Educational Research and Reviews Vol. 7(20), pp. 445-463, 25 July, 2012
Available online at http://www.academicjournals.org/ERR
DOI: 10.5897/ERR11.303
ISSN 1990-3839 ©2012 Academic Journals

Full Length Research Paper

State of the art in off-line writer identification of handwritten text and survey of writer identification of Arabic text

Sameh M. Awaida[1]* and Sabri A. Mahmoud[2]

[1]Qassim University, Qassim 51452, Saudi Arabia.
[2]King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia.

Accepted 6 July, 2012

In this paper we present the state of the art in writer identification and verification of handwritten text. In addition, a special and extensive survey of writer identification and verification of Arabic handwritten text is also included. Different feature extraction techniques are addressed showing the different research groups' efforts as well as individual efforts. The different classification approaches, e.g. minimum distance classifiers and statistical classifiers, used for identification by writer and verification by different groups and individuals are presented. Identification results of surveyed publications are investigated and tabulated for ease of reference. Examples of writer identification and verification of others languages are addressed. An extensive survey of databases used in writer identification and verification for Latin and Arabic text is presented. Conclusions relevant to writer identification of Arabic text are discussed and future directions stated.

Key words: Writer identification, writer verification, handwritten database, feature extraction, handwriting analysis, distance measures, literature survey.

INTRODUCTION

This paper presents the state of the art in writer identification and verification of handwritten text with a special survey on writer identification and verification of Arabic handwritten text. For advances in the field prior to the year 1990, reference can be made to the study of Plamondon and Lorette (1989). Due to technique similarities and inherent connection, signature verification and handwriting recognition surveys; the state of the art in writer identification and verification are discuss as well (Plamondon, 1994; Plamondon and Srihari, 2000). However, there is lack of literature surveys that specifically target writer identification and verification.

Writer identification is the process of determining from a set of possible writers, an author through samples of his/her handwriting (Schlapbach, 2007). Writer verification is the process of comparing questioned handwriting with samples of handwriting obtained from known sources for the purposes of determining authorship or non-authorship (Bradford, 1992). Writer verification involves accept/reject decision-making criteria whilst writer identification involves a one-to-many classification problem and hence is considered more challenging (Gibbons et al., 2005; Zaher and Abu-Rezq, 2010). In recent years, writer identification and verification has become a common application used in confirming the document authenticity in the financial district as well as revealing the identity of suspected criminals, etc. In May 13, 1999, the United States vs. Paul decided that handwritten analysis qualifies as expert testimony and is therefore admissible (Srihari et al., 2002).

Over the past two decades, automatic offline writer identification has enjoyed renewed interest. One of the

*Corresponding author. E-mail: s.awaida@qu.edu.sa. Tel: +966-5-48242577.

driving forces for this surge is the increasing need for writer identification techniques by forensic document examiners to identify criminals based on their handwriting (Zhang et al., 2004). Furthermore, threats of terrorist attacks have increased the use of writer identification and other biometric recognition techniques to identify the assailants (Schlapbach, 2007).

One of the main applications of writer identification and verification is its use in forensic sciences (Franke et al., 2003; Franke and Koppen, 2001; Niels et al., 2007; Srihari et al., 2002; Zhang et al., 2004). The identification of a person on the basis of an arbitrary handwritten sample is a useful application. Writer identification allows for determining the suspects in conjunction with the inherent characteristic of a crime, e.g. the case of threat letters. This is different than other biometric methods, where the relation between the evidence material and the details of an offense can be quite remote (Schomaker and Bulacu, 2004). In addition to forensic applications of writer identification and verification, several other applications exist. Some examples include ink type recognition (Franke et al., 2002), script and language identification (Hochberg et al., 1999), forgery detection (Leedham and Chachra, 2003), writer identification on medieval and historical documents (Bar-Yosef et al., 2007; Bulacu and Schomaker, 2007a; Panagopoulos et al., 2009; Schomaker et al., 2007), writer identification on handwritten musical scores (Fornes et al., 2008), and personalized handwriting text recognizers (Rodríguez-Serrano et al., 2010).

Writer identification can be divided into two categories; text-dependent and text-independent writer identification. Text-dependent writer identification systems require certain known text to be written, whereas text-independent writer identification systems can work on any given text. In this work, research involving text-dependent and text-independent writer identification of offline handwritten text is surveyed.

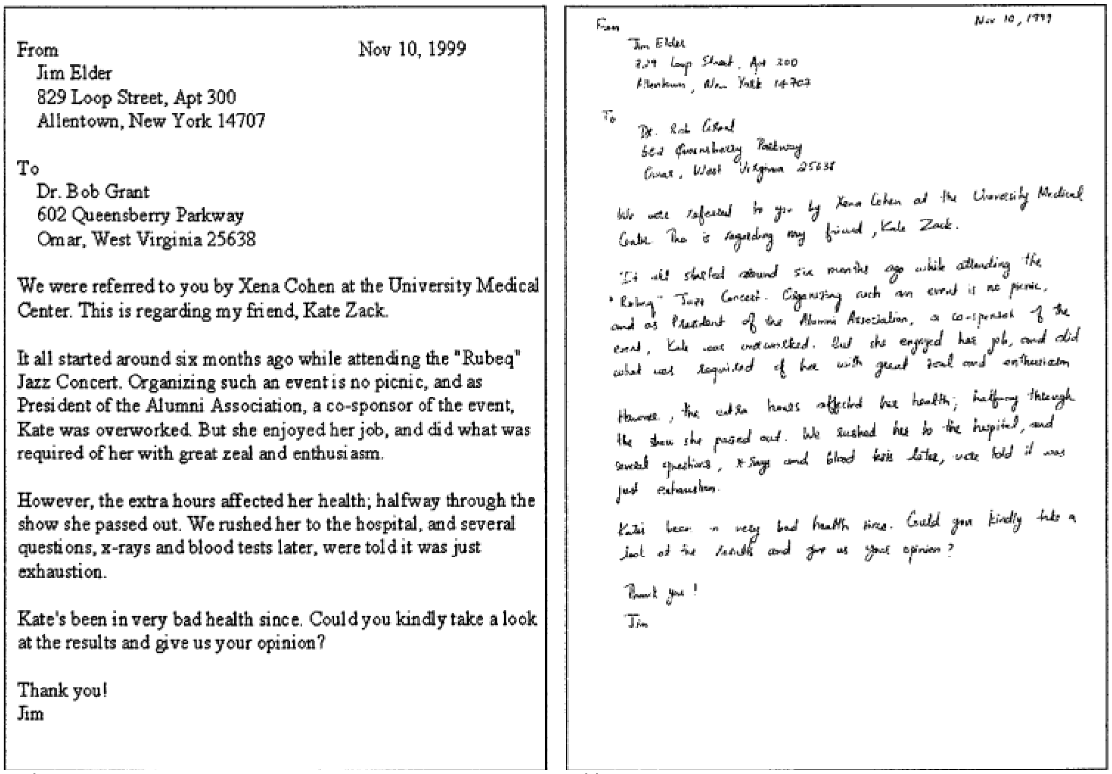
We have included almost 100 accessible and published publications on the field of writer identification and verification. However, we cannot claim that we have addressed all published work for writer identification and verification of Latin or other languages. We tried our best to include the work of all the major research groups and individuals in the field. In surveying writer identification and verification of Arabic text, we included all the papers we had access to and also incorporated research on Persian (Farsi) text for its similarity to the Arabic script. The same claim can be stated about the included databases for writer identification and verification.

Although, research in writer identification and verification is still predominantly aimed for the English language, research of other languages includes Chinese (Cong et al., 2002; He et al., 2005; 2008a; b; He and Tang, 2004; Li and Ding, 2009; Li et al., 2006; Liu et al., 1995; Su et al., 2007; Wang et al., 2003; Zhu et al., 2000), Dutch (Brink et al., 2010; Maaten and Postma, 2005; Schomaker and Bulacu, 2004), Greek (Zois and Anastassopoulos, 2000), French (Bensefia et al., 2002; 2003a; 2003b; 2004; 2005a; Siddiqi and Vincent, 2009), Japanese (Yoshimura, 1988), Uyghur (Ubul et al., 2008), Myanmar (Mar and Thein, 2005), Arabic (Abdi et al., 2009; Al-Dmour and Zitar, 2007; Al-Ma'adeed et al., 2008; Bulacu et al., 2007; Gazzah and Ben, 2006; 2007; 2008; Srihari and Ball, 2008), Persian (Helli and Moghaddam, 2008a; b; 2009, 2010; Ram and Moghaddam, 2009a; b; Shahabi and Rahmati, 2006, 2007), numerals (Leedham and Chachra, 2003), as well as historical manuscripts and inscriptions in different ancient languages (Bar-Yosef et al., 2007; Bensefia et al., 2003b; Bulacu and Schomaker, 2007a; Panagopoulos et al., 2009; Schomaker et al., 2007).

DATABASES FOR WRITER IDENTIFICATION AND VERIFICATION OF WESTERN SCRIPT

In this section the main databases used for writer identification and verification of handwritten Latin and other western scripts are addressed. The CEDAR letter was developed in the University of Buffalo (Cha and Srihari, 2000), and is considered one of the first large databases developed for writer identification and verification of handwritten Latin scripts. The CEDAR Letter, as shown in Figure 1, is concise (it has just 156 words) yet still each alphabet occurs in the beginning of a word as a capital and as a small letter in the middle and end of a word. In addition, it also contains punctuation, numerals, and some letter and numeral combinations (for example, ff, tt, oo, 00). The CEDAR letter was written by 1 000 individuals three times each. Noticeably, (Srihari et al., 2002) reported that the CEDAR letter was written by 1,500 writers.

The IAM-database (Marti and Bunke, 2002) consists of handwritten English sentences that are based on the Lancaster-Oslo/Bergen (LOB) corpus (Johansson et al., 1978). The corpus is a collection of texts that comprise about one million word instances. The database originally included 1 066 forms produced by approximately 400 different writers, and was later extended to include 1 539 forms produced by 657 different writers. The database consists of full English sentences. Figure 2 shows a sample filled form of the IAM database. Due to its public availability, flexible structure, and large number of writers involved, the IAM database has been commonly used for Latin writer identification/verification by a number of researchers (Bensefia et al., 2005a, 2005b; Brink et al., 2008; Bulacu, 2007; Bulacu and Schomaker, 2006, 2007b; Helli and Moghaddam, 2009; Schlapbach and Bunke, 2004a; b; 2007; Schlapbach et al., 2005; Schomaker and Bulacu, 2004; Siddiqi and Vincent, 2007, 2008, 2009).Researchers have used the IAM database alone(Brink et al., 2008; Schlapbach and Bunke, 2007; Siddiqi and Vincent, 2008) or combined/compared it with



a)

From Nov 10, 1999
Jim Elder
829 Loop Street, Apt 300
Allentown, New York 14707

To
Dr. Bob Grant
602 Queensberry Parkway
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

It all started around six months ago while attending the "Rubeq" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

Kate's been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!
Jim

b)

From Nov 10, 1999
Jim Elder
829 Loop Street, Apt 300
Allentown, New York 14707

To
Dr. Bob Grant
602 Queensberry Parkway
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

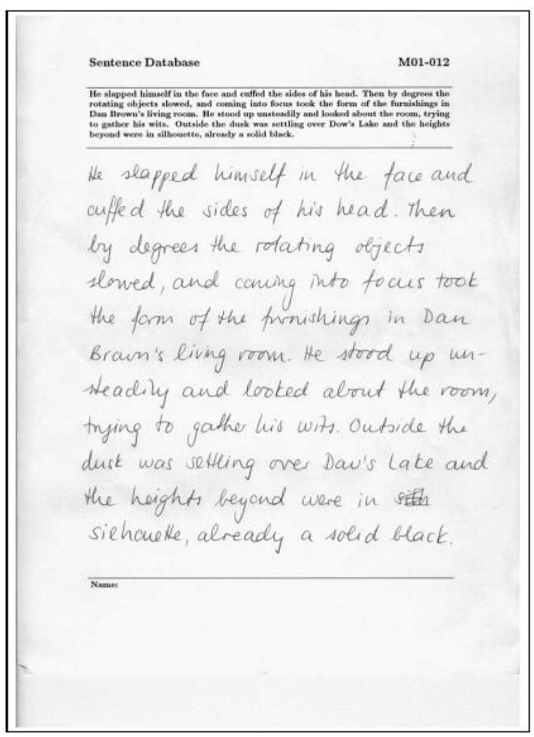
It all started around six months ago while attending the "Rubeq" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

Kate's been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!
Jim

Figure 1. CEDAR letter a) source document, b) scanned sample (Srihari et al., 2002).



