Text-Independent Writer Identification and Verification Using Textural and Allographic Features

Marius Bulacu, *Student Member*, *IEEE*, and Lambert Schomaker, *Member*, *IEEE*

Abstract—The identification of a person on the basis of scanned images of handwriting is a useful biometric modality with application in forensic and historic document analysis and constitutes an exemplary study area within the research field of behavioral biometrics. We developed new and very effective techniques for automatic writer identification and verification that use probability distribution functions (PDFs) extracted from the handwriting images to characterize writer individuality. A defining property of our methods is that they are designed to be independent of the textual content of the handwritten samples. Our methods operate at two levels of analysis: the *texture level* and the *character-shape (allograph) level*. At the *texture level*, we use contour-based joint directional PDFs that encode orientation and curvature information to give an intimate characterization of individual handwriting style. In our analysis at the *allograph level*, the writer is considered to be characterized by a stochastic pattern generator of ink-trace fragments, or graphemes. The PDF of these simple shapes in a given handwriting sample is characteristic for the writer and is computed using a common shape codebook obtained by grapheme clustering. Combining multiple features (directional, grapheme, and run-length PDFs) yields increased writer identification and verification performance. The proposed methods are applicable to free-style handwriting (both cursive and isolated) and have practical feasibility, under the assumption that a few text lines of handwritten material are available in order to obtain reliable probability estimates.

Index Terms—Handwriting analysis, writer identification and verification, behavioral biometrics, joint directional probability distributions, grapheme-emission probability distribution.

1 Introduction

This paper addresses the problem of automatic person identification on the basis of scanned images of handwriting. We present a number of new and very effective statistical pattern recognition methods for automatic writer identification and verification using offline handwriting. Our methods are experimentally evaluated using data sets with handwriting samples collected from up to 900 subjects. There are two distinguishing characteristics of our approach: *human intervention is minimized* in the writer identification process and we encode individual handwriting style using *features designed to be independent of the textual content* of the handwritten sample. Writer individuality is encoded using probability distribution functions extracted from handwritten text blocks and, in our methods, the computer is completely unaware of what has been written in the samples. The development of our writer identification techniques takes place at a time when many biometric modalities are undergoing a transition from research to real full-scale deployment. Our methods also have practical feasibility and hold the promise of concrete applicability.

Physiological biometrics (e.g., iris, fingerprint, hand geometry, retinal blood vessels, DNA) are strong modalities for person identification due to the reduced variability and high complexity of the biometric templates used. However, these physiological modalities are usually more invasive and require cooperating subjects. On the contrary, *behavioral biometrics* (e.g., voice, gait, keystroke dynamics, signature, handwriting) are less invasive, but the achievable identification accuracy is less impressive due to the large variability of the behavior-derived biometric templates. Writer identification pertains to the category of behavioral biometrics and has applicability in the forensic and historic document analysis fields.

Writer identification is rooted in the older and broader domain of automatic handwriting recognition [1], [2]. For automatic handwriting recognition, invariant representations are sought which are capable of eliminating variations between different handwriting in order to classify the shapes of characters and words robustly. The problem of writer identification, on the contrary, requires a specific enhancement of these variations, which are characteristic to a writer's hand. Handwriting recognition and writer identification therefore represent two opposing facets of handwriting analysis. It is important, however, to also mention the idea that writer identification could aid the recognition process if information on the writer's general writing habits and idiosyncrasies is available to the handwriting recognition system.

Research in writer identification and verification has received significant interest in recent years due to its forensic applicability (e.g., the case of the anthrax letters). A *writer identification* system performs a one-to-many search in a large database with handwriting samples of known authorship and returns a likely list of candidates (see Fig. 1a1). This represents a special case of image retrieval, where the retrieval process is based on features capturing handwriting individuality. The

- *The authors are with the AI Institute, University of Groningen, Grote Kruisstraat 2/1, 9712 TS, Groningen, The Netherlands.*
E-mail: {bulacu, schomaker}@ai.rug.nl.

Manuscript received 30 Jan. 2006; revised 24 May 2006; accepted 25 May 2006; published online 18 Jan. 2007.
Recommended for acceptance by S. Prabhakar, J. Kittler, D. Maltoni, L. O'Gorman, and T. Tan.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org and reference IEEECS Log Number TPAMISI-0056-0106.
Digital Object Identifier no. 10.1109/TPAMI.2007.1009.

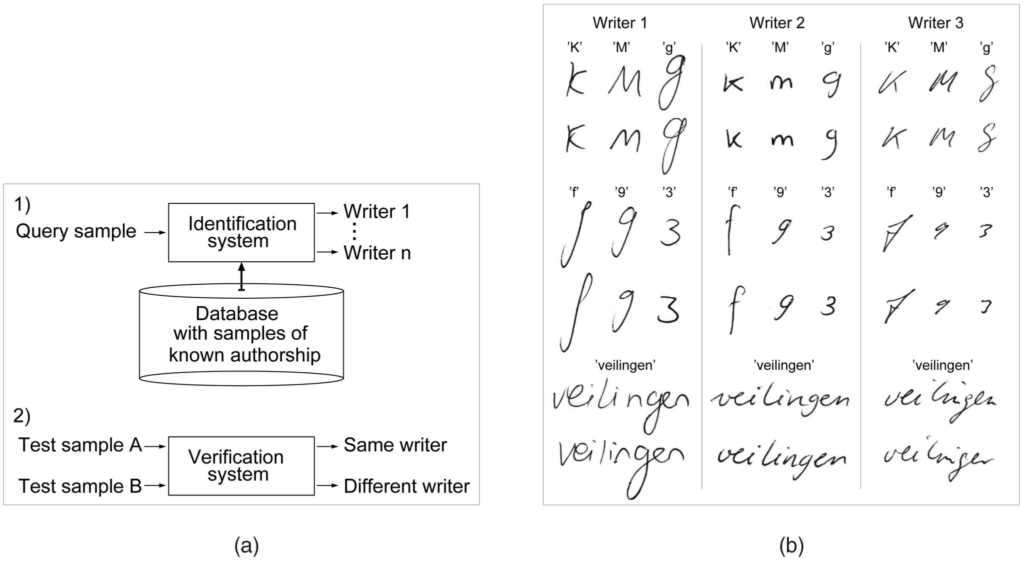


Fig. 1. (a1) A *writer identification* system retrieves, from a database containing handwritings of known authorship, those samples that are the most similar to the query. The hit list is then analyzed in detail by a human expert. (a2) A *writer verification* system compares two handwriting samples and makes an automatic decision as to whether or not the input samples were written by the same person. (b) A comparison of handwritten characters (allographs) and handwritten words from three different writers. The between-writer variation exceeds the within-writer variability and provides the basis for writer identification and verification.

hit list is further scrutinized by the forensic expert who makes the final decision regarding the identity of the author of the questioned sample. Writer identification is therefore possible only if there exist previous samples of handwriting by that person enrolled in the forensic database. *Writer verification* involves a one-to-one comparison with a decision as to whether or not the two samples are written by the same person (see Fig. 1a2). The decidability of this problem gives insight into the nature of handwriting individuality. Writer verification has potential applicability in a scenario in which a specific writer must be automatically detected in a stream of handwritten documents. The *target performance* for forensic writer identification systems is a near 100 percent recall of the correct writer in a hit list of 100 writers, computed from a database on the order of 10k samples, which is the size of the current European forensic databases. This target performance still remains an ambitious goal. Contrary to other forms of biometric person identification used in forensic labs, automatic writer identification often allows for determining identity in conjunction with the intentional aspects of a crime, such as in the case of threat or ransom letters. This is a fundamental difference from other biometric methods, where the relation between the evidence material and the details of an offense can be quite remote.

Writer identification and verification are only possible to the extent that the variation in handwriting style between different writers exceeds the variations intrinsic to every single writer considered in isolation (see Fig. 1b). The results reported in this paper ultimately represent a statistical analysis of the relationship opposing the *between-writer* variability and the *within-writer* variability in feature space. The present study assumes that the handwriting was produced using a natural writing attitude. Forged or disguised handwriting is not addressed in our approach. The forger tries to change the handwriting style, usually by changing the slant and/or the chosen letter shapes. Using detailed manual analysis, forensic experts are sometimes able to correctly identify a forged handwritten sample [3], [4]. On the other hand, our proposed algorithms operate on the scanned handwriting faithfully, considering all graphical shapes encountered in the image under the premise that they are created by the habitual and natural script style of the writer.

With regard to the theoretical basis of our approach, handwriting can be described as a hierarchical psychomotor process: At a high level, an abstract motor program is recovered from long-term memory; parameters are then specified for this motor program, such as size, shape, timing; finally, at a peripheral level, commands are generated for the biophysical muscle-joint systems [5]. The writer tries to maintain his/her preferred slant and letter shapes over the complete range of motion in the biomechanical systems thumb-fingers and hand-wrist [5] and in a manner that is also independent of changes in the horizontal progression motion [6]. Due to neural and neuromechanical propagation delays, a handwriting process based upon a continuous feedback mechanism alone would evolve too slowly [7]. Therefore, handwriting is not a feedback process; the brain is continuously planning series of ballistic movements ahead in time in a feed-forward manner and a character is assumed to be produced by a "motor program" [8]. Every person uses personalized and characteristic shapes, called *allographs*, when writing a chosen letter of the alphabet (see Fig. 1b). In this paper, we propose writer identification methods that aim to capture peripheral and also more central aspects of the writing behavior of an individual. Our methods operate at two levels of analysis: the *texture level* and the *allograph (character-shape) level*. The *texture-level features* are informative for the habitual pen-grip and preferred writing slant, while the *allograph-level features* reveal the character shapes engrained in the motor memory of the writer as a result of educational, cultural, and memetic factors [9]. Furthermore, very effective writer identification and verification is achievable by combining texture-level and allograph-level features that together offer a fuller description of a person's stable and discriminatory unconscious practices in writing.

The paper is organized as follows: In Section 2, we survey the recent research work on offline writer identification and verification. Section 3 describes our experimental data sets. Sections 4 and 5 describe the algorithms for extracting the texture-level and the allograph-level features, respectively. The distances used for feature matching and the feature fusion technique are explained in Section 6. Section 7 gives the experimental results, followed by a discussion in Section 8. Conclusions are then drawn in Section 9.

2 A Survey of Recent Research in the Field

A comprehensive review covering the research work in automatic writer identification until 1989 is given in [10]. Here, we will survey the approaches proposed in the last several years as a result of the renewed interest in the scientific community for this research topic.

Writer identification and verification methods fall into two broad categories: *text-dependent* versus *text-independent* methods. The *text-dependent* methods are very similar to signature verification techniques and use the comparison between individual characters or words of known semantic (ASCII) content (see Fig. 1b). These methods therefore require prior localization and segmentation of the relevant information, which is usually performed interactively by a human user. The *text-independent* methods for writer identification and verification use statistical features extracted from the entire image of a text block. A minimal amount of handwriting (e.g., a paragraph containing a few text lines) is necessary in order to derive stable features insensitive to the text content of the samples. Our approach falls in this latter category. From the application point of view, the notable advantage is that human intervention is minimized.

Said et al. [11], [12] propose a text-independent approach and derive writer-specific texture features using multichannel Gabor filtering and gray-scale co-occurrence matrices. The method requires uniform blocks of text that are generated by word deskewing, setting a predefined distance between text lines/words and text padding. Two sets of 20 writers, 25 samples per writer are used in the evaluation. Nearest centroid classification using weighted Euclidean distance and Gabor features achieved 96 percent writer identification accuracy. A similar approach has also been used on machine-print documents for script [13] and font [14] identification.

Zois and Anastassopoulos [15] perform writer identification and verification using single words. Experiments are performed on a data set containing 50 writers. The word "characteristic" was written 45 times by each writer, both in English and in Greek. After image thresholding and curve thinning, the horizontal projection profiles are resampled, divided into 10 segments, and processed using morphological operators at two scales to obtain 20-dimensional feature vectors. Classification is performed using either a Bayesian classifier or a multilayer perceptron. Accuracies around 95 percent are obtained both for English and Greek words.

