Simple and Fast Geometrical Descriptors for Writer Identification

Angelika Garz*[1] *, angelika.garz@unifr.ch
Marcel Würsch*[1] *, marcel.wuersch@unifr.ch
Andreas Fischer*[1,2] *, andreas.fischer@unifr.ch
Rolf Ingold*[1] *, rolf.ingold@unifr.ch

[1] ***Document, Image, and Voice Analysis Group, University of Fribourg, 1700 Fribourg, Switzerland***
[2] ***iCoSys Institute, University of Applied Sciences and Arts Western Switzerland, 1700 Fribourg, Switzerland***

Abstract
Recent advances in writer identification push the limits by using increasingly complex methods relying on sophisticated preprocessing, or the combination of already complex descriptors. In this paper, we pursue a simpler and faster approach to writer identification, introducing novel descriptors computed from the geometrical arrangement of interest points at different scales. They capture orientation distributions and geometrical relationships of script parts such as strokes, junctions, endings, and loops. Thus, we avoid a fixed set of character appearances as in standard codebook-based methods. The proposed descriptors significantly cut down processing time compared to existing methods, are simple and efficient, and can be applied out-of-the-box to an unseen dataset. Evaluations on widely-used datasets show their potential when applied by themselves, and in combination with other descriptors. Limitations of our method relate to the amount of data needed to obtain reliable models.

Introduction
Handwriting is considered a behavioral biometric feature unique to a person [1]: a person can be identified by their writing style which is characterized by the reproduction of certain recurrent patterns and unconscious practices inherited and adopted culturally [2]. However, unlike other biometric features such as DNA, or fingerprints, which do not change throughout a person's life, handwriting is subject to stochastic natural variation, i.e., no two writings of a single writer are identical [3]. Sources of natural variations include the writer's age, their personal condition, the materials used, the environmental writing conditions, and especially skill and practice [3].

While the natural variation of handwriting potentially leads to a different overall appearance of a document, a writer still uses the same adopted patterns and unconscious practices. This fact is exploited by forensic document examiners, who provide legal evidence of authorship or authenticity on a questioned document in court of law [4]. However, the US National Research Council criticized the current practice in forensic document examination as subjective [5,6]. They express the need for quantifiable measures: "*the scientific basis for handwriting comparisons needs to be strengthened [...] there has been only limited research to quantify the reliability and replicability of the practices used by trained document examiners*" [5]. Automatic writer identification systems can offer objective methods and quantifiable measures.

Writer identification in this context refers to determining a document's writer from a given list of samples of known authorship (1:N comparison). Figure 1 visualizes this concept: a query document (left) is given to the automatic identification system (center), which has access to a database of samples of known writers. The identification system extracts features from the query document and compares them with the database. According to some

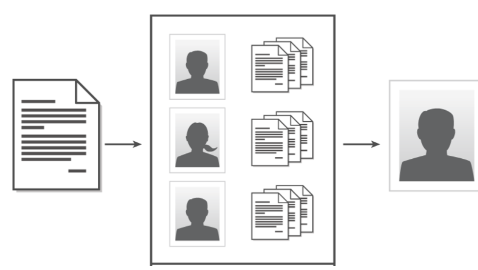


Figure 1. *Concept of writer identification (Figure taken from [7])*

similarity measurement the system returns the most likely author of the query document (right).

Other fields of applications of writer identification include historical document analysis and handwriting recognition. In historical document analysis, knowledge of the writer can be used to classify and authenticate a manuscript, establish its origin, or identify the number of contributors [8–10]. Handwriting recognition systems benefit indirectly from writer identification: knowledge of the writer allows for writer-specific models for recognition rather than a generic model [11,12].

Current attempts in writer identification push the performance by increasing complexity, either by combining already complex descriptors [13,14], or introducing additional (pre)processing steps and heuristics [15,16]. In this paper, we pursue a simpler and faster approach to writer identification, introducing novel descriptors computed from the geometrical Interest Point (IP) arrangement at different scales. IPs offer a multi-scale sparse representation of the image as they are only detected at locations of a 2D signal change in the image. On document images, IPs correspond to small script parts such as stroke sections, junctions, and endings, up to bigger structures like loops, characters, or spaces. Based on the IPs we introduce a set of descriptors that capture the distribution of elementary script characteristics: orientations (angles) of script parts and their geometrical relationships (angles between script parts).

Our approach has two characteristics distinguishing it from existing work: firstly, we do not require image preprocessing such as binarization or segmentation – open research problems themselves which potentially introduce errors to the identification process [17]; and secondly, our features are low-dimensional, simple, and fast to compute, reducing the complexity when compared with existing state-of-the-art systems.

This paper is an extension of previous work [18], where we have firstly introduced scale-specific IP orientations for writer identification, putting a focus on human interpretability of the results by forensic experts. Three additional contributions are presented in this paper. First, we provide a more detailed description of the method and derive different geometrical IP descriptors in a systematic way. Secondly, we conduct a more comprehensive and more detailed experimental evaluation on several benchmark datasets. The proposed descriptors are compared with different state-of-the-art methods. Thirdly, we combine our descriptors with others to improve the performance and to demonstrate that they capture complementary information.

The remainder of this paper is structured as follows. The next section gives an overview of selected related work, followed by the definition of the proposed set of descriptors. The datasets, evaluation procedure, results, and speed assessment are reported in the last section. The paper closes with a conclusion and an outlook to future work.

Related Work

Writer identification methods can broadly be split into two categories: text-dependent and text-independent methods [19,20]. The former rely on the comparison of individual character or word images with known textual content, and require exact localization and segmentation of the respective entities. The latter extracts statistical features from a segmented text block. In order to achieve independence of the textual content, a minimal amount of text is needed [21]. Text-independent methods have the advantage that the identification task is performed without the need of handwriting recognition, or interaction of a user transcribing and annotating character images.

Several comprehensive surveys provide a broad overview of the efforts done in text-dependent [12,22–26] and text-independent writer identification [2,12]. In the following, we summarize selected related work on text-independent writer identification, which we broadly classify into texture-, structural-, and allograph-based methods. Finally, we draw a focus on recent work on IP-based writer identification.

