# Text-independent Writer Identification Using SIFT Descriptor and Contour-directional Feature

Yu-Jie Xiong, Ying Wen, Patrick S P Wang, Yue Lu
Shanghai Key Laboratory of Multidimensional Information Processing
Department of Computer Science and Technology
East China Normal University, Shanghai 200241, China

*Abstract*—**This paper presents a method for text-independent writer identification using SIFT descriptor and contour-directional feature (CDF). The proposed method contains two stages. In the first stage, a codebook of local texture patterns is constructed by clustering a set of SIFT descriptors extracted from images. Using this codebook, the occurrence histograms are calculated to determine the similarities between different images. For each image, we obtain a candidate list of reference images. The next stage is to refine the candidate list using the contour-directional feature and SIFT descriptor. The proposed method is evaluated with two datasets: the ICFHR2012-Latin dataset and the ICDAR2013 dataset. Experimental results show that the proposed method outperforms the state-of-the-art algorithms and archives the best performance.**

***Keywords*—*Writer identification; text-independent; codebook; SIFT descriptor and contour-directional feature***

## I. Introduction

Writer identification is to find out the author of a questioned handwriting from a set of writers. Due to its common interest and potential applications, it is very attractive to both industry and academia and many methods have been proposed in the past few years. In general, writer identification can be categorized into two classes: on-line and off-line. Furthermore, off-line writer identification is subdivided into text-dependent and text-independent approaches. This present work focuses on the task of off-line text-independent writer identification. According to the features extracted, the current popular approaches of off-line text-independent writer identification can be classified into two categories: texture-based and structure-based.

Texture-based approaches extract textural information as features for writer identification. He et al. [1] treated the task of writer identification as a texture analysis problem, and used Gabor wavelet techniques and mesh fractal dimension techniques. Two dimensional discrete wavelet transform with dynamic time warping (DTW) lifting scheme was also utilized [2]. Bertolini et al. [3] discussed the use of texture descriptors to perform writer verification and identification. Djeddi et al. [4] discussed the problem of writer identification in a multi-script condition using the feature extracted from Grey Level Run Length (GLRL) matrices.

Structure-based approaches seem to be commendable in the case that the amount of characters/words or text lines is finite. These methods extract structure feature from the handwriting through statistical analysis, which is steady for allographic variation [5]. Structure-based feature retains more local information of the character than textural feature. The former contains not only the directions of contours but also the relationships of stroke structures. Based on the idea of the local structure, edge-based directional probability distributions and connected component contours were proposed for the writer identification task [6]. Bensefia et al. [7] introduced grapheme as the feature for describing the individual properties of handwriting. Bulacu and Schomaker [8] compared three different clustering methods for generating the grapheme codebook. Jain and David [9] took advantage of contour gradient which can capture local shapes and curvatures to create a pseudo-alphabet for identification.

This paper presents an approach based on SIFT descriptor and contour-directional feature for writer identification. The flowchart of proposed method is given in Fig. 1. SIFT descriptors are extracted to generate a codebook by clustering algorithm. Using this codebook, its occurrence histograms of the query and reference document images are calculated to determine their similarities. For each query image, we obtain a candidate list of reference images after the coarse identification. In the fine identification, we utilize this candidate list to reduce calculation time, which means the query image only compares with images in the list rather than all reference images. Both contour-directional feature and SIFT descriptor are extracted to calculate a feature vector to determine the similarities of images. This vector is composed of the occurrence histogram of SIFT descriptor and the probability density function of contour-directional feature.

## II. The proposed method

The proposed method has two stages. The first stage is coarse identification using SIFT descriptor, and the second stage is fine identification using both contour-directional feature and SIFT descriptor.

### A. Coarse Identification

*1) Keypoint detection:* Scale Invariant Feature Transform (SIFT) [10] is proposed by Lowe, which is successfully applied to object detection. It is also used for writer retrieval and writer identification [11].

