

Principal Component 2-D Long Short-Term Memory for Font Recognition on Single Chinese Characters

Dapeng Tao, Xu Lin, Lianwen Jin, *Member, IEEE*, and Xuelong Li, *Fellow, IEEE*

Abstract—Chinese character font recognition (CCFR) has received increasing attention as the intelligent applications based on optical character recognition becomes popular. However, traditional CCFR systems do not handle noisy data effectively. By analyzing in detail the basic strokes of Chinese characters, we propose that font recognition on a single Chinese character is a sequence classification problem, which can be effectively solved by recurrent neural networks. For robust CCFR, we integrate a principal component convolution layer with the 2-D long short-term memory (2DLSTM) and develop principal component 2DLSTM (PC-2DLSTM) algorithm. PC-2DLSTM considers two aspects: 1) the principal component layer convolution operation helps remove the noise and get a rational and complete font information and 2) simultaneously, 2DLSTM deals with the long-range contextual processing along scan directions that can contribute to capture the contrast between character trajectory and background. Experiments using the frequently used CCFR dataset suggest the effectiveness of PC-2DLSTM compared with other state-of-the-art font recognition methods.

Index Terms—Font recognition, long short-term memory, neurodynamic models, optical character recognition, recurrent neural networks (RNNs).

Manuscript received November 2, 2014; revised January 18, 2015; accepted March 12, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 61125106, Grant 61075021, and Grant 61472144, in part by the National Science and Technology Support Plan under Grant 2013BAH65F01 and Grant 2013BAH65F04, in part by the Guangdong Natural Science Funds under Grant 2014A030310252, in part by the Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme under Grant 2011, in part by the Guangdong Scientific and Technology Research Plan under Grant 2012A010701001 and Grant 2012B091100396, in part by the Shenzhen Technology Project under Grant JCYJ20140901003939001, in part by the Research Fund for the Doctoral Program of Higher Education of China under Grant 20120172110023, in part by the Shaanxi Key Innovation Team of Science and Technology under Grant 2012KCT-04, and in part by the Opening Project of State Key Laboratory of Digital Publishing Technology. This paper was recommended by Associate Editor L. Shao.

D. Tao is with the School of Information Science and Engineering, Yunnan University, Kunming 650091, China (e-mail: dapeng.tao@gmail.com).

X. Lin and L. Jin are with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510640, China.

X. Li is with the Center for OPTICAL Imagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China (e-mail: xuelong_li@opt.ac.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCYB.2015.2414920

I. INTRODUCTION

CHINESE character font recognition (CCFR) has been extensively studied in the field of optical character recognition [15], [34], [55]. One aim of CCFR is to acquire the font information from a given text image for high-performance document recovery, in which the typeface needs to be known. In general, font recognition can be regarded as a special form of the image classification problem, i.e., visual descriptors are extracted from the typeface in the training set and classification is performed on these features.

Similar to other visual object recognition tasks [1], [4], [7], [10], [24], [25], [28], [37], [40], [53], feature descriptors [6], [14], [39], [43], [45], [46], [51], [58] are important for font recognition, in particular wavelet features. Using a group of Gabor filters, Zhu *et al.* [57] proposed a method to recognize a text block combined of a few Chinese characters. Ding *et al.* [11] extracted wavelet feature descriptors from multiresolution Chinese character images to improve font recognition performance. Recently, a more general solution was proposed which concatenated different features as a vector. Tao *et al.* [44] combined different local binary patterns (LBPs) into multiview features to improve font recognition accuracy in text blocks. Moussa *et al.* [32] proposed a new texture descriptor based on fractal geometry to recognize Arabic fonts, while Zramdini and Ingold [59] presented an automatic Arabic font recognition system that utilized the scale invariant feature transform algorithm [30], [31]. All these methods have one characteristic in common, namely that the performance of the typeface feature descriptors mainly dictated the accuracy of font recognition.

Classifier selection is another important step in font recognition [1]–[3], [16], [23], [35], [41], [49], [52], [56]. For instance, Zhu *et al.* [57] utilized simple a Euclidean distance classifier to identify Chinese font information. Modified quadratic discriminant functions (MQDF) [26] are also widely used in Chinese character recognition. Ding *et al.* [11] combined linear discriminant analysis [13], [47] and MQDF to improve the accuracy of CCFR. In addition, Slimane *et al.* [38] used Gaussian mixture models to build an Arabic font recognition system.

Although font recognition has been extensively studied, existing methods do not adequately deal with noisy data. As noise increases, the texture pattern is disrupted, and, as a result, real intelligent applications have seldom provided a font recognition function. Here, by analyzing the

Chinese character writing process, we identify correlations between continuous points on the written stroke, and we propose that font recognition using single Chinese characters is a sequence classification problem, which is important for resolving noisy data.

Recently, a large number of neurodynamic models [5], [8], [36], [42], [48] have been presented for solving the sequence classification problem. Recurrent neural networks (RNNs) [12], [20], [35], [50], [53], which allow cyclical connections, belong to a class of very important dynamical systems and can be applied to the sequence classification problem. RNNs were originally designed for 1-D data; however, the majority of real-world problems are multidimensional. In order to apply RNNs to multidimensional problems, multidimensional RNNs (MDRNNs) have been developed, which permits the use of RNNs architecture in image analysis, video processing, and medical data analytics. Graves *et al.* [17] presented multidimensional long short-term memory (MDLSTM), which is an improvement on standard MDRNNs that deals with long-range contextual processing along several axes.

In this paper, we present principal component 2-D long short-term memory (PC-2DLSTM) to address the compromised accuracy of CCFR caused by random noise. PC-2DLSTM includes two key steps: 1) as a preprocessing layer, a principal component layer (PCL) is presented to obtain complete font information and 2) 2-D long short-term memory (2DLSTM) [18] is used to solve the sequence classification problem. The main contribution of this paper includes: 1) we propose a RNNs approach PC-2DLSTM which integrates a PCL with the 2DLSTM and 2) to demonstrate the effectiveness of method, we provide extensive experimentations on the frequently used CCFR (FUCCFR) dataset to verify that the effectiveness of the newly developed PC-2DLSTM even on noise data.

The remainder of this paper is organized as follows. The proposed PC-2DLSTM is detailed in Section II, experimental results on the FUCCFR dataset are presented in Section III, and we conclude this paper in Section IV.

