University of Groningen

Co-occurrence features for writer identification

He, Sheng; Schomaker, Lambert

Published in:
Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR

DOI:
[10.1109/ICFHR.2016.0027](https://doi.org/10.1109/ICFHR.2016.0027)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

He, S., & Schomaker, L. (2017). Co-occurrence features for writer identification. In *Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR* (pp. 78-83). (Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR). Institute of Electrical and Electronics Engineers Inc.. <https://doi.org/10.1109/ICFHR.2016.0027>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Co-occurrence features for writer identification

Sheng He, Lambert Schomaker

Institute of Artificial Intelligence and Cognitive Engineering (ALICE)

University of Groningen, The Netherlands

Email: heshengxgd@gmail.com, L.Schomaker@ai.rug.nl

Abstract—In this paper, we propose two novel textural-based features for writer identification: CoHinge and QuadHinge which are based on the spatial and attribute co-occurrence of the Hinge kernel. The CoHinge feature is the joint distribution of the Hinge kernel on two different pixels of writing contours and the QuadHinge feature is the joint distribution of angles and curvature information of contour fragments. We evaluate the proposed features on five benchmark data sets and their combined large set and the experimental results demonstrate the discriminative and powerful of the proposed features.

Keywords—Co-occurrence features, writer identification, writer retrieval

I. INTRODUCTION

Writer identification is an important problem for handwritten document understanding and forensic application. In historical document analysis, recognizing the writers of historical materials is helpful to understand their historical context [30]. In addition, it is of interest in discovering the writing styles of the authors of historical documents for historical document dating [14], [15]. As a personal behavioral biometric, handwriting reflects the author's unique writer style, which can be used to determine the authorship of a handwritten document by forensic experts [31]. Therefore, it attracts extensive researches [28], [11], [7], [16], [36] on different scripts, such as Arabic [9], [37], [1], English and Dutch [8], [34], Chinese [38], [39], [18], [40], Persian [19], Farsi [27] and Indic scripts [2]. Many features [36], [7], [28], [1], [18] have been developed in the last decade and achieved state-of-the-art performance.

The similarity between handwritten documents is measured by features extracted from document images. Automatic writer identification takes the basic assumption that the handwritten documents from the same author should have a high similarity value in a given feature space. From this point of view, hand-crafted features play a significant role in handwritten document analysis, especially on a small scale database. The scale of existing data sets for writer identification is quite small. For example, the largest database for writer identification is the QUWI dataset [3], which contains both Arabic and English handwriting from 1300 different writers. Therefore, manually designing effective features is still the main task for writer identification.

It has been shown that the spatial co-occurrence features are more powerful. For example, the Gray-Level

Co-occurrence Matrices (GLCM) [12] provides promising performance for writer identification and the co-occurrence Local Binary Pattern [32], [29] provides promising results for texture classification. In addition, the joint distribution of the relation between different properties (attribute co-occurrence) has specific meanings and is also discriminative. Several features have been proposed following this principle, such as the Hinge feature [8] which is the joint distribution of two orientations of two legs attached on a common end, the Quill feature [7] which is the joint distribution of the ink direction and the ink width and the oriented Basic Image Feature Columns (oBIF Columns) [28] which is the joint distribution of six Derivative-of-Gaussian filters at two scales.

Inspired by the above observation, we propose two novel features: CoHinge which is the joint distribution of the Hinge kernel [8] on two different points on the ink contours based on the **spatial** co-occurrence local features and QuadHinge which is the joint distribution of hinge angles and curvature information of contour fragments based on the **attribute** co-occurrence of the Hinge kernel [8].

II. CO-OCCURRENCE FEATURES

Handwriting documents can be considered as special textural images and textural features have been successfully applied for writer identification. Two important textural properties of writing styles of handwriting are the slant [23] and curvature [36]. The Hinge feature has been designed to capture the ink slant as well as the ink curvature based on ink edges [34] or ink contours [8]. The main idea is to consider two edges or contour fragments on a center point and compute the joint orientational probability of the two fragments [10], like a hinge lying on the edge of the ink trace.

