Rotation-free Online Handwritten Character Recognition using Dyadic Path Signature Features, Hanging Normalization, and Deep Neural Network

Weixin Yang[1], Lianwen Jin[1*], Hao Ni[2], Terry Lyons[2]
[1] College of Electronic and Information Engineering, South China University of Technology, Guangzhou, China
[2] Mathematical Institute and Oxford-Man Institute of Quantitative Finance, University of Oxford, Oxford, UK
wxy1290@163.com, *lianwen.jin@gmail.com, hao.ni@maths.ox.ac.uk, tlyons@maths.ox.ac.uk

***Abstract*—The path signature feature (PSF) which was initially introduced in rough paths theory as a branch of stochastic analysis, has recently been successfully applied to the field of pattern recognition for extracting sufficient quantity of information contained in a finite trajectory, but with potentially high dimension. In this paper, we propose a variation of path signature representation, namely the dyadic path signature feature (D-PSF), to fully characterize the trajectory using a hierarchical structure to solve the rotation-free online handwritten character recognition (OLHCR) problem. We adopt the deep neural network (DNN) as classifier, and investigate three hanging normalization methods to improve the robustness of the DNN to rotational distortions. Extensive experiments on digits, English letters, and Chinese radicals demonstrated that the proposed D-PSF, jointly with hanging normalization and DNN, achieved very promising results for rotated OLHCR, significantly outperforming previous methods.**

Keywords—rotation-free recognition; online handwritten character recognition; path signature feature; rotation normalization; neural network

I. INTRODUCTION

With the development and commercialization of pen-based or touch-screen devices, online handwritten character recognition (OLHCR) is becoming increasingly important, and has been intensively studied to maintain a small footprint [1–3], yield high recognition accuracy [4–7], yet maintain robustness in its use for unconstrained handwriting styles. Many existing handwriting input method editors (IME) do fulfill most of the above goals, but they often fail to achieve robustness to rotational distortion. This is because rotational distortion may reduce the recognition accuracy by generating confusion for some characters that are originally distinguishable. However, despite the possible reduction in accuracy, a rotation-free recognizer can deliver a compelling user experience in special cases, such as supporting multi-user operations in association with the device, recognizing handwritings written in free styles, or during emergencies, and so on. In order to facilitate these applications of the handwriting IME, rotated OLHCR merits more attention.

In OLHCR, the popular directional feature extracted from the decomposed directions of adjacent sampling points, is

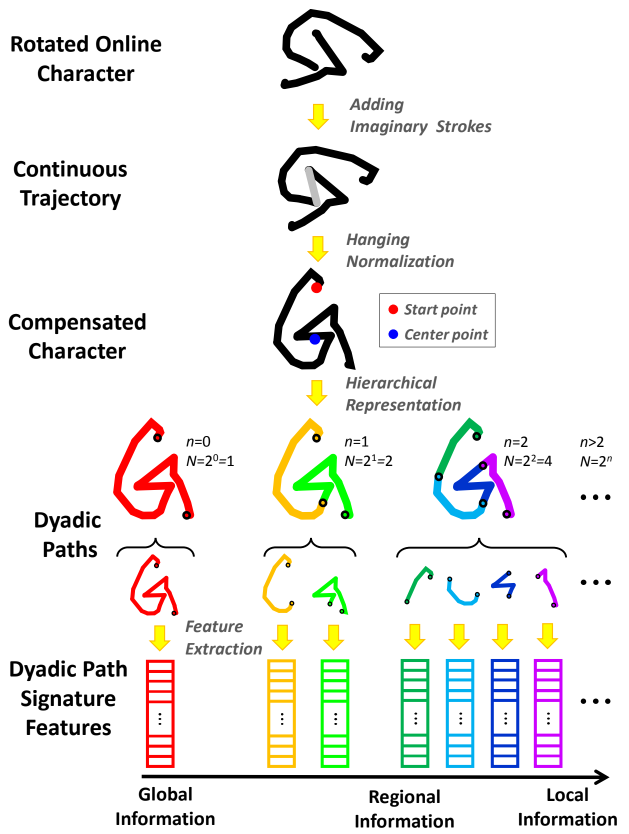

Fig. 1. Illustration of the proposed dyadic path signature features. The hanging normalization in [14] is adopted as preprocessing step in this figure. The parameter n is the hierarchical level, and N is the number of dyadic paths at the corresponding level. When the level n is small, the features represent global information. When n increases, more local information is involved. These hierachical features help us to obtain global, regional, and local information.