Sentence Database **M01-012**

He slapped himself in the face and cuffed the sides of his head. Then by degrees the rotating objects slowed, and coming into focus took the form of the furnishings in Dan Brown's living room. He stood up unsteadily and looked about the room, trying to gather his wits. Outside the dusk was settling over Dow's Lake and the heights beyond were in silhouette, already a solid black.

He slapped himself in the face and
cuffed the sides of his head. Then
by degrees the rotating objects
slowed, and coming into focus took
the form of the furnishings in Dan
Brown's living room. He stood up un-
steadily and looked about the room,
trying to gather his wits. Outside the
dusk was settling over Dow's Lake and
the heights beyond were in ~~sith~~
silhouette, already a solid black.

Name:

Figure 2. A sample IAM filled form (Marti and Bunke, 2002).

other databases (Bulacu, 2007; Bulacu and Schomaker, 2006, 2007b; Schomaker and Bulacu, 2004; Siddiqi and Vincent, 2009).

The Firemaker dataset (Schomaker and Vuurpijl, 2000) consists of 1 008 scanned pages of handwritten Dutch texts written by 252 students, four pages each. Page 1 contains a copied text in natural writing style; Page 2 contains copied upper-case text; Page 3 contains copied forged text while Page 4 contains a self-generated description of a cartoon image in free writing style. The text to be copied has been designed to cover a sufficient amount of different letters from the alphabet while it still conveniently remains writable for the majority of writers. Figure 3 shows an example of Page 2. Since the Firemaker database was not publicly available for some time, it has been mostly used by the researchers in the University of Groningen (Brink et al., 2008; Bulacu, 2007; Bulacu and Schomaker, 2005, 2006, 2007b; Bulacu et al., 2003; Schomaker et al., 2004; Schomaker et al., 2003; Schomaker et al., 2007) with few exceptions (Maaten and Postma, 2005). Lately, the Firemaker database has been publicly available (Int. Unipen Foundation, 2011). It should be noted that Schomaker et al. have combined parts of the Firemaker database with parts of the IAM database to make a western script database of 900 writers (Brink et al., 2008; Bulacu, 2007; Bulacu and Schomaker, 2006, 2007b; Schomaker and Bulacu, 2004). Table 2 shows the databases used in writer identification and verification of handwritten text, the number of writers of each database, the language of the text, and published research work in which these databases are used.

Other public western handwritten databases used in writer identification/verification include the UNIPEN dataset, the Trigraph slant dataset, the HIFCD2 dataset, IRONOFF dataset, and the RIMES dataset. A brief description for each database follows next.

The UNIPEN project (Guyon et al., 1994) described a format and methodology for creating a database for online handwritten text from several countries and languages, and has organized the collection of more than 5 million handwritten characters of more than 2 200 writers. Offline images has been derived from the UNIPEN online database and has been used in writer identification (Bulacu, 2007; Bulacu and Schomaker, 2005, 2006, 2007b; Niels et al., 2007; Schomaker et al., 2004, 2007). The TriGraph Slant Dataset is a recent database that contains images for 47 writers of handwriting, produced under conditions of normal and disguised slant (Brink et al., 2010). The HIFCD2 database contains handwritten samples for the word ‘characteristic’ and its equivalent Greek word written 45 times for each writer, for 50 total writers (Zois and Anastassopoulos, 2000). The IRESTE On/Off (IRONOFF) dual handwriting database (Viard-Gaudin et al., 1999) contains French letters and words for 700 writers. It is dual in the sense that it contains both online data (pen trajectory) and offline data (digital images) for the same writing. The RIMES French database contains more than 5 600 real mails written by 1 300 writers completely annotated, as well as, secondary databases of isolated characters, handwritten words (300 000 snippets) and logos (Grosicki et al., 2008). Figure 4 shows samples of the UNIPEN, TriGraph, HIFCD2, IRONOFF, and RIMES databases, respectively. As mentioned previously, all of these databases are available publicly for research purposes.

FEATURE EXTRACTION APPROACHES

Researchers used different types of features for writer identification. Some of these features are also used in automatic handwritten text recognition. This section presents the types of features that have been used in writer identification and verification. Features used by groups of researchers in writer identification and verification will be presented in conjunction followed by other researchers’ work. Categorizing features by research group allows the reader to see the combination of features in their appropriate scope. It also indicates how these features are developed over time and the different applications or used data of these features.

Bensefia (Bensefia et al., 2002; Bensefia et al., 2003a, 2005a) used graphemes that are generated by segmenting handwritten text into graphemes to identify writers. These graphemes are then clustered using sequential clustering algorithm. Clustering is repeated and graphemes that fall in the same clusters in these repeated clustering are kept in these clusters. Graphemes that change clusters are kept in separate clusters. First-level graphemes, bi-grams and tri-grams are used. Bi-grams and tri-grams of graphemes are connected and features extracted. This technique is applied to two datasets containing different number of writers; a self-built database of 88 writers and 150 writers of the IAM database (Marti and Bunke, 2002). Recognition rates on their own database of 93, 95.45 and 80% were achieved using first-level graphemes, bi-grams, and tri-grams respectively.

Schomaker et al. used two level analysis for feature extraction; texture level and character-shape (allograph) level (Brink et al., 2008, 2010; Bulacu, 2007; Bulacu and Schomaker, 2005, 2006, 2007a, 2007b; Bulacu et al., 2003,2007;; Franke et al., 2003; Niels et al., 2007; Schomaker and Bulacu, 2004; Schomaker et al., 2004, 2003, 2007). At the texture level, they used contour-direction Probability Distribution Function(PDF) ($p(\phi)$, where ϕ is the contour direction as shown in Figure 5(a), contour-hinge PDF (($p(\phi_1, \phi_2)$, where ϕ_1, ϕ_2 are the angles of the two sides of the hinge as shown in Figure 5 (b)), direction co-occurrence ($p(\phi_1, \phi_3)$, where ϕ_1, ϕ_3 are the angles with the horizontal- and vertical-run, as shown in Figure 5 (c)), the probability distribution of the white run

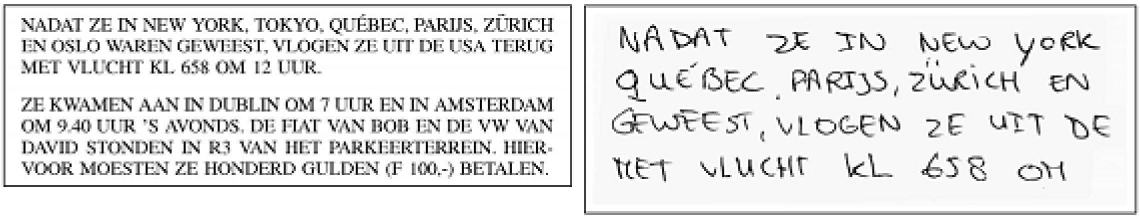


Figure 3. Firemaker page 2 (source document and scanned sample (Schomaker and Bulacu, 2004)).

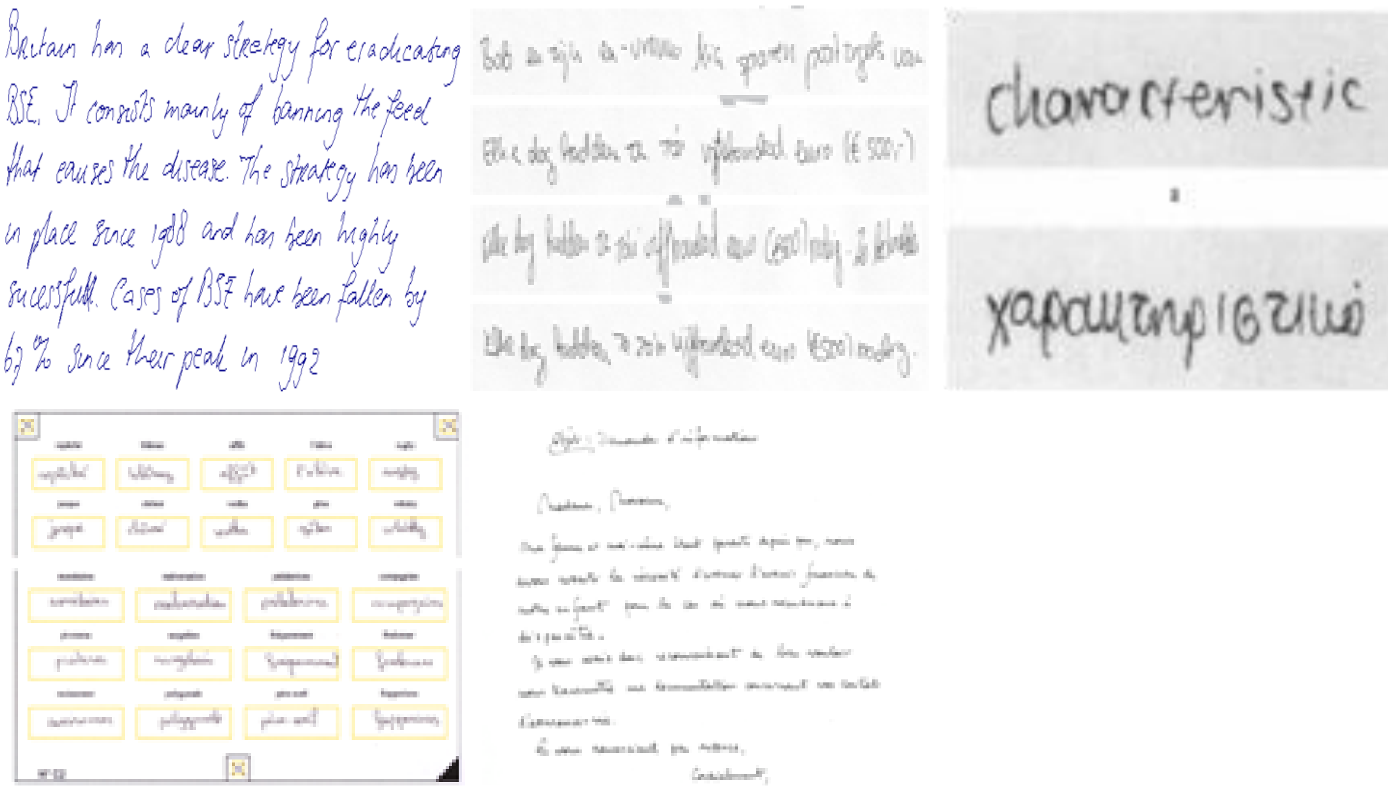


Figure 4. (Left to right) samples of the UNIPEN, TriGraph, HIFCD2, IRONOFF, and RIMES DBs.

lengths PDFs, and autocorrelation in horizontal scan. The contour-direction PDF features are assumed to capture orientation and curvature information, the contour-hinge PDF to capture the curvature of the contour, and the direction co-occurrence to measure the roundness of the written characters.

