Writer Style from Oriented Edge Fragments

Marius Bulacu and Lambert Schomaker

AI Institute, Groningen University, The Netherlands
{bulacu, schomaker}@ai.rug.nl

Abstract. In this paper we evaluate the performance of edge-based directional probability distributions extracted from handwriting images as features in forensic writer identification in comparison to a number of non-angular features. We compare the performances of the features on lowercase and uppercase handwriting. 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.

1 Introduction

This paper deals with the problem of writer identification from scanned images of handwriting. Image-based (off-line) writer identification has its principal application mainly confined to the forensic area. 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 vastly 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). As a rule of thumb, in forensic writer identification one strives for close to 100% recall of the correct writer in a hit list of 100 writers, computed on 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.

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 enhancement of these variations, which are characteristic to a writer's hand. At the same time, such features should,

[0] This paper 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

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.

Three categories of image-based features are usually integrated in operational forensic writer identification systems: 1) features extracted automatically on regions of interest from the script image, 2) features measured manually by forensic experts, and 3) character-based features capturing allograph-shape information. 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.

We analyze in this paper only category 1: features automatically extractable from the handwriting image without any human intervention. 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.

In this paper 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.

2 Data

We conducted our study using the *Firemaker* dataset [1]. A number of 250 Dutch subjects, predominantly students, were required to write 4 different A4 pages. 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 practical reasons, lineation guidelines were used on the response sheets using a special color "invisible" to the scanner. The added drawback is that vertical line distance can not be used as a discriminatory writer characteristic. However, we gain two important advantages that we will effectively use: 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. Recording conditions were standardized: the same kind of paper, ballpoint pen and support were used for all subjects. As a consequence, this also implies that the variations in ink-trace thickness and blackness will be more due to writer differences than due to the recording conditions. The response sheets were scanned with an industrial quality scanner at 300 dpi, 8 bit/pixel, gray-scale.

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 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

Table 1. Features used for writer identification and the used distance function $\Delta(\boldsymbol{u}, \boldsymbol{v})$ between a query sample $\boldsymbol{u}$ and a database sample $\boldsymbol{v}$. All features are computed in two scenarios "entire-lines" and "split-lines" (see text for details)

	Feature	Explanation	Dimensions		$\Delta(\boldsymbol{u}, \boldsymbol{v})$
			entire	split	
f1	$p(\phi)$	Edge-direction PDF	16	32	χ^2
f2	$p(\phi_1, \phi_2)$	Edge-hinge PDF	464	928	χ^2
f3	$p(rl)$	Horiz. run-length on background PDF	100	200	EUCLID
f4	$p(\phi_1, \phi_3)$	Horiz. edge-angle co-occurrence PDF	256	512	χ^2
f5	$p(brush)$	Ink-density PDF	225	450	χ^2

that the use of PDFs is a sensitive and effective way of representing a writer's uniqueness [2]. Another important advantage of using PDFs is that they allows 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 1 shows the features and the corresponding best-performing distance measures used in nearest-neighbor matching.

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 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 (fig. 1b). 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 describe further the extraction methods for the five considered features.

3.1 Edge-direction distribution (f1)

It has long been known from on-line handwriting research [3] that the distribution of directions in handwritten traces, as a polar plot, yields useful information for writer identification or coarse writing-style classification [4].

We recently developed an off-line and edge-based version of the directional distribution. Computation of this feature starts with conventional edge detection: convolution with two orthogonal differential kernels (Sobel), followed by thresholding. This procedure 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 the logical AND operator, in all directions emerging from the central pixel and ending on the periphery of the neighborhood for the presence of an entire edge fragment (fig. 1a). All the verified instances are counted into a histogram that is 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. The orientation is quantized in n directions, n being the number of bins in the histogram and

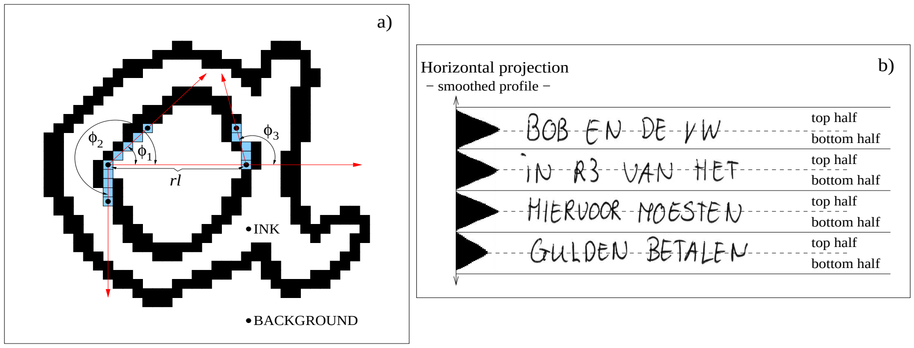

Fig. 1. a) Feature extraction on letter "a", b) Line segmentation and splitting

the dimensionality of the feature vector (see [2] for a more detailed description of the method). A number $n = 16$ directions performed best and will be used in the sequel.