Srihari et al. [16], [17] propose a large number of features divided into two categories. Macrofeatures operate at document/paragraph/word level: gray-level entropy and threshold, number of ink pixels, number of interior/exterior contours, number of four-direction slope components, average height/slant, paragraph aspect ratio and indentation, word length, and upper/lower zone ratio. Microfeatures operate at word/character level: gradient, structural, and concavity (GSC) attributes, used originally for handwritten digit recognition [18]. Text-dependent statistical evaluations are performed on a data set containing 1,000 writers who copied a fixed text of 156 words (the CEDAR letter) three times. This is the largest data set used up to the present in writer identification studies. Microfeatures are better than macrofeatures in identification tests with a performance exceeding 80 percent. A multilayer perceptron or parametric distributions are used for writer verification with an accuracy of about 96 percent. Writer discrimination was also evaluated using individual characters [19], [20] and words [21], [22].

Bensefia et al. [23], [24], [25], [26] use graphemes generated by a handwriting segmentation method to encode the individual characteristics of handwriting independent of the text content. Our allograph-level approach is similar to the work reported in these studies. Grapheme clustering is used to define a feature space common for all documents in the data set. Experimental results are reported on three data sets containing 88 writers, 39 writers (historical documents), and 150 writers, with two samples (text blocks) per writer. Writer identification is performed in an information retrieval framework, while writer verification is based on the mutual information between the grapheme distributions in the two handwritings that are compared. Concatenations of graphemes are also analyzed in the mentioned papers. Writer identification rates around 90 percent are reported on the different test data sets.

Marti et al. [27] and Hertel and Bunke [28] take text lines as the basic input unit from which text-independent features are computed using the height of the three main writing zones, slant and character width, the distances between connected components, the blobs enclosed inside ink loops, the upper/lower contours, and the thinned trace processed using dilation operations. A feature selection study is also performed in [29]. Using a k-nearest-neighbor classifier, identification rates exceeding 92 percent are obtained in tests on a subset of the IAM database [30] with 50 writers, five handwritten pages per writer. The IAM data set will also be used in the current study.

Schlapbach and Bunke [31] use HMM-based handwriting recognizers [32] for writer identification and verification. The recognizers are specialized for a single writer by training using only handwriting originating from the chosen person. This method uses the output log-likelihood scores of the HMMs to identify the writer on separate text lines of variable content. Results of 96 percent identification with 2.5 percent error in verification are reported on a subset of the IAM database containing 100 writers, five handwritten pages per writer.

In [33], we proposed a texture-level approach using edge-based directional PDFs as features for text-independent writer identification. The joint PDF of "hinged" edge-angle combinations outperformed all the other evaluated features. Further improvements are obtained through also incorporating location information by extracting separate PDFs for the upper and lower halves of text lines and then adjoining the feature vectors [34]. Our allograph-level approach [9], [35] assumes that every writer acts as a stochastic generator of ink-blob shapes, or graphemes. The grapheme occurrence PDF is a discriminatory feature between different writers and it is computed on the basis of a common shape codebook obtained by clustering [36]. Full details regarding our methods will be given further. An independent confirmation of our early experimental results is given in [37]. The contribution of this paper is that we offer a coherent overall view of our methods and present significant algorithmic extensions and improvements, together with a thorough experimental evaluation. We consider both problems of writer identification and verification and we provide a comprehensive analysis of feature combinations.

An interactive approach involving character retracing and DTW matching is proposed in [38]. A layered architecture for forensic handwriting analysis systems is proposed in [39]. The relevance of biometrics in the area of document analysis and recognition is discussed in [40].

From the studies reviewed in this section, two main conclusions can be drawn. First, in the text-dependent approach, high performance is achievable even with very small amounts of available handwritten material (on the order of a few words). However, serious drawbacks are the limited applicability due to the assumption of a fixed text or the need for human intervention in localizing the objects of interest. The text-independent approach involves less human work and has broader applicability, but it requires larger amounts of handwriting in order to derive stable statistical features. Second, training writer-specific parametric models leads to significant improvements in performance under the assumption, however, that sufficiently large amounts of handwriting are available for every writer.

The current paper proposes text-independent methods for writer identification and verification. Our approach is sparse-parametric, it involves minimal training, and the testing conditions are relevant to the forensic application. In our experimental data sets, there are only two samples per writer usually containing an amount of handwriting on the order of one paragraph of text.

3 Experimental Data Sets

We conducted our experimental study using three data sets: Firemaker, IAM, and ImUnipen.

The Firemaker set [41] contains handwriting collected from 250 Dutch subjects, predominantly students, who were required to write four different A4 pages. On page 1, they were asked to copy a text of five paragraphs using normal handwriting (i.e., predominantly lowercase with some capital letters at the beginning of sentences and names). On page 2, they were asked to copy another text of two paragraphs using only uppercase letters. Page 3 contains "forged" text and these sample are not used in the current study. On page 4, the subjects were asked to describe the content of a given cartoon in their own words. These samples consist of mostly lowercase handwriting of varying text content and the amount of written ink varies significantly, from two lines up to a full page. The documents were scanned at 300 dpi, 8 bits/pixel, gray-scale. In the writer identification and verification experiments reported in this paper, we performed searches/matches of page 1 versus 4 (Firemaker lowercase) and paragraph 1 versus 2 from page 2 (Firemaker uppercase).

The IAM database [30] consists of forms with handwritten English text of variable content, scanned at 300 dpi, 8 bits/pixel, gray-scale. Besides the writer identity, the images are accompanied by extensive segmentation and ground-truth information at the text line, sentence, and word levels [42]. This data set includes a variable number of handwritten pages per writer, from one page (350 writers) to 59 pages (one writer). In order to have comparable experimental conditions across all data sets, we modified the IAM set to always contain two samples per writer: We kept only the first two documents for those writers who contributed more than two documents to the original IAM data set and we have split the document roughly in half for those writers with a unique page in the original set. Our modified IAM set therefore contains lowercase handwriting from 650 people, two samples per writer. The amount of ink is roughly equal in the two samples belonging to one writer, but varies between writers from three lines up to a full page.

The ImUnipen set contains handwriting from 215 subjects, two samples per writer. The images were derived from the Unipen database [43] of online handwriting. The time sequences of coordinates were transformed to simulated 300 dpi images using a Bresenham line generator and an appropriate brushing function. The samples contain lowercase writing with varying text content and amount of ink. This set was not directly used in the writer identification and verification tests reported here due to the different origin of the images. A part of this data set containing 65 writers (130 samples) was used in our allograph-level approach for training the shape codebooks needed for computing the writer-specific grapheme emission probability.

We merged the Firemaker lowercase and IAM data sets to obtain a combined set which we named "Large." The Large data set therefore contains 900 writers, two samples per writer, lowercase handwriting. This combined set is comparable, in terms of the number of writers, to the largest data set used in writer identification and verification until the present [16]. It is significant to mention here that our approach to writer identification and verification is text-independent and does not require human effort for labeling. This gave us the noteworthy advantage of being able to easily extend our methods to other data sets and to collect data from multiple sources and different languages in a common framework. Table 1 gives an overview of all data sets used in our tests.

4 Texture-Level Features

Asserting writer identity based on handwriting images requires three main processing phases: 1) feature extraction, 2) feature matching/feature combination, and 3) writer identification and verification. In this and in the following sections of the paper, we present the feature extraction methods. We use probability distribution functions (PDFs) extracted from the handwriting images to characterize writer individuality in a text-independent manner. The term "feature" is used to denote such a complete PDF: not a single value, but an entire vector of probabilities capturing a facet of handwriting uniqueness.

An overview of all the features used in our study is given in Table 2. In our analysis, we will consider a number of features that we have designed (*f2*, *f3*, *f4*) and also a number of other features (*f1*, *f5*, *f6*) classically used for writer identification and verification. For the present paper, we have selected the most discriminative features from a larger

TABLE 1
Overview of the Experimental Data Sets, the Number of Writers Contained, and Some of Their Properties

	Dataset	Nwriters	Handwriting	Obs.
1	Firemaker	250	-lowercase -UPPERCASE	-page 1 and 4 -paragraph 1 and 2 of page 2
2	IAM	650	-lowercase	-original IAM dataset modified to contain 2 samples per writer
3	ImUnipen	215	-lowercase	-derived from online data, not used in writer identification and verification tests, 130 samples by 65 writers used for generating the grapheme codebooks
4	Large	900	-lowercase	-merger between Firemaker lowercase and IAM datasets

TABLE 2
Overview of the Considered Features, Their Dimensionalities, and the Distance Functions Used in Identification and Verification

	Feature	Explanation	Ndims	Dist	Computed from
f1	$p(\phi)$	Contour-direction PDF	12	χ^2	contours
f2	$p(\phi_1, \phi_2)$	Contour-hinge PDF	300	χ^2	contours
		Direction co-occurrence PDFs			contours
f3h	$p(\phi_1, \phi_3)$ h.	- horizontal run	144	χ^2	
f3v	$p(\phi_1, \phi_3)$ v.	- vertical run	144	χ^2	
f4	$p(g)$	Grapheme emission PDF	400	χ^2	connected components
		Run-length on background PDFs			binary image
f5h	$p(rl)$ h.	- horizontal run	60	χ^2	
f5v	$p(rl)$ v.	- vertical run	60	χ^2	
f6	ACF	Autocorrelation in horiz. raster	60	L_2	gray-scale image

Features are grouped into four different categories: directional PDFs (f1, f2, f3h, f3v), grapheme emission PDF (f4), run-length PDFs (f5h, f5v), and autocorrelation (f6).

number of features tested in a previous study (normalized entropy, ink-density PDF, wavelets) [9].

A succession of image processing steps applied on the handwriting image will provide a number of alternate base representations which will then be used for feature computation. The initial gray-scale images containing the scanned samples of handwriting are binarized using Otsu's method [44]. The binary images, in which only the ink pixels are "on," undergo connected component detection (labeling) using 8-connectivity. Further, for all connected components, the inner and outer contours are extracted using Moore's contour-following algorithm. The contours will contain the sequence of coordinates (x_k, y_k) of all of the pixels located exactly on the ink-background boundary. This is a very effective vectorial representation that will allow a fast computation of the directional features considered further in this section. Four primary representations of the handwritten document will therefore be used for feature computation: the gray-scale image, the binary image, the connected components, and the contours.

The current study implicitly assumes that the foreground/background separation can be realized in a preprocessing phase, yielding a white background with (near) black ink. This separation will often fail on the smudged and texture-rich fragments sometimes collected in forensic practice, where the ink trace is often hard to identify. However, the complete process of forensic writer identification is never fully automatic and present image processing methods allow for advanced semi-interactive solutions to the foreground/background separation problem.

Our writer identification and verification methods work at two levels of analysis: the *texture* level and the *allograph* level. Further, in this section, we describe the extraction methods for the texture-level 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.

4.1 Contour-Direction PDF (*f1*)

The most prominent visual attribute of handwriting that reveals individual writing style is slant. Handwriting slant

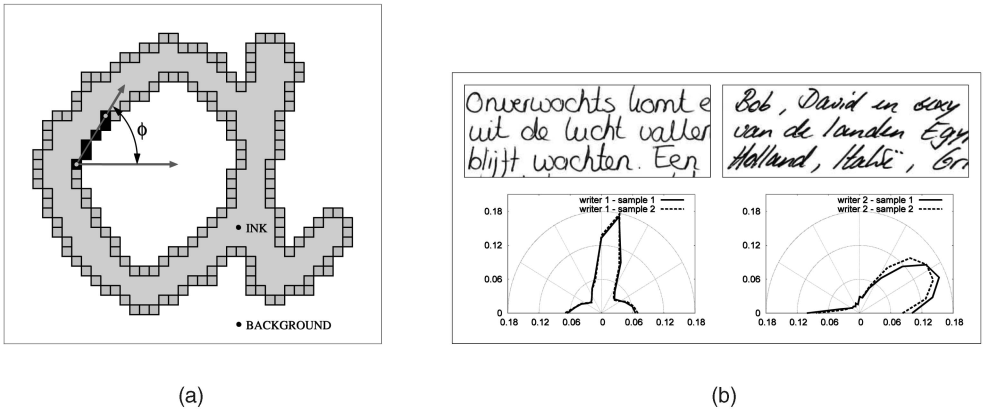


Fig. 2. (a) Schematic description for the extraction method of the contour-direction PDF (feature *f1*). The handwritten letter "a," provided as an example, would be roughly twice as large in reality. (b) Examples of lowercase handwriting from two different subjects. We superposed the polar diagrams of the direction distribution $p(\phi)$ extracted from the two handwritten samples for each of the two subjects. There is a large overlap between the directional PDFs extracted from samples originating from the same writer, while there is a substantial variation in the directional PDFs for different writers. The examples were chosen for visual clarity.

is also a very stable personal characteristic [6], [5]. It has long been known in handwriting research that the distribution of directions in the script provides useful information for writer identification [45], coarse writing-style classification [46], or signature verification [47]. This directional distribution can be computed very fast using the contour representation, with the additional advantage that the influence of the ink-trace width is also eliminated.