Keypoint detection is accomplished by searching for stable points over all scales, using a Gaussian function of scale space. As a consequence, potential interest points are invariant to scale. The scale space function $L(x, y, \sigma)$ of an image is defined as the convolution of a variable-scale Gaussian

978-1-4799-1805-8/15/$31.00 ©2015 IEEE

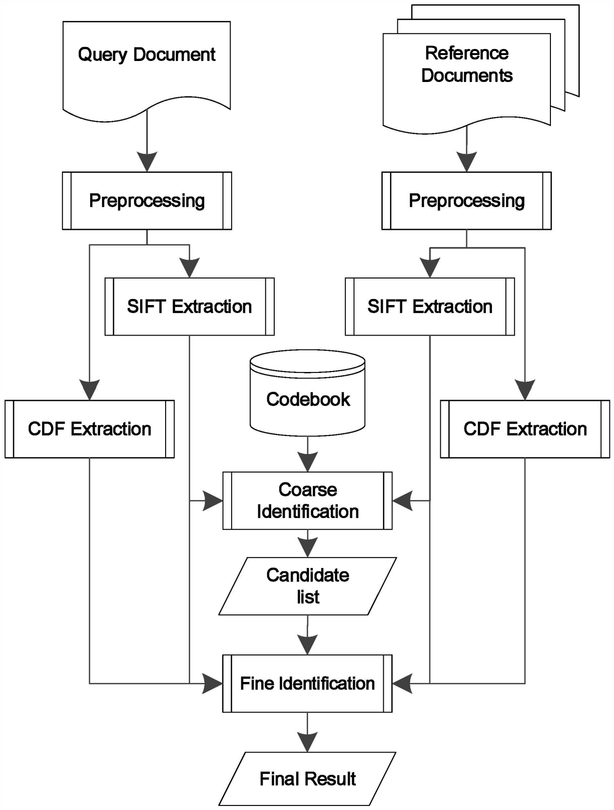



Fig. 1: System flow chart

$G(x, y, \sigma)$ with an input image $I(x,y)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

where $*$ is the convolution operation, and

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

The difference-of-Gaussian function $D(x, y, \sigma)$, which is calculated by the subtraction of two nearby scales separated by a constant multiplicative factor $k$, is used for extrema detection:

$$D(x, y, \sigma) = L(x, y, k * \sigma) - L(x, y, \sigma)$$

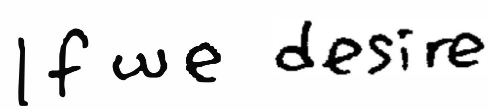
After the difference-of-Gaussian images are created, each sample point of images is compared to its eight neighbors in the current image and nine neighbors in the scale above and below. Only if it is larger or smaller than all of these neighbors is selected as the keypoint.

*2) Keypoint selection:* Compared with natural scene image, handwriting image is lack of gray scale variation, so the traditional SIFT does not work well in the handwriting image. Based on the characteristic of handwriting, we improve the keypoint detection algorithm to overcome this shortcoming. In general, the allographic information of character exists in the stroke of the character rather than the background area. Therefore, we apply the background point elimination to remove these useless keypoints. This procedure is performed according to the following criterions: If $S_n < n - 1$, $p$ is regarded as a keypoint in the background area; if $S_n \geq n-1$, $p$ is regarded as a keypoint in the stroke area, where $p$ is a detected keypoints obtained by traditional SIFT, and $S_n$ is the number of the black pixel in the $n \times n$ spatial neighbor grid of $p$. The keypoint in background area will be removed. Through the above modification, the SIFT descriptors of remaining keypoints are credible representation of the individuality of the handwriting. Fig. 2 is an example of the background point elimination. Fig. 2(a) is an original handwriting image, and Fig. 2(b) is the image after extrema detection, in which the colorized circles with different size are the potential keypoints in different scales. The red circles are keypoints in background area, while the cyan ones are keypoints in the stroke area. There are about 150 potential keypoints in Fig. 2(b). After background point elimination, less than 62 keypoints remain and are shown in the Fig. 2(c).