Approaches based on texture analysis consider a document image simply as an image. Features are extracted globally from image patches in writing areas, e.g., Gabor features [19], angular histograms [27] capturing local stroke directions, or combinations which cover slant and curvature [20]. Newel and Griffin [28] propose a writer identification system using oriented Basic Image Feature Columns (oBIF Columns), a texture-based scheme previously used for character recognition.

Minor changes in writing styles such as differences in word- and line spacing, and strokes having different thickness alter texture, and thus, pose problems to texture-based methods.

Structural methods attempt to capture script properties such as heights of writing zones, the slope, and white runs. They are predominantly extracted from Projected Profiles (PPs) and Connected Components (CCs), which require prior binarization and are problematic in case of touching components of consecutive lines. Marti et al. [29] report a method based on twelve features extracted from PP, CC, and white runs of binarized

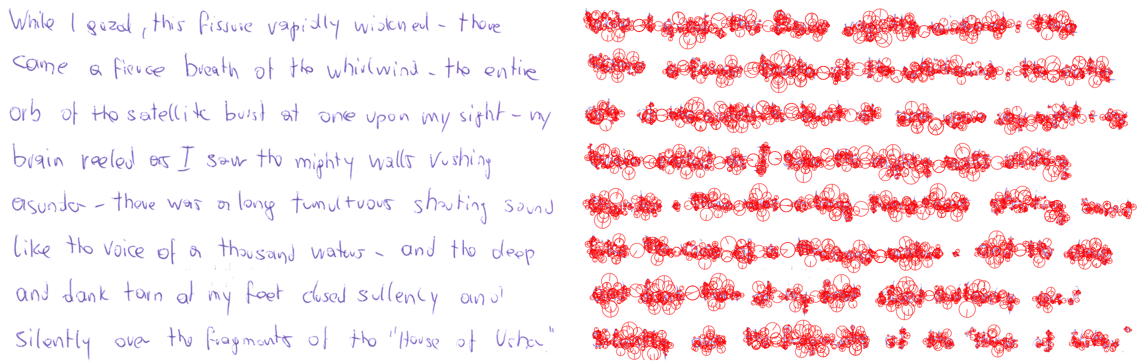


Figure 2. *Original document image (left), and overlaid with the IP detected (right) that encode the information in a sparse manner: IP are found on the writing and in spaces between characters and words. However, no IP are detected on the background between lines or in the document margin.*

segmented text lines. Schlapbach and Bunke [30] propose a stochastic approach using a series of Hidden Markov Model (HMM)-based handwriting recognizers and Gaussian Mixture Models (GMMs), where for each scribe exactly one model is trained. The most likely author is determined by the log-likelihood scores of the recognizers.

Allograph-based methods segment handwriting into characters and extract features from contours or shapes both require binarization and segmentation, still open topics in research [17]. Bulacu and Schomaker [20] introduce a method which combines textural [27] with allographic features [31]. Bensefia et al. [32] characterize handwriting based on a set of invariant features extracted from parts of characters segmented from CC based on the analysis of minima of upper contours of binarized text. Niels et al. [33] hierarchically cluster manually segmented allographs by Dynamic Time Warping. A histogram of the frequency of occurrence of allographs in relation to each character is created, which represents a prototype of a writer's style.

Approaches based on Neural Networks, HMMs, or a set of shapes (e.g., codebooks, allographs) risk having an average yet incomplete representation; however, brute force methods which consider all possible character shapes tend to be computationally over-expensive [34].

Asserting writer identity based on IPs avoids binarization and segmentation required by other approaches. IPs offer the advantage of quickly narrowing down processing to areas with relevant information (c.f. Figure 2). Woodward et al. [35] first proposed to use an occurrence histogram of vector-quantized IP descriptors for writer identification, relying on probabilistic latent semantic analysis (pLSA) to identify the writer.

A codebook of Scale Invariant Feature Transform (SIFT)-descriptors created with k-means is used in a method based on Bag-of-SIFT-Words (BoS) [7]. With respect to the codebook, a histogram of occurrences is created which characterizes a writer. Using the χ^2 distance as similarity measure between histograms, the authors report a 90.8 % top-1 and 97.5 % top-10 identification rate on the *IAM database* [29]. Wu et al. [15] combine a SIFT-descriptor codebook-based approach computed from IP descriptors with a histogram of scales and orientations of IPs, relying on binarization and heuristic word segmentation. They use a descriptor codebook and generate a scales and orientation histogram of the IPs. Experiments on the *IAM dataset* show a 98.5 % top-1 identification rate. Jain and Doermann [13] combine features of contour gradients, edge-base features extracted from character contours, and Speeded Up Robust Features (SURF). They use the Fisher Vector for feature pooling and a linear combination of the distances to fuse the features. They report a 99.2 % and 97.4 % identification rate on the *ICDAR 2013 Greek* [36] and the *ICDAR 2013 English* [36] dataset respectively.

Descriptors

The hypothesis of this paper is that the writers' repeating patterns and practices are revealed in fundamental information of script parts and their respective relationships. Thus, we propose a set of simple descriptors based on IPs that capture such information at multiple scales: the distribution of orientations present in writing, and the geometrical relationships of script parts (e.g., stroke segments, loops, crossings, characters). The orientation is computed from the so-called dominant orientation, which

Table 1. Descriptors and their dimensions.

Descriptor	Name	Dimension
$p(I_\phi)$	Local-Angle Distribution	12
$p(I_\theta)$	Orientation Distribution	12
$p(I_\theta, I_\phi)$	Orientation-Local-Angle Distribution	144
$p(I_s, I_\phi)$	Scale-Local-Angle Distribution	108
$p(I_s, I_\theta)$	Scale-Orientation Distribution	108
$p(I_{BOS})$	Bag-of-SIFT-Words (BOS) Descriptor	300

describes the prevailing direction of local gradients in an image patch around an IP. Geometric relations between IPs are expressed in terms of angles within a local neighborhood. Additionally, we combine these features with the IPs scales, i.e., their spatial extend, allowing for a multi-scale description of the writing. Our descriptors are computed as 1D or 2D histograms, and normalized to a Probability Density Function (PDF). They are listed in Table 1.

