Writer Identification Using Edge-Based Directional Features

Marius Bulacu Lambert Schomaker
AI Institute, Groningen University
The Netherlands
(bulacu, schomaker)@ai.rug.nl

Louis Vuurpijl
NICI, Nijmegen University
The Netherlands
vuurpijl@nici.kun.nl

Abstract

This paper evaluates the performance of edge-based directional probability distributions as features in writer identification in comparison to a number of non-angular features. It is noted that the joint probability distribution of the angle combination of two "hinged" edge fragments outperforms all other individual features. Combining features may improve the performance. Limitations of the method pertain to the amount of handwritten material needed in order to obtain reliable distribution estimates. The global features treated in this study are sensitive to major style variation (upper- vs lower case), slant, and forged styles, which necessitates the use of other features in realistic forensic writer identification procedures.

1. Introduction

In the process of 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 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. In the extreme case, only a single word or a signature should suffice to identify the writer. Three groups of features can be identified in forensic writer identification: 1) global measures, computed automatically on a region of interest (ROI); 2) local measures, of layout and spacing features entered by human experts, and 3) measures related to individual character shapes. We analyze in this paper 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 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 [8]. In the spatial domain, a natural coding, therefore, is expressed by angular information along the handwritten curve [5]. It has long been known [4, 3, 2] 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 paper 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.

2. Data

We evaluated the effectiveness of different features in terms of writer identification using the *Firemaker* data set [7]. 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 page 2 they were asked to describe the content of a given cartoon in their own words. Pages 3 and 4 of this database contain upper case and forged-style samples and are not used here. Lineation 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 is has the same sensed

[0]This paper was published as: Marius Bulacu, Lambert Schomaker, Louis Vuurpijl, *Writer identifi cation using edge-based directional features*, Proc. of 7th Int. Conf. on Document Analysis and Recognition (ICDAR 2003), IEEE Press, 2003, pp. 937-941, vol. II, 3-6 August, Edinburgh, Scotland

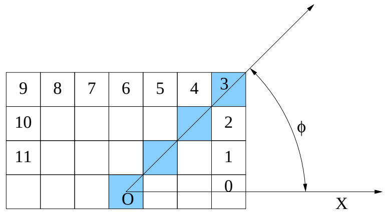

Figure 1. Extraction of edge-direction distribution.

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).

3. Features

In this section we describe the extraction methods for five 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 in this paper.

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 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 (see fig. 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. 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.

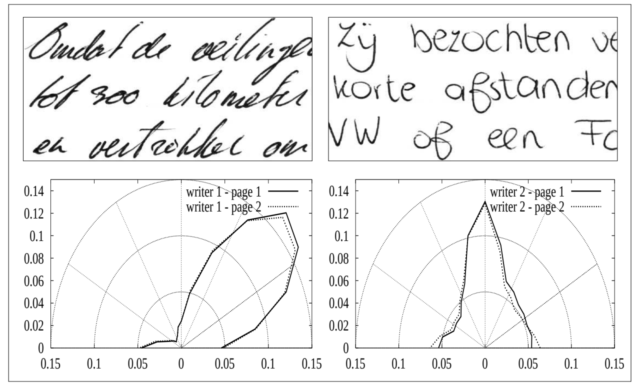

Figure 2. Two handwriting samples from two different subjects. We superposed the polar diagrams of the edge-direction distribution $p(\phi)$ corresponding to pages 1 and 2 contributed to our data set by each of the two subjects.

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 [4] to describe differences between upward and downward strokes. It was also used off-line [2] 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, 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.

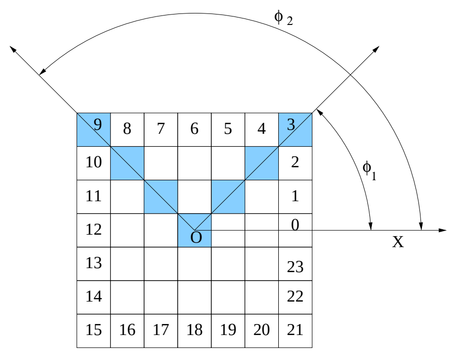

Figure 3. Extraction of edge-hinge distribution.

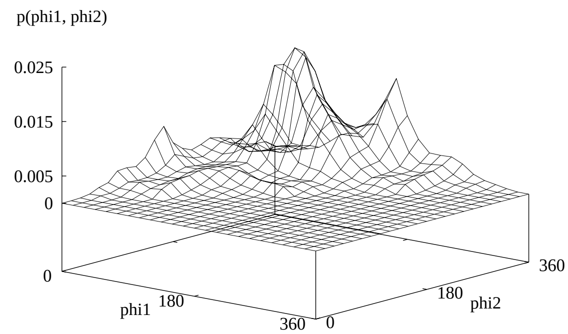

Figure 4. Graphical representation of the edge-hinge probability distribution. One half of the 3D plot (situated on one side of the main diagonal) is flat because we only consider the angle combinations with $\phi_2 > \phi_1$.

A new feature: edge-hinge distribution

Our goal is to generate 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 neighbourhood, 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 picture 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 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. 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.

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

Run-length distributions

Run lengths, first proposed for writer identification by Arazi [1], 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 data set is of about 120 pixels).

