Text-Independent Writer Identification and Verification on Offline Arabic Handwriting

Marius Bulacu Lambert Schomaker Axel Brink
Artificial Intelligence Institute, University of Groningen, The Netherlands
(bulacu, schomaker, a.a.brink)@ai.rug.nl

Abstract

In this paper, we evaluate the performance on Arabic handwriting of the text-independent writer identification methods that we developed and tested on Western script in recent years. We use the IFN/ENIT data in the experiments reported here and our tests involve 350 writers. The results show that our methods are very effective and the conclusions drawn in previous studies remain valid also on Arabic script. High performance is achieved by combining textural features (joint directional probability distributions) with allographic features (grapheme-emission distributions).

1. Introduction

Two important natural factors are in direct conflict in the attempt to identify a person based on samples of handwriting: *between-writer variation* as opposed to *within-writer variability*. Therefore, in automatic writer identification, it is necessary to use computer representations (features) with the ability to maximize the separation between different writers, while remaining stable over samples produced by the same writer. In recent years, we proposed a number of new and very effective statistical features for automatic writer identification using offline handwriting [2, 3, 8]. Our features are probability distribution functions (PDFs) extracted from handwritten text blocks and characterize writer individuality *independently of the textual content* of the written samples. In our methods, the computer is completely agnostic of the actual text written in the samples.

Two fundamental sources of information regarding the individuality of handwriting are exploited by our techniques functioning at two levels of analysis. First, handwriting slant, curvature and roundness, as determined by habitual pen grip, are captured by joint directional probability distributions operating at the *texture* level. Second, the personalized set of letter shapes, *allographs*, that a writer has learned to use under educational, cultural and memetic influences is captured by a grapheme-emission probability distribution operating at the character level. By combining texture-level and allograph-level features, we achieved very high writer identification and verification performance in extensive tests carried out using large datasets (containing up to 900 subjects) of Western handwriting [3].

The purpose of the current paper is to test the effectiveness of our features on Arabic script using the IFN/ENIT dataset [5]. In our experimental evaluation, we will consider both tasks of *writer identification* (one-to-many search in a handwriting database with the return of a likely list of candidates) and *writer verification* (one-to-one comparison with an automatic decision whether or not the two samples were written by the same person).

Research in automatic writer identification has received renewed attention in the last several years [1, 7, 10]. However, despite this increased interest, writer identification on Arabic handwriting has been studied surprisingly little until the present. The only paper we could find that directly treats this topic is [9], where the authors determine the performance of Gabor-based features (initially proposed in [6]) on a Farsi dataset comprising 25 writers.

In this paper, we compactly describe our features and comprehensively evaluate their writer identification and verification performance on the IFN/ENIT Arabic data. We also consider the problem of combining features for improved results. Further, we show how the identification rate depends on two factors: 1) the number of writers and 2) the number of samples per writer contained in the test set.

2. Experimental dataset

The IFN/ENIT database [5] consists of forms with handwritten Arabic town/village names collected from 411 subjects (binary images at 300 dpi resolution). Most writers filled in 5 forms. This dataset was designed for training / testing recognition systems for handwritten words and was used for the ICDAR 2005 Arabic OCR competition [4].

The IFN/ENIT data can be used also for writer identification because the writer information was also recorded. We used some fixed cutting coordinates to extract the handwrit-

[0]This paper was published as: Marius Bulacu, Lambert Schomaker, Axel Brink, *Text-independent writer identification and verification on offline Arabic handwriting*, Proc. of 9th Int. Conference on Document Analysis and Recognition (ICDAR 2007), IEEE Computer Society, 2007, pp. 769-773, vol. II, 23 - 26 September, Curitiba, Brazil

Table 1. Overview of features and their dimensionalities.

	Feature	Explanation	Dim
f1	$p(\phi)$	Contour-direction PDF	12
f2	$p(\phi_1, \phi_2)$	Contour-hinge PDF	300
		Direction co-occurrence PDFs	
f3h	$p(\phi_1, \phi_3)$ h	- horizontal run	144
f3v	$p(\phi_1, \phi_3)$ v	- vertical run	144
f4	$p(g)$	Grapheme emission PDF	400
		Run-length on white PDFs	
f5h	$p(rl)$ h	- horizontal run	60
f5v	$p(rl)$ v	- vertical run	60

ing from the scanned forms. The text content is variable and the samples contain a limited amount of handwriting: only 12 names of Tunisian towns/villages. We have split the dataset into two parts. The handwriting from 61 writers was used to train the shape codebook used in our allograph-level method, as will be shown further. The largest part of the dataset, 350 writers with 5 samples per writer, was used in the writer identification and verification tests.