In this section, we propose two novel features inspired by the co-occurrence features [32]: the spatial co-occurrence CoHinge feature and the attribute co-occurrence QuadHinge feature.

A. Spatial co-occurrence feature

In this section, we propose the spatial co-occurrence feature. Let x_i and x_j be two pixels on ink contours with the Manhattan distance l of the given handwritten image (see an example in Fig. 1), and $f(x_i)$ and $f(x_j)$ be the local features

Figure 1. Illustration of the spatial co-occurrence feature. x_i and x_j are two pixels on the ink contour with a Manhattan distance l .

extracted on the pixel x_i and x_j , respectively. Inspired by [32], the spatial co-occurrence feature is defined as:

$$Cof(x_i, x_j) = [f(x_i), f(x_j)]_{co} \quad (1)$$

Although any local features can be used to derive different co-occurrence features, in this paper, we take the Hinge feature [8] as an example and derive the CoHinge feature:

$$\text{CoHinge}(x_i, x_j) = [\text{Hinge}(x_i), \text{Hinge}f(x_j)]_{co} \quad (2)$$

The Hinge kernel on each pixel x_i of handwriting contours is defined as $(\alpha_{x_i}, \beta_{x_i})$, where α_{x_i} and β_{x_i} are the two angles of a hinge lied on the ink contours [8] (see Fig. 2). Therefore, the CoHinge kernel can be defined as:

$$\text{CoHinge}(x_i, x_j) = [\alpha_{x_i}, \beta_{x_i}, \alpha_{x_j}, \beta_{x_j}] \quad (3)$$

All CoHinge kernels on writing contours of the input document can be quantized into a 4-D histogram to form the feature vector. In this paper, we set the Manhattan distance l to 7 and the number of bins of each orientation to 10. Finally, the dimension of the CoHinge feature is $10^2 \times 10^2 = 10^4$.

Note that the CoHinge feature is different from our previous $\Delta^n\text{Hinge}$ feature [16]. The kernel of the $\Delta^n\text{Hinge}$ is defined as:

$$\Delta^n \text{Hinge} = \left[\frac{d^n(\alpha_{x_i}, \alpha_{x_j})}{dl^n}, \frac{d^n(\beta_{x_i}, \beta_{x_j})}{dl^n} \right] \quad (4)$$

where $\frac{d^n(\alpha_{x_i}, \alpha_{x_j})}{dl^n}$ and $\frac{d^n(\beta_{x_i}, \beta_{x_j})}{dl^n}$ are the differential operations with order n between two angles of the Hinge kernel between two pixels x_i and x_j . The $\Delta^n\text{Hinge}$ feature captures curvature information of writing contours. Therefore, it is rotation-invariant but with a lower performance. The CoHinge feature is also different from the horizontal co-occurrence of edge angles proposed in [8], which is the joint probability distribution of the Hinge kernel at both ends of a run-length on white pixels. The definition of the CoHinge feature is more general and can be easily implemented.

Figure 2. The left and right figures show two contour fragments with different fragment curvature values $\mathcal{C}(\mathcal{F}_c)$. However, their Hinge kernels are same on the center points ($\alpha_1=\alpha_2$ and $\beta_1=\beta_2$). Therefore, there is a loss of detailed curvature information (especially for long leg lengths) in the Hinge kernel.

B. Attribute co-occurrence feature

In this section, we propose the attribute co-occurrence feature. Let x_i be the pixel on ink contours and $f_1(x_i)$ and $f_2(x_i)$ are two different local features which represent different attributes on the pixel x_i . The attribute co-occurrence feature is defined as:

$$Cof(x_i) = [f_1(x_i), f_2(x_i)]_{co} \quad (5)$$

In fact, more than two attributes can be involved in the co-occurrence feature $Cof(x_i)$. In this paper, we take the Hinge feature as an example to build the attribute co-occurrence feature. The Hinge feature achieves reasonable results for writer identification [34], [8]. However, due to the fact that the length of edges or contour fragments are usually set to a small value (for example, it has been set to 5 in [8]), the Hinge captures more slant than curvature information of the ink trace (see an example in Fig. 2). In addition, only orientations of the small fragments are considered, leading to the limitations of the performance.

