Neurocomputing 129 (2014) 159–167

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Sparse Discriminative Information Preservation for Chinese character font categorization

Dapeng Tao [1], Lianwen Jin *, Shuye Zhang, Zhao Yang, Yongfei Wang

School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510640, China

ARTICLE INFO

Article history:
Received 31 May 2013
Received in revised form
6 September 2013
Accepted 21 September 2013
Communicated by Ass. prof.L. Shao
Available online 25 October 2013

Keywords:
Chinese character font categorization
LBP
Harris corner
Dimension reduction
Sparse learning

ABSTRACT

With the rapid development of optical character recognition (OCR), font categorization becomes more and more important. This is because font information has very wide usage and researchers came to know this point recently. In this paper, we propose a new scheme for Chinese character font categorization (CCFC), which applies LBP descriptor based Chinese character interesting points for representing font information. Specifically, it classifies Chinese character font through the cooperation between a new Sparse Discriminative Information Preservation (SDIP) for feature selection and NN classifier. SDIP focus three aspects as follows: (1) it preserves the local geometric structure of the intra-class samples and maximizes the margin between the inter-class samples on the local patch simultaneously; (2) it models the reconstruction error to preserve the prior information of the data distribution; and (3) it introduces the L1-norm penalty to achieve the sparsity of the projection matrix. We conduct experiments on our new collect text block images which include 25 popular Chinese fonts. The average recognition demonstrates the robustness and effectiveness of SDIP for CCFC.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Optical character recognition (OCR) receives intensive attentions for its commercial interest [5,7,13,19,42]. Although, a large number of OCR systems have been developed in the last decades, few of them discussed the aspect of font recognition. With the significant improvement of OCR's accuracy, this situation is starting to change. In general, font categorization can be used the following practical occasions. First, OCR system's performance will benefit from font information. The divergence of different fonts will make the character recognition difficult. If the font is known, we can build a mono-font character recognition scheme to achieve the better accuracy. Second, font categorization contributes to improve the document understanding at the semantic level, because the title, abstract and main body are usually edited in some specific fonts. Third, it is necessary for high-performance document recovery. It is not only need to get the document content, but also need to know the typeface.

Many schemes have been proposed to apply to the western font categorization. Moussa et al. [25,26] proposed a new global font feature based fractal geometry to for Arabic font categorization. Cooperman [8] used some local detectors to get the font properties, such as serif, boldness, etc. Shi and Pavlidis [33] view word length and stroke slopes as font feature. Zramdini and Ingold's [59] method was based on typographical features. In Chinese font field, Zhu et al. [56] used a group of Gabor filters to extract the image's texture feature. The method requests the input image is a text block which is combined by a few characters. Ding's method [9] can work on a single Chinese character. Ding extracted the wavelet feature as font feature. Most methods of Chinese font categorization view font categorization as the whole image texture classification.

Although existing algorithms have been applied to font categorization, there is still room to improve the classification precision for Chinese character font categorization (CCFC) directly, because Chinese font has its own characters.

By analyzing a lot of samples, we can find out two follow facts. First there are frequently four class points, i.e., starting points, ending points, turning points and overlapping points, in Chinese characters. Fig. 1 illustrates the four classes of interesting points. But it is not strictly to distinguish the four class points, because a point may belong to some categories. For example, the overlapping point labeled is an ending point simultaneous in Fig. 1. Second the mainly divergence of different fonts are survival in the texture of these interesting points region. We can observe this phenomenon in Fig. 2. It illustrates the four most common Chinese fonts included Kai, Xihei, Song, and Fangsong. Hence, font information has no relationship with character order. In other

* Corresponding author. Tel.: +86 20 87113540.
E-mail addresses: dapeng.tao@gmail.com (D. Tao), eelwjin@scut.edu.cn, lianwen.jin@gmail.com (L. Jin), zsylitchee@gmail.com (S. Zhang), eezhao.yang@gmail.com (Z. Yang), yongfei.wang.ee@gmail.com (Y. Wang).
[1] Tel.: +86 13 560476206.

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.09.044

Fig. 1. Four classes of interesting points.

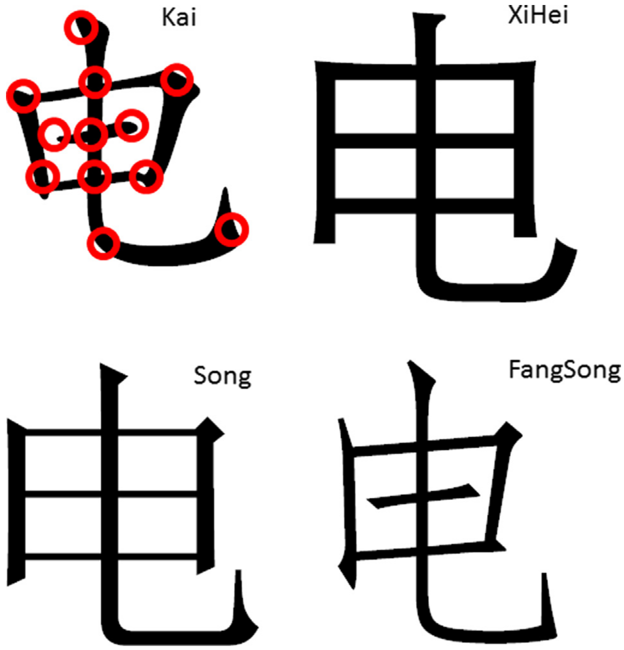


Fig. 2. Divergence of interesting points on four Chinese character fonts.

Fig. 3. The results of Harris corner detection.

word, it is content-independent. On the facts above, we can pay our attention on the texture of the region around interesting points, instead of the whole text block image. In the application of CCFC, we can robustly find these interesting points by utilizing Harris corner detector [29]. In Fig. 3, we illustrate some results of Harris corner detection on four common Chinese fonts.