At the allograph level, graphemes were used. These features were initially applied to uppercase letters with success (Schomaker and Bulacu, 2004), and were later applied to cursive text. The handwriting was segmented at the minima in the lower contour with the distance to the upper contour close to the writing line thickness. The graphemes were extracted as connected components. For each connected component, its contour was computed using Moore's algorithm (Gonzalez and Woods, 2007).Inner contours were discarded. The PDF of these connected components (graphemes) was computed using a common codebook obtained by clustering the graphemes of the data. Figure 5(d) shows an illustration of the used graphemes. K-means and Kohonen self-organization feature maps (Kohonen, 1989) were used to generate the code book.

In their research work, Shomaker et al. addressed both text-dependent and text-independent approaches for writer identification. They have concluded that text-dependent approaches achieve high performance even with small amounts of data. However, this has limited applicability due to the need of specific text and human intervention (Bulacu and Schomaker, 2007b). It is worth adding that having a successful text-independent writer

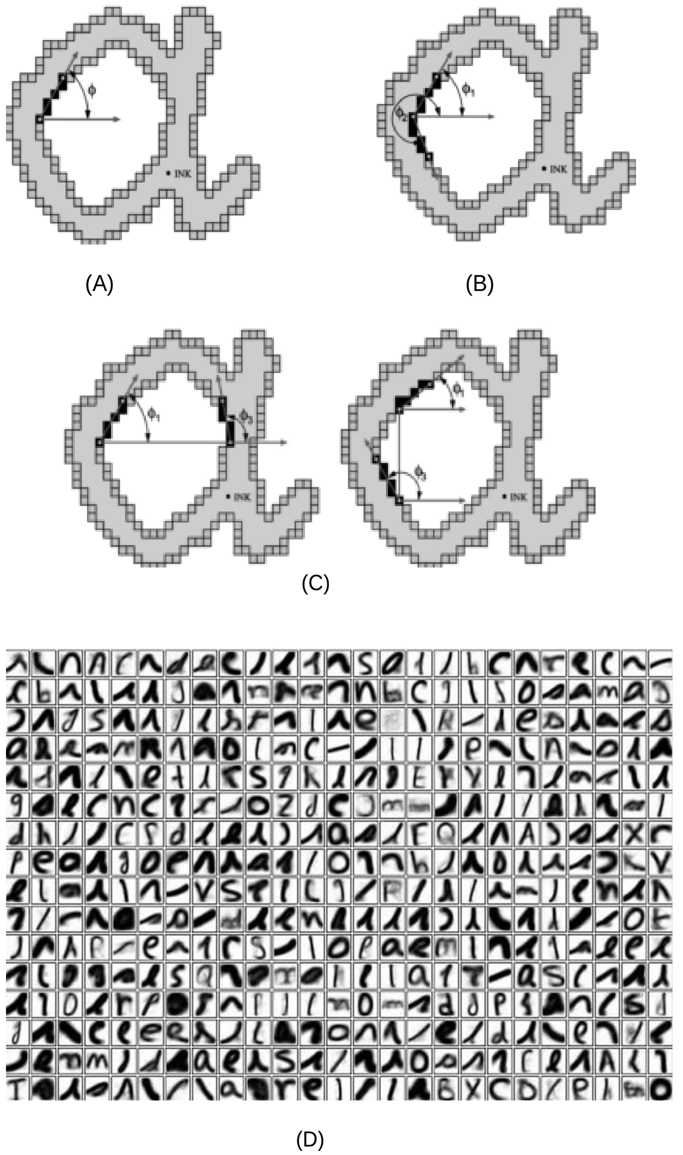


Figure 5. Texture; (a) Contour-direction ϕ. (b) Contour-hinge (ϕ_1, ϕ_2). (c) Direction co-occurrence (horizontal and vertical scans) (d) Shape code-book samples.

identification system can correspondingly operate on dependent-texts without any major modifications to the system, and not vice versa. Schlapbach et al. used features that are normally used for text recognition (Schlapbach, 2007; Schlapbach and Bunke, 2004a, 2004b, 2006, 2007; Schlapbach et al., 2005). In one of their research works (Schlapbach and Bunke, 2004b), they used Hidden Markov Models (HMM) for writer identification and verification by recognizing a text line, using a number of HMMs. They determined the identity of the writer by choosing the HMM of the writer that provided the best confidence measure of the recognized text line. As each HMM was trained with the data of one writer, the HMM that produced higher confidence measure for the text line identified the writer.

For feature extraction, Schlapbach et al. used a sliding window which is commonly practice with HMM classifiers. A window of one pixel wide is shifted from left to right over the line of text. At each position, nine geometrical features are extracted; three global features and six local features. Global features represent the number of black pixels in the window, the center of gravity and the second order moment of the black pixels. The six remaining local features are the position and contour direction of the upper and lower-most pixels, the number of black-to-white transitions in the window, and the fraction of pixels between the upper and lower-most black pixels.

Srihari et al. used statistical features that are extracted at different levels of resolution (Srihari, 2000; Srihari and Ball, 2008; 2009; Srihari et al., 2005; Srihari et al., 2002; Srihariet al., 2007; Srihariet al., 2007; Tomai and Srihari, 2004; Zhang, 2003; Zhang et al, 2003). At the macro level, thirteen global features are extracted, that is to say, measures of pen pressure (entropy of gray values, gray-value threshold, number of black pixels); measures of writing movement (number of interior contours, number of exterior curves); measures of stroke formation (number of vertical, horizontal, positive, and negative strokes); average line height and average slant per line; stroke width, and average word gap.

At the micro level, Gradient, Structural and Concavity features (GSC) are extracted. First, the image is divided $n \times m$ grids with equal number of foreground pixels for each of n rows, and equal number of foreground pixels for each of m columns. Then for each grid cell, the GSC features column vector is extracted. The gradient features are computed by convolving two 3×3 Sobel operators with the binary image. These operators approximate the x and y derivatives in the image data pixel position. The vector addition of the operators' output is used to compute the gradient of the image. Since the gradient is vector valued with magnitude and direction, only the direction is used in the computation of a feature vector, which is stored in a gradient feature map. A histogram of gradient directions is taken at each pixel of the region, where each histogram value corresponds to the count of each gradient direction in the region.

The structural features capture certain patterns embedded in the gradient direction map. These patterns are "mini-strokes" of the image. A set of 12 rules are applied to each pixel. These rules operate on the eight nearest neighbours of the pixel. Each rule examines a particular pattern of the neighbouring pixels for allowed gradient ranges. For example, rule S_1 states that if neighbour (N_0) and neighbour (N_4) of a pixel both have a gradient range of 61° to 150°, then the rule is satisfied and its corresponding value in the feature vector is incremented by 1.The concavity features are the coarsest of the GSC set. They can be broken down into three sub-

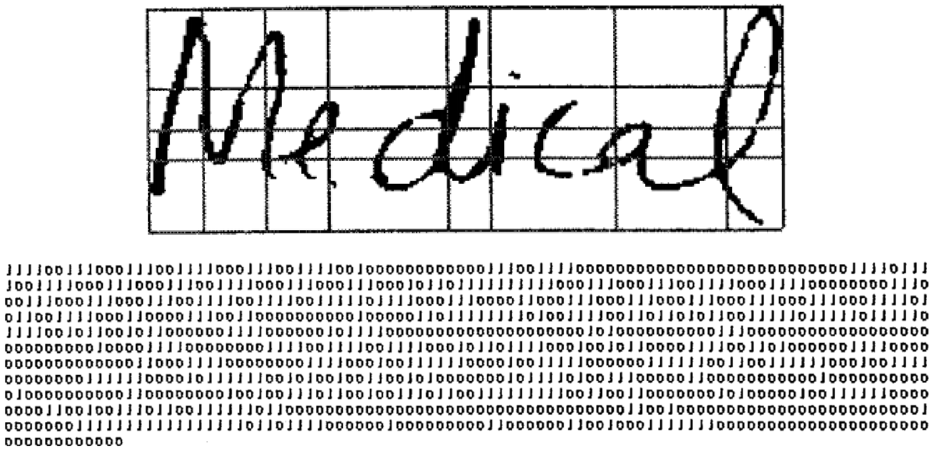


Figure 6. Exemplar word image with 4 × 8 divisions using GSC (Zhang, 2003).

classes of features: segment density, large strokes, and concavity shape. The full list of rules for the GSC features is shown in Table 1. Figure 6 shows an example of the GSC features vector for the word “Medical” for *4* × *8* grid divisions.

Siddiqi and Vincent (2007; 2008; 2009) divided each image into a large number of small sub-images using a window, and clustered these sub-images. They used these clusters as features. They also extracted the histograms of the chain code, the first and second order differential chain codes, and the histogram of the curvature indices at each point of the contour of handwriting. Leedham and Chachra, (2003) used a combination of local and global features These included pixel density, fixed point distance and angular measure, center of gravity, gradient features, height to width ratio, number of end-points, number of junctions, number of loops, and degree of slant.

Ram and Moghaddam (2009a; b) used gradient features, grapheme features; connected components contours, area features, and a collection of local features. Said et al. (1998) used grey scale co-occurrence matrices. Franke et al. (2002) used co-occurrence features like energy, correlation, inverse difference moment, and entropy. Bar-Yosef used the ratio between the area of each dominant background set and the convex hull, and the aspect ratio of the enclosing ellipse (Bar-Yosef et al., 2007). Mar and Thein (2005) used mean and standard deviation of Region of Interests (ROIs). Cha (2001) used sliding windows to extract both local and global features. Wang et al. (2003) used distribution of directional elements (gradient). Liu et al. (1995) used features derived from 2nd and 3rd order moments. Zois and Anastassopoulos (2000) used erosion and dilation function on the horizontal projection.

Researchers have also used image transformations as features. For example, Gabor filters were used in Al-Dmour and Zitar (2007), Cong et al. (2002), He and Tang (2004), Helli and Moghaddam (2008a; 2008b; 2009; 2010), Liu et al. (1995), Said et al. (1998), Shahabi and Rahmati (2006; 2007), Siddiqi and Vincent (2008), Ubul et al. (2008) Zhu et al. (2000), wavelet transforms in Gazzah and Ben (2006; 2007; 2008), He et al. (2005; 2008a; b) and contourlet transformations in He et al. (2005).

It is worth noting that some of the same successful feature extraction techniques have been used by different research groups. For example, taking the histogram of the pixel angle was originally applied for writer identification by both Srihari (2000) and Schomaker et al. (2003), and since then was used by their own research groups as shown previously and by other researchers (Al-Ma’adeed et al., 2008; Leedham and Chachra, 2003; Li and Ding, 2009; Ram and Moghaddam, 2009a, b; Wang et al., 2003).

Measuring slant (at least at the pixel level) using gradient distributions, although there have been experimental results that question the effect of slant on writer identification/verification (Brink et al., 2010). Using parts of letters (graphemes) was originally applied by Benesefia et al. (2002), and since then has been implemented by different researchers as well (Al-Ma’adeed et al., 2008; Bulacu, 2007; Bulacu and Schomaker, 2006; Leedham and Chachra, 2003; Ram and Moghaddam, 2009b; Schomaker and Bulacu, 2004; Schomaker et al., 2004).