The contour-direction distribution is extracted by considering the orientation of local contour fragments. The analyzing fragment is determined by two contour pixels taken a certain distance apart (see Fig. 2a) and the angle that the fragment makes with the horizontal is computed using (1). As the algorithm runs over the contours, the orientation of the local contour fragments is computed and an angle histogram is thereby built. The angle histogram is then normalized to a probability distribution $p(\phi)$ which gives the probability of finding in the handwriting image a contour fragment oriented at the angle ϕ measured from the horizontal.

$$\phi = arctan\left(\frac{y_{k+\epsilon} - y_k}{x_{k+\epsilon} - x_k}\right). \tag{1}$$

The parameter ϵ controls the length of the analyzing contour fragment. 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 angle ϕ resides in the first two quadrants because, without online information, we do not know which way the writer "traveled" along the probing contour fragment. The number of histogram bins spanning the interval 0°-180° was set to $n = 12$ through experimentation: $15°/\mathrm{bin}$ gives a sufficiently detailed and, at the same time, sufficiently robust description of handwriting to be used in writer identification and verification. These settings will be used for all of the directional features presented in this paper.

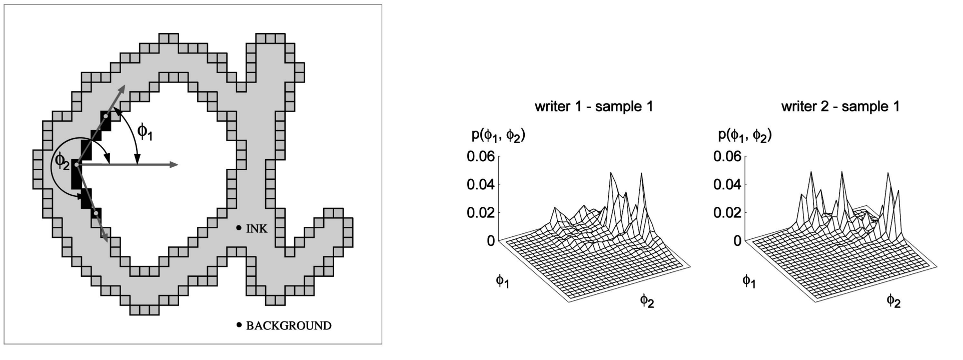
The prevalent direction in $p(\phi)$ (see Fig. 2b) corresponds, as expected, to the slant of writing. In handwriting recognition, this can be used to deslant the script using a shear transform prior to applying the statistical recognizer. Note that not only the slant (the mode of the angular PDF), but the entire distribution is informative for writer identification. For example, even for the same slant angle, a more round handwriting will have a different directional PDF (more spread) than a more pointed handwriting and it will still be possible to distinguish between them using $p(\phi)$.

4.2 Contour-Hinge PDF (*f2*)

The directional distribution $p(\phi)$ represented our starting point in designing more complex features that give a more intimate characterization of the individual handwriting style and ultimately yield significant improvements in writer identification and verification performance. In order to also capture, besides orientation, 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, subsequently, compute the joint probability distribution of the orientations of the two legs of the obtained "contour-hinge" (see Fig. 3a). To have an intuitive picture of this feature, imagine having a hinge laid on the surface of the image. Place its junction on top of every contour pixel, then open the hinge and align its legs along the contour. Consider the angles ϕ_1 and ϕ_2 that the legs make with the horizontal and count the found instances in a two-dimensional array of bins indexed by ϕ_1 and ϕ_2. The final normalized histogram gives the joint PDF $p(\phi_1, \phi_2)$ quantifying the chance of finding in the image two "hinged" contour fragments oriented at the angles ϕ_1 and ϕ_2, respectively.

In contrast to feature *f1* for which spanning the upper two quadrants (180°) was sufficient, we now have to span all four quadrants (360°) around the central junction pixel when assessing the angles of the two fragments. The orientation is now quantized in $2n$ directions for every leg of the "contour-hinge." From the total number of combinations of two angles ($4n^2$), we consider only nonredundant ones ($\phi_2 \geq \phi_1$). The final number of combinations is $C_{2n}^2 + 2n = n(2n+1)$. For $n = 12$, the contour-hinge feature vector has 300 dimensions.

The feature $p(\phi_1, \phi_2)$ is a bivariate PDF capturing both the orientation and the curvature of contours. Examples are given in Fig. 3b. Additionally, the joint probability $p(\phi_1, \phi_2)$ is proportional to the conditional probability $p(\phi_2|\phi_1)$ that can be interpreted as the transition probability from state ϕ_1 to state ϕ_2 in a simple Markov process. Feature *f2* is highly discriminative and gives very satisfying results in writer identification and verification.



(a) (b)

Fig. 3. (a) Schematic description for the extraction method of the contour-hinge PDF (feature *f2*). (b) Surface plots of the contour-hinge PDF $p(\phi_1, \phi_2)$ for two writers. Every writer has a different "probability landscape." One half of the 3D plot (on one side of the main diagonal) is flat because we only consider angle combinations with $\phi_2 \geq \phi_1$.

4.3 Direction Co-Occurrence PDFs (*f3h*, *f3v*)

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 combination of contour-angles occurring at the ends of run-lengths on the background (see Fig. 4). The joint PDF $p(\phi_1, \phi_3)$ of the two contour-angles occurring at the ends of a run-length on white captures longer range correlations between contour directions and gives a measure of the roundness of the written characters. Horizontal runs along the rows of the image generate *f3h* and vertical runs along the columns of the image generate *f3v*. The PDFs *f3h* and *f3v* have n^2 dimensions, namely, 144 in our implementation. These features derive conceptually from the directional distribution *f1* presented above and the run-length distributions *f5h* and *f5v*, which will be described further.

The features presented thus far, *f1*, *f2*, and *f3*, are directional PDFs constructed using oriented contour fragments that act like local phasors and perform, in Fourier terms, a local phase analysis at the scale of the ink-trace width. The local phase correlations are collected in the joint probability distributions that are generic texture descriptors characterizing individual handwriting style independently of the text content of the written samples.

4.4 Other Texture-Level Features: Run-Length PDFs (*f5h*, *f5v*), Autocorrelation (*f6*)

Run-lengths were first proposed for writer identification in [48], [49] and were also used on historical documents in [50]. Run lengths are determined on the binary image taking into consideration either the black pixels corresponding to the ink trace or the white pixels corresponding to the background. The statistical properties of the black runs are significantly influenced by the ink width and, therefore, by the type of pen used for writing. The white runs capture the regions enclosed inside the letters and also the empty spaces between letters and words. The probability distribution of white lengths (runs on background) will be used in our writer identification and verification tests. There are two basic scanning methods: horizontal along the rows of the image (*f5h*) and vertical along the columns of the image (*f5v*). Similarly to the contour-based directional features presented above, the histogram of run lengths is normalized and interpreted as a probability distribution. Our particular implementation considers only run-lengths of up to 60 pixels to prevent the vertical measurements from going in between successive text lines (the height of a written line in our data set is about 120 pixels).

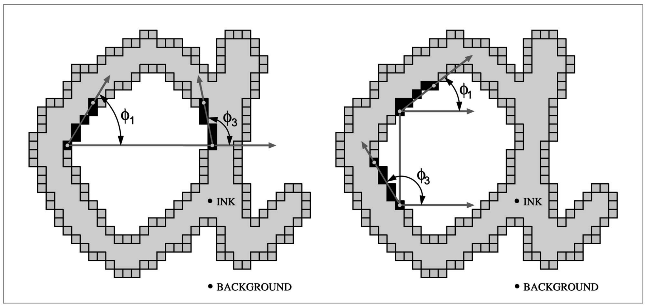


Fig. 4. Schematic description for the extraction methods of the direction co-occurrence PDFs (horizontal scan—feature *f3h* on the left; vertical scan—feature *f3v* on the right).

To compute the autocorrelation feature (*f6*), every row of the image is shifted onto itself by a given offset and then the normalized dot product between the original row and the shifted copy is calculated. The original gray-scale image is used in the computation and the maximum offset ("delay") corresponds to 60 pixels. For every offset, the autocorrelation coefficients are then averaged across all image rows. The autocorrelation function detects the presence of regularity in writing: Regular vertical strokes will overlap in the original row and its horizontally shifted copy for offsets equal to integer multiples of the spatial wavelength of handwriting. This results in a large dot product contribution to the final autocorrelation function. Autocorrelation is the only feature in our analysis that is not a probability distribution function and it will require a different distance measure than the other features, Euclidean (L_2 norm) rather than χ^2.

We note here that the autocorrelation and the power spectrum are Fourier transform pairs. Therefore, in effect, the autocorrelation function performs a Fourier analysis directly in image space along the pixel rows. The amplitude information is retained and averaged across all image rows, while all phase information is discarded. Directional features (*f1*, *f2*, and *f3*) are essentially built on local phase information, while autocorrelation encodes only amplitude information. It will be interesting to consider a performance comparison in the experimental results.

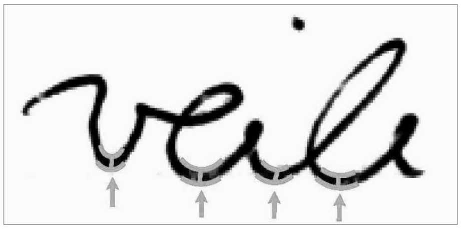


Fig. 5. Handwriting segmentation at the minima in the lower contour that are proximal to the upper contour.

The features presented in this section are generic texture-level descriptors that, when applied to handwriting, capture writer individuality, thus providing the basis for writer identification. Their virtue resides in the local computation on the image and, as such, they are generally applicable and do not impose additional constraints. Using the contour representation for extracting the directional distributions offers definite advantages regarding computation speed and control of feature dimensionality. The PDFs can be estimated, even from samples with very reduced amounts of written ink. In our data, many handwritten samples contain as little as three lines of text.

5 Allograph-Level Features

In this section, we present our allograph-level approach to writer identification and verification. Our method is based on assuming that the writer acts as a stochastic generator of ink-blob shapes, or graphemes. The probability distribution of grapheme usage is characteristic of each writer and is computed using a common codebook of shapes obtained by clustering. This approach was first applied to isolated uppercase handwriting [9] and later it was extended to lowercase cursive handwriting by using a segmentation method [35]. This writer identification and verification method involves three processing stages: handwriting segmentation, shape codebook generation by grapheme clustering, and computation of the writer-specific grapheme usage PDF on the basis of the common shape codebook. These stages will be described in detail further.

5.1 Handwriting Segmentation Method

In free-style cursive handwriting, connected-components may encompass several characters or syllables. A segmentation method that perfectly isolates individual characters remains an elusive goal for handwriting research. Nevertheless, several heuristics can be applied, yielding graphemes (sub or supra-allographic fragments) that may or may not overlap a complete character. While this represents a fundamental problem for handwriting recognition, the fraglets generated by the segmentation procedure can still be effectively used for writer identification. The essential idea is that the ensemble of these simple graphemes still manages to capture the shape details of the allographs emitted by the writer. This segmentation stage makes our allograph-level method applicable to free-style handwriting, both cursive and isolated.