After obtaining the robust interesting points, we need a feature extraction scheme for local feature [32,35,38,39], because local features can effectively obtain representations over interest regions. Scale invariant feature transform (SIFT) [23] proposed by Lowe shows good invariance ability to image deformations. However it ignores to model the spatial information. Speeded up robust features (SURF) [3] greatly improved the computational burden of interest points. It is well known that local binary patterns (LBP) [28] are proposed for texture classification originally. Based on a non-parametric method, it can effectively estimate the local geometric structure of a picture. In the previous literatures, LBP has been vastly used in facial image description [1,15,16,18]. In this paper, we introduce LBP to estimate the local geometric structure of the region around interesting points. The goal of CCFC is to achieve better accuracy, thus it is desirable to encode the different LBP features as the multi-view features. Wang et al. [44,45] proposed a new framework considering Grassmannian Regularized Structured to fuse the multi-view features. Ando et al. [2] proposed a semi-supervised scheme which utilized the two-view feature generation model to address the small sample size problem. However, to concatenate different features as a vector [50,51,53] is a more general solution.

Besides robust features for text block images, an efficient method of feature representation is necessary to obtain the most effective features for pattern recognition. Spatial pyramid matching (SPM) [22] and locality-constrained linear coding (LLC) [43,52] were used to represent the samples by using the local SIFT features and obtained the significant successfully in the image classification systems. Song et al. [34] proposed online coupled dictionary learning to solve the problem of facial of sketch-to photo synthesis. In addition, research and development of dimension reduction technology have made tremendous advances to greatly improving the quality of feature representation in the last few decades. Patch alignment framework (PAF) [54] and its semi-supervised version [49] are proposed to unify popular dimension reduction algorithms. However, we can further improve the efficiency and stability from three below aspects. First, CCFC is a typical small sample size (SSS) problem [10,37]. The most time consuming work is to label the different Chinese font, because some fonts are very similar. Considering the geometric structure of local patch, manifold learning [4,14,20,24,30,31,36,40,41,46,47,55] based dimension reduction tools are efficient and effective. Second, sparse learning schemes have been demonstrated the parsimony property and psychological interpretations in recent years. Third, in order to utilize the prior information of the data distribution, reconstruction error is directly considered in our design scheme. In this paper, we introduce the sparse learning to improve manifold learning for Chinese character font recognition and present a new dimension reduction algorithm termed Sparse Discriminative Information Preservation (SDIP).

Based on the above descriptions, stages in the CCFC are constructed: (1) using Harris corner detector to find the interesting point on the text block images; (2) applying local binary patterns to represent the local geometric structure of the region around interesting points; (3) training Sparse Discriminative Information Preservation projection matrix by using labeled samples; and (4) classifying the SDIP projected samples. The main contribution of this paper include: (1) the newly developed SDIP improves the performance for Chinese character font categorization. In addition, SDIP is an algorithm of sparse learning essentially and needs the low computational cost in the test stage. (2) We introduce the LBP descriptor to Chinese character font categorization. This strategy makes our scheme much more rapid than that in [9], which extracted wavelet features from the character images. (3) We conduct experiments to demonstrate our scheme robust and effectiveness on 25 Chinese font categories. The other parts will not be discussed in detail, because they are easy to implement by referring to the literatures cited therein. Fig. 4 shows the

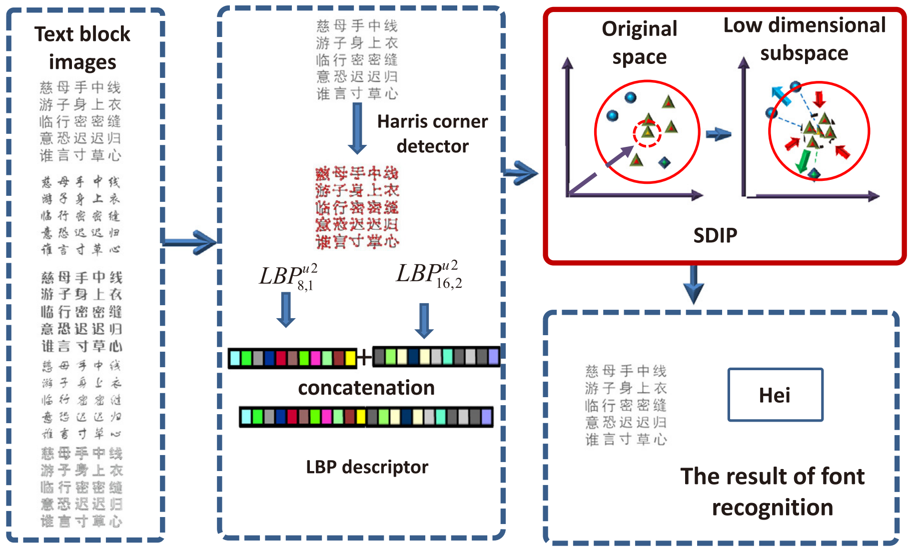


Fig. 4. The framework of Sparse Discriminative Information Preservation for Chinese character font categorization. This scheme contains the following four main components: (1) using Harris corner detector to find the interesting point on the text block images; (2) applying local binary patterns to represent the local geometric structure of the region around interesting points; (3) training Sparse Discriminative Information Preservation projection matrix by using labeled samples; and (4) classifying the SDIP projected samples.

Table 1
Important notations used in this paper and their description.

