

Statistical Pattern Recognition for Automatic Writer Identification and Verification

Marius Lucian Bulacu

Front cover: A schematic description of the extraction of the "hinge" feature.
Back cover: A 20x20 Kohonen self-organized map of handwritten graphemes.

RIJKSUNIVERSITEIT GRONINGEN

**Statistical Pattern Recognition for Automatic
Writer Identification and Verification**

Proefschrift

ter verkrijging van het doctoraat in de
Gedrags- en Maatschappijwetenschappen
aan de Rijksuniversiteit Groningen

op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
donderdag 15 maart 2007
om 16.15 uur

door

Marius Lucian Bulacu

geboren op 8 oktober 1973
te Boekarest (Roemenië)

Promotor: Prof. dr. L. Schomaker

Beoordelingscommissie: Prof. dr. J. Daugman
Prof. dr. F. Groen
Prof. dr. J. Roerdink

ISBN: 90-367-2912-2

To my parents

Contents

1	Introduction	1
1.1	Writer identification as a behavioral biometric	2
1.2	Writer identification in forensics	2
1.3	Writer identification vs. Handwriting recognition	4
1.4	Writer identification vs. Writer verification	4
1.5	Text-dependent vs. Text-independent methods	5
1.6	Within-writer variance vs. Between-writer variation	6
1.7	Factors causing variability in handwriting	7
1.8	Factors determining individuality of handwriting	9
1.9	A survey of recent research in the field	12
1.10	Main assumptions underlying the methods proposed in the thesis	15
1.11	Overview of the thesis	16
I	Texture-Level Approach	19
2	Writer Identification Using Edge-Based Directional Features	21
2.1	Introduction	21
2.2	Experimental data	22
2.3	Features	23
2.3.1	Edge-direction distribution	23
2.3.2	A new feature: edge-hinge distribution	25
2.3.3	Run-length distributions	27
2.3.4	Autocorrelation	28
2.3.5	Entropy	28
2.4	Results	28

2.4.1	Evaluation method	28
2.4.2	Analysis of performances	30
2.4.3	Stability test	32
2.5	Conclusions	32
3	Characterizing Handwriting Individuality Using Localized Oriented Edge Fragments	35
3.1	Introduction	35
3.2	Experimental data	36
3.3	Feature extraction	37
3.3.1	Edge-direction distribution	39
3.3.2	Edge-hinge distribution	40
3.3.3	Run-length distributions	41
3.3.4	A new feature: horizontal co-occurrence of edge angles	41
3.3.5	Brush function: ink density distribution	42
3.4	Results	42
3.4.1	Comparison lower- vs. upper-case and entire- vs. split-line	43
3.4.2	Voting feature combination	48
3.5	Conclusions	48
II	Allograph-Level Approach	51
4	Grapheme Clustering for Writer Identification and Verification	53
4.1	Introduction	53
4.2	Theoretical model	55
4.3	Datasets	57
4.4	Segmentation method	59
4.5	Grapheme codebook generation	60
4.6	Computing writer-specific grapheme-emission PDFs	61
4.7	Results	62
4.7.1	Writer identification	62
4.7.2	Writer verification	64
4.8	Conclusions	67
5	Feature Fusion for Text-Independent Writer Identification and Verification	69
5.1	Introduction	69
5.2	Experimental datasets	73
5.3	Textural features	75
5.3.1	Contour-direction PDF (f_1)	78

Contents

5.3.2	Contour-hinge PDF (f_2)	79
5.3.3	Direction co-occurrence PDFs (f_{3h}, f_{3v})	81
5.3.4	Other texture-level features: run-length PDFs (f_{5h}, f_{5v}), autocorrelation (f_6)	81
5.4	Allographic features	83
5.5	Feature matching and fusion for writer identification and verification . . .	85
5.6	Results	88
5.6.1	Performances of individual features	88
5.6.2	Performances of feature combinations	91
5.7	Discussion	95
5.8	Conclusions	100
6	Concluding the thesis	101
6.1	Summary and Contributions	101
6.2	Further research directions	105
A	GRAWIS: Groningen Automatic Writer Identification System	107
A.1	Introduction	107
A.2	Visualization tool	107
A.3	Examples of writer identification hit lists	108
A.4	Examples of writer verification errors	108
	Bibliography	117
	Index	127
	Publications	131
	Samenvatting	133
	Acknowledgements	139

Chapter 1

Introduction

Truth is what stands the test of experience.

Albert Einstein

This thesis addresses the problem of automatic person identification using scanned images of handwriting. Identifying the author of a handwritten sample using automatic image-based methods is an interesting pattern recognition problem with direct applicability in the forensic and historic document analysis fields. Approaching this challenging problem raises a number of important research themes in computer vision:

- How can individual handwriting style be characterized using computer algorithms?
- What representations or features are most appropriate and how can they be combined?
- What performance can be achieved using automatic methods?

The current study describes a number of new and very effective techniques that we have developed for automatic writer identification and verification. The goal of our research was to design state-of-the-art automatic methods involving only a reduced number of adjustable parameters and to create a robust writer identification system capable of managing hundreds to thousands of writers.

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 undergo a transition from research to real full-scale deployment. Our methods also have practical feasibility and hold the promise of concrete applicability.

The writer identification techniques proposed in this thesis have possible impact in forensic science. Our methods are statistically evaluated using large datasets with handwriting samples collected from up to 900 subjects.

1.1 Writer identification as a behavioral biometric

Biometric modalities are classified into two broad categories: *physiological biometrics* that perform person identification based on measuring a physical property of the human body (e.g. fingerprint, face, iris, retinal blood vessels, hand geometry, DNA) and *behavioral biometrics* that use individual traits of a person's behavior for identification (e.g. voice, gait, keystroke dynamics, signature, handwriting). Writer identification therefore pertains to the category of behavioral biometrics. From the physical body property or the individual behavior traits, biometric templates are extracted and used in the identification process. Biometric identification is performed by comparing the biometric template measured at the moment when the identification of an unknown person is needed with templates previously enrolled in a database and linked with certainty to known persons.

Physiological biometrics, like fingerprint (Jain et al. 1997, Moler et al. 1998), iris (Daugman 1993, Daugman 2003) or DNA (Devlin et al. 1992, Benecke 1997), 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 are less invasive, but the achievable performance is less impressive due to the large variability of the behavior-derived biometric templates.

Leading a worrisome life among the harder forms of biometrics, the identification of a person on the basis of handwriting samples still remains a useful biometric modality, mainly due to its applicability in the forensic field.

1.2 Writer identification in forensics

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.

The *target performance* for writer identification systems is less impressive than in the case of DNA or iris-based person identification. In forensic writer identification, as a rule of thumb, one strives for a near-100% recall of the correct writer in a hit list of

100 writers, computed from a database in the order of 10^4 samples, the size of search sets in current European forensic databases. This amount is based on the pragmatic consideration that a number of one hundred suspects is just about manageable in the criminal-investigation process. This target performance still remains an ambitious goal.

Recent advances in image processing, pattern classification and computer technology at large have provided the context in which our research was carried out. The writer identification techniques that we developed accomplished substantial improvements in performance and have potential applicability in forensic practice.

There exist three groups of script-shape features which are derived from scanned handwritten samples in forensic procedures:

1. Fully automatic features computed from a region of interest in the image;
2. Interactively measured features by human experts using a dedicated graphical user-interface tool;
3. Character-based features which are related to the allograph subset which is being generated by each writer.

The complete process of forensic writer identification is never fully automatic. The features pertaining to groups 2 and 3 require some form of intensive human involvement in executing predefined measuring actions on the script image or in isolating and labeling individual characters or words. Two examples of actual forensic writer identification systems are *Fish* (Philipp 1996) and *Script* (de Jong et al. 1994).

Although requiring less human labor, the first group of features has been treated with some skepticism by practitioners within the application domain, given the complexity of the real-life scanned samples of handwriting that are collected in practice. Indeed, automatic foreground/background separation will often fail on the smudged and texture-rich fragments, where the ink trace is often hard to identify. However, there are recent advances in image processing using "soft computing" methods, i.e., combining tools from fuzzy logic and genetic algorithms, which allow for advanced semi-interactive solutions to the foreground/background separation process (Franke and Köppen 1999, Franke and Köppen 2001, Köppen and Franke 1999). Under these conditions, and assuming the presence of sufficient computing power, the use of automatically computed image features (group 1 from above) is becoming feasible.

The current thesis sets out to explore precisely this category of automatic features. It is implicitly assumed that a crisp foreground/background separation has already been realized in a pre-processing phase, yielding a white background with (near-) black ink.

1.3 Writer identification vs. Handwriting recognition

Writer identification is rooted in the older and broader automatic handwriting recognition domain. For automatic handwriting recognition, invariant representations are sought which are capable of eliminating variations between different handwritings in order to classify the shapes of characters and words robustly. The problem of writer identification, on the contrary, requires a specific enhanced representation of these variations, which, per se, are characteristic to a writer's hand.

Due to its very large applicability, handwriting recognition has always dominated research in handwriting analysis. Writer identification received renewed interest in the last several years, after 9/11 and the anthrax letters (Schomaker and Bulacu 2004, Srihari et al. 2002, Bensefia et al. 2005b, Schlapbach and Bunke 2004, Said et al. 2000, Zois and Anastassopoulos 2000).

The goal in handwriting recognition is to obtain invariance and generalization. For writer identification, one strives for quite the opposite with the aim to maximally expose the specificity of individual handwriting style for writer discrimination.

It is important, however, to mention the idea that writer identification could reduce certain ambiguities in the pattern recognition process if information on the writer's general writing habits and idiosyncrasies is available to the handwriting recognition system (Maarse 1987, Crettez 1995).

A number of references on handwriting recognition are given in the bibliography section of this thesis: (Schomaker 1993, Bunke et al. 1995, Mohamed and Gader 1996, Parisse 1996, Schomaker 1998, Senior and Robinson 1998, Steinherz et al. 1999, El-Yacoubi et al. 1999, Plamondon et al. 1999, Plamondon and Srihari 2000, Marti and Bunke 2001, Favata 2001, Jaeger et al. 2001, Liu and Gader 2002, Xue and Govindaraju 2002, Vinciarelli 2002, Koerich et al. 2003, Vuurpijl et al. 2003, Gunter and Bunke 2004, Nosary et al. 2004, Vinciarelli et al. 2004).

1.4 Writer identification vs. Writer verification

Asserting writer identity based on handwriting images requires three main operational phases after image preprocessing:

- feature extraction
- feature matching / feature combination
- writer identification and verification

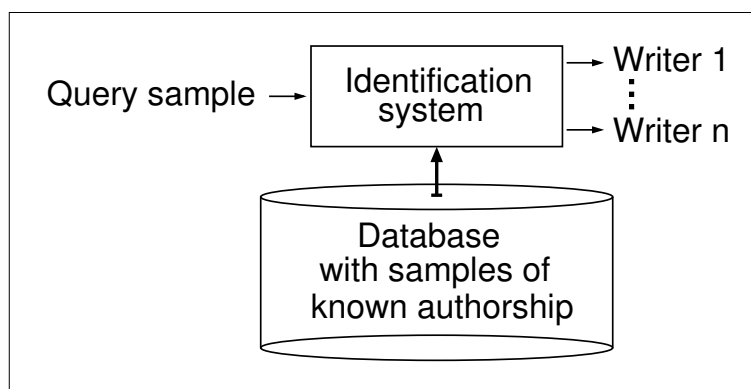


Figure 1.1: A writer identification system retrieves, from a database containing handwritings of known authorship, those samples that are most similar to the query. The hit list is then analyzed in detail by a human expert.

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. 1.1). This list is further scrutinized by the forensic expert who takes the final decision regarding the identity of the author of the questioned sample. *Writer verification* involves a one-to-one comparison with a decision whether or not the two samples are written by the same person (see Fig. 1.2). The decidability of this problem gives insight into the nature of handwriting individuality.

In writer identification searches, all the samples in the dataset are ordered with increasing dissimilarity (or distance) from the query sample. This represents a special case of image retrieval, where the retrieval process is based on features capturing handwriting individuality. In writer verification trials, if the distance between two chosen samples is smaller than a predefined threshold, the samples are deemed to have been written by the same person. Otherwise, the samples are considered to have been written by different writers. Writer verification has potential applicability in a scenario in which a specific writer must be automatically detected in a stream of handwritten documents. In forensic practice, both identification and verification play a central role.

1.5 Text-dependent vs. Text-independent methods

Writer identification and verification approaches fall into two broad categories: *text-dependent* vs. *text-independent* methods (Plamondon and Lorette 1989).

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. 1.3). These methods therefore require the prior localization and segmentation of the relevant information. This is usually performed interactively by a

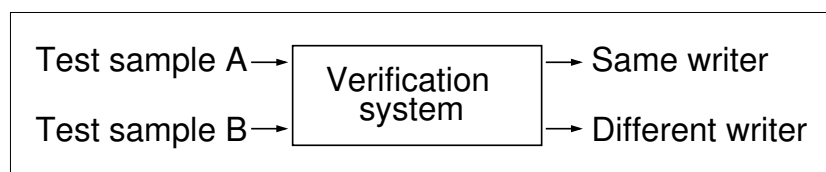


Figure 1.2: A writer verification system compares two handwriting samples and takes an automatic decision whether or not the input samples were written by the same person.

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.

Typical for the text-independent approaches and therefore a defining property of our approach as well, the features used for writer identification provide a lumped description of the whole region containing handwriting by discarding location information. For this reason, it is questionable to use text-independent methods also in the cases where the textual content of the samples is fixed and known.

1.6 Within-writer variance vs. Between-writer variation

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. The results reported in this thesis ultimately represent statistical analyses, on our datasets, of the relationship opposing the *between-writer variation* and the *within-writer variability* in feature space.

The present study assumes that the handwriting was produced using a natural writing attitude. It is important to observe that 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 allographs. Using detailed manual analysis, forensic experts are sometimes able to correctly identify a forged handwritten sample (Huber and Headrick 1999, Morris 2000). On the other hand, our proposed algorithms operate on the scanned handwriting faithfully considering all the graphical shapes encountered in the image under the premise that they are created by the habitual and natural script style of the writer.



Figure 1.3: A comparison of handwritten characters (allographs) and handwritten words from three different writers.

1.7 Factors causing variability in handwriting

Figure 1.4 shows four factors causing variability in handwriting (Schomaker 1998).

The first factor concerns the *affine transforms* (Fig. 1.4a), which are under voluntary control by the writer. Transforms of size, translation, rotation and shear are a nuisance, but not a fundamental obstacle in handwriting recognition or writer identification. In particular, *slant* (shear) constitutes a habitual parameter determined by pen grip and orientation of the wrist subsystem versus the fingers (Dooijes 1983).

The second factor concerns the *neuro-biomechanical variability* (Fig. 1.4b) which is sometimes referred to as "sloppiness space": the local context and physiological state determine the amount of effort that is spent on character-shape formation and determine the legibility of the written sample. In realizing the intended shape, a writer must send motor-control patterns which compensate for the low-pass filtering effects of the biomechanical end-effector. This category of variability sources also contains tremors and effects of psychotropic substances on motor-control processes in writing. As such,

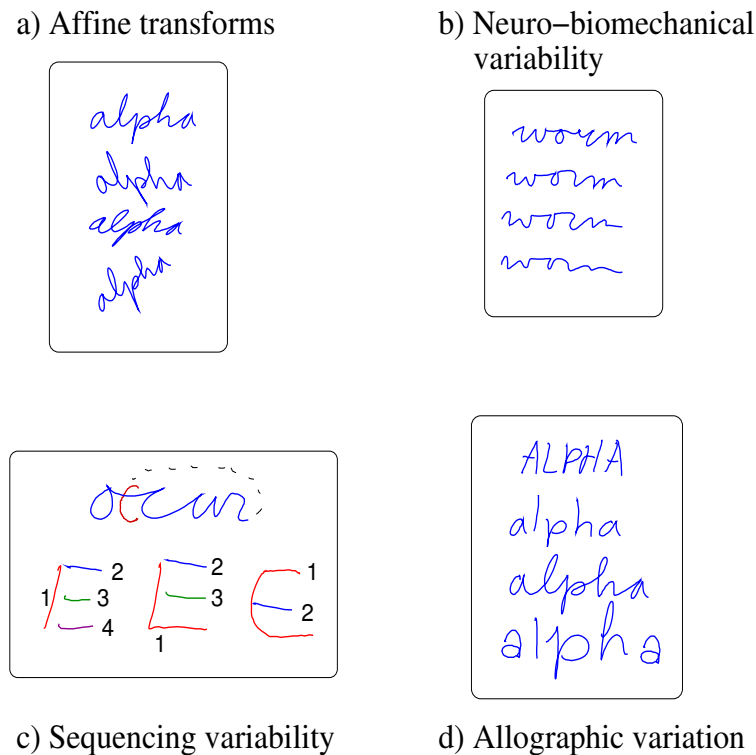


Figure 1.4: Factors causing handwriting variability: (a) Affine transforms are under voluntary control. However, writing slant constitutes a habitual parameter which may be exploited in writer identification; (b) neuro-biomechanical variability refers to the amount of effort which is spent on overcoming the low-pass characteristics of the biomechanical limb by conscious cognitive motor control; (c) sequencing variability becomes evident from stochastic variations in the production of the strokes in a capital *E* or of strokes in Chinese characters, as well as stroke variations due to slips of the pen; (d) allographic variation refers to individual use of character shapes. Factors b) and c) represent system state more than system identity. Affine transforms (a) and **allographic variation** (d) are the most useful sources of information in writer identification and verification.

this factor is more related to system state than system identity.

The third factor is also highly dependent on the instantaneous system state during the handwriting process and is represented by *sequencing variability* (Fig. 1.4c): the stroke order may vary stochastically, as in the production of a capital *E*. A four-stroked *E* can be produced in $4! * 2^4 = 384$ permutations. In the production of some Asian scripts, such as Hanzi, stochastic stroke-order permutations are a well-known problem in handwriting recognition (even though the training of stroke order at schools is rather strict). Finally, spelling errors may occur and lead to post-hoc editing strokes in the writing sequence. Although sequencing variability is generally assumed to pose a problem only for handwriting recognition based on temporal (on-line) signals, the example of post-hoc editing

(Fig. 1.4c) shows that static, optical effects are also a possible consequence of this form of variation.

The fourth factor, *allographic variation* (Fig. 1.4d and Fig. 1.3), refers to the phenomenon of writer-specific character shapes, which produces most of the problems in automatic script recognition, but at the same time provides essential information for automatic writer identification.

The handwriting of a person also changes with age and this constitutes another important variability factor. As a child grows, his handwriting becomes more comfortable, rapid, smooth, continuous, rhythmic and without hesitation. The amount of time a person spends writing may determine his general skill level and speed of writing. At older age, handwriting may become impaired due to chronic conditions that affect hand strength and dexterity.

1.8 Factors determining individuality of handwriting

As the writer matures, he departs from the copybook style learned in the classroom and progressively incorporates into his writing his own individuality. Especially nowadays, when there is less emphasis on penmanship in school.

There exist two fundamental factors contributing to the individuality of script: *genetic* (biological) and *memetic* (cultural) factors.

The first fundamental factor consists of the *genetic* make up of the writer. Genetic factors are known or may be hypothesized to contribute to handwriting style individuality:

- The biomechanical structure of the hand, i.e., the relative sizes of the carpal bones of wrist and fingers and their influence on pen grip;
- The left or right handedness (Francks et al. 2003);
- Muscular strength, fatigability, peripheral motor disorders (Gulcher et al. 1997);
- Central-nervous system (CNS) properties, i.e., aptitude for fine motor control and the CNS stability in motor-task execution (Van Galen et al. 1993).

The second factor consists of *memetic* or culturally transferred influences (Moritz 1990) on pen-grip style and the character shapes (allographs) which are trained during education or are learned from observation of the writings of other persons. Although the term *memetic* is often used to describe the evolution of ideas and knowledge, there does not seem to be a fundamental objection to view the evolution and spreading of character shapes as a memetic process: the fitness function of a character shape depends

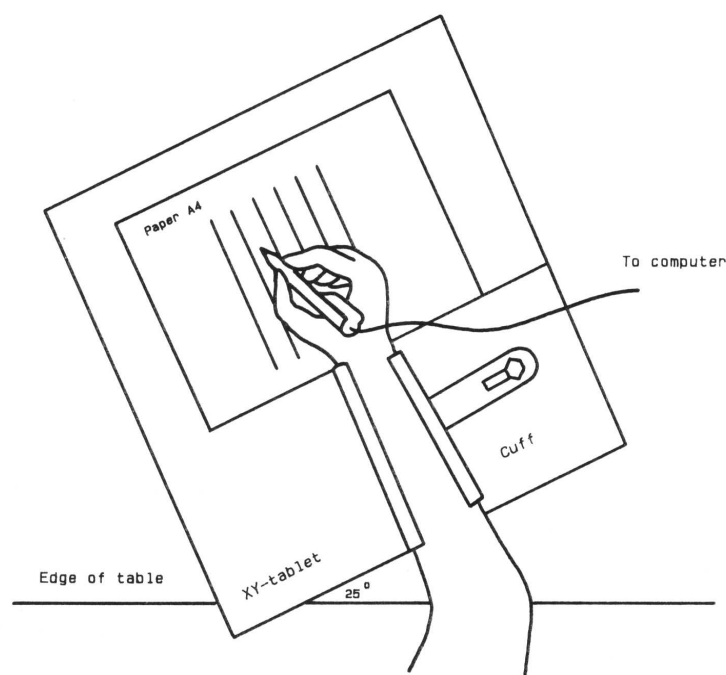


Figure 1.5: *Experimental set-up used by Maarse to isolate the thumb-fingers and hand-wrist biomechanical systems from the forearm motion. Reprinted from (Maarse 1987), with kind permission from the author.*

on the conflicting influences of (a) legibility and (b) ease of production with the writing tools (Jean 1997) which are available within a culture and society. The distribution of allographs over a writer population is heavily influenced by writing methods taught at school, which in turn depend on factors such as geographic distribution, religion and school types.

Together, the genetic and memetic factors determine a habitual writing process, with recognizable shape elements at the local level in the writing trace, at the level of the character shape as a whole and at the level of character placement and page layout. In this thesis, we will focus on the local level in the handwritten trace and on the character level.

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 (Maarse 1987). Writing consists of rapid movements of the fingers and the hand, and superimposed on this a slow progressive horizontal movement of the lower arm. In experiments performed by fixing the lower arm (Maarse 1987) (see Fig. 1.5), Maarse has studied on-line handwriting produced by using only two biophysical systems: one consisting of the

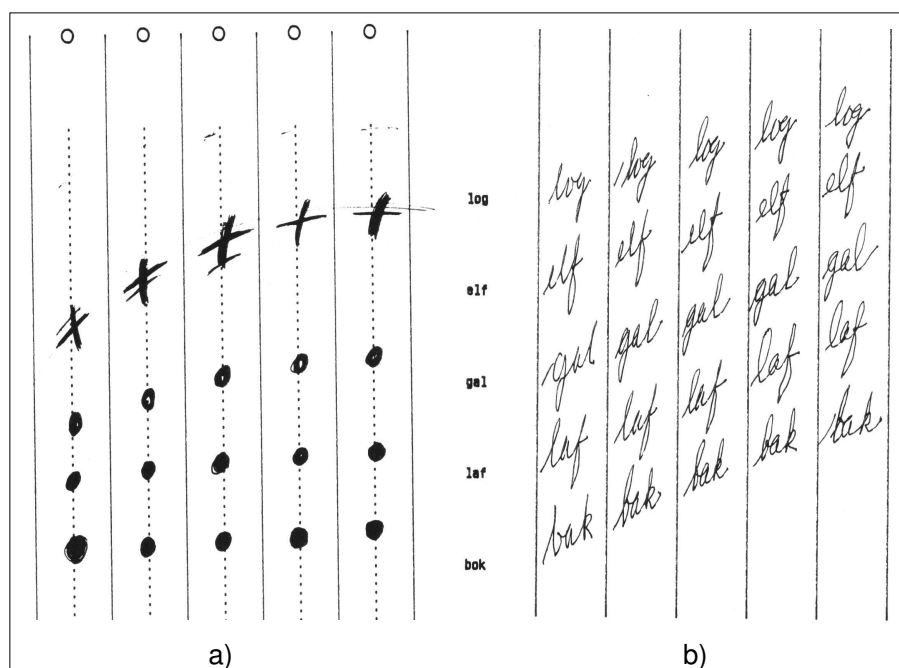


Figure 1.6: The experimental set-up from Fig. 1.5 was used to record on-line signals of simple movements and complete handwriting. a) Recorded movements: the top traces show hand rotating around the wrist (X' direction) and finger (Y' direction) movements reflecting the biomechanical geometry and predominant writing direction. b) Recorded handwriting: writing slant changes are considerably smaller than the orientation changes of the Y' system. Handwriting slant is held constant by an intricate interaction between the X' and Y' subsystems. Reprinted from (Maarse 1987), with kind permission from the author.

thumb and fingers (Y'), and the other consisting of the entire hand rotating around the wrist (X'), from radial abduction to ulnar abduction. Maarse shows (see Fig. 1.6) that the changes observed in writing directions are less than the changes in the orientation of the effector subsystems, with the conclusion that "this unexplained slant constancy may be caused by a setting of writing slant in a motor program at a higher level" (Maarse 1987). The writer therefore 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.

Additional evidence regarding the constancy of individual writing habits is provided in another study (Maarse and Thomassen 1983) observing that changes in the horizontal progression motion affect predominantly the up strokes, while the down strokes maintain their direction correlated with the perceived slant of the manuscript. Up strokes contain practically all the connecting strokes between letters, whereas down strokes appear relatively much more as parts of actual letters or graphemes. Maarse affirms that down strokes "might change less if there is a tendency to keep the grapheme features unchanged, either because the visual appearance of the product is preferably

held constant, or because the motor program for the characteristic of the graphemes remains more or less constant", as a result of writing education (Maarse 1987).

The writer produces a pen-tip trajectory on the writing surface in two dimensions (x,y), modulating the height of the pen tip above the surface by vertical movement (z). Displacement control is replaced by force control (F) at the moment of landing. The pen-tip trajectory in the air between two pen-down components contains valuable writer-specific information, but its shape is not known in the case of off-line scanned handwritten samples. Similarly, pen-force information is highly informative of a writer's identity, but is not directly known from off-line scans (Schomaker and Plamondon 1990).

An important theoretical basis for the usage of handwritten shapes for writer identification is the fact that handwriting is not a feedback process which is largely governed by peripheral factors in the environment. Due to neural and neuromechanical propagation delays, a handwriting process based upon a continuous feedback mechanism alone would evolve too slowly (Schomaker 1991). Hence, the brain is continuously planning series of ballistic movements ahead in time, i.e., in a feed-forward manner. A character is assumed to be produced by a "motor program" (Schmidt 1975), i.e., a configurable movement-pattern generator which requires a number of parameter values to be specified before being triggered to produce a pen-tip movement yielding the character shape (Schomaker et al. 1989, Plamondon and Maarse 1989, Plamondon and Guerfali 1998) by means of the ink deposits (Doermann and Rosenfeld 1992, Franke and Grube 1998, Franke 2005). The final resulting shape on paper represents a variation around the "master pattern" stored centrally, in the motor memory of the writer. Although the process described thus far is concerned with continuous variables such as displacement, velocity and force control, the linguistic basis of handwriting allows for postulating a discrete symbol from an alphabet to which a given character shape refers.

This thesis will show that very effective writer identification and verification is achievable by combining local directional features informative about habitual pen grip and slant with allograph shape features informative about the character forms engrained in the motor memory of the writer.

1.9 A survey of recent research in the field

In this section, we present a review of the recent papers published on the topic of automatic writer identification in order to provide a general literature background for our own research work contained in this thesis. A comprehensive review covering the period until 1989 is given in (Plamondon and Lorette 1989) and we provide a number of references in the bibliography section: (Arazi 1977, Kuckuck et al. 1979, Klement et al. 1980, Kuckuck 1980, Steinke 1981, Klement 1981, Dinstein and Shapira 1982, Naske

1982, Arazi 1983, Klement 1983, Maarse et al. 1988).

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. Throughout the survey, we will make clear the distinction between *text-dependent* versus *text-independent* approaches.

(Said et al. 2000, Said et al. 1998) 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% writer identification accuracy. A similar approach has also been used on machine-print documents for script (Tan 1998) and font (Zhu et al. 2001) identification.

(Zois and Anastassopoulos 2000) perform writer identification and verification using single words. Experiments are performed on a dataset 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 re-sampled, 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% are obtained both for English and Greek words.

(Srihari et al. 2002, Srihari et al. 2005) propose a large number of features divided into two categories. Macro-features operating at document / paragraph / word level: gray-level entropy and threshold, number of ink pixels, number of interior / exterior contours, number of 4-direction slope components, average height / slant, paragraph aspect ratio and indentation, word length and upper / lower zone ratio. Micro-features operating at word / character level: gradient, structural and concavity (GSC) attributes, used originally for handwritten digit recognition (Favata and Srikanth 1996). Text-dependent statistical evaluations are performed on a dataset containing 1000 writers who copied 3 times a fixed text of 156 words (the CEDAR letter). This is the largest dataset used up to the present in writer identification studies. Micro-features are better than macro-features in identification tests with a performance exceeding 80%. A multilayer perceptron or parametric distributions are used for writer verification with an accuracy of about 96%. Writer discrimination was also evaluated using individual characters (Zhang et al. 2003, Srihari et al. 2003) and words (Zhang and Srihari 2003, Tomai et al. 2004).

(Bensefia et al. 2005b, Bensefia et al. 2005a, Bensefia et al. 2002, Bensefia et al. 2003) 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 dataset. Experimental results are reported on three datasets containing 88 writers, 39 writers (historical documents) and 150 writers, with 2 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% are reported on the different test datasets.

(Marti et al. 2001) and (Hertel and Bunke 2003) 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 (Schlapbach et al. 2005). Using a k-nearest-neighbor classifier, identification rates exceeding 92% are obtained in tests on a subset of the IAM database (Marti and Bunke 2002) with 50 writers, 5 handwritten pages per writer. The IAM dataset will also be used in our experiments described in Chapter 5 of the thesis.

(Schlapbach and Bunke 2004) use HMM-based handwriting recognizers (Marti and Bunke 2001) 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% identification with 2.5% error in verification are reported on a subset of the IAM database containing 100 writers, 5 handwritten pages per writer.

In (Bulacu et al. 2003), 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 incorporating also location information by extracting separate PDFs for the upper and lower halves of text lines and then adjoining the feature vectors (Bulacu and Schomaker 2003). Our allograph-level approach (Schomaker and Bulacu 2004, Schomaker et al. 2004) 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 (Bulacu and Schomaker 2005a). An independent confirmation of our early experimental results is given in (van der Maaten and Postma 2005). In this thesis, we collect our published work in a coherent overall scene. We provide full details regarding our features, together with their extensive experimental evaluation. We also provide a comprehensive analysis of feature combi-

nations. On a large dataset containing 900 writers with 2 samples per writer, our best performing feature combinations yield writer identification rates of Top-1 85-87% and Top-10 96% with an error rate around 3% in verification.

An interactive approach involving character retracing and DTW matching is proposed in (van Erp et al. 2003). A layered architecture for forensic handwriting analysis systems is proposed in (Franke and Köppen 2001). The relevance of biometrics in the area of document analysis and recognition is discussed in (Fairhurst 2003).

From the studies reviewed in this section, two main conclusions can be drawn. Firstly, in the text-dependent approach, high performance is achievable even with very small amounts of available handwritten material (in 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. Secondly, 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 thesis 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 domain. In our experimental datasets there are only two samples per writer containing usually an amount of handwriting in the order of one paragraph of text.

1.10 Main assumptions underlying the methods proposed in the thesis

There are three fundamental assumptions at the basis of the research work reported in this thesis. We make them explicit here.

- *Natural writing attitude*: our proposed statistical features capture the general distinctive aspects of the scanned script as it visually appears in the dataset samples. These global features can only be linked to a certain person under the assumption that the collected handwriting is genuine and no attempt has been made by the writer to disguise or forge his / her natural writing.
- *Foreground / background separation*: the input to the feature extraction algorithms described in this thesis are images containing handwritten text blocks with (near-) black ink on white background. We assume that the handwritten trace was separated from the document background and from all other graphical material that

may be present in the scanned images. It was possible to perform this separation automatically on the documents contained in our experimental datasets. In more concrete situations, this may actually require some limited form of human involvement.

- *Sufficient amount of ink*: in order to derive stable and text-independent estimates for the probability distributions used as writer identification features, a sufficient amount of handwritten material, in the order of a few text lines, must be present in the samples. The majority of the samples used in our experiments contain more than three handwritten text lines.

Under these assumptions, the general computer vision task of identifying a person on the basis of scanned images of handwriting becomes a tractable pattern recognition problem (Duda et al. 2001, Jain et al. 2000). The present thesis describes novel and effective statistical methods to automatically solve this interesting biometric identification problem.

In our view, the three assumptions underlying our pattern recognition methods are reasonable and not too restrictive. To a first approximation, handwriting may be considered as a natural binary image. Therefore, the techniques proposed in this thesis have practical applicability. Nevertheless, further research may be directed at eliminating part (ideally all) of these assumptions.

1.11 Overview of the thesis

The writer identification and verification methods presented in this thesis operate at two levels of analysis: the *texture level* and the *character-shape (allograph) level*. The body of the thesis is therefore divided into two main parts treating these two important aspects.

Chapter 2 and Chapter 3 cover the *texture level* features. Chapter 4 and Chapter 5 present the *allograph level* features and the method used to combined multiple features for improving the final performance of our writer identification and verification system. The substance of this thesis resides in the design of new and effective statistical features. An important characteristic that distinguishes our approach is that the proposed features are text-independent: the handwriting is merely seen as a texture characterized by some directional probability distributions or as a simple stochastic shape-emission process characterized by a grapheme occurrence probability.

Chapter 2 introduces the idea of using the directionality of script as a fundamental source of information for writer identification. We show that using edge-angles, and especially edge-angle combinations, to build directional probability distributions is an

effective way to capture individual handwriting style with good performance in the task of writer identification.

Chapter 3 shows that further improvements in performance are obtained by capturing, besides orientation, also location information in the computation of our joint directional probability distributions. A comparison in terms of writer identification performance is carried out between lowercase and uppercase handwriting.