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

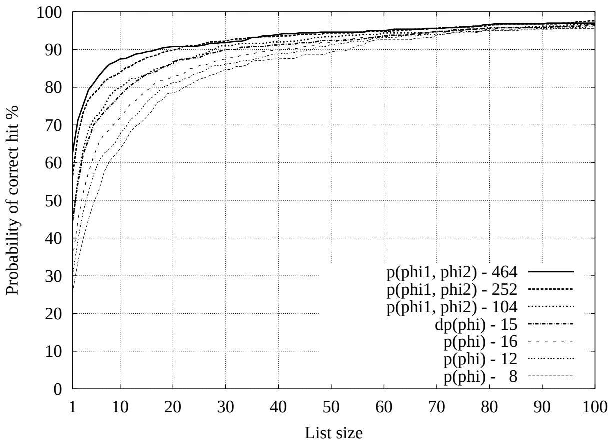

Figure 5. Performance curves of the edge-based features $p(\phi)$ and $p(\phi_1, \phi_2)$ for different direction quantizations (features are ordered with most effective at the top).

List size	$p(\phi)$ *8*	*12*	*16*	$dp(\phi)$ *15*	$p(\phi_1, \phi_2)$ *104*	*252*	*464*	*comb. 564*
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

Table 1. Writer identification accuracy (in percentages) on the Firemaker data set (250 writers, 2 pages / writer).

equal to integer multiples of the local wavelength. This results in a large dot product contribution to the final histogram.

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.

4. Results

Evaluation method

The efficacy of the considered features has been evaluated using nearest-neighbor classification in a leave-one-out strategy. Explicitly, one page is chosen and extracted from the total of 500 pages (notice that the data set 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. 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.

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 [6] for results on 10 writers, 15 documents / writer). As a consequence of these circumstances our results are more conservative.

Analysis of performances

We present the performance curves of the edge-based directional features in fig. 5 and the numerical values in Table 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.

Examination of the family of curves in fig. 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 6 gives a general overview of the comparative performance for all the features considered in this paper.

The edge-based directional features perform significantly better than the other features because they give a

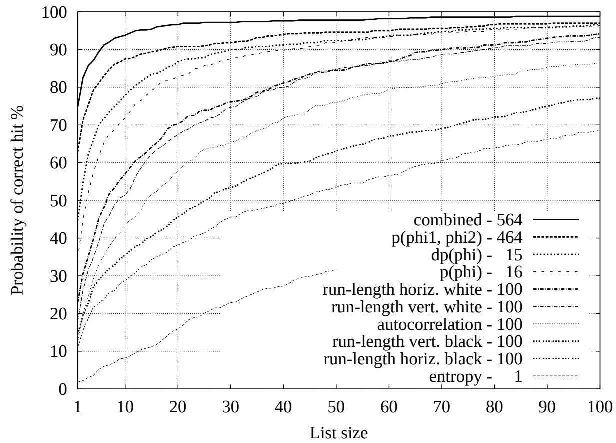

Figure 6. Performance curves for the evaluated features (features are ordered with the most effective at the top).

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 [4].

The presented features are not totally orthogonal, but nevertheless they do offer different points of view on our data set. It is therefore natural to try to combine them for improving the accuracy of writer identification. This topic will be a major interest in our future research. We will present here though, for exemplification, some initial 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 1).

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 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.

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

Table 2. Feature performance degradation with decreasing amounts of written text (writer identification accuracy in percentages for list size = 10).

5. Conclusion

We proposed a new edge-based feature for writer identification that characterizes the changes in direction undertaken during writing. It 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 results reported here, based on a very clean data set, have an academic relevance as to the usefulness of different features and they do not fully address problems like size invariance, non-uniform background, degraded documents.

Our future research interest will focus on reducing the dimensionality of the feature vectors while still keeping their discriminatory power and on further improving performance by combining different features in order to exploit their intrinsic degree of orthogonality.

References

[1] B. Arazi. Handwriting identification by means of run-length measurements. *IEEE Trans. Syst., Man and Cybernetics*, SMC-7(12):878–881, 1977.

[2] J.-P. Crettez. A set of handwriting families: style recognition. In *Proc. of the Third International Conference on Document Analysis and Recognition*, pages 489–494, Montreal, August 1995. IEEE Computer Society Press.

[3] F. Maarse, L. Schomaker, and H.-L. Teulings. Automatic identification of writers. In G. van der Veer and G. Mulder, editors, *Human-Computer Interaction: Psychonomic Aspects*, pages 353–360. Springer, New York, 1988.

[4] F. Maarse and A. Thomassen. Produced and perceived writing slant: differences between up and down strokes. *Acta Psychologica*, 54(1-3):131–147, 1983.

[5] R. Plamondon and F. Maarse. An evaluation of motor models of handwriting. *IEEE Trans. Syst. Man Cybern*, 19:1060–1072, 1989.

[6] H. Said, G. Peake, T. Tan, and K. Baker. Writer identification from non-uniformly skewed handwriting images. In *Proc. of the 9th British Machine Vision Conference*, pages 478–487, 1998.

[7] L. Schomaker and L. Vuurpijl. Forensic writer identification: A benchmark data set and a comparison of two systems [inter-

nal report for the Netherlands Forensic Institute]. Technical report, Nijmegen: NICI, 2000.

[8] L. R. B. Schomaker, A. J. W. M. Thomassen, and H.-L. Teulings. A computational model of cursive handwriting. In R. Plamondon, C. Y. Suen, and M. L. Simner, editors, *Computer Recognition and Human Production of Handwriting*, pages 153–177. Singapore: World Scientific, 1989.