II. PRINCIPAL COMPONENT 2-D LONG SHORT-TERM MEMORY

Detailed analysis of the writing of Chinese characters reveals that Chinese characters consist of basic strokes. These strokes are continuous, single shapes. The five basic strokes of seven Chinese typefaces are shown in Table I, from which we can conclude the following. First, continuous points on the stroke are closely related, and second, the shape of closer points are more closely correlated than farther points. Existing font recognition systems have improved recognition accuracy by exploiting the entire image's texture features and do not exploit correlations between continuous points on the stroke. By suitably arranging the input ordering of the points, recognition of the texture features can be converted into a sequence classification problem, i.e., each input sequence is assigned to a single class label. It is therefore necessary to introduce the 2DLSTM [18], [19] technique to appropriately model

the correlation between continuous points. 2DLSTM borrows from long short-term memory [21], which can handle long-range contexts. 2DLSTM is a special case of MDLSTM, which is built by adding memory blocks to the recurrent part of MDRNNs [17].

In addition, although the pixels forming a Chinese character can be directly used for classification, it is preferable to use a preprocessing layer for feature representation in a font recognition system. The features produced by the preprocessing layer should be complete, rational, and simple, so that the feature contains all the discriminative information needed for classification and preprocessing (and subsequent font recognition) is fast. Here, principal component analysis (PCA) [22] is used to construct the preprocessing layer and improve the traditional MDRNNs for font recognition on single Chinese characters. As noted above, the proposed PC-2DLSTM contains two main components: 1) a PCL and 2) a 2DLSTM.

A. Principal Component Layer

In the principal component layer (PCL), PCA [22], [27] is introduced to construct the preprocessing layer and to improve the accuracy of font recognition. PCA has two main advantages: 1) PCA is a global unsupervised algorithm that does not require sophisticated parameter selection and 2) by maximizing the trace of the total scatter, PCA can easily reconstruct Gaussian-distributed data.

With respect to font recognition on single Chinese characters, a training set $\{P_i\}_{i=1}^N = [P_1, P_2, \dots, P_N]$ consists of N single Chinese character images, and each character sample P_i is m -pixel spaced in both horizontal and vertical directions. An $n \times n$ square patch can be created by using an arbitrary pixel and its neighbors in each character sample and scanning the patch into a long vector. After all the overlapping $m \times m$ patches are collected, we can obtain $m \times m \times N$ vectors $X = [x_1, x_2, \dots, x_{mmN}] \in R^{nm \times mmN}$. Afterwards, using PCA, we can represent X and learn the projection matrix $U \in R^{nm \times d}$ that minimizes the reconstruction error, that is

$$\begin{aligned} \arg \min \| (X - \eta) - UU^T(X - \eta) \|^2 \\ \text{s.t. } U^T U = I_d \end{aligned} \quad (1)$$

where $\mu = (1/mmN) \sum_{i=1}^{mmN} x_i$ is the mean of all vectors. I_d is a $d \times d$ identity matrix. We further define the correlation matrix

$$\sum_X = E(X - \eta)(X - \eta)^T. \quad (2)$$

We can obtain the solution of (1) by means of the generalized eigenvector problem

$$\sum_X u_j = \lambda_j u_j. \quad (3)$$

The decreasing order of d eigenvalues are listed as follows:

$$\lambda_1 > \lambda_2 > \dots > \lambda_j \dots > \lambda_{d-1} > \lambda_d \quad (4)$$

and the corresponding d eigenvectors are

$$U = [u_1, u_2, \dots, u_j \dots, u_{d-1}, u_d]. \quad (5)$$

TABLE I
BASIC STROKE OF CHINESE CHARACTER

	Heng	Shu	Na	Pie	Zhe
Song	—		\	丿	→
Fang	—		\	㇇	→
Yahei	—		\	丿	↗
Kai	—		\	丿	→
Lishu	—		\	丿	→
Xingkai	—		\	丿	→
Youyuan	—		\	丿	↗

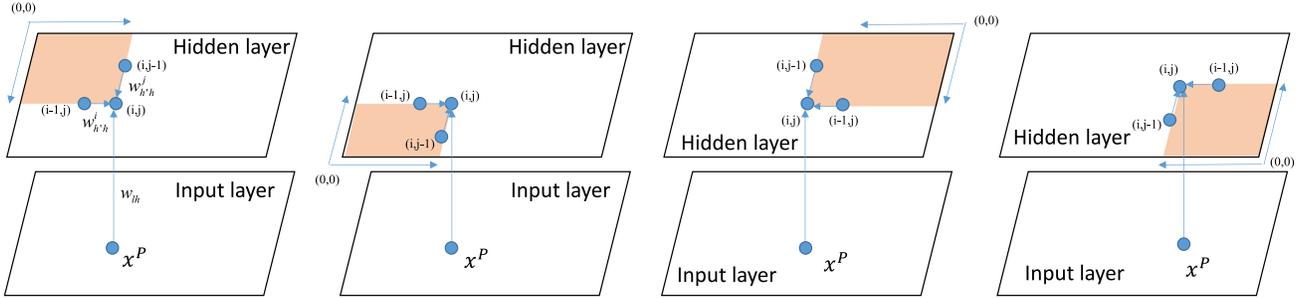


Fig. 1. Scanning directions of 2DLSTM.

The principal component network filters the $m \times m$ input Chinese character with d kernels of size $n \times n$, which are built by eigenvectors

$$k_j = f(u_j) \quad (6)$$

where $f(u_j)$ is a function that converts vector u_j to matrix k_j . Thus, the PCL convolution is defined as

$$P'_i = P_i * k_j, i = 1, 2, \dots, N \quad (7)$$

where $*$ is the 2-D discrete convolution operator.

In the PCL convolution procedure, we set the distance between the receptive field centers as one pixel. In addition, performance is improved when the size of the input field is larger than the recognized object. Therefore, we can enlarge the character sample images by padding the boundary of the input with zero prior to PCL convolution, after which the d feature maps capture the main information of all the patches.

B. 2-D Long Short-Term Memory

Graves and Schmidhuber [18] demonstrated that MDLSTM handles data with long-range interdependencies extremely well. Here, we briefly describe 2DLSTM.