Chapter 4 presents our allograph-level method for automatic writer identification and verification. This theoretically founded method assumes that each writer is characterized by a stable probability of occurrence of some simple ink-trace shapes. We use the term *graphemes* for these sub- or supra-allographic ink fragments resulting from a handwriting segmentation procedure. Three clustering algorithms are compared on the task of generating the common shape codebook needed for estimating the writer-specific grapheme occurrence probability.

Chapter 5 considers the problem of fusing multiple features for improving the combined performance. Algorithmic refinements are described for the directional texture-level features and the experiments are extended to larger datasets. Our largest test set contains 900 writers and it is comparable in size to the largest dataset used in writer identification studies until the present.

Chapter 6 summarizes the research results presented in this thesis, draws the final general conclusions and sketches the future research directions opened by the work reported here. In the closing appendix, we use an HTML-based visualization tool to show some representative results generated by our software, named GRAWIS, an acronym from Groningen Automatic Writer Identification System.

Part I

Texture-Level Approach

A modified version of this chapter was published as: Marius Bulacu, Lambert Schomaker, Louis Vuurpijl, – “Writer identification using edge-based directional features,” Proc. of 7th Int. Conf. on Document Analysis and Recognition (ICDAR 2003), IEEE Computer Society, 2003, pp. 937-941, vol. II, 3-6 August, Edinburgh, Scotland

Chapter 2

Writer Identification Using Edge-Based Directional Features

Science is what we understand well enough to explain to a computer. Art is everything else we do.

Donald Knuth

Abstract

This chapter evaluates the performance of edge-based directional probability distributions as features in writer identification in comparison to a number of other texture-level features encoding non-angular information. We introduce here a new feature: the joint probability distribution of the angle combination of two “hinged” edge fragments. It is noted that the “edge-hinge” distribution outperforms all other individual features. Combining features yields improved performance. Limitations of the studied global features pertain to the amount of handwritten material needed in order to obtain reliable distribution estimates. A stability test is carried out showing the dependence of writer identification accuracy on the amount of handwritten material used for feature extraction.

2.1 Introduction

In the process of automatic handwriting recognition, invariant representations (features) are sought which are capable of eliminating variations between different handwritings in order to classify the shapes of characters and words robustly. The problem of writer identification, on the contrary, requires a specific enhancement of the variations which are characteristic to a writer’s hand. At the same time, such representations or features should, ideally, be independent of the amount and the semantic content of the written material. Slant represents a very stable characteristic of individual handwriting and gives the distinctive visual appearance of a handwritten text block under a general succinct view. The slant angle corresponds to the dominant direction in the handwritten script. In this chapter we will use the complete probability distribution of directions in the ink trace for writer identification. This distribution will be computed using edge fragments along the written ink. The edge-direction distribution will

constitute the building block for designing more complex features yielding increased performance.

Three groups of features can be identified in forensic writer identification:

- global measures, computed automatically on a region of interest (ROI)
- local measures, of layout and spacing features entered by human experts
- measures related to individual character shapes

We reiterate that, in this thesis, we analyze only features that are automatically extractable from the handwriting image without any human intervention. Furthermore, it is assumed that a crisp foreground/background separation has been realized in a pre-processing phase, yielding a white background with near-black ink. As a rule of thumb, in forensic writer identification one strives for 100% recall of the correct writer in a hit list of 100 writers, computed from a database of more than 10^4 samples. This amount is based on the pragmatic consideration that a number of one hundred suspects is just about manageable in criminal investigation. Current systems are not powerful enough to attain this goal.

As regards the theoretical foundation of our approach, the process of handwriting consists of a concatenation of ballistic strokes, which are bounded by points of high curvature in the pen-point trajectory. Curved shapes are realized by differential timing of the movements of the wrist and the finger subsystem (Schomaker et al. 1989). In the spatial domain, a natural coding, therefore, is expressed by angular information along the handwritten curve (Plamondon and Maarse 1989). It has long been known (Maarse and Thomassen 1983, Maarse et al. 1988, Crettez 1995) that the distribution of directions in handwritten traces, as a polar plot, yields useful information for writer identification or coarse writing-style classification. It is the goal of this chapter to explore the performance of angular-distribution directional features, relative to a number of other features which are in actual use in forensic writer-identification systems. The edge-based probability distributions operate at the scale of the ink-trace width, they give a texture-level view of the handwritten sample and they are informative, in general, about the habitual pen grip and biomechanical makeup of the writing hand.

2.2 Experimental data

We evaluated the effectiveness of different features in terms of writer identification using the *Firemaker* dataset (Schomaker and Vuurpijl 2000). A number of 250 Dutch subjects, predominantly students, were required to write four different A4 pages. On page 1 they were asked to copy a text presented in the form of machine print characters. On

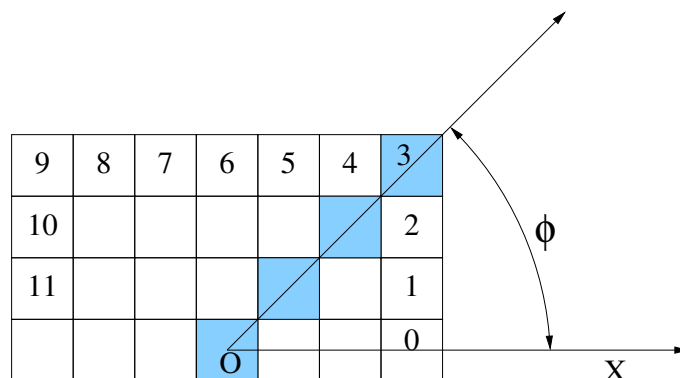


Figure 2.1: Extraction of edge-direction distribution.

page 4 they were asked to describe the content of a given cartoon in their own words. Pages 2 and 3 of this database contain upper case and forged-style samples and are not used here. Liation guidelines were used on the response sheets using a dropout color, i.e., one that fully reflects the light spectrum emitted by the scanner lamp such that it has the same sensed luminance as the white background. The added drawback is that the vertical line distance can no longer be used as a discriminatory writer characteristic. The recording conditions were standardized: the same kind of paper, pen and support were used for all the subjects. As a consequence, this also implies that the ink trace thickness variations will be more due to writer differences than due to recording conditions. The response sheets were scanned with an industrial quality scanner at 300 dpi, 8 bit / pixel, gray-scale. Our experiments are entirely image-based, no on-line information is available (e.g. speed of writing, order of different strokes).

2.3 Features

In this section we describe the extraction methods for five texture-level features used in writer identification. The first two features are edge-based directional distributions. We will focus our attention on the second one of them which is a new feature proposed and analyzed by us in recent publications.

2.3.1 Edge-direction distribution

Feature extraction starts with conventional edge detection (convolution with two orthogonal differential kernels, we used Sobel, followed by thresholding) that generates a binary image in which only the edge pixels are "on". We then consider each edge pixel in the middle of a square neighborhood and we check (using logical AND oper-

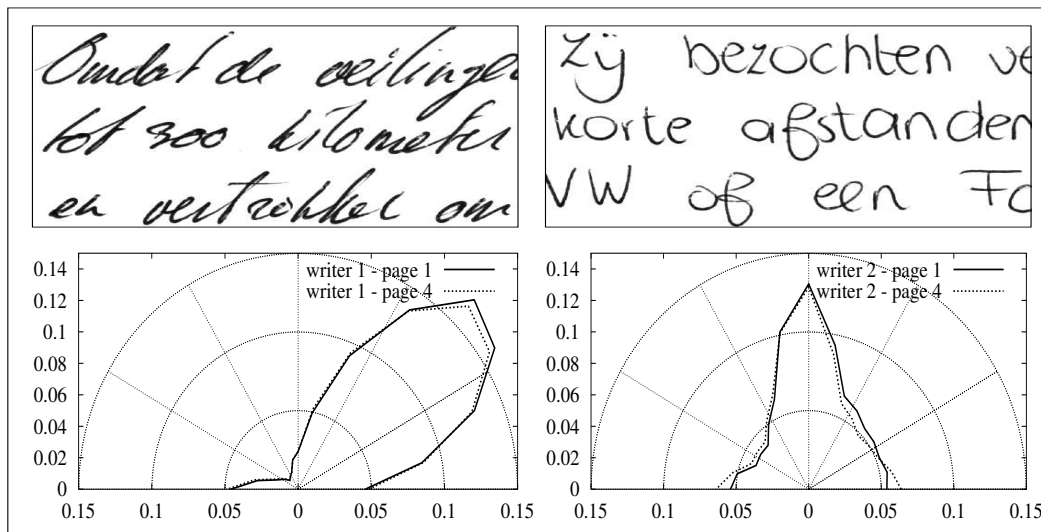


Figure 2.2: Two handwriting samples from two different subjects. We superposed the polar diagrams of the edge-direction distribution $p(\phi)$ corresponding to pages 1 and 4 contributed to our dataset by each of the two subjects. There is a large overlap between the directional distributions extracted from samples originating from the same writer, while there is a substantial variation in the directional distributions for different writers.

ator) in all directions emerging from the central pixel and ending on the periphery of the neighborhood for the presence of an entire edge fragment (see Fig. 2.1). All the verified instances are counted into a histogram that is finally normalized to a probability distribution $p(\phi)$ which gives the probability of finding in the image an edge fragment oriented at the angle ϕ measured from the horizontal.

In order to avoid redundancy, the algorithm only checks the upper two quadrants in the neighborhood because, without on-line information, we do not know which way the writer "traveled" along the found oriented edge fragment.

In the experiments, we considered 3, 4 and 5-pixel long edge fragments. Their orientation is quantized in $n = 8, 12$ and 16 directions respectively (Fig. 2.1 is an example for $n=12$). Clearly, n is also the number of bins in the histogram and the dimensionality of the final feature vector.

The distribution of the writing directions is characteristic of a writer's style. The polar probability density function was used in an on-line study of handwriting (Maarse and Thomassen 1983) to describe differences between upward and downward strokes. It was also used off-line (Crettez 1995) as a preliminary step to handwriting recognition that allows a partition of the writers by unsupervised fuzzy clustering in different groups.

While in the mentioned studies the directional histogram was computed on the written trace itself, for the present work we computed it based on the edges. Edges follow

the written trace on both sides and they are thinner, effectively reducing the influence of trace thickness.

We must mention an important practical detail: our generic edge detection does not generate 1-pixel wide edges, but they can usually be 1-3 pixels wide and this introduces smoothing into the histogram computation because the "probing" edge fragment can fit into the edge strip in a few directions around a central main direction. This smoothing taking place in the pixel space has been found advantageous in our experiments.

As can be noticed in Fig. 2.2, the predominant direction in $p(\phi)$ corresponds, as expected, to the slant of writing. Even if idealized, the example shown can provide an idea about the "within-writer" variability and "between-writer" variability in the feature space.

By analyzing the data, we found out that differentiation of the feature vector ($dp(\phi)$) results in a significant performance improvement. Besides removing the DC component, the differentiated directional probability distribution conveys information about the changes in writing direction. Along this line of thinking came the idea of a more complex feature capable of bringing forth more information about the local writer specificities by computing locally on the image the probability distribution of changes in direction.

2.3.2 A new feature: edge-hinge distribution

Our goal is to design a feature characterizing the changes in direction undertaken during writing with the hope that it will be more specific to the writer and consequently making possible more accurate identification. The method of feature extraction is similar to the one previously described, but it has added complexity. The central idea is to consider in the neighborhood, not one, but two edge fragments emerging from the central pixel and, subsequently, compute the joint probability distribution of the orientations of the two fragments.

To have a more intuitive idea of the feature that we are proposing, imagine having a hinge laid on the surface of the image. Place its junction on top of every edge pixel, then open the hinge and align its legs along the edges. Consider then 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 probability distribution $p(\phi_1, \phi_2)$ quantifying the chance of finding in the image two "hinged" edge fragments oriented at the angles ϕ_1 and ϕ_2 .

As already mentioned, in our case edges are usually wider than 1-pixel and therefore we have to impose an extra constraint: we require that the ends of the hinge legs should be separated by at least one "non-edge" pixel. This makes certain that the hinge is not positioned completely inside the same piece of the edge strip. This is an important

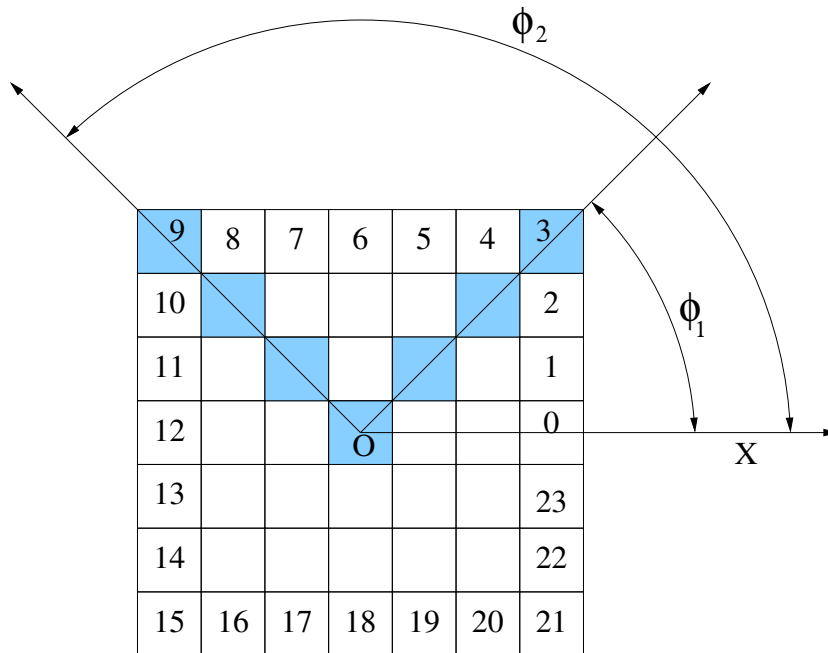


Figure 2.3: Extraction of edge-hinge distribution.

detail, as we want to make sure that our feature properly describes the shapes of edges (and implicitly the shapes of handwriting) and avoids the senseless cases.

If we consider an oriented edge fragment AB, the arrangement of the hinge is different whether a second oriented edge fragment attaches in A or in B. So we have to span all the four quadrants (360°) around the central junction pixel when assessing the angles of the two fragments. This contrasts with the previous feature for which spanning the upper two quadrants (180°) was sufficient because AB and BA were identical situations.

Analogously to the previous feature, we considered 3, 4 and 5-pixel long edge fragments. This time, however, their orientation is quantized in $2n = 16, 24$ and 32 directions respectively (Fig. 2.3 is an example for $2n = 24$). From the total number of combinations of two angles we will consider only the non-redundant ones ($\phi_2 > \phi_1$) and we will also eliminate the cases when the ending pixels have a common side. Therefore the final number of combinations is $C(2n, 2) - 2n = n(2n - 3)$ and, accordingly, our “hinge” feature vectors will have 104, 252 and 464 components.

Figure 2.4 shows a 3D plot of the bivariate edge-hinge distribution $p(\phi_1, \phi_2)$. Every writer has a different “probability landscape” and this provides the basis for very effective writer identification.

For the purpose of comparison, we evaluated also three other features widely used for writer identification:

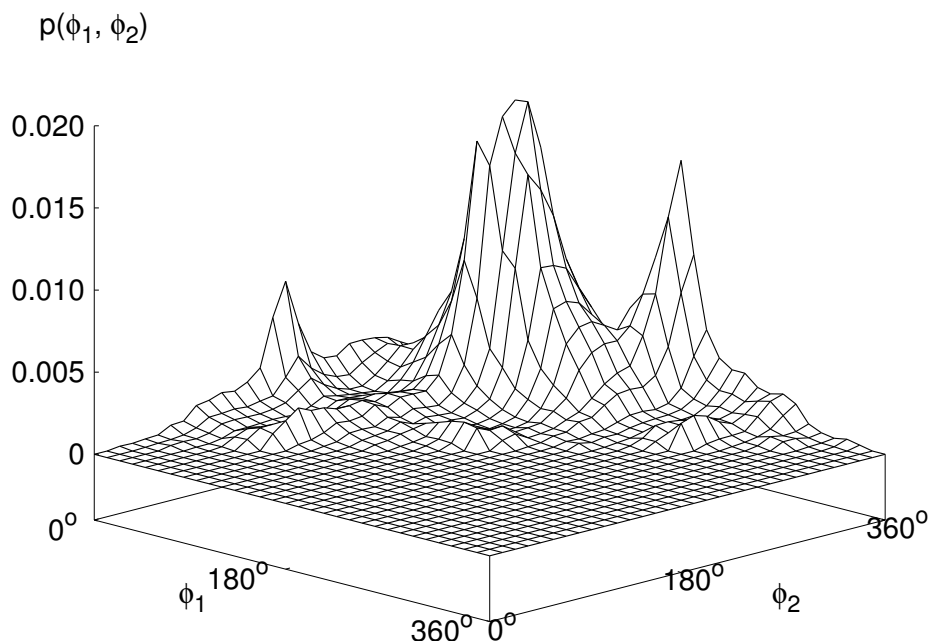


Figure 2.4: Graphical representation of the edge-hinge joint probability distribution. One half of the 3D plot (situated on one side of the main diagonal) is flat because we only consider the angle combinations with $\phi_2 > \phi_1$.

2.3.3 Run-length distributions

Run lengths, first proposed for writer identification by Arazi (Arazi 1977), are determined on the binarized image taking into consideration either the black pixels corresponding to the ink trace or, more beneficially, the white pixels corresponding to the background. Whereas the statistical properties of the black runs mainly pertain to the ink width and some limited trace shape characteristics, the properties of the white runs are indicative of character placement statistics. There are two basic scanning methods: horizontal along the rows of the image and vertical along the columns of the image. Similarly to the edge-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 100 pixels (the height of a written line in our dataset is about 120 pixels).

2.3.4 Autocorrelation

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 computed. The maximum offset ('delay') corresponds to 100 pixels. All autocorrelation functions are then accumulated for all rows and the sum is normalized to obtain a zero-lag correlation of 1. 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 local wavelength. This results in a large dot product contribution to the final histogram.

2.3.5 Entropy

The entropy measure used here focuses on the amount of information, normalized by the amount of ink (black pixels) in the regions of interest. This was realized by using the normalized file size of ROI files after Lempel-Zif compression. The size of the resulting file (in bytes) is divided by the total number of black pixels, which closely estimates the amount of ink present on the page. The obtained feature gives an estimate of the entropy of the ink distribution on the page.

2.4 Results

2.4.1 Evaluation method

The efficacy of the considered features has been evaluated using nearest-neighbor classification (Cover and Hart 1967) in a leave-one-out strategy. Explicitly, one page is chosen and extracted from the total of 500 pages (notice that the experimental data contains 2 pages written by each of 250 subjects). Then the Euclidean distances are computed between the feature vector of the chosen page and the feature vectors of all of the remaining 499 pages. These distances are ranked starting with the shortest one (Press et al. 1992). Ideally, the first ranked page should be the pair page produced by the same writer: an ideal feature extraction making classification effortless and a remapping of the feature space unnecessary. If one considers, not only the nearest neighbor (rank 1), but rather a longer list of neighbors starting with the first and up to a chosen rank (e.g. rank 10), the chance of finding the correct hit increases with the list size. The curve depicting the dependency of the probability of a correct hit vs. the considered list size gives an illustrative measure of performance. Better performance means higher probability of correct hit for shorter list sizes which is equivalent to a curve drawn as much as possible toward the upper-left corner.

Table 2.1: Writer identification accuracy (in percentages) on the Firemaker dataset (250 writers, 2 pages / writer, page 1 vs. page 4). The numbers in the second row of the table header denote the dimensionality of the feature vectors, i.e. the number of bins in the feature histograms. The rightmost column shows the performance obtained by concatenating the edge-hinge PDF and the horizontal run-length PDF.

List size	$p(\phi)$			$dp(\phi)$	$p(\phi_1, \phi_2)$			comb.
	8	12	16	15	104	252	464	
1	26	30	35	45	45	57	63	75
2	34	39	45	55	55	67	71	83
3	40	47	52	62	64	73	75	86
4	45	52	57	66	69	77	79	87
5	49	57	62	70	72	78	81	89
6	53	60	65	72	73	80	83	91
7	58	63	68	74	75	82	85	92
8	60	64	69	75	78	83	86	93
9	62	65	71	76	79	83	87	93
10	64	68	72	78	80	84	88	94
11	66	69	74	79	81	85	88	94
12	68	72	76	81	82	86	88	95
13	70	73	77	82	82	87	89	95
14	71	74	78	83	83	87	89	95
15	72	76	79	84	83	88	89	95
16	74	77	80	84	84	88	90	96
17	76	79	82	84	85	89	90	96
18	77	80	82	85	85	89	90	96
19	78	81	82	86	86	90	91	97
20	79	81	83	87	86	90	91	97

We point out that we do not make a separation between a training set and a test set, all the data is in one suite. This is actually a more difficult and realistic testing condition, with more distractors: not 1, but 2 per false writer and only one correct hit. Error rates are approximately halved when using the traditional train/test set distinction. Note also the added fact that we only have 2 samples per writer (more labeled samples increasing the chance of a correct identification of the author - see reference (Said et al. 1998) for results on 10 writers, 15 documents / writer). As a consequence of these circumstances, our results are more conservative.

It is also important to mention that the text (ASCII) content is different in the two

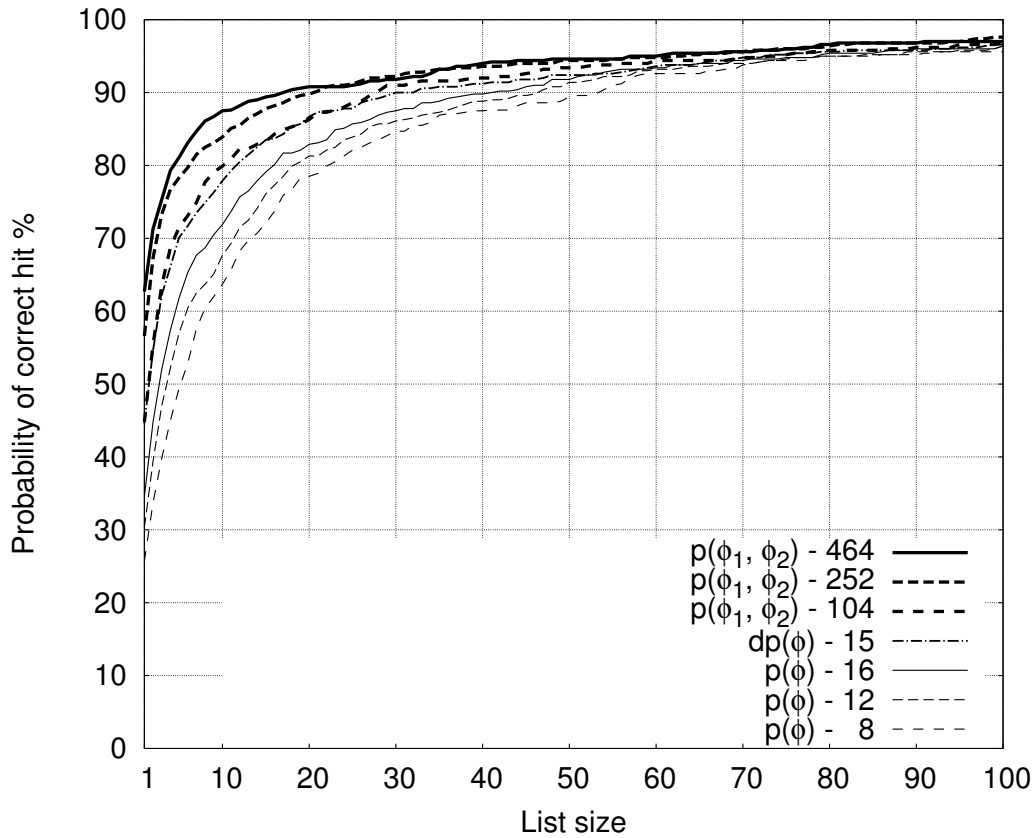


Figure 2.5: Performance curves of the edge-based features $p(\phi)$ and $p(\phi_1, \phi_2)$ for different direction quantizations (features are ordered with most effective at the top).

samples originating from the same writer: page 1 contains copied text, while page 4 contains self generated text describing a cartoon. The proposed features give a content-independent description of the texture of handwriting.

2.4.2 Analysis of performances

We present the performance curves of the edge-based directional features in Fig. 2.5 and the numerical values in Table 2.1. The dimensionality of every feature is mentioned in the figure and in the heading of the table.

Confirming our initial expectations, the improvement in performance yielded by the new feature is very significant despite the excessive dimensionality of the feature vectors (verified by PCA analysis). As a second-order feature, the hinge angular probability distribution captures larger range correlations from the pixel space and therefore it characterizes more intimately the handwriting style providing for more accurate writer identification.

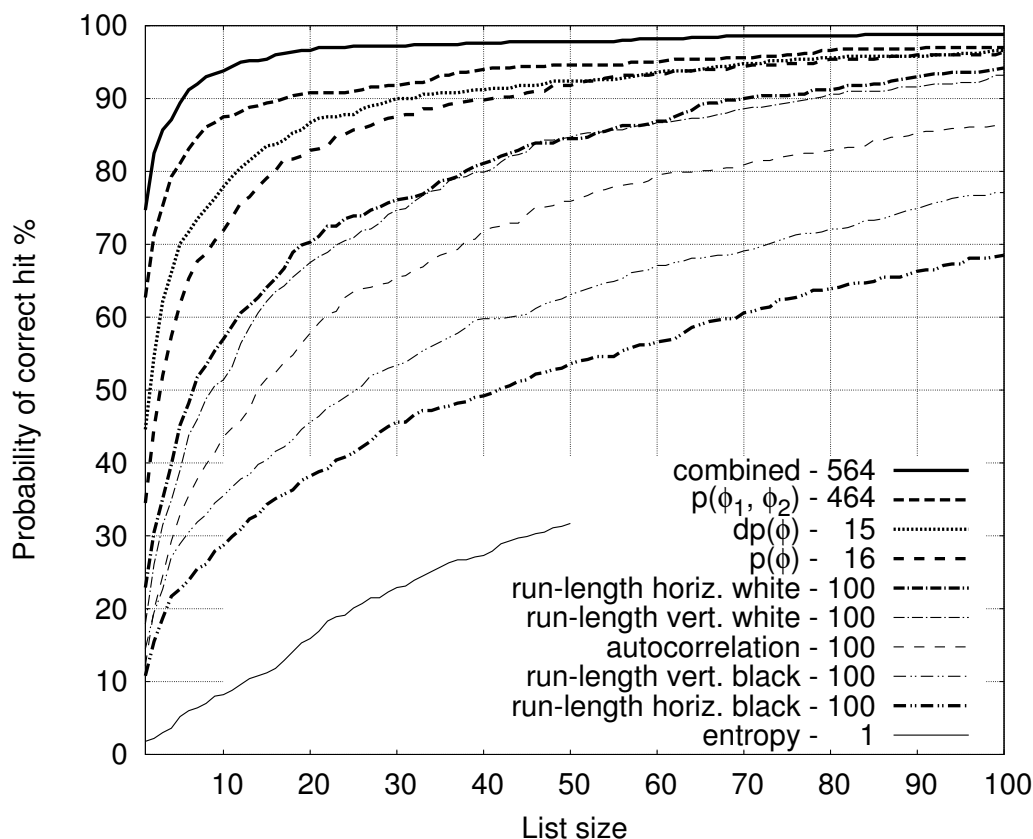


Figure 2.6: Performance curves for the evaluated features (features are ordered with the most effective at the top).

Examination of the family of curves in Fig. 2.5 attests that finer quantized directions result in improved performance at the expense of an increase in feature vector dimensionality (much more sizeable for the edge-hinge feature $p(\phi_1, \phi_2)$).

Figure 2.6 gives a general overview of the comparative performance for all the features considered in this chapter.

The edge-based directional features perform significantly better than the other features because they give a more detailed and intimate information about the peculiarities of the shapes that the writer produces (slant and regularity of writing, roundness or pointedness of letters).

An interesting observation is that the vertical run lengths on ink are more informative than the horizontal ones. This correlates with an established fact from on-line handwriting recognition research stating that the vertical component of strokes carries more information than the horizontal one (Maarse and Thomassen 1983).

The presented features are not totally orthogonal, but nevertheless they do offer different points of view on our dataset. It is therefore natural to try to combine them for

Table 2.2: Feature performance degradation with decreasing amounts of written text (writer identification accuracy in percentages for list size = 10). The PDFs are extracted from samples containing diminishing amounts of handwritten ink: whole page (w), half page (top (t) and bottom (b)), and the first line (l).

Feature	w	t	b	l
$p(\phi_1, \phi_2)$	88	81	84	53
$p(\phi)$	72	66	69	36
run-length horiz. white	57	42	42	18
run-length vert. white	51	39	42	16
run-length vert. black	36	33	33	13
entropy	8	4	6	5

improving the accuracy of writer identification. This topic will be analyzed more thoroughly later in this thesis. We will present here though, for exemplification, the results obtained by concatenating the edge-hinge distribution with the horizontal run lengths on white into a single feature vector that was afterward used for nearest-neighbor classification (last column in Table 2.1).

2.4.3 Stability test

An important question arises: what is the degradation in performance with decreasing amounts of handwritten material? We provide three reference points: whole page (w), half page (top (t) and bottom (b)), and the first line (l). The answer to this question has major bearing for forensic applications (where, in many cases, the available amount of handwritten material is sparse, e.g. the filled-in text on a bank invoice or the address on a perilous letter).

We consider writer identification accuracies for hit lists up to rank 10 (deemed as a more reliable anchor point). Our results from Table 2.2 show significant degradation of performance when very little handwritten material is available. However, it is interesting to observe that the performance standings of the different features with respect to each other remain the same, independent of the amount of text.

2.5 Conclusions

In this chapter, a number of texture-level features have been described and evaluated on the task of text-independent writer identification. The edge-based directional fea-

tures give an overall better performance than run-length, autocorrelation and entropy features.

We described here a new edge-based feature for writer identification that characterizes the changes in direction undertaken during writing. The edge-hinge feature performs markedly better than all the other evaluated features.

Our stability test show that the best performing features when a large amount of text is available still perform best compared to the others when little text is available, despite having considerably higher dimension.

The next chapter of this thesis will focus on increasing the discriminatory power of the feature vectors by including also location information. We will also study how further improvements in performance can be obtained by combining different features in order to exploit their intrinsic degree of orthogonality.

A modified version of this chapter was published as: Marius Bulacu, Lambert Schomaker, – “*Writer style from oriented edge fragments*,” Proc. of 10th Int. Conf. on Computer Analysis of Images and Patterns (CAIP 2003): LNCS 2756, Springer, 2003, pp. 460-469, 25-27 August, Groningen, The Netherlands

Chapter 3

Characterizing Handwriting Individuality Using Localized Oriented Edge Fragments

*Computer Science is no more about computers than astronomy
is about telescopes.*

Edsger Dijkstra

Abstract

In this chapter, we compare the performances of a number of texture-level features on lowercase and uppercase handwriting. We propose a new directional distribution feature that considers the edge angle combinations co-occurring at the extremities of run-lengths. In an effort to gain location-specific information, new versions of the features are computed separately on the top and bottom halves of text lines and then fused. The new features deliver significant improvements in performance. We report also on the results obtained by combining features using a voting scheme.

3.1 Introduction

This chapter continues our analysis of texture-level features automatically extracted from handwriting images for the purpose of writer identification. Image-based (off-line) writer identification has its principal application mainly confined to the areas of forensic and historic document analysis. It is in the same class with other behavioral biometrics (on-line signature dynamics, voice) which, in contrast, enjoy much wider applicability together with the more powerful, but also more intrusive, physiological biometrics (face, hand geometry, fingerprint, iris pattern, retinal blood vessels).

An essential requirement for the forensic application area is that the writer identification system should have, not only verification capability (authentication in a one-to-one comparison), but also the more demanding identification capability (one-to-many search in a large database with handwriting samples of known authorship and return of a likely list of candidates).

Writer identification is rooted in the older and broader automatic handwriting recognition domain. For automatic handwriting recognition, the input patterns are normalized in a number of ways (e.g. by deskewing, deslanting, size normalization) before they are passed to the statistical recognizers. The role of the statistical recognizers is to further eliminate variations between different handwritings in order to classify the shapes of characters and words robustly. Much of the information that is thrown away in this process is, in fact, essential in writer identification. Handwriting recognition and writer identification represent therefore two opposing facets of handwriting analysis. Handwriting recognition research constitutes fertile ground for inspiration in the quest of improving writer identification.

The complete process of forensic writer identification is never fully automatic, due to a wide range of scan-quality, scale and foreground/background separation problems. However, features that are automatically extractable from selected text blocks or ROIs from the handwritten samples can find useful applicability in large-scale searches over large databases for selecting a reduced list of likely candidates. At the same time, such features should, ideally, be independent of the amount and semantic content of the written material. In the extreme case, a single word or the signature should suffice to identify the writer from his individual handwriting style. The automatic methods (or features) analyzed in this thesis are intended as an additional support to the manual measurements and manual feature extraction of traditional forensic handwriting expertise.

In this chapter we summarize the extraction methods for five features: three edge-based directional features, one run-length feature and one ink-distribution feature. In order to gain location-specific information, new versions of the features are computed separately on the top and bottom halves of text lines and then fused. We make a cross comparison of the performance of all features when computed on lowercase and uppercase handwritten text. We report also on results obtained using a voting scheme to combine the different features into a single final ranked hit list.

3.2 Experimental data

We conducted our study using another part of the *Firemaker* dataset (Schomaker and Vuurpijl 2000) than the one used in the previous chapter. We used here page 1 and page 2 from our dataset that contains a total of 250 enrolled subjects. On page 1 they were asked to copy a text of 5 paragraphs using normal handwriting style (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 2 paragraphs using only uppercase letters. Pages 3 and 4 contain forged- and normal-style handwriting and are not used here. For

Table 3.1: Features used for writer identification and the used distance function $\Delta(\vec{u}, \vec{v})$ between a query sample \vec{u} and a database sample \vec{v} . All features are computed in two scenarios "entire-lines" and "split-lines" (see text for details).

Feature	Explanation	Dimensions		$\Delta(\vec{u}, \vec{v})$
		entire	split	
$p(\phi)$	Edge-direction PDF	16	32	χ^2
$p(\phi_1, \phi_2)$	Edge-hinge PDF	464	928	χ^2
$p(rl)$	Horiz. run-length on background PDF	100	200	EUCLID
$p(\phi_1, \phi_3)$	Horiz. edge-angle co-occurrence PDF	256	512	χ^2
$p(brush)$	Ink-density PDF	225	450	χ^2

practical reasons, lineation guidelines were used on the response sheets using a special color "invisible" to the scanner. This gives us two important advantages that we will effectively use in the experiments performed in this chapter: automatic line segmentation can be performed reliably and handwriting is never severely skewed. In addition, the subjects were asked to leave an extra blank line between paragraphs making possible automatic paragraph extraction.

