A Connected Component-Based Approach for Text-Independent Writer Identification

Tayeb Bahram

November 16, 2019

A connected component-based approach for text-independent writer identification

1st Tayeb Bahram
GACA Laboratory
Dr Tahar Moulay University Saïda, Algeria
tayeb.bahram@univ-usto.dz

***Abstract*—Writer identification from off-line images of handwriting is a challenging task. In this work we assess the performance of textural descriptors for writer identification on different writing styles. The proposed method is different from the existing texture based methods: the earlier methods extract texture information at page, paragraph, fragment level to get a document descriptor, while the proposed method exploits the texture at a connected-component level. Specifically, an improvement of the textural features, the contour direction, angle, and length distributions are explored. Using these texture descriptors, the occurrence histograms are calculated in order to determine the similarities between different images. For the evaluation, the IFN/ENIT (411 writers) and IAM (657 writers) datasets were used. In our characterization of the individual and combination of textural features of connected-components we show that the proposed method outperforms the state-of-the-art algorithms and archives the best performance.**

***Index Terms*—handwriting analysis, forensic document examination, writer identification, feature extraction, texture**

I. Introduction

Handwriting is a special type of behavioral biometrics, which can be used mainly for writer identification and verification [23]. Overall, this category can be used efficiently and effectively in forensic document analysis with applications (e.g. a ransom note, a fraudulent letter, a suicide note, a will, a threatening letter) [20], [23]. Writer identification for short, is the process of determining the author of sample handwriting from a group of writers (a one-to-many comparison). Writer verification is the task of analyzing and deciding whether or not two handwritten documents are produced by the same hand or not (a one-to-one comparison) [21].

In the recent years, off-line writer identification work has been focused on two main streams: *codebook-based* and *textural-base*d techniques. In the first group, the codebook (or bag of features) can be used as feature vector to characterize writer individuality. The grapheme occurrence probability is a characteristic of each writer and can be employed to distinguish between inter-class et intra-class distances [3]. Several technique are proposed in the literature to create the codebook, such as the Fraglets (Graphemes or small patterns) [4], [5], [12], [22], Bagged Discrete Cosine Transform (BDCT) descriptors [14], Line Fragments [10], and the Elliptic Graphemes [1].

On the other hand, textural-based approaches take into account that the handwritten texts of each person as a very different texture to identify writers. Moreover, the textural features are extracted from the entire texture image, connected components, or the writing fragments (blocks) to build descriptors. Probability Distribution Functions (PDFs) are calculated using the entire handwriting image (document) for textural features like the hinge-direction, direction co-occurrence, run-length in [4], [5], [7], and the Oriented Basic Image (OBI) in [9]. The contour direction(angle)-length [3], direction(angle) co-probability [3], run-lengths of Local Binary Pattern (LBPruns) [13], and Cloud Of Line Distribution (COLD) [13] extracted from the connected-components (CCs) are used for writer identification. Recently, in [11] three kind of textural descriptors, Local Binary Patterns (LBP), Local Ternary Patterns (LTP), and Local Phase Quantization (LPQ) are computed from the writing fragments. In this study [11] each writing fragment is considered as a texture. Similar identification technique is proposed in [6] using the Block Wise Local Binary Count (LBW-BC) descriptors.

The focus of this study is a new proposal of textural-based feature that is extracted from binary images of connected components (or just CCs) of handwriting texts. The interesting aspect of this work is the computation of texture measures from the complete section of connected trace (connected component), delimited only by where the writer has lifted the pen. The essential attributes, such as writing direction, curvature, ink-trace width, and size of letters are estimated. The remained of this paper is structured as follows. The next section describes the proposed method of extracting the textural feature (LCP-LOC). Section III presents our experimental results and a discussion on the IFN/ENIT [19] and IAM [15] datasets. Finally, Section IV concludes this work.

II. Feature extraction

First, each handwritten document image is binarized using Otsu's thresholding algorithm [18]. From the result, the black connected components were extracted basing on the image 8-connectedness (8-pixels), where a connected-component is a complete region (or ink-blob) of ink-trace, delimited by two successive liftings the pen [3]. The black pixels correspond to the ink and the white pixels correspond to the background in a binary image of handwriting. In this step, some common connected components (CCs) between writers, such as the small CCs, dots, and the scatter noise (binary image) were removed. Further, for each binary connected-component, the

Inner and Outer Contours (IC(s) and OC(s)) are extracted using Moore's method starting at the left-most pixel in a counter-clockwise manner (see Fig. 3 (b)). The following subsections describe the feature considered in this study.

