General Pattern Run-Length Transform for Writer Identification

Sheng He, Lambert Schomaker
Institute of Artificial Intelligence and Cognitive Engineering, University of Groningen, the Netherlands
Email: heshengxgd@gmail.com, L.Schomaker@ai.rug.nl
www.ai.rug.nl/~sheng/dflib

***Abstract*—In this paper we present a novel textural-based feature for writer identification: the General Pattern Run-Length Transform (GPRLT), which is the histogram of the run-length of any complex patterns. The GPRLT can be computed on the binary images (GPRLT^{bin}) or on the gray scale images (GPRLT^{gray}) without using any binarization or segmentation methods. Experimental results show that the GPRLT^{gray} achieves even higher performance than the GPRLT^{bin} for writer identification. The writer identification performance on the challenging CERUG-EN data set demonstrates that the proposed methods outperform state-of-the-art algorithms. Our source code and data set are available on www.ai.rug.nl/~sheng/dflib.**

I. Introduction

Writing style analysis is an important problem in document understanding and has a number of potential applications, such as document dating [1], [2], [3] and writer identification [4]. Given the scanned document images, writing style analysis is performed on certain features extracted from images. Therefore, features play an important role in handwriting style analysis. In this paper, we focus on the typical writer identification problem, which recognizes the authors of the handwritten text according to writing styles.

The basic assumption of writer identification is that the handwriting is individualistic and each individual has consistent writing style which is distinct from the handwriting of another individual [5]. The distinctness is from several aspects: the shapes of specific letters, width and tendency of margins or distance between written words and lines.

Run-lengths were first proposed in [6], [7] and applied for writer identification in [4], [8]. The computation of the run-length feature is very efficient without any segmentation. Traditionally, the run-length feature is extracted on the binary images in the horizontal, vertical or diagonal directions. The horizontal ink run-length reflects the average width of the text and the horizontal background run-length gives the information of the space between letters and words. The vertical ink run-length, however, reflects the structure of the letters (such as the average size of the letters) and the vertical background run-length captures the information of the space between lines. These external properties of handwritings can be used for writer identification [9].

However, the traditional run-length methods only compute the runs of values 0 and 1 on binary images. The patterns 0 and 1 are too simple to capture the complex structures. In this paper, we propose a general pattern run-length transform (GPRLT) to compute the run-lengths of more complex patterns on binary images (GPRLT^{bin}) and on gray scale images (GPRLT^{gray}) and use them for writer identification.

II. Related work

Features for writer identification can be coarsely divided into textural-based and grapheme-based groups. Several textural-based features have been used for writer identification. The first texture feature is the run-lengths proposed in [6]. Several filter-based features have been proposed for writer identification, such as the Gabor filer [10], XGabor [11] and the oriented Basic Image Features [12]. The edge-based directional probability distribution and the joint distribution of two angles of a Hinge kernel has been proposed in [13] and it has been extended to the contour-Hinge in [4] based on the contours of the handwritten text, Quill [14] which combines the ink width with the contour-Hinge and Δ^nHinge [15] which has the rotation-invariant property.

The grapheme-based features extract some ink-blob shapes and map them into a common space to build descriptors. The COnected-COmponent-COntours CO^3 was first proposed in [13] for isolated letters and it has been extended to Fraglets [4] for cursive handwritings. The small patterns without any semantic meanings has been used as graphemes in [16]. Recently, the synthesized graphemes based on the beta-elliptic model are proposed in [17] for Arabic writer identification. The junction features (Junclets) are also used for cross-script writer identification in [18] between Chinese and English.

III. General Pattern Run-Length Transform (GPRLT)

A. GPRLT on binary images

The "*run*" is defined as a sequence of connected pixels which have the some property (such as the gray value) in a given scanning line [8]. In a binary image which contains two types of pixels ('0'/'1'), there are runs of value '0' or '1' given the scanning line along a certain direction. The length of these runs can be quantized into a histogram and the normalized histogram is considered as the feature which characterizes the writer. For example, in the binary sequence "0001111010011" the run lengths of value '0' are '3,1,2' and the run lengths of value '1' are '4,1,2'.