Being recorded in optimal conditions, the *Firemaker* dataset contains very clean data. This is obviously an idealized situation compared to the conditions in practice. However, the dataset serves well our purpose of evaluating the usefulness for writer identification of different features encoding the ink-trace shape.

3.3 Feature extraction

All the features used in the present analysis are probability density functions (PDFs) extracted empirically from the handwriting image. Our previous experiments confirmed that the use of PDFs is a sensitive and effective way of representing a writer's uniqueness (Bulacu et al. 2003). Another important advantage of using PDFs is that they allow for homogeneous feature vectors for which excellent distance measures exist. Experiments have been performed with different distance measures: Hamming, Euclidean, Minkowski up to 5th order, Hausdorff, χ^2 and Bhattacharyya. Table 3.1 shows the features and the corresponding best-performing distance measures used in nearest-neighbor matching. Figure 3.1 gives a schematic description of the extraction methods for the directional features used in this study.

In the present study, all the features will be computed in two scenarios: either on the entire text lines or separately on the top-halves and the bottom halves of all the

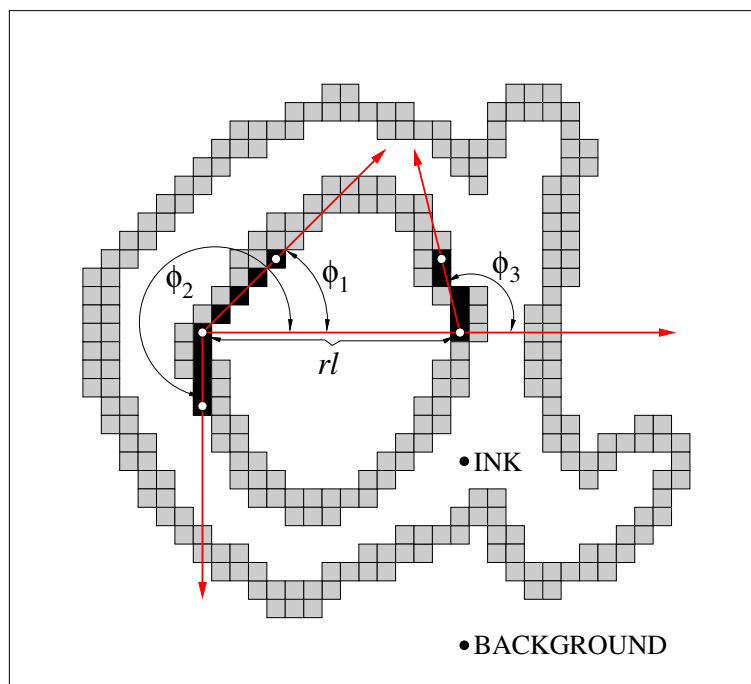


Figure 3.1: Schematic description of the extraction methods for the directional and run-length features. The letter “a”, provided as an example, would be roughly twice as large in actual reality.

text lines. In the first scenario, features are computed on the image without any special provisions. For the second scenario, all text lines are first segmented using the minima of the smoothed horizontal projection. Afterwards, the maxima are used to split horizontally every individual text line into two halves (see Fig. 3.2). All features are then computed separately for the top-halves and the bottom-halves and the resulting two vectors are concatenated into a single final feature vector. Clearly, the “split-line” features have double dimensionality compared to their “entire-line” counterparts.

While feature histograms are accumulated over the whole image providing for a very robust probability distribution estimation, they suffer the drawback that all position information is lost. Line splitting is therefore performed in an effort to localize more our features and gain back some position information together also with some writer specificity. What we must pay is the sizeable increase in feature dimensionality.

We will consider five features in this study and we describe further their extraction methods. We introduce a new edge-based directional feature similar in design to the edge-hinge feature presented in the previous chapter. We also consider an extra feature informative about the pen pressure and inking patterns found in the handwritten trace.

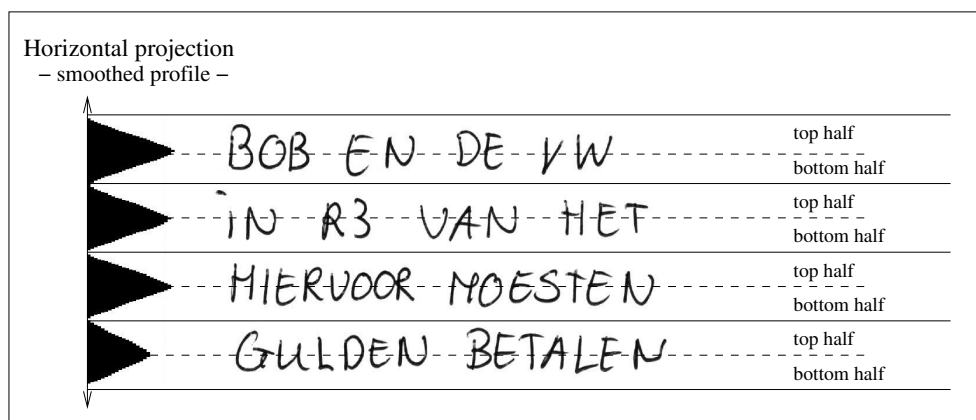


Figure 3.2: Line segmentation and splitting.

3.3.1 Edge-direction distribution

The distribution of the writing directions is characteristic for the identity of the writer, as shown in the previous chapter. This feature was first developed and used in on-line handwriting research (Maarse and Thomassen 1983, Crettez 1995). It was also used for signature verification in combination with k-nearest-neighbors, threshold and neural network classifiers (Sabourin and Drouhard 1992, Drouhard et al. 1995).

We use an off-line and edge-based version of the directional distribution. Computation of this feature (see Fig. 3.1) involves the following processing steps: convolution with two orthogonal differential kernels (Sobel), thresholding, direction estimation of edge-fragments using local edge-pixel neighborhoods, histogram accumulation and normalization to a probability distribution. Please refer to the previous chapter for a more detailed description of the method. In the edge-direction probability distribution $p(\phi)$, the angle is quantized in n bins spanning the first and second quadrants. A number of $n = 16$ directions performed best and will be used here.

The mode of the directional distribution $p(\phi)$ corresponds to the slant of writing (see Fig. 3.3). It is interesting to note that there is an asymmetry between the directional diagrams for the top halves and the bottom halves of the text lines. This observation is precisely the underpinning of our approach to split the lines in an attempt to recover this writer specific positional information. There is a correlation also with the known fact from on-line handwriting research that upward strokes are slightly more slanted than the downward strokes because they contain also the horizontal progression motion (Maarse and Thomassen 1983). The example shown was selected to make visually very clear the difference between "within-writer" variability and "between-writer" variability in the feature space.

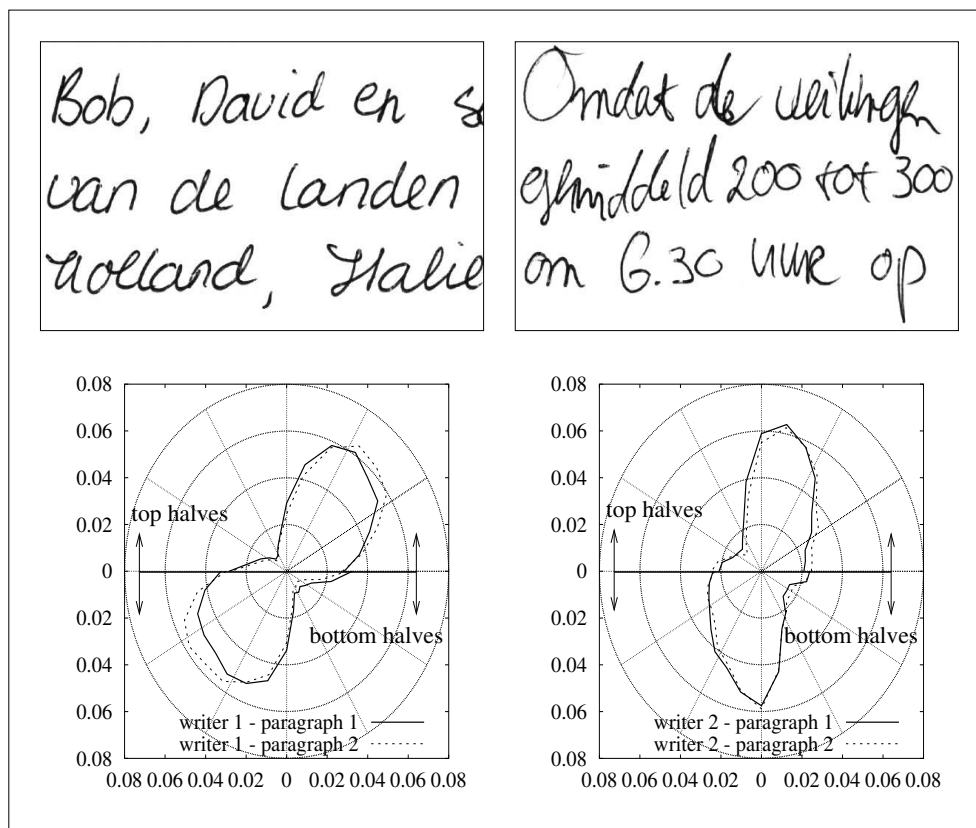


Figure 3.3: Examples of lowercase handwriting from two different subjects. We superposed the polar diagrams of the “split-line” direction distribution $p(\phi)$ extracted from the two lowercase handwriting samples for each of the two subjects. The “between-writer” variation in $p(\phi)$ is larger than the “within-writer” variation. There exists an asymmetry between the directional diagrams for the top halves and the bottom halves of the text lines and this provides extra information for writer identification.

3.3.2 Edge-hinge distribution

In order to capture the curvature of the ink trace, which is very discriminatory between different writers, we designed the edge-hinge feature using local angles along the edges. By counting all the angle combinations of two hinged edge fragments encountered in the image (see Fig. 3.1), the joint probability distribution $p(\phi_1, \phi_2)$ is built and used as a writer characteristic. Orientation is quantized in $2n$ directions for every leg of the “edge-hinge” spanning all four quadrants and the dimensionality of the feature vector is $C_{2n}^2 - n = n(2n - 3)$. For $n = 16$, the edge-hinge feature vector will have 464 dimensions. Please refer to the previous chapter for a more detailed description of the method.

3.3.3 Run-length distributions

Run-lengths have long been used for writer identification. They are determined on the binarized image by scanning either horizontally (along the image rows) or vertically (along the image columns) and taking into consideration either the black pixels (the ink) or, more beneficially, the white pixels (the background). The run-lengths on white are obviously more informative about the characteristics of handwriting as they capture the regions enclosed inside the letters and also the empty spaces between letters and words. Vertical run-lengths on black are more informative than the horizontal run-lengths on black, as shown in the previous chapter. We compute run-lengths of up to 100 pixels, comparable to the height of a written line. This feature is not size invariant, however, size normalization could be performed by hand prior to feature extraction. We will consider here only the horizontal run-lengths on white to be able to directly compute this feature both in the "entire-line" and "split-line" scenarios.

3.3.4 A new feature: horizontal co-occurrence of edge angles

This new feature that we developed and proposed in recent publications derives naturally from the previous two. It is a variant of the edge-hinge feature, in that the combination of edge-angles is computed at the ends of run-lengths on white (see Fig. 3.1). The joint probability distribution $p(\phi_1, \phi_3)$ of the two edge-angles occurring at both ends of a run-length on white captures longer range correlations between edge-angles and gives a measure of the roundness of the written characters. This feature has n^2 dimensions, namely 256 in our implementation.

The edge-angle co-occurrence feature capitalizes on the same idea used for designing the edge-hinge feature: building a joint PDF using combinations of local oriented edge fragments to characterize, at a texture level, the writing predilections peculiar to the author of a given handwritten sample.

The edge-based features that we propose here for writer identification are general texture descriptors and, as such, they have wider applicability (e.g. we used $p(\phi)$ for the pose-estimation of camera-captured machine-print text (Bulacu and Schomaker 2005c)). A more detailed and wider assessment of the applicability of our directional features beyond the realm of writer identification is, indeed, desirable. However, such an analysis would require a breadth that cannot be encompassed in the framework of the present thesis.

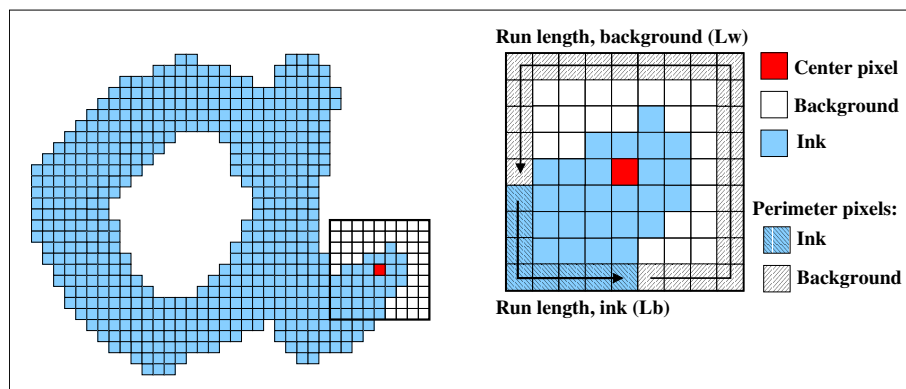


Figure 3.4: Extraction of the brush feature (as originally proposed in (Schomaker et al. 2003)) on the tail stroke of a lowercase letter "a". The size of the analyzing window is actually 15×15 pixels.

3.3.5 Brush function: ink density distribution

It is known that axial pen force ('pressure') is a highly informative signal in on-line writer identification (Schomaker and Plamondon 1990). Force variations will be reflected in saturation and width of the ink trace. Additionally, in ink traces of ballpoint pens, there exist lift-off and landing shapes in the form of blobs or tapering (Doermann and Rosenfeld 1992), which are due to ink-depositing processes. In order to capture the statistics of this process, we use another PDF feature originally proposed in (Schomaker et al. 2003). A convolution window of 15×15 pixels is used, only accumulating the local image if the current region obeys the following constraints: a supraliminal ink intensity in the center of the window, co-occurring with a long run of white pixels along minimally 50% of window perimeter and an ink run of at least 5% of window perimeter (see Fig. 3.4). After scanning all the image, the accumulator window is normalized, yielding a PDF describing ink distribution (see Fig. 3.5). This feature is clearly not size invariant (the window of 15^2 pixels was chosen for capturing the 5-7 pixel-wide ink traces usual in our images), but we use it because the recording conditions have been standardized for all subjects in our dataset.

3.4 Results

We compare the performance of our new "split-line" versions of the features with their former "entire-line" versions. We are also interested to compare the performance of all the features when computed on lowercase as opposed to uppercase handwriting. In order to perform all these comparisons, handwriting samples have been extracted from the database. Two paragraphs have been extracted from page 1 obtaining in this

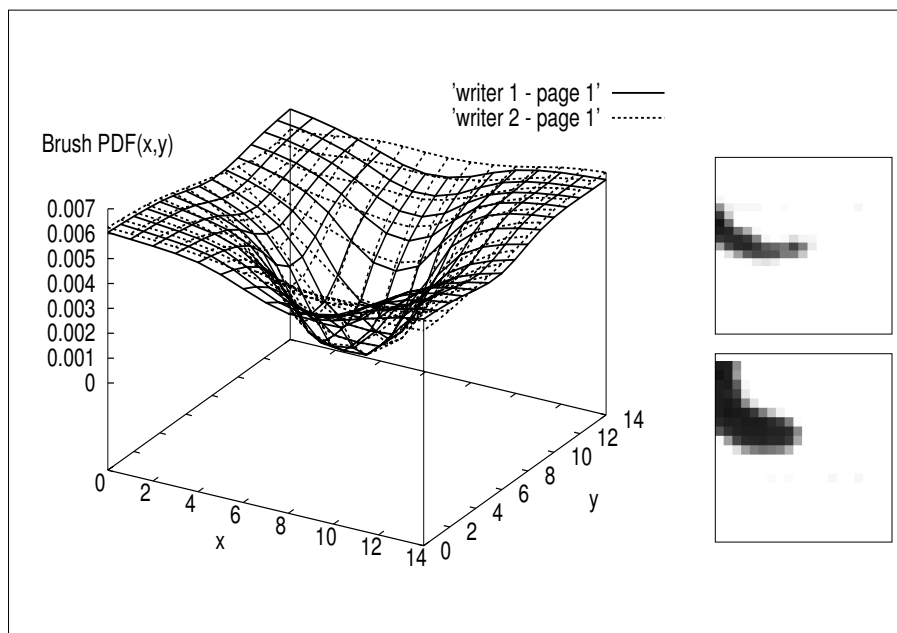


Figure 3.5: Superimposed brush PDFs for two writers and examples of an “a” tail for two writers (see reference (Schomaker et al. 2003)).

way two separate samples in lowercase for every subject. Similarly, from page 2 we extracted separately the two paragraphs in uppercase handwriting. Special care has been taken to have roughly the same amount of text in lowercase and uppercase (approx. 100 characters in the first paragraphs and approx. 150 characters in the second ones). An important observation is that the text content is different in the two samples that are supposed to be matched

Using nearest-neighbor matching (Cover and Hart 1967) in a leave-one-out strategy, the writer identification performance has been evaluated for lowercase and uppercase handwriting using both the “entire-line” and the “split-line” versions of our PDF features. The numerical results for the four possible combinations are given in Table 3.2. These results are also graphically represented in figures 3.6, 3.7, 3.8 and 3.9 to allow for a visual evaluation.

3.4.1 Comparison lower- vs. upper-case and entire- vs. split-line

Visual cross-comparisons of the performance curves given in figures 3.6, 3.7, 3.8 and 3.9 show that there are important differences in writer identification accuracy for the different features considered in this study. The edge-hinge feature surpasses all the other features and, quite remarkably, it performs better on uppercase than on lowercase, opposite to the situation for all the other features. This may result from the fact that the “hinge” can capture the sharp angularities present in uppercase letters. Another impor-

tant observation is that the differences in feature performance between lowercase and uppercase are not as large as one might intuitively expect, thinking that it is always easier to identify the author of lowercase rather than uppercase handwriting. Our results therefore contradict the generally assumed idea that uppercase characters contain less writer-specific information than does connected-cursive handwritten script. This assumption is corroborated by the observation that the automatic classification of uppercase isolated characters is easier than the recognition of cursive script. However, much of the difference in the recognition performance between uppercase characters versus free-style words can be attributed squarely to the character segmentation problem.

In mixed searches (e.g. lowercase query sample / uppercase dataset) writer identification performance is very low. The features used encode the shape of handwriting and, naturally, they are sensitive to major style variations.

The split-line features perform significantly better than their entire-line counterparts, fully justifying the extra cost in terms of dimensionality and computation. The exception is the brush feature on uppercase and this is due to the fact that there are not sufficient image sampling points on the bottom half of uppercase that comply with the imposed constraints and the PDF estimate is not sufficiently reliable. We emphasize that regaining location specific information, especially for the edge-based orientation PDF features, is a substantiated way of improving writer identification accuracy.

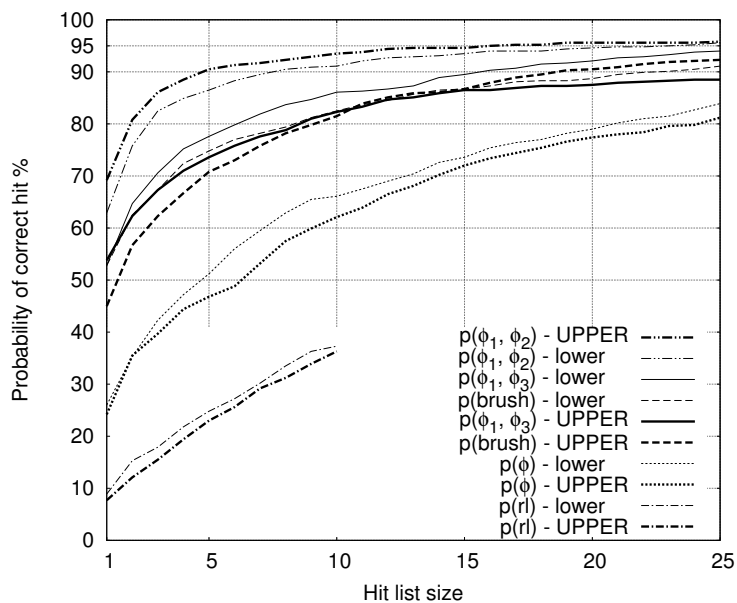


Figure 3.6: *Writer identification performance comparison: lower- vs. UPPER-case using entire-line features.*

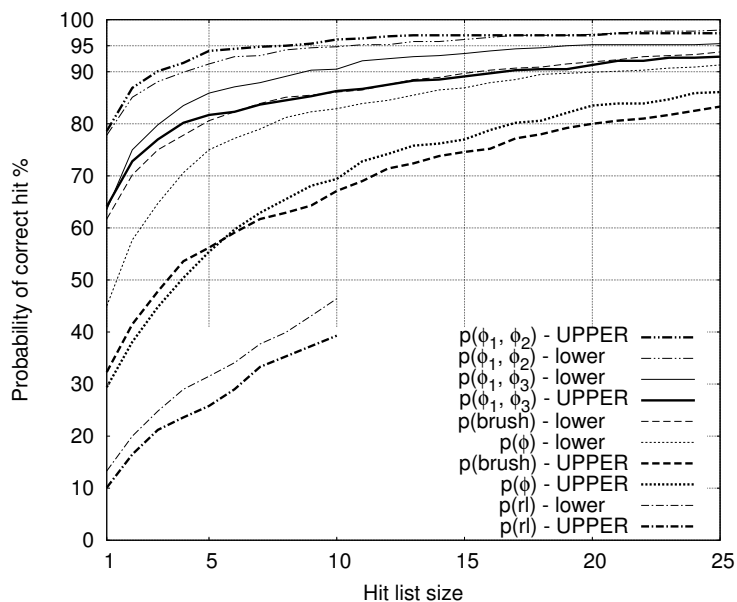


Figure 3.7: *Writer identification performance comparison: lower- vs. UPPER-case using split-line features.*

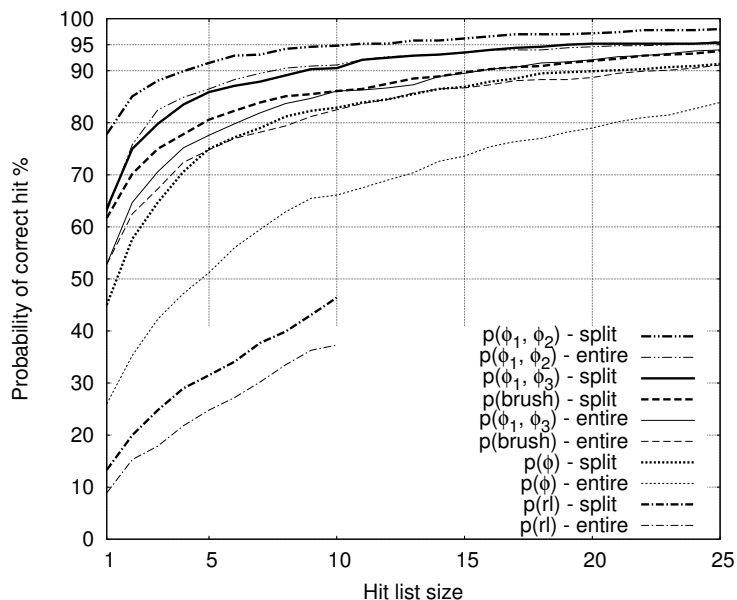


Figure 3.8: Writer identification performance comparison: split- vs. entire-line features on lower-case text.

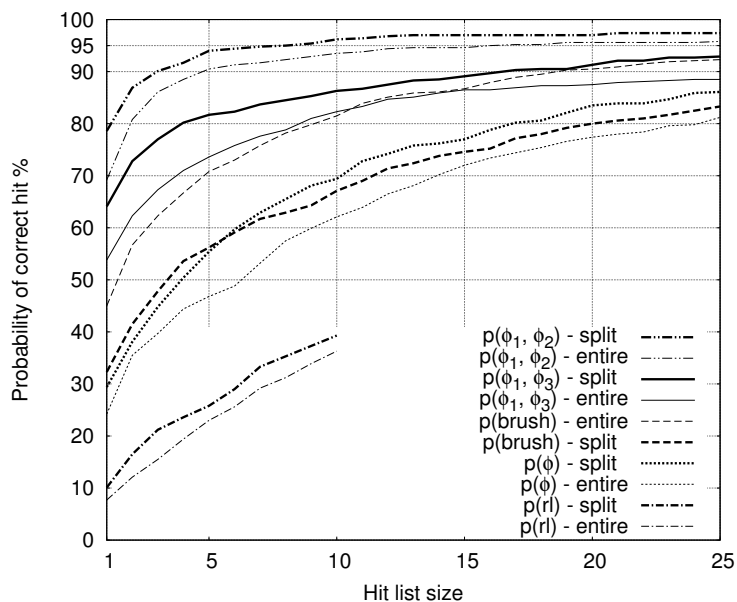


Figure 3.9: Writer identification performance comparison: split- vs. entire-line features on UPPER-case text.

Table 3.2: Writer identification accuracy (in percentages) on the Firemaker dataset (250 writers). One selected sample is matched against the remaining 499 samples that contain only one target sample (the pair) and 498 distractors. In the cells, performance figures for lowercase are in the upper-left corner and for uppercase in the lower-right (with boldface characters). 95% confidence limits: $\pm 4\%$.

Hit list size	$p(\phi)$		$p(\phi_1, \phi_2)$		$p(\tau l)$		$p(\phi_1, \phi_3)$		$p(\text{brush})$	
	entire	split	entire	split	entire	split	entire	split	entire	split
1	26	45	63	78	9	13	53	64	53	62
	24	29	69	79	8	10	54	64	45	32
2	35	58	76	85	15	20	65	75	63	70
	36	38	81	87	12	17	62	73	57	42
3	42	65	83	88	18	25	71	80	67	75
	40	45	86	90	16	21	67	77	62	48
4	47	71	85	90	22	29	75	84	72	78
	44	50	89	92	19	24	71	80	67	54
5	51	75	87	92	25	32	78	86	75	81
	47	55	91	94	23	26	74	82	71	56
10	66	83	91	95	37	46	86	91	83	86
	62	69	94	96	36	39	82	86	82	67

Table 3.3: Writer identification accuracy (in percentages) after feature combination using the Borda "min" voting method. Please refer to Table 3.2 for more details.

Hit list size	1	2	3	4	5	10
entire	67	77	83	86	87	91
	72	82	87	89	91	94
split	80	86	89	90	92	95
	79	87	91	92	94	96

3.4.2 Voting feature combination

It is important to note that no single feature will be powerful enough for the performance target defined by the forensic application, necessitating the use of classifier-combination schemes. In the present study we explored the Borda count method that considers every feature as a voter and then computes an average rank for each candidate over all voters. Different ranked voting schemes have been tested: min, plurality, majority, median, average, max (e.g. using the median instead of the average). The only voting method that brought some improvement in performance over the top-performing edge-hinge feature was the "min" method (results in Table 3.3). In this method, the decision of the voter (feature) giving the lowest rank is considered as the final decision.

In the current context, because the individual features have widely different performance, all the other voting schemes lead to some average performance higher than that of the weakest feature, but certainly lower than that of the strongest feature. An additional drawback is that the considered features are not totally orthogonal. Results reported elsewhere (Schomaker et al. 2003) confirm that another effective method of combining heterogeneous features is to consider a sequential scheme in which the stronger features vote at later stages against the accumulated votes from the weaker features.

The improvement in performance obtained with Borda "min" voting method is marginal: 0-4% for top 1 and vanishing for longer list sizes. It is however worthwhile mentioning that eliminating some of the weaker features from voting results nevertheless in slight performance drops.

3.5 Conclusions

We must emphasize that the method for writer identification presented here is automatic and sparse-parametric (no learning takes place) and this approach possesses major advantages in forensic applications given the appreciable size and time-variant con-

tent of the sample databases. Automatic computation-intensive approaches in this application domain will allow for convenient search in large sample 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.

Localized angular joint probability distributions are very effective features in capturing handwriting individuality. The χ^2 distance measure is mostly a natural choice for our feature vectors which are PDFs (Press et al. 1992). In our experiments, writer identification accuracies are comparable on lowercase and uppercase handwriting. Incorporating location information in our directional features yields improved results.

It is quite evident that texture-level features extracted from the handwriting image will never suffice in attaining the performance requirements in the forensic writer identification domain. Detailed character shape knowledge is needed as well. In this respect, it is important to note also the recent advances (Srihari et al. 2002) that have been made at the detailed allographic level, when character segmentation is performed by hand. In contrast, we will adopt an automatic character-level approach in the second part of this thesis. We will also maintain the requirement for a text-independent solution. Further, we will be interested in combining the texture-level and the allograph-level approaches for obtaining a robust overall system with improved writer identification and verification performance.

Part II

Allograph-Level Approach

A modified version of this chapter was published as: Marius Bulacu, Lambert Schomaker, – “A Comparison of Clustering Methods for Writer Identification and Verification,” Proc. of 8th Int. Conf. on Document Analysis and Recognition (ICDAR 2005), IEEE Computer Society, 2005, pp. 1275-1279, vol. II, 29 August - 1 September, Seoul, Korea

Chapter 4

Grapheme Clustering for Writer Identification and Verification

*Perfect clarity would profit the intellect, but damage the will.
We arrive at truth, not by reason only, but also by the heart.*

Blaise Pascal

Abstract

This chapter introduces our allograph-level method for writer identification and verification. The fundamental underpinning of this effective method is the idea of assuming that each writer acts as a stochastic generator of ink-trace fragments, or graphemes. The probability distribution of these simple shapes in a given handwriting sample is characteristic for the writer and is computed using a common codebook of graphemes obtained by clustering. Originally proposed in (Schomaker and Bulacu 2004), the theoretical model that supports this approach is also provided here in its essential aspects. While in other studies (Schomaker and Bulacu 2004, Schomaker et al. 2004), contours were employed to encode the graphemes, the current work explores a complementary shape representation using normalized bitmaps. The most important aim of the present study is to compare three different clustering methods for generating the grapheme codebook: k-means, Kohonen SOM 1D and 2D. Large scale computational experiments show that the proposed allograph-level writer identification method is robust to the underlying shape representation used (whether contours or normalized bitmaps), to the size of codebook used (stable performance for sizes from 10^2 to 2.5×10^3) and to the clustering method used to generate the codebook (essentially the same performance was obtained for all three clustering methods).

4.1 Introduction

Research in writer identification and verification has received significant interest in recent years due to its forensic applicability (Zois and Anastassopoulos 2000, Said et al. 2000, Srihari et al. 2002, Schomaker and Bulacu 2004, Schlapbach and Bunke 2004, Bensefia et al. 2005b). In *writer identification*, for a query sample of unknown author, a one-to-many search is performed over a large database with handwritten samples of

known authorship. The system retrieves a reduced list of candidates containing the samples most similar to the query in terms of individual handwriting style. This reduced list will be further scrutinized by the forensic expert in order to take the final decision regarding the identity of the author for the questioned sample. Writer identification is therefore possible only if there exist previous samples of handwriting by that person enrolled in the forensic database. In *writer verification*, a one-to-one comparison is performed with an automatic decision whether or not the two compared samples are written by the same person. The decidability of this problem reflects the nature of handwriting individuality (Srihari et al. 2002) and also the discrimination power of the features used for the writer verification task.

While *texture-level* methods that use directional PDFs (capturing slant, size, curvature, regularity) prove to be very efficient as seen in the previous chapters, they must be complemented by *allograph-level*, i.e. character-shape based approaches in order to obtain adequate and robust results. The discriminatory power of singular characters is analyzed in (Srihari et al. 2003) and (Zhang et al. 2003), under the assumption that the individual characters are perfectly separated and labeled using some form of human intervention. Other new results also show that writer-specialized handwriting recognizers can be used for writer identification and verification (Schlapbach and Bunke 2004).

In recent work, Schomaker has proposed an effective writer identification method in which the writer is assumed to act as a stochastic generator of ink-blob shapes, or graphemes (Schomaker and Bulacu 2004). The probability distribution of grapheme usage is characteristic of each writer and is computed using a common codebook obtained by clustering. A brief account of the underlying theoretical model of this approach will be given in the next section of this chapter. This theoretically founded approach was initially applied to isolated uppercase handwriting (Schomaker and Bulacu 2004) and later it was extended to lowercase cursive handwriting by using a segmentation method (Schomaker et al. 2004).

In these previous studies, we have used contours for shape representation and a 2D Kohonen self-organizing map (KSOM) for generating the grapheme codebook. While contours possess definite advantages for shape matching, they are nevertheless susceptible to problems regarding the starting point, open/closed loops or the presence of multiple inner contours. On the other hand, pixel-based representations can be more robustly extracted from the handwriting images, but the matching process becomes more vulnerable in this case, e.g., due to quantization in rescaling. The first purpose of this work is to explore the use of normalized bitmaps as the underlying shape representation. In this respect, our study comes closest to the work reported in (Bensefia et al. 2002, Bensefia et al. 2003, Bensefia et al. 2004) where an information-retrieval framework is used for writer identification. In contrast, our approach uses explicit probability distributions constructed on the basis of the shape codebooks to characterize writer individuality.

The second and most important purpose of the current work is to compare three different clustering methods for generating the grapheme codebook: k-means, Kohonen SOM 1D and 2D. We have run large scale computational experiments for comparing these three clustering methods over a large range of codebook sizes. Both writer identification and verification will be considered in our evaluation.

4.2 Theoretical model

The process of handwriting consists of a concatenation of ballistic strokes bounded by points of high curvature in the pen-tip trajectory (Schomaker 1991). Curved shapes are realized by differential timing in the movements of the wrist and finger subsystems (Schomaker et al. 1989). Handwriting is not a feedback process governed by peripheral environment factors. As a consequence of neural and neuro-muscular propagation delays, handwriting would be too slow if based upon continuous feedback (Schomaker 1991). Rather, the brain is planning series of ballistic strokes ahead in time in a feed-forward manner (Plamondon and Maarse 1989, Plamondon and Guerfali 1998). A character is assumed to be produced by a "motor program" (Schmidt 1975) that can be triggered to produce a pen-tip movement yielding the character shape on paper (Schomaker et al. 1989).