Notation	Description	Notation	Description
X	High-dimensional robust visual feature vectors	k_1	Number of closest intra-class samples
Y	Achieved succinct representations	k_2	Number of closest inter-class samples
D	Dimension of original feature vectors	β	Scaling factor
d	Reduced dimension	$L_{G(i)}$	Representation of local geometric structure preservation
N	Size of the feature vectors X	$L_{M(i)}$	Representation of discriminative information preservation
C_i	Class label	η	Trade-off parameter
X_i	Local patch	λ	Trade-off parameter
S_i	Selection matrix	U	Projection matrix

framework of Sparse Discriminative Information Preservation for Chinese character font categorization.

We organize the rest of the paper as follows: in Section 2, we detail the newly proposed Sparse Discriminative Information Preservation; Section 3 shows the experimental results and comparisons between proposed SDIP and other classical schemes and Section 4 concludes the paper.

2. Sparse Discriminative Information Preservation

Existing manifold learning schemes have improved dimension reduction a lot, they were considered on the specific knowledge and experience for their own aims. They exhibit advantages in handling cases for the problems of nonlinearity and small sample size which has been brought about high dimension data.

In this section, we propose a new sparse learning dimension reduction scheme, named Sparse Discriminative Information Preservation (SDIP), for Chinese character font categorization. It uses the sparse learning for intrinsic manifold estimation to improve the subsequent classification. This is because a useful characteristic of sparse learning that makes the learned model is more interpretive and helps reduce the computation cost. For convenience, Table 1 lists important notations used in this paper.

In Chinese character font categorization, the visual information of N samples by using a fixed length of features, i.e., $X = [x_1, x_2, \cdots, x_N] \in R^{D\times N}$ with a D-dimensional robust visual feature vector $x_i \in R^D$. Each sample has the corresponding class label $C_i \in Z^n$. The linear dimension reduction scheme aims to find the matrix $U \in R^{D\times d}$ that projects samples from the high-dimensional space R^D to a low-dimensional subspace R^d and achieve succinct representations $Y = U^T X = [y_1, y_2, \cdots y_N] \in R^{d\times N}$, where $d < D$. In general, the succinct represents Y is very helpful in improving the classification results.

Under the patch alignment framework (PAF) [54], we have a convenient way to encode an assumption of the low-dimensional sub-manifold. PAF builds a local patch $X_i = [x_i, x_{i^1}, \ldots x_{i^{k_1}}, x_{i_1}, \ldots, x_{i_{k_2}}] \in R^{D\times(k_1+k_2+1)}$ by utilizing an arbitrary sample x_i and its k_1 closest intra-class samples $x_{i^1}, \ldots, x_{i^{k_1}}$ and k_2 closest inter-class samples $x_{i_1}, \ldots, x_{i_{k_2}}$. Thus, the corresponding low-dimensional representation is $Y_i = [y_i, y_{i^1}, \ldots y_{i^{k_1}}, y_{i_1}, \ldots, y_{i_{k_2}}] \in R^{D\times(k_1+k_2+1)}$. Then the optimized objectives of patches are integrated into a whole one. Thus, in our new scheme, there are two aspects to consider for the estimated manifold: (1) the local geometric structure of the intra-class samples is preserved as much as possible; and (2) the margin between the inter-class samples is maximized.

2.1. Local geometric structure preservation

The information of intra-class local geometric structures is meaningful for classification. We have obtained inspiration from the cosine similarity which is wildly used in information retrieval [21]. Thus, the structure of intra-class local geometric can be preserved according to

$$G(y_i) = \sum_{j=1}^{k_1} \|y_i - y_{i^j}\|^2 (w_i)_j. \tag{1}$$

The weight factor $(w_i)_j$ can defined as

$$(w_i)_j = \begin{cases} \frac{x_i^T x_{i^j}}{\|x_i\| \cdot \|x_{i^j}\|} & \text{if} \quad x_{i^j} \in N_{k_1}(x_i); \\ 0 & \text{otherwise}. \end{cases} \tag{2}$$

We further deduce (1) to

$$\begin{aligned} G(y_i) &= \mathrm{tr}\left\{ \begin{bmatrix} (y_i - y_{i^1})^T \\ \vdots \\ (y_i - y_{i^j})^T \end{bmatrix} \mathrm{diag}(w_i)[y_i - y_{i^1}, \ldots, y_i - y_{i^j}] \right\} \\ &= \mathrm{tr}(Y_{G(i)} L_{G(i)} Y_{G(i)}^T), \end{aligned} \tag{3}$$

where

$$L_{G(i)} = \begin{bmatrix} -e_{k_1}^T \\ I_{k_1} \end{bmatrix} \mathrm{diag}(w_i)[-e_{k_1} \quad I_{k_1}],$$

$$e_{k_1} = \left[\overbrace{1, \cdots, 1}^{k_1} \right]^T,$$

$$I_{k_1} = \mathrm{diag}\left(\overbrace{1, \cdots, 1}^{k_1} \right), \quad \text{and}$$

$$Y_{G(i)} = [y_i, y_{i^1}, \cdots, y_{i^{k_1}}]$$

2.2. Discriminative information preservation

In most supervised manifold learning schemes, discriminative information plays a critical role. Note that the dimension reduction always brings out variations in the original distribution. Therefore, ignoring the inter-class local geometric structures intentionally will benefit the preservation of intra-class geometric structures. In our SDIP design, we consider to define the margin is the square of the distance between the center of the inter-class and the center of the intra-class.