invariant to local directional variations of strokes, which is one of the valuable characteristics for recognition. However, the directional feature is hard to absorb the rotational distortion which can be regarded as a global transformation of characters. In [10], to exploit global information in a rotated trajectory, the path signature feature (PSF in brief) developed in [8–9], is introduced to extract features by considering the entire character as a path. It has been proven to be effective, but the dimensionality increases fast when higher degrees of a signature feature are required to describe more detailed (local) information in the entire path.

In this paper, we propose a variation of PSF using a hierarchical path structure, namely the dyadic path signature feature (D-PSF), to overcome the aforementioned problems. At a certain hierarchical level, we equally divide a path to generate two smaller paths for the next level of the hierarchy. Thus, the number of paths is equal to 2^n at the n^{th} level of the hierarchy. The D-PSF is then considered as the combination of the PSFs of both the entire path (i.e., the 0^{th} level) and all the derived smaller paths. The illustration of the D-PSF is presented in Fig. 1. This modification allows the features to capture both the global (or local) information and the regional information. Since the large dimensional features make different contributions to rotation-free recognition, we adopt the deep fully-connected neural network as classifier for weight adjustment. Moreover, we employ the efficient hanging normalization methods to further improve the performance of the rotated, isolated OLHCR. The elicited experimental results on three subsets (i.e., digits, English letters, and Chinese radicals) of the CASIA-OLHWDB1.1 [4] demonstrate that the proposed D-PSF outperforms other features and achieves a higher accuracy in association with hanging normalization and DNN.

The rest of the paper is organized as follows. The proposed dyadic path signature feature is given in Section II. The hanging normalization methods are discussed in Section III. The explanation of the adoption of DNN is presented in Section IV. Two relevant handwritten character recognition technologies we used are described in Section V. Extensive experimental results are reported in Section VI. Finally, we conclude the paper in Section VII.

II. DYADIC PATH SIGNATURE FEATURE

The theory of rough paths can be thought as a non-linear extension of classical theory of controlled differential equations driven by very irregular paths, and the essential object in rough paths theory is the path signature, which was first studied by a geometer Chen in the form of iterated integrals [8] and was further developed by Lyons [9, 11]. The PSF is capable to extract important analytic and geometrical properties from a path. Diehl [12] first used iterated integrals of a curve in OLHCR, and Graham [5] then introduced this feature to a large-scale recognition task that allowed him to win the ICDAR2013 competition [13].

First, we present a brief introduction of the PSF. An online trajectory (path) of finite length can be described as a continuous mapping function $P:[0,T]\to\mathbb{R}^d$ within the time interval $[0,T]\subset\mathbb{R}$. The dimension d is equal to two if the trajectory is on a plane, or larger than two if more information of the trajectory is considered. For an integer $k\geq 1$, and a collection of indices $i_1,i_2,...,i_k\in\{1,2,...,d\}$, the k-fold iterated integral of the path along the indices $i_1,i_2,...,i_k$ can be defined as

$$S(P)_{0,T}^{i_1,i_2,...,i_k}=\int_{0<\tau_k<T}\cdots\int_{0<\tau_2<\tau_3}\int_{0<\tau_1<\tau_2}dP_{\tau_1}^{i_1}dP_{\tau_2}^{i_2}...dP_{\tau_k}^{i_k}, \quad (1)$$

where $0<\tau_1<\tau_2<...<\tau_k<T$. The signature of P consists of all the iterated integrals of P, and the dimension of $S(P)$ is infinite. For practical use, the signature truncated at degree m, also known as the m^{th} degree truncated signature, is adopted to ensure that the signature feature is finite dimensional. The truncated signature is the collection of the iterated integrals whose degree is no greater than m,

$$\begin{aligned}TS(P)_{0,T}^{(m)}=(1,S(P)_{0,T}^{1},...S(P)_{0,T}^{d},S(P)_{0,T}^{1,1},...,\\ S(P)_{0,T}^{d,d},...,S(P)_{0,T}^{i_1,i_2,...,i_m},...,S(P)_{0,T}^{d,d,...,d})\end{aligned} \quad (2)$$

By simple calculation, the dimension is equal to $(d^{m+1}-1)(d-1)^{-1}$. It is noted that the higher degree of truncated signature is required in order to capture a finer description of a path, but it leads to that the dimension of signature increases exponentially.