The allographic shape variations reflecting the character forms engrained in the motor memory of the writer allows for very effective writer identification and verification. Schomaker proposed a theoretical model and provided an experimental evaluation for this allograph-level approach to writer identification (Schomaker and Bulacu 2004). Here we will present the main aspects of this model; more details can be found in (Schomaker and Bulacu 2004).

Assume there exists a finite list S of allographs for a given alphabet L . Each allograph s_{li} is considered to be the i th allowable shape (style) variation of a letter $l \in L$ which should in principle be legible at the receiving end of the writer-reader communication line (Kondo and Attachoo 1986). The source of allographic variation may be located in teaching methods and individual preferences. The human writer is thus considered to be a pattern generator, stochastically selecting each allograph shape s_{li} when a letter l is about to be written (Shannon 1948). It is assumed that the probability density function $p_w(S)$, i.e., the probability of allographs being emitted by writer w , will be informative in the identification of writer w if it holds that

$$w \neq v \implies p_w(S) \neq p_v(S) \quad (4.1)$$

where w and v denote writers, S is a common allograph codebook and $p(\cdot)$ represents

the discrete PDF for allograph emission. This (i.e. eq. 4.1) will be realizable if for handwritten samples u emitted by w and characterized by

$$\vec{x}_{wu} = p_{wu}(S) \quad (4.2)$$

and assuming that the sample u is representative

$$\vec{x}_{wu} \approx p_w(S) \quad (4.3)$$

it holds that

$$\forall a, b, c, w, v \neq w : \Delta(\vec{x}_{wa}, \vec{x}_{wb}) < \Delta(\vec{x}_{wa}, \vec{x}_{vc}) \quad (4.4)$$

where Δ is an appropriate distance function on PDFs \vec{x} , v and w denote writers, as before, and a, b, c are handwriting-sample identifiers. Equation 4.4 states that, in feature space, the distance between any two samples of the same writer is smaller than the distance between any two samples by different writers. In ideal circumstances, this relation would always hold, leading to perfect writer identification. Note that in this model (eq. 4.1), the implication is unidirectional: in case of forged handwriting, $p_w(S)$ does not equal $p_v(S)$ but writer w imposes as v ($w = v$).

A problem at this point is that an exhaustive list S of allographs for a particular script and alphabet is difficult to obtain in order to implement this stochastic allograph-emission model. Clustering of character shapes with a known letter label is possible and has been realized (Vuurpijl and Schomaker 1997). However, the amount of handwritten image data for which no character ground truth exists vastly exceeds the size of commercial and academic training sets which are labeled at the level of individual characters. At this point in time, a commonly accepted list of handwritten allographs (and their commonly accepted names, e.g., in Latin, such as in the classification of species in the field of biology) does not exist, as yet. In this respect, it is noteworthy that for machine-print fonts, with their minute shape differences in comparison to handwriting variation, named font categories exist (e.g., Times-Roman, Helvetica, etc.), whereas we do not use generally agreed names for handwritten character families.

Therefore, it would be conducive to use an approach which avoids expensive character labeling at both training and operational stages. Unfortunately, automatic character segmentation in handwriting cannot be performed reliably. As a consequence, we will use a generic and imperfect segmentation method that generates ink fragments (graphemes) that do not overlap complete characters. These glyphs are usually sub-allographic parts of characters and have the advantage that they can be extracted reliably and in a non-parametric manner from a handwritten sample.

Despite not being complete characters, the graphemes generated by the heuristic over-segmentation method are nevertheless informative of the allographic character

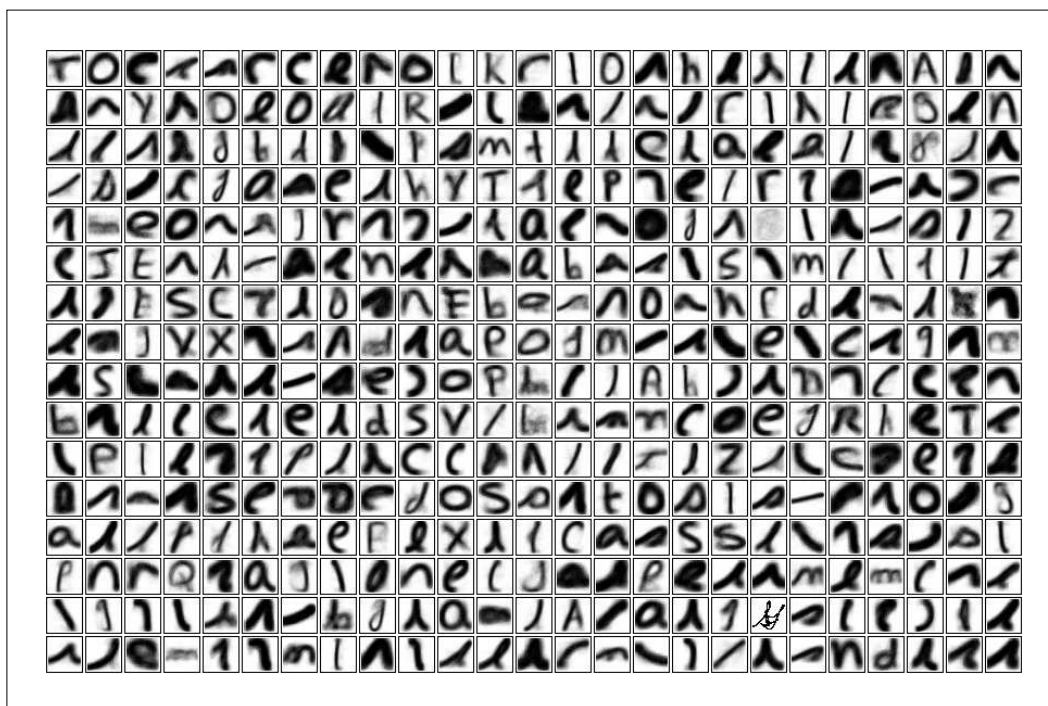


Figure 4.1: Example of shape codebook with 400 graphemes obtained by *kmeans* clustering applied to a set of 41k patterns produced by 65 writers from the *ImUnipen* database. The codebook graphemes have been placed 25 in a row.

variant of which they are a component. We will show that such sub-allographic script fragments are usable for writer identification. We will use the empirical distribution of grapheme occurrence as an approximation for the writer-specific allograph-emission probability.

We will assume a finite set or codebook C of sub- or supra-allographic shapes and we will estimate and use $p_w(C)$ as the writer descriptor in our identification and verification tests. In the next sections, we will describe the construction of the grapheme codebook C , the computation of an estimate of the writer-specific pattern-emission PDF $p_w(C)$, and an appropriate distance function Δ for these PDFs.

4.3 Datasets

The writer identification and verification study reported here was performed using two datasets: *Firemaker* and *ImUnipen*.

For the tests carried out in this chapter, we use pages 1, 2 and 4 from the *Firemaker* set (Schomaker and Vuurpijl 2000) comprising handwriting collected from 250 Dutch subjects, predominantly students. Page 1 contains 5 paragraphs of copied text in lower-

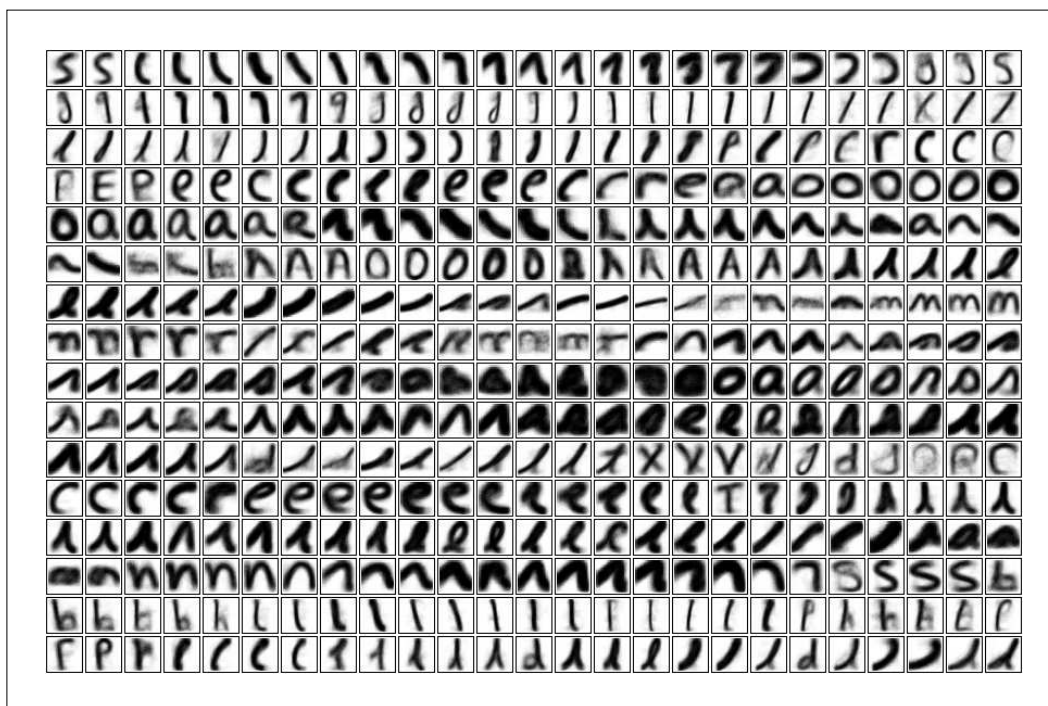


Figure 4.2: Example of shape codebook with 400 graphemes obtained by *ksom1D* clustering. The graphemes have been placed 25 in a row. This codebook displays a 1D order, in contrast to the “disorderly” codebook obtained by *kmeans* (see Fig. 4.1).

case handwriting. On page 2 there are 2 paragraphs of copied text in uppercase handwriting. The category of page 3 (“forged”) samples was not used. Page 4 contains a self-generated description of the content of a given cartoon. These samples consist of mostly lowercase handwriting of varying text content and the amount of written ink varies significantly, from 2 lines up to a full page. The scanned images have a resolution of 300 dpi, 8 bits / pixel, gray-scale. In the writer identification and verification experiments reported here, we performed searches/matches of page 1 vs. 4 (Firemaker lowercase) and paragraph 1 vs. 2 from page 2 (Firemaker uppercase).

The ImUnipen set contains handwriting from 215 subjects, 2 samples per writer. The images were derived from the Unipen database (Guyon et al. 1994) of on-line 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 handwriting with varying text content and amount of ink. The dataset was divided in two parts: 65 writers (130 samples) were used for training the grapheme codebook and the rest of 150 writers (300 samples) were used for testing.

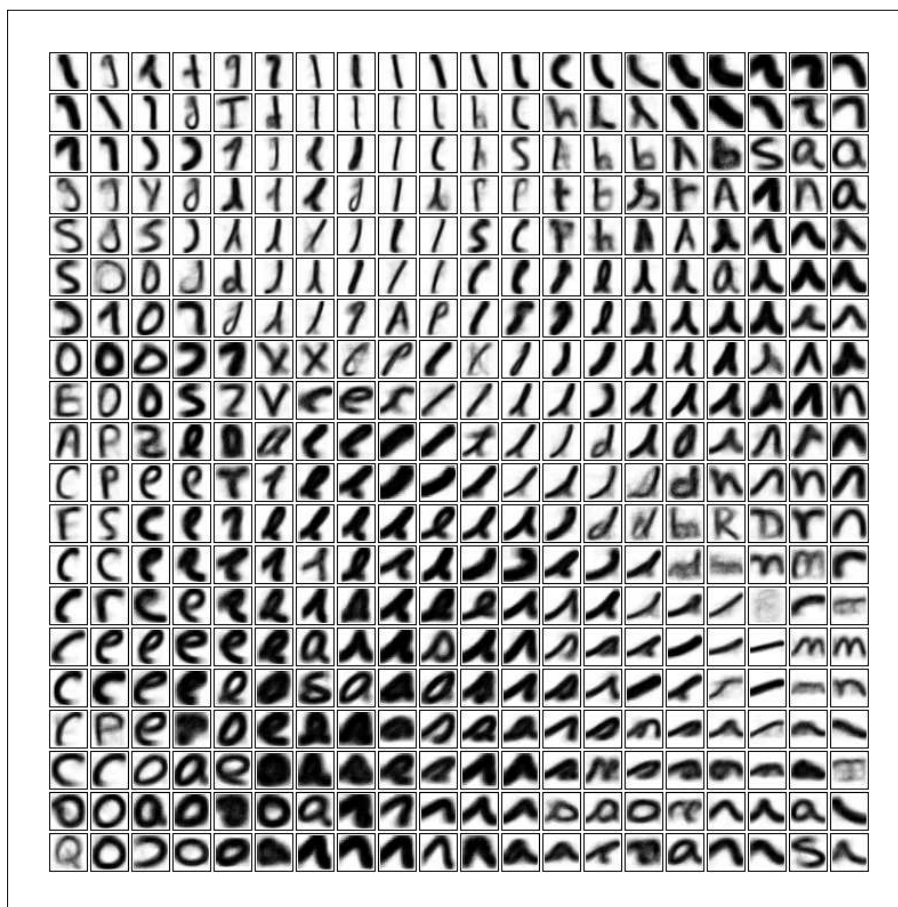


Figure 4.3: Example of shape codebook with 400 graphemes obtained by *ksom2D* clustering without wraparound. The 20x20 original SOM organization has been maintained and the 2D order of the codebook is clearly visible.

4.4 Segmentation method

In free-style cursive handwriting, connected-components may encompass several characters or syllables. A segmentation method that 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.

We segment handwriting at the minima in the lower contour with the added condition that the distance to the upper contour is in the order of the ink-trace width

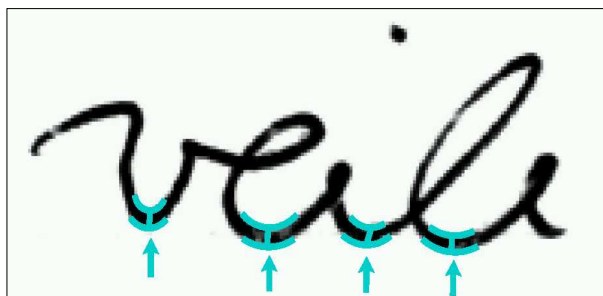


Figure 4.4: Segmentation at the minima in the lower contour that are proximal to the upper contour.

(see Fig. 4.4). For contour extraction we use Moore’s algorithm. After segmentation, graphemes are extracted as connected components, followed by a size normalization to 30x30 pixel bitmaps, preserving the aspect ratio of the original pattern.

4.5 Grapheme codebook generation

A number of 130 samples corresponding to 65 writers have been taken from the ImU-nipen dataset. The graphemes have been extracted from these samples using the described procedure yielding a training set containing a total of 41k patterns (normalized bitmaps).

Three clustering methods will be used to generate 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 (Kohonen 1988, Duda et al. 2001).

The size of the codebook (the number of clusters used) yielding optimal performance is an important parameter in our method. In the experiments, we will explore a large range of codebook sizes. This will allow a thorough comparison of the considered clustering algorithms.

Figures 4.1, 4.2 and 4.3 show examples of shape codebooks that have been obtained by training using each of the three clustering methods. The two grapheme 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.

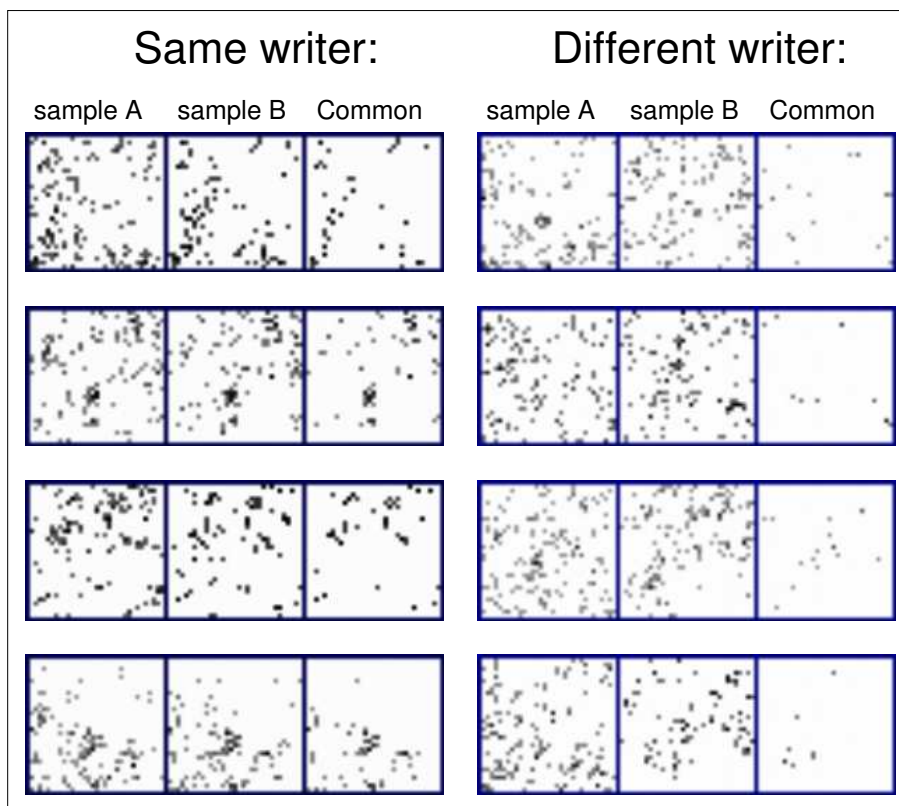


Figure 4.5: Density plots of grapheme emission PDFs computed using a *ksom2D* codebook. Two samples *A* and *B* from the same writer (left panel) yield a much larger common density than samples from different writers (right panel). The common density resulting from the overlap between the two sample PDFs is depicted in the third column ('Common').

4.6 Computing writer-specific grapheme-emission PDFs

The writer is considered to be characterized by a stochastic pattern generator, producing a family of basic shapes (Schomaker and Bulacu 2004). 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 j (the winner) is found using the Euclidean distance and this occurrence is counted into the corresponding histogram bin:

$$j = \operatorname{argmin}_n[\operatorname{dist}(g, C_n)], h_{ij} \leftarrow h_{ij} + 1 \quad (4.5)$$

where n is an index that runs over the shapes in the codebook C . In the end, the his-

togram h_i is normalized to a probability distribution function p_i that sums up to 1. This PDF (see Fig. 4.5) is the writer descriptor used for identification and verification.

4.7 Results

We performed large scale computational experiments to compare the three clustering methods over a large range of codebook sizes. The number of clusters used was varied from 9 (3x3) to 2500 (50x50). A number of 200 epochs have been used for training the Kohonen SOMs. Computations have been performed on a Beowulf high-performance Linux cluster with 1.7GHz/0.5GB nodes. Training times for codebooks of size 400: k-means - 1 hrs, ksom1D - 10 hrs, ksom2D - 17 hrs. Computation times for the grapheme emission PDF on codebooks of size 400: 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 approx. 800 CPU hrs.

4.7.1 Writer identification

Writer identification results are computed using nearest-neighbor classification (Cover and Hart 1967) in a leave-one-out strategy. For a query sample q , the distances to all the other samples $i \neq q$ are computed. Then all the samples i are ordered in a sorted hit list with increasing distance to the query q (Press et al. 1992). Ideally, the first ranked sample (Top 1) should be the pair sample produced by the same writer (in all our experiments there are 2 samples per writer).

An appropriate dissimilarity measure between the grapheme PDFs is the χ^2 distance (Press et al. 1992):

$$\chi_{qi}^2 = \sum_{n=1}^k \frac{(p_{qn} - p_{in})^2}{p_{qn} + p_{in}} \quad (4.6)$$

where p are entries in the PDF, n is the bin index and k is the number of bins in the PDF (equal to the size of the grapheme codebook). In our experiments, χ^2 outperformed other distance measures: Hamming, Euclid, Minkowski order 3, Bhattacharya.

We point out that our writer identification results are realistic and rather conservative because we do not make a separation between a training set and a test set. Keeping all the data in one batch makes the testing conditions actually more difficult and realistic, with more distractors: not 1, but 2 per false writer and only one correct hit.

Figures 4.6, 4.7 and 4.8 show our results obtained on the experimental datasets. Writer identification performance (Top-1 and Top-10) reaches a plateau for codebook

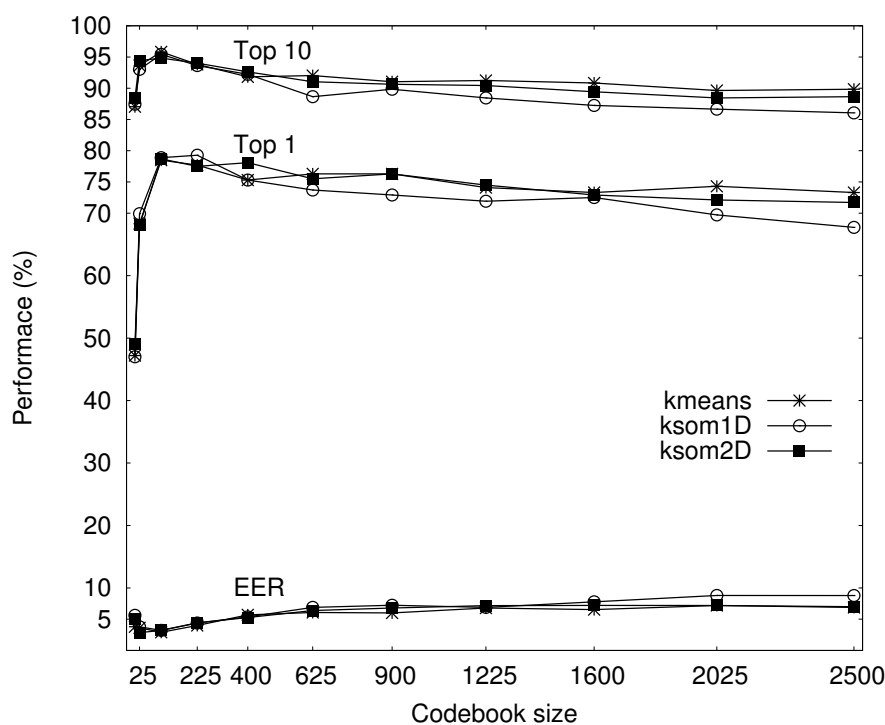


Figure 4.6: *Writer identification and verification performance on the Firemaker lowercase dataset as a function of codebook size.*

sizes larger than about 100 (10x10) shapes. More remarkable is the fact that the same performance is achieved by all three clustering methods. Table 4.1 gives numerical results for codebooks of size 400 which was chosen as an anchor point.

The writer identification performance is stable over a very large range of codebook sizes, from 100 to 2500. Therefore the codebook size does not represent a critical parameter for our allograph-level writer identification approach. Additionally, the three clustering algorithms used to generate the shape codebook yielded the same level of performance. We are thus confident that our results are robust and reproducible.

The lower performance obtained on the Firemaker uppercase dataset can be explained by two factors: the amount of handwriting in these samples is very reduced (only one paragraph of 100-150 characters) and the codebooks have been generated based on samples that contain almost exclusively lowercase (cursive) handwriting. Nevertheless, the overall performance levels achieved on lowercase and uppercase are quite comparable. In the previous chapter, using edge-based directional features under the condition that approximately the same amount of ink is present in all samples, the performance level achieved on lowercase and uppercase was roughly the same (Bulacu and Schomaker 2003). Here again, the empirical results contradict the intuition that writer identification is more effective on lowercase rather than uppercase handwriting.

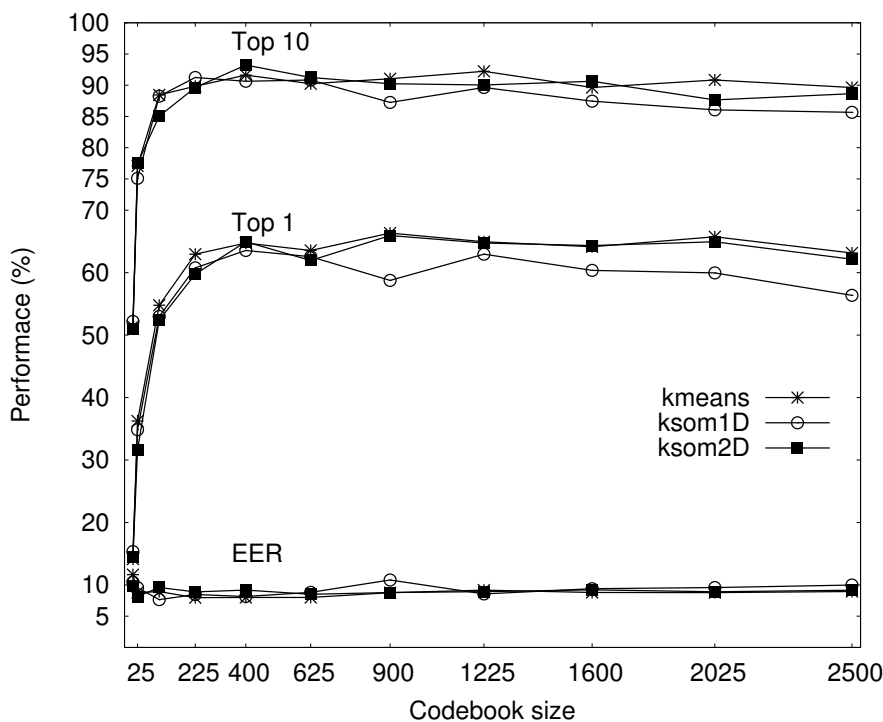


Figure 4.7: Writer identification and verification performance on the Firemaker uppercase dataset as a function of codebook size.

The slightly higher performance obtained on ImUnipen is due to the smaller number of writers contained in the dataset.

The writer identification results presented here are in the same ballpark as the ones we reported in a previous study using contours for shape representation and Kohonen 2D for codebook training (Schomaker et al. 2004). This constitutes additional evidence regarding the robustness of the proposed method of using grapheme emission PDFs for writer identification.

4.7.2 Writer verification

In the writer verification task, the distance ξ between two given handwriting samples is computed using the grapheme PDFs. 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

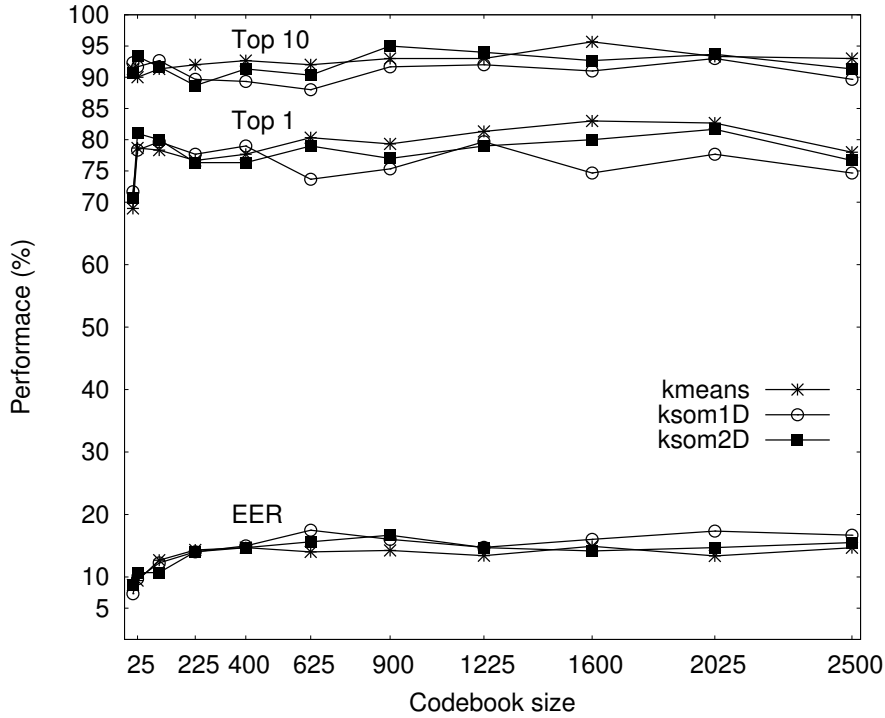


Figure 4.8: Writer identification and verification performance on the ImUnipen dataset (150 writers, 2 samples / writer, not used for training the grapheme codebook) as a function of codebook size.

of documents, FAR becomes false alarm rate, while FRR becomes 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 $P_S(\xi)$ and the probability distribution of distances between samples written by different persons $P_D(\xi)$:

$$FAR = \int_0^T P_D(\xi) d\xi \quad (4.7)$$

$$FRR = \int_T^\infty P_S(\xi) d\xi. \quad (4.8)$$

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.

For the Firemaker dataset, $P_S(\xi)$ has been constructed using the 250 same-writer distances, while $P_D(\xi)$ has been constructed using all the $C_{500}^2 - 250 = 124500$ different-writer distances arising in the dataset. Similarly for ImUnipen. In figures 4.6, 4.7 and

Table 4.1: Writer identification and verification accuracies (percentages) for codebooks of size 400 (20x20). The same writer identification and verification performance is achieved by all three clustering methods. The performance levels are consistent across the three datasets.

Dataset / Method		<i>kmeans</i>	<i>ksom1D</i>	<i>ksom2D</i>
Firemaker lowercase (250 writers)	Top 1	75.3	75.3	78.1
	Top 10	91.8	92.2	92.6
	EER	5.7	5.4	5.3
Firemaker uppercase (250 writers)	Top 1	64.7	63.6	64.9
	Top 10	91.6	90.6	93.2
	EER	8.0	8.2	9.2
ImUnipen (150 writers)	Top 1	77.7	79.0	76.3
	Top 10	92.7	89.3	91.3
	ERR	14.7	15.0	14.7

4.8, the lower family of curves show the ERR as a function of codebook size. Here again the same performance is achieved by all three clustering methods.

For Firemaker uppercase, the EER hovers around 8%. For Firemaker lowercase, the EER reaches a minimum of about 3% for a codebook size of 100 and increases to about 7% for larger codebooks. A similar increase in the ERR for larger codebooks can be seen also for the ImUnipen set, from 8% (codebook with 9 shapes) to 14% (codebooks with 10^3 shapes). 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 becomes gradually more difficult to find a unique threshold distance that separates the sample pairs written by the same person from those written by different persons. 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. A slight decrease in the writer identification performance with increasing codebook size can however be noticed in Fig. 4.6.

We must point out that the essence of the proposed method does not consist in an exhaustive enumeration of all possible allographic part shapes. Rather, the grapheme codebook spans up a shape space by providing a set of nearest-neighbor attractors for

the ink fraglets extracted from a given handwritten sample. The three clustering methods considered in this chapter seem to perform this task equally well.

4.8 Conclusions

The use of grapheme emission PDFs in writer identification and verification yields valuable results. Ultimately, writing style is determined by allographic shape variations and small style elements which are present within a character are the result of the writer's physiological make up as well as education and personal preference. The proposed method proves to be robust to the underlying shape representation used (whether contours or normalized bitmaps), to the size of codebook used (stable performance for sizes from 10^2 to 2.5×10^3) and to the clustering method used to generate the codebook (essentially the same performance was obtained for k-means, ksom1D and ksom2D).

In the next chapter of the thesis, we will combine the texture-level and allograph-level approaches to improve the performance and robustness of our writer identification and verification system. We will also extend our experiments to bigger datasets containing more writers.

A modified version of this chapter was published as: Marius Bulacu, Lambert Schomaker, – “Text-independent writer identification and verification using textural and allographic features,” IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), Special Issue - Biometrics: Progress and Directions, IEEE Computer Society, vol. 29, no. 4, pp. 701-717, April 2007

Chapter 5

Feature Fusion for Text-Independent Writer Identification and Verification

If the brain were so simple we could understand it, we would be so simple we couldn't.

Lyall Watson

Abstract

In the previous chapters, we presented the development of 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 independently of the textual content of the written samples. This chapter presents a coherent overview of all our features and specifically considers the problem of combining multiple features for text-independent writer identification and verification. Our experiments are also extended to larger datasets containing up to 900 writers.

Our features operate at two levels of analysis: the texture level and the character-shape (allograph) level. For computing the directional texture level features, here we use contours, rather than edges, with definite advantages regarding computation speed and control of feature dimensionality. The contour-based joint directional PDFs 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 texture-level and allograph-level features yields very high writer identification and verification performance, with usable rates for datasets containing 10^3 writers.

5.1 Introduction

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. In this chapter, we present an overview of our statistical pattern recognition methods for automatic writer identification and verification using off-line handwriting. We specifically consider the problem of combining multiple features for improving performance on both tasks of writer identification and verification, a topic that was not fully addressed in previous chapters. Here we provide an extensive analysis of feature combinations and report our experimental results obtained on larger datasets containing up to 900 writers.

There are two general characteristics distinguishing our approach: *human intervention is minimized* in the writer identification and verification 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 agnostic of what has been written in the samples. The development of our writer identification techniques takes place at a time when many biometric modalities undergo 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 (Plamondon and Srihari 2000, Vinciarelli 2002). For automatic handwriting recognition, invariant representations are sought which are capable of eliminating variations between different handwritings 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 represent therefore two opposing facets of handwriting analysis. It is important, however, to mention also 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*

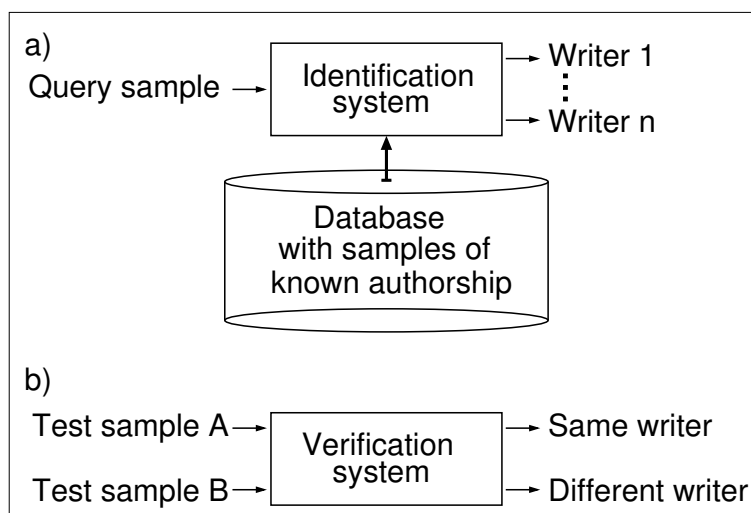


Figure 5.1: a) A writer identification system retrieves, from a database containing handwritings of known authorship, those samples that are most similar to the query. The hit list is then analyzed in detail by a human expert. b) A writer verification system compares two handwriting samples and takes an automatic decision whether or not the input samples were written by the same person.