As can be seen in fig. 2, the predominant direction in $p(\phi)$ corresponds, as expected, to the slant of writing. 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 [3]. Even if idealized, the example shown can provide an idea about the "within-writer" variability and "between-writer" variability in the feature space.

3.2 Edge-hinge distribution (f2)

In order to capture the curvature of the ink trace, which is very discriminatory between different writers, we designed a novel feature using local angles along the edges. Computation of this feature 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 edge fragments. All instances found in the image are counted and 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 angles ϕ_1 and ϕ_2 respectively. Orientation is quantized in $2n$ directions for every leg of the "edge-hinge". From the total number of combinations of two angles ($4n^2$) we will consider only non-redundant ones ($\phi_2 > \phi_1$) and we will also eliminate the cases when the ending pixels have a common side (see [2] for a more detailed description of the method). The final number of combinations is $C^2_{2n} - n = n(2n-3)$. For $n = 16$, the edge-hinge feature vector will have 464 dimensions.

3.3 Run-length distributions (f3)

Run-lengths have long been used for writer identification. They are determined on the binarized image taking into consideration either the black pixels (the ink) or, more beneficially, the white pixels (the background). There are two basic scanning methods: horizontal along the rows of the image and vertical along the columns of the image.

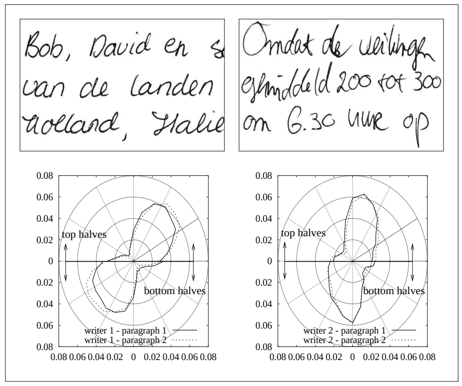

Fig. 2. Examples of lowercase handwriting from two different subjects. We superposed the polar diagrams of the "split-line" direction distribution $p(\phi)$ extracted from the two lowercase handwriting samples for each of the two subjects

Similar to the edge-based directional features presented above, the histogram of run-lengths is normalized and interpreted as a probability distribution. 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 [2] as the vertical component of handwriting strokes carries more information than the horizontal one [3]. Our particular implementation considers only 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.

3.4 Horizontal co-occurrence of edge angles (f4)

This new feature that we recently developed 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. 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.

3.5 Brush function: Ink density PDF (f5)

It is known that axial pen force ('pressure') is a highly informative signal in on-line writer identification [5]. 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 [6], which are due to ink-depositing processes. In order to capture the statistics of this process, we designed another new feature. A convolution window of 15x15 pixels was 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. After scanning

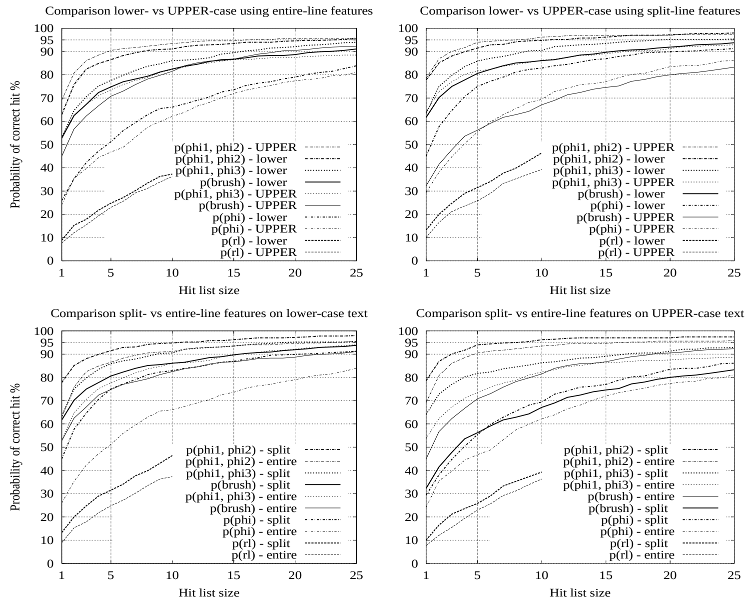

Fig. 3. Performance curves (features are ordered with the most effective at the top)

all the image, the accumulator window is normalized, yielding a PDF describing ink distribution. This feature is clearly not size invariant (the window of 15^2 pixels was chosen for capturing the 6-7 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.