To alleviate this problem, the proposed D-PSF replaces the higher degrees of PSF with a collection of the lower degrees of PSF extracted from smaller pieces of the entire path. More specifically, we divide the entire path into dyadic pieces and set up a hierarchical representation of the path. The n^{th} hierarchical level of D-PSF is the collection of the signatures of N dyadic pieces,

$$D^nS(P)_{0,T}=(S(P)_{0,\frac{1}{N}T},...,S(P)_{\frac{i-1}{N}T,\frac{i}{N}T},...,S(P)_{\frac{N-1}{N}T,T}), \quad (3)$$

where $N=2^n$, $n\in\mathbb{N}$. It is obvious that when $n=0$, $D^0S(P)_{0,T}=S(P)_{0,T}$ is the PSF of the entire path, so it can be regarded as the global representation of path P. When n increases, the dyadic pieces of a path have shorter length, so $D^nS(P)_{0,T}$ provides the regional information. If n is large, then local information can be extracted from $D^nS(P)_{0,T}$. If we would like to keep all these information, then the D-PSF is given by:

$$DS(P)_{0,T}=(D^0S(P)_{0,T},D^1S(P)_{0,T},...,D^nS(P)_{0,T}). \quad (4)$$

The number of dyadic pieces in eq. (4) is $2^{n+1}-1$. The level n cannot be infinite because the path is finite in the first place. The k^{th} degree signature of this path can be calculated by that of the dyadic pieces, according to Chen's identity [8] which states that

$$S(P)_{0,T}^{(k)}=\sum_{j=0}^{k}S(P)_{0,S}^{(j)}\otimes S(P)_{S,T}^{(k-j)}, \quad (5)$$

where $0<S<T$, and $\otimes$ denotes a tensor product and the superscript denotes the length of indices. By extension theorem in rough paths theory [21], the information provided by the higher degree signature of the entire path can be well approximated by the lower degree signatures over the dyadic partition of this path when the hierarchical level n is large.

Note that for practical use, functions $S(P)$ in eq. (3) and (4) are usually replaced by $TS(P)$ as follows:

$$DTS(P)_{0,T}^{(m)}=(D^0TS(P)_{0,T}^{(m)},D^1TS(P)_{0,T}^{(m)},...,D^nTS(P)_{0,T}^{(m)}). \quad (6)$$

The dimension of (6) is $(2^{n+1}-1)(d^{m+1}-1)(d-1)^{-1}$, which increases linearly in terms of the number of dyadic partition ($2^{n+1}-1$), but exponentially with respect to m.

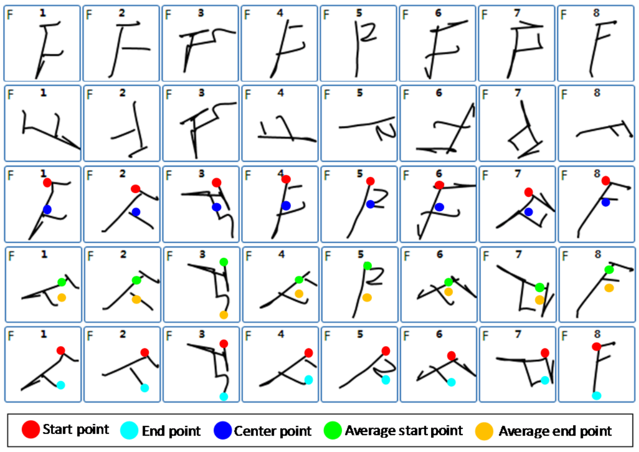

Fig. 2. Examples of the handwritten letter ‘F’ using different hanging normalization methods. The first row denotes the original characters (prototypes), and the second row is the series of the corresponding randomly rotated characters. The last three rows represent the compensated characters using the SC, ASE, and SE hanging normalization methods respectively. Note that only the 4th letter follows the standard stroke order, and the 3rd letter is completed in one stroke.