Considering a 2-D font image, 2DLSTM has four scanning models, as shown in Fig. 1, and we can achieve a sequence pattern from each scanning model. In particular, the 2DLSTM forward pass starts at the original point $(0, 0)$, follows the arrow directions, and scans through the input $\{x^p\}$ in 1-D sequence. Let a_h^p and b_h^p be the input and activation of the LSTM unit at point $p = (i, j)$ in a 2-D input sequence $\{x^p\}$, respectively. We define

$$p_i^- = (i - 1, j) \quad (8)$$

and

$$p_j^- = (i, j - 1) \quad (9)$$

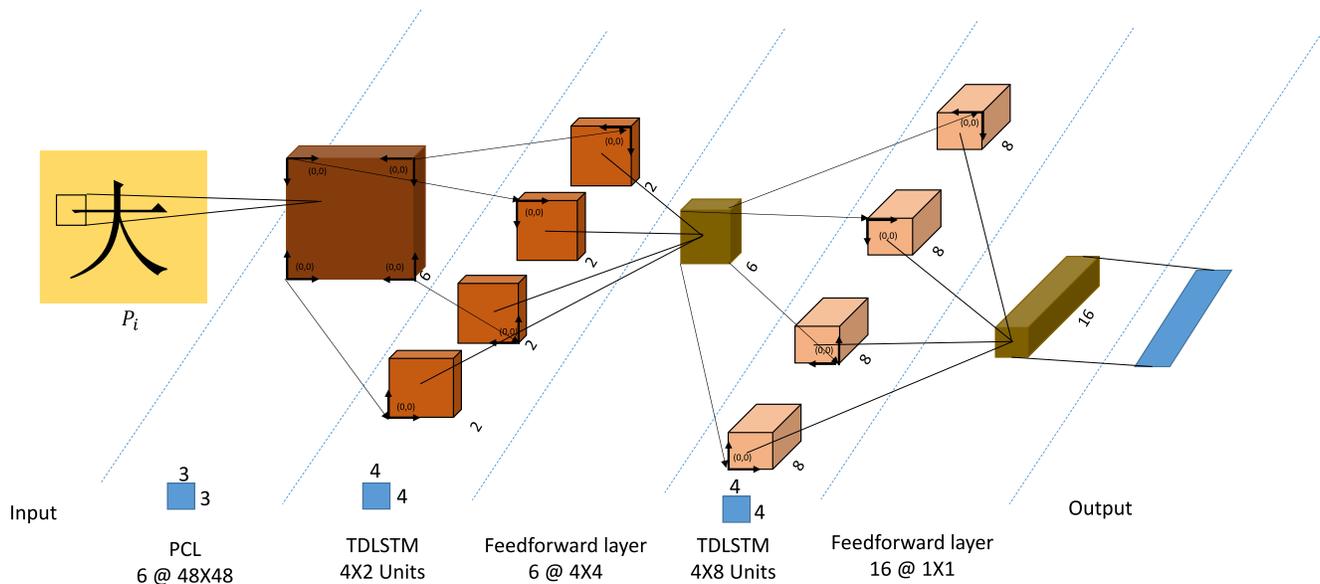


Fig. 2. Networks architecture of PC-2DLSTM.

Algorithm 1 PC-2DLSTM

- Step 1: The input layer is made up of 52×52 sensory nodes. It receive the Chinese character images.
- Step 2: The first hidden layer, i.e., PCL, conducts principal component convolution. It consists of six feature maps of size 48×48 . The size of the receptive field is 5×5 .
- Step 3: The PCL map is partitioned into pixel blocks of size 3×3 . Each block is scanned into a vector as a single input of the second hidden layer, i.e., 2DLSTM, which scans the pixel blocks in four directions;
- Step 4: The second hidden layer map is partitioned into pixel blocks of size 4×4 . The blocks are input into the third hidden layer, i.e., the feed-forward layer, which consists of six feature maps of size 4×4 .
- Step 5: The fourth hidden layer, i.e., 2DLSTM, scans the map of the third hidden layer in four directions.
- Step 6: The map of fourth hidden layer is partitioned into pixel blocks of size 4×4 . The blocks are input for the fifth hidden layer, i.e., the feed-forward layer, which consists of 16 dimensional feature.
- Step 7: The output of the last map is fed into an N -way softmax, where N is the number of font categories.

w_{lh} is the weight of the feed-forward connection from input unit l to hidden unit h . $w_{h'h}^i$ and $w_{h'h}^j$ are the weights of the recurrent connections from hidden unit h' to unit h along each scanning direction. w_{lh} , $w_{h'h}^i$, and $w_{h'h}^j$ are marked in Fig. 1.

The forward equations for 2DLSTM with L input units and H hidden summation units are as follows:

$$a_h^p = \sum_{l=1}^L x_l^p w_{lh} + \sum_{\substack{h'=1 \\ i>0}}^H b_{h'}^{p_i} w_{h'h}^i + \sum_{\substack{h'=1 \\ j>0}}^H b_{h'}^{p_j} w_{h'h}^j \quad (10)$$

and

$$b_h^p = \theta_h(a_h^p) \quad (11)$$

where θ_h is the activation function of hidden unit h .

Due to space constraints, the other parts of 2DLSTM (the input gate, forget gate, output gate, and the memory cell) are not detailed here but are easy to implement [17], [18].

C. Architecture of Networks

In this section, we describe the PC-2DLSTM architecture. In most computer vision problems, the hierarchical design choices, i.e., the outputs of one layer used as the inputs to the next layer, are suitable for feature extraction.

Due to the decrease in feature resolution, complex global feature representations can be achieved. In addition, the widely used “inverted pyramid” structure is adopted in PC-2DLSTM, in which bigger layers are arranged at the top and smaller layers at the bottom. Fig. 2 shows the network architecture of PC-2DLSTM, which comprises an input layer, a PCL, two 2DLSTM and feed-forward layers, and an output layer. Based on the above discussions, PC-2DLSTM is summarized in Algorithm 1.

III. EXPERIMENTAL RESULTS

Since there are no standardized, publicly-available single CCFR datasets, we collected Chinese character samples to create a dataset, which we name the FUCCFR dataset. Comparative experiments were conducted using different algorithms on the dataset to evaluate the proposed PC-2DLSTM font recognition system.

We measured system performance by calculating the average accuracy for each Chinese font category. In addition, we further analyzed a confusion matrix to illustrate where the method failed. The details of experimental setup are presented below.

	Song	Fang	Yahei	Kai	Lishu	Xingkai	Youyuan
Normal	安	安	安	安	安	安	安
Bold	安	安	安	安	安	安	安
Italic	安	安	安	安	安	安	安
Bold-Italic	安	安	安	安	安	安	安

Fig. 3. Typical samples of seven Chinese typefaces combined with four font styles at noise levels: SNR = 22.