In this paper, we propose a general pattern run-length method, which computes the run length of complex patterns,

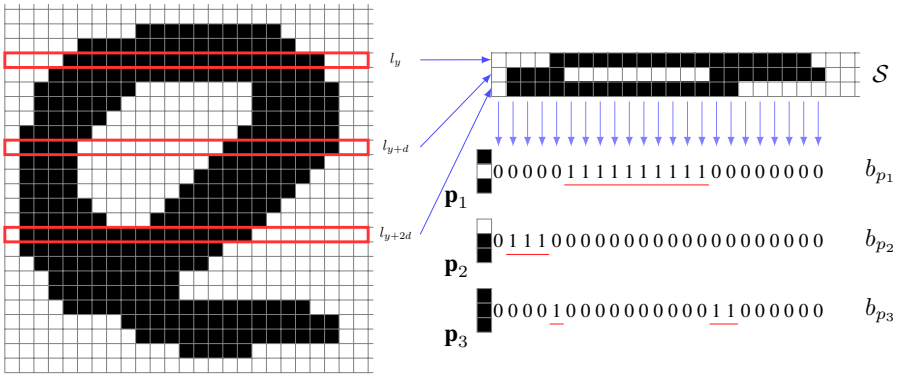

Fig. 1. The run-length of the more complex patterns $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$ on the scanning line $\mathcal{S}$ formed by the three lines l_y, l_{y+d}, l_{y+2d} with distance d.

instead of the simple '0' and '1'. Given n scanning parallel lines (the combined lines) with the inter-line distance d in the binary image, each vertical position can be described by one of the 2^n possible combinations of pixels (an example is shown in Fig. 1). For example, if there are 2 scanning lines, the types of patterns are 4, which are $\mathcal{P} = \{(0,0),(0,1),(1,0),(1,1)\}$. Given a certain pattern $\mathbf{p} \in \mathcal{P}$, the combined scanning line $\mathcal{S}(x)$ which is the sequence of patterns from $\mathcal{P}$ can be converted into 0/1 string line $b_{\mathbf{p}}(x)$ by:

$$b_{\mathbf{p}}(x) = \begin{cases} 1 & if \quad \mathcal{S}(x) = \mathbf{p} \\ 0 & otherwise \end{cases} \tag{1}$$

where x is the index of the sequence. The run length of the pattern $\mathbf{p}$ in the scanning line $\mathcal{S}$ can be computed by the run length of the value '1' in the converted string line $b_{\mathbf{p}}(x)$. Fig. 1 shows an example of the run-length with 3 scanning lines and the corresponding converted string lines of three patterns: $(0,1,0)$,$(0,1,1)$ and $(1,1,1)$. From the figure we can find that the complexity of the patterns are determined by the number of scanning lines. If there is only one scanning line, the proposed method is the traditional run-length method. The combined scanning line $\mathcal{S}$ is determined by the distance d between the nearby scanning lines. If $d = 1$, most of patterns are the "uniform" patterns which is defined as the number of spatial transitions (bitwise 0/1 changes) in the pattern is no higher than 2 (see detailed information in [19]). When d is large, the patterns on the scanned line $\mathcal{S}$ are tend to be arbitrary. We term the d as the spatial frequency and it is learned in the data set.

Compared to the traditional run-length methods, the proposed general pattern run-length method can compute the run-length of more complex patterns and captures more structure and texture information in a large space of the images. In fact, our proposed GPRLT captures the spatial co-occurrence of the binary pixels which has been shown in [20] that "the spatial co-occurrence among features could increase the discriminative power of features".

B. GPRLT on gray scale images

In this section, we present a method to extract the general pattern run-length transform on gray scale images without using any binarization method. Given a center scanning line in a gray scale image, we find m "previous" scanning lines with the inter-line distance d and m "succeeding" scanning lines. If the center scanning line is l_y, then other scanning lines are given by $\mathcal{L} = \{l_{y-md}, l_{y-(m-1)d}, \cdots, l_{y+(m-1)d}, l_{y+md}\}$, where y denotes the position of lines.

A binary test is then used between the center scanning line l_y and the scanning line $l \in \mathcal{L}$ to obtain a binary string b:

$$b(x) = \begin{cases} 1 & if \quad g^{l_y}(x) - g^{l}(x) < \theta \\ 0 & otherwise \end{cases} \tag{2}$$

where $g^l(x)$ is the pixel value on the scanning line l, x is the index and θ is the threshold. Fig. 2 illustrates a center scanning line with other four neighbors. Finally, 2^{2m} binary strings are obtained and formed the combined scanning line $\mathcal{S}$, similar as the combined scanning line obtained in binary images. Then the scanning line $\mathcal{S}$ can be converted into a binary string $b(x)$. The run-length of a given pattern is computed by counting the runs of the value '1' in the binary string $b(x)$.