Using these simple descriptors extracted directly from the original image, we avoid errors originating in preceding image processing steps. Furthermore, when compared with codebook-based methods, we circumvent the risk of having an incomplete feature representation and the computationally expensive codebook generation and matching. As proof of concept, we adopt the Difference-of-Gaussian (DoG) IP detector introduced by Lowe [37], yet other detectors such as SURF, Harris, or Laplace could also be used.

In order to assess the potential of descriptor combination as well as run-time differences, we furthermore include a simple SIFT-codebook-based descriptor $p(I_{BOS})$ proposed by [7].

Scale-Independent Descriptors

In the following, we describe three descriptors which are computed on all IPs regardless of their scale.

Local-Angle Distribution $p(I_\phi)$

This histogram captures the geometric relationships between IPs in terms of the angles formed between their connecting line and the horizontal line. The locations of IPs depend on the writing style, thus their spatial distribution and geometric relationships captures information about the writing.

We create an angle histogram of IP locations. Figure 3 visualizes the computation, with circles denoting

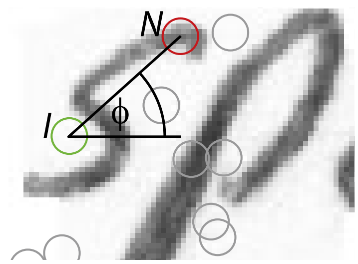


Figure 3. *Angle computation for $p(I_\phi)$. The line connecting the current IP I (green) with its neighbor N (red) and the horizontal line enclose angle ϕ. Further neighboring IPs are depicted grey.*

IPs: the angle ϕ is formed by the intersection of the line connecting the current IP I (green) with its neighbor N (red) and the horizontal line. For each IP I we compute the angle ϕ to its n-nearest neighbors N. Angles are mapped onto $[0°, 180°]$ and quantized with angle step α resulting in a histogram with $^{180}/_{\alpha}$ bins.

Orientation Distribution $p(I_\theta)$

This descriptor captures the distribution of the stroke orientations present as well as orientations of larger structures such as characters and words (e.g., the orientation of the baseline, or mean line of a word, c.f. blue vectors in Figure 4).

We create a histogram of dominant orientations quantized with angle step β, resulting in a histogram with $^{360}/_{\beta}$ bins. An IP's dominant orientation θ is computed as the maximum peak of a histogram h of magnitude-weighted gradient orientations in a local image patch. At locations with multiple peaks of similar magnitude ($\geq 0.8 * max(h)$), several IPs are created with the same location and scale but different orientations (see Figure 4, which shows a handwriting sample with IPs depicted as vectors).

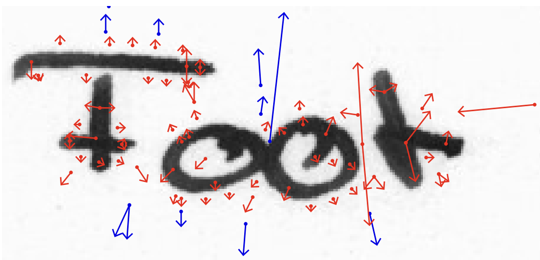


Figure 4. Handwriting sample with dominant orientations I_θ denoted as vectors with their length, angle, and origin indicating scale, orientation, and location, respectively. The blue vectors capture the orientation and position of the base, mean, and toplines.

Orientation-Local-Angle Distribution $p(I_\theta, I_\phi)$

The combination of the two previous descriptors captures geometric relations of IPs sharing the same dominant orientation, i.e., it sets structures of similar orientation θ into a geometrical relationship expressed by their angle ϕ. To create the 2D histogram of size $^{360}/_{\beta} \times {}^{180}/_{\alpha}$, for each IP with a given quantized orientation I_θ, its neighbors sharing the same orientation are determined and the angle ϕ is calculated.

Multi-Scale Descriptors

The following two descriptors are multi-scale views of the previously presented descriptors, i.e., the respective angle distributions $p(I_\phi)$ and dominant orientations $p(I_\theta)$ are observed separately for each scale. This allows setting into relation structures of similar size.

Both descriptors $p(I_\phi)$ and $p(I_\theta)$ are computed as 2D histograms of size $X \times Y$, with X denoting the number of quantized scales and Y being the number of orientations or angle distributions, respectively. The scales are determined from the parameters of the IP: using DoG, we decompose an image into T octaves and S sub-levels, resulting in a scale space of $X = T \times S$. The scale s_i of IP i is in the interval $1 \leq s_i \leq T \times S \times \sigma \times m_i$, with m_i being the extrema magnitude, and σ being the standard deviation of the Gaussian kernel for creating the scale space; it is quantized using step size σ.

Scale-Local-Angle Distribution $p(I_s, I_\phi)$

This descriptor sets IPs into geometrical relation that have a similar scale (i.e., structures of similar size) and expresses the relationship in terms of angles as $p(I_\phi)$.

We compute angles between neighbors sharing the same scale, rather than considering all nearest neighbors: at each scale level I_s, we compute I_ϕ as defined previously, resulting in a 2D histogram.

Scale-Orientation Distribution $p(I_s, I_\theta)$

This descriptor captures the orientations present at different granularity (scales), i.e., instead of collecting all orientations into one descriptor, the distributions are built according to the scale of the respective IPs. This allows describing orientations of stroke segments (scale of IPs similar to the stroke width) separately to the orientations that describe larger structures, e.g., full characters or words.

Using the same strategy as for $p(I_s, I_\phi)$, we generate a 2D histogram collecting the dominant orientations I_θ for each scale level I_s, where I_θ is computed as defined previously.

Additional Descriptor for Combination

In order to assess the potential performance gains of descriptor combination, we implemented a **Bag-of-SIFT-Words (BoS) descriptor** $p(I_{BOS})$ based on [7]. The descriptor is an occurrence histogram of SIFT features derived from a codebook of frequent features extracted from an independent training set.