Table 3 details the published work of writer identification and verification including used features, classifiers and best reported top-1 accuracy results. Some resear-chers tried their writer identification system on multiple databases, and hence more than one accuracy result is reported per publication. For more information about the used databases, readers are referred to Table 2.

Table 1. GSC feature definitions.

Gradient		Structural				Concavity	
ID	Angle	ID	Description	Neighbour 1 (Range)	Neighbour 2 (Range)	ID	Description
G_1	1° to 30°	S_1	Horizontal line (a)	N_0 (61° to 150°)	N_4 (61° to 150°)	CD	Pixel density
G_2	31° to 60°	S_2	Horizontal line (b)	N_0 (241° to 330°)	N_4 (241° to 330°)	CHRL	Horizontal run length
G_3	61° to 90°	S_3	Vertical line (a)	N_2 (151° to 240°)	N_6 (151° to 240°)	CVRL	Vertical run length
G_4	91° to 120°	S_4	Vertical line (b)	N_2 (-29° to 60°)	N_6 (-29° to 60°)	CCH	Hole concavity
G_5	121° to 150°	S_5	Diagonal rising (a)	N_5 (121° to 210°)	N_1 (121° to 210°)	CCU	Upward concavity
G_6	151° to 180°	S_6	Diagonal rising (b)	N_5 (-59° to 30°)	N_1 (-59° to 30°)	CCD	Downward concavity
G_7	181° to 210°	S_7	Diagonal falling (a)	N_3 (31° to 120°)	N_7 (31° to 120°)	CCR	Right concavity
G_8	211° to 240°	S_8	Diagonal falling (b)	N_3 (211° to 300°)	N_7 (211° to 300°)	CCL	Left concavity
G_9	241° to 270°	S_9	Comer (a)	N_2 (151° to 240°)	N_0 (241° to 330°)		
G_{10}	271° to 300°	S_{10}	Comer (b)	N_6 (151° to 240°)	N_0 (61° to 150°)		
G_{11}	301° to 330°	S_{11}	Comer (c)	N_4 (241° to 330°)	N_2 (-29° to 60°)		
G_{12}	331° to 360°	S_{12}	Comer (d)	N_6 (-29° to 60°)	N_4 (61° to 150°)		

Table 1. Databases used in writer identification/verification.

DB#	DB Name	DB reference	Database used in	Public	Language	Type	#Writers
DB01	*na**	Gazzah and Ben (2006)	Gazzah and Ben (2006,2007,2008)	No	Arabic	Text	60
DB02	*na**	Al-Dmour and Zitar (2007)	Al-Dmour and Zitar (2007)	No	Arabic	Text	20
DB03	IFN/ENIT	El Abed and Märgner (2007)	Abdi et al. (2009); Bulacu et al. (2007)	Yes	Arabic	Words	411
DB04	AHDB	Al-Ma'adeed et al. (2008b)	Al-Ma'adeed et al. (2008a; b)	No	Arabic	Words/Phrases	100
DB05	*na**	Srihari and Ball (2008)	Srihari and Ball (2008)	No	Arabic	Text	10
DB06	*na**	Liu et al. (1995)	Liu et al. (1995)	No	Chinese	Characters	20
DB07	*na**	Zhu et al. (2000)	Zhu et al. (2000)	No	Chinese	Text	17
DB08	*na**	Cong et al. (2002)	Cong et al., (2002)	No	Chinese	Text	50
DB09	*na**	He and Tang (2004)	He and Tang (2004)	No	Chinese	Text	50
DB10	*na**	He et al. (2005a)	He et al. (2005a; b)	No	Chinese	Text	10
DB11	HIT-MW	Su et al. (2007)	Li and Ding (2009)	Yes	Chinese	Text	240
DB12	*na**	He et al (2008b)	He (2008a; b)	No	Chinese	Text	500
DB13	SET1	Wang et al. (2003)	Wang et al. (2003)	No	Chinese	Characters	25
	SET2						626

Table 2. Contd.

DB14	Firemaker	Schomaker and Vuurpijl (2000)	Brink et al. (2008); Bulacu, (2007); Bulacu and Schomaker (2005, 2006, 2007b); Bulacu et al. (2003); Maaten and Postma (2005); Schomaker and Bulacu (2004); Schomaker et al. (2004, 2003, 2007)	Yes	Dutch	Text	250
DB15	Unipen	Guyon et al. (1994)	Bulacu (2007); Bulacu and Schomaker (2005, 2006, 2007b); Niels et al. (2007); Schomaker et al. (2004, 2007)	Yes	Various	Text	215
DB16	IAM	Marti and Bunke (2002)	Brink et al. (2008); Bulacu (2007; Bulacu and Schomaker (2006, 2007b); Helli and Moghaddam (2009); Schlapbach and Bunke (2007); Schomaker and Bulacu (2004); Siddiqi and Vincent (2008, 2009)	Yes	English	Text	657
DB17	*na**	Bulacu and Schomaker (2007a)	Bulacu and Schomaker (2007a)	No	Medieval English	Text	10
DB18	Trigraph	Brink et al. (2010)	Brink et al. (2010)	Yes	Dutch	Text	47
DB19	*na**	Hull (1994)	Srihari (2000)	No	English	Words, digits	*na**
DB20	Cedar Letter	Cha and Srihari (2000)	Srihari and Ball (2009); Srihari et al. (2002, 2005, 2007); Tomai and Shrihari (2004); Zhang et al. (2003)	No	English	Text	1000
DB21	*na**	Matsuura and Qiao (1989)	Matsuura and Qiao (1989)	No	English	Words	2
DB22	*na**	Said et al., (1998)	Said et al., (1998)	No	English	Text	20
DB23	*na**	Leedham and Chachra (2003)	Leedham and Chachra (2003)	No	English	Digits	15
DB24	HIFCD2	Zois and Anastassopoulos (2000)	Zois and Anastassopoulos (2000)	Yes	English and Greek	Words	50
DB25	IRONOFF	Viard-Gaudin et al. (1999)	Tan et al. (2008)	Yes	French	Letters/Words	700
DB26	*na**	Bensefia et al. (2002)	Bensefia et al. (2002); Bensefia (2003b)	No	French	Text	88
DB27	RIMES	Grosicki et al. (2008)	Siddiqi and Vincent (2009)	Yes	French	Text	1300
DB28	*na**	Bar-Yosef et al. (2007)	Bar-Yosef et al. (2007)	No	Historical Hebrew	Characters	34
DB29	*na**	Mar and Thein (2005)	Mar and Thein (2005)	No	Myanmar	Characters	20
DB30	*na**	Shahabi and Rahmati (2006)	Shahabi and Rahmati (2006, 2007)	No	Persian	Text	40
DB31	PD100	Helli and Moghaddam (2008b)	Helli and Moghaddam (2008a, 2008b, 2009, 2010)	No	Persian	Text	100
DB32	*na**	Ram and Moghaddam (2009a)	Ram and Moghaddam (2009a, 2009b)	No	Persian	Text	50
DB33	*na**	Ubul et al. (2008)	Ubul et al. (2008)	No	Uyghur	Text	23

*na**: Information not available.

CLASSIFICATION APPROACHES

The research of writer identification and verification used different classifier approaches. Friedman et al. (1999) categorize classifier types into five kinds; minimum distance classifiers, statistical classifiers, neural networks, fuzzy classifiers, and syntactic classifiers. Using this categorization, this section addresses the

Table 3. Writer identification/verification features and classifiers.

Citation	Feature	Classifier	DB	#Wr	Info	Dep./Ind.	Top-1 (%)
Gazzah and Ben (2006)	Entropy as global features, Wavelet transforms, and a set of structural features	Neural networks.	DB01	60	3 docs/wr	Dep.	94.73
Gazzah and Ben (2007)	See Gazzah and Ben Amara (2006)	Neural networks.	DB01	60	3 docs/wr	Dep.	95.68
Gazzah and Ben (2008)		SVM, neural networks.	DB01	60	3 docs/wr	Dep.	94.00
Al-Dmour and Zitar (2007)	Gabor filters	Weighted Euclidean, SVM, LDC.	DB02	20	*na**	*na**	90.00
Abdi et al. (2009)	Length, height/width ratio, and curvature of strokes.	Euclidean, Square, Manhattan, X^2, Chebechev, Hamming, Minkowski, and Mahalanobis distance.	DB03	40	> 100 words/wr	Ind.	92.50
Al-Ma'adeed et al., (2008a)	Edge-hinge features. Grapheme features.	Euclidean distance.	DB04	10	2 docs/wr	Dep.	90.00
Al-Ma'adeed et al., (2008b)	Edge-direction distribution. Moment Invariants, Area, length, Height, Length from Baseline to Upper Edge, Baseline to the Lower Edge.	Euclidean distance.	DB04	100	20 docs/wr	Dep.	93.80
Liu et al. (1995)	Gabor filters. Features from 2nd and 3rd order moments.	Manhattan distance.	DB06	20	7 docs/wr	Ind.	100.0
Zhu et al. (2000)		Weighted Euclidean.	DB07	17	1 doc/wr	Ind.	95.70
Cong et al. (2002)	Gabor filters	Euclidean distance.	DB08	50	110 scripts	Ind.	97.60
He and Tang (2004)		Weighted Euclidean.	DB09	50	2 docs/wr	Both	90.00
He et al. (2005b)	Contourlet transforms.	Kullback-Leibler Distance	DB10	10	2 docs/wr	Ind.	90.00
He et al. (2005a)		Kullback-Leibler Distance	DB10	10	2 docs/wr	Ind.	80.00
He et al. (2008b)	Wavelet transforms	Hidden Markov Tree model	DB12	500	2 docs/wr	Ind.	36.40
He (2008a)		Kullback-Leibler distance	DB12	500	2 docs/wr	Ind.	39.20
Li and Ding (2009)	Histogram of contour-hinge.	Weighted Euclidean, and modified X^2 distance measure	DB11	240	1 doc/wr	Ind.	95.00
Wang et al. (2003)	Distribution of directional elements (gradient).	Euclidean distance	DB13	25 626	16*34 char/wr 20 char/wr	Dep. Dep.	96.12 82.16

Table 3. Contd.