The edge-based features (f1, f2, f4) that we propose here for writer identification are general texture descriptors and, as such, they have wider applicability (e.g., we use then for the analysis of machine-print as well). However, a more detailed discussion can not be encompassed in the framework of the present paper.

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 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). Using nearest-neighbor matching 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 2.

Table 2. Writer identification accuracy (in percentages) on the Firemaker data set (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	f1: $p(\phi)$		f2: $p(\phi_1, \phi_2)$		f3: $p(rl)$		f4: $p(\phi_1, \phi_3)$		f5: $p(brush)$	
size	entire	split	entire	split	entire	split	entire	split	entire	split
1	26 / **24**	45 / **29**	63 / **69**	78 / **79**	9 / **8**	13 / **10**	53 / **54**	64 / **64**	53 / **45**	62 / **32**
2	35 / **36**	58 / **38**	76 / **81**	85 / **87**	15 / **12**	20 / **17**	65 / **62**	75 / **73**	63 / **57**	70 / **42**
3	42 / **40**	65 / **45**	83 / **86**	88 / **90**	18 / **16**	25 / **21**	71 / **67**	80 / **77**	67 / **62**	75 / **48**
4	47 / **44**	71 / **50**	85 / **89**	90 / **92**	22 / **19**	29 / **24**	75 / **71**	84 / **80**	72 / **67**	78 / **54**
5	51 / **47**	75 / **55**	87 / **91**	92 / **94**	25 / **23**	32 / **26**	78 / **74**	86 / **82**	75 / **71**	81 / **56**
10	66 / **62**	83 / **69**	91 / **94**	95 / **96**	37 / **36**	46 / **39**	86 / **82**	91 / **86**	83 / **82**	86 / **67**

4.1 Comparison lower- vs upper-case and entire- vs split-line

The performance curves have been drawn in fig. 3 to allow a quick visual cross-comparison. There are important differences in performance for the different features. The edge-hinge feature (f2) 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 important 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. In mixed searches (e.g. lowercase query sample / uppercase dataset) writer identification 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 (f5) 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 promising way of improving writer identification accuracy.

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 feature (f2) was the "min" method (results in Table 3). In this method, the decision of the voter (feature) giving the lowest rank is considered as the final decision.

Table 3. Writer identification accuracy (in percentages) after feature combination using the Borda "min" voting method. Please refer to Table 2 for more details.

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

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 [7] 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.

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 content of the sample databases. Nevertheless, our future research interest will include also parameter-greedy methods (e.g. multi-layer perceptron or support vector machine) as more data necessary to train the system becomes available. Although results are far from the requirements in the forensic application domain, it quite evident that global features extracted from the handwriting image will never suffice in writer identification. Detailed character shape knowledge is needed as well. In this respect, it is important to note also the recent advances [8] that have been made at the detailed allographic level, when character segmentation is performed by hand. Only a combination of features at trace-level, allograph-level and text-line-level [9] will yield adequate results.

References

1. Schomaker, L., Vuurpijl, L.: Forensic writer identification [internal report for the Netherlands Forensic Institute]. Technical report, Nijmegen: NICI (2000)
2. Bulacu, M., Schomaker, L., Vuurpijl, L.: Writer identification using edge-based directional features. In: Proc. of ICDAR 2003 [accepted]. (2003)
3. Maarse, F., Thomassen, A.: Produced and perceived writing slant: differences between up and down strokes. Acta Psychologica **54** (1983) 131–147
4. Crettez, J.P.: A set of handwriting families: style recognition. In: Proc. of ICDAR 1995, Montreal, IEEE Computer Society (1995) 489–494
5. Schomaker, L.R.B., Plamondon, R.: The Relation between Pen Force and Pen-Point Kinematics in Handwriting. Biological Cybernetics **63** (1990) 277–289

6. Doermann, D., Rosenfeld, A.: Recovery of temporal information from static images of handwriting. In: Proc. of CVPR92. (1992) 162–168
7. Schomaker, L., Bulacu, M., van Erp, M.: Sparse-parametric writer identification using heterogeneous feature groups. In: Proc. of ICIP 2003 [accepted]. (2003)
8. Srihari, S., Cha, S., Arora, H., Lee, S.: Individuality of handwriting. Journal of Forensic Sciences **47** (2002) 1–17
9. Marti, U.V., Messerli, R., Bunke, H.: Writer identification using text line based features. In: Proc. of ICDAR'01, Seattle, IEEE Computer Society (2001) 101–105