	Song	Fang	Yahei	Kai	Lishu	Xingkai	Youyuan
Normal	安	安	安	安	安	安	安
Bold	安	安	安	安	安	安	安
Italic	安	安	安	安	安	安	安
Bold-Italic	安	安	安	安	安	安	安

Fig. 4. Typical samples of seven Chinese typefaces combined with four font styles at noise levels: SNR = 14.

A. Dataset

Although there are some data in this area [11], [57], to the best of our knowledge, there are no publicly-available datasets for CCFR. We therefore collected Chinese character samples to create the FUCCFR dataset as follows. First, utilizing the seven most popular Chinese typefaces (Song, Fang, Yahei, Kai, Lishu, Xingkai, and Youyuan) combined with four font styles (normal, bold, italic, and bold-italic), we generated Chinese character samples using a computer. The size of the original Chinese character was 48×48 . Each font set consists of 3755 different simplified Chinese characters in GB-2312-80 standard. Second, we converted the samples to JPG images. Third, for further comparison on noisy data, we introduce different noise levels to the sample images and the standard test method according to [11]. In particular, white Gaussian noise was added to further simulate real-world applications. The signal-to-noise ratio (SNR) was defined as

$$\text{SNR} = 10 \log \frac{\sum I_{m,n}^2}{\sum (\bar{I}_{m,n} - I_{m,n})^2} \quad (12)$$

where $I_{m,n}$ is the input of the original image and $\bar{I}_{m,n}$ is the input of the noisy image. We set SNR = 22, 17, 14, 12, and 10. Typical examples are shown in Figs. 3–5. It can be observed

that with the increase of the noise level, the contrast decreases between character trajectory and background.

In our experiments, we randomly selected $p_{tr} = 2000$ per font category for training and $p_v = 1000$ per font category for validation; the remaining samples ($p_{ts} = 755$ per font category) were used as test data. The process was repeated ten times, and the average accuracy and standard deviation are reported.

B. Baseline Methods

We compared PC-2DLSTM to Ding *et al.*'s [11] method, Zhu *et al.*'s [57] method, and MDLSTM [18], using their original parameter settings [11], [18], [57]. In addition, we also compared our method with the histograms-of-oriented-gradients (HOG) [9] and LBP [33] methods, which are widely used for image classification.

HOG describes the exact appearance and shape information of the target. The basic procedure and parameter settings of HOG were as follows: the character image was divided into 6×6 cells. In a sliding fashion, 2×2 cells were integrated into a block, and each image therefore consisted of 7×7 overlapping blocks. The nine-bin HOG descriptor was extracted from each block to obtain a 36-dimensional

	Song	Fang	Yahei	Kai	Lishu	Xingkai	Youyuan
Normal	安	安	安	安	安	安	安
Bold	安	安	安	安	安	安	安
Italic	安	安	安	安	安	安	安
Bold-Italic	安	安	安	安	安	安	安

Fig. 5. Typical samples of seven Chinese typefaces combined with four font styles at noise levels: SNR = 10.

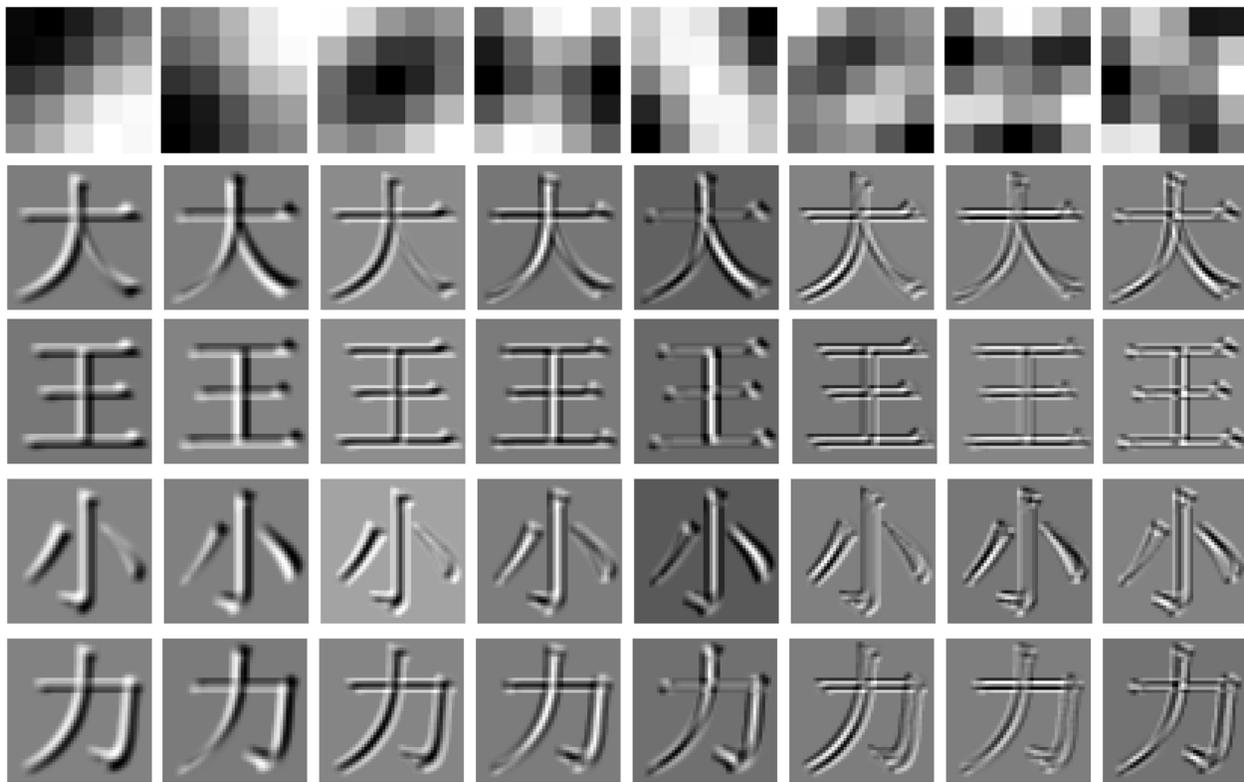


Fig. 6. PCL kernels and the corresponding convolution results of the Chinese characters 大, 王, 小, and 力.

representation. All the features were concatenated to a single vector, a 1764-dimensional representation. We further conducted PCA to obtain a 512-dimensional representation to suppress the Gaussian noise. MQDF [26], which is the most promising Chinese character recognition classifier, was used at the classification stage.