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. 5.1a). This represents a special case of image retrieval, where the retrieval process is based on features capturing handwriting individuality. The hit list is further scrutinized by the forensic expert who takes 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 whether or not the two samples are written by the same person (see Fig. 5.1b). 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% recall of the correct writer in a hit list of one hundred writers, computed from a database in 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.



Figure 5.2: 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.

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. 5.2). The results reported in this thesis 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. 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 underpinnings 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 mo-

tor program, such as size, shape, timing; finally, at a peripheral level, commands are generated for the biophysical muscle-joint systems (Maarse 1987). 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 (Maarse 1987) and in a manner that is also independent of changes in the horizontal progression motion (Maarse and Thomassen 1983). Due to neural and neuromechanical propagation delays, a handwriting process based upon a continuous feedback mechanism alone would evolve too slowly (Schomaker 1991). 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" (Schmidt 1975). Every person uses personalized and characteristic shapes, called *allographs*, when writing a chosen letter of the alphabet (see Fig. 5.2). In this thesis, 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 (Schomaker and Bulacu 2004). Furthermore, very effective writer identification and verification is achievable by combining texture-level and allograph-level features, which together offer a fuller description of a person's stable and discriminatory unconscious practices in writing.

This chapter is organized as follows. Section 5.2 describes the datasets used in the experiments reported in this chapter. Sections 5.3 and 5.4 give an overall coherent overview of 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 5.5. Section 5.6 gives the experimental results, followed by a discussion in Section 5.7. Conclusions are then drawn in Section 5.8.

5.2 Experimental datasets

The experiments reported in this chapter were conducted using three datasets: Firemaker, IAM and ImUnipen. The Firemaker and ImUnipen datasets were described previously in the thesis, while IAM is a large dataset newly introduced in this chapter. The IAM dataset (Marti and Bunke 2002) is available on the Internet and was extensively used for off-line handwriting recognition. In addition to the annotation of the textual content, the IAM set contains also writer identity information needed in writer identification studies. For completeness, we provide here brief descriptions of all the datasets

used in the experiments reported in this chapter.

The Firemaker set (Schomaker and Vuurpijl 2000) contains handwriting collected from 250 Dutch subjects, predominantly students, who were required to write 4 different A4 pages. On page 1 they were asked to copy a text of 5 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 2 paragraphs using only uppercase letters. Page 3 contains "forged" text and these samples 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 2 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 chapter, we performed searches / matches of page 1 vs. 4 (Firemaker lowercase) and paragraph 1 vs. 2 from page 2 (Firemaker uppercase).

The IAM database (Marti and Bunke 2002) 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 (Zimmermann and Bunke 2002). This dataset includes a variable number of handwritten pages per writer, from 1 page (350 writers) to 59 pages (1 writer). In order to have comparable experimental conditions across all datasets, we modified the IAM set to contain always 2 samples per writer: we kept only the first 2 documents for those writers who contributed more than 2 documents to the original IAM dataset 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 persons, 2 samples per writer. The amount of ink is roughly equal in the two samples belonging to one writer, but varies between writers from 3 lines up to a full page.

The ImUnipen set contains handwriting from 215 subjects, 2 samples per writer. The images were derived from the Unipen database (Guyon et al. 1994) of on-line handwriting. The time sequences of coordinates were transformed to simulated 300 dpi images using a Bresenham line generator and an appropriate thickening function. The samples contain lowercase handwriting with varying text content and amount of ink. This set was not directly used in the writer identification and verification tests reported in this chapter. However, a part of this dataset 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 datasets to obtain a combined set which we named "Large". The Large dataset therefore contains 900 writers, 2 samples per writer, lowercase handwriting. This combined set is comparable, in terms of number

Table 5.1: Overview of the experimental datasets, the number of writers contained and some of their properties.

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

of writers, to the largest dataset used in writer identification and verification until the present (Srihari et al. 2002). 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 datasets and to collect data from multiple sources and different languages in a common framework. Table 5.1 gives an overview of all datasets used in our tests.

5.3 Textural features

Asserting writer identity based on handwriting images requires three main processing phases:

- 1) feature extraction,
- 2) feature matching / feature combination,
- 3) writer identification and verification.

In this and in the following sections of this chapter, we present the feature extraction methods in a general coherent framework. We use probability distribution functions

Table 5.2: Overview of the considered features, their dimensionalities and the distance functions used in identification and verification. Features are grouped into four different categories: directional PDFs ($f1$, $f2$, $f3h$, $f3v$), grapheme emission PDF ($f4$), run-length PDFs ($f5h$, $f5v$) and autocorrelation ($f6$).

Feature	Explanation	N dims	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
$f3h: p(\phi_1, \phi_3) h$ $f3v: p(\phi_1, \phi_3) v$	Direction co-occurr. PDFs \rightarrow horiz. run \rightarrow vert. run	144 144	χ^2 χ^2	contours
$f4: p(g)$	Grapheme emission PDF	400	χ^2	connected components
$f5h: p(rl) h$ $f5v: p(rl) v$	Run-length on white PDFs \rightarrow horiz. run \rightarrow vert. run	60 60	χ^2 χ^2	binary image
$f6: ACF$	Autocorr. horiz.	60	L_2	gray-scale image

(PDFs) extracted from the handwriting images to characterize writer individuality in a text-independent manner. The term "feature" will be 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 5.2. 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 chapter, we have selected the most discriminative features from a larger number of features tested in a previous paper (normalized entropy, ink-density PDF, wavelets) (Schomaker and Bulacu 2004).

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 (Otsu 1979). The binary images, in which only the ink pixels are "on", undergo connected component detection (labeling) using

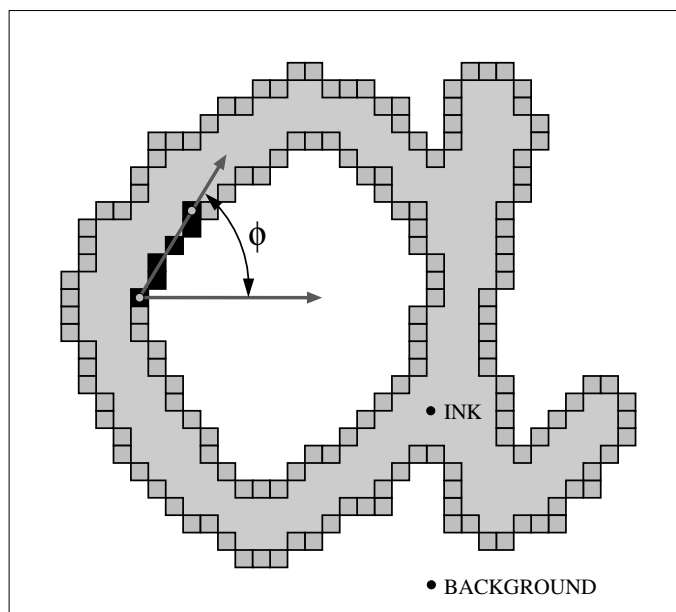


Figure 5.3: Schematic description for the extraction method of the contour-direction PDF (feature f_1). The handwritten letter "a", provided as an example, would be roughly twice as large in reality.

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 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. These features were computed using the edge image in the previous chapters of the thesis. 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 pre-processing 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 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 used in writer identification and verification. 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.

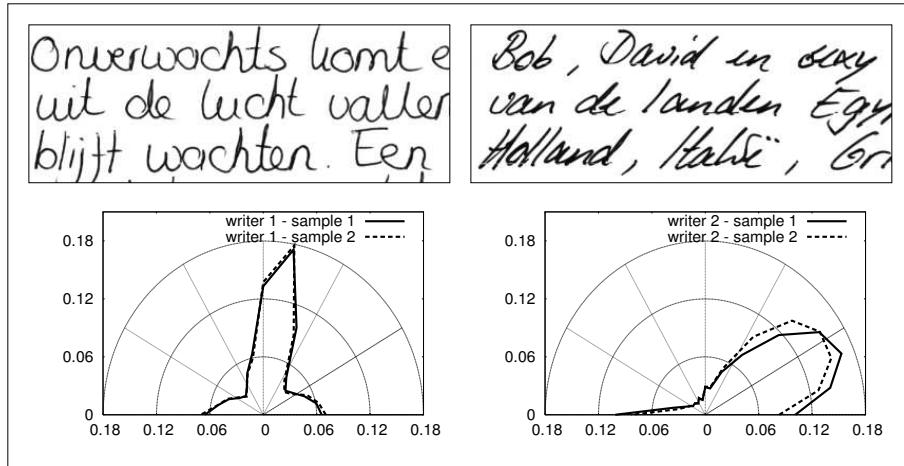


Figure 5.4: Examples of lowercase handwriting from two different subjects. We superposed the polar diagrams of the direction distribution $p(\phi)$ extracted from the two handwritten samples for each of the two subjects. There is a large overlap between the directional PDFs extracted from samples originating from the same writer, while there is a substantial variation in the directional PDFs for different writers. The examples were chosen for visual clarity.

5.3.1 Contour-direction PDF (f_1)

The most prominent visual attribute of handwriting that reveals individual writing style is slant. Handwriting slant is also a very stable personal characteristic (Maarse and Thomassen 1983, Maarse 1987). It has long been known in handwriting research that the distribution of directions in the script provides useful information for writer identification (Maarse et al. 1988), coarse writing-style classification (Crettez 1995) or signature verification (Drouhard et al. 1995). 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. 5.3) and the angle that the fragment makes with the horizontal is computed using equation 5.1. As the algorithm runs over the contours, the orientation of the local contour fragments is computed and an angle histogram is built thereby. 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.

$$\phi = \arctan\left(\frac{y_{k+\epsilon} - y_k}{x_{k+\epsilon} - x_k}\right) \quad (5.1)$$

The parameter ϵ controls the length of the analyzing contour fragment. In our im-

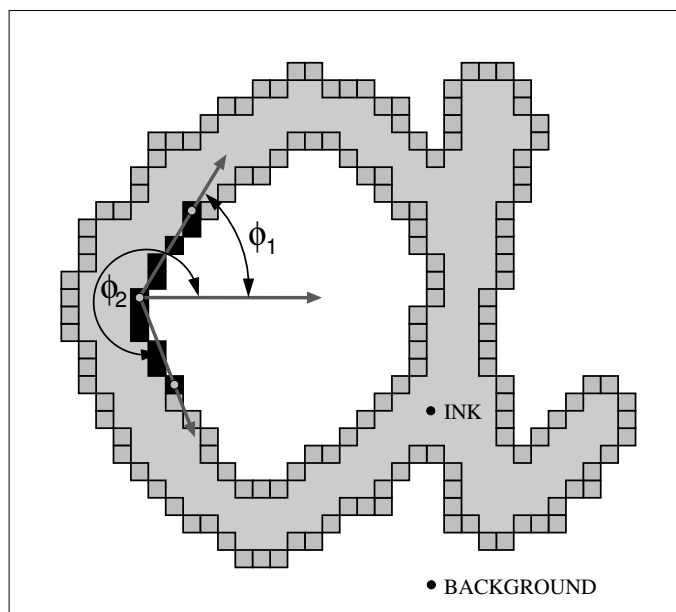


Figure 5.5: Schematic description for the extraction method of the contour-hinge PDF (feature f_2).

plementation $\epsilon = 5$ 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 $0^\circ - 180^\circ$ was set to $n = 12$ through experimentation: $15^\circ / \text{bin}$ 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 the directional features presented in this chapter.

The prevalent direction in $p(\phi)$ (see Fig. 5.4) 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 the distribution $p(\phi)$.

5.3.2 Contour-hinge PDF (f_2)

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 verifica-

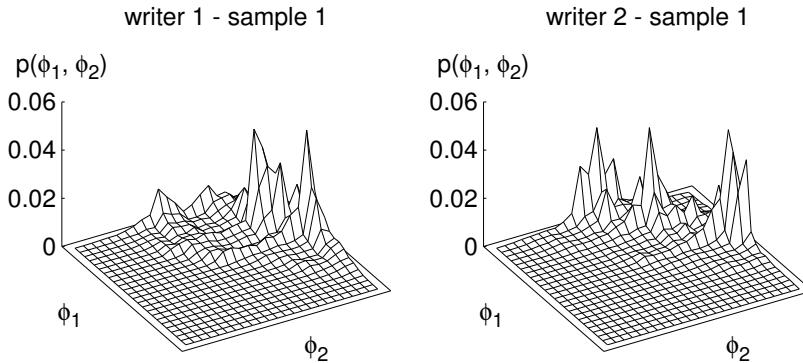


Figure 5.6: Surface plots of the contour-hinge PDF $p(\phi_1, \phi_2)$ for two writers. One half of the 3D plot (on one side of the main diagonal) is flat because we only consider angle combinations with $\phi_2 \geq \phi_1$.

tion performance. In order to capture, besides orientation, also 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. 5.5). 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 with feature $f1$ for which spanning the upper two quadrants (180°) was sufficient, we now have to span all the 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 will consider only non-redundant 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 will have 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. 5.6. 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.

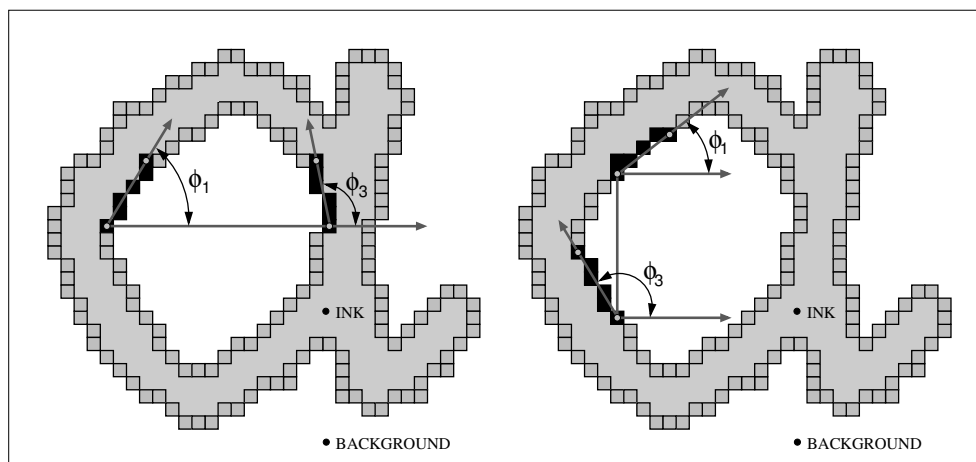


Figure 5.7: Schematic description for the extraction methods of the direction co-occurrence PDFs (horizontal scan - feature $f3h$ on the left and vertical scan - feature $f3v$ on the right).

5.3.3 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. 5.7). 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. Examples of $p(\phi_1, \phi_3)h$ for two writers are given in Fig. 5.8.

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.

5.3.4 Other texture-level features: run-length PDFs ($f5h, f5v$), autocorrelation ($f6$)

Run lengths were first proposed for writer identification in (Arazi 1977, Arazi 1983) and were also used on historical documents in (Dinstein and Shapira 1982). Run lengths

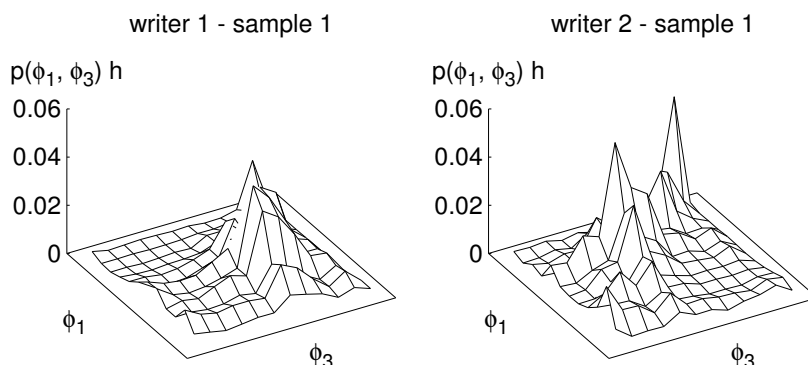


Figure 5.8: Surface plots of the contour-direction co-occurrence PDF $p(\phi_1, \phi_3)h$ for two writers. Every writer has a different “probability landscape”.

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 dataset is about 120 pixels).

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.

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.

5.4 Allographic features

In this section, we briefly reiterate our allograph-level approach to writer identification and verification. Our method, similar to the approach described in (Bensefia et al. 2005b), 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 (Schomaker and Bulacu 2004) and later it was extended to lowercase cursive handwriting by using a segmentation method (Schomaker et al. 2004).

This writer identification and verification method was fully described in the previous chapter of the thesis and involves three processing stages:

1) Handwriting segmentation: the ink is cut at the minima in the lower contour for which the distance to the upper contour is comparable to the ink-trace width (see Fig. 4.4). The graphemes are then extracted as connected components, followed by size normalization to 30x30 pixel bitmaps, preserving the aspect ratio of the original pattern. This segmentation stage makes our allograph-level method applicable to free-style handwriting, both cursive and isolated.

2) Shape codebook generation: grapheme clustering was applied to a training set containing 41k graphemes extracted from 130 samples (65 writers) from the ImUnipen set. On the new Large dataset, the three clustering algorithms used previously will be compared for a large range of codebook sizes: k-means, Kohonen SOM 1D and 2D (Kohonen 1988, Duda et al. 2001). Fig. 5.9 shows three examples of shape codebooks

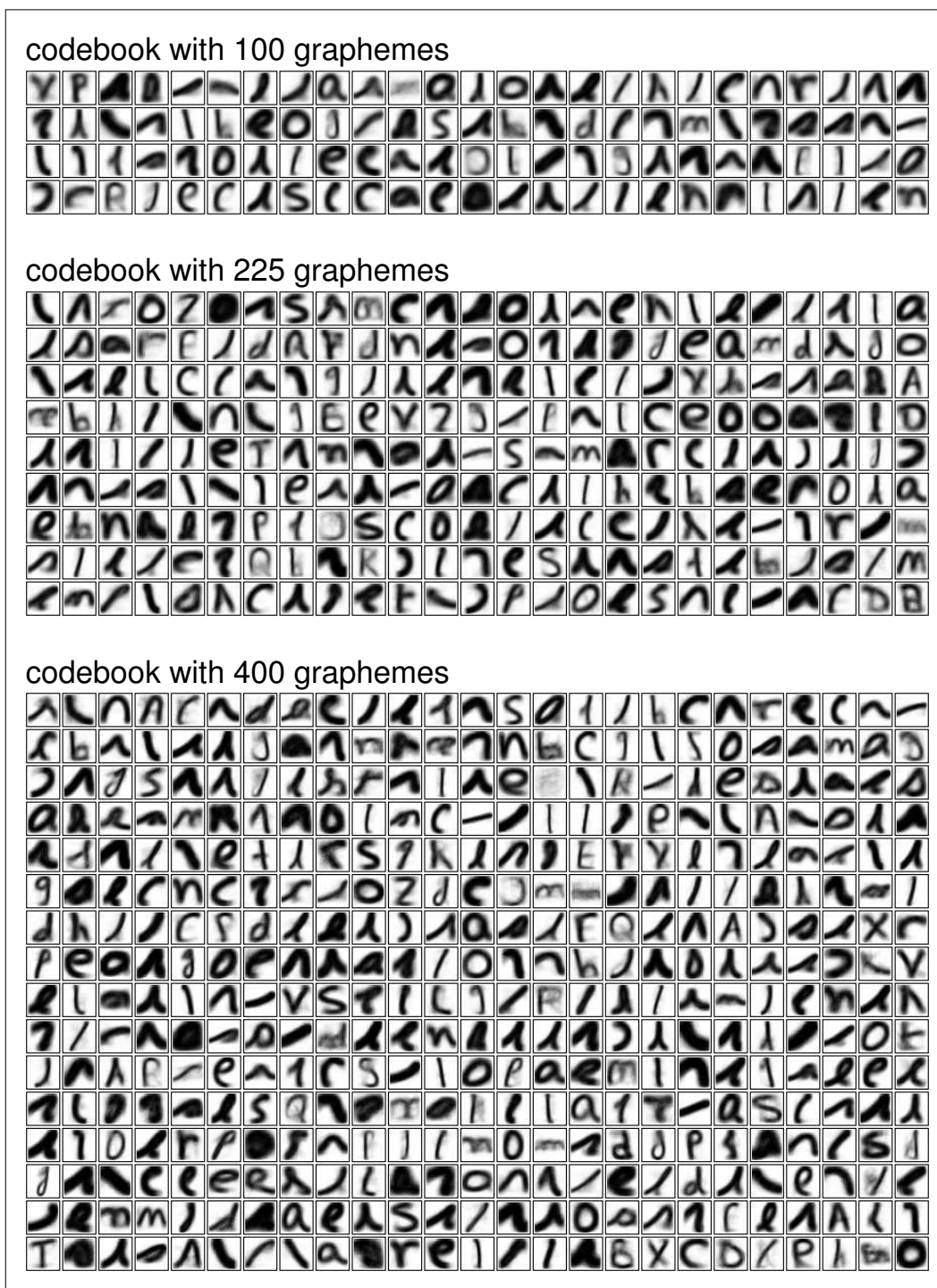


Figure 5.9: Examples of shape codebooks generated by *k*-means clustering and containing an increasing number of graphemes (increasing values of the parameter *k* were used in training).

generated by k-means clustering for increasing values of k . The codebook graphemes act as prototype shapes representative for the types of shapes to be expected as a result of handwriting segmentation.

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

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

The nature of the proposed method does not consist in an exhaustive enumeration of all possible allographic part shapes. Rather, the grapheme codebook spans up a shape space by providing a set of nearest-neighbor attractors for the ink fraglets extracted from a given handwritten sample. The occurrence PDF of these sub-allographic script fragments constitutes a very effective feature for writer identification and verification.

5.5 Feature matching and fusion for writer identification and verification

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, Hausdorff. We will report however only on the best performing ones.

For the PDF features (f_1, f_2, f_3, f_4, f_5), the χ^2 distance (Press et al. 1992) is used for matching a query sample q and any other sample i from the database:

$$\chi_{qi}^2 = \sum_{n=1}^{Ndim_s} \frac{(p_{qn} - p_{in})^2}{p_{qn} + p_{in}} \quad (5.2)$$

where p are entries in the PDF, n is the bin index and $Ndim_s$ 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 (f_6).

Writer identification is performed using nearest-neighbor (Cover and Hart 1967) classification in a "leave-one-out" strategy. For a query sample q , the distances to all 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 (Press et al. 1992). 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 experiments, we do not make a separation between a training set and a test set, all the data is in one suite. This is actually a more difficult and realistic testing condition, with more distractors: not 1, but 2 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 (Neyman and Pearson 1933). For writer verification, the distance ξ 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 false alarm rate, while FRR becomes 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 $P_S(\xi)$ and the probability distribution of distances between samples written by different persons $P_D(\xi)$:

$$FAR = \int_0^T P_D(\xi) d\xi \quad (5.3)$$

$$FRR = \int_T^\infty P_S(\xi) d\xi. \quad (5.4)$$

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 (Bulacu and Schomaker 2006),

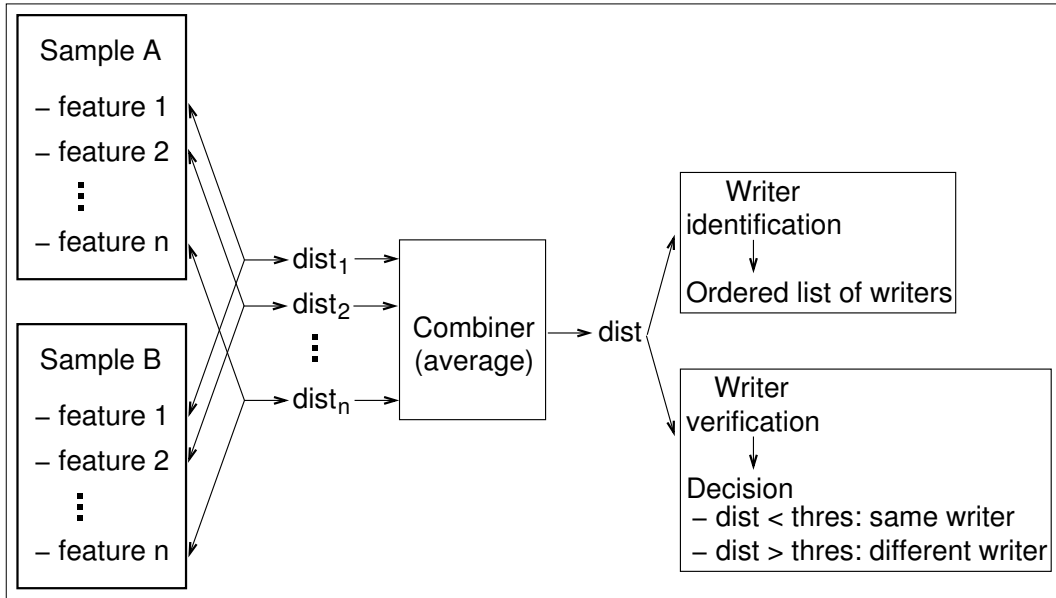


Figure 5.10: Feature combination scheme: the distances generated by the individual features are averaged (using simple or weighted average) and the result is then used in writer identification and verification.

this being the main focus of the present chapter of the thesis. 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. 5.10).

In feature combinations, Hamming distance performed best:

$$H_{qi} = \sum_{n=1}^{Ndims} |p_{qn} - p_{in}| \quad (5.5)$$

The χ^2 distance, due to the denominator (see eq. 5.2), 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 $P_S(\xi) \propto e^{-\xi/\sigma}$. The decay constants σ control the weights that different features take on in the combination. While this basic probabilistic model will almost certainly be violated in reality, experimental results show that significant performance improve-

ments 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 (Kittler et al. 1998) or multi-modal biometrics (Maltoni et al. 2003, Roli et al. 2002). Information can be combined at three levels in the biometric identification or verification process: *sensor fusion*, *similarity-score fusion* and *decision-level fusion* (Daugman 2000). Combining similarity scores ("soft" fusion) seems to be the method of choice in multi-modal 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.

5.6 Results

In this section of the chapter, 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 5.2 of this chapter, four datasets are considered in the experimental evaluation (see Table 5.1): Firemaker uppercase (250 writers), Firemaker lowercase (250 writers), IAM (650 writers) and Large (900 writers obtained by merging Firemaker lowercase and IAM datasets). All datasets contain 2 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 f_4) was built using part of the ImUnipen dataset (65 writers, 2 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 datasets. 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.

5.6.1 Performances of individual features

Table 5.3 gives the writer identification and verification performance of the individual features considered in this 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 datasets. The best performer is the contour-hinge PDF (f_2) followed by the grapheme-emission PDF (f_4).

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 datasets, 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 Chapter 3 in experiments where the amount of ink in the samples was controlled (Bulacu and Schomaker 2003). These findings contradict the idea, 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 dataset), performance is very low.

The writer identification performances obtained on Firemaker lowercase and IAM are very similar, albeit the large difference in the number of writers contained in the two datasets. This is probably due to differences in the writer distributions underlying the two datasets. The Firemaker dataset 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 dataset. Under these conditions, when these two datasets are combined, only a slight decrease in writer identification performance on the Large dataset is noticed. The dependence of the writer identification rate on number of writers contained in the dataset is discussed in the following section of this chapter. For the size of the datasets used here, the writer identification percentages are subject to a 3-4% confidence interval at a 95% confidence level.

Table 5.3: Writer identification and verification performance of individual features. The χ^2 distance was used in matching. The features are explained in Table 5.2.

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

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 (Daugman 1993, Jain et al. 1997).

Further more, 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 datasets). An analysis of writer identification performance vs. amount of ink contained in the samples is given in Chapter 2 (Bulacu et al. 2003).

The run length PDFs, despite having the worst performance among the echelon of features selected in this study, in fact do perform better than a number of other known writer identification features, e.g. entropy, wavelets (see (Schomaker and Bulacu 2004) for a wide analysis).

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 datasets. They constitute the gist of our text-independent approach to writer identification and verification.

5.6.2 Performances of feature combinations

The features considered in this thesis 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 lead us to prefer simplicity and robustness here and report the feature combination results obtained by plain distance averaging.

The features studied here can be grouped into four broad categories (see Table 5.2): 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.

Table 5.4: Writer identification and verification performance of feature combinations. 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 dataset.

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

First, we consider the natural combinations $f3h$ with $f3v$ and $f5h$ with $f5v$ (first two rows of Table 5.4). 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 5.4, the combined performance exceeds the performances of all individual features involved in the combination, with only one exception marked with parenthesis. As can be noticed, the performance of feature combinations is generally consistent over the four experimental datasets.

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

In Fig. 5.11a, we show the results obtained by considering a weighted combination between features $f2$ and $f4$: $d = (1 - \lambda)d_2 + \lambda d_4$, where λ is the mixing coefficient. Similarly, in Fig. 5.11b, we consider the combination $f3$ and $f4$: $d = (1 - \lambda)d_3 + \lambda d_4$. 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 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 up 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 autocorrelation feature $f6$ which requires weighting with respect to the other features. This has lead, however, only to minor additional improvements in performance, only about 1% increase in Top-1 identification rate.

We mention that we replaced the linear distance combiner with an SVM (Joachims 1999, Burges 1998, Cristianini and Shawe-Taylor 2000) 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 in Chapter 3 with only marginal performance improvements (Bulacu and Schomaker 2003).

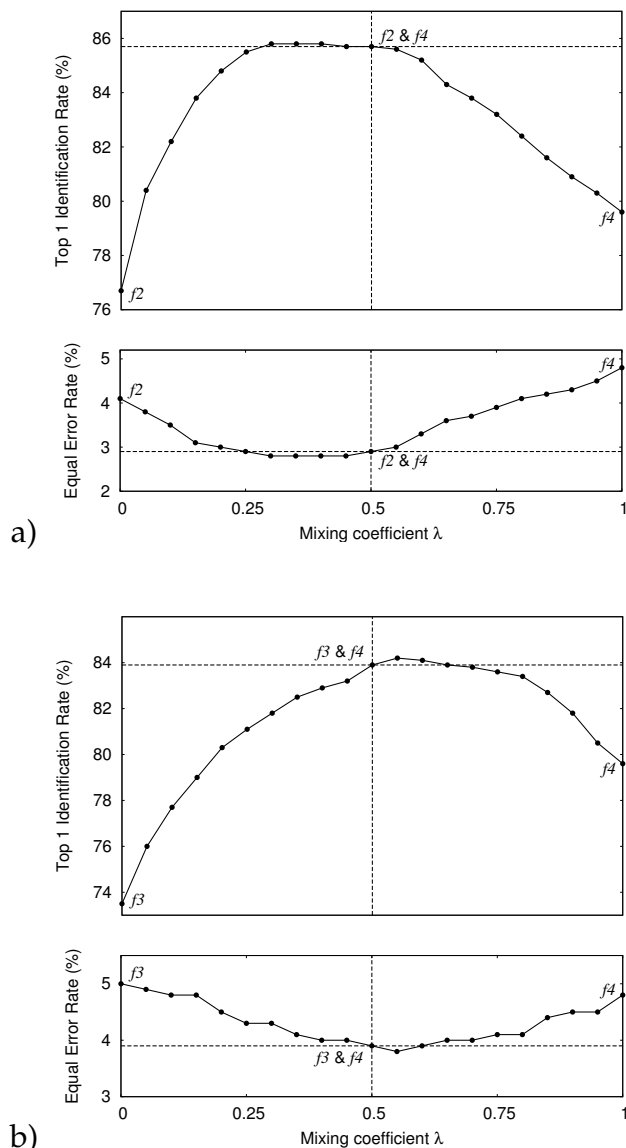


Figure 5.11: Writer identification and verification performance on the Large dataset for a weighted combination of features a) f_2 and f_4 , b) f_3 and f_4 . Only marginal improvements are obtainable over the performance levels of the simple average combination represented by the horizontal lines and corresponding to a mixing coefficient $\lambda = 0.5$.

Fig. 5.12 gives a graphical overview of the writer identification results on the Large dataset 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 tables 5.3 and 5.4. Fig. 5.13 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.

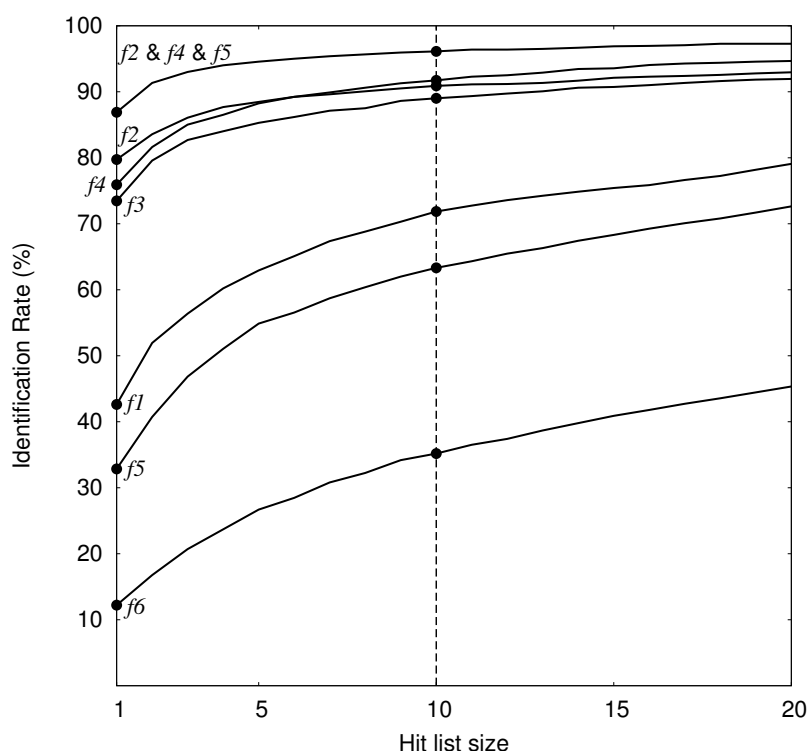


Figure 5.12: Writer identification performance as a function of hit list size. The results were obtained on the Large dataset containing 900 writers, 2 samples per writer.

5.7 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.