Moreover, we can generate the proposed method to compute the run-length of any patterns on any scanning lines. The binary test can be defined as:

$$b(x) = \begin{cases} 1 & if \quad D\Big(\mathcal{S}(x), \mathbf{p}\Big) < \theta \\ 0 & otherwise \end{cases} \tag{3}$$

where $\mathbf{p}$ is the given pattern and $\mathcal{S}(x)$ is the element in the position x of the scanning line $\mathcal{S}$, $D\big(\mathcal{S}(x), \mathbf{p}\big)$ is the defined distance function and θ is a threshold. This method can convert the scanning line $\mathcal{S}$ into a binary string given the pattern $\mathbf{p}$. Fig. 3 illustrates an example of the processing of converting a scanning line into a binary string. We will leave this method for future works.

IV. Experimental Results

A. Data set

There are several data sets used for writer identification, such as the Firemaker [21], IAM [22] and the relatively new CERUG [18] data set. The CERUG data set contains documents written by 105 Chinese subjects with Chinese and English. The data set can be divided into three subsets: CERUG-CN which contains Chinese handwritings, CERUG-EN which

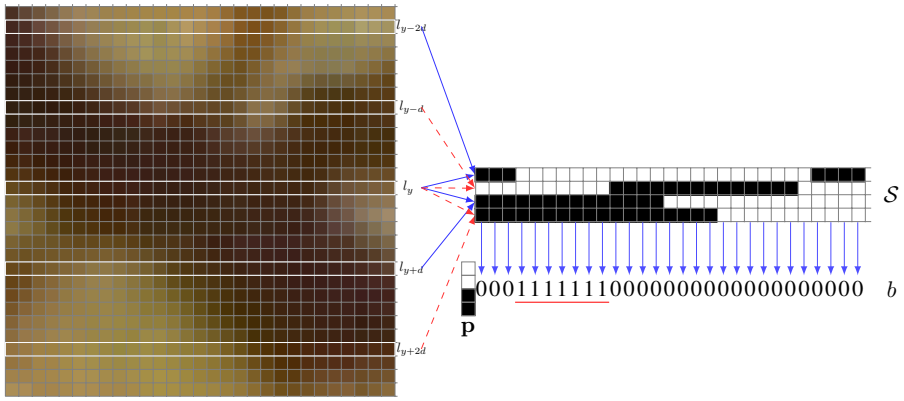

Fig. 2. The general pattern run-length computation in a gray scale image with $d = 6$.

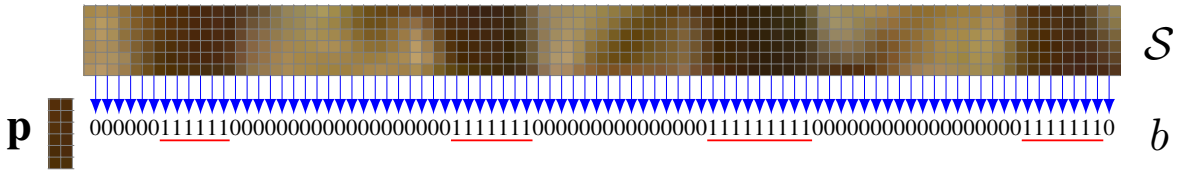

Fig. 3. The general pattern run-length of the arbitrary pattern **p** on the scanning line $\mathcal{S}$ and b is the converted binary string.

contains English handwritings and CERUG-MIXED which contains handwritings in both English letters and Chinese characters. The CERUG-EN data set is a more challenging data set in which the traditional textural-based methods failed. For example, the top-1 performance of the Hinge [4] in the CERUG-EN data set is only 12.3% with fixed parameters [18]. The main reason we have analysed in [18] is that the curvature and slant information are not very important in the English text written by Chinese people because the handwriting samples contain more long lines compared to those written by native-speaker subjects. Fig. 4 shows handwriting samples from the CERUG-EN data set. We also evaluate the proposed method on the benchmarking dataset used on ICDAR2013 Competition on Writer Identification [23]. This data set contains 1000 handwritings written by 250 writers and each writer contributed four pages (two Greek and two English). Therefore, it can be split into two subsets: ICDAR2013-English and ICDAR2013-Greek.