$$\begin{aligned} M(y_i) &= \left\| \frac{1}{k_1+1}\left(y_i + \sum_{j=1}^{k_1} y_{i^j} \right) - \frac{1}{k_2} \sum_{p=1}^{k_2} y_{i_p} \right\|^2 \\ &= \mathrm{tr}(Y_{M(i)} m_i m_i^T Y_{M(i)}^T) \\ &= \mathrm{tr}(Y_{M(i)} L_{M(i)} Y_{M(i)}^T), \end{aligned} \tag{4}$$

where $Y_{M(i)} = [y_i, y_{i^1}, \cdots, y_{i^{k_1}}, y_{i_1}, \cdots, y_{i_{k_2}}]$, $L_{M(i)} = m_i m_i^T$, and

$$m_i = \left[\overbrace{\frac{1}{(k_1+1)} \cdots \frac{1}{(k_1+1)}}^{k_1+1} \overbrace{-\frac{1}{k_2} \cdots -\frac{1}{k_2}}^{k_2} \right]^T.$$

2.3. From part optimization to whole alignment

In part optimization stage, we have the optimizing objective in a local patch and can be written as

$$\arg\min_{y_i} (G(y_i) - \beta M(y_i)) \tag{5}$$

where β is a scaling factor to combine different motivations, and $0 \le \beta \le 1$.

Under the PAF, the low-dimensional representations Y_i can be integrated into a new coordinate. This stage can be realized by using a selection matrix $S_i \in R^{N \times (k_1+k_2+1)}$. The low-dimensional representations of the local patch Y_i can be achieved by

$$Y_i = Y S_i. \tag{6}$$

Thus, we define the selection matrixes as

$$(S_{G(i)})_{pq} = \begin{cases} 1, & \text{If} \quad p = \mathrm{F}_{G(i)}\{q\}; \\ 0 & \text{else}. \end{cases} \tag{7}$$

$$(S_{M(i)})_{pq} = \begin{cases} 1, & \text{If} \quad p = \mathrm{F}_{M(i)}\{q\}; \\ 0 & \text{else}. \end{cases} \tag{8}$$

where $F_{G(i)} = \{i, i^1, \cdots, i^{k_1}\}$ and $F_{M(i)} = \{i, i^1, \cdots, i^{k_1}, i_1, \cdots i_{k_2}\}$ are the index sets. According to (7) and (8), the optimizing objective in a local patch can be rewritten as

$$\begin{aligned} &\arg\min_{Y_i} (G(y_i) - \beta M(y_i)) \\ &= \arg\min_{Y_{G(i)}, Y_{M(i)}} (\mathrm{tr}(Y_{G(i)} L_{G(i)} Y_{G(i)}^T) - \beta \mathrm{tr}(Y_{M(i)} L_{M(i)} Y_{M(i)}^T)) \\ &= \arg\min_{Y} (\mathrm{tr}(Y S_{G(i)} L_{G(i)} S_{G(i)}^T Y^T) - \beta \mathrm{tr}(Y S_{M(i)} L_{M(i)} S_{M(i)}^T Y^T)) \end{aligned} \tag{9}$$

In whole alignment stage, all the part optimizations defined in (9) over all samples are summed over to integrated into the whole alignment objective function as

$$\begin{aligned} &\arg\min_{Y} \sum_{i=1}^{N} (\mathrm{tr}(Y S_{G(i)} L_{G(i)} S_{G(i)}^T Y^T) - \beta \mathrm{tr}(Y S_{M(i)} L_{M(i)} S_{M(i)}^T Y^T)) \\ &= \arg\min_{Y} \mathrm{tr}\left(Y \left(\sum_{i=1}^{N} (S_{G(i)} L_{G(i)} S_{G(i)}^T) - \beta \sum_{i=1}^{N} (S_{M(i)} L_{M(i)} S_{M(i)}^T) \right) Y^T \right) \\ &= \arg\min_{Y} \mathrm{tr}(Y L Y^T), \end{aligned} \tag{10}$$

where $L = \sum_{i=1}^{N} (S_{R(i)} L_{R(i)} S_{R(i)}^T) - \beta \sum_{i=1}^{N} (S_{M(i)} L_{M(i)} S_{M(i)}^T) \in R^{N \times N}$ is the alignment matrix.

2.4. Reconstruction error minimization

Although the discriminative information has been well modeled in (10), reconstruction error is not directly considered in SDIP. It is worth emphasizing that reconstruction error is as small as possible, because it will preserve the prior information of the data distribution.

PCA can achieve the subspace which has the projected direction of largest variance. We aim to minimize the divergence between the subspace of PCA and the objective subspace. For given a low-dimensional representation P obtained by PCA, we can restrict the objective representation Y according to

$$\arg\min_{Y} \|P - Y\|^2. \tag{11}$$

The objective incorporated with the reconstruction error minimization can be written as

$$\arg\min_{U} \mathrm{tr}(U^T X L X^T U) + \eta \|P - U^T X\|^2, \tag{12}$$

2.5. Sparsity penalty term

The objective (12) can acquire a projection matrix U in which elements are all (or most of) non-zero. In Chinese character font categorization, the feature dimension is greater than the size of labeled samples in general. Thus, by introducing L1-norm of the projection matrix, we can control the model complexity. The whole objective function can be written as

$$\arg\min_{U} \mathrm{tr}(U^T X L X^T U) + \eta \|P - U^T X\|^2 + \lambda \|U\|_1. \tag{13}$$

2.6. Solution for Sparse Discriminative Information Preservation

In order to achieve the objective solution, we transform (13) to the standard quadratic form with the L1-norm penalty. It can be easily solved by using the least angle regression (LARS). We further

Fig. 5. Some typical samples from our dataset. There are same-poem samples which belong to different Chinese font category. There is a category label below the corresponding text block image.