In the rotated OLHCR experiments in Section VI, we employ the features described in (6) and evaluate the effects of the truncated degree *m,* and the hierarchical level *n,* on the performance of the networks using D-PSF. Since the 0th degree of truncated signature is defined as 1 by convention, so we do not include it in our features and the actual dimension is $(2^{n+1}-1)(d^{m+1}-d)(d-1)^{-1}$.

III. Hanging Normalization methods

In rotated OLHCR, the hanging normalization first proposed in [14], is one of the most intuitive and efficient ways to achieve rotation-invariance. It assumes that the relative positions of some key points in an online character are stable under the rotation. Thus, the direction between the key points can be used to compensate the distorted characters and transform them to a relatively stable position for recognition. These key points often appear in pairs, such as the start and center points [14] (SC for short), the start and end points [15] (SE), the average start and end points of all the strokes [3] (ASE), etc. Fig. 2 presents some examples before and after hanging normalization.

Without loss of generality, let us consider the SC hanging normalization as an example. Given an online handwritten character sample with T sampling points $P=(P_1,P_2,\dots,P_t,\dots,P_T)$, where $1\le t\le T$, we define the x-y plane as the rotation plane, and (x_t,y_t) as the only two coordinates associated with the rotation. The start point and the center point can then be defined as

$$S=P_1=(x_1,y_1), \tag{7}$$

$$C=(\bar{x},\bar{y})=\Big(\frac{1}{T}\sum\nolimits_1^T x_i,\frac{1}{T}\sum\nolimits_1^T y_i\Big). \tag{8}$$

For a character, hanging normalization is the transformation of each point to a new position to change the direction from C to S as follows:

$$\begin{aligned} P_t' &= (x_t',y_t')=(x_t,y_t)\begin{pmatrix}\cos(\theta-\frac{\pi}{2}) & -\sin(\theta-\frac{\pi}{2})\\ \sin(\theta-\frac{\pi}{2}) & \cos(\theta-\frac{\pi}{2})\end{pmatrix}\\ &=(x_t,y_t)\begin{pmatrix}\sin\theta & \cos\theta\\ -\cos\theta & \sin\theta\end{pmatrix}, \end{aligned} \tag{9}$$

where

$$\cos\theta=\frac{\bar{x}-x_1}{\sqrt{(\bar{x}-x_1)^2+(\bar{y}-y_1)^2}}, \tag{10}$$

$$\sin\theta=\frac{\bar{y}-y_1}{\sqrt{(\bar{x}-x_1)^2+(\bar{y}-y_1)^2}}. \tag{11}$$

After the hanging normalization, the compensated character P' may not be in the original regular position, but the intra-class characters can be similar with each other if they are normalized in the same way.

However, we must emphasize that these informative key points that benefited from the stability of the online strokes, are most affected by the conditions of the strokes. In unconstrained handwritings, the order, direction, quantity, and connection of the strokes are no longer regular, so the key points of hanging normalization may become unstable, as shown in Fig. 2. Thus, the selection of different hanging normalization methods should depend on the stroke conditions of the handwritten objects. Further analysis, based on experiments, will be presented in Section VI.

IV. Deep neural network

The extraction of D-PSF provides sufficient features which have different importance in recognizing rotated characters, we thus adopt the deep fully-connected neural network as classifier to iteratively adjust the weights of features. For DNN, the ability in classification is not only dependent on the structure or hyper parameters of network, but also influenced by the difficulty of the task. A challenging task can make full use of the learning capacity of neural networks so that effective features can be selected through training [16]. In rotated OLHCR, the rotational variations of samples give rise to the task difficulty and facilitate the DNN to use rotation-invariant features for recognition. However, in some cases of our implementation, the hanging normalization may reduce the variations of samples and cause an information loss. Therefore, we suggest that the normalization operations can be considered as a fine-tuning step instead of a preprocessing step when facing a challenging task, for example in rotated OLHCR, the rotated characters are used for training, and the compensated characters are used for fine-tuning and testing.

For all the experiments in this paper, our fully-connected neural network with six hidden layers, is trained by back-propagation. The number of neurons is set to 128 for the first layer, and increases in increments of 128 for the following layers. The mini-batch size is set to 100. We use the method of stochastic gradient descent with a momentum value equal to 0.9. The learning rate updates in accordance to $\alpha(t)=\alpha(0)\cdot\exp(-\lambda t)$, where $\alpha(0)=0.002,\ \lambda=5\times10^{-4}$. The dropout [20] rates for the last five layers are set to 0.1, 0.2, 0.3, 0.4, and 0.5, respectively.