3. Feature extraction methods

An overview of all features used in this paper is given in table 1. The term "feature" denotes a complete PDF (an entire vector of probabilities). We have designed features *f2*, *f3* and *f4*, while features *f1*, *f5* are classically known. Please refer to previous papers [3, 2, 8] for more details.

The primary binary images are processed by extracting the connected components and their inner and outer contours (using Moore's algorithm). Our methods work at two levels of analysis: the *texture level* and the *allograph level*.

3.1. Textural features

In these features, the handwriting is merely seen as a texture described by some probability distributions computed from the image and capturing the distinctive visual appearance of the written samples.

The distribution of directions in handwriting provides useful information for writer identification. The directional PDF can be computed very fast using the contours by considering the orientation of local contour fragments determined by two contour pixels taken a certain distance apart (Fig. 1). As the algorithm runs over the contours, the angle that the analyzing fragment makes with the horizontal is computed using equation 1 and an angle histogram is built thereby. This histogram is then normalized to a probability distribution $p(\phi)$ that constitutes the feature (*f1*) used in writer identification and verification.

$$\phi = arctan(\frac{y_{k+\epsilon} - y_k}{x_{k+\epsilon} - x_k}) \qquad (1)$$

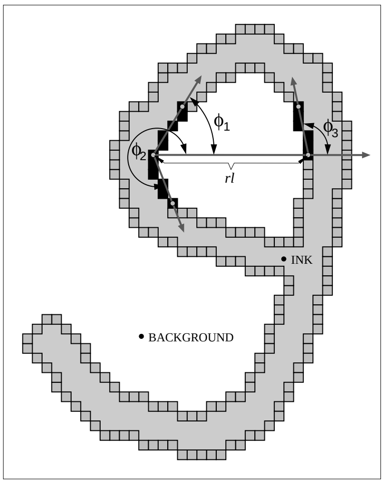

Figure 1. Schematic description for the feature extraction methods of directional and run-length PDFs.

In our implementation $\epsilon = 5$ and this value was selected such that the length of the contour fragment is comparable to the thickness of the ink trace (6 pixels). The number of histogram bins spanning the interval 0° - 180° was set to $n = 12$ through experimentation. These settings will be used for all the directional features.

The directional PDF $p(\phi)$ was our starting point in designing more complex and more effective features. In order to capture the curvature of the ink trace, which is very discriminatory between different writers, we designed the "hinge" feature. The central idea is to consider, not one, but two contour fragments attached at a common end pixel and then compute the joint PDF of the orientations of the two legs of the "contour-hinge" (Fig. 1). The feature $p(\phi_1, \phi_2)$ is therefore a bivariate PDF capturing both the orientation and the curvature of contours.

Building upon the same idea of combining oriented contour fragments, we designed another feature: the directional co-occurrence PDF. For this feature, we consider the combinations of contour-angles occurring at the ends of run-lengths on the background (Fig. 1). The joint PDF $p(\phi_1, \phi_3)$ of the two contour-angles occurring at the ends of a white run-length captures longer range correlations between script directions and gives a measure of the roundness of handwriting. Horizontal runs along the image rows generate *f3h* and vertical runs along the image columns generate *f3v*.

Run lengths are determined on the binary image taking into consideration either the black pixels (the ink) or the white pixels (the background). We consider the white runs that capture the regions enclosed inside letters and also the empty spaces between letters and words. There are two basic scanning methods: horizontal along the image rows (*f5h*) and vertical along the image columns (*f5v*). Similarly to the directional features presented above, the histogram of run lengths is normalized and interpreted as a PDF.