deduce (13) to

$$\begin{aligned}
&\arg\min_U \operatorname{tr}(U^T X L X^T U) + \eta \|P - U^T X\|^2 + \lambda \|U\|_1 \\
&\quad = \arg\min_U \operatorname{tr}(U^T X L X^T U) \\
&\qquad + \eta \operatorname{tr}((P - U^T X)(P - U^T X)^T) + \lambda \|U\|_1 \\
&\quad = \arg\min_U \operatorname{tr}(U^T X (L + \eta I) X^T U - \eta P X^T U - \eta U^T X P^T) + \lambda \|U\|_1 \\
&\quad = \arg\min_U \operatorname{tr}\left(U^T X \left(\frac{1}{\eta} L + I\right) X^T U - P X^T U - U^T X P^T\right) + \frac{\lambda}{\eta} \|U\|_1 \\
&\quad = \arg\min_U \operatorname{tr} A + \frac{\lambda}{\eta} \|U\|_1, \qquad (14)
\end{aligned}$$

$$A = U^T X \left(\frac{1}{\eta} L + I\right) X^T U - P X^T U - U^T X P^T. \qquad (15)$$

We can conduct standard eigenvalue decomposition on $(1/\eta)L + I$, because the alignment matrix L is symmetric. Thus, we get

$$\frac{1}{\eta} L + I = B \operatorname{diag}(\Lambda_i) B^T = B \Lambda B^T, \qquad (16)$$

where B is the eigenvector matrix, Λ_i is the ith eigenvalue, and $\Lambda = diag(\Lambda_i)$ is the diagonal eigenvalue matrix.

Substituting (16) back into (15), we get

$$\begin{aligned}
&U^T X \left(\frac{1}{\eta} L + I\right) X^T U - P X^T U - U^T X P^T \\
&\quad = U^T X (B \Lambda^{1/2})(\Lambda^{1/2} B^T) X^T U - P (B \Lambda^{1/2})(B \Lambda^{1/2})^{-1} X^T U \\
&\qquad - U^T X (B \Lambda^{1/2})(B \Lambda^{1/2})^{-1} P^T. \qquad (17)
\end{aligned}$$

By using (17), we further deduce (14) to

$$\begin{aligned}
&\arg\min_U \operatorname{tr} A + \frac{\lambda}{\eta} \|U\|_1 \\
&\quad = \arg\min_U \|(B \Lambda^{1/2})^{-1} P^T - (\Lambda^{1/2} B^T) X^T U\|^2 \\
&\qquad - \operatorname{tr}(P (B \Lambda B^T)^{-1} P^T) + \frac{\lambda}{\eta} \|U\|_1 \qquad (18)
\end{aligned}$$

The constant item $\operatorname{tr}(P(B\Lambda B^T)^{-1} P^T)$ can be ignored. We get a new objective function

$$\arg\min_U \|(B \Lambda^{1/2})^{-1} P^T - (\Lambda^{1/2} B^T) X^T U\|^2 + \frac{\lambda}{\eta} \|U\|_1$$

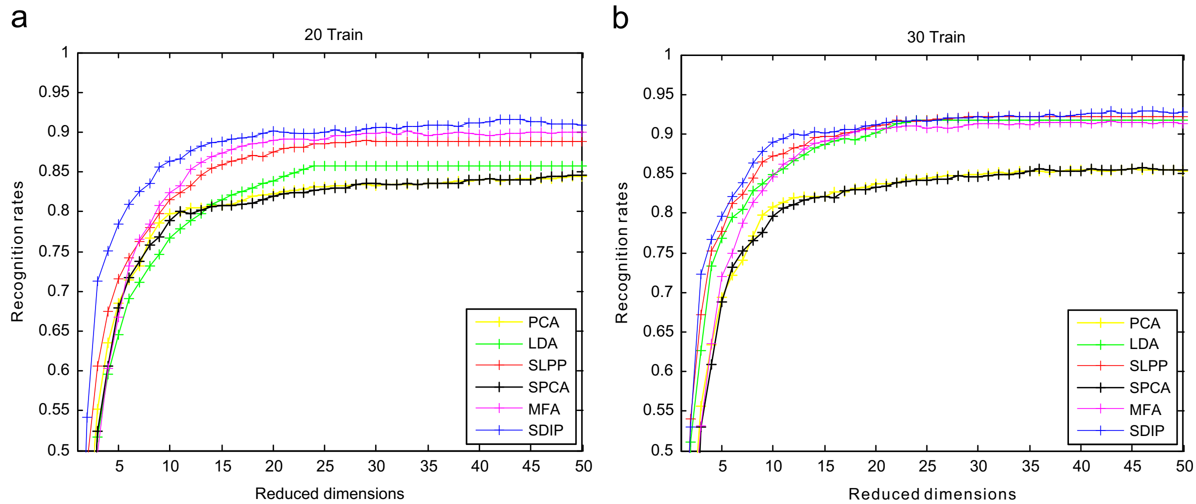


Fig. 6. We compare SDIP with PCA, SPCA, LDA, SLPP and MFA on two different training datasets: (a) 20 samples of each category for training; and (b) 30 samples of each category for training.

$$= \arg\min_{U} \|\hat{P}-\hat{X}U\|^2+\frac{\lambda}{\eta}\|U\|_1, \tag{19}$$

where $\hat{P}=[(B\Lambda^{1/2})^{-1}P^T]=[p_1,p_2,\cdots,p_d]\in R^{N\times d}$, $\hat{X}=(\Lambda^{1/2}B^T)X^T\in R^{N\times D}$, and $U=[u_1,u_2,\ldots u_d]\in R^{D\times d}$. We can rewrite (19) to

$$\arg\min_{U} \|\hat{P}-\hat{X}U\|^2+\frac{\lambda}{\eta}\|U\|_1=\sum_{i=1}^{d}\left(\arg\min_{u_i}\left(\|\hat{p}_i-\hat{X}u_i\Big|^2+\frac{\lambda}{\eta}\|u_i\|_1\right)\right). \tag{20}$$

Thus, in order to obtain the optimal solution of u_i, we can use LARS [11] to solve (20).