A. Local Contour Pattern operator

The Local Binary Pattern (LBP) operator proposed by Ojala et al. in [17] for texture description, is a simple and efficient generic texture descriptor. This operator labels with decimal number each pixel p_c of coordinates (x, y) located in the image by thresholding its surrounding pixels. The gray-scale LBP operator can be defined as follows:

$$LBP_{P,R}(x,y) \quad = \quad \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{1}$$

Where g_c and g_p $(p = 0, \dots, P-1)$ represent the gray values of the center pixel p_c and the P surrounding pixels on a circle of radius R $(R > 0)$, respectively, P is the number of neighbors, and $s(t)$ is defined as follows:

$$s(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases} \tag{2}$$

As shown in Fig 1, the gray-scale LBP code is computed by comparing the center pixel intensity with its surrounding 3×3 pixels, with $P = 8$ and $R = 1.0$.

The LBP histograms [11], [13] give a rough idea about the texture information of writing but this might not be very effective in capturing the fine details in writing; thus we propose a new local geometrical structure, named Local Contour Pattern (LCP). The fundamental objective of this study is the computation of texture measures from the outer contours $\mathcal{C} = \{OC_0, OC_1, \dots, OC_{N-1}\}$ of black connected-components $\mathcal{D} = \{CC_0, CC_1, \dots, CC_{N-1}\}$ to characterize a handwritten sample. Where N is the number of connected-components in the document $\mathcal{D}$ (related to amount of text).

The binary LCP operator labels each black pixel p_b of coordinates (x, y) in connected-component CC_i ($p_b \in CC_i$ and $p_b \notin OC_i$) by examining its surrounding P pixels $(n_0, \dots, n_{P-1})$. This binary code LCP will be obtained by reading the values in a clockwise fashion and can be defined as:

$$LCP_{i,P,R}(x,y) \quad = \quad \sum_{p=0}^{P-1} f(n_p)2^p \tag{3}$$

where the coordinates of neighbors can be calculated as

$$(x_{n_p}, y_{n_p}) \quad = \quad \left(x + R\cos(\frac{2\pi p}{P}), y + R\sin(\frac{2\pi p}{P})\right) \tag{4}$$

the notation (i, P, R) denotes a neighborhood of P points on a circle of radius R of the connected-component CC_i, and the function $f(n_p)$ is defined as follows:

$$f(n_p) = \begin{cases} 1, & if \quad n_p \in OC_i \\ 0, & otherwise \end{cases} \tag{5}$$

The Fig. 2 (b) shows an example of LCP code calculation, the central black pixel p_b is labelled as 01101100 in binary or 54 in decimal ($P = 8$ and $R = 1.0$).

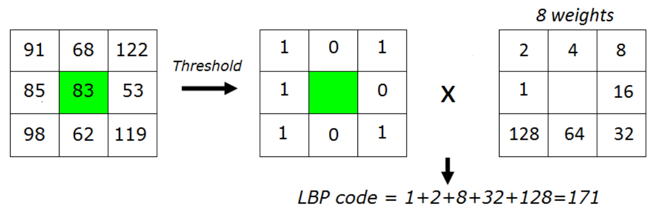

Fig. 1. An example of gray-scale LBP code ($P = 8$ and $R = 1.0$).