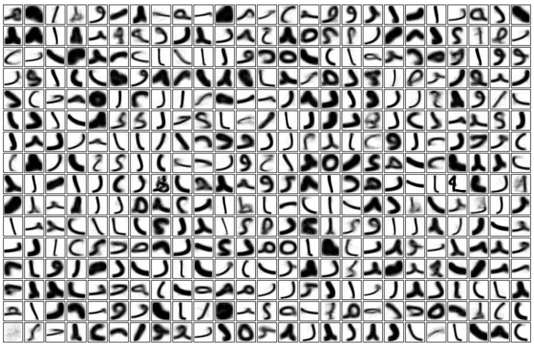

Figure 2. Shape codebook generated by k-means clustering and containing 400 Arabic graphemes.

3.2. Allographic features

We assume that every writer is a stochastic generator of ink-blob shapes (graphemes) [8, 1]. The PDF of grapheme usage in a given sample is characteristic of each writer and is computed using a common shape codebook obtained by clustering [2]. To make this approach applicable to free-style handwriting (cursive and isolated), a segmentation method is used yielding graphemes (sub- or supra-allographic fragments) that often will not overlap a whole character. This method involves three processing stages:

1) Handwriting segmentation: the ink is cut at the minima in the lower contour for which the distance to the upper contour is comparable to the ink-trace width. The graphemes are then extracted as connected components, followed by size normalization to 30x30 pixel bitmaps.

2) Shape codebook generation: clustering was applied to a training set containing 35k graphemes extracted from the handwriting of 61 writers. We will compare three clustering methods: k-means, Kohonen self-organizing maps (KSOM) 1D and 2D. The size of the codebook was set to 400 (20x20) shapes. This value was used also in our previous studies [2]. Fig. 2 shows the shape codebook generated by k-means clustering. The codebook graphemes act as prototype shapes representative for the types of shapes to be expected as a result of handwriting segmentation.

3) Grapheme-usage PDF computation: one bin is allocated to every grapheme in the codebook and a shape occurrence histogram is computed for every handwritten sample. For every ink fraglet extracted from a sample after segmentation, the nearest codebook grapheme g is found using Euclidean distance and this occurrence is counted into the corresponding histogram bin. The histogram is normalized to a PDF $p(g)$ that acts as the writer descriptor.

The perfect segmentation of individual characters in free-style script is unachievable and this represents a fundamental problem for handwriting recognition. Nevertheless, the ink fraglets generated by our imperfect segmentation can still be effectively used for writer identification.

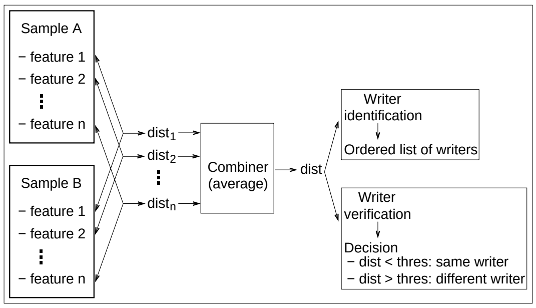

Figure 3. Feature fusion method: the distances generated by the individual features are averaged (using simple or weighted average) and the result is then used in writer identification and verification.

4. Feature matching and fusion for writer identifi cation and verifi cation

Writer identification is performed using nearest-neighbor classification in a "leave-one-out" strategy: one sample is chosen as the query and all the other samples from the test set (350 x 5 - 1 = 1749) are ordered with increasing distance from the query, using a selected feature. Ideally the first ranked (Top-1) sample should be one of the other 4 samples produced by the writer of the query. If a longer hit list is considered (Top-10) the chance of finding the correct writer increases with the list size. The χ^2 distance is used in matching the individual features. This represents a natural choice for our PDFs and also it performed best in our tests.

Writer verification is performed in the classical Neyman-Pearson framework of statistical decision theory. By varying the decision threshold, Receiver Operating Characteristic (ROC) curves are computed for all features. The Equal Error Rate (EER) is used to quantify in a single number the writer verification performance.

The considered features capture different aspects of handwriting individuality and operate at different scales. Combining features yields improved performance. In our feature combination scheme, the final unique distance between any two handwritten samples is computed as the average (simple or weighted) of the distances due to the individual features participating in the combination (Fig. 3). In feature combinations, Hamming distance performed best.