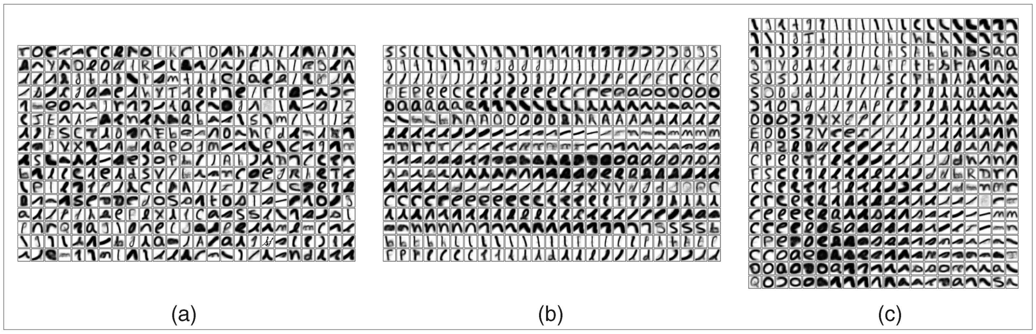
We segment the ink at the minima in the lower contour with the added condition that the distance to the upper contour is on the order of the ink-trace width (see Fig. 5). After segmentation, graphemes are extracted as connected components, followed by a size normalization to 30×30 pixel bitmaps, preserving the aspect ratio of the original pattern.

5.2 Grapheme Codebook Generation by Clustering

A number of 130 samples from 65 writers have been taken from the ImUnipen data set. The graphemes have been extracted from these samples using the described procedure yielding a training set containing a total of 41k patterns (normalized bitmaps). Clustering was applied to this training set in order to obtain a reduced list of graphemes that will be used as a shape codebook. The graphemes in the codebook act as prototype shapes representative of the types of shapes that are to be expected in handwriting as a result of segmentation.

We will compare three clustering techniques for generating the grapheme codebook: k-means, Kohonen SOM 1D, and 2D. We use standard implementations of these methods. Complete and clear descriptions of the algorithms can be found in references [51], [52]. The size of the codebook (the number of clusters used) yielding optimal performance is an important parameter in our method and we will explore a large range of codebook sizes in the experiments.

Fig. 6 shows examples of codebooks that have been obtained using the three clustering methods. The two codebooks obtained using Kohonen training show spatial order, while the one obtained using k-means is "disorderly." The ksom1D codebook must be understood as a long linear string of shapes and gradual transitions can be observed if the map is "read" in left-to-right top-to-bottom order. The ksom2D codebook shows a clear bidimensional organization.



(a) (b) (c)

Fig. 6. Examples of codebooks with 400 graphemes. For (a) kmeans and (b) ksom1D the graphemes have been placed 25 in a row, while, for (c) ksom2D, the 20×20 original SOM organization has been maintained.

5.3 Computing Writer-Specific Grapheme Emission PDF (*f4*)

The writer is considered to be characterized by a stochastic pattern generator, producing a family of basic shapes. The individual shape emission probability is computed by building a histogram in which one bin is allocated to every grapheme in the codebook.

For every sample i of handwriting, the graphemes are extracted using the segmentation/connected-component-detection/size-normalization procedure described before. For every grapheme g in the sample, the nearest codebook prototype w (the winner) is found using the Euclidean distance and this occurrence is counted into the corresponding histogram bin:

$$w = argmin_n[dist(g, C_n)], h_{iw} \leftarrow h_{iw} + 1, \tag{2}$$

where n is an index that runs over the shapes in the codebook C. In the end, the histogram h_i is normalized to a probability distribution function p_i that sums to 1. This PDF is the writer descriptor used for identification and verification (feature *f4*).

The essence of this method does not consist of an exhaustive enumeration of all possible allographic part shapes. Rather, the grapheme codebook spans a shape space by providing a set of nearest-neighbor attractors for the ink fraglets extracted from a given written sample. The grapheme occurrence PDF captures the individual shape usage preference of the writer.

6 Writer Identification and Verification by Feature Matching and Fusion

After the handwritten samples have been mapped onto features capturing writer individuality, an appropriate distance measure between the feature vectors is needed to compute the (dis)similarity, in individual handwriting style, between any two chosen samples. A large number of distance measures were tested in our experiments: Minkowski up to order 5, χ^2, Bhattacharya and Hausdorff. We will report however only on the best performing ones.

For the PDF features (*f1*,*f2*,*f3*,*f4*,*f5*), the χ^2 distance [53] is used for matching a query sample q and any other sample i from the database:

$$\chi^2_{qi} = \sum_{n=1}^{Ndims} \frac{(p_{qn} - p_{in})^2}{p_{qn} + p_{in}}, \tag{3}$$

where p are entries in the PDF, n is the bin index, and $Ndims$ is the number of bins in the PDF (the dimensionality of the feature). χ^2 is a natural choice as a distance measure for the PDF features. Euclidean distance is used for the autocorrelation (*f6*).

Writer identification is performed using nearest-neighbor classification in a "leave-one-out" strategy. For a query sample q, the distances to all of the other samples $i \neq q$ are computed using a selected feature. Then, all the samples i are ordered in a sorted hit list with increasing distance to the query q [53]. Ideally, the first ranked sample should be the pair sample produced by the same writer. If one considers not only the nearest neighbor (Top 1), but rather a longer list of neighbors starting with the first and up to a chosen rank (e.g., Top 10), the chance of finding the correct hit (the recall) increases with the list size. We point out that, in our experiments, we do not make a separation between a training set and a test set all of the data is in one suite. This is actually a more difficult and realistic testing condition, with more distractors: not one, but two per false writer and only one correct hit.

Writer verification, as all biometric verification tasks, can be perfectly placed into the classical Neyman-Pearson framework of statistical decision theory [54]. For writer verification, the distance x between two given handwriting samples is computed using a chosen feature. Distances up to a predefined decision threshold T are deemed sufficiently low for considering that the two samples have been written by the same person. Beyond T, the samples are considered to have been written by different persons. Two types of error are possible: falsely accepting (FA) that two samples are written by the same person when, in fact, this is not true or falsely rejecting (FR) that two samples are written by the same person when, in fact, this is the case. The associated error rates are FAR and FRR. In a scenario in which a suspect must be found in a stream of documents, FAR becomes the false alarm rate, while FRR becomes the miss rate. These error rates can be empirically computed by integrating, up to/from the decision threshold T the probability distribution of distances between samples written by the same person $P_S(x)$ and the probability distribution of distances between samples written by different people $P_D(x)$:

$$FAR = \int_0^T P_D(x)\, dx, FRR = \int_T^\infty P_S(x)\, dx. \tag{4}$$

By varying the threshold T, a Receiver Operating Characteristic (ROC) curve is obtained that illustrates the inevitable trade-off between the two error rates. The Equal Error Rate (EER) corresponds to the point on the ROC curve where FAR = FRR and it quantifies in a single number the writer verification performance.

The features considered in the present study are not totally orthogonal, but, nevertheless, they do offer different points of view on a handwritten sample. It is therefore natural to try to combine them for improving performance. In our feature combination scheme, the final unique distance between any two handwritten samples is computed as the average (simple or weighted average) of the distances due to the individual features participating in the combination (see Fig. 7).

In feature combinations, Hamming distance performed best:

$$H_{qi} = \sum_{n=1}^{Ndims} |p_{qn} - p_{in}|. \tag{5}$$

The χ^2 distance, due to the denominator (see (3)), gives more weight to the low probability regions in the PDFs and maximizes performance for each individual feature. On the other hand, Hamming distance generates comparable distance values for the different PDF features and offers a common ground with slight advantages in feature combinations.

The Bayesian framework underlying the feature combination scheme proposed here entails two fundamental assumptions: Features are independent and the probability of two samples having been written by the same person assumes an exponential distribution with respect to the distance between the two samples as generated by a chosen feature $P_S(x) \propto e^{-x/\sigma}$. The decay constants σ control the weights that different features take on in the combination. While this basic probabilistic model will almost certainly be

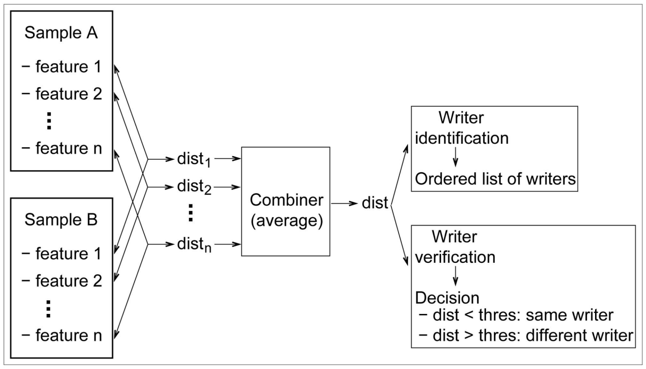


Fig. 7. Feature combination scheme: The distances generated by the individual features are averaged (using simple or weighted average) and the result is then used in writer identification and verification.

violated in reality, experimental results show that significant performance improvements are nevertheless achievable by using the proposed feature combination method.

In a more general perspective, feature fusion for writer identification and verification pertains to the broader theme of classifier combination [55] or multimodal biometrics [56], [57]. Information can be combined at three levels in the biometric identification or verification process: *sensor fusion*, *similarity-score fusion*, and *decision-level fusion* [58]. Combining similarity scores ("soft" fusion) seems to be the method of choice in multimodal biometrics. This is also confirmed in our experiments: We obtained the best feature fusion results by combining the distances (or similarity scores) generated by the individual features.

7 Results

In this section, we present our experimental results. The performance measures used are the Top-1 and Top-10 identification rates and the Equal-Error-Rate (EER) for verification. As explained in Section 3 of this paper, four data sets are considered in the experimental evaluation (see Table 1): Firemaker uppercase (250 writers), Firemaker lowercase (250 writers), IAM (650 writers), and Large (900 writers obtained by merging the Firemaker lowercase and IAM data sets). All data sets contain two samples per writer and writer identification searches are performed in a "leave-one-out" manner. The shape codebook necessary for computing the grapheme occurrence probability (feature *f4*) was built using part of the ImUnipen data set (65 writers, two samples/writer, 41k bitmap patterns). This ensures a complete separation, at the level of the writers, between the training and the testing data. For the results reported in this section, we used a grapheme codebook generated by k-means clustering and containing 400 prototype shapes.

We are interested in a comparative performance analysis of the different features across the four test data sets. We are also interested in the improvements in performance obtained by combining multiple features. First, we shall consider the individual features and then their combinations. Afterward, we will present an HTML-based visualization tool that we developed in order to directly examine the output generated by our writer identification and verification system.

7.1 Performances of Individual Features

Table 3 gives the writer identification and verification performance of the individual features considered in the present study. While there are important differences in performance among the different features, it can be noticed that, for a chosen feature, performance is consistent across the four experimental data sets. The best performer is the contour-hinge PDF (*f2*), followed by the grapheme-emission PDF (*f4*).

The results obtained on Firemaker uppercase are comparable to those obtained on Firemaker lowercase. Although the amount of ink contained in the samples varies between the two data sets, this result is nevertheless interesting because, in our data, the uppercase samples generally contain less handwriting than the lowercase ones. Similar results were reported in experiments where the amount of ink in the samples was controlled [34]. These findings contradict the idea that one might intuitively expect that it is always easier to identify the author of lowercase rather than uppercase handwriting. Naturally, the features used are sensitive to major style variations and, in mixed searches (e.g., lowercase query sample/uppercase data set), performance is very low.

The writer identification performances obtained on Firemaker lowercase and IAM are very similar, albeit the large

TABLE 3
Writer Identification and Verification Performance of Individual Features

	Dataset	Firemaker - UPPERCASE 250 writers			Firemaker - lowercase 250 writers			IAM - lowercase 650 writers			Large - lowercase 900 writers		
	Feature	Top 1	Top 10	EER	Top 1	Top 10	EER	Top 1	Top 10	EER	Top 1	Top 10	EER
f1	$p(\phi)$	43	81	7.6	48	79	7.7	46	76	7.1	43	72	7.1
f2	$p(\phi_1, \phi_2)$	84	96	4.1	81	92	4.8	81	92	5.0	80	91	4.8
f3h	$p(\phi_1, \phi_3)$ h.	51	84	8.3	68	86	6.4	68	87	5.5	65	84	5.9
f3v	$p(\phi_1, \phi_3)$ v.	37	72	16.0	66	89	7.6	65	84	9.6	59	82	9.1
f4	$p(g)$	65	92	8.0	75	92	5.7	80	94	5.6	76	92	5.8
f5h	$p(rl)$ h.	8	32	17.2	18	50	14.4	10	32	17.0	8	29	16.6
f5v	$p(rl)$ v.	9	39	14.8	16	44	16.3	8	31	15.5	10	34	12.1
f6	ACF	16	47	15.6	16	48	15.3	13	38	16.1	12	35	14.7

The χ^2 distance was used in matching. The features are explained in Table 2.