LBP can describe the exact texture distribution information of the target. The basic procedure and parameter settings of LBP were as follows: the image was partitioned into a regular 6×6 grid. From the grid, the LBP descriptor was extracted from overlapping blocks of size 12×12 . Thus, a 59-dimensional LBP representation was obtained. All the features were concatenated to a single vector,

a 2891-dimensional representation. We further conducted PCA to obtain a 512-dimensional representation to suppress the Gaussian noise. For classification, the MQDF classifier was again used.

C. PC-2DLSTM Parameters

For PC-2DLSTM, we further enlarged the character samples to 52×52 by padding the boundary of the input with zero. For the basic procedure and parameter settings, refer to Section II-C. Since the number of kernels has a real impact on PCL performance, the following experiments were designed. The number of kernels was varied from 1 to 8, and the first row of Fig. 6 shows the eight PCL kernels learned on the

TABLE II
ACCURACY OF RECOGNITION OF PC-TDLSTM ON TEST SET WITHOUT NOISE FOR VARYING NUMBER OF KERNELS

kernels	1		2		4		6		8		10	
	mean	std	mean	std								
PC-TDLSTM	96.18	0.42	97.21	0.54	97.77	0.45	97.63	0.46	97.62	0.55	97.5	0.44

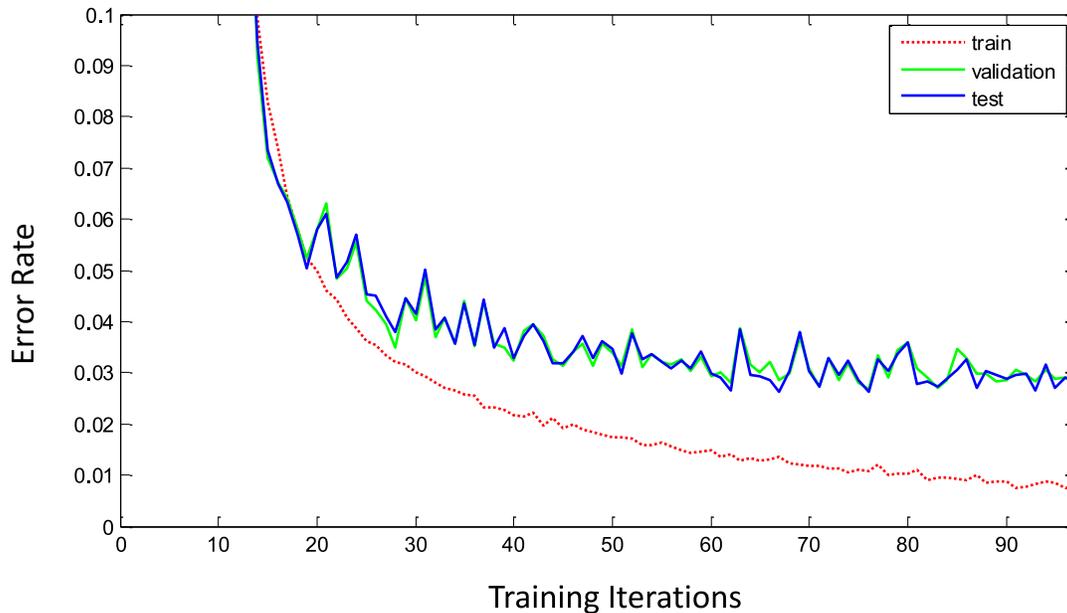


Fig. 7. PC-2DLSTM font recognition error rate during training at a noise level of SNR = 22.

TABLE III
ACCURACY OF RECOGNITION UNDER DIFFERENT NOISE LEVELS

SNR	Without Noise		22		17		14		12		10	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
PC-TDLSTM	97.77	0.45	97.3	0.57	95.55	0.67	95.0	0.64	94.2	0.66	93.41	0.71
Ding's [11]	93.15	0.52	92.8	0.61	88.0	0.49	81.1	0.57	64.7	0.84	38.3	0.95
Zhu's [57]	71.28	0.78	55.47	0.88	43.23	0.84	38.15	0.92	30.02	0.87	25.91	0.96
HOG+MQDF	94.34	0.73	80.80	0.82	70.96	1.2	58.53	0.75	47.52	0.52	34.08	0.65
LBP+MQDF	93.79	0.68	63.95	0.74	53.02	0.78	44.22	0.84	37.05	0.62	29.46	0.71
MDLSTM [18]	96.48	0.41	95.38	0.54	94.87	0.62	93.52	0.62	92.51	0.85	91.87	0.88

FUCCFR dataset; the remaining rows show the corresponding convolution results of the Chinese characters “大,” “王,” “小,” and “力.” Table II shows the accuracy of PC-LSTM recognition on the test set without noise for varying numbers of kernels. It is unsurprising that the performance ceased to improve with more than four kernels, because invalid or even harmful features are introduced into the recognition system with increased numbers of feature maps; we therefore selected the top four kernels learned on the training sets. During training, PC-2DLSTM was trained using online gradient descent [17] with a learning rate of 10^{-4} and a momentum of 0.9. The validation error was assessed after each round of training, and the termination criterion for training was judged by the new lowest value for the validation error in 20 iterations.

D. Experimental Results and Analysis

PC-2DLSTM has an important parameter, i.e., the number of kernels d . Thus, we first studied the effects of the parameter on the recognition accuracy in this section. Table II shows the

recognition accuracy of PC-2DLSTM on the test set without noise for varying number of kernels. We observe that the best font recognition rates are obtained when $d = 4$.

The comparison of PC-2DLSTM with Ding *et al.*'s [11] method, Zhu *et al.*'s [57] method, MDLSTM [18], HOG [9], and LBP [33] on the FUCCFR datasets is shown in Table III. It can be seen that PC-2DLSTM outperforms the other methods in terms of recognition accuracy. Even with a noise level of SNR = 10, our method is still effective. PC-2DLSTM is capable to handle the noisy data, because: 1) the PCL convolution operation helps to remove the noise and get a rational and complete font information and 2) simultaneously, 2DLSTM deals with the long-range contextual processing along scan directions that can contribute to capture the contrast between character trajectory and background. Fig. 7 shows PC-2DLSTM font recognition error rate during training at a noise level of SNR = 22. Note that although the training error rates decrease constantly, the validation error rate and test error rate do not drop after 70 epochs.