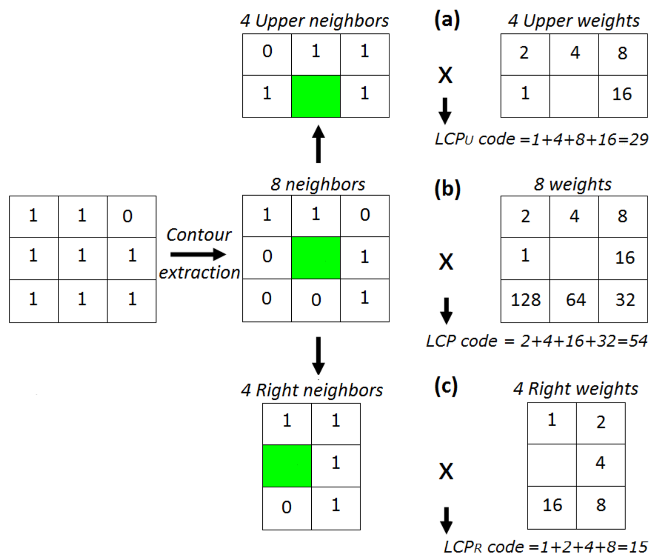

Fig. 2. Computing the binary LCP code and its splitting into Upper and Right LCP codes ($P = 8$ and $R = 1.0$).

B. Length measure

The run-length distribution proposed by Arazi [2] for automatic writer identification is one of the oldest and very effective features which still attracts the attention of researchers in the field of biometrics [13]. A *run* can be defined as a sequence of connected pixels with a repeated property (e.g. pixels of the same color) along a given direction [4], [8]. Run-lengths can be computed on the black ink-trace (black pixels) [4], [5], the background (white pixels) [13], or the gray scale images [13]. The lengths are usually determined by counting the number of pixels for the horizontal (0°) or vertical (90°) scan direction, or the number of pixels multiplied by $\sqrt{2}$ for the right-diagonal (45°) or left-diagonal (135°) scan direction. For example, if the binary sequence was "00001110010011", the Black (value '1') and White (value '0') run-lengths would be coded respectively as $(3, 1, 2)$ and $(4, 2, 2)$. On the other hand, Bahram et al. in [3] gave explanations on the determination of two lengths based on the black connected-components and their contours: Black length (L_B) and length between two outer points (L_O).

In this paper, we use length between outer contour pixels (LOC). If $S = \{p_u, p_{u+1}, \ldots, p_{v-1}, p_v\}$ a sequence of connected pixels of black connected-component CC_i in a given direction (or a *run*), p_u ($p_u \in OC_i$) and p_v ($p_v \in OC_i$) are two pixels of outer contour OC_i and $\{p_j\}$ ($p_j \in \{0,1\}$, $j = u+1 \ldots v-1$) are the binary pixels between p_u and p_v. In this case, the length LOC_i represents in CC_i the distance between two outer contour pixels, p_u and p_v, and can be expressed as follows:

$$LOC_i = \sqrt{(y_v - y_u)^2 + (x_v - x_u)^2} \quad (6)$$

where (x_u, y_u) and (x_v, y_v) are the coordinates of outer contour pixels p_u and p_v, respectively. In Fig. 3, length LOC from a connected-component is illustrated.

C. Local Contour Pattern - Length Outer Contour (LCP-LOC)

The distribution of Local Contour Patterns (LCPs) gives a rough idea about the orientation (direction) and curvature (angle) information of the writing, which can be applied in a discriminatory way between different writers. The Length (LOC) histograms could serve well to provide the overall information about the handwriting shapes, size of letters, and the ink-trace width (habitual pen gripe).

In this approach, we propose a new complex feature that is capable of introducing more information: Local Contour Pattern - Length Outer Contour of connected components ($LCP-LOC$) feature. The main idea is to consider, not one, but two locally measured LCP_i and length LOC_i for each pixel p_b in the connected-component CC_i. Every measurement (LCP_i,LOC_i) is agglomerated in a local histogram H_i which represents the local texture of the CC_i. Our proposed technique works at the accumulated connected-component histograms H_i, which are two primary representation of the handwritten document: the black connected-components and their contours. They will be employed for the final texture feature computation and this histogram H can be defined as following:

$$H(k) = \sum_{i=0}^{N-1} H_i(k) \quad (7)$$

where N is the total number of connected-components in the document $\mathcal{D}$ and the local histogram H_i of CC_i is

$$H_i(k) = \sum_{p_b \in CC_i} \delta\{k, ((LCP_i - 1)max_L + (LOC_i - 1))\},$$
$$k = 0 \ldots (max_P \times max_L - 1) \quad (8)$$

Here, max_P and max_L represent the maximum numbers of LCP and LOC, respectively, and δ is the Kronecker delta and given by

$$\delta(k,l) = \begin{cases} 1, & if \quad k = l \\ 0, & otherwise \end{cases} \quad (9)$$

Then, the resulting histogram H_i is a ($max_P \times max_L$) dimensional feature vector.