To compute the descriptor, we extract the SIFT features, match them against a codebook to count the occurrences of each codebook entry ("word"). The resulting descriptor is called "bag of words". The codebook of size 300 is computed on the training set of the CVL database [39] using k-means clustering (with 7 documents of 27 writers, 189 documents in total).

Parameters

Parameter evaluations have been performed on the training set of the *ICDAR 2011 Writer Identification Contest* [38], a small but representative dataset for free-form handwritten documents with homogeneous background. An extensive parameter evaluation and more details on the importance of adapting the IPs to the task at hand is provided in [18].

Optimal DoG parameters were found to be $[T = 3, S = 6, \sigma = 1.3, th = 5, r = 0]$, with T being the number of octaves, S the number of sub-levels, σ the standard deviation of the Gaussian kernel, th the detection threshold and r the edge threshold [37]. The value $r = 0$ means that IPs located on edges are permitted.

The descriptor parameters adopted based on the evaluation are $[\alpha = 15°, \beta = 30°, n = 10]$, with α and β being the angle steps of $p(I_\phi)$ and $p(I_\theta)$, respectively, and n being the number of neighbors considered for $p(I_\phi)$. The number of neighbors is fixed and does not consider the writing size.

Our experiments show that truncating the descriptor vectors $p(I_s, I_\phi)$ and $p(I_s, I_\theta)$ from 216 to the first 108

Table 2. Specifications of the evaluation datasets with the short name, the reference, number of writers, documents written by one writer, lines per document and the scripts and/or languages the documents are written in. Every writer contributed to each of the languages.

Dataset	Ref	# writers	# docs/writer	# lines/doc	Languages and scripts
IAM modified	[29]	657	2	3-14	English
ICDAR'11F ICDAR'11C	[38]	26	8	13-23 2	6 Latin (2 English, 2 French, 2 German), 2 Greek
ICDAR'13	[36]	250	4	4	2 English, 2 Greek

scale levels improves the performance. Note that changing the size X of the scale space and truncating are not the same operation since the actual scale of an IP additionally depends on its extrema magnitude.

Evaluation

This section first lists the datasets used for the evaluation, followed by the description of the evaluation procedure and results. We close the section with a note on the processing time and speed gains with respect to existing IP-based methods that use a codebook.

Datasets

The evaluations of the descriptors were performed on the following four datasets (an overview of the specifications is provided in Table 2).

IAM modified contains a subset of documents of the *IAM database* [29]. Since the number of documents available of each writer ranges from 1 to 58 documents in the original database, we modified the dataset to have exactly two documents for each writer according to the procedure proposed in [15] in order to facilitate comparisons with the state of the art. This means, for writers who contributed more than two documents we only keep the first two, and for writers with only one contribution, we cut the respective documents roughly in half. This results in 1314 documents with 8.5 text lines per document on average.

ICDAR'11F The benchmark set of the *ICDAR 2011 Writer Identification Contest* [38]. This dataset consists of 26 writers each of whom has written the same 8 texts resulting in 208 documents.

ICDAR'11C The same dataset as above but cropped such that only the first two text lines are kept.

ICDAR'13 The benchmark set of the *ICDAR 2013 Competition on Writer Identification* [36]. The dataset consists of 4 texts written by 250 writers resulting in 1000 documents. The documents were cropped to contain 4 text lines per document.

Assessing the complexity of these datasets, *ICDAR'11F* can be rated lowest, despite the fact that it contains several different scripts. However, it consists of a rather small number of writers each of whom has written several documents. *ICDAR'11C* has a higher complexity as the amount of available data is significantly less (2 lines per document). *IAM modified* has the highest number of writers while still providing a fair amount of data per writer, written in only one language. *ICDAR'13* has the highest complexity: it is a reasonably large dataset with two different scripts, while only providing a small amount of data per document and writer.

Evaluation Procedure

Our evaluation design follows the procedure of the *ICDAR 2011 Competition* [38]. For identification, we employ a naïve nearest neighbor approach in a leave-one-out manner using the χ^2 distance metric. For a query document q and a given reference document r, we compute the distance between their respective descriptors $h(q)$ and $h(r)$ as

$$\chi^2(q,r) = \sum_{i=1}^{H} \frac{(h(q)_i - h(r)_i)^2}{h(q)_i + h(r)_i}, \tag{1}$$

where i is the index of the bin and H is the dimensionality of the histogram.

Table 3. *Top-1* performance of all descriptors evaluated individually and in combination on all datasets (we report only the 5 best-performing combinations).
Top-1: an error is reported if a document of a different writer is returned as first result.

		Descriptor	IAM mod.	ICDAR'11F	ICDAR'11C	ICDAR'13
Individual	(1)	$p(I_\phi)$	19.5	69.2	20.2	7.8
	(2)	$p(I_\theta)$	49.2	87.5	63.0	34.9
	(3)	$p(I_\theta, I_\phi)$	68.5	97.1	69.7	58.0
	(4)	$p(I_s, I_\phi)$	26.9	90.3	40.9	22.1
	(5)	$p(I_s, I_\theta)$	**81.9**	**98.6**	**87.5**	**81.4**
	(6)	$p(I_{BOS})$	82.3	99.5	80.7	80.1
Combination	(1,2)	$p(\phi)$, $p(I_\theta)$	55.6	88.4	63.9	39.0
	(3,4)	$p(I_s, \phi)$, $p(I_s, I_\theta)$	80.4	98.5	84.1	81.3
	(2,5)	$p(I_\theta)$, $p(I_s, I_\theta)$	77.9	**98.6**	82.2	77.3
	(5,6)	$p(I_s, I_\theta)$, $p(I_{BOS})$	**86.9**	**98.6**	**87.3**	**85.5**
	(1-6)	All descriptors	86.1	**98.6**	86.7	84.1

Table 4. Results on the *ICDAR'11F* and *'11C* dataset.
***Soft evaluation:* an error is reported if *no document of the same writer* is returned in the n first results.**
***Hard evaluation:* an error is reported if *a document by a different writer* is returned in the n first results.**