Bulacu et al. (2003)	Edge-direction distribution, edge-hinge distribution, run-length distributions, autocorrelation, and entropy.	Euclidean distance	DB14	250	2 docs/wr	Ind.	75.00
Bulacu et al. (2003)	See (Bulacu et al., 2003).	X^2, Hamming, Minkowski, Bhattacharyya, and Hausdorff distance	DB14	251	2 docs/wr	Ind.	88.00
Schomaker et al. (2003)	See (Schomaker et al., 2003). Grapheme emission PDFs.	X^2 and Hamming distance	DB14	150	1 doc/wr	Dep.	87.00
Schomaker and Bulacu (2004)			DB14	150	1 doc/wr	Dep.	97.00
		Euclidean distance	DB14	250	2 docs/wr	Ind.	78.10
Bulacu and Schomaker (2005)			DB14	250	1 doc/wr	Dep.	64.90
			DB15	150	2 docs/wr	Ind.	76.30
	See Schomaker and Bulacu, (2004).						
Bulacu and Schomaker (2006)			DB14,15,16	900	2 docs/wr	Ind.	87.00
Bulacu (2007)		X^2 and Hamming distance	DB14,15,16	900	2 docs/wr	Ind.	87.00
Bulacu and Schomaker (2007a)			DB17	10	2 regions/wr	Ind.	89.00
Bulacu and Schomaker (2007b)			DB14,15,16	900	2 docs/wr	Ind.	87.00
Schomaker et al. (2007)	See Schomaker and Bulacu (2004). Writer information: handedness, sex, age, and style.	X^2 distance	DB14	150	1 doc/wr	Dep.	80.00
Bulacu et al., (2007)		X^2 and Hamming distance	DB04	350	5 docs/wr	Ind.	88.00
Brink et al. (2008)	See Schomaker and Bulacu (2004).	*na**	DB14,16	498	2 docs/wr	Ind.	Varies
Brink et al. (2010)		X^2 distance	DB18	47	4 docs/wr	Dep.	97.00-100
Schlapbach and Bunke (2004b)	Sliding window.	Hidden Markov Models (HMM)	DB16	50	5 docs/wr	Ind.	94.23
Schlapbach and Bunke (2004a)	See Schlapbach and Bunke (2004a).	Hidden Markov Models (HMM)	DB16	100	5 docs/wr	Ind.	96.56
Schlapbach et al. (2005)	100 simple features: slant, skew angle, fractal features…	Euclidean distance	DB16	50	5 docs/wr	Ind.	98.36
Schlapbach and Bunke (2006)		HMM, Gaussian Mixture Models	DB16	100	5 docs/wr	Ind.	98.46
Schlapbach (2007)	See Schlapbach and Bunke (2004a).	HMM, Gaussian Mixture Models	DB16	100	5 docs/wr	Ind.	97.03
Schlapbach and Bunke, (2007)		HMM, Gaussian Mixture Models	DB16	100	5 docs/wr	Ind.	97.03
Srihari et al. (2002)	Gradient, structural, and concavity histograms. Eleven macro features.	Euclidean distance Correlation measure	DB20	1500	3 docs/wr	Dep.	98.00

Table 3. Contd.

Zhang (2003)		Euclidean distance Correlation measure	DB20	1000	3 docs/wr	Dep.	98.06
Tomai and Srihari (2004)	See Srihari et al. (2002).	Manhattan and Correlation measure	DB20	1000	3 docs/wr	Dep.	99.00
Srihari et al. (2007)		Manhattan and Correlation measure	DB20	1000	3 docs/wr	Dep.	96.10
Srihari and Ball (2008)		Manhattan and Correlation measure. Log-likelihood ratio	DB05	10	10 docs/wr	Ind.	99.30
Srihari and Ball (2009)		Log-likelihood ratio	DB20	1000	3 docs/wr	Dep.	*na**
Matsuura and Qiao 1(989)	Impulse response of image	Euclidean distance	DB21	2	5 words/wr	Dep.	100.0
Said et al. (1998)	Gabor filters. Grey Scale Co-occurrence Matrices.	Weighted Euclidean	DB22	20	25 block/wr	Ind.	95.30
Leedham and Chachra (2003)	Pixel density, fixed point distance and angular measure, center of gravity, gradient features, connected components contours, and a collection of local features.	Hamming distance	DB23	15	10 strings/wr	Ind.	100.0
Zois and Anastassopoulos (2000)	Erosion and dilation function.	Linear Bayes classifier Neural networks	DB24	50	90 words/wr	Dep.	> 95.0
Tan et al. (2008)	x and y co-ordinates, the directions of x and y co-ordinates, the curvatures of x and y co-ordinates and the Pen-up or Pen- down information.	Fuzzy classifiers	DB25	120	Characters, online	*na**	98.30
Bensefia et al. (2003b)	Grapheme clustering.	Correlation similarity measure	DB26	88	1 doc/wr	Dep.	97.70
Siddiqi and Vincent (2007)	Modified sliding window.	Bayesian classifier	DB16	50	2 docs/wr	Ind.	94.00
Siddiqi and Vincent (2008)	Gabor filters.	Mahalanobis distance	DB16	100	2 docs/wr	Ind.	92.00
Siddiqi and Vincent (2009)	Chain code histograms.	Euclidean, X^2, Hamming, and Bhattacharyya distance	DB16 DB27	650 225	2 docs/wr 2 docs/wr	Ind. Ind.	86.00 79.00
Bar-Yosef et al. (2007)	The ratio between the area of the background and the convex hull. The aspect ratio of the enclosing ellipse. Concavity features. Ellipse aspect ratio. Moment features.	Euclidean distance and Linear Bayes classifier	DB28	34	20 characters/wr	Dep.	100.0

Table 3. Contd.

Mar and Thein (2005)	Mean and standard deviation of ROIs	Weighted Euclidean	DB29	20	2 docs/wr	Dep.	97.50
Shahabi and Rahmati (2006)		Weighted Euclidean.X^2 distance	DB30	25	4 blocks/wr	Dep.	88.00
Shahabi and Rahmati (2007)		Euclidean and X^2 distance	DB30	40	3 docs/wr	Dep.	82.50
Helli and Moghaddam (2008b)		Longest Common Subsequence	DB31	100	5 docs/wr	Ind.	95.00
Helli and Moghaddam (2008a)	Gabor filters	Weighted Euclidian distance	DB31	70	5 docs/wr	Ind.	77.00
Helli and Moghaddam (2009)		Longest Common Subsequence	DB31	100	5 docs/wr	Ind.	89.00
			DB16	30	7 docs/wr	Ind.	94.40
Helli and Moghaddam (2010)		Graph similarity	DB31	100	5 docs/wr	Ind.	98.00
Ram and Moghaddam (2009a)	Gradient features.	Neural networks	DB32	50	5 docs/wr	Ind.	94.00
Ram and Moghaddam, (2009b)	Grapheme features. Gradient features. Used area features.	Fuzzy classifiers	DB32	50	5 docs/wr	Ind.	90.00
Ubul et al. (2008)	Gabor filters.	Euclidean distance, weighted Euclidean, and SVM	DB33	23	2 docs/wr	Dep.	88.00

*na**: Information not available

classifier types used in writer identification and verification.

Minimum distance classifiers

Minimum distance classifiers classify a new pattern by measuring its distance from the test sample to the training patterns and choosing the K-nearest classes to which the nearest neighbors belong (Friedman and Kandel, 1999). Various distance measures have been attempted; with the Euclidean distance measure remains the most commonly used distance measure for writer identification and verification. Researchers who used the Euclidean distance measure include (Abdi et al., 2009; Al-Ma'adeed et al., 2008a; Al-Ma'adeed et al., 2008b; Bar-Yosef et al., 2007; Bulacu and Schomaker, 2005; Bulacu et al., 2003; Cong et al., 2002; Matsuura and Qiao, 1989; Siddiqi and Vincent, 2009; Srihari et al., 2002; Ubul et al., 2008; Wang et al., 2003; Zhang, 2003). By adding weights to each feature value, researchers also used the weighted Euclidean distance measure (Al-Dmour and Zitar, 2007; He and Tang, 2004; Li and Ding, 2009; Mar and Thein, 2005; Said et al., 1998; Shahabi and Rahmati, 2006, 2007; Ubul et al., 2008; Zhu et al., 2000).

Other used distance measures for writer identification/verification include: square Euclidean distance (Abdi et al., 2009), Manhattan also known as city block - distance measure (Abdi et al., 2009; Liu et al., 1995; Srihari and Ball, 2008, 2009; Srihari et al., 2007a; Srihari et al., 2007b; Tomai and Srihari, 2004), X^2 distance measure (Abdi et al., 2009; Brink et al., 2010; Bulacu, 2007; Bulacu and Schomaker, 2006, 2007a, 2007b; Bulacu et al., 2007a; Bulacu et al. 2007b; Schomaker and Bulacu, 2004; Schomaker et al., 2003; Shahabi and Rahmati, 2006, 2007; Siddiqi and Vincent, 2009), a modified version of the χ^2 distance measure (Li and Ding, 2009), Chebechev distance measure (Abdi et al., 2009), Hamming distance measure(Abdi et al., 2009; Bulacu, 2007; Bulacu and Schomaker, 2006, 2007a, 2007b; Bulacu et al., 2007a; Bulacu et al., 2007b; Leedham and Chachra, 2003; Schomaker and Bulacu, 2004; Schomaker et al., 2004, 2003; Siddiqi and Vincent, 2009), Minkowski (Abdi et al., 2009; Schomaker et al., 2003; Tomai and Srihari, 2004), the Mahalanobis distance measure (Abdi et al., 2009; Siddiqi and Vincent, 2008), correlation measure (Bensefia et al., 2002; 2003b; Srihari and Ball, 2008; 2009; Srihari et al., 2005; 2007a; 2007b; Tomai and Srihari, 2004; Zhang, 2003; Zhang et al., 2003), Bhattachalyya distance (Schomaker et al., 2003; Siddiqi and Vincent, 2009), the Hausdorff distance (Schomaker et al., 2003), and the Longest Common Subsequence (LCS) algorithm (Helli and Moghaddam, 2008b, 2009).

Since the performance of distance measured heavily rely on the features' nature, it is often hard to conclude the best distance measure for writer identification/ verification. Nevertheless, many researchers have reported that the χ^2 distance measure reported highest accuracy for their features when compared with other distance measures (Brink et al., 2008, 2010; Bulacu, 2007; Bulacu and Schomaker, 2005, 2006, 2007a; Bulacu et al., 2007b; 2003a; Franke et al., 2003; Niels et al., 2007; Schomaker and Bulacu, 2004; Schomaker et al., 2004, 2003, 2007). In addition, Srihari et al. used binary feature vectors for writer identification and verification, and hence relied on (dis) similarity computation for classification (Srihari, 2000; Srihari and Ball, 2008; 2009; Srihari et al., 2005; 2002; Srihari et al., 2007a; b; Tomai and Srihari, 2004; Zhang, 2003; Zhang et al., 2003). They conducted various experiments to select the best performing (dis)similarity measure and concluded that the correlation distance measure provided the best results (Zhang, 2003).

Statistical classifiers

Minimum distance classifiers are based on the assumption that training samples form distinct clusters. However, this is not usually the case. Training samples of various classes overlap, and in this case a statistical approach is more appropriate assuming that the samples come from statistical distribution (Friedman and Kandel, 1999). Examples of statistical classifiers used in writer identification and verification include Linear Bayes classifier (Bar-Yosef et al., 2007; Zois and Anastassopoulos, 2000), Support Vector Machines (SVM) (Al-Dmour and Zitar, 2007; Franke et al., 2002; Gazzah and Ben Amara, 2008; Ubul et al., 2008), Hidden Markov Models (HMM) (Schlapbach and Bunke, 2004a, 2007), Hidden Markov Tree (HMT) model (He et al., 2008b), Gaussian Mixture Models (GMM) (Schlapbach, 2007), Kull back Leibler distance (KLD) between two PDFs (He et al., 2008a), Cumulative Distribution Functions of the log-likelihood ratio (LLR) of the same and different writers (Srihari and Ball, 2008; Srihari et al., 2005), and the linear discriminant classifier (LDC) (Al-Dmour and Zitar, 2007).

Other classifiers

Researchers have also used neural networks (Gazzah and Ben, 2006, 2007, 2008; Ram and Moghaddam, 2009a; Zois and Anastassopoulos, 2000), fuzzy classifiers (Ram and Moghaddam, 2009b; Tan et al., 2008).Structural classifiers are used less frequently and with less significant accuracy results (Helli and Moghaddam, 2010).