5. Results

Table 2 gives the writer identification and verification performance of the individual features considered here. The best performing feature is the contour-hinge PDF (feature *f2*: Top-1 82%, Top-10 97%, EER 7.5%), followed by the grapheme PDF (feature *f4*: Top-1 60%, Top-10 90%, EER 11.0%). The same performance is achieved by the three clustering methods (kmeans, ksom1D and ksom2D) used

Table 2. Writer identification and verification performance of individual features on the IFN/ENIT dataset of Arabic handwriting (350 writers, 5 samples per writer). The features are explained in Table 1.

	Feature	Identification		Verification
		Top 1	Top 10	EER
f1	$p(\phi)$	31	70	14.4
f2	$p(\phi_1, \phi_2)$	82	97	7.5
f3h	$p(\phi_1, \phi_3)$ h.	38	75	17.3
f3v	$p(\phi_1, \phi_3)$ v.	39	74	15.6
	- kmeans	61	89	11.0
f4	$p(g)$ - ksom1D	60	89	11.3
	- ksom2D	59	90	11.1
f5h	$p(rl)$ h.	3	19	29.4
f5v	$p(rl)$ v.	3	19	29.6

for generating the grapheme codebook. This behavior was observed previously in our studies on Western script [2].

The angle combination features *f2*, *f3h* and *f3v* perform better than the basic directional PDF *f1*. We obtain thus a confirmation also on Arabic script that joint PDFs capture more individuality information from the handwriting. Despite their higher dimensionality, reliable probability estimates can be obtained from samples containing a reduced amount of ink, this being the case in the IFN/ENIT set.

The run length PDFs have the worst performance among the considered features. Nevertheless, they provide additional information that will used in feature combinations.

The features studied in the paper can be grouped into 3 broad categories (see table 1): contour-based directional PDFs (*f1*, *f2*, *f3h*, *f3v*), grapheme emission PDF (*f4*) and run-length PDFs (*f5h*, *f5v*). We analyzed combinations of features within and between these broad feature groups. As stated earlier, feature fusion is performed by distance averaging. Assigning distinct weights to the different features participating in the combination yielded significant performance improvements only for the combination *f2* & *f4*. For the other combinations, we preferred simplicity / robustness and used plain distance averaging.

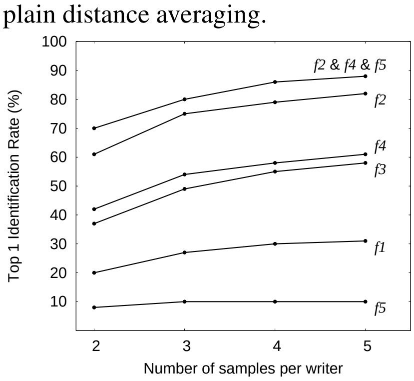

Figure 4. Top-1 identification rate vs. number of samples per writer contained in the dataset.

Table 3. Writer identification and verification performance of feature combinations on the IFN/ENIT dataset.

Feature combination	Identification		Verification
	Top 1	Top 10	EER
f3: *f3h* & *f3v*	58	87	12.4
f5: *f5h* & *f5v*	10	38	23.3
f1 & *f4*	71	94	7.6
f1 & *f5*	41	81	13.3
f2 & *f4*	86	98	5.6
f3 & *f4*	80	97	7.4
f3 & *f5*	63	91	11.1
f4 & *f5*	69	93	10.1
f1 & *f4* & *f5*	76	96	7.5
f2 & *f4* & *f5*	88	99	5.8
f3 & *f4* & *f5*	84	98	7.5

Features *f3* and *f5* (first two rows of table 3) are obtained by combining the two orthogonal directions of scanning the input image. They perform markedly better compared to their single horizontal or vertical counterparts.

The experiments showed that improvements are obtained by combining features from different feature groups (see table 3). The best performing feature combination fuses directional, grapheme and run-length information yielding identification rates of Top-1 88% and Top-10 99% with an EER around 5-6% in verification.

Figure 4 shows how identification rate depends on the number of samples per writer: as every writer has more enrolled samples, the chance of a correct hit increases, despite the fact that the number of distractors involved in our leave-one-out test also has increased. The returns in performance are however diminishing for every new sample added.