Method	Soft Evaluation				Hard Evaluation			
	'11F		'11C		'11F		'11C	
	Top-1	Top-10	Top-1	Top-10	Top-2	Top-7	Top-2	Top-7
ECNU	84.6	88.9	65.9	86.5	51.0	0.0	39.4	0.0
QUQA-a	90.9	99.0	74.0	96.2	76.4	20.2	52.4	3.4
QUQA-b	98.1	**100.0**	67.3	94.7	92.3	41.4	47.6	6.3
TSINGHUA	**99.5**	**100.0**	90.9	99.5	95.2	41.4	79.8	12.5
GWU	93.8	99.0	74.0	95.2	80.3	20.2	51.4	6.3
CS-UMD	**99.5**	99.5	66.8	89.9	91.8	22.1	51.9	3.4
TEBESSA	98.6	**100.0**	87.5	99.5	97.1	50.0	76.0	14.4
MCS-NUST	99.0	99.5	82.2	97.6	93.3	38.9	71.6	11.1
Wu et al. [15]	**99.5**	**100.0**	**95.2**	**100.0**	**98.6**	**63.9**	**88.5**	**31.3**
$p(I_s, I_\theta)$	98.6	**100.0**	87.5	98.6	92.7	43.3	73.6	9.1
$p(I_s, I_\theta)$, $p(I_{BOS})$	99.0	**100.0**	87.3	99.0	94.7	52.9	74.2	11.1

We report the *Top-n* results as soft- and hard-evaluation. Soft evaluation means that at least one document of the same writer has to be returned within the n first results, while hard evaluation is stricter, and counts an error if a document by a different writer is returned in the n first results. Significance is measured with the *Chi-square* test ($\alpha = 0.05$).

Results

In the following, we first report the results of the individual descriptors introduced in the previous section, and the results of their respective combinations. This is followed by a detailed evaluation and comparison of our best-performing descriptors with the state of the art.

Table 5. Results on the ICDAR'13 dataset.
Soft evaluation: **an error occurs if** ***no document of the same writer*** **is returned in the n first results.**
Hard evaluation: **an error occurs if** ***a document by a different writer*** **is returned in the n first results.**

Method	Soft Evaluation				Hard Evaluation	
	Top-1	Top-2	Top-5	Top-10	Top-2	Top-3
CS-UMD-a	**95.1**	**97.7**	**98.6**	**99.1**	19.6	7.1
CVL-IPK	90.9	93.6	97.0	98.0	44.8	24.5
HANNOVER-b	91.5	94.2	97.0	98.0	54.3	27.3
Wu et al. [15]	94.8	96.7	98.0	98.3	**63.2**	**36.5**
QATAR-b	78.4	85.8	91.5	95.1	34.6	16.5
TEBESSA-c	93.4	96.1	97.8	98.5	62.6	36.5
$p(I_s, I_\theta)$	81.4	88.2	92.8	96.2	36.3	16.9
$p(I_s, I_\theta)$, $p(I_{BOS})$	85.5	92.1	95.6	97.7	41.1	19.3

Table 3 presents the *Top-1* performance of all descriptors on all datasets. The descriptors proposed and the BOS descriptor in the first half, and best-performing combinations in the second. Descriptor combination is performed using linear combination with equal weights.

Table 3 reveals that geometric relations between IPs $p(I_\phi)$ as well as dominant orientations $p(I_\theta)$ as single descriptors are not discriminative enough to characterize a writer. Combining local angles and orientations $p(I_\theta, I_\phi)$ does not yield competitive scores. However, introducing scale as a second dimension in $p(I_s, I_\phi)$ and $p(I_s, I_\theta)$ leads to a significant increase in performance on all datasets.

Combining all descriptors (1-6) does not yield better results than the best-performing descriptor combination (5,6), because their performance is lowered by the less distinctive descriptors. However, on large datasets such as *IAM* and *ICDAR'13*, combination with the BOS approach significantly boosts the performance with respect to both the BOS-method[1] and the geometrical descriptors introduced in this paper, showing that our descriptors capture complementary information.

[1]Note that the performance reported on the IAM dataset (90.8) by the authors is not directly comparable to the literature since they used only a subset of documents for evaluation: writers with only one sample are not evaluated, and 2 to 58 reference samples are kept for identification, while for each writer we keep only one reference sample in our evaluation. It is inherent that fewer writers and more reference samples result in better performance. To assess whether our implementation is comparable to the original, we evaluated it according to the strategy explained in their paper [7], achieving a slightly better identification rate of 92.4 compared to the rate they reported (90.8). Thus, we assume our implementation to be comparable.

Table 6. Top-1 performance comparison of our best-performing descriptors and two complex IP-based state-of-the-art methods on the IAM modified.

Method	Top-1	Top-5	Top-10
Wu et al. [15]	98.5	99.1	99.5
Jain & Doermann [13]	94.7	98.1	98.7
$p(I_s, I_\theta)$	81.9	90.4	94.1
$p(I_s, I_\theta), p(I_{BOS})$	86.9	91.6	94.7

Table 4 and Table 5 compare the performance of our best-performing descriptor $p(I_s, I_\theta)$ and feature combination $p(I_s, I_\theta), p(I_{BOS})$ to the participants of the *ICDAR 2011 Competition* and the *ICDAR 2013 Competition* [36], respectively. On *ICDAR'11F* and *'11C* the performance of our method is competitive with the state of the art, being only weaker compared to the top method on the cropped dataset. On *ICDAR'13* and also *IAM modified* (see Table 6) our descriptors are still able to produce good results, especially when considering their simplicity.

A simple ranking of all documents of a dataset with respect to their distance to a query document gives insight into the performance of the features. Figure 5 shows such a ranking with respect to the descriptor $p(I_s, I_\theta)$ for all documents of *ICDAR'11F* with respect to a randomly selected query document[2].Each document is depicted as data point color-coded according to its language. We adopted the color-coding to allow assessing potential dif-

[2]Document 2 of writer 9, ICDAR'11F (English text).