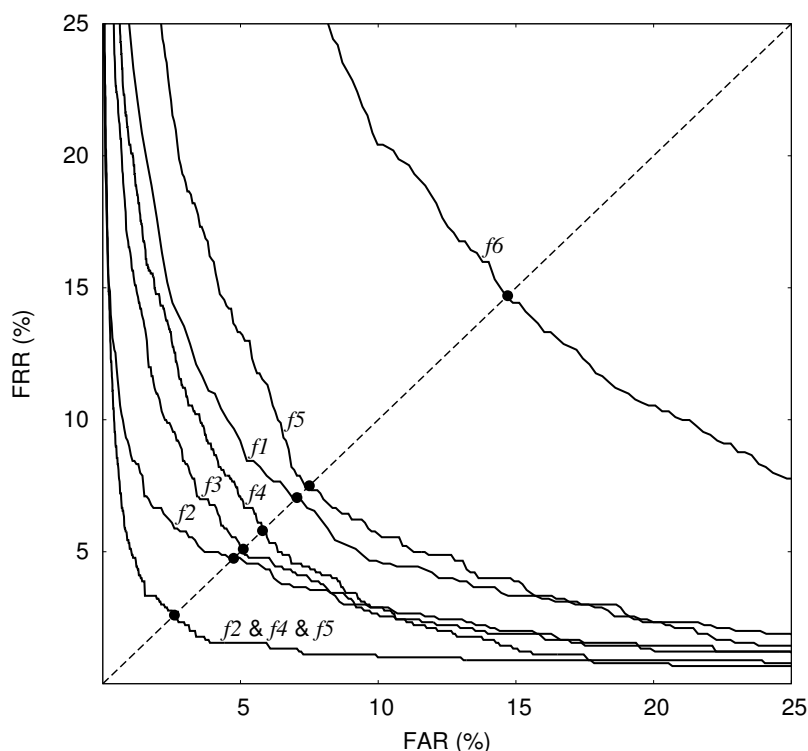


Figure 5.13: Writer verification ROC curves obtained on the Large dataset containing 900 writers, 2 samples per writer. The EER operational points lie on the dotted diagonal.

Similar to the previous chapter, we accomplished a more in-depth analysis of the performance of our allograph-level method on the Large dataset. The computation of feature f_4 depends on two important issues: the size of the shape codebook and the clustering algorithm used to generate 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. Figures 4.1, 4.2 and 4.3 show examples of shape codebooks that have been generated by the three clustering methods. Figure 5.9 shows examples of codebooks of increasing size generated by k-means clustering.

In the experiments, the number of clusters used was varied from 9 (3x3) to 2500 (50x50). A number of 200 epochs have been used for training the Kohonen SOMs. Computations have been run on a Beowulf high-performance Linux cluster with 1.7GHz / 0.5GB nodes. Training times for codebooks of size 400: k-means - 1 hrs, ksom1D - 10 hrs, ksom2D - 17 hrs. Computation times for the grapheme emission PDF on codebooks of size 400: 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 results obtained on the Large dataset confirm our previous findings from Chap-

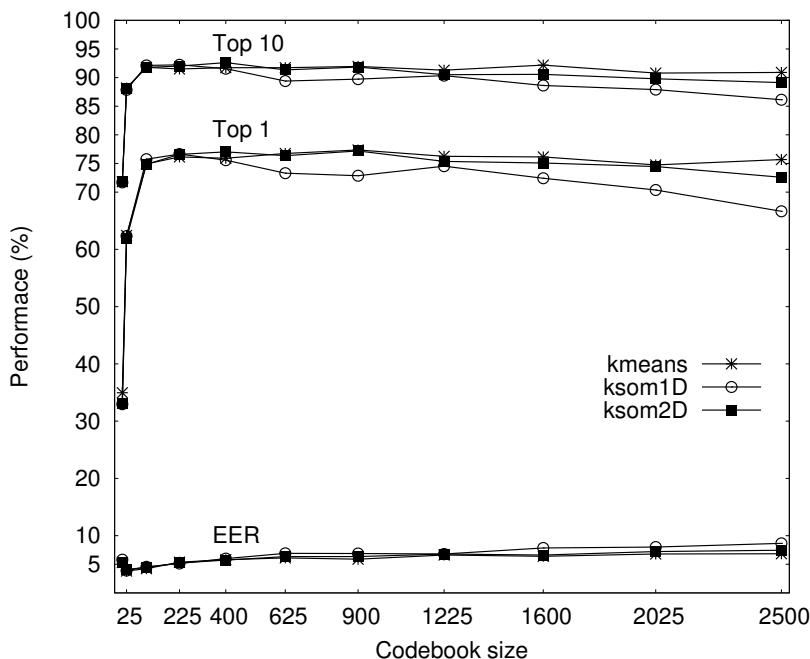


Figure 5.14: Performance vs. clustering method and codebook size for the grapheme-based writer identification and verification method (feature f_4) on the Large dataset.

ter 4. Fig. 5.14 shows that the same performance is achieved by all three clustering methods and that performance is stable over the range of codebook sizes covered in the experiments. Writer identification rates (Top-1 and Top-10) reach a plateau for codebook sizes larger than about 100 (10x10) shapes. The writer verification EER reaches a minimum of about 4% for a codebook size of 100 and increases to about 7% for larger codebooks.

These results can be explained considering that, as the codebook size increases, it contains a larger variety of shapes and therefore becomes more discriminatory between writers, with the inevitable drawback that PDF estimation becomes more difficult given the limited amount of handwriting present in our samples. As observed also previously in the experiments reported in Chapter 4, the increase in the EER is probably due to the fact that, for larger codebooks, the dimensionality of the grapheme emission PDFs increases and consequently a unique decision threshold is no longer appropriate for all the sample-to-sample distances used in writer verification. The writer verification system commits to a global decision threshold before actually being confronted with the two samples that must be compared. An individualized threshold would be required, taking into account the within-writer variability specific to the two samples being matched in a chosen writer verification trial. However, considering the limited amount of handwritten material contained in our samples, estimating this within-writer variability is a difficult problem that requires further research. It is important to observe

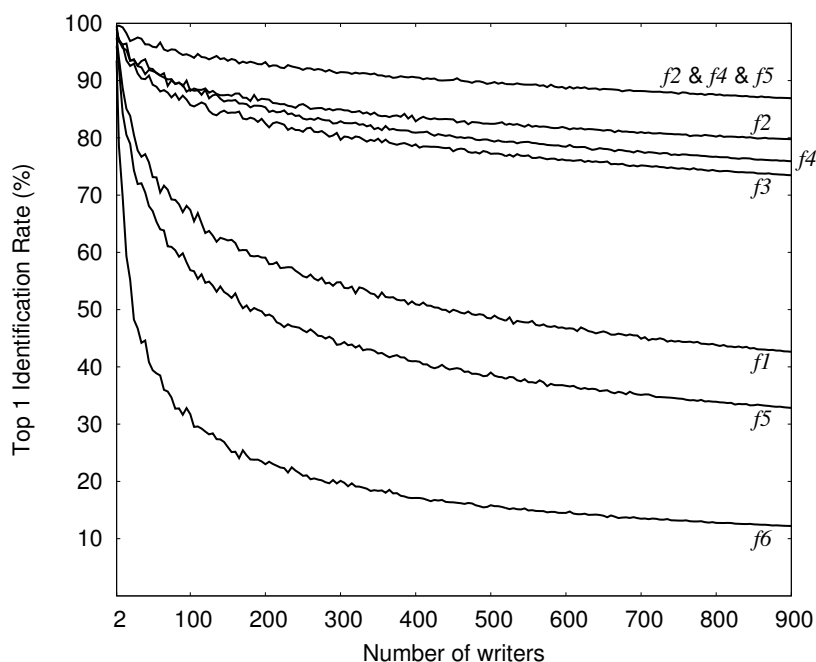


Figure 5.15: Top-1 identification rate vs. number of writers contained in the test. For every size of the writer set, the results were averaged over fifty random draws from the Large dataset. For the complete dataset (1800 samples by 900 writers), the writer identification percentages are subject to a $\pm 3\%$ confidence interval at a 95% confidence level.

that the described dimensionality problem does not significantly affect the writer identification performance because the query sample constitutes a vantage point with respect to which the distance rankings of the other samples remain essentially stable with the increase in codebook size. Similar results were found also on the other test datasets in the previous chapter (Bulacu and Schomaker 2005a).

The results reported for the grapheme-emission PDF (feature f_4) in the previous sections of the chapter 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 grapheme codebook spans up 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 equally well the task of selecting representative graphemes adequate for constructing a shape-occurrence PDF informative about the writer identity.

We can confidently conclude that the proposed allograph-level method is robust to the underlying shape representation used (whether contours or normalized bitmaps), to the size of codebook used (stable performance for sizes from 10^2 to 2.5×10^3) and to the clustering method used to generate the codebook (essentially 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 dataset. We determined this relationship by experiment using the Large dataset: for each size of the writer set (up to 900 writers), fifty identification tests were performed on random selections of writers and the results were averaged. Fig. 5.15 shows the Top-1 identification rate as a function of the number of writers for individual features and for the feature combination f_2 & f_4 & f_5 . 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 f_2 & f_4 & f_5 , we observe that the Top-1 identification rate drops by approx. 2-3% for every doubling of the number of writers in the dataset. 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 thesis 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 3 samples per writer for about a quarter of the total of 650 writers incorporated in the set. For the best performing feature combination f_2 & f_4 & f_5 , we obtained writer identification rates of Top-1 92% and Top-10 98%. These values exceed the identification rates obtained on our modified IAM set that always contained only two samples per writer (see Table 5.4).

In another study performed on a subset comprising 100 writers from the Firemaker dataset, our methods largely outperformed two actual systems used in current forensic practice (Schomaker and Bulacu 2004). 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 thesis are equally applicable to handwriting as well as machine print: writer identification vs. 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.

5.8 Conclusions

The writer identification and verification methods described in this thesis exploit two essential sources of behavioral information regarding handwriting individuality. Firstly, 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. Secondly, 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 datasets 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.

It is your work in life that is the ultimate seduction.

Pablo Picasso

6.1 Summary and Contributions

There are two fundamental dogmas underpinning handwriting identification. Their clear-cut statements are as follows:

- No two people write exactly alike.
- No one person writes exactly the same way twice.

These two principles, albeit oversimplified and disputable, unequivocally highlight the two natural factors that are in direct conflict in the attempt to identify a person based on samples of handwriting: *between-writer variation* as opposed to *within-writer variability*.

Our goal in this thesis was to automate the process of writer identification using scanned images of handwriting and thereby to provide a computer analysis of handwriting individuality. In this endeavor, a third computational factor takes center stage: the design and use of *appropriate representations*, computable features capturing the writing style of a person from the scanned handwritten samples. The power of such a representation or feature relies in its ability to maximize the separation between different writers, while remaining stable over samples produced by the same writer. We present in this thesis novel and very effective features for automatic writer identification on the basis of scanned images of handwriting. The similarity in handwriting style between any two samples is computed by using appropriate distance measures between their corresponding feature vectors. Our features and writer classification operate in the general framework of statistical pattern recognition (Duda et al. 2001, Jain et al. 2000).

Two fundamental sources of information regarding the individuality of handwriting are exploited by our methods functioning at two levels of analysis. First, handwriting slant, curvature and roundness, as determined by habitual pen grip, are captured by

joint directional probability distributions operating at the texture level. Second, the personalized set of letter shapes, called *allographs*, that a writer has learned to use under educational, cultural and memetic influences is captured by a grapheme-emission probability distribution operating at the character level. Combining texture-level and allograph-level features provides a very intimate and comprehensive characterization of the individual handwriting style of a person. Our methods achieved very high writer identification and verification performance in extensive tests carried out using large datasets with handwriting samples collected from up to 900 subjects.

In our methods, writer individuality is robustly encoded using probability distribution functions extracted from handwritten text blocks. 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 samples. In our methods the computer is completely agnostic of the actual text written in the samples. The handwriting is merely seen as a texture characterized by some directional probability distributions or as a simple stochastic shape-emission process characterized by a grapheme occurrence probability distribution. Our techniques have practical feasibility and hold the potential of concrete use in real applications.

Chapter 1 of the thesis introduces writer identification as a behavioral biometric modality and presents the fundamental genetic and cultural factors causing the individuality of handwriting. The task of *writer identification* is equivalent to answering the question: "Who wrote this sample?" A writer identification system performs a one-to-many search in a large database with samples of known authorship and returns a likely list of candidates containing the handwritings most similar to the questioned one. The hit list is further scrutinized by a human expert. The task of *writer verification* is equivalent to answering the question: "Were these two samples written by the same person?" A one-to-one comparison is performed and an automatic yes / no decision is taken. In the introductory chapter, a connection is also drawn between writer identification and the related, but much broader, field of handwriting recognition. In handwriting recognition, the variations between different handwritings must be eliminated to obtain invariance and generalization. In writer identification, on the contrary, these same variations must be enhanced to obtain writer specificity and discrimination. Further in Chapter 1, a survey of recent publications in the field makes clear the distinction between text-dependent versus text-independent approaches and provides the necessary context in which to place our own research work.

The thesis then shows the progression of our writer identification research from low level textural features to higher level allographic features. The thesis is divided into two main parts. Chapter 2 and Chapter 3 describe our texture-level approach. Chapter 4 and Chapter 5 present our allograph-level approach and the fusion method used to combine

textural and allographic features for improved writer identification performance.

Chapter 2 shows that using the orientation of short fragments of edges along the written trace provides the basis for building several directional probability distributions that are very effective features for writer identification. The first angular feature constructed using oriented edge fragments is the edge-direction distribution, a classically known descriptor for writer identification. The mode of this distribution, i.e. the dominant direction in the script, corresponds to the slant of handwriting, which is a stable personal trait and a discriminatory characteristic between different writers. We propose further a new and potent method that considers the angle combinations of two "hinged" edge fragments and builds a joint directional probability distribution that simultaneously encodes both orientation and curvature information. This novel "edge-hinge" feature is a bivariate probability function that delivers a very significant improvement in writer identification and performance over the simple edge-angle distribution. The edge-based directional distributions, as a group of related features, outperform a number of non-angular features (run-length distributions, autocorrelation, entropy). Reducing the amount of ink in the test samples leads to an overall decrease in performance for all features, but the performance standings of the different features with respect to each other remain the same.

Chapter 3 carries on the idea of using the directionality of the script as an effective source of information for text-independent writer identification. And another new and strong feature is designed that considers the edge-angle combinations co-occurring at the extremities of run-lengths. Further performance improvements are obtained by incorporating also location information into the basic features. This is achieved by extracting two probability distributions separately from the top and bottom halves of text lines and then adjoining the two feature vectors. The asymmetry between top and bottom distributions provides extra information regarding writer identity. The experimental study is performed as a comparison between lowercase and uppercase handwriting on test samples containing controlled amounts of ink. We obtain similar writer identification performance for lowercase and uppercase handwriting for the battery of features considered in the analysis.

Chapter 4 introduces our allograph-level method for writer identification and verification. This theoretically founded approach assumes that each writer is characterized by the occurrence probability of elementary shapes from a common shape codebook. These elementary shapes, or *graphemes*, are obtained by applying a heuristic segmentation procedure on the written ink. The common shape codebook is generated by clustering the set of graphemes extracted from the handwritings of a sufficiently large number of writers, kept separate from those used in identification and verification tests. The graphemes resulting from handwriting segmentation may, but usually will not, overlap a complete character. This is a fundamental problem for handwriting recognition.

Nevertheless, the ensemble of these sub- or supra-allographic shapes is very descriptive about the identity of the writer who generated them. And therefore is very effective in writer identification. In large scale computational experiments, we compare three clustering algorithms used for generating the common grapheme codebook: k-means, Kohonen Self-Organizing Maps 1D and 2D. The results prove the robustness of the proposed allograph-level writer identification method: similar good performance is obtained for all three clustering algorithms over a large range of codebook sizes.

Chapter 5 performs an extensive analysis of feature combinations. It is natural to try to combine the proposed features for improving the performance and robustness of our writer identification and verification system: while not totally orthogonal, the different features do offer different points of view on a handwritten sample and operate at different levels of analysis and also at different scales. In our fusion scheme, the final unique distance between two handwritten samples is computed as the average of the distances due to the individual features participating in the combination. In this chapter, more efficient algorithms are proposed for computing the directional features using contours, rather than edges. The functioning of the considered features is also put in an overall Fourier perspective that better explains also their relative performance merits. The evaluation experiments are extended to bigger datasets. The largest dataset comprises 900 writers and is comparable in size to the largest dataset used in writer identification studies until the present. The experimental results, consistent across the different test datasets, show that fusing multiple features yields increased writer identification and verification performance. The best performing feature combinations fuse directional, grapheme and run-length information yielding, on the large dataset containing 900 subjects, writer identification rates of Top-1 85-87% and Top-10 96% with an error rate around 3% in verification.

Chapter 6 concludes the thesis and Appendix A presents an HTML-based visualization tool developed with the purpose of visually assessing our writer identification and verification system called GRAWIS, an acronym from Groningen Automatic Writer Identification System.

The present thesis analyzes in depth the algorithmic aspects of automatic writer identification and verification. The proposed text-independent methods have possible impact in forensic science: they allow the search in a large dataset with handwritten samples with the retrieval of only those documents that pictorially look similar to the query in terms of handwriting style. In this way, the hit list containing the likely candidates is reduced to a size than can be analyzed in detail by the forensic expert to finally establish the writer identity for the questioned document. Part of the texture-level methods described in this thesis have already been used in a concrete industrial setting. Nevertheless, the wider application beyond the realm of academic research of our writer identification and verification techniques still remains a challenge for the future.

6.2 Further research directions

Considered in the general context of biometrics, automatic writer identification and verification is presently a thriving research topic. It is also a very engaging one. Here we sketch a number of further research ideas.

The writer identification methods presented in this thesis require the separation of the ink from the background of the document (image binarization). They also require the separation of handwriting from other graphical objects that might be present also in the document (layout analysis). Our academic datasets did not exhibit the full range of problems that must be solved in a complete document analysis and recognition system. A more extended examination is therefore needed of the document processing steps preceding feature extraction for writer identification and verification.

It is important to observe that the full variability of a person's handwriting (*within-writer variability*) is not completely exposed in our datasets. For example, the long term changes occurring over years in the handwriting of an individual would require longitudinal studies.

While the writer identification and verification techniques presented in this thesis make extensive use of probabilities, our approach is not manifestly Bayesian. Nevertheless, our methods can be cast into a Bayesian framework and a more extensive analysis along this line is needed. Regarding the adoption of a Bayesian approach to writer identification and verification in the forensic application domain, a word of caution must be said about using prior probabilities: a maximum likelihood (ML) solution that ignores priors and rests on the shape evidence alone might be preferable to a full maximum a posteriori (MAP) solution. Weighting the shape evidence results with priors is considered to take place outside the scope of the current research. In this thesis, within the context of feature combinations presented in Chapter 5, we discussed the underlying Bayesian feature fusion model. Throughout most of our work however, we used vectorial representations and distances, rather than probability multiplication. During our research, the methods were developed using explanations and encodings that were close to the actual physical meaning of the features and the intuitive interpretation of the information they convey about the specific handwriting style of an individual (Bulacu and Schomaker 2007).

The experimental studies presented in this thesis were performed on Western script. Considering that our techniques are fairly generic and text-independent, their applicability to other scripts, e.g. Chinese, Arabic, Indian, is a pertinent and interesting research question.

In this thesis, we have designed and evaluated a number of writer identification features belonging to the category of *fully automatic features* computed from a region of interest (i.e. a handwritten text block) in the image. In forensic praxis, two other cat-

egories of features derived from scanned samples are additionally used: *interactively measured features* by human experts using a dedicated graphical user-interface tool and *character-based features* related to the allograph subset that is being generated by each writer and requiring human work to isolate and label individual handwritten characters. Further exploration is required of the text-dependent methods, applicable for samples containing very limited amounts of handwriting, where probability estimation becomes unreliable. A performance comparison between the automatic features and the features requiring human involvement is still a fundamental open problem. This represents, in fact, the main topic of a new project, called *Trigraph* and financed by NWO, that will continue our research in the area of writer identification and verification.

Automatic writer identification can be applied to historical documents (Schomaker et al. 2007). While recognizing the actual text content of the documents is clearly more worthwhile, the identification of the writer can nevertheless have a degree of relevance in historical studies of paleography and codicology. Our methods are equally applicable to handwriting and machine print, writer identification versus font identification. Automatic script identification on historical documents would, in principle, open a number of interesting possibilities. It would allow to identify the scribe in case of handwritten documents or identify the printing house in case of machine printed documents. This would allow for automatic manuscript dating and/or authentication. Also, manuscript indexing and retrieval based on script style (graphical, rather than textual content) would become possible. Different types of calligraphy with their corresponding historical period could also be identified in a collection of documents. Because it provides some form of content enrichment, automatic script identification on historical documents might become a useful tool for the historian. This topic, placed at the confluence between computer science and humanities, can be a rewarding future research direction.

A modified version of this appendix was published as: Marius Bulacu, Lambert Schomaker, – “GRAWIS: Groningen Automatic Writer Identification System,” Proc. of 17th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2005), 2005, pp. 413-414, 17-18 October, Brussels, Belgium

Appendix A

GRAWIS: Groningen Automatic Writer Identification System

Teach yourself programming in 10 years.

Peter Norvig

Abstract

In this appendix, we present an HTML-based visualization tool that we built in order to be able to directly see and assess the results generated by our writer identification and verification software.

A.1 Introduction

Research in writer identification and verification received renewed interest in recent years, stimulated also by 9/11 and the anthrax letters. Our work has been performed in the framework of the *Wanda* project financed by Fraunhofer Institute, Berlin. In the future, the *Trigraph* project, financed by NWO, will continue and extend the research on this topic.

We use *textural* and *allographic* features to characterize individual handwriting style independently of the textual (ASCII) content of the handwritten samples (Bulacu and Schomaker 2004). Our experimental evaluations were performed on several large datasets with consistently good results. In *writer identification* searches, all the samples in the test dataset are ordered with increasing distance from the query sample. In *writer verification* trials, if the distance between two chosen samples is smaller than a predefined threshold, the samples are deemed to have been written by the same person. Otherwise, the samples are considered to have been written by different writers.

A.2 Visualization tool

Besides the empirically measured accuracy percentages or error rates on the experimental datasets, a subjective evaluation of the performance is always an important

component in the development and the final assessment of any pattern recognition system. We built an HTML-based visualization tool to directly see the results generated by our writer identification and verification system, dubbed GRAWIS for Groningen Automatic Writer Identification System (Bulacu and Schomaker 2005b). After feature extraction and feature matching, our programs generate HTML files containing numerical results (distances, ranks, identity codes, thresholds) and hyperlinks to the scanned samples. A web browser can then be used to visualize these HTML files. This HTML-based approach allows a quick development of the visualization tool without the considerable programming effort needed to construct a complete graphical user interface.

A.3 Examples of writer identification hit lists

Figures A.1, A.2, A.3, A.4 and A.5 show five examples of hit lists generated by our system. For a chosen query sample, writer identification searches can be 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 a very handy navigation in the space of individual handwriting styles. Fig. A.1 shows a hit list generated by feature f_5 (the combination of horizontal and vertical run-length PDFs - see Table 5.2) applied on the Large dataset (900 writers, 2 samples per writer, lowercase handwriting - see Table 5.1). 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. A.2 shows a successful hit list generated by the best performing feature combination f_2 & f_4 & f_5 (contour-hinge PDF & grapheme-emission PDF & run-length PDFs) for the same query sample. The correct hit is now in position 1 and the handwriting style is homogeneous across the hit list. Fig. A.3 shows another successful hit list generated by f_2 & f_4 & f_5 for a different query. Figures A.4 and A.5 show hit lists generated by the contour-hinge feature f_2 applied on the Fire-maker dataset (250 writers, 2 samples per writer, lowercase and uppercase handwriting respectively).

A.4 Examples of writer verification errors

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. A.6 shows two examples of false reject errors. Fig. A.7 shows two examples of 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 Bayes error rate in writer verification.

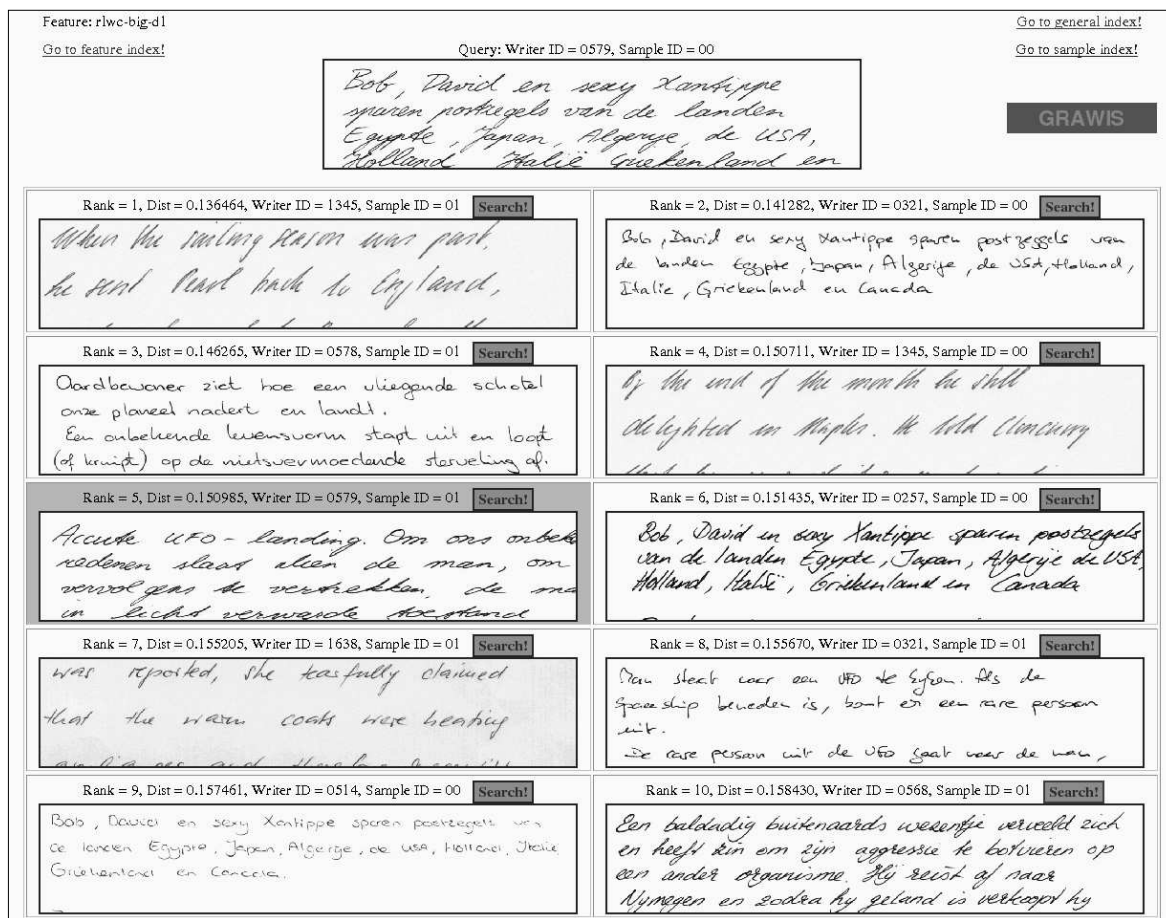


Figure A.1: Writer identification hit list generated by the moderately performing feature f_5 (combined horizontal and vertical run-length PDFs) on the Large dataset (900 writers, 2 samples per writer, lowercase handwriting). The query sample is in the top-center position and the correct hit is on rank 5 (marked with a darker frame). The handwriting style is heterogeneous across the hit list.

Feature: hinge-cocopdf-rlwc-big-d1 [Go to general index!](#)
 Go to feature index! [Go to sample index!](#)

Query: Writer ID = 0579, Sample ID = 00

Bob, David en Gary Kantsjpe
 sparen postzegels van de landen
 Egypte, Japan, Algerije, de USA,
 Holland, Italië, Griekenland en

GRAWIS

<p>Rank = 1, Dist = 0.338160, Writer ID = 0579, Sample ID = 01 Search!</p> <div style="border: 1px solid black; padding: 2px;"> <p>Accuse UFO-landing. Om ons onbekende redenen staat allen de man, om wereld was te verstoren, de ma in licht verwarde te stand</p> </div>	<p>Rank = 2, Dist = 0.367954, Writer ID = 1639, Sample ID = 00 Search!</p> <div style="border: 1px solid black; padding: 2px;"> <p>Between all the interested parties, the final agreement had been delayed long enough, for this 1900 miles to cover all the money which</p> </div>
<p>Rank = 3, Dist = 0.373618, Writer ID = 1639, Sample ID = 01 Search!</p> <div style="border: 1px solid black; padding: 2px;"> <p>Discouraging one churchilla jacket, a Robt-Royce and a Essex minor house, all</p> </div>	<p>Rank = 4, Dist = 0.422287, Writer ID = 1475, Sample ID = 01 Search!</p> <div style="border: 1px solid black; padding: 2px;"> <p>him. Nick was certain in his own mind that he had really seen Weston's hand</p> </div>
<p>Rank = 5, Dist = 0.430005, Writer ID = 1475, Sample ID = 00 Search!</p> <div style="border: 1px solid black; padding: 2px;"> <p>"It certainly was, Sam, Nick would agree, and go on to say with the best of self-possession: "No</p> </div>	<p>Rank = 6, Dist = 0.440450, Writer ID = 1592, Sample ID = 00 Search!</p> <div style="border: 1px solid black; padding: 2px;"> <p>"Yes. I travelled down from London as the dawn was breaking, and</p> </div>
<p>Rank = 7, Dist = 0.447057, Writer ID = 0557, Sample ID = 00 Search!</p> <div style="border: 1px solid black; padding: 2px;"> <p>Bob, David en Gary Kantsjpe sparen postzegels van de landen Egypte, Japan, Algerije, de USA, Holland, Italië, Griekenland en Canada.</p> </div>	<p>Rank = 8, Dist = 0.453766, Writer ID = 0257, Sample ID = 00 Search!</p> <div style="border: 1px solid black; padding: 2px;"> <p>Bob, David en Gary Kantsjpe sparen postzegels van de landen Egypte, Japan, Algerije de USA, Holland, Italië, Griekenland en Canada</p> </div>
<p>Rank = 9, Dist = 0.459697, Writer ID = 1598, Sample ID = 01 Search!</p> <div style="border: 1px solid black; padding: 2px;"> <p>The others were all dancing and as Jay drew out a chair and sat down Gavin withdrew a wood sword</p> </div>	<p>Rank = 10, Dist = 0.466966, Writer ID = 1592, Sample ID = 01 Search!</p> <div style="border: 1px solid black; padding: 2px;"> <p>never been inside. Long since of me," said Mr. Copphome. The eyes had the same in Flackin and</p> </div>

Figure A.2: A successful writer identification search using the best performing feature combination f_2 & f_4 & f_5 (contour-hinge PDF & grapheme-emission PDF & run-length PDFs) for the same query sample as in Fig. A.1. 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.

Feature: hinge-cocopdf-rlwc-big-d1 [Go to general index!](#)
 Go to feature index! [Go to sample index!](#)

Query: Writer ID = 1299, Sample ID = 00

None of these classes, except the yeomen, paid or expected to pay direct taxes. During the four-
 month conference the traditional methods of

DRAWIS

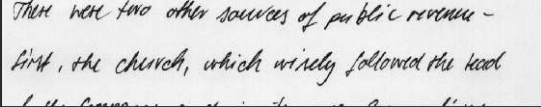
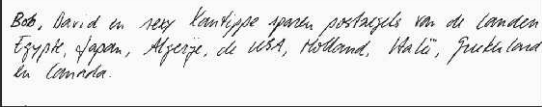
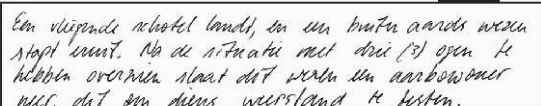
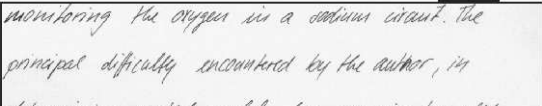
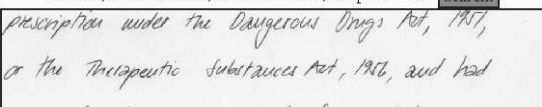
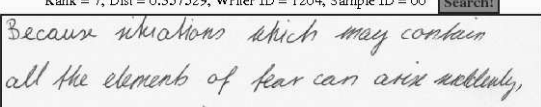
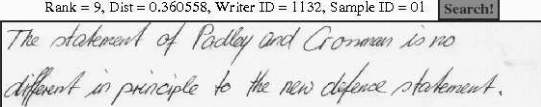
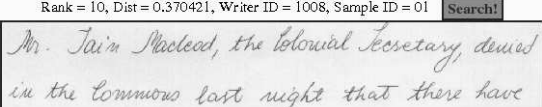
Rank = 1, Dist = 0.231317, Writer ID = 1299, Sample ID = 01 Search! 	Rank = 2, Dist = 0.326768, Writer ID = 0583, Sample ID = 00 Search! 
Rank = 3, Dist = 0.328085, Writer ID = 0583, Sample ID = 01 Search! 	Rank = 4, Dist = 0.333898, Writer ID = 1398, Sample ID = 01 Search! 
Rank = 5, Dist = 0.348692, Writer ID = 1620, Sample ID = 00 Search! 	Rank = 6, Dist = 0.354522, Writer ID = 1365, Sample ID = 01 Search! 
Rank = 7, Dist = 0.357529, Writer ID = 1204, Sample ID = 00 Search! 	Rank = 8, Dist = 0.357642, Writer ID = 1620, Sample ID = 01 Search!
Rank = 9, Dist = 0.360558, Writer ID = 1132, Sample ID = 01 Search! 	Rank = 10, Dist = 0.370421, Writer ID = 1008, Sample ID = 01 Search! 

Figure A.3: Another successful writer identification search on the Large dataset using the best performing feature combination f2 & f4 & f5 for a different query sample.