TABLE IV
PC-TDLSTM CLASSIFICATION CONFUSION MATRIX FOR ONE TEST SPLIT UNDER SNR = 10

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	93.5	0.0	0.8	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0
2	0.0	89.4	0.0	0.1	0.0	1.6	0.0	0.0	1.5	0.0	0.0	0.0	0.0	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.6	0.0	0.1
3	0.3	0.0	92.7	0.0	0.1	0.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0
4	0.0	0.1	0.0	90.7	0.0	0.0	0.0	2.0	0.0	0.0	1.6	0.0	0.0	0.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.4
5	0.5	0.0	0.3	0.0	96.0	0.0	1.6	0.0	0.0	0.0	0.0	0.0	1.5	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.1	2.1	0.0	0.1	0.0	94.7	0.0	0.9	1.0	0.0	0.1	0.0	0.3	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.4	0.0	0.5	0.0	96.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.1	1.9	0.0	0.7	0.0	94.3	0.0	0.0	0.4	0.0	0.0	0.0	1.7	0.0	0.0	0.1	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.3	0.0	0.0
9	0.0	2.3	0.0	0.0	0.0	0.0	0.0	0.0	83.4	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	13.6	0.0	0.4
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.8	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0
11	0.0	0.1	0.0	1.5	0.0	0.0	0.0	1.1	0.3	0.0	91.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	5.4	0.0
12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	98.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.9
13	1.6	0.0	0.0	0.0	2.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	93.5	0.0	1.5	0.0	0.5	0.0	0.0	0.0	0.4	0.0	0.1	0.0	0.0	0.0	0.0	0.0
14	0.0	1.9	0.0	0.3	0.0	3.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	90.5	0.0	0.4	0.4	0.0	0.0	0.0	2.5	0.1	0.4	0.0	0.0	0.0	0.0	0.0
15	0.0	0.0	1.5	0.0	0.0	0.0	2.5	0.1	0.0	0.0	0.0	0.0	0.5	0.0	94.3	0.0	0.0	0.0	0.5	0.0	0.1	0.0	0.4	0.0	0.0	0.0	0.0	0.0
16	0.0	0.0	0.0	2.5	0.0	0.0	0.0	4.1	0.0	0.0	0.0	0.0	0.0	0.9	0.0	89.9	0.0	0.0	0.7	0.0	0.0	0.0	1.9	0.0	0.0	0.0	0.0	0.0
17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.8	0.0	0.0	94.7	0.4	2.4	0.0	1.3	0.0	0.1	0.0	0.0	0.0	0.0	0.0
18	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	1.1	94.6	0.0	1.7	0.1	2.1	0.0	0.3	0.0	0.0	0.0	0.0
19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	1.2	1.2	0.0	95.5	0.3	0.1	0.0	1.6	0.0	0.0	0.0	0.0	0.0
20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	1.2	0.7	96.4	0.0	0.0	0.1	1.1	0.0	0.0	0.0	0.0	0.0
21	0.0	0.0	0.0	0.0	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.4	2.6	0.0	0.1	0.3	0.0	0.1	0.0	92.8	0.5	2.0	0.0	0.0	0.0	0.0	0.0
22	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.6	0.0	0.0	0.0	0.7	0.0	0.0	1.1	95.5	0.0	1.1	0.0	0.0	0.0	0.0
23	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.4	4.5	0.0	0.0	0.8	0.0	1.3	0.0	91.8	0.3	0.0	0.0	0.0	0.0
24	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	1.1	0.0	2.0	1.2	95.2	0.0	0.0	0.0	0.0	0.0
25	4.1	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	94.6	0.0	0.7	0.0
26	0.0	3.3	0.0	0.0	0.0	0.3	0.0	0.0	8.3	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	87.8	0.0	0.0
27	0.3	0.0	3.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	95.4	0.0
28	0.0	0.1	0.0	3.0	0.0	0.0	0.0	0.1	0.1	0.0	3.2	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	92.8	0.0

1: Song_Normal, 2: Song_Bold, 3: Song_Italic, 4: Song_BoldItalic, 5: Fang_Normal, 6: Fang_Bold, 7: Fang_Italic, 8: Fang_BoldItalic, 9: Yahei_Normal, 10: Yahei_Bold, 11: Yahei_Italic, 12: Yahei_BoldItalic, 13: Kai_Normal, 14: Kai_Bold, 15: Kai_Italic, 16: Kai_BoldItalic, 17: Lishu_Normal, 18: Lishu_Bold, 19: Lishu_Italic, 20: Lishu_BoldItalic, 21: Xingkai_Normal, 22: Xingkai_Bold, 23: Xingkai_Italic, 24: Xingkai_BoldItalic, 25: Youyuan_Normal, 26: Youyuan_Bold, 27: Youyuan_Italic, 28: Youyuan_BoldItalic.

Table IV depicts the classification confusion matrix for one test split at noise level SNR = 10. It can be seen that “Yahei_Normal” is confused with “Youyuan_Bold” because they are very similar.

The main observations from the experiments can be summarized as follows.

- 1) PC-2DLSTM and MDLSTM represent promising solutions to the CCFR problem because they model the correlation of continuous points on the written stroke. This advantage is even effective at a noise level of SNR = 10. In addition, PC-2DLSTM outperforms MDLSTM because the PCL convolution operation captures the important information for font recognition.
- 2) Ding *et al.*'s [11] method performs moderately well, because the method designs typeface features that take noise into account. In addition, Ding *et al.*'s [11] system utilized the promising MQDF Chinese character recognition classifier [29] to handle Gaussian-like distributions of features.
- 3) HOG and LBP work well without noise, but these features perform poorly with noise.
- 4) Zhu *et al.*'s [57] method performs poorly, because the method was designed to recognize the font using Chinese text blocks and is not suitable for font recognition using a single character.

IV. CONCLUSION

CCFR is critically important for a number of Chinese character-based intelligent applications. Considering the

complexity of many practical applications, a suitable algorithm that improves the recognition accuracy in noisy situations is required. Over recent years, many algorithms have been proposed that, to some extent, address this difficulty, but these algorithms fail to handle noisy data adequately.