3. Experimental results

We collected some samples to create a dataset that includes 25 most popular Chinese character fonts. The contrastive experiments were conducted for different schemes on the dataset and the associated results were illustrated to demonstrate the effectiveness of the new proposed SDIP.

For each text block image, the LBP features are extracted from regions on interesting points. We measure the performance of our scheme by applying the average accuracy for each Chinese font category. In order to further better understanding of where the method fails, we use the confusion matrix to express the results of Chinese font categorization. The experimental setup and performance evaluation will be shown in greater details later.

3.1. Dataset

As to our knowledge there is no available public dataset for Chinese font categorization. The dataset for evaluation of Chinese font is created as follows. First, we collect 40 Tang poems and edit these materials in Microsoft Word software with different font category. Second, we use A4 papers to print them and use scanner to scan the papers and save then in JPG format. Third, we convert the JPG image into gray scale image and align the gray scale image by Hough transformation. In our creation processing, 25 Chinese font categories are considered, i.e., Hei, YaHei, XiHei, YueHei, MeiHei, YaYuan, XingKai, Kai, FangSong, Song, ZhongSong, ZongYi, HuoYi, CuQian, GuangBiao, HuangCao, HuPo, LingXin, Shu, WeiBei, XinWei, YaoTi, YouYuan, LiShu, ShuangXian. Fig. 5 shows the partial samples of 25 Chinese font categories.

In our experiment, we randomly select $p=20$ and $p=30$ samples per font category respectively to train SDIP, while the remaining samples $p_s=20$ and $p_s=10$ are used as the test data. The process was repeated 10 times, and then the average accuracy is reported. In addition, we chose their original parameter settings [6,48,57] for the methods compared in our paper.

3.2. Feature descriptor

Text block images are processed as follows to obtain the feature descriptors. First, we utilize Harris corner detector to achieve the interesting points on each text block image. Second, we count for the uniform binary patterns [28] on the interesting points for a text block, and then we get a histogram of uniform binary patterns. Third, we normalize histogram and set the normalized histogram as the font feature. Thus, each dimension indicates frequency of the uniform binary patterns. In our experiment, we use $LBP_{8,1}^{u2}$ and $LBP_{16,2}^{u2}$ ($LBP_{P,R}^{u2}$, $u2$ means utilizing uniform operator and the maximum transitions number is 2. P is the pixel number in the neighbor set, R is the radius). The dimension number of $LBP_{8,1}^{u2}$ is 59 and that of $LBP_{16,2}^{u2}$ is 243. Finally, $LBP_{8,1}^{u2}$ and $LBP_{16,2}^{u2}$ representation are concatenated as a new font feature vector.

3.3. Baselines and performance evaluation

This section evaluates the performance of SDIP by comparing it with five representative algorithms, including principal component analysis (PCA) [17], sparse PCA (SPCA) [27,57,58], linear discriminant analysis (LDA) [12], supervised LPP (SLPP) [4,6] and Marginal Fisher's Analysis (MFA) [48]. Each algorithm selected has its own relative merits. PCA is an unsupervised algorithm and familiar with us for its wide usage. SPCA is a sparse learning version of PCA. LDA, SLPP and MFA are supervised algorithms. Before conducting LDA, SLPP, DLA and SDIP, the PCA projection is applied as a first stage. We need to take note of the several points as follows: (1) if the dimensions of the original features are much larger than the dimensions of training samples in LDA [12], we conduct PCA to retain $N-C$ dimensions in order to ensure that within-scatter matrix S_w is non-singular, where N is size of training samples, C is number of classes; (2) according [48], we retain $N-C$ dimensions to ensure that $X(D^p-W^p)X^T$ is non-singular; (3) considering the acceleration of the learning process, we retain $N-1$ dimensions in SLPP and SDIP by utilizing PCA. In

Table 2
Best average recognition rates of six algorithms on different training datasets.

Training set	PCA	LDA	SLPP	SPCA	MFA	SDIP
20 Samples	0.845 (50)	0.857 (24)	0.889 (29)	0.845 (49)	0.901 (33)	0.916 (42)
30 Samples	0.855 (45)	0.918 (24)	0.922 (30)	0.858 (46)	0.916 (43)	0.930 (47)