B. Writer identification

We use the proposed general pattern run-length transform to map the handwriting documents into feature spaces which contain the external properties of writing styles. Because the feature is a normalized histogram, the χ^2 distance is usually used to compute the distance between two documents.

The writer identification is performed with a "leave-one-out" strategy by using a nearest neighbor classification. The distances between the query document and the rest ones are computed and ordered in a sorted hit list. The query document is identified as the same author of the "Top-x" document, corresponding to "Top-x" performance. Usually, the Top-1 and Top-10 performance are reported in the literature.

TABLE I
THE TOP-1 PERFORMANCE OF WRITER IDENTIFICATION IN THE CERUG-EN DATA SET FOR NUMBER OF LINES n AND INTER-LINE DISTANCE d USING THE GENERAL PATTERN RUN-LENGTH TRANSFORM IN THE BINARY IMAGES.

System		d [pixels]									
		1	2	3	4	5	6	7	8	9	10
n	2	27.1	30.5	32.4	36.2	38.1	36.2	35.2	33.8	32.8	31.9
	3	16.7	41.4	47.6	47.6	49.1	45.7	41.4	37.6	33.8	28.6
	4	17.1	49.5	59.5	59.1	61.9	53.3	44.2	37.6	33.3	30.9
	5	11.4	49.5	64.7	68.1	71.4	62.8	53.8	47.6	43.3	36.2
	6	9.5	25.2	47.1	70.5	**75.2**	66.2	57.1	44.8	38.6	30.9
	7	2.8	11.4	30.5	59.1	62.9	57.1	41.4	31.4	27.6	22.4

TABLE II
THE TOP-1 PERFORMANCE IN THE CERUG-EN DATA SET FOR NUMBER OF LINES n AND INTER-LINE DISTANCE d USING THE GENERAL PATTERN RUN-LENGTH TRANSFORM IN THE GRAY SCALE IMAGES WITH $\theta = 90$.

System		d [pixels]									
		1	2	3	4	5	6	7	8	9	10
m	1	23.8	50.0	53.8	55.2	61.4	61.4	61.9	63.3	63.8	60.9
	2	3.3	29.1	73.3	**91.4**	**91.4**	85.2	82.4	76.2	72.3	67.1
	3	5.2	14.3	26.7	52.4	49.5	49.6	47.1	33.3	28.6	29.1
	4	10.5	20.5	25.7	33.8	35.2	29.1	28.6	28.1	30.0	23.8
	5	17.6	31.4	32.9	24.5	32.4	27.1	30.0	23.8	15.7	14.8

C. Parameter evaluation

One run-length histogram is computed from each type of patterns and all the normalized histograms of the 2^n patterns are combined together. The feature vectors of the horizontal and vertical scanning lines are concatenated together to form the final feature vector of the given document. The maximum length of runs is empirically set to 100 following the work [8].

The GPRLT^{bin} contains two parameters: the number of scanning lines n and the distance d. We use a grid search method to find the best combination of these two parameters.

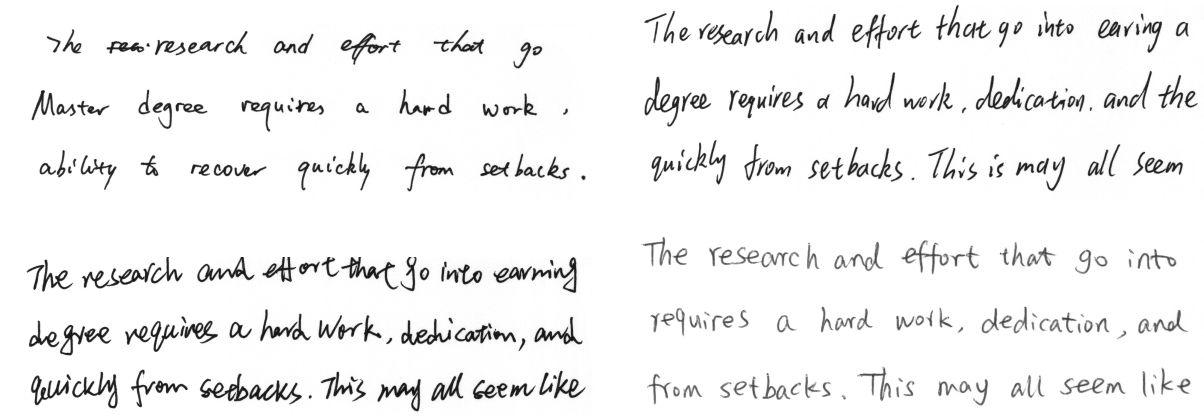

Fig. 4. Samples of handwritings from the CERUG-EN dat set.