In order to compute the curvature information of the contour fragments in the Hinge kernel, we give the following definition for a fragment curvature measurement (FCM) $\mathcal{C}(\mathcal{F}_c)$ for contour fragments, inspired by [5], [20]:

Definition. Let $\mathcal{F}_c$ be a contour fragment on the ink trace, $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$ are the Cartesian coordinates of the two end points. Then the fragment curvature measurement $\mathcal{C}(\mathcal{F}_c)$ is defined as the proportion of the Euclidean distance $\mathbf{d}_2(p_1, p_2)$ between two end points to the length of the contour fragment s .

$$\mathcal{C}(\mathcal{F}_c) = \mathbf{d}_2(p_1, p_2)/s \quad (6)$$

where $\mathbf{d}_2(p_1, p_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$.

All pixels of contour fragments are represented in the Cartesian coordinate space and the Chebyshev distance between two neighbor pixels is equal to 1. Therefore, the fragment curvature measurement $\mathcal{C}(\mathcal{F}_c)$ has the following property: $\mathcal{C}(\mathcal{F}_c) \in (0, \sqrt{2}]$ for all contour fragments $\mathcal{F}_c$. (see examples in Fig. 3). The $\mathcal{C}(\mathcal{F}_c)$ achieves the maximal

Figure 3. Examples of the fragment curvature measurements for two contour fragments with $s=10$ (with $s+1$ pixels). The $\mathcal{C}(\mathcal{F}_c)=1.41$ for the first fragments and $\mathcal{C}(\mathcal{F}_c)=0.98$ for the second one.

value $\sqrt{2}$ when pixels form a line in the diagonal direction (see the left figure in Fig. 3).

Unlike the traditional methods which measure the curvature on a point of the curves using the first and second-regularized derivatives [26] or the integral invariants [24], the $\mathcal{C}(\mathcal{F}_c)$ measures the curvature of a contour fragment instead of the curvature on a point. The $\mathcal{C}(\mathcal{F}_c)$ is also known as the *straightness index* [5], which is the simplest one of the linearity measures of an open curve segment [33]. This method has also been used for shape classification in [20].

The original Hinge kernel only considers the orientations α and β (see Fig. 2) formed by the two contour fragments with the same Manhattan distance s on the ink contour. In this paper, we propose a novel Quadruple Hinge kernel, which integrates the $\mathcal{C}(\mathcal{F}_c)$ into the original Hinge kernel, defined as: $\mathcal{H}(p, s) = \{\alpha, \beta, \mathcal{C}(\mathcal{F}_1), \mathcal{C}(\mathcal{F}_2)\}$, where p is the center point, s is the fragment length, $\mathcal{C}(\mathcal{F}_1)$ and $\mathcal{C}(\mathcal{F}_2)$ are the fragment curvature measurements of the two contour fragments, respectively (see the examples in Fig. 2). Adding the curvature information of the two contour fragments can improve the discriminative of the Hinge kernel. For example, the Hinge kernels $\{\alpha, \beta\}$ of the left and right fragments in Fig. 2 are the same. However, the curvature of the fragments are different, yielding different Quadruple Hinge kernels. Finally, the Quadruple Hinge kernels on all contour pixels are collected and quantized into a 4-D histogram. There are two parameters involved in this processing: the number of angle bins N_a for α and β , and the number of curvature bins N_l for $\mathcal{C}(\mathcal{F}_1)$ and $\mathcal{C}(\mathcal{F}_2)$. The resulting histogram is normalized into a joint probability $p_s(\alpha, \beta, \mathcal{C}(\mathcal{F}_1), \mathcal{C}(\mathcal{F}_2))$ which is the Quadruple Hinge feature representation.