By studying the Chinese character writing process, we observe that the continuous points on the written stroke are closely related and the shape of closer points is more correlated than farther points. Based on this, we present PC-2DLSTM, which converts the problem of recognition of an image's texture features to a sequence classification problem. In addition, by utilizing PCA, we present a PCL to improve the performance of the network. Seamless integration of PCL and 2DLSTM improves font recognition on single Chinese characters in our experiments.

REFERENCES

- [1] A. A. Ahmed and I. Traoré, “Biometric recognition based on free-text keystroke dynamics,” *IEEE Trans. Cybern.*, vol. 44, no. 4, pp. 458–472, Apr. 2014.
- [2] Z. Bai, G.-B. Huang, D. Wang, H. Wang, and M. B. Westover, “Sparse extreme learning machine for classification,” *IEEE Trans. Cybern.*, vol. 44, no. 10, pp. 1858–1870, Oct. 2014.
- [3] A. Bauer, N. Görmitz, F. Biegler, K.-R. Müller, and M. Kloft, “Efficient algorithms for exact inference in sequence labeling SVMs,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 5, pp. 870–881, May 2014.
- [4] D. Cai and X. He, “Manifold adaptive experimental design for text categorization,” *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 4, pp. 707–719, Apr. 2012.
- [5] A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*. Chichester, U.K.: Wiley, Nov. 1994.

- [6] J. Chen and J. Yang, "Robust subspace segmentation via low-rank representation," *IEEE Trans. Cybern.*, vol. 44, no. 8, pp. 1432–1445, Aug. 2014.
- [7] Y. Chen, J. Zhang, D. Cai, W. Liu, and X. He, "Nonnegative local coordinate factorization for image representation," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 969–979, Mar. 2013.
- [8] L. Cheng, Z.-G. Hou, M. Tan, and X. Wang, "Necessary and sufficient conditions for consensus of double-integrator multi-agent systems with measurement noises," *IEEE Trans. Autom. Control*, vol. 56, no. 8, pp. 1958–1963, Aug. 2011.
- [9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, San Diego, CA, USA, 2005, pp. 886–893.
- [10] R. Diao, F. Chao, T. Peng, N. Snooke, and Q. Shen, "Feature selection inspired classifier ensemble reduction," *IEEE Trans. Cybern.*, vol. 44, no. 8, pp. 1259–1268, Aug. 2014.
- [11] X.-Q. Ding, C. Li, and W. Tao, "Character independent font recognition on a single Chinese character," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 195–204, Feb. 2007.
- [12] J. L. Elman, "Finding structure in time," *Cogn. Sci.*, vol. 14, no. 2, pp. 179–211, 1990.
- [13] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.*, vol. 7, no. 2, pp. 179–188, 1936.
- [14] W. Fu, M. Johnston, and M. Zhang, "Low-level feature extraction for edge detection using genetic programming," *IEEE Trans. Cybern.*, vol. 44, no. 8, pp. 1459–1472, Aug. 2014.
- [15] H. Fujisawa, "Forty years of research in character and document recognition—An industrial perspective," *Pattern Recognit.*, vol. 41, no. 8, pp. 2435–2446, 2008.
- [16] E. S. García-Treviño and J. A. Barria, "Structural generative descriptions for time series classification," *IEEE Trans. Cybern.*, vol. 44, no. 10, pp. 1978–1991, Oct. 2014.
- [17] A. Graves, S. Fernandez, and J. Schmidhuber, "Multidimensional recurrent neural networks," in *Proc. Int. Conf. Artif. Neural Netw.*, Porto, Portugal, Sep. 2007, pp. 549–558.
- [18] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Whistler, BC, Canada, 2009, pp. 1–8.
- [19] A. Graves *et al.*, "A novel connectionist system for unconstrained handwriting recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 855–868, May 2009.
- [20] B. Hammer, "On the approximation capability of recurrent neural networks," *Neurocomputing*, vol. 31, nos. 1–4, pp. 107–123, 2000.
- [21] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [22] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, no. 6, pp. 417–441, 1933.
- [23] S.-C. Huang and B.-H. Do, "Radial basis function based neural network for motion detection in dynamic scenes," *IEEE Trans. Cybern.*, vol. 44, no. 1, pp. 114–125, Jan. 2014.
- [24] B. Hyun, P. Kabamba, and A. Girard, "Optimal classification by mixed-initiative nested thresholding," *IEEE Trans. Cybern.*, vol. 45, no. 1, pp. 29–39, Jan. 2015.
- [25] B. Jiang, M. Valstar, B. Martínez, and M. Pantic, "A dynamic appearance descriptor approach to facial actions temporal modeling," *IEEE Trans. Cybern.*, vol. 44, no. 2, pp. 161–174, Feb. 2014.
- [26] F. Kimura, K. Takashina, S. Tsuruoka, and Y. Miyake, "Modified quadratic discriminant functions and the application to Chinese character recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 1, pp. 149–153, Jan. 1987.
- [27] L. I. Kuncheva and W. J. Faithfull, "PCA feature extraction for change detection in multidimensional unlabeled data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 1, pp. 69–80, Jan. 2014.
- [28] L. Laporte, R. Flamary, S. Canu, S. Dejean, and J. Mothe, "Nonconvex regularizations for feature selection in ranking with sparse SVM," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 6, pp. 1118–1130, Jun. 2014.
- [29] C.-L. Liu, F. Yin, D.-H. Wang, and Q.-F. Wang, "Online and offline handwritten Chinese character recognition: Benchmarking on new databases," *Pattern Recognit.*, vol. 46, no. 1, pp. 155–162, 2013.
- [30] L. Liu, M. Yu, and L. Shao, "Multiview alignment hashing for efficient image search," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 956–966, Mar. 2015.
- [31] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [32] S. B. Moussa, A. Zahour, A. Benabdelhafid, and A. M. Alimi, "New features using fractal multi-dimensions for generalized Arabic font recognition," *Pattern Recognit. Lett.*, vol. 31, no. 5, pp. 361–371, 2010.
- [33] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [34] L. Peng *et al.*, "Multi-font printed Mongolian document recognition system," *Int. J. Doc. Anal. Recognit.*, vol. 13, no. 2, pp. 93–106, 2010.
- [35] M. Perez-Illarbe, "New discrete-time recurrent neural network proposal for quadratic optimization with general linear constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 2, pp. 322–328, Feb. 2013.
- [36] A. H. Phan, P. Tichavský, and A. Cichocki, "Fast alternating LS algorithms for high order CANDECOMP/PARAFAC tensor factorizations," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4834–4846, Oct. 2013.
- [37] P. Qi and X. Hu, "Learning nonlinear statistical regularities in natural images by modeling the outer product of image intensities," *Neural Comput.*, vol. 26, no. 4, pp. 693–711, Apr. 2014.
- [38] F. Slimane, S. Kanoun, J. Hennebert, A. M. Alimi, and R. Ingold, "A study on font-family and font-size recognition applied to Arabic word images at ultra-low resolution," *Pattern Recognit. Lett.*, vol. 34, no. 2, pp. 209–218, 2013.
- [39] L. Shao, L. Liu, and X. Li, "Feature learning for image classification via multiobjective genetic programming," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 7, pp. 1359–1371, Jul. 2014.
- [40] L. Shao, X. Zhen, D. Tao, and X. Li, "Spatio-temporal Laplacian pyramid coding for action recognition," *IEEE Trans. Cybern.*, vol. 44, no. 6, pp. 817–827, Jun. 2014.
- [41] U. Srinivas, N. M. Nasrabadi, and V. Monga, "Graph-based sensor fusion for classification of transient acoustic signals," *IEEE Trans. Cybern.*, vol. 45, no. 3, pp. 576–587, Mar. 2015.
- [42] D. Stewart, R. Seymour, A. Pass, and J. Ming, "Robust audio-visual speech recognition under noisy audio-video conditions," *IEEE Trans. Cybern.*, vol. 44, no. 2, pp. 175–184, Feb. 2014.
- [43] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Trans. Syst., Man, Cybern.*, vol. 8, no. 6, pp. 460–473, Jun. 1978.
- [44] D. Tao, L. Jin, S. Zhang, Z. Yang, and Y. Wang, "Sparse discriminative information preservation for Chinese character font categorization," *Neurocomputing*, vol. 129, pp. 159–167, Apr. 2014.
- [45] D. Tao, L. Jin, Y. Wang, and X. Li, "Person reidentification by minimum classification error-based KISS metric learning," *IEEE Trans. Cybern.*, vol. 45, no. 2, pp. 242–252, Feb. 2015.
- [46] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
- [47] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Geometric mean for subspace selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 260–274, Feb. 2009.
- [48] Y. Xia, T. Chen, and J. Shan, "A novel iterative method for computing generalized inverse," *Neural Comput.*, vol. 26, no. 2, pp. 449–465, Feb. 2014.
- [49] Y. Xia, C. Sun, and W. Zheng, "Discrete-time neural network for fast solving large linear L1 estimation problems and its application to image restoration," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 5, pp. 812–820, May 2012.
- [50] J. Xiao, Z. Zeng, and A. Wu, "New criteria for exponential stability of delayed recurrent neural networks," *Neurocomputing*, vol. 134, pp. 182–188, Jun. 2014.
- [51] C. Xu, D. Tao, and C. Xu, "Large-margin multi-view information bottleneck," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1559–1572, Aug. 2014.
- [52] Y. Yang, Q. He, and X. Hu, "A compact neural network for training support vector machines," *Neurocomputing*, vol. 86, pp. 193–198, Jun. 2012.
- [53] J. Yu, Y. Rui, Y. Tang, and D. Tao, "High-order distance-based multi-view stochastic learning in image classification," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2431–2442, Dec. 2014.
- [54] Z. Zeng, T. Huang, and W. Zheng, "Multistability of recurrent neural networks with time-varying delays and the piecewise linear activation function," *IEEE Trans. Neural Netw.*, vol. 21, no. 8, pp. 1371–1377, Aug. 2010.
- [55] H. Zhang, D.-H. Wang, and C.-L. Liu, "Character confidence based on N-best list for keyword spotting in online Chinese handwritten documents," *Pattern Recognit.*, vol. 47, no. 5, pp. 1880–1890, 2014.