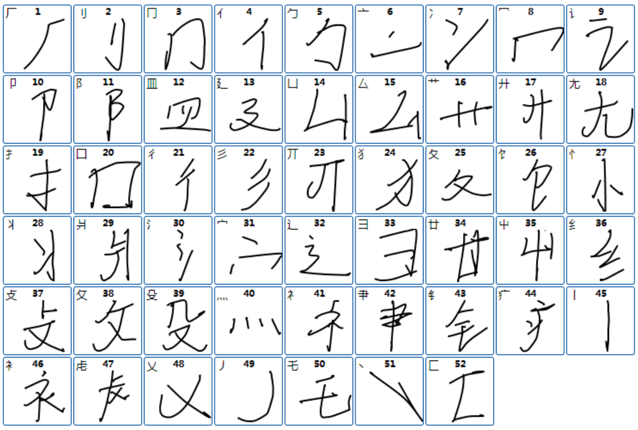

Fig. 3. Handwritten examples of the 52 Chinese radicals. The actual label of each example is shown in the upper left corner. Note that many radicals are similar with each other even without rotational distortion.

V. Handwritten Character Recognition Technologies

A. Imaginary stroke technique

The imaginary strokes [17–18], which are those pen-moving trajectories when the pen tip leaves the writing plane, are sometimes virtual straight lines since they are not recorded in the data. One intuitive way in which imaginary strokes embed to a multi-stroke character, while being distinguishable from the real strokes, is adding an "ink" dimension [15] to the original 2D points. The release of "ink" increases in the pen-down state and ceases in the pen-up state, so the entire 3D trajectory is continuous without breaking or overlapping points.

B. 8-directional feature extraction

The 512-dimensional 8-directional feature extraction method in [19] is adopted. In order to incorporate the imaginary stroke technique, we extract a second 512-dimensional vector from the character in which all the real strokes and imaginary ones are concatenated into a single stroke. Finally, a total of 1024-dimensional, 8-directional features are obtained for each character.

VI. Experiments

A. Experimental datasets

The experiments are conducted on digits (10 classes), English upper letters (26 classes) and Chinese radicals (52 classes, as shown in Fig. 3) of CASIA-OLHWDB1.1 (DB1.1 in brief), because these three datasets contain online characters with different structural complexities and various average stroke numbers. Each class has approximately 240 prototype samples for training and 60 for testing. In the training stage of rotated OLHCR, the rotated characters are generated artificially from the prototypes, and the rotation angle is randomized at each iterative step. The training of the digits, letters, and radicals, are completed after 3000, 4000, and 5000 epochs, respectively. In the test stage, each prototype generates 30 testing samples through sequential rotations by an angle increment of 12° to fully simulate different rotations. Thus, the rotated test set is 30 times larger than the prototype test set.

TABLE I. Results of D-PSF on digits of DB1.1

Test error rate (%)		Hierarchical Level of D-PSF			
		***n:0* [15]**	***n:0–1***	***n:0–2***	***n:0–3***
Truncated Degree of PSF	***m:1***	60.25	4.97	1.52	1.43
	m:2	10.03	1.67	0.92	0.97
	m:3	2.86	1.35	0.77	**0.73**
	m:4	1.50	1.19	1.07	0.97

TABLE II. Results of D-PSF on English letters of DB1.1

Test error rate (%)		Hierarchical Level of D-PSF			
		***n:0* [15]**	***n:0–1***	***n:0–2***	***n:0–3***
Truncated Degree of PSF	***m:1***	84.88	23.79	6.14	5.12
	m:2	28.76	7.02	4.49	4.50
	m:3	14.23	5.44	4.49	4.21
	m:4	9.04	4.76	4.50	**4.09**

TABLE III. Results of D-PSF on Chinese radicals of DB1.1

Test error rate (%)		Hierarchical Level of D-PSF			
		***n:0* [15]**	***n:0–1***	***n:0–2***	***n:0–3***
Truncated Degree of PSF	***m:1***	91.14	38.33	11.81	9.44
	m:2	42.18	11.56	8.86	8.37
	m:3	18.48	10.87	8.68	**7.91**
	m:4	15.51	10.25	8.89	7.98