The fragment length s affects the Quadruple Hinge kernel. When s is small, the values of the $\mathcal{C}(\mathcal{F}_1)$ and $\mathcal{C}(\mathcal{F}_2)$ are approximately equal to a constant value because the fragments are more like a line, which is the case of the original Hinge kernel. However, when s is large, the Quadruple Hinge kernel reflects more curvature than the orientation information of contours because the values of α and β close to small ones. In order to balance these two aspects, it is necessary to agglomerate the Quadruple Hinge kernels with multiple scales. In this paper, the set of scales are computed

as:

$$s = s_0 * (t + 1) \quad (7)$$

where s_0 is the basic fragment length, $t = \{0, 1, \dots, T\}$ and T is the maximum of the scale. We set the basic fragment length to $s_0 = 5$, the number of angle bins to $N_a = 12$, the number of $\mathcal{C}(\mathcal{F})$ bins to $N_l = 6$ and $T = 10$. Finally, the dimension of the QuadHinge feature is $6^2 \times 12^2 = 5,184$.

III. EXPERIMENTAL RESULTS

Writer identification is performed with a one-to-many search in the database. Given the query handwritten document without the writer information, all the handwriting in the database are sorted according to the distance computed based on the given feature to form a hit list. The query document is recognized as the writer of the document on the top x of the hit list, corresponding to the Top- x performance. The χ^2 distance is applied for all features in this paper:

$$\chi^2(f, g) = \sum_{i=1}^{dims} \frac{(f_i - g_i)^2}{f_i + g_i} \quad (8)$$

where f and g are two histograms and $dims$ is the size of the histogram.

We also present a performance comparison of our proposed two features with six recent features:

Run-length histogram (RLH) [4]: The run-lengths of ink and white pixels are computed on both horizontal and vertical directions with the maximum length 100. Therefore, the dimension of RLH is $2 \times 2 \times 100 = 400$.

Run-lengths of Local Binary Pattern (LBPruns) [17] LBPruns is the run-lengths of local binary pattern computed on n parallel scanning lines with inter-line distance d on the binarized images. In this paper, we set n and d to 5 and compute the run-lengths on the horizontal and vertical directions based on binarized images with maximum length 100 and the dimension of LBPruns is: $2 \times 2^5 \times 100 = 6,400$.

Hinge [8] Hinge feature is the joint probability of two orientations of legs of a hinge kernel on the ink contours. Following the original work [8], we set the number of bins to 23 and the leg length to 7 and the feature dimension is 253.

Δ^n Hinge [16] The Δ^n Hinge kernel is defined in Equ.(4). In this paper, we only consider the Δ^1 Hinge feature, and the feature dimension is 780.

Quill [7] Quill is the joint probability distribution of the ink direction and ink width which can capture the property of the writing instrument. The parameters of Quill are set following the original paper [7] and the dimension is 1,600.

Junction feature (Junclets) [18] The Junclets feature is the stroke length distribution on 120 directions spanning all four quadrants (2π) on the junction points in the binarized images. In this paper, we train a codebook with 900 code-words and finally the dimension of the Junclets feature is 900.

Figure 4. Example handwritten documents from different data sets with the corresponding languages.

A. Data sets

In this paper, we evaluate the performance of the proposed CoHinge and QuadHinge features on five databases for writer identification, such as the Firemaker [35], IAM [25], CERUG [18], ICFHR2012 Arabic data set [13] and ICDAR2013 [22]. The Firemaker set contains four pages of handwriting written by 250 Dutch subjects: page 1 and page 4 contain the lower-case letters, page 2 was written by only uppercase letters, and Page 3 contains the “forged” text. We use the page 1 vs 4 in our experiments, similar as works in [8], [7], [18]. The IAM set contains 650 writers written in English, modified following the work [8] from the original IAM database [25]. The CERUG set is a cross-script data set, written by 105 Chinese subjects on four pages: page 1 and page 2 were written in Chinese, page 3 contains the English text and page 4 contains both Chinese and English characters. The data set used for the ICFHR2012 competition on writer identification with Arabic scripts [13] contains 204 writers and we only use the first two paragraphs to perform writer identification. The data set used for the ICDAR2013 competition on writer identification [22] contains 250 writers with four pages (2 English and 2 Greek).