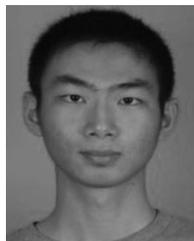
- [56] L. Zhang, X. Zhen, and L. Shao, "Learning object-to-class kernels for scene classification," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3241–3253, Aug. 2014.
- [57] Y. Zhu, T. Tan, and Y. Wang, "Font recognition based on global texture analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1192–1200, Oct. 2001.
- [58] F. Zhu and L. Shao, "Weakly-supervised cross-domain dictionary learning for visual recognition," *Int. J. Comput. Vis.*, vol. 109, nos. 1–2, pp. 42–59, Aug. 2014.
- [59] A. Zramdini and R. Ingold, "Optical font recognition using typographical features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 877–882, Aug. 1998.



Dapeng Tao received the B.E. degree from Northwestern Polytechnical University, Fremont, CA, USA, and the Ph.D. degree from the South China University of Technology, Guangzhou, China.

He is currently an Engineer with the School of Information Science and Engineering, Yunnan University, Kunming, China. His current research interests include machine learning, computer vision, and cloud computing. He has authored and co-authored over 30 scientific articles.

Mr. Tao has served on over ten international journals, including the *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, the *IEEE TRANSACTIONS ON MULTIMEDIA*, the *IEEE SIGNAL PROCESSING LETTERS*, and *PLOS-ONE*.



Xu Lin received the B.Eng. degree from the South China University of Technology, Guangzhou, China.

His current research interests include machine learning and computer vision.



Lianwen Jin (M'98) received the B.S. degree from the University of Science and Technology of China, Anhui, China, and the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 1991 and 1996, respectively.

He is currently a Professor with the College of Electronic and Information Engineering, South China University of Technology. His current research interests include handwriting analysis and recognition, image processing, machine learning, computer vision, and intelligent systems. He has

authored over 100 scientific papers.

Prof. Jin was a recipient of the Award of New Century Excellent Talent Program of Ministry of Education in 2006 and the Guangdong Pearl River Distinguished Professor Award in 2011. He served as a Program Committee Member for a number of international conferences, including International Conference on Pattern Recognition 2010–2014, International Conference on Frontiers in Handwriting Recognition 2006–2014, International Conference on Document Analysis and Recognition 2009–2015, International Conference on Image Processing 2014, China Summit and International Conference on Signal and Information Processing 2013–2015, International Conference on Multimedia and Expo 2014–2015, and International Joint Conference on Artificial Intelligence 2015.

Xuelong Li (M'02–SM'07–F'12) is a Full Professor with the Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China.