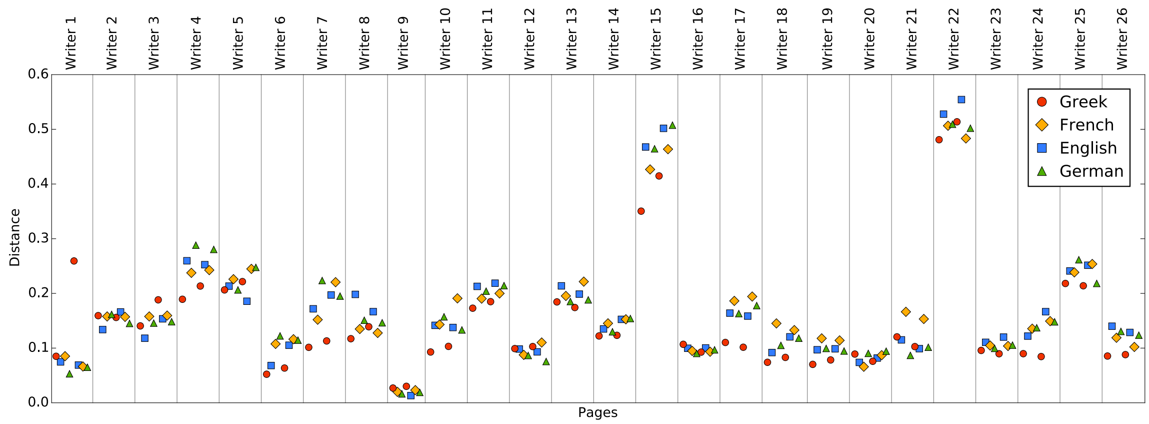
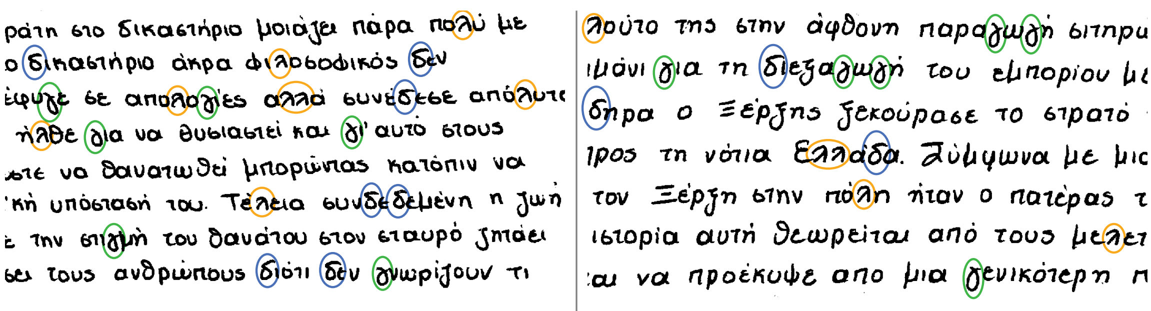


Figure 5. *Ranking of all documents of ICDAR'11F compared to document 2 of writer 9 (randomly selected) according to the descriptor $p(I_s, I_\theta)$. The distances of the documents by writer 9 are smaller than for all other writers, and the documents for each respective writer are clustered with respect to the query document. Note the outlier in writer 1's documents.*



(a) writer 1, crop of page 1

(b) writer 1, two crops of page 5

Figure 6. *Comparison of the two Greek documents of writer 1 from ICDAR'11F. The characters γ, δ, and λ are highlighted with color-coded ellipses.*

ferences in languages and scripts. Greek script is different to all other Latin scripts, and accents in French writings might change the script in a way different to German and English. This ranking allows several observations: Firstly, as expected, documents of the same writer have a smaller distance with respect to the query document than the documents of all other writers. Secondly, we can see that the documents of each writer form clusters, even across different languages and scripts. For the majority of writers, the respective scripts cluster within a writer's cluster (differences between German and English writing are negligible). This cluster behavior of documents written by the same author can be interpreted as a performance indicator: we can expect a high identification performance based on such a discriminative distance measure.

In our data analysis, we found a single outlier in the *ICDAR'11F dataset*: the fifth document of writer 1 (c.f. Figure 5, writer 1). While all other documents cluster well with respect to their writer, we found this specific outlier. In order to allow assessing the differences, we included the two Greek documents of writer 1 in Figure 6. We can observe consistent differences in the character shapes between the two documents, such as γ, δ, or λ. While the writing style of the characters is consistent within a document, it is different between the documents. In order to assess the inherent variance in the writer's documents, we made several plots shown in Figure 7, that clearly demonstrate document 5 as an outlier. Figure 7 (a) shows that documents written in another language and script to the query document (Greek) are closer than the outlier; (b) shows the overall larger distances of all documents with respect to the outlier, and (c) illustrates the distances with respect to a document written in English.

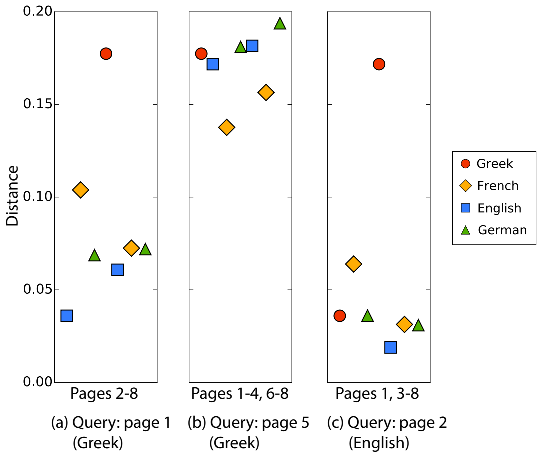


Figure 7. *Ranking of the documents of writer 1 according to the descriptor $p(I_s, I_\theta)$. According to their distance to (a) document 1, (b) document 5, (c) document 2.*

Table 7. Comparison of the processing time of two IP-based methods: our best performing descriptor $p(I_s, I_\theta)$ and the BoS descriptor inspired by [7] expressed in O-Notation and measured on *ICDAR'11F*.