TABLE 4
Writer Identification and Verification Performance of Feature Combinations

Dataset	Firemaker - UPPERCASE 250 writers			Firemaker - lowercase 250 writers			IAM - lowercase 650 writers			Large - lowercase 900 writers		
Combination	Top 1	Top 10	EER	Top 1	Top 10	EER	Top 1	Top 10	EER	Top 1	Top 10	EER
f3: *f3h* & *f3v*	63	89	7.6	75	92	4.8	77	91	5.3	73	89	5.0
f5: *f5h* & *f5v*	29	70	8.1	42	75	9.6	31	60	9.0	33	63	7.5
f1 & *f4*	69	95	5.6	79	93	4.1	84	95	3.3	81	94	3.3
f1 & *f5*	66	93	4.1	70	91	4.6	68	91	4.0	67	90	3.6
f2 & *f4*	(77)	97	4.5	83	94	3.2	88	97	2.8	86	95	2.9
f3 & *f4*	75	95	5.2	82	94	3.8	86	96	3.9	84	95	3.9
f3 & *f5*	79	95	4.0	80	94	3.4	82	94	3.9	80	94	3.7
f4 & *f5*	76	97	4.7	79	94	3.7	85	96	3.1	83	95	3.2
f1 & *f4* & *f5*	82	98	4.0	82	95	3.2	87	96	2.8	85	96	2.8
f2 & *f4* & *f5*	85	98	3.7	83	95	3.2	89	97	2.8	87	96	2.6
f3 & *f4* & *f5*	86	97	3.6	83	95	3.2	89	96	3.3	87	96	3.3

The $Hamming$ distance was used in matching. Combining features from different feature groups yields improvements in performance over the best individual feature participating in the combination. There is one exception marked with parentheses: Top-1 identification rate for f2 & f4 on Firemaker uppercase data set.

difference in the number of writers contained in the two data sets. This is probably due to differences in the writer distributions underlying the two data sets. The Firemaker data set was collected from a rather uniform population in terms of age and education, predominantly Dutch students, and, as a consequence, there is less variation in writing styles compared to the IAM data set. Under these conditions, when these two data sets are combined, only a slight decrease in writer identification performance on the Large data set is noticed. The dependence of the writer identification rate on the number of writers contained in the data set is discussed in the following section of this paper. For the size of the data sets used here, the writer identification percentages are subject to a 3-4 percent confidence interval at a 95 percent confidence level.

From the point of view of Fourier analysis, it is important to observe that the contour-direction feature *f1* encoding local phase information performs much better than the autocorrelation feature *f6* encoding amplitude information. In computer vision, it is commonly acknowledged that phase information is predominantly used for identification, while amplitude information is generally used for recognition, mainly due to the shift-invariance of the power spectrum. Phase demodulation and phase-based representations are pervasive in biometric identification [59], [60].

Furthermore, the contour-angle combination features *f2*, *f3h*, and *f3v*, based on local phase correlations, deliver significant improvements in performance over the basic directional PDF *f1*. This confirms the general principle that joint probability distributions do capture more information from the input signal. And, despite their higher dimensionalities, reliable probability estimates can be obtained for the proposed joint PDFs when a few handwritten text lines are available (usually more than three in our data sets). An analysis of writer identification performance versus amount of ink contained in the samples is given in [33].

In brief, our results show that the contour-based angle-combination PDFs (*f2*, *f3h*, *f3v*) and the grapheme-emission PDF (*f4*) outperform the other features over the four test data sets. They constitute the gist of our text-independent approach to writer identification and verification.

7.2 Performances of Feature Combinations

The features considered in the present study capture different aspects of handwriting individuality and operate at different levels of analysis and also at different scales. While our features are not completely orthogonal, combining multiple features proves, nevertheless, to be beneficial.

As stated previously, feature fusion is performed by distance averaging. Assigning distinct weights to the different features participating in the combination yields only very small performance improvements, as will be shown further. This has led us to prefer simplicity and robustness here and report the feature combination results obtained by plain distance averaging.

The features studied in the paper can be grouped into four broad categories (see Table 2): contour-based directional PDFs (*f1*, *f2*, *f3h*, *f3v*), grapheme emission PDF (*f4*), run-length PDFs (*f5h*, *f5v*), and autocorrelation (*f6*). We will analyze combinations of features within and between these broad feature groups.

First, we consider the natural combinations *f3h* with *f3v* and *f5h* with *f5v* (first two rows of Table 4). Features *f3* and *f5* are therefore obtained by combining the two orthogonal directions of scanning the input image. Compared to their single horizontal or vertical counterparts, the fused features perform markedly better and they will be used, as such, in future combinations.

It is important to note that further combining directional features (*f1* & *f2*, *f1* & *f3*, *f2* & *f3*, or *f1* & *f2* & *f3*) did not produce extra improvements over the performance of the best feature involved in the combination. Rather, the experimental results show that improvements are obtained by combining features from different feature groups. In the results given in Table 4, the combined performance exceeds the performances of all individual features involved in the combination, with only one exception, marked with parentheses. As can be noticed,

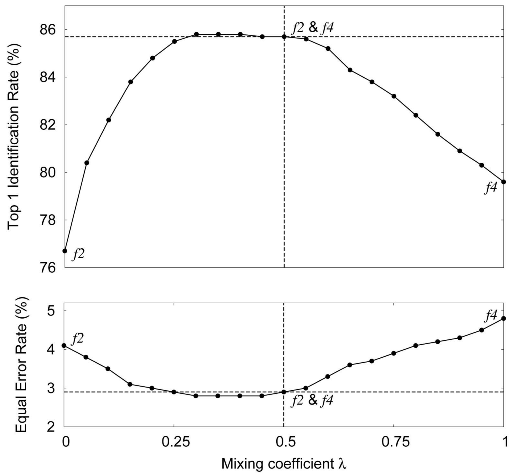


Fig. 8. Writer identification and verification performance for a weighted combination of features *f2* and *f4* on the Large data set. Only marginal improvements are obtainable over the performance levels of the simple average combination represented by the horizontal lines and corresponding to a mixing coefficient $\lambda = 0.5$.

the performance of feature combinations is generally consistent over the four experimental data sets.

The best performing feature combinations fuse directional, grapheme, and run-length information, yielding, on the Large data set, writer identification rates of 85-87 percent Top-1 and 96 percent Top-10 with an EER around 3 percent for verification.

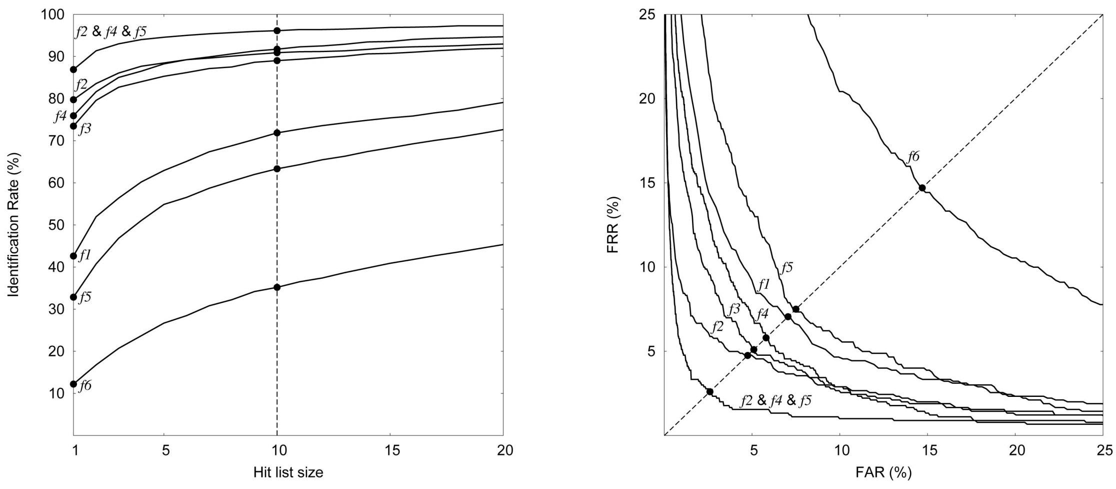
In Fig. 8, we show the results obtained by considering a weighted combination between features *f*2 and *f*4: $d = (1 - \lambda)d_2 + \lambda d_4$, where λ is the mixing coefficient. Only marginal improvements are attainable over the performance, corresponding to simple distance averaging at $\lambda = 0.5$. These results are, in fact, representative for extensive weight optimization tests carried out on different feature combinations and generating, in the end, very small overall additional performance improvements. Such a direct feature combination by simple distance averaging is possible in our case because the fused features are PDFs (that sum to 1) and, for a chosen pair of samples, the Hamming distances produced by the different features lie roughly within the same range. The only exception is the autocorrelation feature *f6* which requires weighting with respect to the other features. This has led, however, to only minor additional performance improvements, only about 1 percent increase in the Top-1 identification rate.

We mention that we replaced the linear distance combiner with an SVM [61], [62] trained for writer verification. The output of the SVM, i.e., the distance to the separating hyperplane in the space induced by the kernel function, was used for writer identification (ordering the samples with increasing distance) and writer verification (decision same/different writer). The linear kernel outperformed the other general purpose kernels (polynomial, radial basis, sigmoid). However, the experimental results were rather dismal, not justifying, in our view, the increase in system complexity and computation time. We also experimented with Borda rank combination schemes with only marginal performance improvements [34].

Fig. 9a gives a graphical overview of the writer identification results on the Large data set for individual features and for the best performing feature combination. The Top-1 and Top-10 recall rates were used as anchor points in reporting the numerical results from Table 3 and Table 4. Fig. 9b gives the writer verification ROC curves. In our case, the EER values are sufficiently descriptive, as a performance measure, for the whole profile of the corresponding ROC curves.

7.3 Visualization Tool

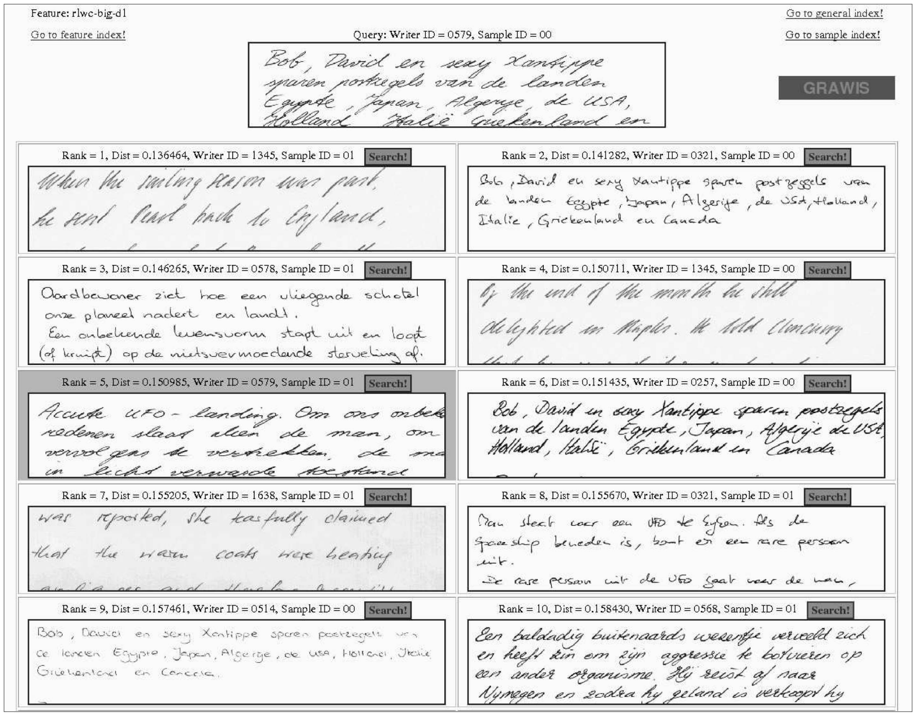
After feature extraction, feature matching, and performance calculation, our programs generate HTML files containing numerical results (distances, ranks, writer/sample identity codes, thresholds) and hyperlinks to the handwritten samples. Figs. 10, 11, and 12 show examples of hit lists generated by our system, dubbed GRAWIS for Groningen Automatic Writer Identification System. A Web browser can then be used to visualize these HTML files. This HTML-based approach allows quick development of the visualization tool without the considerable programming effort needed to construct a complete graphical user interface. For a chosen query sample, writer identification searches can be



(a)

(b)

Fig. 9. (a) Writer identification performance as a function of hit list size. (b) Writer verification ROC curves. These results were obtained on the Large data set containing 900 writers, two samples per writer.