Feature: hinge-firemaker-14-ndirs-016-d2 [Go to general index!](#)

[Go to feature index!](#) Query: Writer ID = 0397, Sample ID = 01 [Go to sample index!](#)

Op het eerste plaats zie ik een vliegende schotel die een man naderd. Er stapte een ~~but~~ buitenaards wezen uit. Vervolgens geeft het de man een ram op het gezicht. De man

GRAWIS

<p>Rank = 1, Dist = 0.009506, Writer ID = 0397, Sample ID = 00 Search!</p> <p>Bob, David en sexy Xantippe sparen postzegels van de landen Egypte, Japan, Algerije, de USA, Holland, Italië, Griekenland en Canada.</p>	<p>Rank = 2, Dist = 0.018781, Writer ID = 0561, Sample ID = 00 Search!</p> <p>Bob, David en Sexy Xantippe sparen postzegels van de landen Egypte, Japan, Algerije, de USA, Holland, Italië, Griekenland en Canada.</p>
<p>Rank = 3, Dist = 0.018988, Writer ID = 0558, Sample ID = 00 Search!</p> <p>Bob, David en sexy Xantippe sparen postzegels van de landen Egypte, Japan, Algerije, de USA, Holland, Italië, Griekenland en Canada.</p>	<p>Rank = 4, Dist = 0.022439, Writer ID = 0558, Sample ID = 01 Search!</p> <p>Een man ziet een vliegende schotel landen. Er komt een alien uit die hem neerstaat. Het wezen heeft 3 ogen, geen neus, maar wel een mond. Verder ziet hij eruit als een soort feesthoed.</p>
<p>Rank = 5, Dist = 0.027279, Writer ID = 0432, Sample ID = 01 Search!</p> <p>Een man staat te kijken naar de ruimtevaartuig dat gaat landen. Nadat het geland is, komt er een vreemd 3 oogig wezentje uit. Het wezentje slaat het monnetje in zijn gezicht! De man is neergeslagen.</p>	<p>Rank = 6, Dist = 0.029763, Writer ID = 0527, Sample ID = 00 Search!</p> <p>Bob, David en sexy Xantippe sparen postzegels van de landen Egypte, Japan, Algerije, de USA, Holland, Italië, Griekenland en Canada.</p>
<p>Rank = 7, Dist = 0.030103, Writer ID = 0561, Sample ID = 01 Search!</p> <p>Een dikke man met pet kijkt naar de landing van een UFO. Wanneer de UFO geland is komt er een vreemd wezen uitgekropen.</p>	<p>Rank = 8, Dist = 0.031255, Writer ID = 0477, Sample ID = 01 Search!</p> <p>Op een mooie dag stond Kees in een weiland waar plotseling een UFO voor z'n neus landde. Verbaasd bleef hij staan en de klep van de UFO</p>
<p>Rank = 9, Dist = 0.032751, Writer ID = 0494, Sample ID = 00 Search!</p> <p>Bob, David en sexy Xantippe sparen postzegels van de landen Egypte, Japan, Algerije, de USA, Holland, Italië, Griekenland en Canada.</p>	<p>Rank = 10, Dist = 0.033579, Writer ID = 0417, Sample ID = 00 Search!</p> <p>Bob, David en sexy Xantippe sparen postzegels van de landen Egypte, Japan, Algerije, de USA, Holland, Italië, Griekenland en Canada.</p>

Figure A.4: A successful writer identification search using the best performing individual feature f_2 (contour-hinge PDF) on the Firemaker lowercase dataset (250 writers, 2 samples per writer). The correct hit is in the first position and the handwriting style is uniform across the hit list.

Feature: hinge-firemaker-2-ndirs-012-d2 [Go to general index!](#)
[Go to feature index!](#) Query: Writer ID = 0277, Sample ID = 02 [Go to sample index!](#)

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.

GRAWIS

<p>Rank = 1, Dist = 0.025531, Writer ID = 0277, Sample ID = 03 Search!</p> <div style="border: 1px solid black; padding: 5px;"> <p>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 PARKEERTERRAIN.</p> </div>	<p>Rank = 2, Dist = 0.031390, Writer ID = 0432, Sample ID = 02 Search!</p> <div style="border: 1px solid black; padding: 5px;"> <p>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.</p> </div>
<p>Rank = 3, Dist = 0.031676, Writer ID = 0397, Sample ID = 03 Search!</p> <div style="border: 1px solid black; padding: 5px;"> <p>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</p> </div>	<p>Rank = 4, Dist = 0.032940, Writer ID = 0397, Sample ID = 02 Search!</p> <div style="border: 1px solid black; padding: 5px;"> <p>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.</p> </div>
<p>Rank = 5, Dist = 0.034000, Writer ID = 0477, Sample ID = 03 Search!</p> <div style="border: 1px solid black; padding: 5px;"> <p>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 PARKEERTERRAIN.</p> </div>	<p>Rank = 6, Dist = 0.034221, Writer ID = 0506, Sample ID = 02 Search!</p> <div style="border: 1px solid black; padding: 5px;"> <p>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.</p> </div>
<p>Rank = 7, Dist = 0.037437, Writer ID = 0506, Sample ID = 03 Search!</p> <div style="border: 1px solid black; padding: 5px;"> <p>ZE KWAMEN IN 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</p> </div>	<p>Rank = 8, Dist = 0.037895, Writer ID = 0432, Sample ID = 03 Search!</p> <div style="border: 1px solid black; padding: 5px;"> <p>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 PARKEERTERRAIN.</p> </div>
<p>Rank = 9, Dist = 0.041108, Writer ID = 0551, Sample ID = 02 Search!</p> <div style="border: 1px solid black; padding: 5px;"> <p>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.</p> </div>	<p>Rank = 10, Dist = 0.041615, Writer ID = 0405, Sample ID = 02 Search!</p> <div style="border: 1px solid black; padding: 5px;"> <p>NADAT ZE IN NEW YORK, TOKYO, QUÉBEC, PARIJS, ZÜRICH EN OSLO WAREN GEWEEST, VLOGEN ZE UIT DE USA TERUG MET VLUCHT KL 658 OM</p> </div>

Figure A.5: A successful writer identification search using the best performing individual feature f_2 on the Firemaker uppercase dataset.

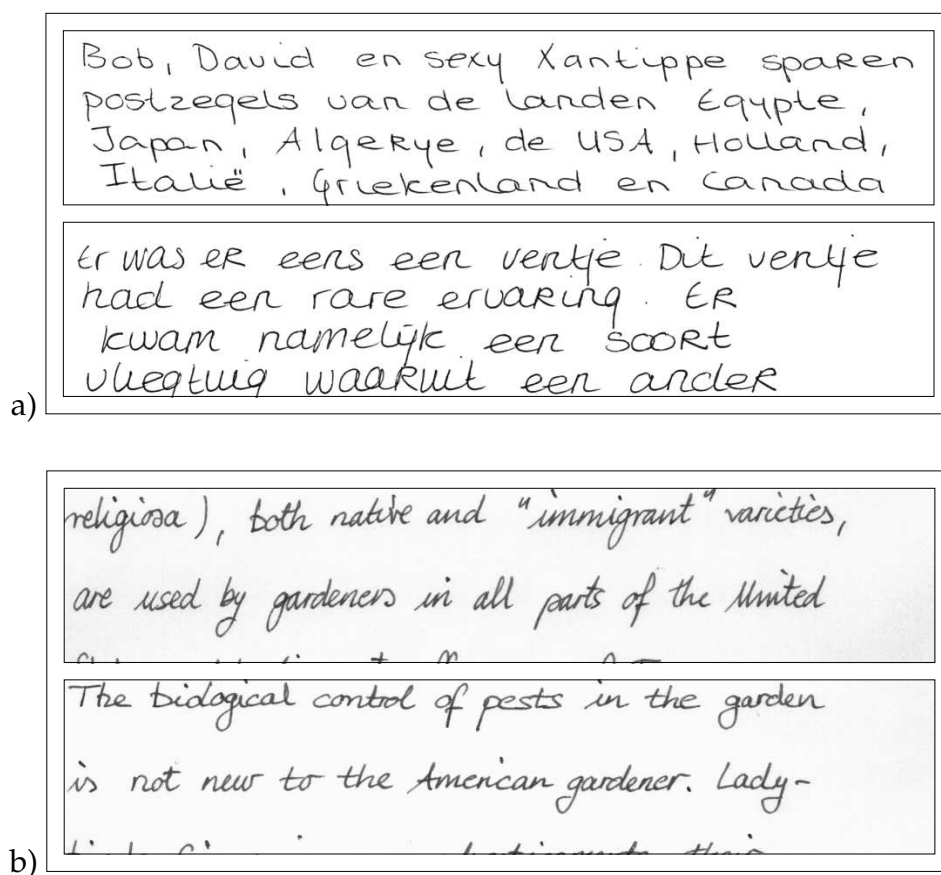


Figure A.6: Two examples - a) and b) - of false reject errors in writer verification: the two samples were written by the same person, but the system wrongly decided the opposite.

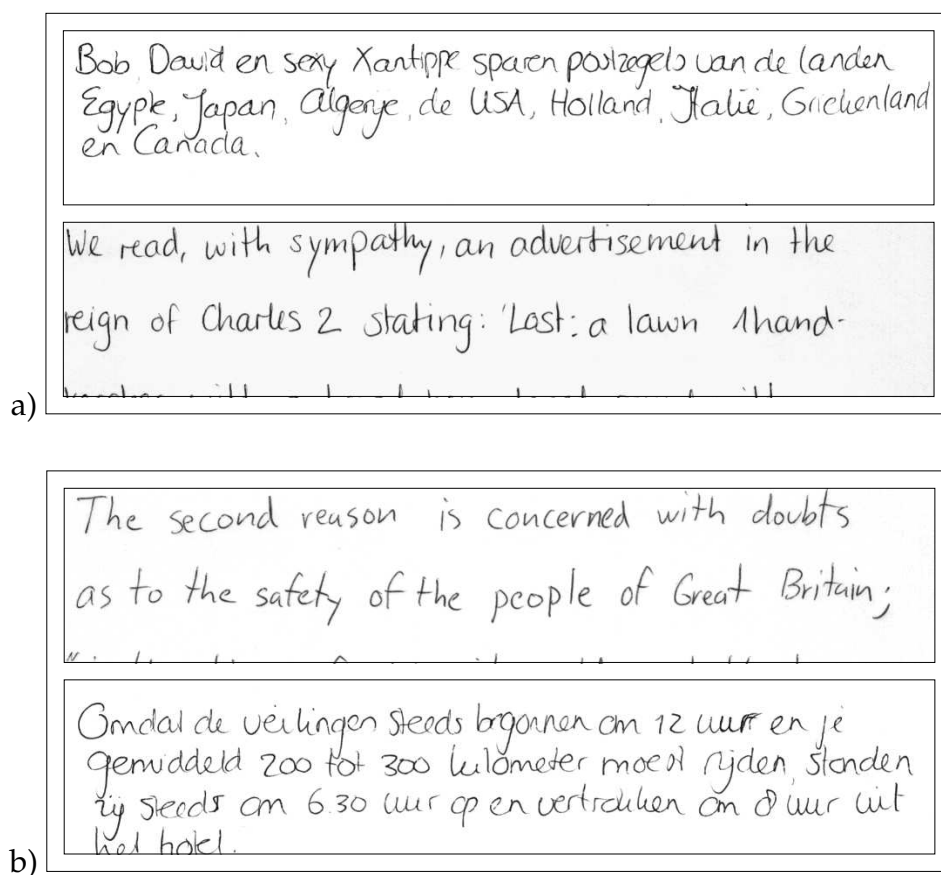


Figure A.7: Two examples - a) and b) - of false accept errors in writer verification: the two samples were written by different persons, but the system wrongly decided the opposite.

Bibliography

- Arazi, B.: 1977, Handwriting identification by means of run-length measurements, *IEEE Trans. Syst., Man and Cybernetics* **SMC-7**(12), 878–881.
- Arazi, B.: 1983, Automatic handwriting identification based on the external properties of the samples, *IEEE Trans. Syst., Man and Cybernetics* **SMC-13**(4), 635–642.
- Benecke, M.: 1997, DNA typing in forensic medicine and in criminal investigations: A current survey, *Naturwissenschaften* **84**(5), 181–188.
- Bensefia, A., Nosary, A., Paquet, T. and Heutte, L.: 2002, Writer identification by writer's invariants, *Proc. of 8th IWFHR*, Niagara-on-the-Lake, Canada, pp. 274–279.
- Bensefia, A., Paquet, T. and Heutte, L.: 2003, Information retrieval based writer identification, *Proc. of 7th ICDAR*, Vol. II, Edinburgh, Scotland, pp. 946–950.
- Bensefia, A., Paquet, T. and Heutte, L.: 2004, Handwriting analysis for writer verification, *Proc. of 9th IWFHR*, Tokyo, Japan, pp. 196–201.
- Bensefia, A., Paquet, T. and Heutte, L.: 2005a, Handwritten document analysis for automatic writer recognition, *Electronic Letters on Computer Vision and Image Analysis* **5**(2), 72–86.
- Bensefia, A., Paquet, T. and Heutte, L.: 2005b, A writer identification and verification system, *Pattern Recognition Letters* **26**(10), 2080–2092.
- Bulacu, M. and Schomaker, L.: 2003, Writer style from oriented edge fragments, *Proc. of 10th Int. Conf. on Computer Analysis of Images and Patterns (CAIP 2003): LNCS 2756*, Springer, Groningen, The Netherlands, pp. 460–469.
- Bulacu, M. and Schomaker, L.: 2004, Analysis of texture and connected-component contours for the automatic identification of writers, *Proc. of 16th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2004)*, Groningen, The Netherlands, pp. 371–372.

- Bulacu, M. and Schomaker, L.: 2005a, A comparison of clustering methods for writer identification and verification, *Proc. of 8th Int. Conf. on Document Analysis and Recognition (ICDAR 2005)*, Vol. II, IEEE Computer Society, Seoul, Korea, pp. 1275–1279.
- Bulacu, M. and Schomaker, L.: 2005b, GRAWIS: Groningen Automatic Writer Identification System, *Proc. of 17th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2005)*, Brussels, Belgium, pp. 413–414.
- Bulacu, M. and Schomaker, L.: 2005c, Text-pose estimation in 3D using edge-direction distributions, *Proc. of Int. Conf. on Image Analysis and Recognition (ICIAR 2005): LNCS 3656*, Springer, Toronto, Canada, pp. 625–634.
- Bulacu, M. and Schomaker, L.: 2006, Combining multiple features for text-independent writer identification and verification, *Proc. of 10th International Workshop on Frontiers in Handwriting Recognition (IWFHR 2006)*, La Baule, France, pp. 281–286.
- Bulacu, M. and Schomaker, L.: 2007, Text-independent writer identification and verification using textural and allographic features, *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), Special Issue - Biometrics: Progress and Directions* **29**(4), 701–717.
- Bulacu, M., Schomaker, L. and Vuurpijl, L.: 2003, Writer identification using edge-based directional features, *Proc. of 7th Int. Conf. on Document Analysis and Recognition (ICDAR 2003)*, Vol. II, IEEE Computer Society, Edinburgh, Scotland, pp. 937–941.
- Bunke, H., Roth, M. and Schukat-Talamazzini, E.: 1995, Off-line cursive handwriting recognition using hidden Markov models, *Pattern Recognition* **28**(9), 1399–1413.
- Burges, C. J.: 1998, A tutorial on support vector machines for pattern recognition, *Knowledge Discovery and Data Mining* **2**(2), 121–167.
- Cover, T. M. and Hart, P. E.: 1967, Nearest neighbor pattern classification, *IEEE Transactions on Information Theory* **IT-13**(1), 21–27.
- Crettez, J.-P.: 1995, A set of handwriting families: style recognition, *Proc. of the 3rd International Conference on Document Analysis and Recognition*, Montreal, Canada, pp. 489–494.
- Cristianini, N. and Shawe-Taylor, J.: 2000, *An introduction to support vector machines*, Cambridge University Press, Cambridge, UK.
- Daugman, J. G.: 1993, High confidence visual recognition of persons by a test of statistical independence, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15**(11), 1148–1160.
- Daugman, J. G.: 2000, Biometric decision landscapes, *Technical report*, University of Cambridge: Computer Laboratory TR482.
- Daugman, J. G.: 2003, The importance of being random: statistical principles of iris recognition, *Pattern Recognition* **36**(2), 279–291.

- de Jong, W., Kroon van der Kooij, L. and Schmidt, D.: 1994, Computer aided analysis of handwriting, the NIFO-TNO approach, *Proc. of the 4th European Handwriting Conference for Police and Government Handwriting Experts*.
- Devlin, B., Risch, N. and Roeder, K.: 1992, Forensic inference from DNA fingerprints, *Journal of the American Statistical Association* **87**(418), 337–350.
- Dinstein, I. and Shapira, Y.: 1982, Ancient hebraic handwriting identification with run-length histograms, *IEEE Trans. Syst., Man and Cybernetics* **SMC-12**(3), 405–409.
- Doermann, D. and Rosenfeld, A.: 1992, Recovery of temporal information from static images of handwriting, *Proc. of CVPR92*, pp. 162–168.
- Dooijes, E.: 1983, Analysis of handwriting movements, *Acta Psychologica* **54**, 99–114.
- Drouhard, J., Sabourin, R. and Godbout, M.: 1995, A comparative study of the k nearest neighbours, threshold and neural network classifiers for handwritten signature verification using an enhanced directional pdf, *Proc. of the 3rd International Conference on Document Analysis and Recognition*, Montreal, Canada, pp. 807–810.
- Duda, R. O., Hart, P. E. and Stork, D. G.: 2001, *Pattern Classification*, second edn, Wiley Interscience, New York.
- El-Yacoubi, A., Gilloux, M., Sabourin, R. and Suen, C.: 1999, An HMM-based approach for off-line unconstrained handwritten word modeling and recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21**(8), 752–760.
- Fairhurst, M.: 2003, Document identity, authentication and ownership: the future of biometric verification, *Proc. of 7th Int. Conf. on Document Analysis and Recognition (ICDAR 2003)*, Vol. II, IEEE Computer Society, Edinburgh, Scotland, pp. 1108–1116.
- Favata, J. and Srikantan, G.: 1996, A multiple feature/resolution approach to handprinted digit and character recognition, *International Journal of Imaging Systems and Technology* **7**, 304–311.
- Favata, J. T.: 2001, Offline general handwritten word recognition using an approximate BEAM matching algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(9), 1009–1021.
- Francks, C., DeLisi, L., Fisher, S., Laval, S., Rue, J., Stein, J. and Monaco, A.: 2003, Confirmatory evidence for linkage of relative hand skill to 2p12-q11, *American Journal of Human Genetics* **72**, 499–502.
- Franke, K.: 2005, *The influence of physical and biomechanical processes on the ink trace: Methodological foundations for the forensic analysis of signatures*, PhD thesis, University of Groningen, Artificial Intelligence Institute, The Netherlands.
- Franke, K. and Grube, G.: 1998, The automatic extraction of pseudo-dynamic information from static images of handwriting based on marked gray value segmentation, *Journal of Forensic Document Examination* **11**, 17–38.

- Franke, K. and Köppen, M.: 1999, Towards an universal approach to background removal in images of bank checks - extended version, in S.-W. Lee (ed.), *Advances in Handwriting Recognition*, World Scientific, pp. 91–100.
- Franke, K. and Köppen, M.: 2001, A computer-based system to support forensic studies on handwritten documents, *International Journal on Document Analysis and Recognition* **3**(4), 218–231.
- Gulcher, J., Jonsson, P., Kong, A. and et al.: 1997, Mapping of a familial essential tremor gene, *fet1*, to chromosome 3q13, *Nature Genetics* **17**(1), 84–87.
- Gunter, S. and Bunke, H.: 2004, HMM-based handwritten word recognition: on the optimization of the number of states, training iteration and Gaussian components, *Pattern Recognition* **37**, 2069–2079.
- Guyon, I., Schomaker, L., Plamondon, R., Liberman, R. and Janet, S.: 1994, UNIPEN project of on-line data exchange and recognizer benchmarks, *Proc. of 12th ICPR*, Jerusalem, Israel, pp. 29–33.
- Hertel, C. and Bunke, H.: 2003, A set of novel features for writer identification, *Proc. of 4th Int. Conf. on Audio- and Video-Based Biometric Person Authentication (AVBPA 2003)*, Guildford, UK, pp. 679–687.
- Huber, R. A. and Headrick, A.: 1999, *Handwriting Identification: Facts and Fundamentals*, CRC Press, Boca Raton.
- Jaeger, S., Manke, S., Reichert, J. and Waibel, A.: 2001, On-line handwriting recognition: the NPen++ recognizer, *International Journal on Document Analysis and Recognition* **3**(3), 169–180.
- Jain, A. K., Duin, R. and Mao, J.: 2000, Statistical pattern recognition: a review, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(1), 4–37.
- Jain, A. K., Hong, L. and Bolle, R.: 1997, On-line fingerprint verification, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(4), 302–314.
- Jean, G.: 1997, *Writing: The story of alphabets and scripts*, Thames and Hudson Ltd., London.
- Joachims, T.: 1999, Making large-scale SVM learning practical, in B. Scholkopf, C. Burges and A. Smola (eds), *Advances in Kernel Methods - Support Vector Learning*, MIT Press.
- Kittler, J., Hatef, M., Duin, R. and Matas, J.: 1998, On combining classifiers, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(3), 226–239.
- Klement, V.: 1981, Forensic writer recognition, in J. Simon and R. Haralick (eds), *Proc. NATO Adv. Stud. Inst.*, pp. 519–524.
- Klement, V.: 1983, An application system for the computer-assisted identification of handwriting, *Proc. Int. Carnahan Conf. on Security Technol.*, Zurich, Switzerland, pp. 75–79.

- Klement, V., Steinke, K. and Naske, R.: 1980, The application of image processing and pattern recognition techniques to the forensic analysis of handwriting, *Proc. 1980 Int. Conf. Security through Sci. Engin.*, West Berlin, Germany, pp. 5–11.
- Koerich, A., Sabourin, R. and Suen, C.: 2003, Large vocabulary off-line handwriting recognition: a survey, *Pattern Anal. Applic.* **6**, 97–121.
- Kohonen, T.: 1988, *Self-Organization and Associative Memory*, second edn, Springer Verlag, Berlin.
- Kondo, S. and Attachoo, B.: 1986, Model of handwriting process and its analysis, *Proc. ICPR*, pp. 562–565.
- Köppen, M. and Franke, K.: 1999, Fuzzy Morphology revisited, *Proc. of 3rd International Workshop on Soft Computing in Industry (IWSCI)*, Muroran, Japan, pp. 258–263.
- Kuckuck, W.: 1980, Writer recognition by spectrum analysis, *Proc. 1980 Int. Conf. Security through Sci. Engin.*, West Berlin, Germany, pp. 1–3.
- Kuckuck, W., Rieger, B. and Steinke, K.: 1979, Automatic writer recognition, *Proc. 1979 Carnahan Conf. on Crime Countermeasures*, University of Kentucky, Lexington, pp. 57–64.
- Liu, J. and Gader, P.: 2002, Neural networks with enhanced outlier rejection ability for off-line handwritten word recognition, *Pattern Recognition* **35**, 2061–2071.
- Maarse, F. J.: 1987, *The study of handwriting movement: Peripheral models and signal processing techniques*, PhD thesis, University of Nijmegen, Department of Experimental Psychology, The Netherlands.
- Maarse, F., Schomaker, L. and Teulings, H.-L.: 1988, Automatic identification of writers, in G. van der Veer and G. Mulder (eds), *Human-Computer Interaction: Psychonomic Aspects*, Springer, New York, pp. 353–360.
- Maarse, F. and Thomassen, A.: 1983, Produced and perceived writing slant: differences between up and down strokes, *Acta Psychologica* **54**(1-3), 131–147.
- Maltoni, D., Maio, D., Jain, A. K. and Prabhakar, S.: 2003, *Handbook of Fingerprint Recognition*, Springer, New York.
- Marti, U.-V. and Bunke, H.: 2001, Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition system, *International Journal of Pattern Recognition and Artificial Intelligence* **15**, 65–90.
- Marti, U.-V. and Bunke, H.: 2002, The IAM-database: an english sentence database for off-line handwriting recognition, *International Journal on Document Analysis and Recognition* **5**(1), 39–46.
- Marti, U.-V., Messerli, R. and Bunke, H.: 2001, Writer identification using text line based features, *Proc. of 6th ICDAR*, Seattle, USA, pp. 101–105.

- Mohamed, M. and Gader, P.: 1996, Handwritten word recognition using segmentation-free hidden Markov modeling and segmentation-based dynamic programming techniques, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**(5), 548–554.
- Moler, E., Ballarin, V., Pessana, F., Torres, S. and Olmo, D.: 1998, Fingerprint identification using image enhancement techniques, *Journal of Forensic Sciences* **43**(3), 689–692.
- Moritz, E.: 1990, Replicator-based knowledge representation and spread dynamics, *IEEE International Conference on Systems, Man, and Cybernetics*, The Institution of Electrical Engineers, pp. 256–259.
- Morris, R. N.: 2000, *Forensic Handwriting Identification: Fundamental Concepts and Principles*, first edn, Academic Press, London.
- Naske, R.: 1982, Writer recognition by prototype related deformation of handprinted characters, *Proc. of the 6th International Conference on Pattern Recognition*, Vol. 2, Munich, Germany, pp. 819–822.
- Neyman, J. and Pearson, E.: 1933, On the problem of the most efficient tests of statistical hypotheses, *Phil. Trans. Roy. Soc. London, Series A* **231**, 289–337.
- Nosary, A., Heutte, L. and Paquet, T.: 2004, Unsupervised writer adaption applied to handwritten text recognition, *Pattern Recognition* **37**, 385–388.
- Otsu, N.: 1979, A threshold selection method from gray-level histogram, *IEEE Trans. Systems, Man and Cybernetics* **9**, 62–69.
- Parisse, C.: 1996, Global word shape processing in off-line recognition of handwriting, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**(4), 460–464.
- Philipp, M.: 1996, Fakten zu Fish, Das Forensische Informations System Handschriften des Bundeskriminalamtes - eine analyse nach uber 5 jahren wirkbetrieb [technical report], *Technical report*, Kriminaltechnisches Institut 53, Bundeskriminalamt, Wiesbaden, Germany.
- Plamondon, R. and Guerfali, W.: 1998, The generation of handwriting with delta-lognormal synergies, *Biological Cybernetics* **78**, 119–132.
- Plamondon, R., Lopresti, D. P., Schomaker, L. R. and Srihari, R.: 1999, On-line handwriting recognition, in J. G. Webster (ed.), *Encyclopedia of Electrical and Electronics Engineering*, Vol. 15, Wiley, New York, pp. 123–146.
- Plamondon, R. and Lorette, G.: 1989, Automatic signature verification and writer identification - the state of the art, *Pattern Recognition* **22**(2), 107–131.
- Plamondon, R. and Maarse, F.: 1989, An evaluation of motor models of handwriting, *IEEE Trans. Syst., Man, and Cybern.* **19**(5), 1060–1072.
- Plamondon, R. and Srihari, S. N.: 2000, On-line and off-line handwriting recognition: a comprehensive survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(1), 63–84.

- Press, W., Teukolsky, S., Vetterling, W. and Flannery, B.: 1992, *Numerical Recipes in C: The Art of Scientific Computing*, second edn, Cambridge Univ. Press, Cambridge.
- Roli, F., Kittler, J., Fumera, G. and Muntoni, D.: 2002, An experimental comparison of classifier fusion rules for multimodal personal identity verification systems, *Proc. Multiple Classifier Systems: LNCS 2364*, Springer, pp. 325–336.
- Sabourin, R. and Drouhard, J.: 1992, Off-line signature verification using directional pdf and neural networks, *Proc. of the 11th International Conference on Pattern Recognition (ICPR 1992)*, The Hague, Netherlands, pp. 321–325.
- Said, H., Peake, G., Tan, T. and Baker, K.: 1998, Writer identification from non-uniformly skewed handwriting images, *Proc. of the 9th British Machine Vision Conference*, pp. 478–487.
- Said, H., Tan, T. and Baker, K.: 2000, Personal identification based on handwriting, *Pattern Recognition* **33**(1), 149–160.
- Schlapbach, A. and Bunke, H.: 2004, Using HMM-based recognizers for writer identification and verification, *Proc. of 9th IWFHR*, IEEE Computer Society, Tokyo, Japan, pp. 167–172.
- Schlapbach, A., Kilchherr, V. and Bunke, H.: 2005, Improving writer identification by means of feature selection and extraction, *Proc. of 8th Int. Conf. on Document Analysis and Recognition (ICDAR 2005)*, Vol. I, IEEE Computer Society, Seoul, Korea, pp. 131–135.
- Schmidt, R.: 1975, A schema theory of discrete motor skill learning, *Psychological Review* **82**, 225–260.
- Schomaker, L.: 1991, *Simulation and recognition of handwriting movements: A vertical approach to modeling human motor behavior*, PhD thesis, University of Nijmegen, NICI, The Netherlands.
- Schomaker, L. and Bulacu, M.: 2004, Automatic writer identification using connected-component contours and edge-based features of uppercase western script, *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)* **26**(6), 787–798.
- Schomaker, L., Bulacu, M. and Franke, K.: 2004, Automatic writer identification using fragmented connected-component contours, *Proc. of 9th International Workshop on Frontiers in Handwriting Recognition (IWFHR 2004)*, IEEE Computer Society, Tokyo, Japan, pp. 185–190.
- Schomaker, L., Bulacu, M. and van Erp, M.: 2003, Sparse-parametric writer identification using heterogeneous feature groups, *Proc. of Int. Conf. on Image Processing (ICIP 2003)*, Vol. I, IEEE Press, Barcelona, Spain, pp. 545–548.
- Schomaker, L., Franke, K. and Bulacu, M.: 2007, Using codebooks of fragmented connected-component contours in forensic and historic writer identification, *Pattern Recognition Letters (PRL), Pattern Recognition in Cultural Heritage and Medical Applications* **28**(6), 719–727.
- Schomaker, L. R. B.: 1993, Using stroke- or character-based self-organizing maps in the recognition of on-line, connected cursive script, *Pattern Recognition* **26**(3), 443–450.

- Schomaker, L. R. B.: 1998, From handwriting analysis to pen-computer applications, *IEE Electronics Communication Engineering Journal* **10**(3), 93–102.
- Schomaker, L. R. B. and Plamondon, R.: 1990, The relation between pen force and pen-point kinematics in handwriting, *Biological Cybernetics* **63**, 277–289.
- Schomaker, L. R. B., Thomassen, A. J. W. M. and Teulings, H.-L.: 1989, A computational model of cursive handwriting, in R. Plamondon, C. Y. Suen and M. L. Simner (eds), *Computer Recognition and Human Production of Handwriting*, Singapore: World Scientific, pp. 153–177.
- Schomaker, L. and Vuurpijl, L.: 2000, Forensic writer identification: A benchmark data set and a comparison of two systems [internal report for the Netherlands Forensic Institute], *Technical report*, Nijmegen: NICI.
- Senior, A. W. and Robinson, A. J.: 1998, An off-line cursive handwriting recognition system, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(3), 309–321.
- Shannon, C. E.: 1948, A mathematical theory of communication, *Bell System Technical Journal* **27**, 379–423 and 623–656.
- Srihari, S., Beal, M., Bandi, K., Shah, V. and Krishnamurthy, P.: 2005, A statistical model for writer verification, *Proc. of 8th Int. Conf. on Document Analysis and Recognition (ICDAR 2005)*, Vol. II, IEEE Computer Society, Seoul, Korea, pp. 1105–1109.
- Srihari, S., Cha, S., Arora, H. and Lee, S.: 2002, Individuality of handwriting, *J. of Forensic Sciences* **47**(4), 1–17.
- Srihari, S., Tomai, C., Zhang, B. and Lee, S.: 2003, Individuality of numerals, *Proc. of 7th ICDAR*, Vol. II, Edinburgh, Scotland, pp. 1096–1100.
- Steinherz, T., Rivlin, E. and Intrator, N.: 1999, Off-line cursive script word recognition - a survey, *International Journal on Document Analysis and Recognition* **2**(2), 90–110.
- Steinke, K.: 1981, Recognition of writers by handwriting images, *Pattern Recognition* **14**(1-6), 357–364.
- Tan, T.: 1998, Rotation invariant texture features and their use in automatic script identification, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(7), 751–756.
- Tomai, C., Zhang, B. and Srihari, S.: 2004, Discriminatory power of handwritten words for writer recognition, *Proc. 17th Int. Conf. on Pattern Recognition (ICPR 2004)*, Vol. II, IEEE Computer Society, Cambridge, UK, pp. 638–641.
- van der Maaten, L. and Postma, E.: 2005, Improving automatic writer identification, *Proc. of 17th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2005)*, Brussels, Belgium, pp. 260–266.
- van Erp, M., Vuurpijl, L., Franke, K. and Schomaker, L.: 2003, The WANDA measurement tool for forensic document examination, *Proc. of the IGS'2003*, Scottsdale, Arizona, pp. 282–285.

- Van Galen, G., Portier, J., Smits-Engelsman, B. and Schomaker, L.: 1993, Neuromotor noise and poor handwriting in children, *Acta Psychologica* **82**, 161–178.
- Vinciarelli, A.: 2002, A survey on off-line cursive word recognition, *Pattern Recognition* **35**, 1433–1446.
- Vinciarelli, A., Bengio, S. and Bunke, H.: 2004, Offline recognition of unconstrained handwritten texts using HMMs and statistical language models, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(6), 709–720.
- Vuurpijl, L. and Schomaker, L.: 1997, Finding structure in diversity: A hierarchical clustering method for the categorization of allographs in handwriting, *Proc. of 4th ICDAR*, IEEE, Ulm, Germany, pp. 387–393.
- Vuurpijl, L., Schomaker, L. and van Erp, M.: 2003, Architecture for detecting and solving conflicts: two-stage classification and support vector classifiers, *International Journal of Document Analysis and Recognition* **5**(4), 213 – 223.
- Xue, H. and Govindaraju, V.: 2002, On the dependence of handwritten word recognizers on lexicons, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(12), 1553–1564.
- Zhang, B. and Srihari, S.: 2003, Analysis of handwritten individuality using word features, *Proc. of 7th ICDAR*, Vol. II, Edinburgh, Scotland, pp. 1142–1146.
- Zhang, B., Srihari, S. and Lee, S.: 2003, Individuality of handwritten characters, *Proc. of 7th ICDAR*, Vol. II, Edinburgh, Scotland, pp. 1086–1090.
- Zhu, Y., Tan, T. and Wang, Y.: 2001, Font recognition based on global texture analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(10), 1192–1200.
- Zimmermann, M. and Bunke, H.: 2002, Automatic segmentation of the IAM off-line database for handwritten english text, *Proc. of 16th Int. Conf. on Pattern Recognition (ICPR 2002)*, Vol. IV, IEEE Computer Society, pp. 35–39.
- Zois, E. and Anastassopoulos, V.: 2000, Morphological waveform coding for writer identification, *Pattern Recognition* **33**(3), 385–398.