B. Performance on the different data sets

Table I shows the performance of different features on the five data sets. It can be seen that the QuadHinge feature achieves the best results on the Firemaker, IAM, CERUG-Chinese data sets and the CoHinge feature provides the comparable results with QuadHinge. On the CERUG-English data set, the best results are achieved by Junclets and Δ^1 Hinge. The main reason is that the handwritten documents on the CERUG-English contain more long lines (as shown in [18]) and less curvature information. Therefore, the Junclets and Δ^1 Hinge, as well as the LBPruns, provide much better results than other textural-based features. The CoHinge feature gives the best result on Arabic handwriting from the ICFHR2012 data set, while QuadHinge provides the best result on ICDAR2013 data set, both on English and Greek handwriting. On all the five data sets with different scripts, the CoHinge and QuadHinge features achieve better

Table II
THE PERFORMANCE OF WRITER IDENTIFICATION ON THE COMBINED DATE SET.

Features	Top-1	Top-10
RLH [4]	67.6	87.8
LBPruns [17]	78.6	92.6
Hinge [8]	82.7	92.6
Δ^1 Hinge [16]	78.7	92.5
Quill [7]	85.9	93.9
Junclets [18]	84.7	94.0
CoHinge	93.8	97.6
QuadHinge	92.2	96.9

results than the Hinge feature, which demonstrates that the co-occurrence features are much more discriminative and powerful than their original ones.

C. Performance of writer identification on the combined data set

In this section, we give the performance of different features on a large and mixed data set. Following the work [8], we merge the Firemaker, IAM, CERUG, CVL, ICFHR2012 and ICDAR2013 data sets to obtain a large and combined data set which contains 5177 handwritten documents from 1760 hands with six languages: Dutch, English, Chinese, German, Arabic and Greek. Some samples are shown in Fig. 4.

Table II shows the performance of different features on this large set. The CoHinge feature achieves the best result and Top-1 accuracy is 93.8%. The QuadHinge feature also provides a comparable result with 92.2%, which is better than the Hinge and Quill features. This indicates that the handwriting styles captured by CoHinge and QuadHinge take more important information than the property of writing instruments captured by the Quill feature on this large data set where handwritten documents are written with different pens.

Table I
THE WRITER IDENTIFICATION PERFORMANCE BASED ON SINGLE-SCRIPT ON FIVE DATA SETS: FIREMAKER, IAM, CERUG, ICFHR2012 AND ICDAR2013. THE COLUMN “DIM” IS THE DIMENSION OF EACH FEATURE.

Feature	Dim	Firemaker 250 writers		IAM 650 writers		CERUG 105 writers				ICFHR2012 204 writers		ICDAR2013 250 writers			
		Dutch		English		Chinese		English		Arabic		English		Greek	
		Top-1	Top-10	Top-1	Top-10	Top-1	Top-10	Top-1	Top-10	Top-1	Top-10	Top-1	Top-10	Top-1	Top-10
RLH [4]	400	59.6	86.2	71.1	89.0	77.1	92.9	25.7	64.8	51.2	80.1	63.0	90.2	74.6	91.8
LBPPruns [17]	6,400	72.2	91.8	81.4	94.4	88.6	95.7	77.1	98.1	42.9	73.0	83.2	97.6	82.4	97.0
Hinge [8]	253	84.8	95.8	85.8	95.1	90.9	95.7	22.8	47.1	75.5	92.4	86.2	95.4	80.4	96.2
Δ^1 Hinge [16]	780	75.6	94.0	82.1	95.7	80.5	93.3	90.5	98.6	61.3	89.5	75.6	93.8	78.6	92.8
Quill [7]	1,600	69.2	86.8	89.1	95.8	87.1	92.4	28.1	67.6	90.4	99.3	92.8	97.2	95.2	97.2
Junclets [18]	900	80.2	93.4	85.8	95.5	93.3	98.1	92.9	97.1	56.4	85.5	91.0	96.2	90.4	97.2
CoHinge	10,000	91.6	96.4	92.4	96.5	95.2	98.1	42.8	78.1	93.6	99.3	93.0	96.2	93.8	98.8
QuadHinge	5,184	92.2	97.2	93.2	96.5	96.2	98.6	46.7	83.3	87.0	98.3	94.2	96.8	95.2	98.6