Feature: rlwc-big-d1

Query: Writer ID = 0579, Sample ID = 00

Rank = 1, Dist = 0.136464, Writer ID = 1345, Sample ID = 01

Rank = 2, Dist = 0.141282, Writer ID = 0321, Sample ID = 00

Rank = 3, Dist = 0.146265, Writer ID = 0578, Sample ID = 01

Rank = 4, Dist = 0.150711, Writer ID = 1345, Sample ID = 00

Rank = 5, Dist = 0.150985, Writer ID = 0579, Sample ID = 01

Rank = 6, Dist = 0.151435, Writer ID = 0257, Sample ID = 00

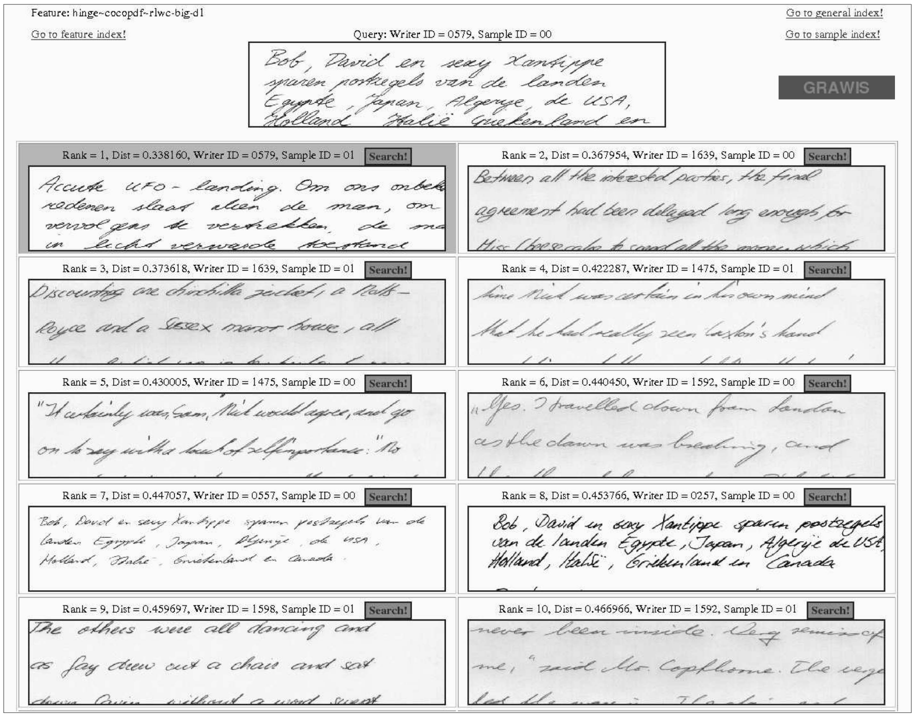
Rank = 7, Dist = 0.155205, Writer ID = 1638, Sample ID = 01

Rank = 8, Dist = 0.155670, Writer ID = 0321, Sample ID = 01

Rank = 9, Dist = 0.157461, Writer ID = 0514, Sample ID = 00

Rank = 10, Dist = 0.158430, Writer ID = 0568, Sample ID = 01

Fig. 10. Writer identification hit list generated by the moderately performing feature *f5* on the Large data set. The query sample is in the top-center position and the correct hit is on rank 5 (marked with a darker frame). The handwriting style is heterogeneous across the hit list.


Feature: hinge-cocopdf-rlwc-big-d1

Query: Writer ID = 0579, Sample ID = 00

Rank = 1, Dist = 0.338160, Writer ID = 0579, Sample ID = 01

Rank = 2, Dist = 0.367954, Writer ID = 1639, Sample ID = 00

Rank = 3, Dist = 0.373618, Writer ID = 1639, Sample ID = 01

Rank = 4, Dist = 0.422287, Writer ID = 1475, Sample ID = 01

Rank = 5, Dist = 0.430005, Writer ID = 1475, Sample ID = 00

Rank = 6, Dist = 0.440450, Writer ID = 1592, Sample ID = 00

Rank = 7, Dist = 0.447057, Writer ID = 0557, Sample ID = 00

Rank = 8, Dist = 0.453766, Writer ID = 0257, Sample ID = 00

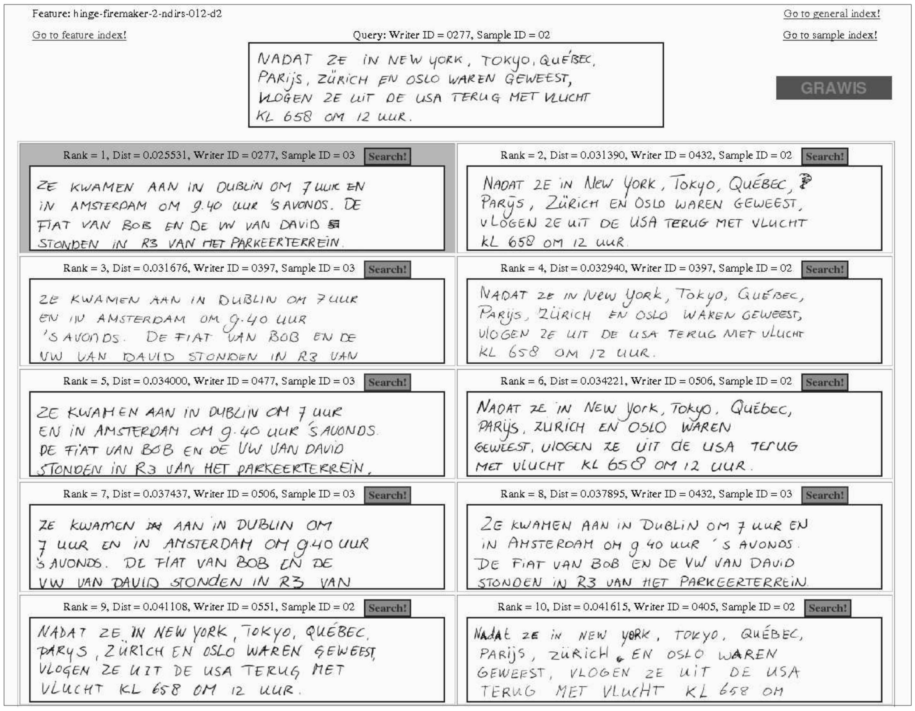
Rank = 9, Dist = 0.459697, Writer ID = 1598, Sample ID = 01

Rank = 10, Dist = 0.466966, Writer ID = 1592, Sample ID = 01

Fig. 11. A successful writer identification search using the best performing feature combination *f2* & *f4* & *f5* for the same query sample as in Fig. 10. The best-matching sample (rank 1) was written by the same writer. A uniform handwriting style can be observed across the query sample and at the top of the hit list.

run using a battery of different features or feature combinations. Every sample from a hit list can, in turn, become the query and this allows very handy navigation in the space of individual handwriting styles.

Fig. 10 shows a hit list generated by feature *f5* (fused horizontal and vertical run-length PDFs) applied on the Large set. The query sample is placed at the top-center and the correct hit (the pair sample written by the same writer) is in position 5. A rather heterogeneous handwriting style is noticeable across the retrieved samples. Fig. 11 shows a successful hit list generated by the best performing feature combination (*f2* & *f4* & *f5*) for the same query sample. The



Feature: hinge-firemaker-2-ndirs-012-d2

Query: Writer ID = 0277, Sample ID = 02

NADAT ZE IN NEW YORK, TOKYO, QUÉBEC,
PARIJS, ZÜRICH EN OSLO WAREN GEWEEST,
VLOGEN ZE UIT DE USA TERUG MET VLUCHT
KL 658 OM 12 UUR.

Rank = 1, Dist = 0.025531, Writer ID = 0277, Sample ID = 03

ZE KWAMEN AAN IN DUBLIN OM 7 UUR EN
IN AMSTERDAM OM 9.40 UUR 'S AVONDS. DE
FIAT VAN BOB EN DE VW VAN DAVID
STONDEN IN R3 VAN HET PARKEERTERREIN.

Rank = 2, Dist = 0.031390, Writer ID = 0432, Sample ID = 02

NADAT ZE IN New York, Tokyo, QUÉBEC,
PARIJS, ZÜRICH EN OSLO WAREN GEWEEST,
VLOGEN ZE UIT DE USA TERUG MET VLUCHT
KL 658 OM 12 UUR.

Rank = 3, Dist = 0.031676, Writer ID = 0397, Sample ID = 03

ZE KWAMEN AAN IN DUBLIN OM 7 UUR
EN IN AMSTERDAM OM 9.40 UUR
'S AVONDS. DE FIAT VAN BOB EN DE
VW VAN DAVID STONDEN IN R3 VAN

Rank = 4, Dist = 0.032940, Writer ID = 0397, Sample ID = 02

NADAT ZE IN New York, Tokyo, QUÉBEC,
PARIJS, ZÜRICH EN OSLO WAREN GEWEEST,
VLOGEN ZE UIT DE USA TERUG MET VLUCHT
KL 658 OM 12 UUR.

Rank = 5, Dist = 0.034000, Writer ID = 0477, Sample ID = 03

ZE KWAMEN AAN IN DUBLIN OM 7 UUR
EN IN AMSTERDAM OM 9.40 UUR 'S AVONDS.
DE FIAT VAN BOB EN DE VW VAN DAVID
STONDEN IN R3 VAN HET PARKEERTERREIN.

Rank = 6, Dist = 0.034221, Writer ID = 0506, Sample ID = 02

NADAT ZE IN NEW York, Tokyo, QUÉBEC,
PARIJS, ZURICH EN OSLO WAREN
GEWEEST, VLOGEN ZE UIT DE USA TERUG
MET VLUCHT KL 658 OM 12 UUR.

Rank = 7, Dist = 0.037437, Writer ID = 0506, Sample ID = 03

ZE KWAMEN AAN IN DUBLIN OM
7 UUR EN IN AMSTERDAM OM 9.40 UUR
'S AVONDS. DE FIAT VAN BOB EN DE
VW VAN DAVID STONDEN IN R3 VAN

Rank = 8, Dist = 0.037895, Writer ID = 0432, Sample ID = 03

ZE KWAMEN AAN IN DUBLIN OM 7 UUR EN
IN AMSTERDAM OM 9 40 UUR 'S AVONDS.
DE FIAT VAN BOB EN DE VW VAN DAVID
STONDEN IN R3 VAN HET PARKEERTERREIN.

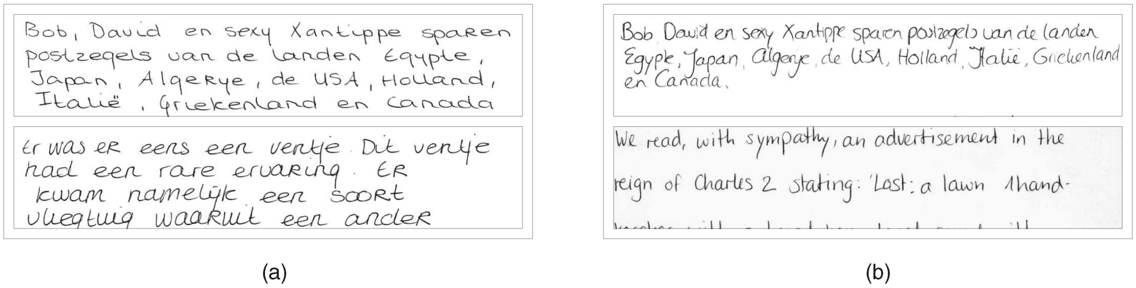
Rank = 9, Dist = 0.041108, Writer ID = 0551, Sample ID = 02

NADAT ZE IN NEW YORK, TOKYO, QUÉBEC,
PARYS, ZÜRICH EN OSLO WAREN GEWEEST,
VLOGEN ZE UIT DE USA TERUG MET
VLUCHT KL 658 OM 12 UUR.