Index

- affine transforms, 7
- allograph, 9, 55, 56, 73, 100
- analysis
 - Fourier, 81, 83, 91
 - scale, 91, 95
 - texture level / allograph level, 12, 16, 73, 77, 100, 101
- angle combinations, 16, 24, 81
- Bayes, 105
 - error rate, 108
 - feature fusion framework, 87
- Beowulf Linux cluster, 62, 96
- biomechanical systems (thumb-fingers and hand-wrist), 10, 55, 73
- biometrics, 1, 100
 - behavioral vs. physiological, 2, 35, 70
 - multi-modal, 88
 - templates, 2, 70
- Bresenham line generator, 58, 74
- classifier
 - combination, 48, 88
 - nearest-neighbor, 28, 43, 62, 86
 - threshold, 64, 86
- clustering
 - k-means, 60, 83, 96, 98
 - Kohonen maps, 60, 64, 83, 96
- comparison
 - lowercase vs. uppercase, 17, 43, 63, 88
 - one-to-many search, 4, 35, 53, 71
 - one-to-one, 4, 35, 54, 71
 - split-line vs. entire-line, 43
- computer
 - algorithms, 1, 104
 - vision, 1, 16, 91
- connected-component, 59
 - 8-connectivity, 77
 - detection (labeling), 76
- contour
 - angle, 78, 81
 - following (Moore), 60, 77
 - fragment, 78, 79, 81
 - inner / outer, 54, 77
 - starting point, 54
- datasets, 13, 74
 - Firemaker, 22, 36, 57, 74
 - IAM, 14, 74
 - ImUnipen / Unipen, 58, 60, 74, 83
 - Large, 74
- dimensionality problem, 66, 97
- dissimilarity
 - between handwritings, 85
 - measures, 62, 85
- distance
 - χ^2 , 37, 49, 62, 85
 - Euclid, 28, 61, 82, 85
 - Hamming, 87, 93
 - sorting / ranking, 28, 62, 86

- document
 - analysis and recognition, 15, 105
 - skew, 37
- dot product, 28, 82
- Dynamic Time Warping (DTW), 15
- edge
 - angle, 16, 25
 - detection (Sobel), 23, 39
 - fragment, 24, 25, 39
- feature, 1
 - automatic, 3, 36, 105
 - completeness, 95
 - discriminatory power, 54, 97
 - groups, 76, 91
 - interactively measured, 3, 22
 - matching, 4, 75
 - orthogonal, 31, 48, 86
 - statistical, 6, 16, 101
 - vector dimensionality, 24, 26, 30, 40, 80, 81
- feature combination, 4, 12, 14, 17, 70, 91, 104
 - average (weighted average), 87, 91, 93
 - SVM, 93
 - vector concatenation, 32, 38
 - voting (Borda), 48
- feature extraction, 4, 37, 75
 - line splitting, 36–38, 103
 - lossy operation, 95
- features, allographic, 3, 14, 22, 83
 - grapheme-emission distribution, 54, 55, 57, 61, 83, 85, 88, 103
- features, textural, 14, 22, 23, 31, 41, 75
 - autocorrelation, 28, 81
 - contour-direction distribution, 78
 - contour-hinge distribution, 79, 88
 - direction co-occurrence distribution, 41, 81, 93, 103
 - edge-direction distribution, 23, 39, 103
 - edge-hinge distribution, 25, 40
 - entropy, 28
 - ink-density distribution, 42
 - run-length distributions, 27, 31, 41, 81, 93
- Fish, 3
- font identification, 13, 99
- foreground/background separation, 3, 15, 36, 77, 105
- forensic experts, 6, 71
- gcc, 62, 96
- grapheme (glyph), 56, 59, 103
 - clustering, 14, 17, 54, 83
 - matching, 61, 85
 - segmentation, 56, 83
- grapheme codebook, 60, 83, 85, 98, 103
 - size, 60, 62, 63, 66, 96–98
 - training, 60, 62, 96, 98
- grapheme representation, 98
 - contour, 54, 64
 - normalized bitmap, 60, 83
- Groningen Automatic Writer Identification System (GRAWIS), 107
- handwriting
 - analysis, 35, 70
 - curvature, 40, 80, 100
 - direction, 10, 16, 103
 - education, 10
 - forged (disguised), 6, 56, 71
 - invariant representations, 4, 21, 70
 - movement, 10–12, 22, 55, 72
 - natural (habitual), 6, 10, 15, 71
 - offline / online, 8, 12, 23
 - questioned sample, 4, 71
 - recognition vs. identification, 4, 21, 24, 35, 70, 100
 - segmentation, 5, 56, 59, 85
 - slant, 7, 10, 21, 25, 39, 78, 79, 100
 - style, 4, 44, 67
 - text-blocks, 1, 6, 15, 36, 102
- handwriting individuality, 9, 71
 - genetic factors, 9

- memetic factors, 9
- handwriting stroke
 - order, 8
 - regularity, 28, 82
 - upward / downward, 11, 24, 39
- handwriting variability, 7, 101
 - between-writer / within-writer, 6, 25, 39, 71, 101
 - neuro-biomechanical, 7
 - sequencing, 8
- HTML, 107
- human intervention, 1, 3, 5, 15, 36, 54, 70, 77
- image
 - binary, 16, 23, 27, 76, 82
 - compression (Lempel-Zif), 28
 - gray-scale, 23, 58, 74
 - processing, 24, 76
 - projection profile, 38
 - region of interest, 3, 22
 - resolution, 23, 58, 74
 - retrieval, 5, 71
 - scanning direction, 27, 81, 82, 93
 - thresholding (Otsu), 76
- information
 - amplitude vs. phase, 91
 - directional, 22
 - fusion, 88, 93
 - location, 6, 36, 38, 39, 44, 103
 - retrieval, 54
- ink-trace width (ink thickness), 23, 25, 78, 79
- Markov process, 80
- methods
 - automatic, 1, 22
 - sparse-parametric, 1, 15
 - text-dependent, 5, 13, 15
 - text-independent, 5, 13, 15, 16, 21, 30, 70, 95, 102
- Optical Character Recognition (OCR), 99
- pattern recognition, 1, 4, 16, 35, 70, 101
- pen
 - force (pressure), 12, 42
 - grip, 9, 22
- performance
 - Equal Error Rate (EER), 65, 66, 86, 93, 97
 - leave-one-out, 28, 43, 62, 86
 - plots, 30, 31, 43, 62, 93, 97, 99
 - Receiver Operating Characteristic (ROC), 65, 86
 - target, 2, 22, 71
 - Top1 / Top 10 identification rate, 30, 48, 62, 93, 97, 99
- phase
 - information, 81, 83, 91
 - local analysis, 81, 91, 95
 - local correlations, 91
- probability distribution
 - bivariate, 26, 80
 - conditional, 80
 - function (PDF), 37, 62, 70, 76, 85, 100
 - histogram counting and normalization, 24, 25, 39, 62, 78, 80, 85
 - joint, 25, 40, 41, 80, 81, 91
 - landscape, 26
 - same-writer / different-writer, 65, 86
 - stable estimates, 15, 16, 32, 44, 91, 97
- Script, 3
- script identification, 13
- shape
 - codebook, 66
 - matching, 54
 - representation, 54, 67
- signature verification, 5, 39, 78
- statistical decision theory (Neyman-Pearson), 86
- stochastic pattern generator, 54, 55, 61, 83
- text-pose estimation, 41
- theoretical model, 55
- Trigraph, 106, 107
- visualization tool, 107

Wanda, 107

writer

 individuality, 100, 101

 population, 89

 search / recall, 86

writer identification, 4, 71, 102

 forensic, 1, 2, 5, 22, 32, 99

 historic, 81, 99, 106

 hit list, 5, 28, 54, 86, 108

writer verification, 4, 71, 102

 decision threshold, 64, 66, 86, 97

 false-accept error (false alarm), 64, 86,
 108

 false-reject error (miss), 64, 86, 108

Publications related to this thesis

Journal papers

Marius Bulacu, Lambert Schomaker (2007) *Text-independent writer identification and verification using textural and allographic features*, IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), Special Issue - Biometrics: Progress and Directions, IEEE Computer Society, vol. 29, no. 4, pp. 701-717, April.

Lambert Schomaker, Katrin Franke, Marius Bulacu (2007) *Using codebooks of fragmented connected-component contours in forensic and historic writer identification*, Pattern Recognition Letters (PRL), Pattern Recognition in Cultural Heritage and Medical Applications, Elsevier Science, vol. 28, no. 6, pp. 719-727, 15 April.

Lambert Schomaker, Marius Bulacu (2004) *Automatic writer identification using connected-component contours and edge-based features of upper-case western script*, IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), IEEE Computer Society, vol. 26, no. 6, pp. 787-798, June.

Refereed conference papers

Marius Bulacu, Lambert Schomaker (2006) *Combining multiple features for text-independent writer identification and verification*, Proc. of 10th International Workshop on Frontiers in Handwriting Recognition (IWFHR 2006), pp. 281-286, 23 - 26 October, La Baule, France.

Marius Bulacu, Lambert Schomaker (2005) *GRAWIS: Groningen Automatic Writer Identification System*, Proc. of 17th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2005), pp. 413-414, 17 - 18 October, Brussels, Belgium.

Marius Bulacu, Lambert Schomaker (2005) *A comparison of clustering methods for writer identification and verification*, Proc. of 8th Int. Conf. on Document Analysis and Recognition (ICDAR 2005), IEEE Computer Society, pp. 1275-1279, vol. II, 29 August - 1 September, Seoul, Korea.

Marius Bulacu, Lambert Schomaker (2004) *Analysis of texture and connected-component contours for the automatic identification of writers*, Proc. of 16th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2004), pp. 371-372, 21 - 22 October, Groningen, The Netherlands.

Lambert Schomaker, Marius Bulacu, Katrin Franke (2004) *Automatic writer identification using fragmented connected-component contours*, Proc. of 9th International Workshop on Frontiers in Handwriting Recognition (IWFHR 2004), IEEE Computer Society, pp. 185-190, 26 - 29 October, Tokyo, Japan.

Marius Bulacu, Lambert Schomaker (2003) *Writer style from oriented edge fragments*, Proc. of 10th Int. Conf. on Computer Analysis of Images and Patterns (CAIP 2003): LNCS 2756, Springer, pp. 460-469, 25 - 27 August, Groningen, The Netherlands.

Lambert Schomaker, Marius Bulacu, Merijn van Erp (2003) *Sparse-parametric writer identification using heterogeneous feature groups*, Proc. of Int. Conf. on Image Processing (ICIP 2003), IEEE Press, pp. 545-548, vol. I, 14 - 17 September, Barcelona, Spain.

Marius Bulacu, Lambert Schomaker, Louis Vuurpijl (2003) *Writer identification using edge-based directional features*, Proc. of 7th Int. Conf. on Document Analysis and Recognition (ICDAR 2003), IEEE Computer Society, pp. 937-941, vol. II, 3 - 6 August, Edinburgh, Scotland.

Other publications

Marius Bulacu, Lambert Schomaker (2005) *Text-pose estimation in 3D using edge-direction distributions*, Proc. of Int. Conf. on Image Analysis and Recognition (ICIAR 2005): LNCS 3656, Springer, pp. 625-634, 28 - 30 September, Toronto, Canada.

Nobuo Ezaki, Bui Truong Minh, Kimiyasu Kiyota, Marius Bulacu, Lambert Schomaker (2005) *Improved text-detection methods for a camera-based text reading system for blind persons*, Proc. of 8th Int. Conf. on Document Analysis and Recognition (ICDAR 2005), IEEE Computer Society, pp. 257-261, vol. I, 29 August - 1 September, Seoul, Korea.

Nobuo Ezaki, Marius Bulacu, Lambert Schomaker (2004) *Text detection from natural scene images: Towards a system for visually impaired persons*, Proc. of 17th Int. Conf. on Pattern Recognition (ICPR 2004), IEEE Computer Society, pp. 683-686, vol. II, 23 - 26 August, Cambridge, UK.

Samenvatting

Information is physical.

Rolf Landauer

Dit proefschrift is een exercitie in het ontwikkelen van methoden voor het identificeren van schrijvers op basis van schriftkenmerken. Aan schrijveridentificatie liggen twee principes ten grondslag.

- Er zijn geen twee personen die exact hetzelfde schrijven.
- Niemand schrijft een zelfde tekst tweemaal exact hetzelfde.

Hoewel dit uitgangspunt ongenueanceerd en betwistbaar is, blijftstaan dat er twee natuurlijke factoren in conflict zijn bij een poging om een persoon te identificeren op basis van handgeschreven documenten, namelijk, (1) de *variatie* in schrijfvormen tussen *verschillende schrijvers* (2) de *variabiliteit* van het schrift *binnen een schrijver*. Het doel van het onderzoek in dit proefschrift is om het proces van schrijveridentificatie te automatiseren met gescande afbeeldingen van handschrift, uitgaande van minimale menselijke tussenkomst. Hiertoe is het nodig om de individualiteit van handschrift met behulp van een computer vast te stellen. Dit brengt ons op een derde aspect van het onderzoek, het ontwerp en gebruik van *geschikte representaties*, d.w.z., berekenbare kenmerken die de schrijfstijl van een persoon vatten uit het 'gescande' beeld van handgeschreven documenten. De kracht van een dergelijk kenmerk (Engels: 'feature') wordt bepaald door zijn vermogen om het onderscheid tussen verschillende schrijvers te maximaliseren en tegelijkertijd de verschillen tussen het schriftmonsters van dezelfde schrijver te negeren. In dit proefschrift worden nieuwe en zeer effectieve kenmerken gepresenteerd voor de automatische schrijveridentificatie op basis van schriftmonsters. De overeenkomst in schrijfstijl tussen twee willekeurige manuscripten wordt berekend met een geschikte afstandsmaat voor de gekozen kenmerken. De gehanteerde methoden vallen binnen het raamwerk van de statistische patroonherkenning (Duda et al. 2001, Jain et al. 2000).

Bij de keuze van informatieve fundamentele kenmerken van schrift worden twee facetten van het schrijven genomen, elk op een eigen niveau van analyse. Ten eerste zijn er de schuinheid, kromming en rondheid van het schrift die in hoge mate bepaald worden door de habituele pen-greep. Deze vormkenmerken worden berekend door samengestelde kansdichtheidverdelingen van richtingen in de textuur van het schriftbeeld. Ten tweede bestaat er een gepersonaliseerde

verzameling van lettervormen, *allografen* genaamd, die een schrijver geleerd heeft te gebruiken onder invloed van genoten onderwijs, cultuur en memetische (nabootsende) processen. Om de gebruikte vormen in de letters van een schrijfstijl te karakteriseren wordt gebruik gemaakt van een stochastisch grafeem-emissie model. Hierbij wordt er van uitgegaan dat de schrijver letters vormt door het plaatsen van stijlspecifieke vormelementen (glyfen). Het combineren van de textuur- en allograafgebaseerde kenmerken voorziet in een zeer nauwkeurige en uitvoerige karakterisering van de individuele schriftstijl van een persoon. De ontwikkelde methoden presteerden zeer goed bij schrijveridentificatie (het zoeken van de juiste schrijver van een betwist document in een verzameling documenten van bekende oorsprong) en bij schrijververificatie (het geven van een gelijkensoordeel, gegeven een betwist en een bekend schriftmonster van een schrijver). Hierbij zijn uitgebreide tests met verzamelingen handgeschreven documenten uitgevoerd, in schrijversgroepen tot 900 proefpersonen.

In onze methoden wordt een individuele schriftstijl robuust gerepresenteerd middels kansverdelingsfuncties die uit het schriftbeeld worden afgeleid. Deze benadering heeft twee karakteristieke eigenschappen: in het proces van schrijveridentificatie wordt menselijke tussenkomst geminimaliseerd en de individuele schriftstijl wordt zoveel mogelijk onafhankelijk van de tekstuele inhoud geëncodeerd. In feite is de computer 'onwetend' met betrekking tot de inhoud van het geschrevene. De ontwikkelde algoritmen hebben het potentiële om ingezet te worden in reële toepassingen. De structuur van het proefschrift is als volgt. Na een inleidend hoofdstuk 1 behandelen de hoofdstukken 2 en 3 de textuurgebaseerde kenmerken. De hoofdstukken 4 en 5 behandelen de allograafgebaseerde methode alsmede een combinatiemethode voor textuur- en allograafkenmerken om te komen tot verbeterde systeemprestaties.

Hoofdstuk 1 van het proefschrift introduceert schrijveridentificatie als een voorbeeld van een 'behavioral biometric', een gedragsgebaseerde biometrie en geeft een overzicht over mogelijke fundamentele genetische en culturele factoren die de individualiteit van schrift bepalen. De taak van *schrijveridentificatie* komt overeen met de vraag: "Wie heeft dit document geschreven?". Een systeem voor schrijveridentificatie zoekt aan de hand van een betwist document in een groot bestand met documenten waarvan de identiteit van de schrijver bekend is en geeft vervolgens een lijst terug met kandidaten, gesorteerd op gelijkens met het betwist document. Deze lijst wordt vervolgens geïnspecteerd door een menselijke (forensische) expert. Een systeem voor *schrijververificatie* beantwoordt de vraag: "Zijn deze twee documenten geschreven door dezelfde persoon?". Twee documenten worden één op één vergeleken en een beslissing "gelijk"/"ongelijk" wordt automatisch genomen. Verder wordt in dit inleidende hoofdstuk een verband gelegd met het gerelateerde en veel bredere gebied van de handschriftherkenning. Bij automatische schriftherkenning moeten de variaties in schriftstijl worden geëlimineerd door gebruik te maken van invariante kenmerken die goed generaliseren over verschillende lettervormen uit het schrift. Bij schrijveridentificatie moet het contrast tussen verschillende lettervariaties juist versterkt worden om schrijvers te kunnen onderscheiden. Er wordt een overzicht gegeven van recente publicaties. Er is een onderscheid tussen tekstafhankelijke en tekstonafhankelijke methoden: het huidige onderzoek valt in de tweede groep.

In hoofdstuk 2 wordt aangetoond dat het gebruik van de orientatie van korte fragmenten van

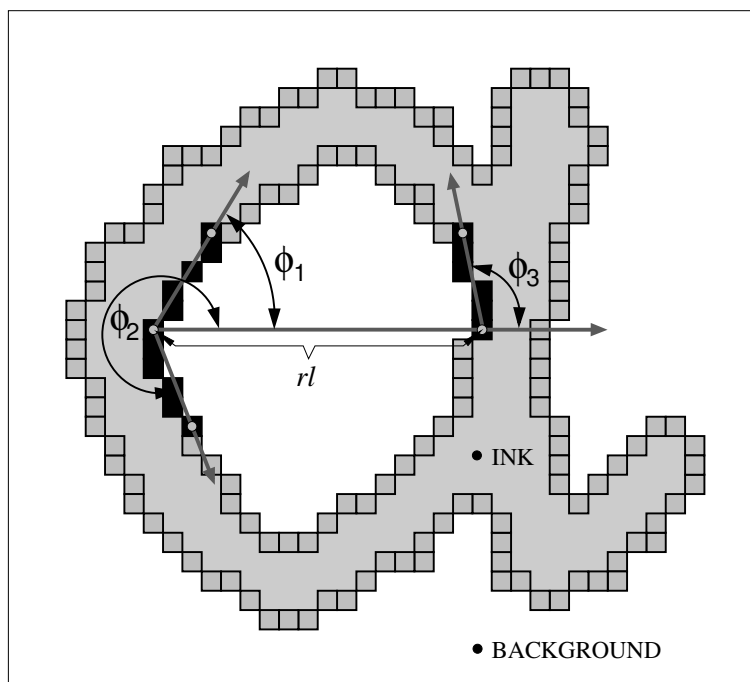


Figure B.1: Schematische weergave van richtingskenmerken in schrift. De hoek ϕ en combinaties van hoeken ϕ_1 , ϕ_2 en ϕ_3 zijn informatief voor persoonlijke schrijfstijl.

de randen ('edges') van het schrijfspoor de basis vormt voor het construeren van kansverdelingen van richtingen die effectief zijn in schrijveridentificatie. De eerste - op hoeken gebaseerde - kenmerksvector beschrijft de kansverdeling van hoeken van lijnstukjes op de randen van het schrijfspoor, een klassieke descriptor in schrijveridentificatie. De modus van deze verdeling, d.w.z. de dominante richting in het schrift, komt overeen met de schuinheid van het schrift ('slant'). Dit is een stabiel en typisch kenmerk voor een schrijver, met een redelijk onderscheidend vermogen tussen schrijvers. Vervolgens wordt een nieuwe krachtige methode voorgesteld die uitgaat van twee hoeken van een scharnier ('hinge') dat langs de randen van het schrijfspoor wordt gelegd. Hiermee wordt een conjuncte kansverdeling van twee hoeken vastgelegd, waarmee gelijktijdig de verdeling van globale richtingen en de verdeling van kromming kan worden gerepresenteerd. Dit nieuwe kenmerk ('edge hinge') is een bivariate kansverdeling die een zeer significante verbetering in schrijveridentificatie opleverde in verhouding tot de klassieke univariate kansverdeling van richtingen. De richtingsgebaseerde kenmerken presteren voor het overige ook beter dan een aantal andere traditionele kenmerken ('run length', autocorrelatie, entropie). Het reduceren van de hoeveelheid inkt in de testmonsters leidt tot een algehele daling in schrijveridentificatie voor alle methoden. De rangorde van de methoden in termen van hun prestatie verandert hierbij niet.

In hoofdstuk 3 wordt dieper ingegaan op de lokale richtingen binnen het handschrift als een effectieve bron van informatie voor schrijveridentificatie. Een derde, nieuw kenmerk werd ontworpen dat de bivariate kansverdeling van richtingen weergeeft, die zich voordoen op tegenoverliggende inkranden met een tussenliggend wit veld. Verdere verbeteringen van prestaties

werden verkregen door locatie-afhankelijkheid van de basiskenmerken mee te nemen. Dit wordt bereikt door de kenmerken onafhankelijk te berekenen voor het bovenste deel en het onderste deel van een tekstregel, en vervolgens de resulterende vectoren samen te voegen. De asymmetrie tussen het bovenste en onderste deel van regels tekst levert blijkbaar additionele schrijverspecifieke informatie op. In het empirisch onderzoek wordt een vergelijkbare prestatie in tekstonafhankelijke schrijveridentificatie vastgesteld voor handschrift bestaande uit uitsluitend hoofdletters en normaal, gemengd handschrift bestaande uit voornamelijk kleine letters en enige hoofdletters, voor de batterij van verschillende kenmerken die hier berekend zijn. Figuur B.1 geeft een schematische weergave van de gebruikte elementaire onderdelen van de hoekverdelingen.

Hoofdstuk 4 introduceert onze allograaf-gebaseerde methode voor schrijveridentificatie en -verificatie. Deze methode heeft een theoretische basis en neemt aan dat iedere schrijver wordt gekarakteriseerd op basis van de productie van elementaire schrijfvormen uit een algemeen code-boek van vormen. Deze elementaire vormen of *glyfen* worden verkregen door een heuristische segmentatie op het inktspoor toe te passen. Een algemeen code-boek van vormen (Fig. B.2) wordt berekend door 'clustering' van de verzameling van glyfen die geëxtraheerd zijn uit het schrift van een aanzienlijk aantal schrijvers. Deze schrijvers worden buitengesloten van de experimenten in verificatie en identificatie, waarvoor een verzameling ongeziene schrijvers wordt gebruikt. De gebruikte segmentatie-methode levert glyfen op die minder, meer, of precies evenveel inkt als een karakter (letter) representeren. Een dergelijke methode zou problemen opleveren indien het doel is om de inhoud van de tekst te herkennen. Desalniettemin blijken de sub- of supra-allografische vormen zeer descriptief te zijn met betrekking tot een schrijfstijl. Deze methode leent zich daarom uitstekend voor de automatische identificatie van schrijvers. In grootschalige reken-experimenten zijn er verschillende methoden vergeleken om een code-boek van basisvormen te bepalen: 'k means', Kohonen zelf-organiserende kaarten (1D en 2D). Het bleek dat de voorgestelde methode vrij ongevoelig is voor de aard van het gebruikte cluster-algoritme. Een vergelijkbare prestatie in schrijver-identificatie kan worden verkregen voor een breed bereik in omvang van het code-boek in termen van aantallen glyfen.

Hoofdstuk 5 gaat in op de vraag of er een verbetering in de prestaties van schrijveridentificatie optreedt indien de eerder gepresenteerde methoden worden gecombineerd. Op theoretische gronden is aangenomen dat de twee hoofdmethoden - tekstuurgebaseerde versus allograaf-gebaseerde kenmerken - twee verschillende informatiebronnen representeren voor de identificatie van een schrijver. In dit hoofdstuk wordt daarom een uitgebreide analyse gepresenteerd over combinatiemethoden om de effectiviteit en robuustheid van automatische schrijveridentificatie te verbeteren. Hoewel de twee methoden niet in absolute zin orthogonaal zijn, leveren ze een zeer verschillend perspectief op één en hetzelfde stuk handschrift, op verschillend niveau van analyse en op een verschillende schaal binnen het schrift. Het is gebleken dat complexe methoden voor de combinatie van kenmerken hier niet nodig zijn. Gegeven een aantal randvoorwaarden betreffende de omvang en de aard van de kenmerksvectoren blijkt het rekenkundig gemiddelde van de twee afstanden tussen een betwist monster en een referentiemonster voor de twee methoden het meest effectief te zijn. Een tweede aspect dat belicht

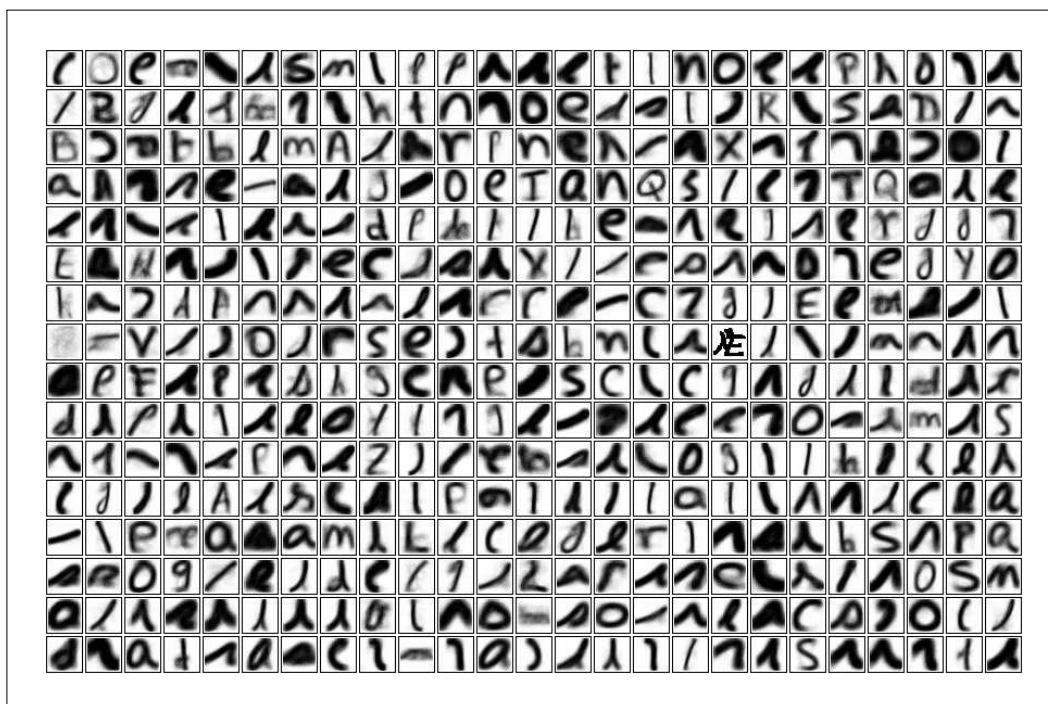


Figure B.2: Voorbeeld van een code-boek met schriftfragmenten (glyfen) berekend over een grote verzameling schrijvers. Elke schrijver gebruikt een kenmerkende deelverzameling van glyfen uit deze referentieverzameling.

wordt in dit hoofdstuk betreft de computationele doelmatigheid. De berekening van de textuurgerelateerde kenmerken hoeft niet plaats te vinden op basis van een (dure) directe convolutie van scharniervormen met het schriftbeeld. Een alternatieve methode is gebaseerd op de detectie van aaneengesloten brokken inkt ('connected components') waarvan de contour bepaald kan worden. Daarna hoeven alleen de richtingen van de contouren te worden bepaald zonder irrelevante delen van het beeld te raadplegen. Een derde facet van dit hoofdstuk betreft de omvang van de bestanden met schrijvers. Experimenten met een bestand van 900 schrijvers leveren een bevestiging van eerder verkregen resultaten. Deze omvang is vergelijkbaar met de grootste verzameling van gegevens die elders gebruikt werd voor schrijver-identificatie. De experimentele resultaten waren consistent over verschillende collecties van bestanden en toonden aan dat een combinatie van kenmerken een verbetering opleverde in zowel de identificatie als de verificatie van schrijvers. De beste combinatie van kenmerken integreert informatie over: richtingen, grafemen en letterplaatsing ('run lengths') met, voor 900 schrijvers, een Top-1 identificatiepercentage van 85-87% en een Top-10 herkenning van 96%. Voor de verificatie van schrijvers werd een fout ('equal-error rate') van 3% gevonden.

Het laatste hoofdstuk 6 geeft conclusies, terwijl Aanhangsel A een overzicht geeft van een HTML-demonstratie van de ontwikkelde methode, met het doel om de prestaties van het systeem GRAWIS (Groningen Automatic Writer Identification System) visueel te kunnen beoordelen.

Dit proefschrift geeft een gedetailleerd overzicht van algoritmische aspecten van automatische schrijveridentificatie en -verificatie. De ontwikkelde tekstonafhankelijke methoden hebben mogelijk impact op toepassingen in forensisch onderzoek: ze maken het mogelijk om in zeer grote bestanden van handgeschreven tekstbeelden te zoeken naar schriftmonsters die lijken op de schrijfstijl van een gegeven betwist document. Zo kan er een lijst van kandidaten worden gepresenteerd aan de forensisch schriftexpert ter nadere analyse om te komen tot een eindoordeel. Een deel van de tekstuurgebaseerde kenmerken uit deze dissertatie is al feitelijk in gebruik bij een industriële toepassing. De verdere verdieping van de fundamentele inzichten uit dit promotie-onderzoek en de uitgewerkte toepassing van de voorgestelde methoden zijn een mooie uitdaging voor nader onderzoek.

Acknowledgments

The PhD years have been very beautiful and intense and the time has come now to conclude this important period in my life. Here I want to thank the people who have influenced my scientific work and helped me reach this moment.

Primary thanks go to my supervisor prof. Lambert Schomaker who has given me the chance to work in his group. Lambert, your prodigious knowledge on handwriting that you generously shared with me during the last years provided the necessary basis without which the current thesis would have been impossible. Your thorough approach to science was an invaluable guide to me during the project and your honest friendship has always provided a tonic spirit to our discussions. I am very fortunate to be able to continue to work together with you in the postdoc period.

I am grateful to the members of the assessment committee, prof. John Daugman, prof. Frans Groen and prof. Jos Roerdink, for reading and evaluating the manuscript. I appreciate their willingness to assume responsibility for the quality of this dissertation.

I thank prof. John Daugman and prof. Linne Mooney for giving me a dataset with images of medieval documents to test the effectiveness of the writer identification algorithms described in this thesis on historical writings.

During my PhD project, I had a very good and productive collaboration with Nobuo Ezaki. I am grateful to him, also because he introduced me to Asian cuisine. For a short period, I also collaborated with Vamsi Madasu on the topic of signature verification. I thank Katrin Franke for her constant interest and implication in the area of forensic document analysis. I am grateful to Louis Vuurpijl for giving me some very useful advice in a critical moment of the project. I also thank the engineers from Prime Vision and the forensics experts from the Netherlands Forensic Institute for very stimulating discussions.

Further, I wish to thank all my colleagues from the Artificial Intelligence Department for their contribution in making my stay in Groningen a very pleasurable one. The trips organized in the department were very nice moments where I could directly experience the Dutch landscape and hospitality. I thank Esther Wiersinga-Post and Geertje Zwarts for having helped me find a house in Groningen. I thank Wouter Teepe for always being very kind and supportive and also for providing some essential pieces of furniture for my house. I also thank Tijn van der Zant for his contribution to the proposal that became the research project in which I am currently involved as a postdoc. I would also like to mention Axel Brink who is making sure that writer identification remains an active area of research in Groningen. I am grateful to Maria Niessen

and Axel Brink who kindly accepted to be my paranymphs for the public defense ceremony.

I also wish to thank the Romanian community in Groningen who always provided very useful information and much needed practical support, especially with the accommodation at the difficult beginning: Mihai Popinciuc, Gabriel Blaj, Bogdan Craus and the families Didraga, Grigorescu, Chezan and Jalba. I also wish to express my thanks to Cosmin Grigorescu and Andrei Jalba for providing the L^AT_EX macros for producing this thesis.

I thank my former colleagues from the Biophysics Department of the University of Bucharest where I worked before coming to the Netherlands. They have provided a very warm and unforgettable environment in which I could develop my academic interests. Especially, I express my sincere gratitude to prof. Aurel Popescu who has introduced me to the fields of Neural Networks and Pattern Recognition and who has also offered me my first university position. I also thank prof. Laura Tugulea for creating the first opportunity for me to come and visit the University of Groningen.

My deepest gratitude goes to my beloved parents, Fevronia and Ilie, for their encouragement, continuous support, and especially for their hard work and unconditional love. Apart from being great parents, you have also shaped my scientific identity through your personal example.

Special thanks go my daughter Ioana who has, reluctantly, accepted all my absent minded episodes and always kept up a good spirit. I also thank my son Stefan for being such a lovely baby and rewarding me with a big smile when I return home after work. Last but not least, I am grateful to my wife Monica for her inexhaustible love and patience and for keeping our family functioning throughout all these busy PhD years. Dear Monica, you have critically listened to all my theories as a true scientific companion and gracefully helped me take the right decisions.

Marius Bulacu
Groningen
December 2006

*Look but coldly on it all,
Should they praise or should they jeer;
Waves that leap like waves must fall,
Do not hope and do not fear.
You imagine and construe
What is well and what is ill;
All is old and all is new,
Days go past and days come still.*

Mihai Eminescu (Translated by Corneliu M. Popescu)