Table III
THE NUMBER OF QUERIES WITH DIFFERENT LINE TEXTS ON THE CERUG AND CVL DATA SETS FOR WRITER RETRIEVAL.

Data set	line 1	line 2	line 3	line 4	line 5
CERUG					
201 writers	2095	1012	617	450	345
Chinese					
CVL					
310 writers	10820	5038	3187	2118	1686
English					

Table IV
THE MAP OF WRITER RETRIEVAL OF DIFFERENT FEATURES ON THE CERUG-CHINESE DATA SET.

Feature	line 1	line 2	line 3	line 4	line 5
RLH [4]	0.282	0.459	0.572	0.663	0.717
LBPPruns [17]	0.354	0.657	0.796	0.852	0.881
Hinge [8]	0.498	0.729	0.830	0.873	0.891
Δ^1 Hinge [16]	0.248	0.492	0.633	0.719	0.762
Quill [7]	0.489	0.696	0.804	0.855	0.878
Junclets [18]	0.355	0.622	0.781	0.879	0.915
CoHinge	0.461	0.745	0.873	0.917	0.946
QuadHinge	0.504	0.770	0.888	0.925	0.950

D. Performance of writer retrieval

In this section, we perform the task of writer retrieval, which retrieves the hit list of handwritten documents from the same hand with the query document on the data set. We use the CVL data set [21], which contains handwritten documents from 310 writers, 27 of which wrote 7 texts and 283 writers have 5 texts. In this experiment, we only use English handwriting from 310 writers. In addition, we also use the Chinese handwriting on the CERUG data set. We segment the handwritten documents into text lines and conduct experiments on handwritten documents with different number of text lines from one to five for writer retrieval. Table III shows the number of queries on two data sets with different text lines.

We use the mean Average Precision (mAP) to measure

Table V
THE MAP OF WRITER RETRIEVAL OF DIFFERENT FEATURES ON THE CVL DATA SET.

Feature	line 1	line 2	line 3	line 4	line 5
RLH [4]	0.238	0.445	0.584	0.680	0.740
LBPPruns [17]	0.298	0.583	0.729	0.811	0.842
Hinge [8]	0.540	0.781	0.853	0.886	0.893
Δ^1 Hinge [16]	0.363	0.679	0.809	0.861	0.887
Quill [7]	0.569	0.828	0.896	0.925	0.933
Junclets [18]	0.372	0.675	0.818	0.892	0.920
CoHinge	0.591	0.874	0.937	0.957	0.959
QuadHinge	0.632	0.883	0.936	0.954	0.956

the performance of writer retrieval, which is defined as:

$$\text{mAP} = \frac{1}{N} \sum_{q=1}^N \text{AveP}(q) \quad (9)$$

where N is the number query samples and $\text{AveP}(q)$ is the average precision of the query q .

Table IV and V show the results on the CERUG Chinese and CVL English data sets, respectively. From the tables we can see that all of features give the reasonable results on the handwritten documents with at least three lines about 100 characters, which is the minimum amount of needed text for writer identification [6]. The QuadHinge feature provides the best performance on the CERUG-Chinese data set and CoHinge provides the best performance on the CVL data set when the number of lines is greater than 2.