TABLE IV. Number of dimensions for various settings of D-PSF

Number of dimensions		Hierarchical Level of D-PSF			
		***n:0* [15]**	***n:0–1***	***n:0–2***	***n:0–3***
Truncated Degree of PSF	***m:1***	3	9	21	45
	m:2	12	36	84	180
	m:3	39	117	273	585
	m:4	120	360	840	1800

B. Evaluation of dyadic path signature features

In the first set of experiments, we study the performance of different combinations of the truncated degrees *m,* and the hierarchical level *n* of D-PSF in rotated OLHCR. We use "0–*n*" to denote the collection of levels of D-PSF. The results are given in Tables I – III and the corresponding number of dimensions are listed in Table IV. In the case when $n = 0$ (i.e., the method used in [15]), the performances improve when higher truncated degrees of PSF are applied, even though features are extracted only from the entire path. When higher levels of dyadic paths are involved to extract regional and local information, the results are significantly improved. Note that from the results of the last columns, when level *n* is large enough, we may not need a very high degree of *m* of PSF to achieve better results. It is because a certain degree truncated PSF of the dyadic small paths contain sufficient regional and local information which can well approximate a higher degree PSF of the entire path. Actually, when degree *m* increases, the number of dimensions increases exponentially. We observe that, most of the time, involving a

TABLE V. RESULTS OF DIFFERENT FEATURES AND DIFFERENT HANGING NORMALIZATION METHODS ON THREE DATASETS OF DB1.1

Test error rates (%)		**Rotation-free OLHCR**			
		No Hanging	***Hanging Normalization***		
			***SC* [14]**	***ASE* [3]**	***SE* [15]**
Digits	Bitmap	17.06	4.03	4.36	4.70
	8-directional [3,19]	1.88	1.85	1.01	1.17
	D-PSF (proposed)	0.73	**0.67**	0.84	1.01
English Upper Letters	Bitmap	18.99	8.35	9.44	10.02
	8-directional [3,19]	4.99	4.57	4.88	4.69
	D-PSF (proposed)	4.21	**3.70**	4.01	4.12
Chinese Radicals	Bitmap	28.14	13.96	14.45	14.68
	8-directional [3,19]	10.57	9.13	10.24	10.44
	D-PSF (proposed)	7.91	8.53 (**7.31**)*	8.89 (7.43)	9.05 (7.57)

* The results in the brackets come from the training of the networks with rotated characters, fine-tuned, and tested with compensated characters.

higher level *n,* results in fewer additional dimensions and better performance than increasing the degree *m*. For the following experiments, we adopt the settings of levels n=0–3 and degree m=3 because of its high performance on average. It is suggested that the dyadic level should not be too high, because if it did, the length of the dyadic paths would be too small to be insensitive to local noise of sampling points.

C. Investigation of different features and hanging normalization methods

We compare three kinds of features including baseline bitmap features, 8-directional features, and D-PSF. The baseline network is supplied with data that is directly rendered from the online characters as 40×40 offline bitmaps without any ordered information, so its input features are of dimension 1600. The 8-directional features has dimension 1024, while the D-PSF we used has only dimension 585. As shown in Tables V, the bitmap features using offline information perform poorer than those using sequential information. Compared with bitmap features and 8-directional features, our proposed D-PSF yields significantly better performance, irrespective on whether the hanging normalization is employed or not.

The hanging normalization can be considered as either a preprocessing step or a fine-tuning step in our DNN. When hanging normalization is a preprocessing step, most results outperform those of the networks without using hanging, except for the results of D-PSF on the Chinese radical dataset. It is attributed to the fact that the rotated radical recognition is a more difficult task than the letter or digit recognition of which the networks easily learn and quickly fit the data even under rotation, so, the hanging normalization in rotated radical recognition reduces the variations of samples, which is harmful for the network to find out effective features to solve the hard problem. To prove this viewpoint, we regard the hanging normalization as a fine-tuning step for rotated radicals, and deliberately cease the training of the networks after 4000 epochs and use the compensated characters to fine tune the DNN for the remaining 1000 epochs. Eventually, the test error rates of using hanging normalization achieve the best performance in our rotated OLHCR.