Rank = 10, Dist = 0.041615, Writer ID = 0405, Sample ID = 02

NAdAt ZE IN NEW YORK, TOKYO, QUÉBEC,
PARIJS, ZÜRICH, EN OSLO WAREN
GEWEEST, VLOGEN ZE UIT DE USA
TERUG MET VLUCHT KL 658 OM

Fig. 12. A successful writer identification search using the best performing individual feature *f2* on the Firemaker uppercase data set. The correct hit is in the first position and the handwriting style is uniform across the hit list.



Bob, David en sexy Xantippe sparen
postzegels van de landen Egypte,
Japan, Algerye, de USA, Holland,
Italië, Griekenland en Canada

Er was er eens een ventje. Dit ventje
had een rare ervaring. Er
kwam namelijk een soort
vliegtuig waaruit een ander

(a)

Bob David en sexy Xantippe sparen postzegels van de landen
Egypte, Japan, Algerye, de USA, Holland, Italie, Griekenland
en Canada.

We read, with sympathy, an advertisement in the
reign of Charles 2 stating: 'Lost: a lawn 1hand-

(b)

Fig. 13. (a) Example of a *false reject* error in writer verification: The two samples were written by the same person, but the system wrongly decided the opposite. (b) Example of a *false accept* error in writer verification: The two samples were written by different people and the system made the wrong decision in this case.

correct hit is now in position 1 and the handwriting style is homogeneous across the hit list. Fig. 12 shows another successful hit list generated by the contour-hinge feature *f2* applied on the Firemaker uppercase data set.

It is also interesting to see the writer verification errors produced by GRAWIS and to visually judge the resemblance between the handwritings being compared. Fig. 13a shows a false reject error and Fig. 13b shows a false accept error. These examples were selected to illustrate very problematic cases where the within-writer variability arguably exceeds the between-writer variability at the fringes of the Bayes decision boundary in the writer verification task.

8 Discussion

The analyzed features are not complete: Feature extraction is a lossy operation and, thus, starting from the feature values, a total reconstruction of the input handwriting image is not possible. On the other hand, this is also not desirable as we are interested in text-independent methods for writer identification and verification. Our features used to encode individual handwriting style are independent of the textual content of the handwritten sample. The handwriting is merely seen as a texture characterized by joint directional probability distributions or as a simple stochastic shape-emission process characterized by a grapheme occurrence probability.

The directional PDFs (*f1*, *f2*, *f3*) operate at the scale of the ink-trace width and implement a local phase analysis yielding results that are significantly better than those of the autocorrelation feature (*f6*) capturing amplitude information. The writer-specific shape-emission PDF (*f4*) operates at the scale of characters. Combining information across multiple scales by feature fusion results in sizeable performance improvements. The presented fusion method based on simple distance averaging diminishes the risk of a biased solution while capturing most of the achievable increases in writer identification and verification performance.

We accomplished a more in-depth analysis of our allograph-level method. The computation of feature *f4* depends on two important issues: the size of the shape codebook and the clustering algorithm used to generate

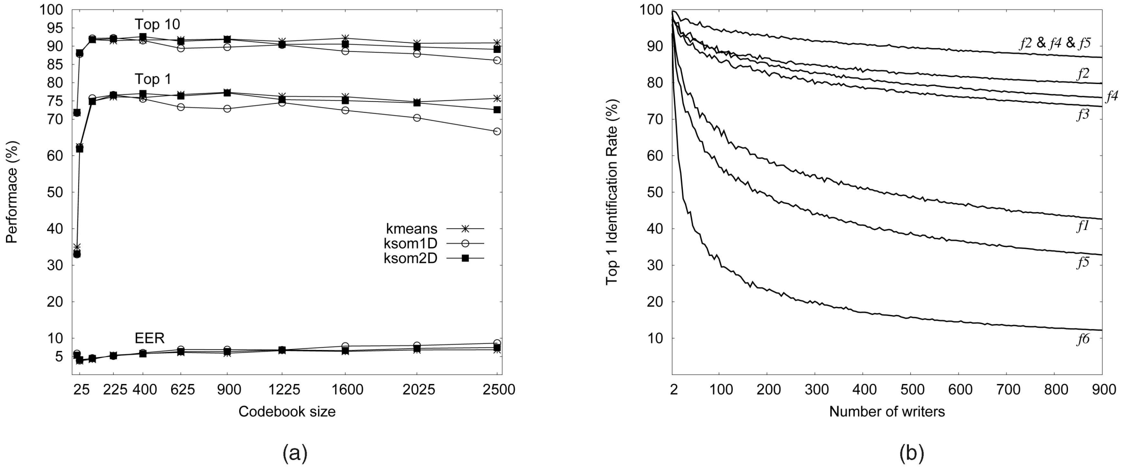


Fig. 14. (a) Performance versus clustering method and codebook size for the grapheme-based writer identification and verification method (feature *f4*) on the Large data set. (b) Top-1 identification rate versus number of writers contained in the test. For every size of the writer set, the results were averaged over 50 random draws from the Large data set.

the codebook. We have run large-scale computational experiments to compare three clustering methods over a large range of codebook sizes: k-means, Kohonen SOM 1D, and 2D (see Fig. 6). The number of clusters used was varied from 9 (3×3) to 2,500 (50×50). A number of 200 epochs has been used for training the Kohonen SOMs. Computations have been run on a Beowulf high-performance Linux cluster with 1.7 GHz/0.5 GB nodes. Training times for codebooks of size 400 were: k-means—1 hrs, ksom1D —10 hrs, ksom2D—17 hrs. Computation times for the grapheme emission PDF on codebooks of size 400 were: k-means—0.5 s/sample, ksom1D—1.5 s/sample, ksom2D—3.1 s/sample. These computation times were obtained using the "gcc" compiler with optimization for single-precision floating-point calculations. The total computation time used in the experiments amounts to approximately 800 CPU hrs.

Fig. 14 a shows our results obtained on the Large data set. It is remarkable that the same performance is achieved by all three clustering methods and that performance is stable over a large range of codebook sizes. Writer identification rates (Top-1 and Top-10) reach a plateau for codebook sizes larger than about 100 (10×10) shapes. The writer verification EER reaches a minimum of about 4 percent for a codebook size of 100 and increases to about 7 percent for larger codebooks. This effect can be explained considering that, as the codebook size increases, the grapheme emission PDFs reside in increasingly higher dimensional spaces that progressively become less and less populated. The distances between the individual handwriting samples increase in relative terms. As a result, it gradually becomes more difficult to find a unique threshold distance that separates the sample pairs written by the same person from those written by different people. Clearly, an individualized threshold is needed that depends on the variability in feature space of the handwriting belonging to that particular person. However, estimating this within-writer variability using a limited amount of handwritten material is a difficult problem that requires further research. The described dimensionality problem does not significantly affect the distance rankings with respect to a chosen sample and, consequently, writer identification performance remains essentially stable over a large range of codebook sizes. Similar results were also found on the other test data sets [36].

The results reported for the grapheme-emission PDF (feature *f4*) in the previous sections of the paper were obtained using a codebook generated by k-means clustering and containing 400 graphemes, which was chosen as an anchor point. The grapheme codebook is obtained much faster using k-means instead of Kohonen training. The writer identification results presented here are in the same ballpark as the ones we reported in a previous study using contours (rather than normalized bitmaps) for shape representation and Kohonen 2D for codebook training [35].

The grapheme codebook spans the shape space of the possible allographic parts encountered in handwritten samples as a result of the ink segmentation procedure. The three clustering methods considered here seem to perform the task of selecting representative graphemes adequate for constructing shape-occurrence PDFs informative about writer identity equally well. The proposed allograph-level method proves to be robust to the underlying shape representation used (whether contours or bitmaps), to the size of codebook used (stable performance for sizes 10^2 to 2.5×10^3), and to the clustering method used to generate the codebook (the same performance was obtained for k-means, ksom1D, and ksom2D).

In order to complete our study, another necessary analysis was carried out evaluating how the identification performance (Top-1 and Top-10) depends on the number of writers contained in the test data set. We determined this relationship by experiment using the Large data set: For each size of the writer set (up to 900 writers), 50 identification tests were performed on random selections of writers and the results were averaged. Fig. 14b shows the Top-1 identification rate as a function of the number of writers for individual features and for the feature combination *f2* & *f4* & *f5*. Naturally, the identification rate decreases as the number of writers grows. However, the decline is not severe. In the range studied, for the best performing feature combination, *f2* & *f4* & *f5*, we observe that the Top-1 identification rate drops by approximately 2-3 percent for every doubling of the number of writers in the data set. Our writer identification system shows usable performance for 10^3 writer sets. Undoubtedly, further experiments with larger numbers of writers are needed in order to approach the 10^4 scale of the actual forensic databases.

The writer identification experiments reported in this paper always involved two samples per writer: One was used as the query, while the other one represented the correct hit that the system was supposed to find in the database. Having more samples per writer enrolled in the database increases the chance of finding in the top positions of the hit list the correct author for a given query. We have run writer identification tests on the original IAM database that included at least three samples per writer for about a quarter of the total of 650 writers incorporated in the set. For the best performing feature combination, *f2* & *f4* & *f5*, we obtained writer identification rates of Top-1 92 percent and Top-10 98 percent. These values exceed the identification rates obtained on our modified IAM set that always contained only two samples per writer (see Table 4).

In another study performed on a subset comprised of 100 writers from the Firemaker data set, our methods largely outperformed two actual systems used in current forensic practice [9]. The use of automatic and computation-intensive approaches in this application domain will allow for massive search in large databases, with less human intervention than is current practice. By reducing the size of a target set of writers, detailed manual and microscopic forensic analysis becomes feasible. In the foreseeable future, the toolbox of the forensic expert will have been thoroughly modernized and extended. Part of our directional texture-level features have already been included in real-life applications.

It is important to note that the methods described in this paper are equally applicable to handwriting as well as machine print: writer identification versus font identification (e.g., for OCR). Besides the forensic field, interesting potential applications are in the domain of historic document analysis: identification of scribes or manuscript dating on medieval handwritten documents or identification of the printing house on historic prints. Furthermore, writer identification may be used in handwriting recognition as a preprocessing step, allowing the use of dedicated recognizers specialized to one writer or to a limit group of writers with similar handwriting styles.

9 CONCLUSIONS

The writer identification and verification methods described in this paper exploit two essential sources of behavioral information regarding handwriting individuality. First, habitual pen grip and preferred writing slant and curvature are reflected in the directional texture-level features that operate in the angular domain at the scale of the ink-trace width. Second, the personalized set of allographs that each person uses in writing is captured by the grapheme occurrence probability. This feature works in the Cartesian domain at the scale of the character shapes.

The proposed features are probability distributions extracted from the handwriting images and offer a text-independent and robust characterization of individual handwriting style. They have practical feasibility and they are applicable to free-style handwriting, both cursive and isolated. Combining texture-level and allograph-level features yields very high writer identification and verification performance, with usable rates for data sets containing 10^3 writers.

The challenge is to integrate the recent developments in this field of behavioral biometrics into the real writer identification systems of the future.

REFERENCES

[1] R. Plamondon and S.N. Srihari, "Online and Offline Handwriting Recognition: A Comprehensive Survey," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 1, pp. 63-84, Jan. 2000.

[2] A. Vinciarelli, "A Survey on Offline Cursive Word Recognition," *Pattern Recognition,* vol. 35, pp. 1433-1446, 2002.

[3] R.A. Huber and A. Headrick, *Handwriting Identification: Facts and Fundamentals.* CRC Press, 1999.

[4] R. Morris, *Forensic Handwriting Identification: Fundamental Concepts and Principles.* Academic Press, 2000.

[5] F.J. Maarse, "The Study of Handwriting Movement: Peripheral Models and Signal Processing Techniques," PhD dissertation, Dept. of Experimental Psychology, Univ. of Nijmegen, The Netherlands, 1987.

[6] F. Maarse and A. Thomassen, "Produced and Perceived Writing Slant: Differences between Up and Down Strokes," *Acta Psychologica,* vol. 54, nos. 1-3, pp. 131-147, 1983.