WRITER IDENTIFICATION AND VERIFICATION OF ARABIC TEXT

Writer identification and verification of Arabic text is still considered a fresh field but seems to be getting a strong momentum lately. To the best of the researchers' knowledge, only limited number of researchers has addressed writer identification and verification of Arabic text. In the following section, we will address the databases used in writer identification and verification of Arabic text then a survey of writer identification and verification of Arabic text follows. In addition, research of writer identification and verification for Persian (Farsi) text was also addressed due the similarities of Arabic and Farsi text.

Databases used in writer identification of Arabic text

The IfN/ENIT database (Pechwitz et al., 2002a; El Abed and Märgner, 2007a) was created by the Institute of Communications Technology (IfN) at Technical University Braunschweig in Germany and the Ecole Nationale d'Inge´nieurs de Tunis (ENIT) in Tunisia. The database consists of 26 459 images of the 937 names of cities and towns in Tunisia, written by 411 different writers. To this date, this database has been widely used by many researchers of Arabic handwritten text recognition (more than 100 research groups from more than 30 countries) and has appeared in several global competitions (Märgner and El Abed, 2007, 2009, 2010, 2011; Märgner et al., 2005). Due to its public availability, researchers have also used the IfN/ENIT database for writer identification of Arabic text (Abdi et al., 2009; Bulacu et al., 2007) although it is limited to city names and thus contains limited vocabulary. Figure 6 shows an example of a filled form of the IfN/ENIT database.

Al-Ma'adeed et al. presented the AHDB (Al-Ma'adeed et al., 2002), which contains Arabic words and texts written by one hundred writers. It also contains the most popular words in Arabic, as well as, sentences used in writing checks with Arabic words. Finally, it contains free

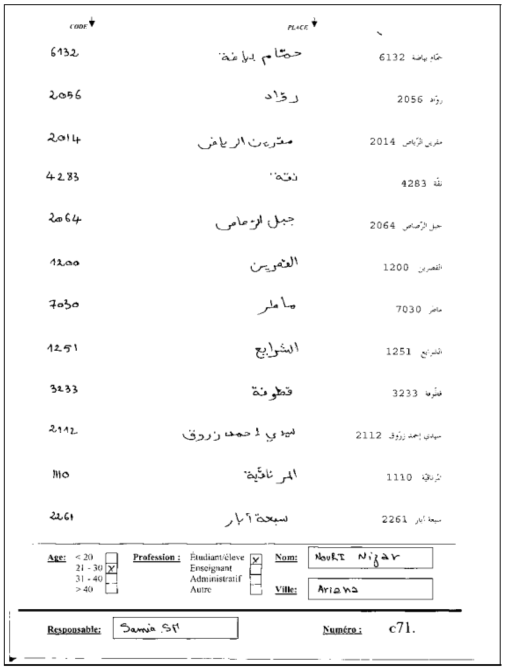


Figure 7. An example of an IFN/ENIT filled form (El Abed and Märgner, 2007).

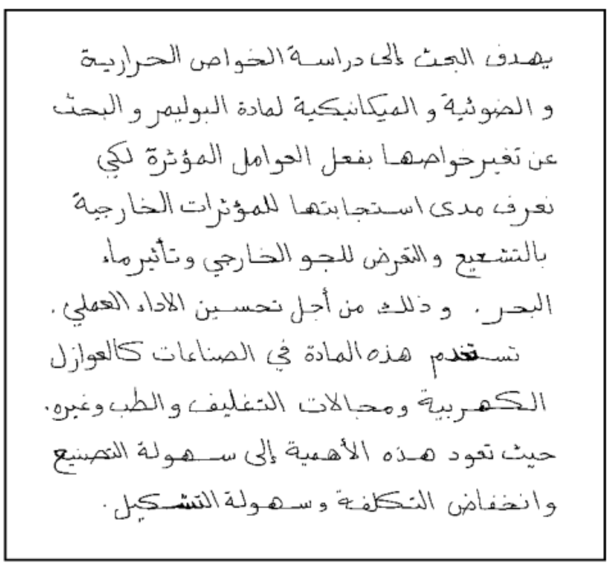


Figure 8. Free handwriting sample from the AHDC dataset (Al-Ma'adeed et al., 2002).

handwriting pages in a topic of interest to the writer. The form was designed in five pages. The first three pages were filled with ninety-six words, sixty-seven of which are handwritten words corresponding to textual words of numbers that can be used in handwritten cheque writing. The other twenty-nine words are from the most popular words in Arabic writing. The fourth page is designed to contain three sentences of handwritten words representing numbers and quantities that can be written on cheques. The fifth page is lined, and designed to be completed by the writer in freehand on any subject of his choice as shown in Figure 8. Further information, like the availability of the dataset, is not clear from the authors' published work. Al-Ma'adeed et al. used their database for Arabic writer identification in (Al-Ma'adeed et al., 2008a; b).

Srihari et al. used a much smaller database for writer identification of Arabic handwritten text (Ball and Srihari, 2008) prepared from 10 different writers, each contributing 10 different full page documents in handwritten Arabic for a total of 100 documents.

Gazzah and Ben (2006, 2007, 2008) designed their own Arabic letter database, which contains 505 characters, 15 numerals and 6 punctuations. The choice of the letter contents was made to ensure the use of the various internal shapes of the letter within a sub-word (isolated, initial, middle and end). Handwriting samples of 60 persons were collected. Each person was required to copy the same letter three times: two samples were used for training and the other for the testing; a total of 180, A4 format sample pages. Finally, Table 2 shows a summary of handwritten text databases used for writer identification.

Writer identification and verification of Arabic text

In this section we present a survey of research of writer identification and verification of Arabic text. It is to be noted that most of the efforts of writer identification and verification of Arabic text are based on the techniques that were used for English text. Most of the features and classifiers were previously used for writer identification of English text. Since Persian (Farsi) text is similar to Arabic, research of writer identification and verification of Persian text will also be presented.

Researchers used a combination of global and structural features (Average line height, Spaces between sub-words, inclination of the ascender, height and the width of each diacritic dot) along with a multilayer perceptron (MLP) classifier (Gazzah and Ben, 2006). They reported an accuracy of 94.73% for 60 writers. (Gazzah and Ben, 2007) used a 2D discrete wavelet transforms for feature extraction along with the MLP classifier with a reported accuracy of 95.68% on the same database. In their latest report work, Support Vector Machines (SVM) classifier was used where they showed that MLP provided slightly better results than SVM (Gazzah and Ben, 2008).

Bulacu et al. (2007) used the IFN/ENIT dataset (Pechwitz et al., 2002), which is limited to Arabic town and city names. For tests involving 350 writers, they reported a best accuracy of 88%. They concluded that the identification and verification results obtained on Arabic text cannot be numerically compared with previous results for Western script because the experimental datasets are different (in terms of the amount of ink contained in the samples among others). They also indicated that the results obtained on Arabic text are generally lower than the ones obtained on Western script. Abdi et al. (2009) used the IFN/ENIT dataset, but with only 40 writers (Pechwitz et al., 2002). Using statistical features (the length, height/width ratio, and the curvature of the strokes to calculate various probability distribution function (PDF) feature vectors) along with Euclidean, Manhattan, and Mahalanobis distance measures and the Borda count ranking algorithm, they reported a top-1 accuracy of 92.5%.

Al-Dmour and Zitar (2007) presented a technique for feature extraction based on hybrid spectral-statistical measures (SSMs) of texture. Correct identification of 90% was reported using Arabic handwriting samples from 20 different writers. Al-Ma'adeed et al. used edge-based statistical features to recognize Arabic handwritten words (Al-Ma'adeed et al., 2008a; b). They used their own generated database as described previously. Some of the phrases scored a Top-10 result of more than 90% accuracy, whereas shorter words scored around 50% accuracy for 100 writers. Srihari and Ball (2008) used a dataset of 10 different writers, each contributing 10 different full page documents in hand written Arabic for a total of 100 pages. Using macro- and micro-features along with likelihood ratio computation, they reported 86% accuracy.

Persian, also known as Farsi, handwriting is very similar to Arabic in terms of strokes and structure. Therefore, a Persian writer identification system can also be used for identification of Arabic text. Farsi character set comprises all of the 28 Arabic characters plus four additional ones, shown in Figure 9. Similar to Arabic, Persian writer identification and verification has been increasingly popular lately. Shahabi and Rahmati (2006, 2007) used features based on Gabor filters for feature extraction, and different distance measures (Euclidean, Weighted Euclidean, and X^2 distance) for classifiers. Their latest work reported a top-1 accuracy of 82.50% for 40 writers. Ram and Moghaddam (2009a, 2009b) used gradient and grapheme features and tested them on a database of 50, writers 5 pages per writer and reported top-1 accuracy of 94.0%.

Helli and Moghaddam (2008a, 2008b, 2009, 2010) used modified Gabor filters for feature extraction and tried different classification techniques for identification. They used a database of 100 writers, 5 pages per writers. The volunteer was free to write anything in the pages, and hence their approach was text independent. They reported top-1 accuracy of 98% for all 100 writers. Quite interestingly, they tried their system on the IAM database (Marti and Bunke, 2002) for 30 writers (7 pages per writer) and reported top-1 accuracy of 94.4%. Since the databases are different, hence their results cannot be compared. Therefore, no conclusion can be drawn based on Latin/Farsi text although the general understanding is that Latin text gives better identification rates. We think that the used data for Arabic text writer identification does not match in representation and naturalness the databases of Latin text.

گ	ژ	چ	پ

Figure 9. Four additional Farsi isolated characters.

CONCLUSIONS AND FUTURE DIRECTIONS

In this paper we presented the state of the art in writer identification and verification of Latin and western texts, the databases used, the feature extraction approaches, and the classifier approaches. The state of the art was grouped by addressing the research work publications of different research groups due to similarities in used features and classifiers. This grouping helps in showing each group's own improvement over time and related to other groups. The published research work was tabulated indicating the used features, the classifiers, the databases used, the best identification rates of each publication, the number of writers and the year of publication. This makes it easier to compare the research work of the different researchers. Tabulation was included for the used databases, the number of writers, samples, etc. This indicates the large number of publications on this topic and increasing number of researchers working in this area.

The paper presented a survey of writer identification and verification of Arabic text. Comparing the work on Arabic text with Latin indicates that limited number of researchers is involved in writer identification of Arabic text. In addition, comparing features and classification approaches indicates that most of the work on Arabic text is based on features and classifiers used for English. For Arabic, most of the databases are researcher generated for their own research with the exception of the IEF/ENIT database, which consists of city names. So far there is no Arabic text database that is freely available, for writer identification of Arabic text. It is clear that the published work related to Arabic text has lower accuracy than Latin. We cannot conclude that Arabic text is less identifiable than Latin text, although, the general understanding in that Latin text gives better identification rates. We think that the used data for Arabic text writer identification does not match in representation and naturalness of the databases of Latin text. The used databases are self-generated (with embedded limitations in size and comprehensive) or the IEF/ENIT, which consists of city names in which researchers had to concatenate a number of city names to make an Arabic text. This is neither a good representation of Arabic text nor comprehensive. To reach to a real conclusion about this issue, more research work needs to be conducted using a more representative and natural databases of Arabic text and use features that take advantage of the characteristics of Arabic text like diacritics, dotted characters, the writing line, etc.

We expect this to change with time. There is a need for an Arabic text database with large number of writers for writer identification and verification. It is also about time that researchers of writer identification and verification of Arabic text design features that are novel and that take the characteristics of Arabic text into considerations. Researchers, as shown above, have indicated that techniques for Latin techniques give lower rates when applied to Arabic due to some characteristics of the language.