It is worth noting that, in general, the left-to-right languages (e.g. Latin, Greek) are written from left to right (horizontal lines) and from top to bottom. Some languages such as the Arabic and Farsi texts run like these, but from right to left. For this purpose, two Local Contour Pattern operators (LCP) are retained: *Right* LCP (or LCP_R) and *Upper* LCP (or LCP_U). In Fig. 4 the LCP_{R_i} and LCP_{U_i} from a connected-component (CC_i) are illustrated. For each pixel p_b at position (x,y) in a connected-component CC_i, the resulting LCP_{R_i} and LCP_{U_i} can be expressed in decimal as follows:

$$LCP_{R_i}(x,y) = \sum_{p=2}^{6} f(n_p) 2^{(p-2)} \quad (10)$$

$$LCP_{U_i}(x,y) = \sum_{p=0}^{4} f(n_p) 2^{p} \quad (11)$$

In order to capture both ink-trace width and structure of the letters (e.g. size of letters) details, our proposed technique exploits two basic scanning methods: Horizontal along the columns (LOC_H) and vertical along rows (LOC_V) of connected-components. The pairs (LCP_{R_i}, LOC_{H_i}) and (LCP_{U_i}, LOC_{V_i}) are computed separately for each black pixel p_b of connected-component CC_i and the histograms H_{R_i} and H_{U_i} are used to characterize the connected-component CC_i. This feature (LCP-LOC) is illustrated in Fig. 4.

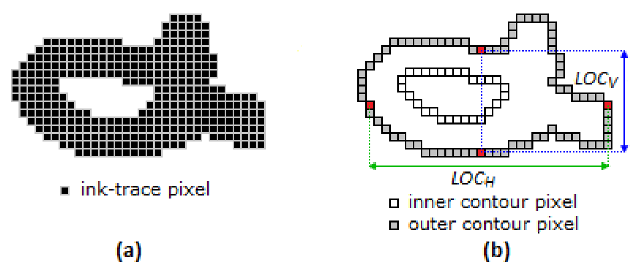

Fig. 3. Computing local length "LOC", (a) binary image and (b) horizontal and vertical lengths between outer contour pixels.

For our writer identification, the histograms H_{R_i} and H_{U_i} are agglomerated separately to form the final histograms $H_R(k) = \sum_{i=0}^{N-1} H_{R_i}(k)$ and $H_U(k) = \sum_{i=0}^{N-1} H_{U_i}(k)$. Indeed, the resulting histogram H_R and H_U are respectively normalized and interpreted in probability distributions ($f1$) and ($f2$) and are used as features in writer characterization. Furthermore, in order to have all the information needed to identifier the writer of a questioned document, the histograms H_R and H_U are concatenated together to form the combined histogram H_C, ($f1$ + $f2$).

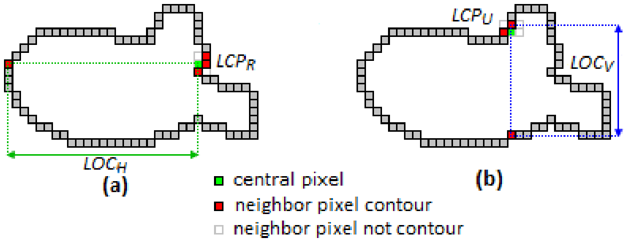

Fig. 4. *LCP-LOC* feature extraction, (a) (LCP_R,LOC_H) and (LCP_U,LOC_V) measures.

III. Experimental results

This section presents the results of the proposed textural-based method for off-line text-independent writer identification. We have evaluated and tested our method using the writing samples from available publicly datasets (IFN/ENIT and IAM).

A. Datasets

1) IFN/ENIT dataset: The IFN/ENIT dataset contains 26,459 images of handwritten Arabic names of 937 Tunisian towns and villages from 411 writers. All images are available as PNG binary image (black/white) and considered at 300 dpi. This dataset used by more researchers of off-line Arabic handwriting recognition [16] and writer identification/verification [1], [3], [5]. In the writer identification experiments reported in this study, we randomly chosen 50 words samples written by 411 writers. For each writer, we used 30 handwritten word images for reference set while the remaining 20 words are used for evaluating set.