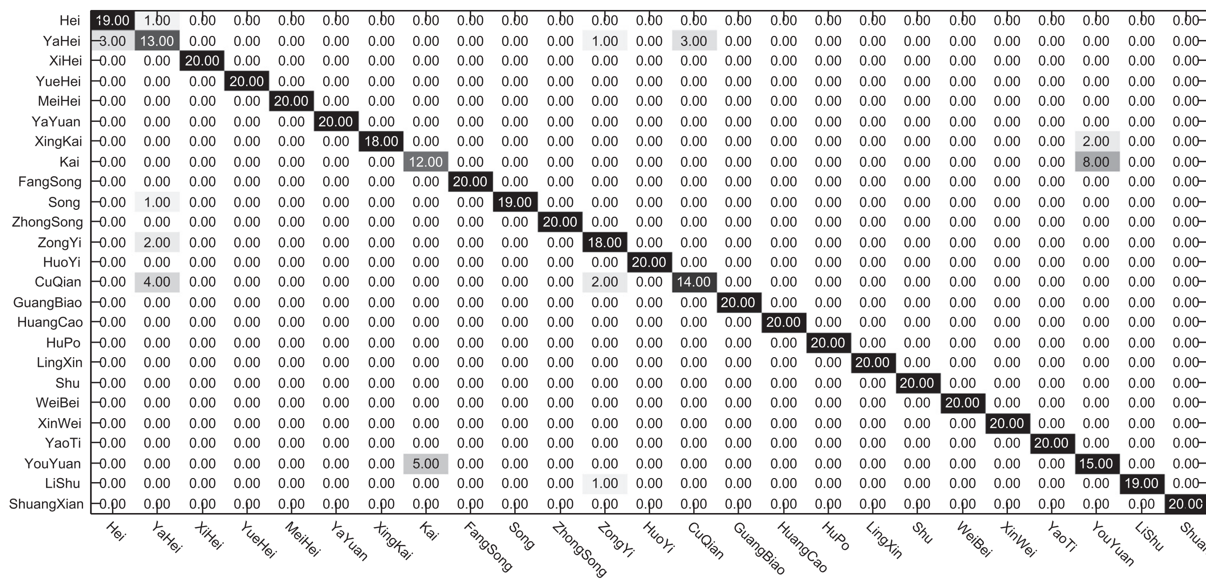


Fig. 7. SDIP classification confusion matrix. We randomly select $p=30$ samples per font category for training, while the remaining samples are used as the test data. The NN classifier is used for recognition.

addition, we use the Nearest Neighbor (NN) classifier in the classification stage. All experiments were conducted 10 times, and then we calculated the average recognition rates for real justice to compare different schemes.

3.4. Experimental results and analysis

In Fig. 6 and Table 2, we compare the proposed SDIP with PCA, SPCA, LDA, SLPP and MFA on our dataset. The average recognition rate is computed on two different training setting and varied with the number of the reduced dimensionalities. The dimensions of LDA subspace are $C-1$, i.e., 24, and the dimension of other algorithms subspace is 50. It can be observed that SDIP outperforms the others in terms of average recognition rates. In addition, the average recognition rate of SDIP has risen faster than other algorithms. It shows that the lasso penalty does select the most valuable LBP descriptors for CCFC.

Fig. 7 depicts the classification confusion matrix of SDIP for one test split. We can convenient observe the right result of classification for individual categories arranged along the diagonal. In addition, it deserves our attention that confusions occur between Kai and YouYuan.

The main observations from the experiment of CCFC can be summarized as follows: (1) SDIP, SLPP and MFA are promising in the situation of SSS, because they consider the local geometric information in own way. LDA becomes useful with the increase of training samples. In short, supervised schemes considered the class label information are superior to PCA and SPCA. (2) We compare the proposed SDIP with baselines in terms of average accuracy on the different dimension. It can be concluded that the SDIP is helpful for improving the accuracy of Chinese font categorization. This can be explained by the sparsity of SDIP projection matrix. If the number of classes is much less the number of features, the sparsity can be viewed as a regularization term to guarantee the robustness of the projection matrix.

4. Conclusions

We have presented a new dimension reduction method termed Sparse Discriminative Information Preservation (SDIP) for Chinese character font categorization. Under the PAF framework, SDIP construct a new sparse learning dimension reduction tool by considering the balance between the local geometric information of the intra-class samples and the margin between the inter-class samples. In addition, considering preserve the prior information of the data distribution will facilitate classification, the reconstruction error is also considered in our new design model. By using a series of mathematical transformations, the whole objective function of SDIP can be viewed a lasso penalized least squares problem and resolved by using the least angle regression (LARS).

In comparison to the frequent dimension reduction algorithms, the proposed SDIP has shown many attractive and competitive properties for Chinese character font categorization.

Acknowledgments

This research was partially supported by NSFC (No. 61075021, 61201348), National science and technology support plan (2013BAH65F01, 2013BAH65F04), GDNSF (No. S2011020000541), GDSTP (No. 2012A010701001) and Research Fund for the Doctoral Program of Higher Education of China (no. 20120172110023).