Among different hanging normalization methods, the SC method [14] usually outperforms others because the character center points are by definition more stable than the start and end points. It is noted that in CASIA-OLHWDB1.1, handwritten characters are either cursive or neat, and most of them do not follow the standard stroke order. Therefore, the ASE method [3] performs better than the SE method [15], since it is insensitive to the stroke order. Furthermore, it is worse than the SC method because it can be inversely affected by the additional connections of cursive strokes, thereby causes some start and end points to vanish. Taking into consideration the stroke conditions in our datasets, the SC method may be a relatively better choice as proven in our experiments.

Finally, to demonstrate the proposed method is robust to rotational distortion and be able to recognize unconstrained handwritten characters, some recognition results are shown in Fig. 4. The output first candidates are shown above each sample together with the corresponding ground truth labels and rotation angles. We can see that even though the rotated characters generate confusions for some classes, the proposed method can find out the detailed differences and achieve rotation-free performance. In addition, the misclassified examples are given in red boxes. Some of them are difficult even for humans to classify.

VII. CONCLUSIONS

In this paper, we have proposed a new dyadic representation of path signature features (D-PSF) to solve the problem of rotated online handwritten character recognition problem. The proposed D-PSF can effectively extract global, regional, and local information, from the online characters for rotation-free recognition tasks. For classification, we have exploited the learning capacity of deep fully-connected neural network. We have also investigated hanging normalization as preprocessing or fine-tuning step in DNN to further improve the performance. Based on the comparisons of different features, the D-PSF has manifested an excellent ability in detailed feature representation of sequential trajectory. In regard to future work, we will use the dyadic path signature feature for more difficult pattern recognition tasks, like handwritten text recognition, writer identification, speaker verification, and others. It would be

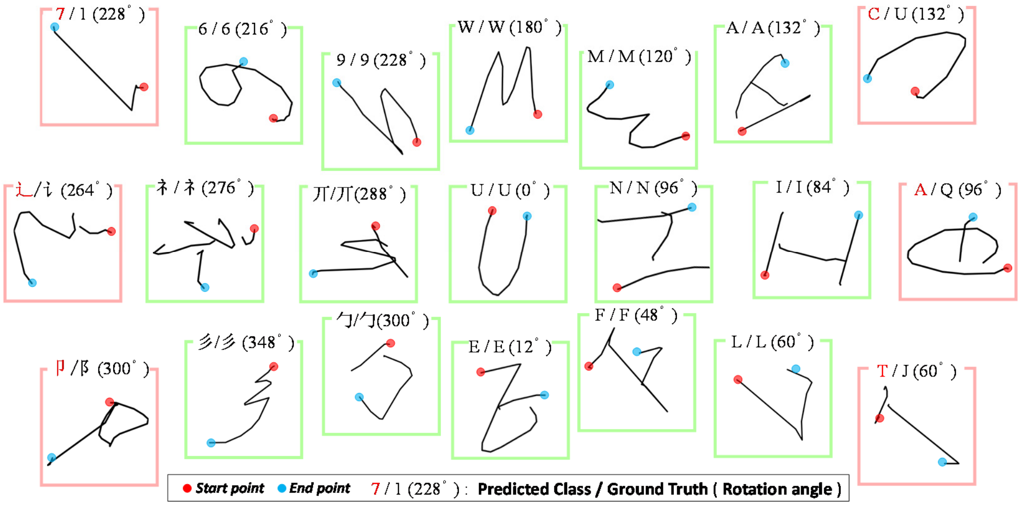

Fig. 4. The first candidate recognition results given by the proposed method for some rotated characters. The start and end points are marked to roughly describe the sequential information of each online character. The predicted classes, ground truth labels, and rotation angles are shown above the characters.

interesting to see the feature representation capability in describing other sequential data in these tasks.

Acknowledgement

This research is supported in part by NSFC (Grant No.: 61472144), GDSTP (Grant No.: 2014A010103012, 2015B010131004) and GDUPS (2011). The last author has been supported by EPSRC (grant Reference: EP/F029578/1), ERC (Grant Agreement No.291244 Esig) and the Oxford-Man Institute of Quantitative Finance. The third author is supported by ERC (Grant Agreement No.291244 Esig) and by the Oxford-Man Institute of Quantitative Finance.