2) IAM dataset: The IAM dataset [15] is one of the large and widely used dataset for writer identification and verification [4], [10]. It is a collection of 1539 pages of handwritten English texts written by 657 different writers with a varied length text (from 2 to 11 lines per page). The collected pages are scanned at 300 dpi and provided as PNG image in gray-scale. We modify the IAM dataset as described in [14] in our experiments by keeping only the first two documents for 657 writers who contribute more than two documents and splitting the document roughly in half for those writers with a unique page in the original set. Therefore, the result dataset contains 2 samples per writer, one of which is used as reference document while the other is used as query document.

Some samples from the IFN/ENIT and IAM datasets are presented in Fig. 5.

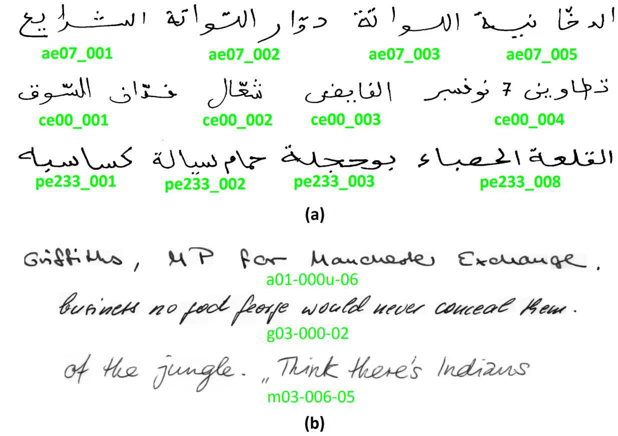

Fig. 5. Samples from the (a) IFN/ENIT Arabic dataset [19] and (b) IAM English dataset [15].

B. Writer identification

Writer identification is carrying out by measuring the overall similarity between the input unknown sample D_q and all samples of the reference base $\mathcal{R} = \{D_r\}$, $r = 0 \dots R-1$, where R is the number of documents in the reference base $\mathcal{R}$. Since the proposed feature *LCP-LOC* of each document is a normalized histogram in the form of ($max_P \times max_L$)-dimensional feature vector, the similarity between two documents is measured by computing the distance between their respective distribution. The $max_P = 31$ and max_L is experimentally fixed at $max_L = 35$ pixels for the IFN/ENIT dataset and $max_L = 25$ for the IAM dataset. We tested and evaluated four distance measures in our experiments including: Chi-square, Manhattan, Euclidean, Minkowski up to order 5. Our writer identification system is performed using Chi-square (χ^2) distance. Then, the distances between the normalized histogram H_q of the query document D_q and the H_r, ($H_r \in \mathcal{H}$) of the reference document D_r are calculated and classified in an organized hit list with decreasing similarity to input sample D_q (or increasing distance to H_q). Thus in fact, the distance between H_q and H_r is computed as

$$\chi^2(H_q, H_r) = \sum_{k=0}^{size-1} \frac{(H_q^k - H_r^k)^2}{H_q^k + H_r^k} \quad (12)$$

where k is the bin index, and $size$ is the number of bins in the histograms. In our implementation $size = max_P \times max_L$ for $f1$ and $f2$ features and $size = 2 \times max_P \times max_L$ for the feature combination ($f1 + f2$). The writer identification rate is given by the average probability of correctly identifying the writer of a queried document, when at least one document of the same writer is among in the top h of the hit list. Therefore, each time, one handwriting sample (or query document) of testing set is compared with all reference handwritten documents based on the nearest neighbor criterion (1-NN) and χ^2 distance. In this approach, the performance measure used is Top-1 identification rate.

C. Experiments with LCP-LOC feature

In the first, *LCP-LOC* feature on writer identification was tested using IFN/ENIT Arabic and IAM English text samples from each datasets independently. Table I gives the identification rates of three types of *LCP-LOC* feature considered here: LCP_R-LOC_H ($f1$), LCP_U-LOC_V ($f2$), and $f1$+$f2$. As this table shows, the combined performance ($f1$+$f2$) exceeds the performance of $f1$ and $f2$ features by using χ^2 distance. Satisfactory results, which were higher than 96%

for top writer identification rate, indicated that the proposed feature LCP-LOC were easily used in writer identification. From the same table, we also observe that the performance of the proposed method drops little from IFN/ENIT dataset to IAM dataset, such as Top-1 performance from 97.81% to 96.04% using $f1 + f2$ features combination.