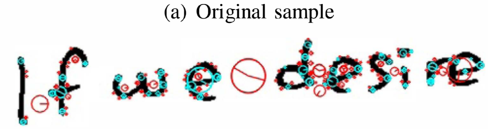
*3) SIFT descriptor extraction and codebook generation:* After keypoint detection, descriptors of keypoints are computed. The keypoint descriptor is represented as orientation histograms over $4 * 4$ sample regions, and every region has 8 orientations. Therefore, the normalized feature vector for each keypoint has 128 elements. Through the above steps, thousands of SIFT descriptors are extracted from the detected keypoints in handwriting. It is hard to calculate the similarity of different handwritings using those descriptors directly. To solve this problem, we cluster the descriptors of keypoints into $N$ classes as the codebook and represent each class by its center $(C_1, C_2...C_N)$. For each SIFT descriptor, we calculate its nearest cluster center $C_i(1 < i \leq N)$ of the codebook based on Euclidean distance, and the occurrence counter corresponding to $C_i$ is incremented by one. After all SIFT descriptors are calculated, the normalized occurrence histogram is treated as feature representation of the handwriting.

In this work, the K-means clustering is used for codebook generation. Some experiments are performed to confirm the optimal number $N$ of cluster center $C_i$. In our experiments, $N$ is set equal to 300, and more details about the selection of $N$ are described in Section 3.

*4) Identification:* For two handwriting image $I_1$ and $I_2$, the occurrence histogram of SIFT descriptor extracted from $I_1$ and $I_2$ are denoted by $OHS_1$ and $OHS_2$. The similarity of $I_1$ and $I_2$ is determined by the distance $D_{OHS}$ between $OHS_1$ and $OHS_2$. Many existed methods based on minimum distance can




(a) Original sample



(b) Potential keypoints before Background point elimination

(c) Keypoints after Background point elimination

Fig. 2: Keypoint detection processing

be used. In this paper, weighted Chi-squared distance [12] are used as distance measurement. For every handwriting image in the database, we calculate its distance to all other handwriting images, and then a distance-based candidate list is obtained by sorting the results from the most similar to the least similar handwriting image.

### B. Fine Identification

*1) Contour detection:* Contour detection is the preprocessing for contour-directional feature extraction. After the image is binarized, morphology method is used to remove the noise. Sobel operator is applied to generate the contour image in our method. The adjacency type of original contour image pixels obtained by Sobel operator is m-adjacency. We can obtain the 4-adjacency contour image and the 8-adjacency contour image from the original contour image easily. The experimental results show 4-adjacency contour image achieves the best performance. More description about contour image with different adjacency is given in Section 3.

*2) Contour-directional feature extraction:* The proposed contour-directional feature (CDF) is similar to the grid microstructure feature (GMF) [12]. Our proposed CDF treats the handwriting from the perspective of edge direction. It focuses on the stroke direction of edge pixel pairs rather than two edge fragments in the neighborhood.

The major difference between CDF and GMF is the way to index the pixels that are neighboring to the edge pixel. Fig. 3 is the comparison of GMF and CDF. The black block is the edge pixel $P$, and the gray ones are edge pixels which are connected to $P$. The pixel $A$ around $P$ is marked with the index $G(A) = B_i$ in Fig. 3(a), where $B$ denotes the larger distance in the horizontal and vertical distance between $A$ and $P$, and $1 \leq i \leq 8 * B$. While the index $C(A)$ of a pixel $A$ in Fig. 3(b) is marked according to the direction $Dir(A)$ of $A$:

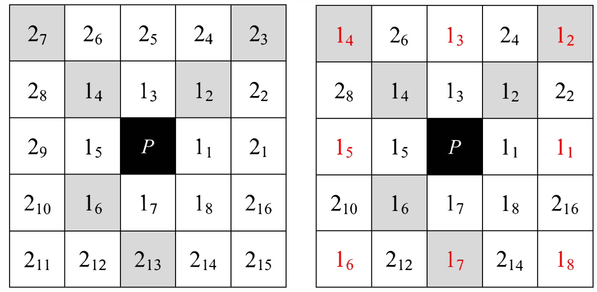
$$\begin{cases} \text{If } Dir(A) \text{ is unique, then } C(A) = G(A) = B_i; \\ \text{If } Dir(A) = Dir(A_1) = ... = Dir(A_n), \\ B(An) < ... < B(A1) < B(A) \\ \text{then } C(A) = C(A_1) = ... = C(A_n) = G(A_n) = B(A_n)_i; \end{cases}$$

where $(A_x, A_y)$ and $(P_x, P_y)$ are the coordinates of $A$ and $P$, and $DirA = \arctan((A_y - P_y)/(A_x - P_x))$. As shown in Fig. 3(b), blocks with the red index have different $C$ and $G$.

For each edge pixel $P$ surrounded by a $(2L+1)*(2L+1)$ sliding window, we record the specific edge pixel pairs $\{\alpha, \beta\}$ which shall meet the following conditions to create the feature vector:

$$\begin{cases} \alpha \text{ and } \beta \text{ are edge pixels,} \\ G(\alpha) = A_i,\ G(\beta) = A_j, \text{ and } i < j, \\ G(\gamma) = A_k, i < k < j,\ \gamma \text{ is not the edge pixel} \end{cases}$$

As a comparison, $\{1_2, 1_4\}$, $\{1_4, 1_6\}$, $\{2_3, 2_7\}$, $\{2_7, 2_{13}\}$ in Fig. 3(a) are recorded as GMF, while the edge pixel pairs $\{1_2, 1_4\}$*2, $\{1_4, 1_6\}$, $\{1_4, 1_7\}$ in Fig. 3(b) are recorded as CDF.




(a) Grid microstructure feature (b) Contour-directional feature

Fig. 3: Comparison of GMF and CDF

*3) Re-identification:* For a given handwriting image $I_Q$, the N most similar images$\{I_1,I_2,...I_N\}$ in the dataset is found out in the previous stage. For two handwriting image $I_Q$ and $I_i$, the contour-directional feature extracted from $I_Q$ and $I_i$ are denoted as $PDF_Q$ and $PDF_i$. The distance of $PDF_Q$ and $PDF_i$ is denoted as $D_{CDF}$. We define the similarity $D_F$ between $I_Q$ and $I_i$ as the sum of $D_{OHS}$ and $D_{CDF}$. In our experiments, N is set equal to 30 empirically.

## III. Experimental results

To evaluate the proposed method, ICFHR2012-Latin dataset and ICDAR2013 dataset are used. The ICFHR2012-Latin Dataset [13] was the benchmarking dataset created for the ICFHR2012 Writer Identification Contest. This dataset was created by 100 writers, and every writer was asked to copy four paragraphs of text in two languages (English and Greek). The ICDAR2013 Dataset [14] was designed for the ICDAR2013 Competition on Writer Identification. This dataset contains 1000 handwriting images written by 250 writers and four pages per writer, and each writer was asked to copy four pages of text in two languages (English and Greek).

### A. Comparative evaluation of CDF and GMF

Experiments are performed to find the best feature extraction criterion of CDF, and validate its performance. Table I gives the writer identification performance (Top-1) of CDF and GMF on the entire ICDAR2013 dataset. It is noted that the best accuracy of CDF is 93.2% while best accuracy of GMF is 91.9%, and CDF performs better than the GMF in every case. In addition, both CDF and GMF show the same trend that features extracted from m-adjacency contour image are better than those extracted from 8-adjacency contour image, while features extracted from 4-adjacency contour image are the best.