IV. CONCLUSION

This paper has proposed two novel features based on the co-occurrence features for writer identification, i.e., CoHinge and QuadHinge. Our proposed spatial and attribute co-occurrence Hinge features achieve much better results than the original Hinge feature on the five data sets. In conclusion, the proposed methods are very promising for capturing the handwriting style of handwritten documents, which have many potential applications, such as historical document dating based on handwriting style analysis [14].

ACKNOWLEDGMENT

This work has been supported by the Dutch Organization for Scientific Research NWO (project No. 380-50-006). The authors would like to thank Shijie Zhao and Yanfang Feng to collect the CERUG data set.

REFERENCES

- [1] M. N. Abdi and M. Khemakhem. A model-based approach to offline text-independent arabic writer identification and verification. *Pattern Recognition*, 48(5):1890–1903, 2014.
- [2] C. Adak and B. Chaudhuri. Writer identification from offline isolated Bangla characters and numerals. In *ICDAR*, pages 486–490, 2015.
- [3] S. Al Máadeed, W. Ayoubi, A. Hassaine, and J. M. Aljaam. QUWI: an arabic and english handwriting dataset for offline writer identification. In *ICFHR*, pages 746–751. IEEE, 2012.
- [4] B. Arazi. Handwriting identification by means of run-length measurements. *IEEE Trans. Syst., Man and Cybernetics*, (12):878–881, 1977.
- [5] S. Benhamou. How to reliably estimate the tortuosity of an animal’s path:: straightness, sinuosity, or fractal dimension? *Journal of theoretical biology*, 229(2):209–220, 2004.
- [6] A. Brink, M. Bulacu, and L. Schomaker. How much handwritten text is needed for text-independent writer verification and identification. In *ICPR*, pages 1–4, 2008.
- [7] A. Brink, J. Smit, M. Bulacu, and L. Schomaker. Writer identification using directional ink-trace width measurements. *Pattern Recognition*, 45(1):162–171, 2012.
- [8] M. Bulacu and L. Schomaker. Text-independent writer identification and verification using textural and allographic features. *TPAMI*, 29(4):701–717, 2007.
- [9] M. Bulacu, L. Schomaker, and A. Brink. Text-independent writer identification and verification on offline Arabic handwriting. In *ICDAR*, pages 769–773, 2007.
- [10] M. Bulacu, L. Schomaker, and L. Vuurpijl. Writer identification using edge-based directional features. In *ICDAR*, pages 937–941, 2003.
- [11] G. Ghiasi and R. Safabakhsh. Offline text-independent writer identification using codebook and efficient code extraction methods. *Image and Vision Computing*, 31(5):379–391, 2013.
- [12] R. K. Hanusiak, L. S. Oliveira, E. Justino, and R. Sabourin. Writer verification using texture-based features. *IJDAR*, 15(3):213–226, 2012.
- [13] A. Hassaine and S. A. Maadeed. ICFHR 2012 competition on writer identification challenge 2: Arabic scripts. In *ICFHR*, pages 835–840, 2012.
- [14] S. He, P. Samara, J. Burgers, and L. Schomaker. Towards style-based dating of historical documents. In *ICFHR 2014*, pages 265–270.
- [15] S. He, P. Samara, J. Burgers, and L. Schomaker. Image-based historical manuscript dating using contour and stroke fragments. *Pattern Recognition*, 58:159–171, 2016.
- [16] S. He and L. Schomaker. Delta-n hinge: Rotation-invariant features for writer identification. In *ICPR*, pages 2023–2028. IEEE, 2014.
- [17] S. He and L. Schomaker. General pattern run-length transform for writer identification. In *DAS*, 2016.
- [18] S. He, M. Wiering, and L. Schomaker. Junction detection in handwritten documents and its application to writer identification. *Pattern Recognition*, 48(12):4036–4048, 2015.