Desc.	Training		Test	
	O	min	O	min
$p(I_s, I_\theta)$	-	-	$O(k)$	2.0
$p(I_{BOS})$	$O(n^{mk+1} \log n)$	81.79	$O(mnk)$	3.3

Processing Time

In order to assess the speed of our approach and the potential gain in processing time with respect to existing IP-based methods, we compared our best-performing descriptor $p(I_s, I_\theta)$ with our implementation of the BoS method [7] on a standard desktop computer (Intel i7 3.4 GHz processor, 32 GB RAM). A summary is presented in Table 7.

The BoS descriptor needs a training phase prior to identification, in which the codebook of size k is created with k-means clustering in $O(n^{m*k+1} \log n)$ time, where m is the dimension of the feature vector, and n the number of IPs to be clustered. The creation of the codebook, using the training set of the CVL dataset takes 82 minutes on the test machine. Our descriptors do not require any training and can be applied out-of-the-box to an unseen dataset.

Since both methods are based on IPs, we do not need to consider the processing time for extracting the IPs from the document images for the run-time comparison. Also both methods need to normalize their histograms, which can be assumed to take linear time with respect to the histogram size.

Methods requiring a codebook are also time-consuming at run-time: for each extracted feature the corresponding codebook entry is searched to create the occurrence histogram. Thus, such methods lead to a total run-time complexity of $O(mnk)$ for one query document, where m is the codebook size, n is the number of IPs on the document, and k is the size of the descriptor. The complexity is composed of the cost of creating the descriptor $O(mn)$ and the distance computation $O(k)$.

In contrast, our method directly computes the descriptor from the IP properties; thus, it has a linear run-time complexity of only $O(k)$ for a k-dimensional descriptor. Thus, in practice we cut down the processing-time by $^1/_3$ compared to the BoS descriptor at run-time measured on *ICDAR'11F* with our test setup.

Conclusion

This paper proposed a set of novel descriptors for writer identification that is simple and fast to compute. The descriptors capture the distribution of elementary stroke orientations at multiple scales, and the geometrical configuration of a local neighborhood of script parts defined by angles. Building upon interest points, we avoid (pre)processing such as binarization and segmentation.

Experiments on widely-used datasets show the potential of our descriptors by themselves and in combination with others. While we cannot always match the results of competition winners, our descriptors have a compelling performance considering their simplicity and low dimensionality. Having a very sparse image representation and no need of matching descriptors against a codebook, we reduce the processing time by about $^1/_3$ compared to traditional codebook-based methods at runtime.

A limitation of our method is the amount of data needed to create a stable identification model especially on large datasets. However, we showed that combining our descriptor set with a simple bag-of-SIFT-words-based method significantly boosts the performance for both methods, demonstrating its ability to capture complementary information.

Future work includes assessing our descriptors' performance when combined with more powerful additional methods, and using advanced machine-learning for the identification. We suggest including the *Scale Orientation Distribution* in future interest-point-based methods for its performance, ability to capture complementary information, and fast computation.

Acknowledgements

This work has been supported by the Swiss National Science Foundation project 205120_150173.

References

[1] V. Pervouchine and G. Leedham, “Extraction and Analysis of Forensic Document Examiner Features Used for Writer Identification,” *Pattern Recognition*, vol. 40, no. 3, pp. 1004–1013, Mar. 2007.

[2] L. Schomaker, “Advances in Writer Identification and Verification,” in *International Conference on Document Analysis and Recognition*, 2007, pp. 1268–1273.

[3] R. Huber and A. Headrick, “The fundamentals of the identification process,” in *Handwriting Identification: Facts and Fundamentals*. CRC Press, 1999.

[4] S. Srihari and H. Arora, “Individuality of Handwriting: A Validation Study,” in *International Conference on Document Analysis and Recognition*, vol. 47, 2002, pp. 106–109.

[5] U. S. Committee on Identifying the Needs of the Forensic Sciences Community of the National Research Council, *Strengthening Forensic Science in the United States: A Path Forward*. National Academies Press, 2009.

[6] H. H. Harralson, E. Waites, and E. J. WILL, “A Survey of Forensic Handwriting Examination Research in Response to the NAS Report,” in *Conference of the International Graphonomics Society (IGS2015)*, 2015, pp. 107–110.

[7] S. Fiel and R. Sablatnig, “Writer Retrieval and Writer Identification Using Local Features,” in *International Workshop on Document Analysis Systems*, 2012, pp. 145–149.

[8] A. Fornés, J. Lladós, G. Sánchez, and H. Bunke, “Writer Identification in Old Handwritten Music Scores,” in *Pattern Recognition and Signal Processing in Archaeometry*, 2012, pp. 27–63.

[9] W. Hofmeister, A. Hofmeister-Winter, and G. Thallinger, “Forschung am Rande des paläographischen Zweifels: Die EDV-basierte Erfassung individueller Schriftzüge im Projekt DAmalS,” in *Kodikologie und Paläographie im digitalen Zeitalter*, 2009, pp. 261–292.

[10] I. Bar-Yosef, I. Beckman, K. Kedem, and I. Dinstein, “Binarization, Character Extraction, and Writer Identification of Historical Hebrew Calligraphy Documents,” *International Journal of Document Analysis and Recognition*, vol. 9, no. 2-4, pp. 89–99, 2007.

[11] S. Connell and A. Jain, “Writer adaptation for online handwriting recognition,” *IEEE Trans. PAMI*, vol. 24, no. 3, pp. 329–346, 2002.

[12] S. M and S. Mary Idicula, “A Survey on Writer Identification Schemes,” *International Journal of Computer Applications*, vol. 26, no. 2, pp. 23–33, Jul. 2011.

[13] R. Jain and D. Doermann, “Combining Local Features for Offline Writer Identification,” in *International Conference on Frontiers in Handwriting Recognition*, 2014, pp. 583–588.

[14] M. Bulacu and L. Schomaker, “Text-Independent Writer Identification and Verification Using Textural and Allographic Features.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 701–17, 2007.

[15] 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*, vol. 9, no. 3, pp. 526–536, 2014.

[16] R. Kumar, B. Chanda, and J. Sharma, “A Novel Sparse Model Based Forensic Writer Identification,” *Pattern Recognition Letters*, vol. 35, pp. 105–112, Jan. 2014.