References

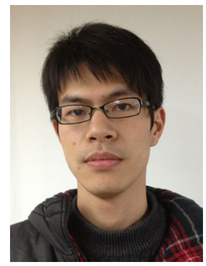
[1] A.T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.
[2] R.K. Ando, T. Zhang, Two-view feature generation model for semi-supervised learning, in: Proceedings of the International Conference on Machine Learning, 2008, pp. 376–383.
[3] H. Bay, A. Ess, T. Tuytelaars, L. Gool, SURF: Speeded Up Robust Features, Comput. Vision Image Understanding 110 (3) (2008) 346–359.
[4] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, Adv. Neural Inf. Process. Syst. 14 (2002) 585–591.
[5] S. Boquera, M. Bleda, J. Gorbe-Moya, F. Zamora-Martinez, Improving offline handwritten text recognition with hybrid HMM/ANN models, IEEE Trans. Pattern Anal. Mach. Intell. 33 (4) (2011) 767–779.
[6] D. Cai, X. He, J. Han, Using Graph Model for Face Analysis, Department of Computer Science Technical Report No. 2636, University of Illinois at Urbana-Champaign, September 2005.
[7] F. Camastra, A. Vinciarelli, Combining neural gas and learning vector quantization for cursive character recognition, Neurocomputing 51 (2003) 147–159.
[8] R. Cooperman, Producing good font attribute determination using error-prone information, electronic imaging'97, Int. Soc. Opt. Photonics 3027 (1997) 50–57.
[9] X.-Q. Ding, C. Li, W. Tao, Character independent font recognition on a single Chinese character, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2) (2007) 195–204.
[10] D.L. Donoho, High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality, American Mathematical Society Lecture, Los Angeles, CA, 2000.
[11] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Ann. Stat. 32 (2) (2004) 407–499.
[12] R.A. Fisher, The use of multiple measurements in taxonomic problems, Ann. Eugen. 7 (1936) 179–188.
[13] H. Fujisawa, Forty years of research in character and document recognition—an industrial perspective, Pattern Recognition 41 (8) (2008) 2435–2446.
[14] B. Geng, D. Tao, C. Xu, L. Yang, X. Hua, Ensemble manifold regularization, IEEE Trans. Pattern Anal. Mach. Intell. 34 (6) (2012) 1227–1233.
[15] B. Geng, D. Tao, C. Xu, DAML: Domain Adaptation Metric Learning, IEEE Trans. Image Process. 20 (10) (2011) 2980–2989.
[16] N. Guan, D. Tao, Z. Luo, J. Shawe-Taylor, MahNMF: Manhattan Non-negative Matrix Factorization, CoRR abs/1207.3438, 2012.
[17] H. Hotelling, Analysis of a complex of statistical variables into principal components, J. Educ. Psychol. 24 (1933) 417–441.
[18] D. Huang, C. Shan, M. Ardabilian, Y. Wang, L. Chen, Local binary patterns and its application to facial image analysis: a survey, IEEE Trans. Syst., Man, Cybern. C 41 (6) (2011) 765–781.
[19] Z. Jiang, X. Ding, C. Liu, Y. Wang, A novel short merged off-line handwritten Chinese character string segmentation algorithm using hidden Markov model, in: Proceedings of the International Conference on Document Analysis and Recognition, 2011, pp. 668–672.
[20] J. Karhunen, T. Hao, J. Ylipaavalniemi, Finding dependent and independent components from related data sets: a generalized canonical correlation analysis based method, Neurocomputing 113 (2013) 153–167.
[21] K. Lang, Newsweeder: learning to filter netnews, in: Proceedings of the 12th International Conference on Machine Learning, 1995, pp. 331–339.
[22] S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2006.
[23] D.G. Lowe, Distinctive image features from scale-invariant key—points, Int. J. Comput. Vision 60 (2) (2004) 91–110.
[24] B. Mokbel, W. Lueks, A. Gisbrecht, B. Hammer, Visualizing the quality of dimensionality reduction, Neurocomputing 112 (2013) 109–123.
[25] S.B. Moussa, A. Zahour, A. Benabdelhafid, A.M. Alimi, New features using fractal multi-dimensions for generalized Arabic font recognition, Pattern Recognition Lett. 31 (5) (2010) 361–371.
[26] S.B. Moussa, A. Zahour, A. Benabdelhafid, A.M. Alimi, Fractal-based system for Arabic/Latin, printed/handwritten script identification, in: Proceedings of the International Conference on Pattern Recognition, 2008, pp. 1–4.
[27] N. Naikal, A. Yang, S.S. Sastry, Informative Feature Selection for Object Recognition via Sparse PCA, Technical Report, Electrical Engineering and Computer Sciences, UCB, No. UCB/EECS-2011-27. ⟨http://www.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-27.html⟩.
[28] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.
[29] E. Rosten, R. Porter, T. Drummond, A machine learning approach to corner detection, IEEE Trans. Pattern Anal. Mach. Intell. 32 (1) (2010) 105–119.
[30] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000) 2323–2326.
[31] S. Si, D. Tao, B. Geng, Bregman divergence-based regularization for transfer subspace learning, IEEE Trans. Knowl. Data Eng. 22 (7) (2010) 929–942.
[32] L. Shao, R. Mattivi, Feature detector and descriptor evaluation in human action recognition, in: Proceedings of the ACM International Conference on Image and Video Retrieval, 2010.
[33] H. Shi, T. Pavlidis, Font recognition and contextual processing for more accurate text recognition, in: Proceedings of the Fourth International Conference on Document Analysis and Recognition, August 1997, pp. 39–44.
[34] M. Song, C. Chen, J. Bu, T. Sha, Image-based facial sketch-to-photo synthesis via online coupled dictionary learning, Inf. Sci. 193 (2012) 233–246.
[35] H. Tamura, S. Mori, T. Yamawaki, Textural features corresponding to visual perception, IEEE Trans. Syst., Man, Cybern. 8 (1979) 460–473.
[36] D. Tao, L. Jin, W. Liu, X. Li, Hessian regularized support vector machines for mobile image annotation on the cloud, IEEE Trans. Multimedia 15 (4) (2013) 833–844.
[37] D. Tao, X. Li, X. Wu, S.J. Maybank, Geometric mean for subspace selection, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 260–274.
[38] D. Tao, X. Li, X. Wu, S.J. Maybank, General tensor discriminant analysis and Gabor features for gait recognition, IEEE Trans. Pattern Anal. Mach. Intell. 29 (10) (2007) 1700–1715.
[39] D. Tao, X. Tang, X. Li, X. Wu, Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 28 (7) (2006) 1088–1099.
[40] J. Tenenbaum, V. Silva, J. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000) 2319–2323.
[41] V. Vapnik, Statistical Learning Theory, Wiley-Interscience, New York, USA, 1998.
[42] D.-H. Wang, C.-L. Liu, X.-D. Zhou, An approach for real-time recognition of online Chinese handwritten sentences, Pattern Recognition 45 (10) (2012) 3661–3675.
[43] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2010, pp. 3360–3367.
[44] X. Wang, W. Bian, D. Tao, Grassmannian regularized structured multi-view embedding for image classification, IEEE Trans. Image Process. 22 (7) (2013) 2646–2660.
[45] X. Wang, Z. Li, D. Tao, Subspaces indexing model on Grassmann manifold for image search, IEEE Trans. Image Process. 20 (9) (2011) 2627–2635.
[46] X. Wang, Z. Li, L. Zhang, J. Yuan, Grassmann hashing for approximate nearest neighbor search in high dimensional space, in: Proceedings of the International Conference on Multimedia and Expo, 2011, pp. 1–6.
[47] X. