin in writer verification. Hannad and Siddiqi[39] utilized LBP for Arabic writer identification, and their subsequent work[40] evaluated the effectiveness of LBP, LPQ and local ternary patterns (LTP).

Some researchers also tried to extract features from text lines. Marti *et al.*[75] extracted 12 features from text line images, including height of the text line, slant information, and slopes of the second and third line segments. Rafiee and Motavalli[82] extracted a set of similar features from the baseline of text line images. Kirli and Gulmezoglu[65] utilized global and local information of three main writing zones such as width, thickness, surface area, size, and density to reveal individual writing style. Schlapbach and Bunke[87–90] utilized hidden Markov model (HMM) and Gaussian mixture model (GMM) for handwriting modeling. A sliding window is used to transform a normalized text line into a sequence of feature vectors. The features are represented by the number of black pixels in the window, center of gravity, second-order moment, number of black-to-white transitions in the window, and so on.

3.2. *Shape features*

Different from spatial distribution features, shape features employ the local closed regions of characters/letters to represent the characteristics of handwriting. Schomaker and Bulacu[93,94] assumed that the writer is a stochastic allograph generator and proposed the connected-component contours (CO^3, see Fig. 6) to describe the shape of allograph. Schomaker and Bulacu[17] also discussed the clustering methods for shape codebook generation (see Fig. 7). Allographic features were successfully applied to historic documents.[18,95] Bensefia *et al.*[10–12] exploited graphemes extracted from the fragments of handwriting as the invariants of each writer for writer recognition. Sreeraj and Sumam[105] extracted graphemes from the Malayalam handwritten documents for writer identification and verification. Abdi and Khemakhem[1] synthesized graphemes using beta-elliptic model rather than clustering original graphemes from the segmented handwriting. Instead of generating one single codebook, Khalifa *et al.*[64] utilized an ensemble of grapheme codebooks to describe properties of handwriting, and improved the performance by the fusion of multiple codebooks.

Smaller patches of the characters/letters are also used to extract shape features. Siddiqi and Vincent[100–102] divided the words into small sub-images which only

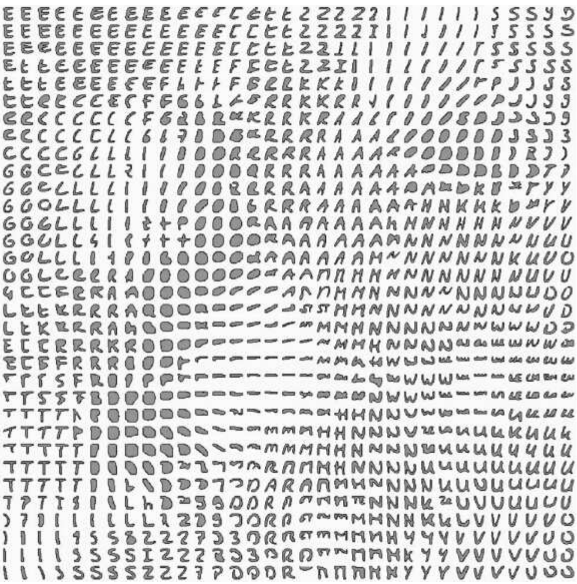


Fig. 6. A Kohonen self-organized map of CO^3.[93]

contain a part of stroke, and utilized the sub-images to represent the redundant patterns which are specific to different writers. Ghiasi and Safabakhsh[36] used small fragments of connected components rather than complex shapes to describe the writing style, and improved the encoding methods of fragments[37] by linear piece-wise approximation. Tang *et al.*[107] proposed a writer identification approach using the stroke fragment histogram combined with the local contour pattern histogram. Jain and Doermann[60,61] utilized k-adjacent segments and contour gradient descriptors (CGD) instead of original character contours or graphemes for shape representation. He *et al.*[48] proposed a method to detect the primitive junctions in the stroke fragments, and used the probability distribution of junctions as the feature to distinguish different writers.

3.3. *Deep learning*

Unlike traditional methods, DNN can learn a feature mapping from training data directly. Compared with hand-designed feature descriptors, learning-based feature representation methods usually show better recognition performance because more data-adaptive information can be exploited in the learned features. Christlein *et al.*[26] used convolutional neural networks (CNN) to learn the activations of the hidden layers and encoded them to feature vectors by GMM supervector encoding for classification. Fiel and Sablatnig[35] used the output of the second-to-last fully connected layer as the feature vector for writer identification based on the Chi-square distance. Unlike above methods which only use CNN for feature extraction,

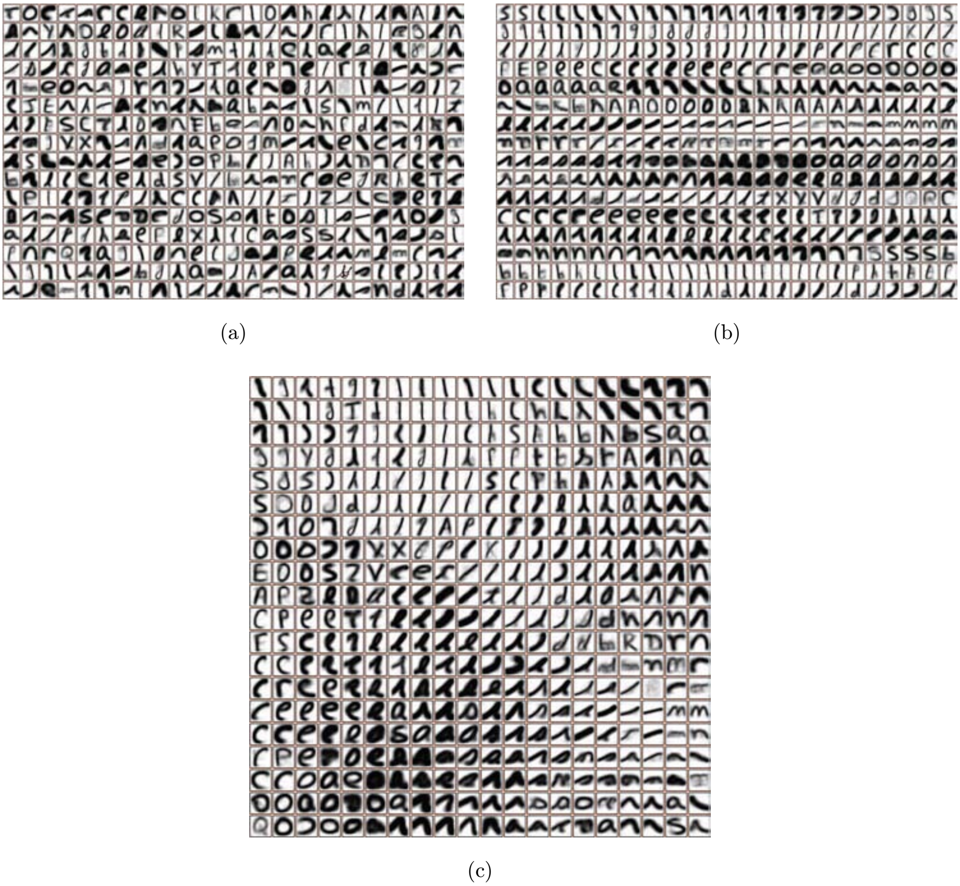


Fig. 7. Examples of codebooks with 400 graphemes, (a) k-means, (b) k-som1D and (c) k-som2D.[17]

Yang *et al.*[117] proposed an end-to-end writer recognition system which employs CNN for both feature extraction and classification directly.

4. Classification

In the classification process, the authenticity of the query is determined by matching its features to those stored in the knowledge base. This process produces a sorted list that states the authenticity of the query. Many classifiers can be used for text-independent writer recognition.

Figure 8 shows some of the most relevant classifiers which can be divided into three categories. The simplest and the most intuitive classifier is based on the concept of distance. In practice, there are not enough handwriting samples to train the parameters of a complex model. Hence, parameter-free distance-based classifiers are appropriate. The unknown pattern of the query is assigned to the pattern which contains the sample with the minimum distance to the query. For a distance-based

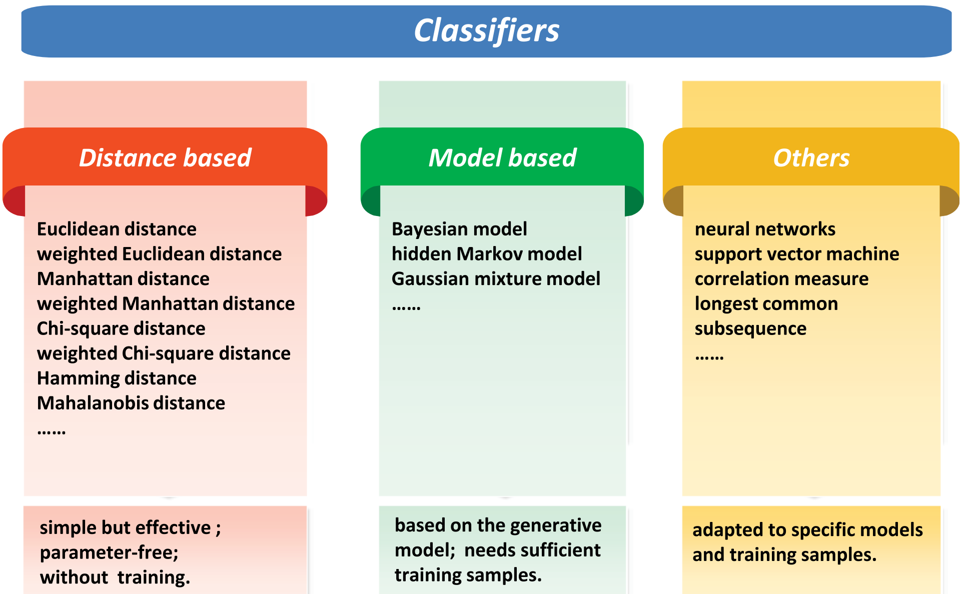


Fig. 8. Classifiers for writer recognition.

classifier, the distance metric is very important. The commonly used distance metrics for off-line text-independent writer recognition include Euclidean distance, weighted Euclidean distance, Manhattan distance, weighted Manhattan distance, Chi-square distance, weighted Chi-square distance, Hamming distance, and Mahalanobis distance. Among them, Euclidean distance is usually used as the benchmark to show the effectiveness of other distance metrics. Wirotius *et al.* [112] employed the minimum distance classifier with Mahalanobis distance for writer identification. Schomaker *et al.*[93,94] used the nearest-neighbor classifier with both Hamming distance and Chi-square distance. In Refs. 45 and 49, the Kullback–Leibler divergence was used to calculate the similarity between probability distributions. The use of Chi-square distance metric yielded the best experimental results in Refs. 21, 33 and 97. However, the use of Hamming distance metric obtained the highest identification accuracy in Ref. 39. This is because the effectiveness of distance metrics is highly related to the features employed. Hence, some approaches adopt different distance metrics[4,19] to calculate the similarity of corresponding features. Compared with original distance metrics, weighted distance metrics attempt to balance the contribution of each component of the feature vector. It is obvious that weighted metric always yields a better result than that of original metric.

The second type of classifier is based on the probabilistic model. Siddiqi and Vincent[100] utilized the naive Bayes classifier which assigns input pattern to the class with the maximum posterior probability for classification. Kirli and Gulmezoglu[65] used normal density discriminant function Bayes classifier to identify the writer.

Due to the capability of adapting the personal variability of different writers, HMM and GMM are applied to handwriting modeling. Schlapbach *et al.*[87,88,90] presented a system for writer recognition using HMM-based recognizers. For each writer in the references, an individual HMM-based recognizer is built and trained with the data coming exclusively from that writer. The query is the input of each recognizer, and the output of each recognizer is a sequence of words with their log-likelihood scores. If the input and the training data come from the same writer, the output of the trained recognizer is more likely to be correctly recognized with the highest score. Therefore, the log-likelihood scores of all the recognizers are sorted and the query is assigned to the corresponding writer with the highest score. Schlapbach *et al.*[89,91] proved that the GMM-based system is conceptually much simpler and trained faster than the HMM-based system under the same condition.

Some other classifiers and similarity metrics are also used for writer recognition. Bensefia *et al.*[10] adopted correlation measure to calculate the similarity of graphemes. Saad[84] utilized fuzzy logic and genetic algorithm to estimate the similarity of different handwriting samples. Longest common subsequence[51,53] was employed to calculate the distance between the references and the query. Helli and Moghaddam[54] employed a graph framework to determine the similarity of the feature relation graph (FRG) generated from the references and the query. When training samples are sufficient, neural networks (NN)[55,75,82,83] and support vector machine (SVM)[5,13,22,23,58,108] were widely applied to writer recognition.

5. Datasets

Public datasets play a crucial role in validating the performance of various approaches, and a growing number of public datasets is the fundamental prerequisite for the development of writer recognition. In early studies, some famous datasets containing writer information for handwriting recognition[74,104] were used for writer identification task directly. As the research goes along, more and more specialized datasets are collected to satisfy the growing research requirements for various conditions. In this section, we review some widely used datasets which contain different image formats, numbers of writers and languages. A brief description is shown in Table 2.

Table 2. Popular datasets for writer recognition.

#	Dataset	Year	Images	Writers	Language
1	IAM[74]	1999	1539	657	English
2	Firemaker[21]	2002	1000	250	Dutch
3	ImUnipen[19]	2007	416	208	English
4	IFN/ENIT[79]	2002	2265	411	Arabic
5	HIT-MW[104]	2007	241	241	Chinese
6	QUWI[6]	2012	5085	1017	Arabic and English
7	HaFT[85]	2013	1800	600	Farsi
8	CVL[66]	2013	1609	311	German and English

5.1. *IAM*

The IAM dataset[74] contains 1539 English handwriting document images from 657 writers. It includes the segmentation and ground-truth information at the text line, sentence, and word levels. IAM includes a variable number of handwritten pages per writer, from 1 page to 59 pages. In order to build a comparable experimental condition, researchers modified the IAM so that there are only two samples for each writer. For those writers who have more than two documents in the original IAM, only the first two documents are kept. For those writers who only have one document, the document is split roughly into half.

5.2. *Firemaker*

The Firemaker dataset[21] contains 1000 handwriting pages collected from 250 Dutch writers and four pages each writer. Page 1 contains five short paragraphs with normal handwriting. Page 2 contains two paragraphs using uppercase letters. Page 3 contains forged handwriting with an unnatural writing attitude. Page 4 contains a description about a given cartoon which is written by writers in their own words. In general, only Pages 1 and 4 are used for writer identification.

5.3. *ImUnipen*

The ImUnipen dataset[19] is the off-line version of an on-line dataset.[38] It contains 416 handwriting samples from 208 writers (two samples per writer). The images were derived from the on-line handwritten data, so the images of ImUnipen are synthetic. This dataset is often used to train codebook rather than test writer identification and verification approaches directly.

5.4. *IFN/ENIT*

The IFN/ENIT dataset[79] consists of 26 459 images of the 937 names of cities and towns in Tunisia, written by 411 different writers. Each writer was asked to fill 5 pages, and each page contains 12 city names. This dataset has been widely used by researchers of Arabic handwritten text recognition. Due to its public availability, researchers have also used the IFN/ENIT dataset for writer identification of Arabic text although it is limited to city names.

5.5. *HIT-MW*

The HIT-MW dataset[104] consists of 853 pages of handwriting and 186 444 characters produced under an unconstrained condition without preprinted character boxes. However, only 254 images from 241 writers are labeled with writer's information. For the labeled images, only the first page for each writer is used and each page is segmented into two commensurate parts.

5.6. *QUWI*

The QUWI dataset[6] contains both Arabic and English off-line handwriting samples. It consists of handwritten documents of 1 017 volunteers of different ages, nationalities, genders and education levels. Each writer was asked to copy a given text and to generate a random handwriting in both languages, which allows QUWI to be used for both text-dependent and text-independent writer recognition tasks.

5.7. *HaFT*

The HaFT dataset[85] is an off-line Farsi handwriting dataset for the purpose of text-independent writer identification and verification tasks. The dataset contains 1800 gray level images of unconstrained handwriting written by 600 writers. There are three eight text-line samples for each writer, each of which was written at a different time on a separate sheet. Four versions of the dataset were created with the sample lengths of 1, 2, 4, and 8 lines. Furthermore, 120 of the 600 participants wrote with the same writing instruments to create another versions of the dataset called UniHaFT.

5.8. *CVL*

The CVL dataset[66] is built for writer retrieval, writer identification and word spotting. It has 311 different writers and seven different handwriting samples (one German and six English) for each writer. CVL stores handwriting samples in the format of color images, and the bounding boxes for each word are also available.

6. Performance Evaluation

In order to evaluate the effectiveness of various writer recognition approaches, we summarize experimental results in the literature to analyze the performance of existing prominent approaches. We also report the performance of several approaches which are the winners and runner-ups of the international competitions[28,41,42,69–71,73,103] in International Conference on Document Analysis and Recognition (ICDAR) and International Conference on Frontiers in Handwriting Recognition (ICFHR) from 2011 to 2015. The emergence of recent competitions provides us the opportunities to fairly compare the approaches in the state of the art with the same training samples and testing samples.

6.1. *Evaluation criteria*

The soft and hard TOP-N criteria[69] are widely used as evaluation criteria for writer identification. On the other hand, false acceptance rate (FAR), and false rejection rate (FRR) are utilized as evaluation criteria for writer verification.

For the soft TOP-N criterion, a correct hit is accumulated when at least one handwriting sample in the first N places of the sorted list is written by the correct writer. As to the hard TOP-N criterion, a correct hit is accumulated only when all

Table 3. Four possible conditions with different inputs and outputs.

Output/Input	C_1	C_2
C_1	TAR	FAR
C_2	FRR	TRR

Notes: (TAR = true acceptance rate, FRR = false rejection rate, FAR = false acceptance rate, TRR = true rejection rate).

handwriting samples in the first N places of the sorted list are written by the correct writer. It is obvious that the hard criterion is more stringent.

Writer verification can be considered as the binary classification problem. Two types of class are C_1: two handwriting samples black belong to the same writer and C_2: two handwriting samples black belong to different writers. For a verification system, there are four possible conditions with different inputs and outputs (see Table 3). For a similarity measure x, the probability density distribution $P_{C_1}(x)$ of class C_1 and the probability density distribution $P_{C_2}(x)$ of class C_2 can be computed, respectively. Subsequently, the following cumulative probability distributions are obtained by changing the threshold θ for acceptance or rejection.

$$\mathrm{TAR}(\theta) = \int_0^{\theta} P_{C_1}(x)dx, \tag{8}$$

$$\mathrm{FRR}(\theta) = \int_{\theta}^{\infty} P_{C_1}(x)dx, \tag{9}$$

$$\mathrm{FAR}(\theta) = \int_0^{\theta} P_{C_2}(x)dx, \tag{10}$$

$$\mathrm{TRR}(\theta) = \int_{\theta}^{\infty} P_{C_2}(x)dx. \tag{11}$$

Among them, two types of errors are contained. The type I error is incorrect rejection of the case that two handwriting samples belong to the same writer, while the type II error is incorrect acceptance of the case that two handwriting samples belong to different writers. Receiver operating curve (ROC) is the plot of FAR and TAR with all possible thresholds, and equal error rate (EER) is the system error rate when FRR equals FAR. Both ROC and EER can also be used for performance evaluation.

6.2. *Performance of existing prominent approaches*

Table 4 is a comprehensive overview of the text-independent writer recognition systems presented in the literature. We can find that the performance of frequency domain features decreases seriously when the number of writers becomes larger, and that both spatial distribution features and shape features are superior to frequency

Table 4. Performance of writer recognition systems on different datasets.

Year	Ref.	Features	Classifiers	Writers	Images (Pages/Lines)	Dataset	Performance
2001	75	Text line based features	KNN, NN	20	100, Train:Test=4:1 (P)	IAM	Top-1: 90.7%
2004	87	Text line based features	HMM	50	2200, Train:Test=3:1 (L)	IAM	Top-1: 94.7%
2004	88	Text line based features	HMM	100	4307, Train:Test=3:1 (L)	IAM	Top-1: 96.6%
2006	89	Text line based features	GMM	100	4103, Train:Test=1:1 (L)	IAM	Top-1: 98.5%
2012	65	Text line based features	BC	93	4075, Train:Test=3:1 (P)	IAM	Top-1: 98.8%
2005	12	Graphemes	CD	150	300, Train:Test=1:1 (P)	IAM	Top-1: 86.0%
				88	176, Train:Test=1:1 (P)	PSI	Top-1: 95.0%
2009	101	Chain code features and polygon features	CHID	150	300, Train:Test=1:1 (P)	IAM	Top-1: 94.0%
				650	1300, Train:Test=1:1 (P)		Top-1: 89.0%
2010	102	Redundant pattern features, curvature	CHID	650	1300 (P, LOO)	IAM	Top-1: 91.0%
		features and orientation features		375	1875 (P, LOO)	RIMES	Top-1: 84.0%
2011	60	K-adjacent segments	KNN	650	1300 (P, LOO)	IAM	Top-1: 92.1%
				302	3020, Train:Test=7:3 (P)	MADCAT	Top-1: 90.0%
2012	33	SIFT descriptors	CHID	650	1300 (P, LOO)	IAM	Top-1: 90.8%
				47	94, Train:Test=1:1 (P)	Private	Top-1: 98.9%
2013	13	LBP and	SVM	650	—, Train:Test=2:1 (P)	IAM	Top-1: 99.6%
		LPQ		315	945, Train:Test=2:1 (P)	BFL	Top-1: 99.4%
2013	61	CGDs	KNN	301	602 (P, LOO)	IAM	Top-1: 96.5%
2013	37	Contour fragment features	CHID	650	1300, Train:Test=1:1 (P)	IAM	Top-1: 94.8%
2013	107	Stroke fragment and contour pattern features	CHID	657	1314 (P, LOO)	IAM	Top-1: 92.6%
2014	76	oBIF columns	ED	301	602 (P, LOO)	IAM	Top-1: 99.0%
2014	106	Triangular and modified SIFT descriptors	CHID	657	1314 (P, LOO)	IAM	Top-1: 97.1%
2004	93	CO^3,	HD	150	300 (P, LOO)	Firemaker	Top-1: 87.0%
		Edge-direction and edge-hinge features	CHID		300, Train:Test=1:1 (P)		Top-1: 95.0%
2004	94	Fragmented CO^3 and edge-hinge features	HD	150	300 (P, LOO)	Firemaker	Top-1: 97.0%
2005	109	Multi-scale edge-hinge features	HD	150	300 (P, LOO)	Firemaker	Top-1: 97.0%

Table 4. (*Continued*)

Year	Ref.	Features	Classifiers	Writers	Images (Pages/Lines)	Dataset	Performance
2007	19	Texture-level and allograph-leave features	CHID, ED	250	500 (P, LOO)	Firemaker	Top-1: 86.0%
				650	1300 (P, LOO)	IAM	Top-1: 89.0%
2007	20	Texture-level and allograph-leave features	CHID, HD	350	1750 (P, LOO)	IFN/ENIT	Top-1: 88%
2009	2	Length, ratio and curvature stroke features	CHID	40	—, Train:Test=2:1 (P)	IFN/ENIT	Top-1: 92.5%
2010	30	Gray level run-length features	ED	130	650 (P, LOO)	IFN/ENIT	Top-1: 71.9%
2012	31	Run-length, edge-hinge features	MD	275	1375 (P, LOO)	IFN/ENIT	Top-1: 93.5%
2015	1	Beta-elliptic model based graphemes	CHID	411	2265, Train:Test=1:1 (P)	IFN/ENIT	Top-1: 90.0%
2013	34	SIFT descriptors	CD	309	1545, (P, LOO)	CVL	Top-1: 97.8%
2014	25	RootSIFT based GMM supervectors	GMM	309	1545, (P, LOO)	CVL	Top-1: 99.2%
2009	68	GMFs	WCHID	240	480, Train:Test=1:1 (P)	HIT-MW	Top-1: 95.0%
2012	110	Fragmented edge structure features	CHID	240	480 (P, LOO)	HIT-MW	Top-1: 95.4%
2014	57	SIFT descriptors	KNN	240	480, Train:Test=1:1 (P)	CASIA	Top-1: 96.3%
2000	86	2-*d* Gabor features	WED, KNN	40	1000, Train:Test=3:2 (P)	Private	Top-1: 96.0%
2000	119	2-*d* Gabor features	WED	17	34, Train:Test=1:1 (P)	Private	Top-1: 95.7%
2002	10	Graphemes	CM	88	264,Train:Test=2:1 (P)	Private	Top-1: 97.7%
2002	99	Gabor wavelet features	KNN	50	110, - (P)	Private	Top-1: 97.6%
2003	21	Edge-direction, edge-hinge and run-length features, autocorrelation, entropy	ED	250	500 (P, LOO)	Private	Top-1: 75%
2003	56	Connected components, local regions, lower and upper contours, fractal features	KNN	50	250, Train:Test=4:1 (P)	Private	Top-1: 99.6%
				20	100, Train:Test=4:1 (P)		Top-1:100.0%
2004	46	2-*d* Gabor features	WED	50	100, Train:Test=1:1 (P)	Private	Top-1: 42.0%
		Autocorrelation function	ED	50	100, Train:Test=1:1 (P)	Private	Top-1: 90.0%
2005	45	Gabor features with GGDM	KLD	10	20, Train:Test=1:1 (P)	Private	Top-1: 80.0%
2007	5	Fourier spectral features	LDC	20	40, Train:Test=1:1 (P)	Private	Top-1: 90.0%
2007	18	Texture-level and allograph-leave features	CHID, ED	10	70, (P, LOO)	Private	Top-1: 89.0%

Table 4. (*Continued*)

Year	Ref.	Features	Classifiers	Writers	Images (Pages/Lines)	Dataset	Performance
2007	58	Steered hermite features	SVM	30	300, Train:Test=1:1 (P)	Private	Top-1: 86.5%
2007	82	Baseline features	NN	20	130, Train:Test=1:1 (P)	Private	Top-1: 83.0%
2007	96	Moment-based Gabor features	CHID	40	120, Train:Test=2:1 (P)	Private	Top-1: 82.5%
2007	100	Sub-image features	CM	50	100, Train:Test=1:1 (P)	Private	Top-1: 94.0%
2008	49	Wavelet features with HMTM	KLD	500	1000, Train:Test=1:1 (P)	Private	Top-1: 36.4%
2008	44	Wavelet features with GGDM	KLD	500	1000, Train:Test=1:1 (P)	Private	Top-1: 39.2%
2008	52	2-*d* extended Gabor features	WED	70	350, Train:Test=3:2 (P)	Private	Top-1: 77.0%
2008	51	2-*d* extended Gabor features	LCS	40	500, Train:Test=3:2 (P)	Private	Top-1: 95.0%
2009	118	Gabor features with mesh fractal dimension	WED	50	100, Train:Test=1:1 (P)	Private	Top-1: 56.0%
2010	36	Connected component fragment features	CHID	180	540 (P, LOO)	Private	Top-1: 99.7%
2013	8	Statistical and structural features	ED	250	500 (P, LOO)	Private	Top-1: 75.0%

Notes: *ED = Euclidean distance, WED = Weighted euclidean distance, KNN = K-nearest neighbor classifier, NN = neural networks, CM = correlation measure, HMM = hidden Markov model, HD = hamming distance, CHID = chi-square distance, CD = cosine distance, KLD = Kullback–Leibler divergence, GMM = Gaussian mixture model, LDC = linear discriminant classifier, SVM = support vector m = machine, GGDM = generalized gaussian density model, HMTM = hidden Markov tree model, LCS = longest common subsequence, WCHID = weighted chi-square distance, MD = mahalanobis distance, BC = Bayesian classifier, LOO = leave one out test.

Table 5. Competitions for writer identification.

Method	Rank&competition	Writers	Features	Classifiers	Performance
UCL	1st, 2011-ICDAR-A[42]	54	oBIF columns[76]	KNN	Top-1: 100.0%
SHASTA	2nd, 2011-ICDAR-A[42]	54	Directions, curvatures, etc.	KNN	Top-1: 81.1%
WAYNE	1st, 2012-ICFHR-A[41]	206	Edge-hinge features,[21] etc.	SVM	Top-1: 95.3%
NEWELL	1st, 2012-ICFHR-A[91]	206	oBIF columns[76]	KNN	Top-1: 95.3%
GMMS	1st, 2014-ICFHR-A[103]	1000	RootSIFT based GMM supervectors[25]	CD	Top-1: 73.4%
OBI/SIFT	2nd, 2014ICFHR-A[103]	1000	SIFT features[32] and oBIF columns	CD, CHID	Top-1: 63.2%
TSINGHUA	1st, 2011-ICDAR-L[71]	26	GMF[68]	WCHID	Top-1: 99.5%
MCS-NUST	2nd, 2011-ICDAR-L[71]	26	Chain-based features, etc.	KNN	Top-1: 99.5%
TEBESSA-C	1st, 2012-ICFHR-L[69]	100	GLRM[30]	MD	Top-1: 94.5%
TEBESSA-A	2nd, 2012-ICFHR-L[69]	100	Edge-hinge features[21]	MD	Top-1: 92.3%
CS-UMD-A	1st, 2013-ICDAR-L[70]	250	CGD[61]	—	Top-1: 95.1%
CS-UMD-B	2nd, 2013-ICDAR-L[70]	250	CGD[61]	—	Top-1: 95.0%
Nuremberg	1st, 2015-ICDAR-M[28]	300 (E,A)	Contour-Zernike moments[24]	SVM	Top-1: 29.0%
		300 (A,E)			Top-1: 55.0%
CVC	2nd, 2015-ICDAR-M[28]	300 (E,A)	LBP	—	Top-1: 14.0%
		300 (A,E)			Top-1: 21.0%

Note: *KNN = K-nearest neighbor classifier, SVM = support vector m = machine, CD = cosine distance, CHID = chi-square distance, WCHID = weighted chi-square distance, MD = Manhattan distance, (E,A) = English samples for training and Arabic samples for test, (A,E) = Arabic samples for training and English samples for test, -A = Arabic handwriting samples, -L = Latin handwriting samples, -M = both Arabic and Latin handwriting samples.

domain features in capturing the individual traits of handwriting. Furthermore, it can be found that the performance of spatial distribution features is better than that of shape features in most cases, which probably are due to two reasons. Firstly, segmenting a handwriting into shape elements leads to the damage of some necessary details. Secondly, the number of shapes is typically very large. We always need to generate a codebook of shapes in order to make the calculation of features computationally feasible. The usage of codebook also leads to the loss of characteristics of handwriting.

6.3. *Performance of approaches in competitions*

Table 5 reports the performance of several writer identification approaches including the winners and runner-ups of the ICDAR and ICFHR competitions. The oBIF columns is regarded as one of the most effective features of text-independent writer identification, in particular for Arabic handwriting. Newell and Griffin[76] developed an approach using oBIF columns with the Delta encoding, and achieved the first place in both 2011-ICDAR-A competition[42] and 2012-ICFHR-A competition.[41] SIFT-based features play an important role in the 2014-ICFHR-A competition.[103] The winner of the competition used the RootSIFT-based GMM supervectors[25] to encode the features for each writer, and the runner-up combined SIFT features with oBIF columns for feature representation. There are many discriminative features for Latin text-independent writer identification, including edge-hinge features,[21] GMF,[68] GLRM,[30] and CGD.[61] It can be found that the participant in the 2011-ICDAR-A competition[42] who only used simple features (directions and curvatures) and plain classifier (KNN) could still obtain relatively good performance (81.1%). The difficulty of 2014-ICFHR-A[103] comes from the very large number of writers (1000 writers). As a consequence, the performance (73.4%) of the winner was much lower than that of previous competitions. Note that the performance of 2015-ICDAR-M[28] fell off seriously when training set and test set belonged to different languages. On the other hand, most of the winners and runner-ups of the ICDAR and ICFHR competitions employ spatial distribution features. It indicates that spatial distribution features are superior to both frequency domain features and shape features in capturing the individual traits. Besides, the fusion of different features is capable of achieving better performance.

7. Discussion

Considerable attention has been paid to the research of off-line text-independent writer recognition and great progress has been achieved in the past years. However, it is still an open issue because the performance of the state of the art is still far from being satisfactory.

In general, most of writer recognition approaches assume that: (a) All images in the dataset usually share the same acquisition condition. (b) Handwriting of the same people is created by the same or very similar writing instruments and materials. But both of them are far away from the real situation. There are challenges in solving the impracticability of those assumptions. It is hard to guarantee that the reference images and query images have the same resolution in practical applications due to the variety of acquisition equipment and collecting condition. In our daily life, a writer could have various handwriting samples due to different writing instruments used. As a consequence, the width of strokes and size of characters from the same writer are alterable in different time, which makes the authentication of the writer more difficult. Every case should be taken into consideration when we attempt to design a practical writer recognition system.

Many writer recognition approaches have been proposed for particular languages (Arabic, Bengali, Chinese, English, French, Oriya, Telugu, Uyghur, etc.), and some of them declare they can deal with different kinds of languages. An interesting but challenging question comes forward: is it even possible to create a language-independent writer recognition system? The system using English samples for training and Greek samples for test already appeared. The competition[28] in 2015-ICDAR-M was organized to study this interesting scenario when training and test samples belong to different languages.

Acknowledgments

This work is partially supported by the Science and Technology Commission of Shanghai Municipality under Research Grant Nos. 14511105500 and 14DZ2260800.

References

1. M. N. Abdi and M. Khemakhem, A model-based approach to offline text-independent Arabic writer identification and verification, *Pattern Recognit.* **48**(5) (2015) 1890–1903.
2. M. N. Abdi, M. Khemakhem and H. Ben-Abdallah, A novel approach for off-line Arabic writer identification based on stroke feature combination, in *Proc. Int. Symp. on Computer and Information Sciences* (Cyprus, North Cyprus, September 14–16, 2009), pp. 597–600.
3. M. N. Abdi, M. Khemakhem and H. Ben-Abdallah, An effective combination of MPP contour-based features for off-line text-independent Arabic writer identification, in *Proc. Int. Conf. on Signal Processing, Image Processing and Pattern Recognition* (Jeju Island, South Korea, December 10–12, 2009), pp. 209–220.
4. M. N. Abdi, M. Khemakhem and H. Ben-Abdallah, Off-line text-independent Arabic writer identification using contour-based features, *Int. J. Signal Image Process.* **1**(1) (2010) 4–11.
5. A. Al-Dmour and R. A. Zitar, Arabic writer identification based on hybrid spectral-statistical measures, *J. Exp. Theor. Artif. Intell.* **19**(4) (2007) 307–332.
6. S. Al-Maadeed, W. Ayouby, A. Hassaine and J. M. Aljaam, QUWI: An Arabic and English handwriting dataset for offline writer identification, in *Proc. Int. Conf. on Frontiers in Handwriting Recognition* (Bari, Italy, September 18–20, 2012), pp. 746–751.
7. S. M. Awaida and S. A. Mahmoud, State of the art in off-line writer identification of handwritten text and survey of writer identification of Arabic text, *Educ. Res. Rev.* **7**(20) (2012) 445–463.

8. S. M. Awaida and S. A. Mahmoud, Writer identification of Arabic text using statistical and structural features, *Cybern. Syst.* **44**(1) (2013) 57–76.
9. H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, Speeded-up robust features (SURF), *Comput. Vis. Image Underst.* **110**(3) (2008) 346–359.
10. A. Bensefia, A. Nosary, T. Paquet and L. Heutte, Writer identification by writer's invariants, in *Proc. Int. Workshop on Frontiers in Handwriting Recognition* (Niagara-on-the-Lake, Canada, August 6–8, 2002), pp. 274–279.
11. A. Bensefia, T. Paquet and L. Heutte, Information retrieval based writer identification, in *Proc. Int. Conf. Document Analysis and Recognition* (Edinburgh, Britain, August 3–6, 2003), pp. 946–950.
12. A. Bensefia, T. Paquet and L. Heutte, A writer identification and verification system, *Pattern Recognit. Lett.* **26**(13) (2005) 2080–2092.
13. D. Bertolini, L. S. Oliveira, E. Justino and R. Sabourin, Texture-based descriptors for writer identification and verification, *Expert Syst. App.* **40**(6) (2013) 2069–2080.
14. K. M. Bin-Abdl and S. Z. M. Hashim, Handwriting identification: A direction review, in *Proc. Int. Conf. Signal and Image Processing Applications* (Kuala Lumpur, Malaysia, November 18–19, 2009), pp. 459–463.
15. A. Brink, M. L. Bulacu and L. Schomaker, How much handwritten text is needed for text-independent writer verification and identification, in *Proc. Int. Conf. Pattern Recognition* (Florida, United States, December 8–11, 2008), pp. 1–4.
16. M. L. Bulacu, Statistical pattern recognition for automatic writer identification and verification, Ph.D. thesis, University of Groningen, 2007.
17. M. L. Bulacu and L. Schomaker, A comparison of clustering methods for writer identification and verification, in *Proc. Int. Conf. on Document Analysis and Recognition* (Seoul, South Korea, August 29-September 1, 2005), pp. 1275–1279.
18. M. L. Bulacu and L. Schomaker, Automatic handwriting identification on medieval documents, in *Proc. Int. Conf. Image Analysis and Processing* (Modena, Italy, September 10–14, 2007), pp. 279–284.
19. M. L. Bulacu and L. Schomaker, Text-independent writer identification and verification using textural and allographic features, *IEEE Trans. on Pattern Anal. Mach. Intell.* **29**(4) (2007) 701–717.
20. M. L. Bulacu, L. Schomaker and A. Brink, Text-independent writer identification and verification on offline arabic handwriting, in *Proc. Int. Conf. of Document Analysis and Recognition* (Curitiba, Brazil, September 23–26, 2007), pp. 769–773.
21. M. L. Bulacu, L. Schomaker and L. Vuurpijl, Writer identification using edge-based directional features, in *Proc. Int. Conf. Document Analysis and Recognition* (Edinburgh, Britain, August 3–6, 2003), pp. 937–941.
22. S. Chanda, K. Franke and U. Pal, Text-independent writer identification for Oriya script, in *Proc. Int. Workshop on Document Analysis Systems* (Gold Coast, Australia, March 27–29, 2012), pp. 369–373.
23. S. Chanda, K. Franke, U. Pal and T. Wakabayashi, Text-independent writer identification for Bengali script, in *Proc. Int. Conf. on Pattern Recognition* (Istanbul, Turkey, August 23–26, 2010), pp. 2005–2008.
24. V. Christlein, D. Bernecker and E. Angelopoulou, Writer identification using VLAD encoded contour-zernike moments, in *Proc. Int. Conf. on Document Analysis and Recognition* (Nancy, France, August 23–26, 2015), pp. 906–910.
25. V. Christlein, D. Bernecker, F. Honig and E. Angelopoulou, Writer identification and verification using GMM supervectors, in *Proc. Winter Conf. on Applications of Computer Vision* (Steamboat Springs, United States, March 24–26, 2014), pp. 998–1005.

26. V. Christlein, D. Bernecker, A. Maier and E. Angelopoulou, Offline writer identification using convolutional neural network activation features, in *Proc. German Conf. on Pattern Recognition* (Aachen, Germany, October 7–10, 2015), pp. 540–552.
27. I. Dinstein and Y. Shapira, Ancient hebraic handwriting identification with run-length histograms, *IEEE Trans. Syst. Man Cybern.* **12**(3) (1982) 405–409.
28. C. Djeddi, S. Al-Maadeed, A. Gattal, I. Siddiqi, L. Souici-Meslati and H. El Abed, ICDAR2015 competition on multi-script writer identification and gender classification using QUWI database, in *Proc. Int. Conf. on Document Analysis and Recognition* (Nancy, France, August 23–26, 2015), pp. 1191–1195.
29. C. Djeddi, I. Siddiqi, L. Souici-Meslati and A. Ennaji, Text-independent writer recognition using multi-script handwritten texts, *Pattern Recognit. Lett.* **34**(10) (2013) 1196–1202.
30. C. Djeddi and L. Souici-Meslati, A texture based approach for Arabic writer identification and verification, in *Proc. Int. Conf. on Machine and Web Intelligence* (Algiers, Algeria, October 3–5, 2010), pp. 115–120.
31. C. Djeddi, L. Souici-Meslati and A. Ennaji, Writer recognition on Arabic handwritten documents, in *Proc. Int. Conf. on Image and Signal Processing* (Agadir, Morocco, June 28–30, 2012), pp. 493–501.
32. D. Fecker, A. Asit, V. Margner, J. El-Sana and T. Fingscheidt, Writer identification for historical Arabic documents, in *Proc. Int. Conf. on Pattern Recognition* (Stockholm, Sweden, August 24–28, 2014), pp. 3050–3055.
33. S. Fiel and R. Sablatnig, Writer retrieval and writer identification using local features, in *Proc. Int. Workshop on Document Analysis Systems* (Gold Coast, Australia, March 27–29, 2012), pp. 145–149.
34. S. Fiel and R. Sablatnig, Writer identification and writer retrieval using the fisher vector on visual vocabularies, in *Proc. Int. Conf. on Document Analysis and Recognition* (Washington, DC, United States, August 25–28, 2013), pp. 545–549.
35. S. Fiel and R. Sablatnig, Writer identification and retrieval using a convolutional neural network, in *Proc. Int. Conf. on Computer Analysis of Images and Patterns* (Valletta, Malta, September 2–4, 2015), pp. 26–37.
36. G. Ghiasi and R. Safabakhsh, An efficient method for off-line text-independent writer identification, in *Proc. Int. Conf. on Pattern Recognition* (Istanbul, Turkey, August 23–26, 2010), pp. 1245–1248.
37. G. Ghiasi and R. Safabakhsh, Offline text-independent writer identification using codebook and efficient code extraction methods, *Image Vis. Comput.* **31**(5) (2013) 379–391.
38. I. Guyon, L. Schomaker, R. Plamondon, M. Liberman and S. Janet, UNIPEN project of on-line data exchange and recognizer benchmarks, in *Proc. Int. Conf. on Pattern Recognition* (Jerusalem, Israel, October 9–13, 1994), pp. 29–33.
39. Y. Hannad, I. Siddiqi and M. E. Y. El Kettani, Arabic writer identification using local binary patterns (LBP) of handwritten fragments, in *Proc. Iberian Conf. on Pattern Recognition and Image Analysis* (Santiago de Compostela, Spain, June 17–19, 2015), pp. 237–244.
40. Y. Hannad, I. Siddiqi and M. E. Y. El Kettani, Writer identification using texture descriptors of handwritten fragments, *Expert Syst. Appl.* **47**(1) (2016) 14–22.
41. A. Hassaine and S. Al-Maadeed, ICFHR 2012 competition on writer identification challenge 2: Arabic scripts, in *Proc. Int. Conf. on Frontiers in Handwriting Recognition* (Bari, Italy, September 18–20, 2012), pp. 835–840.

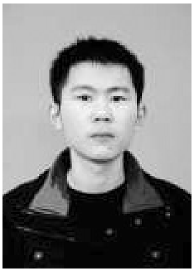
42. A. Hassaine, S. Al-Maadeed, J. M. Aljaam, A. Jaoua and A. Bouridane, ICDAR 2011 Arabic writer identification contest, in *Proc. Int. Conf. on Document Analysis and Recognition* (Beijing, China, September 18–21, 2011), pp. 1470–1474.
43. A. Hassaine, S. Al-Maadeed and A. Bouridane, A set of geometrical features for writer identification, in *Proc. Int. Conf. on Neural Information Processing* (Doha, Qatar, November 12–15, 2012), pp. 584–591.
44. Z. Y. He and Y. Y. Tang, A wavelet-based statistical method for Chinese writer identification, in *Applied Pattern Recognition*, H. Bunke, A. Kandel and M. Last (eds.), Studies in Computational Intelligence, Vol. 91 (Springer, NY, 2008), pp. 203–220.
45. Z. Y. He, B. Fang, J. W. Du, Y. Y. Tang and X. G. You, A novel method for offline handwriting-based writer identification, in *Proc. Int. Conf. on Document Analysis and Recognition* (Seoul, South Korea, August 29-September 1, 2005), pp. 242–246.
46. Z. Y. He and Y. Y. Tang, Chinese handwriting-based writer identification by texture analysis, in *Proc. Int. Conf. on Machine Learning and Cybernetics* (Shanghai, China, August 26–29, 2004), pp. 3488–3491.
47. S. He and L. Schomaker, Delta-n hinge: Rotation-invariant features for writer identification, in *Proc. Int. Conf. Pattern Recognition* (Stockholm, Sweden, August 24–28, 2014), pp. 2023–2028.
48. S. He, M. Wiering and L. Schomaker, Junction detection in handwritten documents and its application to writer identification, *Pattern Recognit.* **48**(12) (2015) 4036–4048.
49. Z. Y. He, X. G. You and Y. Y. Tang, Writer identification of Chinese handwriting documents using hidden markov tree model, *Pattern Recognit.* **41**(4) (2008) 1295–1307.
50. Z. Y. He, X. G. You and Y. Y. Tang, Writer identification using global wavelet-based features, *Neurocomputing* **71**(10) (2008) 1832–1841.
51. B. Helli and M. E. Moghaddam, A text-independent Persian writer identification system using LCS based classifier, in *Proc. Int. Symp. on Signal Processing and Information Technology* (Sarajevo, Bosnia and Herzegovina, December 16–19, 2008), pp. 203–206.
52. B. Helli and M. E. Moghadam, Persian writer identification using extended Gabor filter, in *Proc. Int. Conf. on Image Analysis and Recognition* (Povoa de Varzim, Portugal, June 25–27, 2008), pp. 579–586.
53. B. Helli and M. E. Moghaddam, A writer identification method based on xgabor and LCS, *IEICE Electron. Express* **6**(10) (2009) 623–629.
54. B. Helli and M. E. Moghaddam, A text-independent persian writer identification based on feature relation graph (FRG), *Pattern Recognit.* **43**(6) (2010) 2199–2209.
55. B. Helli and M. E. Moghaddam, An off-line text-independent persian writer identification method, *Int. J. on Artif. Intell. Tools* **20**(3) (2011) 489–509.
56. C. Hertel and H. Bunke, A set of novel features for writer identification, in *Proc. Int. Conf. Audio and Video-Based Biometric Person Authentication* (Guildford, Britain, June 9–11, 2003), pp. 679–687.
57. Y. J. Hu, W. M. Yang and Y. B. Chen, Bag of features approach for offline text-independent Chinese writer identification, in *Proc. Int. Conf. on Image Processing* (Paris, France, October 27–30, 2014), pp. 2609–2613.
58. A. Imdad, S. Bres, V. Eglin, H. Emptoz and C. Rivero-Moreno, Writer identification using steered hermite features and SVM, in *Proc. Int. Conf. on Document Analysis and Recognition* (Curitiba, Brazil, September 23–26, 2007), pp. 839–843.
59. D. Impedovo and G. Pirlo, Automatic signature verification: The state of the art, *IEEE Trans. Syst. Man Cybern. C, Appl. Rev.* **38**(5) (2008) 609–635.
60. R. Jain and D. Doermann, Offline writer identification using k-adjacent segments, in *Proc. Int. Conf. Document Analysis and Recognition* (Beijing, China, September 18–21, 2011), pp. 769–773.

61. R. Jain and D. Doermann, Writer identification using an alphabet of contour gradient descriptors, in *Proc. Int. Conf. on Document Analysis and Recognition* (Washington, DC, United States, August 25–28, 2013), pp. 550–554.
62. A. K. Jain, L. Hong and S. Pankanti, Biometric identification, *Commun. ACM* **43**(2) (2000) 90–98.
63. A. K. Jain, A. Ross and S. Pankanti, Biometrics: A tool for information security, *IEEE Trans. on Inf. Forensics Sec.* **1**(2) (2006) 125–143.
64. E. Khalifa, S. Al-maadeed, M. Tahir, A. Bouridane and A. Jamshed, Off-line writer identification using an ensemble of grapheme codebook features, *Pattern Recognit. Lett.* **59**(1) (2015) 18–25.
65. O. Kirli and M. B. Gulmezoglu, Automatic writer identification from text line images, *Int. J. Doc. Anal. Recognit.* **15**(2) (2012) 85–99.
66. F. Kleber, S. Fiel, M. Diem and R. Sablatnig, CVL-database: An on-line database for writer retrieval, writer identification and word spotting, in *Proc. Int. Conf. Document Analysis and Recognition* (Washington, DC, United States, August 25–28, 2013), pp. 560–564.
67. S. Kore and S. Apte, The current state of art: Handwriting a behavioral biometric for person identification and verification, in *Proc. Int. Conf. on Advances in Computing, Communications and Informatics* (Chennai, India, August 3–5, 2012), pp. 925–930.
68. X. Li and X. Q. Ding, Writer identification of Chinese handwriting using grid microstructure feature, in *Proc. Int. Conf. on Biometrics* (Alghero, Italy, June 2–5, 2009), pp. 1230–1239.
69. G. Louloudis, B. Gatos and N. Stamatopoulos, ICFHR 2012 competition on writer identification challenge 1: Latin/Greek documents, in *Proc. Int. Conf. on Frontiers in Handwriting Recognition* (Bari, Italy, September 18–20, 2012), pp. 829–834.
70. G. Louloudis, B. Gatos, N. Stamatopoulos and A. Papandreou, ICDAR 2013 competition on writer identification, in *Proc. Int. Conf. on Document Analysis and Recognition* (Washington, DC, United States, August 25–28, 2013), pp. 1397–1401.
71. G. Louloudis, N. Stamatopoulos and B. Gatos, ICDAR 2011 writer identification contest, in *Proc. Int. Conf. Document Analysis and Recognition* (Beijing, China, September 18–21, 2011), pp. 1475–1479.
72. D. G. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vis.* **60**(2) (2004) 91–110.
73. M. I. Malik, M. Liwicki, L. Alewijnse, W. Ohyama, M. Blumenstein and B. Found, ICDAR 2013 competitions on signature verification and writer identification for on- and off-line skilled forgeries (sigwicomp 2013), in *Proc. Int. Conf. on Document Analysis and Recognition* (Washington, DC, United States, August 25–28, 2013), pp. 1477–1483.
74. U. V. Marti and H. Bunke, The IAM-database: an English sentence database for off-line handwriting recognition, *Int. J. Document Analysis Recognition* **5**(1) (2002) 39–46.
75. U. V. Marti, R. Messerli and H. Bunke, Writer identification using text line based features, in *Proc. Int. Conf. on Document Analysis and Recognition* (Seattle, United States, September 10–13, 2001), pp. 101–105.
76. A. J. Newell and L. D. Griffin, Writer identification using oriented basic image features and the delta encoding, *Pattern Recognit.* **47**(6) (2014) 2255–2265.
77. A. Nicolaou, M. Liwicki and R. Ingolf, Oriented local binary patterns for writer identification, in *Proc. Int. Workshop on Automated Forensic Handwriting Analysis* (Washington DC, United States, August 22–23, 2013), pp. 15–20.
78. T. Ojala, M. Pietikainen and T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7) (2002) 971–987.

79. M. Pechwitz, S. S. Maddouri, V. Margner, N. Ellouze and H. Amiri, IFN/ENIT-database of handwritten Arabic words, in *Proc. Colloque Int. Francophone sur l'Ecrit et le Document* (Hammamet, Tunisia, October 21–23, 2002), pp. 127–136.
80. R. Plamondon and G. Lorette, Automatic signature verification and writer identification - the state of the art, *Pattern Recognit.* **22**(2) (1989) 107–131.
81. R. Plamondon and S. N. Srihari, Online and off-line handwriting recognition: A comprehensive survey, *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(1) (2000) 63–84.
82. A. Rafiee and H. Motavalli, Off-line writer recognition for Farsi text, in *Proc. Mexican Int. Conf. Artificial Intelligence Special Session* (Aguascalientes, Mexico, November 4–10, 2007), pp. 193–197.
83. S. S. Ram and M. E. Moghaddam, A Persian writer identification method based on gradient features and neural networks, in *Proc. Int. Congress on Image and Signal Processing* (Tianjin, China, October 17–19, 2009), pp. 1–4.
84. S. M. Saad, Application of fuzzy logic and genetic algorithm in biometric text-independent writer identification, *IET Inf. Sec.* **5**(1) (2011) 1–9.
85. R. Safabaksh, A. R. Ghanbarian and G. Ghiasi, HaFT: A handwritten Farsi text database, in *Proc. Iranian Conf. on Machine Vision and Image Processing* (Zanjan, Iran, September 10–12, 2013), pp. 89–94.
86. H. E. S. Said, T. N. Tan and K. D. Baker, Personal identification based on handwriting, *Pattern Recognit.* **33**(1) (2000) 149–160.
87. A. Schlapbach and H. Bunke, Off-line handwriting identification using HMM based recognizers, in *Proc. Int. Conf. on Pattern Recognition* (Cambridge, Britain, August 23–26, 2004), pp. 654–658.
88. A. Schlapbach and H. Bunke, Using HMM based recognizers for writer identification and verification, in *Proc. Int. Workshop on Frontiers in Handwriting Recognition* (Tokyo, Japan, October 26–29, 2004), pp. 167–172.
89. A. Schlapbach and H. Bunke, Off-line writer identification using gaussian mixture models, in *Proc. Int. Conf. on Pattern Recognition* (Hong Kong, China, August 20–24, 2006), pp. 992–995.
90. A. Schlapbach and H. Bunke, A writer identification and verification system using HMM based recognizers, *Pattern Anal. and Appl.* **10**(1) (2007) 33–43.
91. A. Schlapbach and H. Bunke, Off-line writer identification and verification using gaussian mixture models, in *Machine Learning in Document Analysis and Recognition*, S. Marinai and H. Fujisawa (eds.), Studies in Computational Intelligence, Vol. 90 (Springer, Berlin, 2008), pp. 409–428.
92. L. Schomaker, Advances in writer identification and verification, in *Proc. Int. Conf. on Document Analysis and Recognition* (Curitiba, Brazil, September 23–26, 2007), pp. 1268–1273.
93. L. Schomaker and M. L. Bulacu, Automatic writer identification using connected-component contours and edge-based features of uppercase western script, *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(6) (2004) 787–798.
94. L. Schomaker, M. L. Bulacu and K. Franke, Automatic writer identification using fragmented connected-component contours, in *Proc. Int. Workshop on Frontiers in Handwriting Recognition* (Tokyo, Japan, October 26–29, 2004), pp. 185–190.
95. L. Schomaker, K. Franke and M. L. Bulacu, Using codebooks of fragmented connected-component contours in forensic and historic writer identification, *Pattern Recognit. Lett.* **28**(6) (2007) 719–727.
96. F. Shahabi and M. Rahmati, A new method for writer identification and verification based on Farsi/Arabic handwritten texts, in *Proc. Int. Conf. on Document Analysis and Recognition* (Curitiba, Brazil, September 23–26, 2007), pp. 829–833.

97. F. Shahabi and M. Rahmati, A new method for writer identification of handwritten Farsi documents, in *Proc. Int. Conf. of Document Analysis and Recognition* (Barcelona, Spain, July 26–29, 2009), pp. 426–430.
98. M. K. Sharma and V. P. Dhaka, Offline scripting-free author identification based on speeded-up robust features, *Int. J. Doc. Anal. Recognit.* **18**(4) (2015) 303–316.
99. C. Shen, X. G. Ruan and T. L. Mao, Writer identification using Gabor wavelet, in *Proc. World Congress on Intelligent Control and Automation* (Shanghai, China, June 10–14, 2002), pp. 2061–2064.
100. I. Siddiqi and N. Vincent, Writer identification in handwritten documents, in *Proc. Int. Conf. on Document Analysis and Recognition* (Curitiba, Brazil, September 23–26, 2007), pp. 108–112.
101. I. Siddiqi and N. Vincent, Combining contour based orientation and curvature features for writer recognition, in *Proc. Int. Conf. on Computer Analysis of Images and Patterns* (Munster, Germany, September 2–4, 2009), pp. 245–252.
102. I. Siddiqi and N. Vincent, Text-independent writer recognition using redundant writing patterns with contour-based orientation and curvature features, *Pattern Recognit.* **43**(11) (2010) 3853–3865.
103. F. Slimane, S. Awaida, A. Mezghani, M. T. Parvez, S. Kanoun, S. A. Mahmoud and V. Mrgner, ICFHR 2014 competition on Arabic writer identification using AHTID/MW and KHATT databases, in *Proc. Int. Conf. on Frontiers in Handwriting Recognition* (Crete, Greece, September 1–4, 2014), pp. 797–802.
104. T. H. Su, T. W. Zhang and D. J. Guan, Corpus-based HIT-MW database for off-line recognition of general purpose Chinese handwritten text, *Int. J. Doc. Anal. Recognit.* **10**(1) (2007) 27–38.
105. M. Srebraj and S. Idicula, A novel approach to writer identification in Malayalam using graphemes, in *Proc. Int. Conf. on Information Processing* (Bangalore, India, August 5–7, 2011), pp. 646–651.
106. Y. B. Tang, W. Bu and X. Q. Wu, Text-independent writer identification using improved structural features, in *Proc. Chinese Conf. on Biometric Recognition* (Shenyang, China, November 7–9, 2014), pp. 404–411.
107. Y. B. Tang, X. Q. Wu and W. Bu, Off-line text-independent writer identification using stroke fragment and contour based features, in *Proc. Int. Conf. on Biometrics* (Madrid, Spain, June 4–7, 2013), pp. 1–6.
108. K. Ubul, A. Hamdulla, A. Aysa, A. Raxidin and R. Mahmut, Research on Uyghur off-line handwriting-based writer identification, in *Proc. Int. Conf. on Signal Processing* (Beijing, China, October 26–29, 2008), pp. 1656–1659.
109. L. Van Der Maaten and E. Postma, Improving automatic writer identification, in *Proc. Belgium-Netherlands Conf. on Artificial Intelligence* (Brussels, Belgium, October 17–18, 2005), pp. 260–266.
110. J. Wen, B. Fang, J. L. Chen, Y. Y. Tang and H. X. Chen, Fragmented edge structure coding for Chinese writer identification, *Neurocomputing* **86** (2012) 45–51.
111. T. S. Wilkinson, D. A. Mighell and J. W. Goodman, Backpropagation and its application to handwritten signature verification, in *Proc. Annual Conf. on Neural Information Processing Systems* (Denver, United States, November 27–30, 1989), pp. 340–347.
112. M. Wirotius, A. Seropian and N. Vincent, Writer identification from gray level distribution, in *Proc. Int. Conf. on Document Analysis and Recognition* (Edinburgh, Britain, August 3–6, 2003), pp. 1168–1168.
113. J. Woodard, M. Lancaster, A. Kundu, D. Ruiz and J. Ryan, Writer recognition of Arabic text by generative local features, in *Proc. Int. Conf. on Biometrics: Theory Applications and Systems* (Washington, DC, United States, September 27–29, 2010), pp. 1–7.

114. X. Q. Wu, Y. B. Tang and W. Bu, Off-line text-independent writer identification based on scale invariant feature transform, *IEEE Trans. Inf. Forensics Sec.* **9**(3) (2014) 526–536.
115. Y.-J. Xiong, Y. Wen, P. S. Wang and Y. Lu, Text-independent writer identification using SIFT descriptor and contour-directional feature, in *Proc. Int. Conf. on Document Analysis and Recognition* (Nancy, France, August 23–26, 2015), pp. 91–95.
116. L. Xu, X. Q. Ding, L. Peng and X. Li, An improved method based on weighted grid micro-structure feature for text-independent writer recognition, in *Proc. Int. Conf. on Document Analysis and Recognition* (Beijing, China, September 18–21, 2011), pp. 638–642.
117. W. X. Yang, L. W. Jin and M. F. Liu, Chinese character-level writer identification using path signature feature, dropstroke, and deep CNN, in *Proc. Int. Conf. on Document Analysis and Recognition* (Nancy, France, August 23–26, 2015), pp. 546–550.
118. J. J. Zhang, Z. Y. He, Y. M. Cheung and X. G. You, Writer identification using a hybrid method combining gabor wavelet and mesh fractal dimension, in *Proc. Int. Conf. on Intelligent Data Engineering and Automated Learning* (Burgos, Spain, September 23–26, 2009), pp. 535–542.
119. Y. Zhu, T. N. Tan and Y. H. Wang, Biometric personal identification based on handwriting, in *Proc. Int. Conf. on Pattern Recognition* (Barcelona, Spain, September 3–8, 2000), pp. 797–800.



Yu-Jie Xiong received his B.E. degree from Central South University, Changsha, China, in 2011. He is currently pursuing his Ph.D in Computer Science and Technology at East China Normal University, Shanghai, China. His research interests include pattern recognition, writer identification, and biometrics.

Yue Lu is a Professor in the Department of Computer Science and Technology, East China Normal University, and is presently serving as the Director of Shanghai Key Laboratory of Multidimensional Information Processing, and the Vice President of the Shanghai Research Institute of China Post. He received his B.S. degree in Wireless Technology and his M.S. degree in Telecommunications and Electronic System, both from Zhejiang University in 1990 and 1993 respectively, and his Ph.D. in Pattern Recognition and Intelligent System from Shanghai Jiao Tong University in 2000. From 1993 to 2000, he was an Engineer at the Third Research Institute of Posts and Telecommunications Ministry of China. Before he joined East China Normal University in 2004, he was a Research Fellow with the Department of Computer Science, National University of Singapore. In 2010, he was a Visiting Scientist at the Centre for Pattern Recognition and Machine Intelligence (CENPARMI), Concordia University of Canada for six months. His research interests include pattern recognition, image processing, natural language processing, biometrics, data mining and intelligent system development. Professor Lu has contributed to more than 100 reviewed publications in journals and conferences, and holds 11 authorized invention patents. He is serving as an editorial board member of Pattern Recognition and associate editor of International Journal of Pattern Recognition and Artificial Intelligence, ACM Transactions on Asian and Low-Resource Language Information Processing.

Patrick S. P. Wang Fellow of IAPR, ISIBM and WASE, is a Professor of Computer and Information Science at Northeastern University, USA, Shanghai East China Normal University ZiJiang Visiting Chair Professor, NSF Visiting Chair Professor at NTUST, Taipei, Taiwan, research consultant at MIT Sloan School, and adjunct faculty of computer science at Harvard University. He received his Ph.D. in Computer Science from Oregon State University, M.S. degree in Information and Computer Science from the Georgia Institute of Technology, M.S. degree in Electrical Engineering from the National Taiwan University and B.S. degree in Electronic Engineering from the National Chiao Tung University (Hsin-chu campus, Taiwan). As IEEE and ISIBM Distinguished Achievement Awardee, Professor Wang was on the Faculty at the University of Oregon and Boston University, and Senior Researcher at Southern Bell, GTE Labs and Wang Labs. Professor Wang was Otto-Von-Guericke Distinguished Guest Professor of Magdeburg University, Germany, and iCORE (Informatics Circle of Research Excellence) Visiting Professor at the University of Calgary, Canada, Honorary Advisor Professor for Sichuan University, Chongqing University, Xiamen University, and Guangxi Normal University, Guilin, Guangxi in China. In addition to his research experience at MIT AI Lab, Professor Wang has been Visiting Professor and has been invited to give lectures, do research and present papers in a number of countries in Europe, Asia and at many universities and industries in the United States and Canada. Professor Wang has published over 200 technical papers and 26 books in Pattern Recognition, A.I. Biometrics and Imaging Technologies and has 3 OCR patents awarded by the US and Europe Patent Bureaus. One of his books is so important and is so widely cited that the United States Department of Homeland Security (DHS) used it as reference for Call For Proposals 2010. For details, please refer to DHS website: Image Pattern Recognition C Synthesis and Analysis of Biometrics (WSP). As IEEE senior member, he has organized numerous international conferences and workshops including conference co-chair of the 18th IAPR ICPR (International Conference on Pattern Recognition) in 2006, Hong Kong, China, and served as reviewer for many journals and NSF grant proposals. Professor Wang is currently founding Editor-in-Chief of IJPRAI (Int. J. of Pattern Recognition and A.I.), and Machine Perception and Artificial Intelligence Book Series by World Scientific Publishing Co. and Imperial College Press, London, UK, and elected chair of IAPR-SSPR (Int. Assoc. for P.R.). Professor Wang has been invited to give talks at many international conferences, including AIA2007, Innsbruck, Austria, IAS2007, Manchester, UK, IEEE-SMC2007, 2009, 2010, Montreal, San Antonio, Istanbul respectively, WorldComp2010, Las Vegas, USA, CIS2007, Harbin, China, eForensics2008, Adelaide, Australia, ISI2008, Taipei, Taiwan, BroadCom2008, Pretoria, South Africa, VISAPP2009, Lisboa, Portugal, UKSim2011, Cambridge, UK, and IADIS2010, 2011, Freiburg, Germany, and Roma, Italy, respectively. Professor Wang received the IEEE Distinguished Achievement Award at IEEE-BIBE2007 at Harvard Medical, for Outstanding Contributions in Bioinformatics and Bioengineering. In addition to his technical achievements and contributions, Professor Wang has also been very active in community services, and has written several articles on Du Fu, Li Bai's poems, Verdi, Puccini, Bizet, and Wagner's operas, and Mozart, Beethoven, Schubert and Tchaikovsky's symphonies. A collection of selected proses was published in his book Harvard Meditation Melody by Jian-Shing Pub. Co., Taipei, Taiwan, which won best publication award by Taiwan. https://sites.google.com/site/mozart200, patwang@ieee.org, pwang@acm.org.