TABLE I: The identification accuracy comparison of CDF and GMF on the entire ICDAR2013 dataset

| Feature \ contour | 4-adjacency | 8-adjacency | m–adjacency |
|---|---|---|---|
| CDF | **93.2%** | 92.7% | 92.8% |
| GMF [12] | **91.9%** | 91.6% | 91.7% |

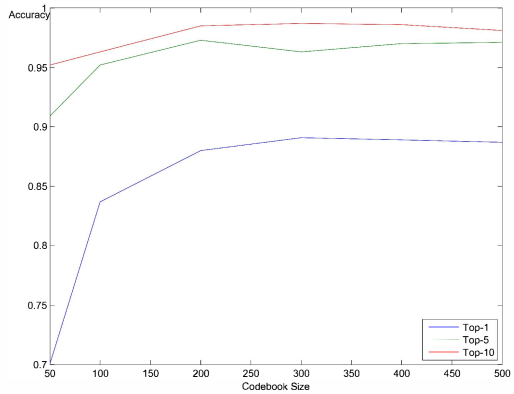



Fig. 4: The influence of the codebook size on the performance on the ICDAR2013 dataset

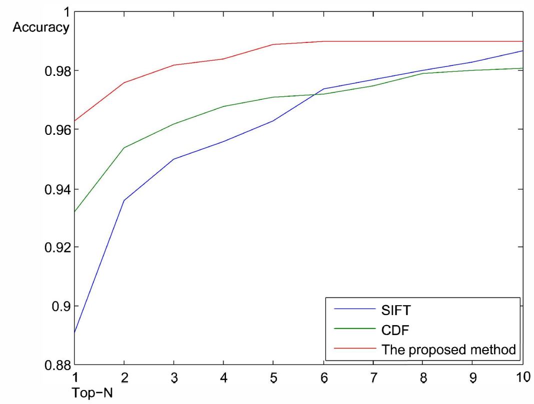



Fig. 6: Soft TOP-N performance of different features on the entire ICDAR2013 dataset

### B. Codebook size

The size of codebook is important to the performance of the proposed method. In this experiment, a range of codebook size are explored to study the impact of the size of codebook to the performance. A number of 100 samples corresponding to 25 writers on the entire ICDAR2013 dataset are used for codebook generation. Fig. 4 shows that the accuracy of identification is improved with the increasing of codebook size, When the size is larger than 300, the accuracy dropping slightly. So the codebook size is set equal to 300 in the following experiments.

### C. Performance on two public datasets

ICFHR2012-Latin dataset and ICDAR2013 dataset are employed for experiments respectively. We take 100 images from one dataset to generate the codebook for the other dataset. The soft and hard TOP-N criterions are used for performance evaluation. The values of N used for the soft criterion in our experiments are 1, 5 and 10 while the N for the hard criterion are 2 and 3.

Fig. 5 and Fig. 6 show the performance of different features in two datasets. As shown in the both Figures, the proposed method achieves the best accuracy. It means that CDF and SIFT descriptor characterize the handwriting from different aspects, so the combination of both features contributes to enhancing the performance effectively.

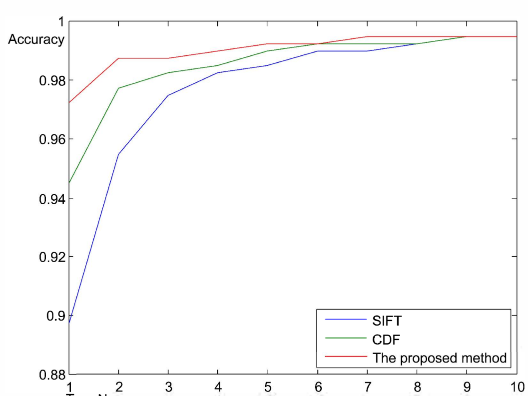



Fig. 5: Soft TOP-N performance of different features on the entire ICFHR2012-Latin dataset

TABLE II: Identification accuracy comparison on the ICFHR2012-Latin dataset of Greek

| Method \ Top-N | S-Top-1 | S-Top-5 | S-Top-10 |
|---|---|---|---|
| TEBESSA-a | 92.0% | 98.8% | 99.0% |
| TEBESSA-c | 93.5% | 99.5% | 99.5% |
| TSINGHUA | 90.0% | 98.5% | 99.0% |
| Proposed method | **97.0%** | **100%** | **100%** |