TABLE I
TOP-1 IDENTIFICATION RATES ON THE IFN/ENIT (411 WRITERS) AND IAM (657 WRITERS) DATASETS.

Feature	**IFN/ENIT Arabic dataset (411 writers)**	**IAM English dataset (657 writers)**
LCP_R-LOC_H $(f1)$	90.02	87.21
LCP_U-LOC_V $(f2)$	89.78	91.17
$f1$+$f2$	97.81	96.04

TABLE II
WRITER IDENTIFICATION PERFORMANCE (TOP-1) OF DIFFERENT SYSTEMS ON THE IFN/ENIT AND IAM DATASETS.

Study	**IFN/ENIT dataset**		**IAM dataset**	
	# writers	**Top-1**	**# writers**	**Top-1**
Bulacu et al. [4], [5]	350	89.00	650	88.00
Siddiqi and Vincent. [22]	-	-	650	91.00
Djeddi et al. [7]	275	91.50	-	-
Ghiasi et al. [10]	-	-	650	93.70
Abdi and Khemakhem [1]	411	90.02	-	-
He and Schomaker [12]	-	-	650	91.10
Hannad et al. [11]	411	94.89	657	89.54
Bahram et al. [3]	350	97.50	-	-
Chahi et al. [6]	411	96.47	657	88.99
Khan et al. [14]	411	76.00	650	**97.20**
He et al. [13]	-	-	650	89.90
Durou et al. [9]	-	-	650	92.03
Our System	411	**97.81**	657	96.04

D. Stability

One major question of the approach adopted in this work is the amount of available text for each writer and its effect on the performance of the text-independent writer identification system. In the case of IFN/ENIT, the influence of amount of available text for each writer is studied on the 411 writers. This allows varying the amount of text from $3/2$ (3 words for the reference set and 2 words for testing set), to $30/20$ words for each writer. Since a document written by a writer in IAM dataset contains text ranging from 2 to 11 lines. In this experiment, we evaluate our system by using half a line (average 4 words), 1 line, 1.5 lines, 2 lines, 2.5 lines, 3 lines, 3.5 lines, 4 lines, half a page, and a complete page for each sample (2 samples $\times$ 657 writers). As shown in Fig. 6 (a) as the number of connected-components (or amount of writings) increases the performance of the system also increases. The performance curves begin to stabilize from $18/12$ words and 3.5 lines (average 24 words) on the IFN/ENIT and IAM datasets, respectively.

We will be interested to study the influence of the number of writers (or size of dataset) on Top-1 identification rate. The evaluations are carried out on the same datasets by varying the number of writers from 10 to the total number of writers of each dataset. As Fig. 6 (b) shows, there was not a sharp decrease in the identification rate as the number of writers increased. While the 100% correct identification rate was obtained for 10 writers per dataset, the 97.81% and 96.04% correct identification rates were obtained respectively on 411 writers of IFN/ENIT Arabic and 657 of IAM English dataset dataset by combining the features LCP_R-LOC_H $(f1)$ with LCP_U-LOC_V $(f2)$.

E. Comparison with other methods

This section presents a comparison between the proposed method and some other methods discussed early on Arabic and English writer identification. In order to have a significant comparison, we only compare the performance of our system with the systems which have been evaluated on the same datasets and which consider a significant number of writers. Table II lists our approach along with the ones surveyed in Section I, in terms of corpus size and writer identification performance (Top-1). As this table shows, the authors in [3] currently hold the best performance results with achievement of 97.5% on 350 writers of IFN/ENIT dataset. We have achieved an identification rate of 97.81% for the top-ranked writer on 411 writers of the same Arabic dataset. In the case of IAM dataset, the best existing performance for 650 writers of IAM dataset is 97.20% [14] which is used Bagged Discrete Cosine Transform (BDCT) descriptors. In case of our system, the proposed method achieves a higher recognition rate of 96.04% on 657 writers of English IAM dataset.