Figure 5 shows the identification rate as a function of the number of writers involved in the test. Naturally, the identification rate decreases as the number of writers grows. However, the decline is not severe for the feature combination *f2* & *f4* & *f5*: Top-1 identification rate drops by $\sim 2.5\%$ for every doubling of the number of writers in the dataset.

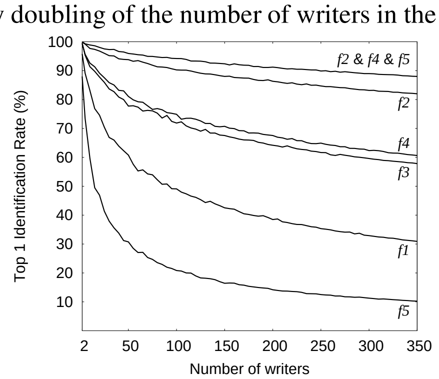

Figure 5. Top-1 identification rate vs. number of writers contained in the test. For every size of the writer set, the results are averaged over fifty random draws.

Figure 6. A successful writer identification hit list generated by GRAWIS using the best performing feature combination *f2* & *f4* & *f5*. The query sample is in the top-center position and the enrolled samples produced by the same writer are marked with a darker border (ranks 1, 2, 5 and 8 in the hit list).

Fig. 6 shows a successful hit list generated by our system, named GRAWIS from Groningen Automatic Writer Identification System.

The identification and verification results obtained on Arabic script and reported here cannot be numerically compared with our previous results for Western script because the experimental datasets are different (in terms of the amount of ink contained in the samples among others). Nevertheless, it seems that the results obtained on Arabic are somewhat lower than the ones obtained on Western script. A visual analysis of our data also suggests that, pictorially, there seems more style variation across individuals in Western handwritings (especially due to slant) compared to Arabic ones. Automatic writer identification on Arabic script appears to be more difficult.

6. Conclusions

The gist of our text-independent approach to writer identification and verification is constituted by the contour-based angle-combination PDFs (*f2*, *f3h*, *f3v*) and the grapheme-emission PDF (*f4*). As observed also earlier, these state-of-the-art features outperform other text-independent methods. Combining textural and allographic features yields very high writer identification rates for datasets containing hundreds of writers.

The observations that we have made in previous studies on Western script have been confirmed also Arabic handwriting. Our statistical methods have a generic nature and generate robust and stable results.

References

[1] A. Bensefia, T. Paquet, and L. Heutte. A writer identification and verification system. *Pattern Recognition Letters*, 26(10):2080–2092, October 2005.

[2] M. Bulacu and L. Schomaker. A comparison of clustering methods for writer identification and verification. In *Proc. of 8th ICDAR*, pp 1275–1279, Seoul, Korea, 2005.

[3] M. Bulacu and L. Schomaker. Combining multiple features for text-independent writer identification and verification. In *Proc. of 10th IWFHR*, pp 281–286, La Baule, France, 2006.

[4] V. Märgner, M. Pechwitz, and H. El Abed. ICDAR 2005 arabic handwriting recognition competition. In *Proc. of 8th ICDAR*, pp 70–74, Seoul, Korea, 2005.

[5] M. Pechwitz, S. Maddouri, V. Märgner, N. Ellouze, and H. Amiri. IFN/ENIT-database of handwritten arabic words. In *Proc. of CIFED*, pp 129–136, Hammamet, Tunisia, 2002.

[6] H. Said, T. Tan, and K. Baker. Personal identification based on handwriting. *Pattern Recognition*, 33(1):149–160, 2000.

[7] A. Schlapbach and H. Bunke. Off-line writer verification: a comparison of a HMM and a GMM based system. In *Proc. of 10th IWFHR*, pp 275–280, La Baule, France, 2006.

[8] L. Schomaker and M. Bulacu. Automatic writer identification using connected-component contours and edge-based features of uppercase western script. *IEEE Trans. on PAMI*, 26(6):787–798, June 2004.

[9] F. Shahabi and M. Rahmati. Comparison of Gabor-based features for writer identification of Farsi/Arabic handwriting. In *Proc. of 10th IWFHR*, pp 545–550, La Baule, France, 2006.

[10] S. Srihari, S. Cha, H. Arora, and S. Lee. Individuality of handwriting. *J. of Forensic Sciences*, 47(4):1–17, July 2002.