The performance of different approaches in single language datasets is also investigated. Table II and III show the performance on ICFHR2012-Latin Greek and English dataset respectively, and Table IV and V show the performance on ICDAR2013 Greek and English dataset. The Top-1 accuracy of the proposed method is better than the state-of-the-art approaches, while its Top-5 and Top-10 performance is not the best in few cases. The reason is that the proposed method is a two-stage scheme, and the final result cannot redeem the correct writer which is lost in the first stage. It means that the Top-1, Top-5 and Top-10 of the second stage cannot higher than the Top-30 of the first stage, so the Top-5 and Top-10 is hard to improve compared with the Top-1. When the samples of one writer are less, this situation is more troublesome.

Table VI and VII respectively represent the soft and hard Top-N results on ICFHR2012-Latin dataset and ICDAR2013 dataset. The results show that the proposed method outper-

TABLE III: Identification accuracy comparison on the ICFHR2012-Latin dataset of English

| Method \ Top-N | S-Top-1 | S-Top-5 | S-Top-10 |
|---|---|---|---|
| TEBESSA-a | 89.5% | 97.0% | **98.5%** |
| TEBESSA-c | 91.5% | **97.5%** | 98.0% |
| TSINGHUA | 94.0% | 95.5% | 98.0% |
| Proposed method | **94.0%** | 97.0% | 97.5% |

TABLE IV: Identification accuracy comparison on the ICDAR2013 dataset of Greek

| Method \ Top-N | S-Top-1 | S-Top-5 | S-Top-10 |
|---|---|---|---|
| CS-UMD-a | 95.6% | 98.6% | **99.2%** |
| CS-UMD-b | 95.2% | **98.8%** | 99.0% |
| HIT-ICG | 93.8% | 97.2% | 97.8% |
| TEBESSA-c | 92.6% | 98.0% | 98.4% |
| Proposed method | **96.2%** | 98.4% | 99.0% |

TABLE V: Identification accuracy comparison on the ICDAR2013 dataset of English

| Method \ Top-N | S-Top-1 | S-Top-5 | S-Top-10 |
|---|---|---|---|
| CS-UMD-a | **94.6%** | **98.4%** | 98.8% |
| CS-UMD-b | 94.4% | **98.4%** | **99.0%** |
| HIT-ICG | 92.2% | 96.4% | 96.8% |
| TEBESSA-c | 91.2% | 96.2% | 96.6% |
| Proposed method | 94.0% | 96.4% | 97.2% |

forms all of approaches submitted to the competitions of ICFHR2012 and ICDAR2013 in most case. It shows that our approach has good capability for writer identification in multi-script environment. The performance of our approach drops much less than others for the tighter hard criterion, which demonstrates that our approach is robust for harsh conditions.

## IV. Conclusions and Future Work

This paper proposes a method for text-independent writer identification based on SIFT descriptor and contour-directional feature. Experiments with two public different language datasets demonstrate the effectiveness of the proposed method. The first stage of identification is based on SIFT descriptor, therefore it is insensitive to the aspect ratio and slant of the characters. Moreover, in the second stage of identification, the discriminability of occurrence histogram of SIFT descriptor is enhanced through the fusion with the contour-directional feature reflecting the structure information of handwriting. Our method is independent of segmentation, which means it is robust for layout variation of text lines. In conclusion, the proposed method outperforms the state-of-the-art approaches and is promising for text-independent writer identification with different applications.