[7] L. Schomaker, "Simulation and Recognition of Handwriting Movements: A Vertical Approach to Modeling Human Motor Behavior," PhD dissertation, Univ. of Nijmegen, The Netherlands, 1991.

[8] R. Schmidt, "A Schema Theory of Discrete Motor Skill Learning," *Psychological Rev.,* vol. 82, pp. 225-260, 1975.

[9] L. Schomaker and M. Bulacu, "Automatic Writer Identification Using Connected-Component Contours and Edge-Based Features of Uppercase Western Script," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 6, pp. 787-798, June 2004.

[10] R. Plamondon and G. Lorette, "Automatic Signature Verification and Writer Identification—The State of the Art," *Pattern Recognition,* vol. 22, no. 2, pp. 107-131, 1989.

[11] H. Said, T. Tan, and K. Baker, "Personal Identification Based on Handwriting," *Pattern Recognition,* vol. 33, no. 1, pp. 149-160, 2000.

[12] H. Said, G. Peake, T. Tan, and K. Baker, "Writer Identification from Non-Uniformly Skewed Handwriting Images," *Proc. Ninth British Machine Vision Conf.,* pp. 478-487, 1998.

[13] T. Tan, "Rotation Invariant Texture Features and Their Use in Automatic Script Identification," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 7, pp. 751-756, July 1998.

[14] Y. Zhu, T. Tan, and Y. Wang, "Font Recognition Based on Global Texture Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 10, pp. 1192-1200, Oct. 2001.

[15] E. Zois and V. Anastassopoulos, "Morphological Waveform Coding for Writer Identification," *Pattern Recognition,* vol. 33, no. 3, pp. 385-398, Mar. 2000.

[16] S. Srihari, S. Cha, H. Arora, and S. Lee, "Individuality of Handwriting," *J. Forensic Sciences,* vol. 47, no. 4, pp. 1-17, July 2002.

[17] S. Srihari, M. Beal, K. Bandi, V. Shah, and P. Krishnamurthy, "A Statistical Model for Writer Verification," *Proc. Eighth Int'l Conf. Document Analysis and Recognition (ICDAR),* pp. 1105-1109, 2005.

[18] J. Favata and G. Srikantan, "A Multiple Feature/Resolution Approach to Handprinted Digit and Character Recognition," *Int'l J. Imaging Systems and Technology,* vol. 7, pp. 304-311, 1996.

[19] B. Zhang, S. Srihari, and S. Lee, "Individuality of Handwritten Characters," *Proc. Seventh Int'l Conf. Document Analysis and Recognition (ICDAR),* pp. 1086-1090, 2003.

[20] S. Srihari, C. Tomai, B. Zhang, and S. Lee, "Individuality of Numerals," *Proc. Seventh Int'l Conf. Document Analysis and Recognition (ICDAR),* pp. 1096-1100, 2003.

[21] B. Zhang and S. Srihari, "Analysis of Handwritten Individuality Using Word Features," *Proc. Seventh Int'l Conf. Document Analysis and Recognition (ICDAR),* pp. 1142-1146, 2003.

[22] C. Tomai, B. Zhang, and S. Srihari, "Discriminatory Power of Handwritten Words for Writer Recognition," *Proc. 17th Int'l Conf. Pattern Recognition,* pp. 638-641, 2004.

[23] A. Bensefia, T. Paquet, and L. Heutte, "A Writer Identification and Verification System," *Pattern Recognition Letters,* vol. 26, no. 10, pp. 2080-2092, Oct. 2005.

[24] A. Bensefia, T. Paquet, and L. Heutte, "Handwritten Document Analysis for Automatic Writer Recognition," *Electronic Letters on Computer Vision and Image Analysis,* vol. 5, no. 2, pp. 72-86, Aug. 2005.

[25] A. Bensefia, A. Nosary, T. Paquet, and L. Heutte, "Writer Identification by Writer's Invariants," *Proc. Eighth Int'l Workshop Frontiers in Handwriting Recognition,* pp. 274-279, Aug. 2002.

[26] A. Bensefia, T. Paquet, and L. Heutte, "Information Retrieval Based Writer Identification," *Proc. Seventh Int'l Conf. Document Analysis and Recognition (ICDAR),* pp. 946-950, Aug. 2003.

[27] U.-V. Marti, R. Messerli, and H. Bunke, "Writer Identification Using Text Line Based Features," *Proc. Sixth Int'l Conf. Document Analysis and Recognition (ICDAR),* pp. 101-105, Sept. 2001.

[28] C. Hertel and H. Bunke, "A Set of Novel Features for Writer Identification," *Proc. Fourth Int'l Conf. Audio and Video-Based Biometric Person Authentication,* pp. 679-687, 2003.

[29] A. Schlapbach, V. Kilchherr, and H. Bunke, "Improving Writer Identification by Means of Feature Selection and Extraction," *Proc. Eighth Int'l Conf. Document Analysis and Recognition (ICDAR),* pp. 131-135, Aug.-Sept. 2005.

[30] U.-V. Marti and H. Bunke, "The IAM-Database: An English Sentence Database for Offline Handwriting Recognition," *Int'l J. Document Analysis and Recognition,* vol. 5, no. 1, pp. 39-46, 2002.

[31] A. Schlapbach and H. Bunke, "Using HMM-Based Recognizers for Writer Identification and Verification," *Proc. Ninth Int'l Workshop Frontiers in Handwriting Recognition,* pp. 167-172, Oct. 2004.

[32] U.-V. Marti and H. Bunke, "Using a Statistical Language Model to Improve the Performance of an HMM-Based Cursive Handwriting Recognition System," *Int'l J. Pattern Recognition and Artificial Intelligence,* vol. 15, pp. 65-90, 2001.

[33] M. Bulacu, L. Schomaker, and L. Vuurpijl, "Writer Identification Using Edge-Based Directional Features," *Proc. Seventh Int'l Conf. Document Analysis and Recognition,* pp. 937-941, Aug. 2003.

[34] M. Bulacu and L. Schomaker, "Writer Style from Oriented Edge Fragments," *Proc. 10th Int'l Conf. Computer Analysis of Images and Patterns,* pp. 460-469, Aug. 2003.

[35] L. Schomaker, M. Bulacu, and K. Franke, "Automatic Writer Identification Using Fragmented Connected-Component Contours," *Proc. Ninth Int'l Workshop Frontiers in Handwriting Recognition (IWFHR),* pp. 185-190, Oct. 2004.

[36] M. Bulacu and L. Schomaker, "A Comparison of Clustering Methods for Writer Identification and Verification," *Proc. Eighth Int'l Conf. Document Analysis and Recognition,* vol. II, pp. 1275-1279, Aug.-Sept. 2005.

[37] L. van der Maaten and E. Postma, "Improving Automatic Writer Identification," *Proc. 17th Belgium-Netherlands Conf. Artificial Intelligence,* pp. 260-266, 2005.

[38] M. van Erp, L. Vuurpijl, K. Franke, and L. Schomaker, "The WANDA Measurement Tool for Forensic Document Examination," *Proc. Conf. Int'l Graphonomics Soc.,* pp. 282-285, 2003.

[39] K. Franke and M. Köppen, "A Computer-Based System to Support Forensic Studies on Handwritten Documents," *Int'l J. Document Analysis and Recognition,* vol. 3, no. 4, pp. 218-231, 2001.

[40] M. Fairhurst, "Document Identity, Authentication and Ownership: the Future of Biometric Verification," *Proc. Seventh Int'l Conf. Document Analysis and Recognition,* vol. II, pp. 1108-1116, Aug. 2003.

[41] L. Schomaker and L. Vuurpijl, "Forensic Writer Identification: A Benchmark Data Set and a Comparison of Two Systems," technical report, Nijmegen: NICI, 2000.

[42] M. Zimmermann and H. Bunke, "Automatic Segmentation of the IAM Offline Database for Handwritten English Text," *Proc. 16th Int'l Conf. Pattern Recognition,* vol. 4, pp. 35-39, 2002.

[43] I. Guyon, L. Schomaker, R. Plamondon, R. Liberman, and S. Janet, "UNIPEN Project of Online Data Exchange and Recognizer Benchmarks," *Proc. 12th Int'l Conf. Pattern Recognition,* pp. 29-33, Oct. 1994.

[44] N. Otsu, "A Threshold Selection Method from Gray-Level Histogram," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 9, pp. 62-69, 1979.

[45] F. Maarse, L. Schomaker, and H.-L. Teulings, "Automatic Identification of Writers," *Human-Computer Interaction: Psychonomic Aspects,* pp. 353-360, 1988.

[46] J.-P. Crettez, "A Set of Handwriting Families: Style Recognition," *Proc. Third Int'l Conf. Document Analysis and Recognition,* pp. 489-494, Aug. 1995.

[47] J. Drouhard, R. Sabourin, and M. Godbout, "A Comparative Study of the k Nearest Neighbours, Threshold and Neural Network Classifiers for Handwritten Signature Verification Using an Enhanced Directional PDF," *Proc. Third Int'l Conf. Document Analysis and Recognition,* pp. 807-810, 1995.

[48] B. Arazi, "Handwriting Identification by Means of Run-Length Measurements," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 7, no. 12, pp. 878-881, 1977.

[49] B. Arazi, "Automatic Handwriting Identification Based on the External Properties of the Samples," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 13, no. 4, pp. 635-642, 1983.

[50] I. Dinstein and Y. Shapira, "Ancient Hebraic Handwriting Identification with Run-Length Histograms," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 12, no. 3, pp. 405-409, 1982.

[51] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification,* second ed. Wiley Interscience, 2001.

[52] T. Kohonen, *Self-Organization and Associative Memory,* second ed. Springer Verlag, 1988.

[53] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C: The Art of Scientific Computing,* second ed. Cambridge Univ. Press, 1992.

[54] J. Neyman and E. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Philosophical Trans. Royal Soc. A,* vol. 231, pp. 289-337, 1933.

[55] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On Combining Classifiers," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 3, pp. 226-239, Mar. 1998.

[56] D. Maltoni, D. Maio, A.K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition.* Springer, 2003.

[57] F. Roli, J. Kittler, G. Fumera, and D. Muntoni, "An Experimental Comparison of Classifier Fusion Rules for Multimodal Personal Identity Verification Systems," *Proc. Conf. Multiple Classifier Systems,* pp. 325-336, 2002.

[58] J.G. Daugman, "Biometric Decision Landscapes," Technical Report TR482, Computer Laboratory, Univ. of Cambridge, Jan. 2000.

[59] J.G. Daugman, "High Confidence Visual Recognition of Persons by a Test of Statistical Independence," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 11, pp. 1148-1160, Nov. 1993.

[60] A. Jain, L. Hong, and R. Bolle, "Online Fingerprint Verification," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 4, pp. 302-314, Apr. 1997.

[61] T. Joachims, "Making Large-Scale SVM Learning Practical," *Advances in Kernel Methods—Support Vector Learning,* MIT Press, 1999.

[62] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines.* Cambridge Univ. Press, 2000.

Marius Bulacu received the BSc and MSc degrees in physics from the University of Bucharest, Romania, in 1997 and 1998, respectively. He did teaching and research in the Biophysics Department, Faculty of Physics, University of Bucharest from 1999 to 2002. Since March 2002, he has been with the Artificial Intelligence Institute of the University of Groningen, The Netherlands, pursuing the PhD degree. He is currently working on developing statistical pattern recognition methods for automatic writer identification and for handwritten historical document retrieval. His scientific interests include computer vision, statistical pattern recognition, biometrics, and document analysis and recognition. He is a student member of the IEEE and the IEEE Computer Society.

Lambert Schomaker received the MSc degree cum laude in psychophysiological psychology in 1983 and the PhD degree on the simulation and recognition of pen movement in handwriting in 1991 from Nijmegen University, The Netherlands. Since 1988, he has worked in several European research projects concerning the recognition of online, connected cursive script, and multimodal multimedia interfaces. Current projects are in the area of image-based retrieval, historical handwritten document analysis, and forensic handwriting analysis systems. Professor Schomaker is a member of the IEEE, the IEEE Computer Society, and the IAPR. He has contributed to more than 60 reviewed publications in journals and books. In 2001, he accepted the position of full professor and director of the AI Institute at Groningen University, The Netherlands.