[17] I. Pratikakis, B. Gatos, and K. Ntirogiannis, “ICDAR 2013 Document Image Binarization Contest (DIBCO 2013),” in *International Conference on Document Analysis and Recognition*. IEEE, Aug. 2013, pp. 1471–1476.

[18] A. Garz, M. Würsch, and R. Ingold, “Training- and Segmentation-Free Intuitive Writer Identification with Task-Adapted Interest Points,” in *Conference of the International Graphonomics Society (IGS2015)*, 2015, pp. 103–106.

[19] H. Said, T. Tan, and K. Baker, “Personal Identification based on Handwriting,” *Pattern Recognition*, vol. 33, no. 1, pp. 149–160, 2000.

[20] M. Bulacu and L. Schomaker, “Text-Independent Writer Identification and Verification Using Textural and Allographic Features,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 701–717, 2007.

[21] A. Brink, M. Bulacu, and L. Schomaker, “How much Handwritten Text is needed for Text-Independent Writer Verification and Identification,” in *International Conference on Pattern Recognition*, 2008, pp. 1–4.

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

[23] F. Leclerc and R. Plamondon, “Automatic Signature Verification: The State of the Art – 1989-1993,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 08, no. 03, pp. 643–660, 1994.

[24] R. Plamondon and S. N. Srihari, “Online and Off-Line Handwriting Recognition: A Comprehensive Survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 63–84, 2000.

[25] D. Impedovo and G. Pirlo, “Automatic Signature Verification: The State of the Art,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 38, no. 5, pp. 609–635, 2008.

[26] D. Impedovo, G. Pirlo, and R. Plamondon, “Handwritten Signature Verification: New Advancements and Open Issues,” in *International Conference on Frontiers in Handwriting Recognition*, 2012, pp. 367–372.

[27] M. Bulacu, L. Schomaker, and L. Vuurpijl, “Writer Identification using Edge-Based Directional Features,” in *International Conference on Document Analysis and Recognition*, vol. 1, 2003, pp. 937–941.

[28] A. J. Newell and L. D. Griffin, “Writer Identification Using Oriented Basic Image Features and the Delta Encoding,” *Pattern Recognition*, vol. 47, no. 6, pp. 2255–2265, Jun. 2014.

[29] U. V. Marti and H. Bunke, “The IAM-Database: an English Sentence Database for Offline Handwriting Recognition,” *International Journal on Document Analysis and Recognition*, vol. 5, no. 1, pp. 39–46, 2002.

[30] A. Schlapbach and H. Bunke, “A Writer Identification and Verification System using HMM Based Recognizers,” *Pattern Analysis Applications*, vol. 10, no. 1, pp. 33–43, 2007.

[31] L. Schomaker, M. Bulacu, and K. Franke, “Automatic Writer Identification Using Fragmented Connected-Component Contours,” in *International Workshop on Frontiers in Handwriting Recognition*, 2004, pp. 185–190.

[32] A. Bensefia, T. Paquet, and L. Heutte, “A Writer Identification and Verification System,” *Pattern Recognition Letters*, vol. 26, no. 13, pp. 2080–2092, 2005.

[33] R. Niels, F. Grootjen, and L. Vuurpijl, “Writer Identification through Information Retrieval: The Allograph Weight Vector,” in *International Conference on the Frontiers of Handwriting Recognition*, 2008, pp. 481–486.

[34] L. Schomaker, “Handling Within-Writer Variability and Between-Writer Variation in the Recognition of Online Handwriting,” Ph.D. dissertation, 1995.

[35] J. Woodard, M. Lancaster, A. Kundu, D. Ruiz, and J. Ryan, “Writer Recognition of Arabic Text by Generative Local Features,” in *International Conference on Biometrics: Theory, Applications and Systems*, 2010, pp. 1–7.

[36] G. Louloudis, B. Gatos, N. Stamatopoulos, and A. Papandreou, “ICDAR 2013 Competition on Writer Identification,” in *International Conference on Document Analysis and Recognition*, 2013, pp. 1397–1401.

[37] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[38] G. Louloudis, N. Stamatopoulos, and B. Gatos, “ICDAR 2011 Writer Identification Contest,” in *International Conference on Document Analysis and Recognition*, 2011, pp. 1475–1479.

[39] F. Kleber, S. Fiel, M. Diem, and R. Sablatnig, “CVL-DataBase: An Off-Line Database for Writer Retrieval, Writer Identification and Word Spotting,” in *International Conference on Document Analysis and Recognition*, 2013, pp. 560–564.

Author Biography

Angelika Garz received her MS in computer science from the Vienna University of Technology, Austria (2011). She has then worked as a research assistant at the Computer Vision Lab, Vienna University of Technology in document analysis. Currently she pursues her PhD at the Document, Image, and Voice Analysis (DIVA) group at the University of Fribourg, Switzerland, on historical document image analysis. Her research focuses on layout analysis, script and writer identification on historical manuscripts.

Marcel Würsch received his MS in computer science from the Universities of Bern, Fribourg and Neuchatel, Switzerland (2014). He is currently a PhD student at the DIVA group (University of Fribourg). His research focus is twofold: writer identification directed at human-interpretable features and developing a web service framework for research. Web services allow researchers to provide a public interface to their methods, data, and evaluations even without necessarily granting access to their code.

Andreas Fischer received his M.S. and Ph.D. degrees in Computer Science in 2008 and 2012 from the University of Bern, Switzerland. Afterwards, he conducted post-doctoral research at CENPARMI and the Ecole Polytechnique in Montreal, Canada. Since 2015, he is a senior researcher and lecturer at the University of Fribourg and the HES-SO in Fribourg, Switzerland. His research is focused on document analysis and graph-based pattern recognition. He is a member of the IAPR.

Rolf Ingold received his PhD degree in Computer Science from the Swiss Federal Institute of Technology Lausanne (EPFL). He joined the University of Fribourg in 1989 as associate professor and was promoted full professor in 1997. He is considered as an international expert in multimodal document analysis. Since 2008, he is Chair of the Swiss Pattern Recognition Association and member of the Governing Board of IAPR. In 2011 he received the James A. Lindner Prize.