TABLE VI: Identification accuracy comparison on the entire ICFHR2012-Latin dataset

| Method \ Top-N | S-Top-1 | S-Top-5 | S-Top-10 | H-Top-2 | H-Top-3 |
|---|---|---|---|---|---|
| TEBESSA-a | 92.3% | 98.8% | 99.0% | 57.5% | 38.0% |
| TEBESSA-c | 94.5% | **99.3%** | 99.3% | 65.0% | 37.8% |
| TSINGHUA | 92.8% | 97.8% | 98.3% | 51.5% | 27.3% |
| Proposed method | **96.8%** | **99.3%** | **99.3%** | **67.8%** | **39.8%** |

TABLE VII: Identification accuracy comparison on the entire ICDAR2013 dataset

| Method \ Top-N | S-Top-1 | S-Top-5 | S-Top-10 | H-Top-2 | H-Top-3 |
|---|---|---|---|---|---|
| CS-UMD-a | 95.1% | 98.6% | 99.1% | 19.6% | 7.1% |
| CS-UMD-b | 95.0% | 98.6% | **99.2%** | 20.2% | 8.4% |
| HIT-ICG | 94.8% | 98.0% | 98.3% | 63.2% | **36.1%** |
| TEBESSA-c | 93.4% | 97.8% | 98.5% | 62.6% | **36.1%** |
| Proposed method | **96.2%** | **98.6%** | 99.0% | **63.5%** | 35.0% |

## V. Acknowledgments

This work is jointly supported by the Science and Technology Commission of Shanghai Municipality under research grants 14511105500, 14DZ2260800, and Shanghai Collaborative Innovation Center of Trustworthy Software for Internet of Things (ZF1213).

## References

[1] Z. Y. He, X. G. You, L. Zhou, Y. M. Cheung, and J. W. Du, "Writer identification using fractal dimension of wavelet subbands in gabor domain," *Integrated Computer-Aided Engineering*, vol. 17, no. 2, pp. 157–165, 2010.

[2] S. Gazzah and N. Ben Amara, "Arabic handwriting texture analysis for writer identification using the dwt-lifting scheme," in *Proceedings of the International Conference on Document Analysis and Recognition*, pp. 1133–1137, 2007.

[3] D. Bertolini, L. S. Oliveira, E. Justino, and R. Sabourin, "Texture-based descriptors for writer identification and verification," *Expert Systems with Applications*, vol. 40, no. 6, pp. 2069–2080, 2013.

[4] C. Djeddi, I. Siddiqi, L. Souici-Meslati, and A. Ennaji, "Text-independent writer recognition using multi-script handwritten texts," *Pattern Recognition Letters*, vol. 34, no. 10, pp. 1196–1202, 2013.

[5] 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.

[6] L. Schomaker and M. Bulacu, "Automatic writer identification using connected-component contours and edge-based features of uppercase western script," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 787–798, 2004.

[7] A. Bensefia, T. Paquet, and L. Heutte, "Handwriting analysis for writer verification," in *Proceedings of the International Workshop on Frontiers in Handwriting Recognition*, pp. 196–201, 2004.

[8] M. Bulacu and L. Schomaker, "A comparison of clustering methods for writer identification and verification," in *Proceedings of the International Conference on Document Analysis and Recognition*, pp. 1275–1279, 2005.

[9] R. Jain and D. Doermann, "Writer identification using an alphabet of contour gradient descriptors," in *Proceedings of the International Conference on Document Analysis and Recognition*, pp. 550–554, 2013.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[11] S. Fiel and R. Sablatnig, "Writer identification and writer retrieval using the fisher vector on visual vocabularies," in *Proceedings of the International Conference on Document Analysis and Recognition*, pp. 545–549, 2013.

[12] X. Li and X. Ding, "Writer identification of chinese handwriting using grid microstructure feature," in *Advances in Biometrics*, pp. 1230–1239, 2009.

[13] G. Louloudis, B. Gatos, and N. Stamatopoulos, "ICFHR 2012 competition on writer identification challenge 1: Latin/greek documents," in *Proceedings of the International Conference on Frontiers in Handwriting Recognition*, pp. 829–834, 2012.

[14] G. Louloudis, B. Gatos, N. Stamatopoulos, and A. Papandreou, "ICDAR 2013 competition on writer identification," in *Proceedings of the International Conference on Document Analysis and Recognition*, pp. 1397–1401, 2013.