There is a need for establishing research groups for Arabic text recognition and identification. This will enable building resources that the research community can utilize. We hope this survey of writer identification of Arabic text, although, limited due to limited publications on Arabic, encourages more researchers to contribute.

ACKNOWLEDGMENT

The authors would like to acknowledge the support provided by King Abdul-Aziz City for Science and technology (KACST) through the Science & Technology Unit at King Fahd University of Petroleum and Minerals (KFUPM) for funding this work through project number(A-T-18-95) and project number (08-INF99-4).

REFERENCES

Abdi M, Khemakhem M, Ben-Abdallah H (2009). A Novel Approach for Off-Line Arabic Writer Identification Based on Stroke Feature Combination. 24th International Symposium on Computer Information Science, IEEE. pp. 597-600.

Al-Dmour A, Zitar R (2007). Arabic Writer Identification Based on Hybrid Spectral-Statistical Measures. J. Exp. Theor. Artif. Intell. 19(4):307-332.

Al-Ma'adeed S, Al-Kurbi A, Al-Muslih A, Al-Qahtani R, Al Kubisi H (2008a). Writer Identification of Arabic Handwriting Documents Using Grapheme Features. IEEE/ACS International Conference on Computing System Application. pp. 923-924.

Al-Ma'adeed S, Elliman D, Higgins C (2002). A data base for Arabic handwritten text recognition research. Proceedings of the International Workshop on Frontiers in Handwriting Recognition. pp. 485-489.

Al-Ma'adeed S, Mohammed E, Al Kassis D (2008b). Writer Identification Using Edge-Based Directional Probability Distribution Features for Arabic Words. IEEE/ACS International Conference on Computing System Applications. pp. 582-590.

Ball G, Srihari S (2008). Writer adaptation in off-line Arabic handwriting recognition. Doc. Recognit. Retr. 68(15):4.

Bar-Yosef I, Beckman I, Kedem K, Dinstein I (2007). Binarization, character extraction, and writer identification of historical Hebrew calligraphy documents. Int. J. Doc. Anal. Recognit. 9(2):89.

Bensefia A, Nosary A, Paquet T, Heutte L (2002). Writer identification by writer's invariants. Eighth International Workshop on Frontiers in Handwriting Recognition. IEEE Computing Society, Ontario, Canada. pp. 274-279.

Bensefia A, Paquet T, Heutte L (2003a). Grapheme based writer verification. 11th Conference of the International Graphonomics Society (IGS2003) Arizona, USA. In: Proceedings of the 11th Conference of the International Graphonomics Society. pp. 274-277.

Bensefia A, Paquet T, Heutte L (2003b). Information Retrieval Based Writer Identification. 7th International Conference on Document Analysis and Recognition IEEE Computing Society, Edinburgh, Scotland. p. 946.

Bensefia A, Paquet T, Heutte L (2004). Handwriting Analysis for Writer Verification. 9th International Workshop on Frontiers in Handwriting Recognition. IEEE, Tokyo, Japan. pp. 196-201.

Bensefia A, Paquet T, Heutte L (2005a). A writer identification and verification system. Pattern Recognit. Lett. 26(13):2080-2092.

Bensefia A, Paquet T, Heutte L (2005b). Handwritten Document Analysis for Automatic Writer Recognition. Electron. Lett. Comput. Vis. Image Anal. 5(2):72-86.

Bradford R (1992). Introduction to Handwriting Examination and Identification. Burnham, Inc.

Brink A, Bulacu M, Schomaker L (2008). How Much Handwritten Text Is Needed for Text-Independent Writer Verification and Identification. 19th International Conference on Pattern Recognition. pp. 1-4.

Brink A, Niels R, van Batenburg R, van Den Heuvel C, Schomaker L (2010). Towards Robust Writer Verification by Correcting Unnatural Slant. Pattern Recognit. Lett. 32(3):449-457.

Bulacu M (2007). Statistical Pattern Recognition for Automatic Writer Identification and Verification. Ph.D. Dissertation, Department of Behaviour and Social Science, University of Groningen, Netherlands, Netherlands. p. 140.

Bulacu M, Schomaker L (2005). A Comparison of Clustering Methods for Writer Identification and Verification. 8th International Conference on Document Analysis and Recognition. 2:1275-1279

Bulacu M, Schomaker L (2006). Combining Multiple Features for Text-Independent Writer Identification and Verification. In: G. Lorette (Ed.), Proceedings of 10th IWFHR La Baule (France): Suvisoft. pp. 281-286.

Bulacu M, Schomaker L (2007a). Automatic Handwriting Identification on Medieval Documents. 14th International Conference on Image Analysis and Processing. pp. 279-284.

Bulacu M, Schomaker L (2007b). Text-Independent Writer Identification and Verification Using Textural and Allographic Features. IEEE Trans. Pattern Anal. Mach. Intell. 29(4): 701-717.

Bulacu M, Schomaker L, Brink A (2007). Text-Independent Writer Identification and Verification on Offline Arabic Handwriting. 9th International Conference on Document Analysis and Recognition. 2:769-773.

Bulacu M, Schomaker L, Vuurpijl L (2003). Writer Identification Using Edge-Based Directional Features. International Conference on Document Anal. Recognit. IEEE Comput. Soc. p. 937.

Bulacu M, van Koert R, Schomaker L, van Der Zant T (2007). Layout Analysis of Handwritten Historical Documents for Searching the Archive of the Cabinet of the Dutch Queen. Ninth Int. Conf. Doc. Anal. Recognit. (ICDAR 2007) 1: 357-361.

Cha S (2001). Use of distance measures in handwriting analysis. Ph.D. dissertation, Dept. of Comp. Sci. and Eng., St. Univ. N.Y Buffalo, NY, p. 239. Retrieved from http://portal.acm.org/citation.cfm?id=933434

Cha S, Srihari S (2000). Assessing the Authorship Confidence of Handwritten Items. Fifth IEEE Workshop on Applications of Computer Vision, California, USA. pp. 42-47.

Cong S, Xiao-Gang R, Tian-Lu M (2002). Writer identification using Gabor wavelet. Proceedings of the 4th World Congress on Intelligent Control and Automation. pp. 2061-2064.

El Abed H, Märgner V (2007). The IFN/ENIT-database - a tool to develop Arabic handwriting recognition systems. 9th International Symposium on Signal Processing and Applications. Sharjah, U.A.E. pp. 1-4.

Fornes A, Llados J, Sanchez G, Bunke H (2008). Writer Identification in Old Handwritten Music Scores. The Eighth IAPR International Workshop on Document Analysis and Systems. pp. 347-353.

Franke K, Bünnemeyer O, Sy T (2002). Ink Texture Analysis for Writer Identification. Proceedings of the 8th International Workshop on Frontiers in Handwriting Recognition. IEEE Computing Society, Los Alamitos, CA, USA. p. 268.

Franke K, Koppen M (2001). A Computer-Based System to Support Forensic Studies on Handwritten Documents. Int. J. Doc. Anal. Recognit. 3(4):218-231.

Franke K, Schomaker L, Veenhuis C, Taubenheim C, Guyon I, Vuurpijl L, van Erp M, Zwarts G (2003). WANDA: a generic framework applied in forensic handwriting analysis and writer identification. Design and application of hybrid intelligent systems. Proceedings of the 3rd International Conference on Hybrid Intelligence Systems. IOS Press. pp. 927-938.

Friedman M, Kandel A (1999). Introduction to Pattern Recognition: Statistical, Structural, Neural and Fuzzy Logic Approaches. World Sci. Publishing Company. p. 329.

Gazzah S, Ben Amara N (2006). Writer Identification Using Modular MLP Classifier and Genetic Algorithm for Optimal Features Selection. In: Wang J., Yi Z, Zurada J, Lu BL and Yin H (Eds.), Advances in Neural Networks. Springer Berlin / Heidelberg. 3972: 271-276.

Gazzah S, Ben Amara N (2007). Arabic Handwriting Texture Analysis for Writer Identification Using the DWT-Lifting Scheme. Ninth International Conference on Document Analysis and Recognition 2:1133-1137.

Gazzah S, Ben Amara N (2008). Neural Networks and Support Vector Machines Classifiers for Writer Identification Using Arabic Script. Int. Arab J. Inform. Technol. 1:13-75.

Gibbons M, Yoon S, Cha S, Tappert C (2005). Evaluation of Biometric Identification in Open Systems. In Kanade T, Jain A and Ratha N (Eds.), Audio- and Video-Based Biometric Person Authentication. Springer Berlin / Heidelberg. 3546:823-831.

Gonzalez R, Woods R (2007). Digital Image Processing (3rd Edition) Prentice Hall. p. 976.

Grosicki E, Carré M, Brodin J, Geoffrois E (2008). RIMES Evaluation Campaign for Handwritten Mail Processing. Proceedings of the 11th International Conference on Frontiers in Handwriting Recognition. Montreal, Canada.

Guyon I, Schomaker L, Plamondon R, Liberman M, Janet S (1994). UNIPEN project of on-line data exchange and recognizer benchmarks. 12th IAPR International Conference on Pattern Recognition. IEEE Computing Society Press, Jerusalem. pp. 29-33.

He Z, Bin F, Jianwei D, Yuan Yan T, Xinge Y (2005). A Novel Method for Off-line Handwriting-based Writer Identification. Eighth International Conference on Document Analysis and Recognition. IEEE. pp. 242-246.

He Z, Tang Y (2004). Chinese handwriting-based writer identification by texture analysis. Proceedings of the International Conference on Machine Learning and Cybernetics. IEEE. 6:3488-3491.

He Z, Tang Y, You X (2005). A contourlet-based method for writer identification. IEEE International Conference on System Manufacturing and Cybernetics. 1:364-368.

He Z, You X, Tang Y (2008a). Writer Identification Using Global Wavelet-Based Features. Neurocomputing 71(10-12):1832-1841.

He Z, You X, Tang Y (2008b). Writer identification of Chinese handwriting documents using hidden Markov tree model. Pattern Recognit. 41(4):1295-1307.

Helli B, Moghaddam M (2008a). Persian Writer Identification Using Extended Gabor Filter. In Campilho A and Kamel M (Eds.), Image Analysis and Recognition. Springer Berlin /Heidelberg. 5112:579-586.

Helli B, Moghaddam M (2008b). A Text-Independent Persian Writer Identification System Using LCS Based Classifier. IEEE International Symposium on Signal Processing Information and Technology. pp. 203-206.

Helli B, Moghaddam M (2009). A Writer Identification Method Based on XGabor and LCS. IEICE Electron. Express 6(10):623-629.

Helli B, Moghaddam M (2010). A Text-Independent Persian Writer Identification Based on Feature Relation Graph (FRG). Pattern Recognit. 43(6):2199-2209.

Hochberg J, Bowers K, Cannon M, Kelly P (1999). Script and Language Identification for Handwritten Document Images. Int. J. Doc. Anal. Recognit. 2(2):45-52.

Hull J (1994). A database for handwritten text recognition research. IEEE Trans. Pattern Anal. Mach. Intell. 16(5):550-554.

Int. Unipen Foundation (2011). Int. Unipen Foundation - iUF. Retrieved from http://unipen.org/products.html

Johansson S, Leech G, Goodluck H (1978). Manual of information to accompany the Lancaster-Oslo/Bergen corpus of British English, for use with digital computers. Department of English, University of Oslo, Oslo.