IV. CONCLUSION

In this work we are concerned with the problem of off-line text-independent writer identification, where the task is to determine the author of sample handwriting from a group of writers. The fundamental objective of this study is the computation of two texture measures (*Local Contour Pattern*, LCP and *Length Outer Contour*, LOC) from outer contours of the black handwriting connected-components to characterize the writer of a handwritten sample. Indeed, complete section of connected ink trace (connected-component), delimited only by where the writer has lifted the pen is inspected in this approach relatively to its aspects of directionality/angularity, ink-trace width , and structure of letters (i.e. size of letters). In order to enhance the performance of our writer identification system, both measures are combined (LCP-LOC). Using these measures, the occurrence histograms are calculated in order to determine the similarities between different handwriting images. The proposed textural-based feature LCP-LOC has practical feasibility and it is applicable to free style handwriting, both cursive and isolated forms. Two publicly available datasets, namely, the IFN/ENIT Arabic and IAM English were used for testing the proposed technique based on the nearest neighbor criterion. The experimental results show that our approach can achieve much higher identification rates than previous state-of-the-art methods. Finally, we will evaluate the proposed methodology in a multi-script environment (i.e. Arabic, English).

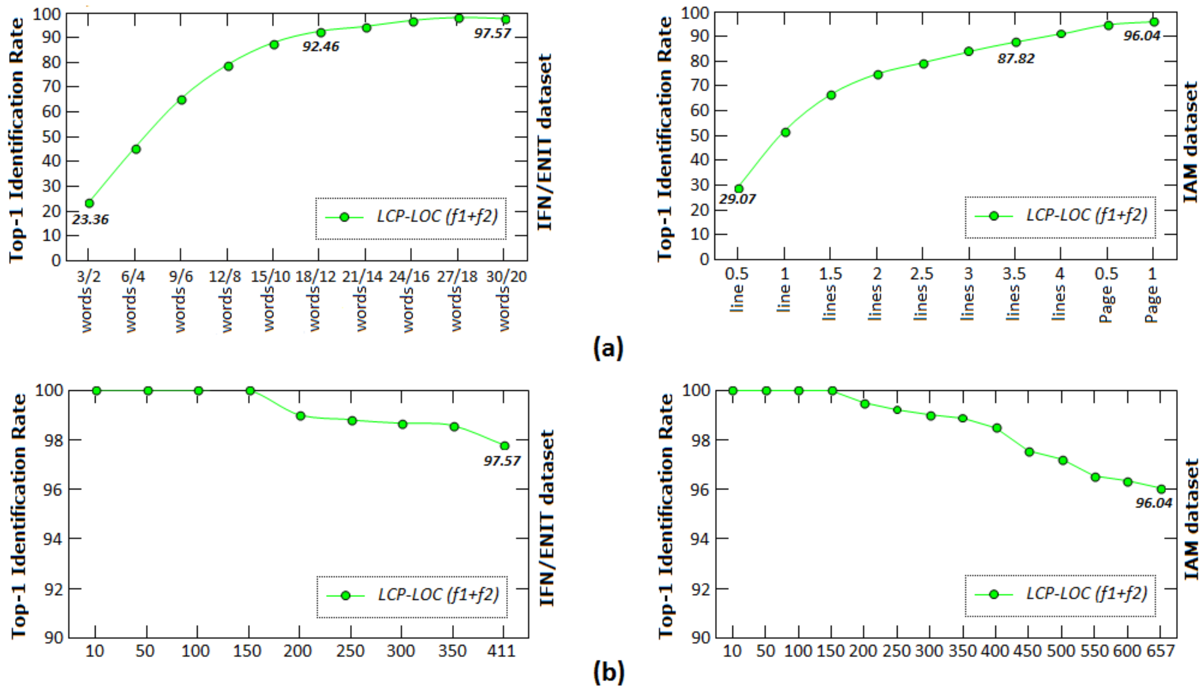

Fig. 6. Top-1 identification rates of $LCP\text{-}LOC$ with respect (a) to amount of text and (b) number of writers on the IFN/ENIT Arabic and IAM English datasets.

REFERENCES

[1] M. N. Abdi and M. Khemakhem, “A model-based approach to offline text independent arabic writer identification and verification,” Pattern Recognition, vol. 48, pp. 1890–1903, June 2015.