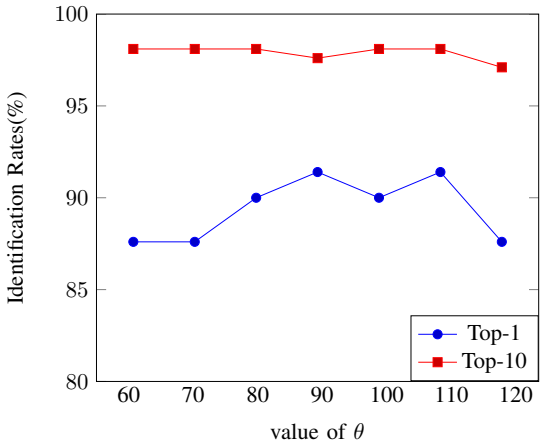

Fig. 5. The Top-1 and Top-10 performance of the GPRLT^{Gray} for the parameter θ with fixed $m = 2$ and $d = 4$.

Table I shows the Top-1 performance of writer identification with different values of n and d. For different n, the best performance is achieved when $d = 5$. The number of n determines the complexity of the patterns and the best performance of Top-1 (75.2%) is achieved when $n = 6$.

The GPRLT^{gray} contains three parameters: the number of "previous" and "succeeding" scanning lines m, the distance d and the threshold θ. Table II shows the Top-1 performance with different values of m and d given a fixed $\theta = 90$ (the gray value of the image is from 0 to 255). From the table we can see that the best Top-1 performance is achieved when $m = 2$ and $d = 4$ or $d = 5$. In this paper, we set d to 4.

Fig. 5 shows the effect of the threshold θ in the gray scale images for writer identification on the CERUG-EN data set. The Top-10 performance is quite stable with different values and the best Top-1 performance (91.4%) is achieved when $\theta = 90$ and $\theta = 110$. In this paper, we set θ to 90.

D. The performance of different patterns

In this section, we evaluate the performance of different patterns using GPRLT^{gray} on the CERUG-EN data set. When the number of scanning lines m is 2, there are $2^{2m} = 16$ types of patterns, such as $\{(0,0,0,0),(1,0,0,0),\cdots,(1,1,1,0),(1,1,1,1)\}$. We use the normalized histogram of the run-length of each pattern as a feature vector for writer identification. Fig. 6 illustrates the Top-1 and Top-10 performance of the run-length transform of each pattern. The pattern $(1,1,0,0)$ provides the best performance (the Top-1 and Top-10 performance) and the pattern $(1,0,1,0)$ gives the lowest performance. In addition, the highest identification rate is only 59.1%. Therefore, using the run-length histogram of single pattern does not achieve the optimal result.

E. Comparison with the traditional run-length methods

In this section, we compare our proposed method with the traditional run-length methods. The following features were evaluated:

- *WRL*$_h$: the white run-lengths on the horizontal scanning line.
- *WRL*$_v$: the white run-lengths on the vertical scanning line.
- *WRL*$_{hv}$: the white run-lengths combined on the horizontal and vertical scanning lines.
- *IRL*$_h$: the ink run-lengths on the horizontal scanning line.
- *IRL*$_v$: the ink run-lengths on the vertical scanning line.
- *IRL*$_{hv}$: the ink run-lengths combined on the horizontal and vertical scanning lines.
- GPRLT$^{bin}_h$(6,5): the proposed general pattern run-length transform based on the horizontal direction on the binary images with $n = 6$ and $d = 5$.
- GPRLT$^{bin}_v$(6,5): the proposed general pattern run-length transform based on the vertical direction on the binary images with $n = 6$ and $d = 5$.
- GPRLT$^{bin}_{hv}$(6,5): the proposed general pattern run-length transform based on the combination of horizontal and

Fig. 6. The Top-1 and Top-10 performance of the run-length of each pattern for writer identification using the GPRLT^{gray}.

TABLE III
THE WRITER IDENTIFICATION PERFORMANCES OF PROPOSED METHODS ON THE CERUG DATA SET.

Feature	CERUG-CN		CERUG-EN		CERUG-MIXED	
	Top1	Top10	Top1	Top10	Top1	Top10
WRL$_h$	22.9	64.8	34.3	76.7	17.1	53.3
WRL$_v$	16.7	54.8	10.0	24.8	1.9	14.3
WRL$_{hv}$	35.2	77.1	22.4	37.1	7.6	25.7
IRL$_h$	52.4	82.4	61.9	90.5	72.8	93.8
IRL$_v$	47.6	82.4	10.4	23.8	64.8	93.8
IRL$_{hv}$	73.8	88.6	20.5	44.3	86.2	97.6
GPRLT$^{bin}_h(6,5)$	77.1	92.4	74.8	95.7	72.9	96.7
GPRLT$^{bin}_v(6,5)$	72.6	92.4	36.2	87.1	65.2	96.2
GPRLT$^{bin}_{hv}(6,5)$	84.8	95.2	75.2	98.1	84.8	99.0
GPRLT$^{gray}_h(2,4,90)$	79.1	92.4	77.1	96.7	74.3	96.2
GPRLT$^{gray}_v(2,4,90)$	77.1	93.8	67.1	96.2	70.9	95.7
GPRLT$^{gray}_{hv}(2,4,90)$	88.1	95.2	91.4	97.6	84.3	99.5

vertical directions on the binary images with $n = 6$ and $d = 5$.

- GPRLT$^{gray}_h$(2,4,90): the proposed general pattern run-length transform based on the horizontal direction on the gray scale images with $m = 2$, $d = 4$ and $\theta = 90$.
- GPRLT$^{gray}_v$(2,4,90): the proposed general pattern run-length transform based on the vertical direction on the gray scale images with $m = 2$, $d = 4$ and $\theta = 90$.
- GPRLT$^{gray}_{hv}$(2,4,90): the proposed general pattern run-length transform based on the combination of horizontal and vertical directions on the gray scale images with $m = 2$, $d = 4$ and $\theta = 90$.

Table III shows the results of the performance of the traditional and the proposed run-length methods on the CERUG data set. Table IV, V, VI and VII give the performances of the different methods on the Firemaker, IAM, ICDAR2013-English and ICDAR2013-Greek data sets, respectively. From these tables we can find that our proposed general pattern run-length transform works much better than the traditional run-length methods. In addition, the performance of the proposed GPRLT^{gray} method outperform the performance of the GPRLT$^{bin}_{hv}$ in these data sets.

F. Comparison with other studies

We summarized the results of other methods proposed in the literature about writer identification on the CERUG data set in Table VIII. We follow the work in [18] to set the parameters of the Hinge [4], Quill [14], QuillHinge [14] and Junclets [18]. The Hinge and Quill features capture the slant and curvature information of the ink contours and these features are failed in the CERUG-EN data set. The Junclets feature is the grapheme-based feature which captures the singular structural

TABLE IV
WRITER IDENTIFICATION PERFORMANCE OF DIFFERENT RUN-LENGTH FEATURES ON THE **FIREMAKER** DATA SET. THE NUMBERS REPRESENT RECOGNITION PERCENTAGES.

Methods	Top-1	Top-10	Methods	Top-1	Top-10
WRL$_h$	21.4	55.2	IRL$_h$	23.4	48.0
WRL$_v$	16.6	51.0	IRL$_v$	33.4	58.8
WRL$_{hv}$	41.2	76.2	IRL$_{hv}$	44.6	67.4
GPRLT$^{bin}_h$(6,5)	47.8	78.2	GPRLT$^{gray}_h$(2,4,90)	60.2	85.6
GPRLT$^{bin}_v$(6,5)	57.4	84.0	GPRLT$^{gray}_v$(2,4,90)	58.2	84.0
GPRLT$^{bin}_{hv}$(6,5)	61.4	88.8	GPRLT$^{gray}_{hv}$(2,4,90)	67.8	89.4

TABLE V
WRITER IDENTIFICATION PERFORMANCE OF DIFFERENT RUN-LENGTH FEATURES ON THE **IAM** DATA SET. THE NUMBERS REPRESENT RECOGNITION PERCENTAGES.

Methods	Top-1	Top-10	Methods	Top-1	Top-10
WRL$_h$	13.7	36.5	IRL$_h$	37.6	68.1
WRL$_v$	13.9	36.5	IRL$_v$	54.8	81.2
WRL$_{hv}$	31.4	58.0	IRL$_{hv}$	71.2	89.0
GPRLT$^{bin}_h$(6,5)	56.5	78.5	GPRLT$^{gray}_h$(2,4,90)	66.5	88.1
GPRLT$^{bin}_v$(6,5)	61.6	83.5	GPRLT$^{gray}_v$(2,4,90)	67.0	89.5
GPRLT$^{bin}_{hv}$(6,5)	69.7	89.3	GPRLT$^{gray}_{hv}$(2,4,90)	78.3	92.5

TABLE VI
WRITER IDENTIFICATION PERFORMANCE OF DIFFERENT RUN-LENGTH FEATURES ON THE **ICDAR2013-ENGLISH** DATA SET. THE NUMBERS REPRESENT RECOGNITION PERCENTAGES.

Methods	Top-1	Top-10	Methods	Top-1	Top-10
WRL$_h$	12.2	42.8	IRL$_h$	42.8	74.4
WRL$_v$	5.4	28.6	IRL$_v$	46.2	78.2
WRL$_{hv}$	18.6	55.0	IRL$_{hv}$	66.4	89.6
GPRLT$^{bin}_h$(6,5)	51.0	82.2	GPRLT$^{gray}_h$(2,4,90)	64.0	91.6
GPRLT$^{bin}_v$(6,5)	60.0	86.6	GPRLT$^{gray}_v$(2,4,90)	60.4	89.4
GPRLT$^{bin}_{hv}$(6,5)	71.0	93.8	GPRLT$^{gray}_{hv}$(2,4,90)	81.2	96.4

TABLE VII
WRITER IDENTIFICATION PERFORMANCE OF DIFFERENT RUN-LENGTH FEATURES ON THE **ICDAR2015-GREEK** DATA SET. THE NUMBERS REPRESENT RECOGNITION PERCENTAGES.

Methods	Top-1	Top-10	Methods	Top-1	Top-10
WRL$_h$	14.4	57.4	IRL$_h$	51.0	81.8
WRL$_v$	10.0	32.0	IRL$_v$	54.0	88.2
WRL$_{hv}$	26.6	62.4	IRL$_{hv}$	78.8	93.4
GPRLT$^{bin}_h$(6,5)	51.0	79.6	GPRLT$^{gray}_h$(2,4,90)	71.4	91.2
GPRLT$^{bin}_v$(6,5)	52.0	82.0	GPRLT$^{gray}_v$(2,4,90)	65.4	91.0
GPRLT$^{bin}_{hv}$(6,5)	67.4	90.0	GPRLT$^{gray}_{hv}$(2,4,90)	82.8	96.6

TABLE VIII
THE WRITER IDENTIFICATION PERFORMANCES OF DIFFERENT METHODS ON THE CERUG DATA SET.

Feature	CERUG-CN		CERUG-EN		CERUG-MIXED	
	Top1	Top10	Top1	Top10	Top1	Top10
Hinge [4]	90.8	96.2	12.3	30.0	84.7	95.7
Quill [14]	82.7	92.3	15.8	48.6	74.8	93.3
QuillHinge [14]	88.5	93.8	45.2	91.0	86.7	98.6
Junclets [18]	**90.4**	**97.1**	87.1	96.2	**85.7**	98.5
$\text{GPRL}_{hv}^{gray}(2,4,90)$	88.1	95.2	**91.4**	**97.6**	84.3	**99.5**

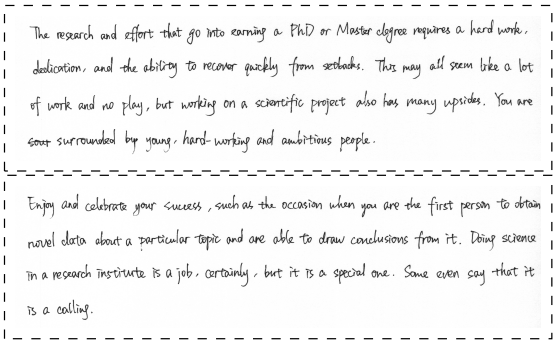

Fig. 7. Two samples written by the same writer, which have been recognized correctly by the proposed method and incorrectly identified by the Hinge feature.

information in the handwritten text. From the Table VIII we can find that the proposed method get the best results on the CERUG-EN data set and obtain the comparable results on CERUG-CN and CERUG-MIXED data sets. In addition, the computation of the proposed methods is more efficient than others. The GPRLT^{gray} method does not need any binarization or segmentation methods. The computational operations only contain are the binary test in Eq. (2) and the counting operation of the run lengths.

G. Analysis

Fig. 7 demonstrates an example which has been correctly recognized by the proposed general pattern run-length transform and false rejected by the Hinge. From this figure, we can find that the structures of these two samples (such as the space between letters and words) are quite same and we can easily judge by the eyes that they are from the same hand.

V. CONCLUSION

This paper has proposed a general pattern run-length transform which counting the runs of the complex patterns and can be used on the binary images or on the gray scale images. The proposed methods are more discriminative than the traditional run-length method. We used the proposed method for writer identification on four public data sets and experimental results have emonstrated that our proposed method outperforms state-of-the-art approaches on the challenging CERUG-EN data set.

ACKNOWLEDGMENTS

This work has been supported by the Dutch Organization for Scientific Research NWO (project No. 380-50-006).

REFERENCES

[1] S. He, P. Sammara, J. Burgers, and L. Schomaker, "Towards style-based dating of historical documents," in *International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 2014, pp. 265–270.
[2] S. He and L. Schomaker, "A polar stroke descriptor for classification of historical documents," in *International Conference on Document Analysis and Recognition (ICDAR)*, 2015, pp. 6–10.
[3] S. He, P. Sammara, J. Burgers, and L. Schomaker, "Historical document dating using unsupervised attribute learning," in *IAPR Workshop on Document Analysis Systems (DAS)*, 2015.
[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, no. 4, pp. 701–717, 2007.
[5] S. N. Srihari, S.-H. Cha, H. Arora, and S. Lee, "Individuality of handwriting," *Journal of Forensic Sciences*, vol. 47, no. 4, pp. 856–872, 2002.
[6] M. M. Galloway, "Texture analysis using gray level run lengths," *Computer graphics and image processing*, vol. 4, no. 2, pp. 172–179, 1975.
[7] B. Arazi, "Handwriting identification by means of run-length measurements," *IEEE Trans. Syst., Man and Cybernetics*, no. 12, pp. 878–881, 1977.
[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, no. 10, pp. 1196–1202, 2013.
[9] B. Arazi, "Automatic handwriting identification based on the external properties of the samples," *IEEE Transactions on Systems, Man and Cybernetics*, no. 4, pp. 635–642, 1983.
[10] H. E. Said, T. N. Tan, and K. D. Baker, "Personal identification based on handwriting," *Pattern Recognition*, vol. 33, no. 1, pp. 149–160, 2000.
[11] B. Helli and M. E. Moghaddam, "A text-independent persian writer identification based on feature relation graph (FRG)," *Pattern Recognition*, vol. 43, no. 6, pp. 2199–2209, 2010.
[12] 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, 2014.
[13] 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.
[14] A. Brink, J. Smit, M. Bulacu, and L. Schomaker, "Writer identification using directional ink-trace width measurements," *Pattern Recognition*, vol. 45, no. 1, pp. 162–171, 2012.
[15] S. He and L. Schomaker, "Delta-n hinge: Rotation-invariant features for writer identification," in *International Conference on Pattern Recognition (ICPR)*, 2014, pp. 2023–2028.
[16] I. Siddiqi and N. Vincent, "Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features," *Pattern Recognition*, vol. 43, no. 11, pp. 3853–3865, 2010.
[17] M. N. Abdi and M. Khemakhem, "A model-based approach to offline text-independent Arabic writer identification and verification," *Pattern Recognition*, vol. 48, no. 5, pp. 1890–1903, 2015.
[18] S. He, M. Wiering, and L. Schomaker, "Junction detection in handwritten documents and its application to writer identification," *Pattern Recognition*, vol. 48, no. 12, pp. 4036–4048, 2015.
[19] 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, no. 7, pp. 971–987, 2002.
[20] X. Qi, R. Xiao, C.-G. Li, Y. Qiao, J. Guo, and X. Tang, "Pairwise rotation invariant co-occurrence local binary pattern," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2199–2213, 2014.
[21] L. Schomaker and L. Vuurpijl, "Forensic writer identificaiton: a benchmark data set and a comparison of two systems," *Technical Report, Nijmegen: NICI*, 2000.
[22] 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.
[23] 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.