- [19] B. Helli and M. E. Moghaddam. A text-independent Persian writer identification based on feature relation graph (FRG). *Pattern Recognition*, 43(6):2199–2209, 2010.
- [20] J. J. d. M. S. Junior and A. R. Backes. Shape classification using line segment statistics. *Information Sciences*, 305:349–356, 2015.
- [21] F. Kleber, S. Fiel, M. Diem, and R. Sablatnig. CVL-database: An off-line database for writer retrieval, writer identification and word spotting. In *ICDAR*, pages 560–564, 2013.
- [22] G. Louloudis, B. Gatos, N. Stamatopoulos, and A. Papan-dreou. ICDAR 2013 competition on writer identification. In *ICDAR*, pages 1397–1401, 2013.
- [23] F. J. Maarse and A. J. Thomassen. Produced and perceived writing slant: Difference between up and down strokes. *Acta psychologica*, 54(1):131–147, 1983.
- [24] S. Manay, D. Cremers, B.-W. Hong, A. J. Yezzi, and S. Soatto. Integral invariants for shape matching. *TPAMI*, 28(10):1602–1618, 2006.
- [25] U.-V. Marti and H. Bunke. The IAM-database: an English sentence database for offline handwriting recognition. *IJDAR*, 5(1):39–46, 2002.
- [26] F. Mokhtarian and A. K. Mackworth. A theory of multi-scale, curvature-based shape representation for planar curves. *TPAMI*, (8):789–805, 1992.
- [27] F. S. Nejad and M. Rahmati. A new method for writer identification and verification based on Farsi/Arabic handwritten texts. In *ICDAR*, pages 829–833, 2007.
- [28] A. J. Newell and L. D. Griffin. Writer identification using oriented basic image features and the delta encoding. *Pattern Recognition*, 47(6):2255–2265, 2014.
- [29] R. Nosaka, Y. Ohkawa, and K. Fukui. Feature extraction based on co-occurrence of adjacent local binary patterns. In *Advances in image and video technology*, pages 82–91. Springer, 2011.
- [30] M. Panagopoulos, C. Papaodysseus, P. Rousopoulos, D. Dafi, and S. Tracy. Automatic writer identification of ancient greek inscriptions. *TPAMI*, 31(8):1404–1414, 2009.
- [31] V. Pervouchine and G. Leedham. Extraction and analysis of forensic document examiner features used for writer identification. *Pattern Recognition*, 40(3):1004–1013, 2007.
- [32] X. Qi, R. Xiao, C.-G. Li, Y. Qiao, J. Guo, and X. Tang. Pairwise rotation invariant co-occurrence local binary pattern. *TPAMI*, 36(11):2199–2213, 2014.
- [33] P. L. Rosin, J. Pantović, and J. Žunić. Measuring linearity of curves in 2d and 3d. *Pattern Recognition*, 2015.
- [34] L. Schomaker and M. Bulacu. Automatic writer identification using connected-component contours and edge-based features of uppercase western script. *PAMI*, 26(6), 2004.
- [35] L. Schomaker and L. Vuurpijl. Forensic writer identification: a benchmark data set and a comparison of two systems. *Technical Report, Nijmegen: NICI*, 2000, 2000.
- [36] I. Siddiqi and N. Vincent. Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. *Pattern Recognition*, 43(11):3853–3865, 2010.
- [37] S. N. Srihari and G. R. Ball. Writer verification of Arabic handwriting. In *DAS*, pages 28–34, 2008.
- [38] J. Wen, B. Fang, J. Chen, Y. Tang, and H. Chen. Fragmented edge structure coding for Chinese writer identification. *Neurocomputing*, 86:45–51, 2012.
- [39] X. Wu, Y. Tang, and W. Bu. Offline text-independent writer identification based on scale invariant feature transform. *IEEE Transactions on Information Forensics and Security*, 9(3):526–536, 2014.
- [40] W. Yang, L. Jin, and M. Liu. Chinese character-level writer identification using path signature feature, dropstroke and deep CNN. In *ICDAR*, pages 546–550, 2015.