[2] B. Arazi, “Handwriting Identification by Means of Run-Length Measurements,” IEEE Transactions on Systems, Man, and Cybernetics, vol. 7, pp. 878–881, December 1977.

[3] T. Bahram, A. Benyettou, and D. Ziadi, “A set of features for text-independent writer identification,” International Review on Computers and Software (I.RE.CO.S.), vol. 11, pp. 898–906, October 2016.

[4] 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, pp. 701–717, April 2007.

[5] M. Bulacu, L. Schomaker, and A. Brink, “Text-independent writer identification and verification on offine arabic handwriting,” 9th International Conference on Document Analysis and Recognition (ICDAR 2007), vol. 2, IEEE, 2007, pp. 769–773.

[6] A. Chahi, I. El-khadiri, Y. El-merabet, Y. Ruichek, and R. Touahni, “Block wise local binary count for off-Line text-independent writer identification,” Expert Systems with Applications, vol. 93, pp. 1–14, March 2017.

[7] C. Djeddi, I. Siddiqi, L. Souici-Meslati, and A. Ennaji, “Multi-script Writer Identification Optimized with Retrieval Mechanism,” 2012 International Conference on Frontiers in Handwriting Recognition, ICFHR 2012, IEEE, 2012, pp. 509–514.

[8] C. Djeddi, I. Siddiqi, L. Souici-Meslati, and A. Ennaji,Text-independent writer recognition using multi-script handwritten texts,” Pattern Recognition Letters, vol. 34, pp. 1996–1202, July 2013.

[9] A. Durou, I. Aref, and S. Al-Máadeed, A. Bouridane, “Writer identification approach based on bag of words with OBI features,” Information Processing and Management, October 2017.

[10] G. Ghiasi and R. Safabakhsh, “Offline text-independent writer identification using codebook and efficient code extraction methods,” Image and Vision Computing, vol. 31, pp. 379–391, May 2013.

[11] Y. Hannad, I. Siddiqi, M. E. El Kettani, “Writer identification using texture descriptors of handwritten fragments,” Expert Systems with Applications, vol. 47, pp. 14–22, April 2016.

[12] S. He, M. Wiering and L. Schomaker, “Junction detection in handwritten documents and its application to writer identification,” Pattern Recognition, vol. 48, pp. 4036–4048, December 2015.

[13] S. He and L. Schomaker, “Writer identification using curvature-free features,” Pattern Recognition, vol. 63, pp. 451–464, March 2017.

[14] F. A. Khan, M. A. Tahir, F. Khelifi, A. Bouridane, and R. Almotaeryi, “Robust off-line text independent writer identification using bagged discrete cosine trans- form features”, Expert Systems with Applications, vol. 17, pp. 404–415, April 2017.

[15] U. V. Marti and H. Bunke, “The IAM-database: an english sentence database for offine handwriting recognition,” International Journal on Document Analysis and Recognition IJDAR, vol. 5, pp. 39–46, November 2002.

[16] H. El Abed and V. Märgner, “ICDAR 2009-Arabic handwriting recognition competition,” International Journal on Document Analysis and Recognition IJDAR, vol. 14, pp. 3–13, Mars 2011.

[17] T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 971–987, August 2002.

[18] N. Otsu. “A Threshold Selection Method from Gray-level Histograms,” IEEE Transactions on Systems, Man and Cybernetics, vol. 9, pp. 62–66, January 1979.

[19] M. Pechwitz, S. S. Maddouri, V. Märgner, N. Ellouze, and H. Amiri, “IFN/ENIT database of handwritten arabic words,” 7th Colloque International Francophone sur l’Ecrit et le Document, CIFED 2002, 2002, pp. 129–136.

[20] H. Said, T. Tan, and K. Baker, “Personal identification based on handwriting,” Pattern Recognition, vol. 33, pp. 149–160, May 2000.

[21] L. Schomaker, “Writer identification and verification,” Advances in Biometrics: Sensors, Algorithms and Systems, Springer London, 2008, pp. 247–264.

[22] I. Siddiqi and N, Vincent, “Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features,” Pattern Recognition, vol. 43, pp. 3853–3865, November 2010.

[23] S. Srihari, S. Cha, S. Lee, and H. Arora, “Individuality of handwriting,” Journal Forensic Science, vol. 47, pp. 1–17, July 2002.