References

[1] T. Long, L. Jin, "Building compact MQDF classifier for large character set recognition by subspace distribution sharing," *Pattern Recognition*, vol. 41, no. 9, pp. 2916-2925, 2008.

[2] Y. Wang, Q. Huo, "A study of designing compact recognizers of handwritten Chinese characters using multiple-prototype based classifiers," *Pattern Recognition (ICPR), IEEE International Conference on*, pp. 1872-1875, 2010.

[3] J. Du, Q. Huo, K. Chen, "Designing compact classifiers for rotation-free recognition of large vocabulary online handwritten Chinese characters," *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on*, pp. 1721-1724, 2012.

[4] C. Liu, F. Yin, D. Wang, et al, "Online and offline handwritten Chinese character recognition: benchmarking on new databases," *Pattern Recognition*, vol. 46, no. 1, pp. 155-162, 2013.

[5] B. Graham, "Sparse arrays of signatures for online character recognition," *arXiv preprint arXiv:1308.0371*, 2013.

[6] W. Yang, L. Jin, Z. Xie, Z. Feng, "Improved deep convolutional neural network for online handwritten Chinese character recognition using domain-specific knowledge," In *Document Analysis and Recognition (ICDAR), IEEE 13th International Conference on*, pp. 551-555, 2015.

[7] W. Yang, L. Jin, D. Tao, et al, "DropSample: a new training method to enhance deep convolutional neural networks for large-Scale unconstrained handwritten Chinese character recognition," *arXiv preprint arXiv:1505.05354*, 2015.

[8] K.T. Chen, "Integration of paths--A faithful representation of paths by noncommutative formal power series," *Trans. of the American Mathematical Society*, pp. 395-407, 1958.

[9] T. Lyons, "Differential equations driven by rough signals," *Revista Matemática Iberoamericana*, vol. 14, no. 2, pp. 215-310, 1998.

[10] M. Nakagawa, M. Onuma, "A model of on-line handwritten Japanese text recognition free from line direction and writing format constraints," *IEICE Trans. on Information and Systems*, vol. 88, no. 8, pp.1815-1822, 2005.

[11] T. Lyons, "Rough paths, Signatures and the modelling of functions on streams," *arXiv:1405.4537*, 2014.

[12] J. Diehl, "Rotation invariants of two dimensional curves based on iterated integrals," *arXiv preprint arXiv:1305.6883*, 2013.

[13] F. Yin, Q. Wang, X. Zhang, et al, "ICDAR 2013 Chinese handwriting recognition competition," In *Document Analysis and Recognition (ICDAR), IEEE International Conference on*, pp. 1464-1470, 2013.

[14] S. Huang, "A study on recognition for rotated isolated online handwritten Chinese character," Master Thesis (in Chinese and supervised by Professor Lianwen Jin), South China University of Technology, China, 2010.

[15] J. Reizenstein, B. Graham, "Signature in online handwriting recognition." http://www2.warwick.ac.uk/fac/cross_fac/complexity/people/students/dtc/students2013/reizenstein/handwriting.pdf, 2014.

[16] Y. Sun, X. Wang, X. Tang, "Deep learning face representation from predicting 10,000 classes," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1891-1898, 2014.

[17] M. Okamoto, A. Nakamura, K. Yamamoto, "On-line Handwriting Character Recognition Method with Directional Features and Directional Change Features", In *Document Analysis and Recognition (ICDAR), IEEE International Conference on*, pp.926-930, 1997.

[18] K. Ding, G. Deng, L. Jin, "An investigation of imaginary stroke technique for cursive online handwriting Chinese character recognition," In *Document Analysis and Recognition (ICDAR), IEEE International Conference on*, pp. 531-535, 2009.

[19] Z. Bai, Q. Huo, "A study on the use of 8-directional features for online handwritten Chinese character recognition," In *Document Analysis and Recognition (ICDAR), IEEE International Conference on*, pp. 262-266, 2005.

[20] G.E. Hinton, N. Srivastava, A. Krizhevsky, et al, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv: 1207.0580*, 2012.

[21] T. Lyons, M. Caruana, T. Lévy, "Differential equations driven by rough paths," *Springer Berlin Heidelberg*, pp. 81-93, 2007.