Kohonen T (1989). Self-Organization and Associative Memory. 3rd Edition, Springer-Verlag New York Inc, NYC, USA. p. 312.

Leedham G, Chachra S (2003). Writer Identification Using Innovative Binarised Features of Handwritten Numerals. Proceedings of the Seventh International Conference on Document Analysis and Recognition. 1:413-416.

Li X, Ding X (2009). Writer Identification of Chinese Handwriting Using Grid Microstructure Feature. In: Tistarelli M and Nixon M (Eds.), Advances in Biometrics. 5558:1230-1239.

Li X, Wang X, Ding X (2006). An Off-line Chinese Writer Retrieval System Based on Text-sensitive Writer Identification. 18th International Conference on Pattern Recognition. IEEE Computing Society. pp. 517-520.

Liu C, Dai R, Liu Y (1995). Extracting individual features from moments for Chinese writer identification. Proceedings of the Third International Conference on Document Analysis and Recognition. IEEE Computing Society. Los Alamitos, CA, USA. 1:438-441.

Maaten L, Postma E (2005). Improving automatic writer identification. Proceedings of the 17th Belgium-Netherlands Conference on Artificial Intelligence. pp. 260-266.

Mar S, Thein N (2005). Myanmar Character Identification of Handwriting between Exhibit and Specimen. Proceedings of the 6th Asia-Pacific Symposium on Information Telecommunication and Technology. pp. 95-98.

Marti U, Bunke H (2002). The IAM-Database: an English Sentence Database for Offline Handwriting Recognition. Int. J. Doc. Anal. Recognit. 5(1):39-46.

Matsuura T, Qiao Y (1989). Writer identification using an impulse response of the system characterizing handwriting motion. IEEE Colloquium Character Recognition Applications. pp. 2/1-2/8.

Märgner V, El Abed H (2007). ICDAR 2007 - Arabic Handwriting Recognition Competition. 9th International Conference on Document Analysis and Recognition. Curitiba - Paraná - Brazil. pp. 1274 - 1278.

Märgner V, El Abed H (2009). ICDAR 2009 Arabic Handwriting Recognition Competition. 10th Int. Conf. Doc. Anal. Recognit. (ICDAR) IEEE. pp. 1383-1387.

Märgner V, El Abed H (2010). ICFHR 2010 - Arabic Handwriting Recognition Competition. 12th International Conference on Frontiers in Handwriting Recognition (ICFHR) IEEE. pp. 709-714.

Märgner V, El Abed H (2011). ICDAR 2011 Arabic Handwriting Recognition Competition. To appear in the 11th International Conference on Document Analysis and Recognition (ICDAR). Beijing, China.

Märgner V, Pechwitz M, Abed H (2005). ICDAR 2005 Arabic handwriting recognition competition. Eighth Int. Conf. Doc. Anal. Recognit. 1:70-74.

Niels R, Vuurpijl L, Schomaker L (2007). Automatic allograph matching in forensic writer identification. Int. J. Pattern Recognit. Artif. Intell. 21:61-81.

Panagopoulos M, Papaodysseus C, Rousopoulos P, Dafi D, Tracy S (2009). Automatic Writer Identification of Ancient Greek Inscriptions. IEEE Trans. Pattern Anal. Mach. Intell. IEEE Comput. Soc. 31(8)1404-14.

Pechwitz M, Maddouri S, Märgner V, Ellouze N, Amiri H (2002). IFN/ENIT - Database of Handwritten Arabic Words. 7th Colloque International Francophone sur l'Ecrit et le Document, Hammamet, Tunis. pp. 129-136.

Plamondon R (1994). Progress in Automatic Signature Verification. World Scientific Publishing Co., Inc. River Edge, NJ, USA. p. 180.

Plamondon R, Lorette G (1989). Automatic Signature Verification and

Writer Identification — State art. Pattern Recognit. 22(2):107-131.
Plamondon R, Srihari S (2000). Online and Off-Line Handwriting Recognition: a Comprehensive Survey. IEEE Trans. Pattern Anal. Mach. Intell. 22(1):63-84.
Ram S, Moghaddam M (2009a). A Persian Writer Identification Method Based on Gradient Features and Neural Networks. 2nd International Congress on Image Signal Processing. IEEE. pp. 1-4.
Ram S, Moghaddam M (2009b). Text-independent Persian Writer Identification Using Fuzzy Clustering Approach. 2009 Int. Conf. Inform. Manage. Eng. IEEE. pp. 728-731.
Rodríguez-Serrano J, Perronnin F, Sánchez G, Llados J (2010). Unsupervised Writer Adaptation of Whole-Word HMMs with Application to Word-Spotting. Pattern Recognit. Lett. 31(8):742-749.
Said H, Baker J, Tan T (1998). Personal identification based on handwriting. Proc. Int. Conf. Pattern Recognit. 2:1761-1764.
Schlapbach A (2007). Writer Identification and Verification. Ph.D. dissertation, Institute of Computer Science and Application in Mathematics. Bern University, Netherlands, The Netherlands. p. 161.
Schlapbach A, Bunke H (2004a). Using HMM Based Recognizers for Writer Identification and Verification. Ninth International Workshop on Frontier Handwriting Recognition. 9:167- 172.
Schlapbach A, Bunke H (2004b). Off-Line Handwriting Identification Using HMM Based Recognizers. Proc. Int. Conf. Pattern Recognit. 2:654-658.
Schlapbach A, Bunke H (2006). Off-line Writer Identification Using Gaussian Mixture Models. Int. Conf. Pattern Recognit. 3:992-995.
Schlapbach A, Bunke H (2007). A Writer Identification and Verification System Using HMM Based Recognizers. Pattern Anal. Appl. 10(1):33-43.
Schlapbach A, Kilchherr V, Bunke H (2005). Improving writer identification by means of feature selection and extraction. Proc. Int. Conf. Doc. Anal. Recognit. 1:131-135.
Schomaker L, Bulacu M (2004). Automatic Writer Identification Using Connected-Component Contours and Edge-Based Features of Uppercase Western Script. IEEE Trans. Pattern Anal. Mach. Intell. 26(6):787-798.
Schomaker L, Bulacu M, Franke K (2004). Automatic writer identification using fragmented connected-component contours. 9th International Workshop on Frontiers Handwriting Recognition. pp. 185-190.
Schomaker L, Bulacu M, van Erp M (2003). Sparse-Parametric Writer Identification Using Heterogeneous Feature Groups. Proceedings of the International Conference on Image Processing. 1:545-8.
Schomaker L, Franke K, Bulacu M (2007). Using codebooks of fragmented connected-component contours in forensic and historic writer identification. Pattern Recognit. Lett. 28(6):719-727.
Schomaker L, Vuurpijl L (2000). Forensic Writer Identification: A Benchmark Data Set and a Comparison of Two Systems (research report). Nijmegen.
Shahabi F, Rahmati M (2006). Comparison of Gabor-Based Features for Writer Identification of Farsi/Arabic Handwriting. 10th International Workshop on Frontiers Handwritten Recognition. pp. 550-545.
Shahabi F, Rahmati M (2007). A New Method for Writer Identification and Verification Based on Farsi/Arabic Handwritten Texts. 9th International Conference on Document Analysis and Recognition. 2:829-833.
Siddiqi I, Vincent N (2007). Writer Identification in Handwritten Documents. 9th International Conference Document Analysis Recognition, IEEE. pp. 108-112.
Siddiqi I, Vincent N (2008). Combining Global and Local Features for Writer Identification. Proceedings of the Eleventh International Conference on Frontiers in Handwriting Recognition. Montreal, Canada. Pp. 48-53
Siddiqi I, Vincent N (2009). A Set of Chain Code Based Features for Writer Recognition. 10th International Conference on Document Analysis Recognition, IEEE. pp. 981-985.
Srihari S (2000). Distance between histograms of angular measurements and its application to handwritten character similarity. Proceedings of the 15th International Conference on Pattern Recognition. IEEE Computing Society. pp. 21-24.
Srihari S, Ball G (2008). Writer Verification of Arabic Handwriting. The Eighth IAPR International Workshop on Document Analysis Systems. pp. 28-34.
Srihari S, Ball G (2009). Comparison of Statistical Models for Writer Verification. Proceedings on Document Recognition and Retrieval XVI San Jose, CA, USA. pp. 7247OE 1-8.
Srihari S, Beal M, Bandi K, Shah V, Krishnamurthy P (2005). A Statistical Model for Writer Verification. Proc. Doc. Anal. Recognit. 2:1105-1109.
Srihari S, Cha S, Arora H, Lee S (2002). Individuality of Handwriting. J. Forensic Sci. 47(4):1-17.
Srihari S, Huang C, Srinivasan H, Shah V (2007). Biometric and Forensic Aspects of Digital Document Processing. Digital Document Processing. Springer, London. pp. 379-405.
Srihari S, Srinivasan H, KD. (2007). Questioned Document Examination using CEDAR-FOX. J. Forensic Doc. Examination 18(2):1-20.
Su T, Zhang T, Guan D (2007). Corpus-based HIT-MW database for offline recognition of general-purpose Chinese handwritten text. Int. J. Doc. Anal. Recognit. 10(1):27-38.
Tan G, Viard-Gaudin C, Kot A (2008). Online Writer Identification Using Fuzzy C-means Clustering of Character Prototypes. International Conference on Frontiers Handwriting Recognit. Montréal, Canada. p. 6.
Tomai C, Srihari S (2004). Discriminatory Power of Handwritten Words for Writer Recognition. Proceed. 17th International Conference on Pattern Recognition. IEEE. 2:638-641.
Ubul K, Hamdulla A, Aysa A, Raxidin A, Mahmut R (2008). Research on Uyghur off-line handwriting-based writer identification. 9th International Conference on Signal Process. IEEE. pp. 1656-1659.
Viard-Gaudin C, Lallican P, Binter P, Knerr S (1999). The IRESTE On/Off (IRONOFF) Dual Handwriting Database. 5th International Conference on Document Analysis Recognit. pp. 455-458.
Wang X, Ding X, Liu H (2003). Writer Identification Using Directional Element Features and Linear Transform. Proceedings of the 7th International Conference on Document Analysis Recognition. IEEE Computer Society. pp. 942 - 945
Yoshimura I (1988). Writer identification based on the arc pattern transformation. 9th International Conference on Pattern Recognition. 1:35-37
Zaher A, Abu-Rezq A (2010). A Hybrid ANN-Based Technique for Signature Verification. Proceedings of the 4th WSEAS International Conference on Computational Intelligence. Bucharest, Romania. pp. 13-19.
Zhang B (2003). Handwriting Pattern Matching and Retrieval with Binary Features. Ph.D. dissertation, Department of Computer Science and Engineering, State University of New York, Buffalo, NY. p. 172.
Zhang B, Srihari S, Lee S (2003). Individuality of Handwritten Characters. 7th International Conference on Document Analysis and Recognition. Proc. IEEE Comput. Soc. 1086-1090
Zhang D, Jain A, Tapiador M, Sigüenza J (2004). Writer Identification Method Based on Forensic Knowledge - Biometric Authentication. Lect. Notes Comput. Sci. 3072:555-561.
Zhu Y, Tan T, Wang Y (2000). Biometric Personal Identification Based on Handwriting. Proceedings. 15th International Conference on Pattern Recognition. IEEE Computing Society./ 2:797-800
Zois E, Anastassopoulos V (2000). Morphological Waveform Coding for Writer Identification. Pattern Recognit. 33(3):385-398.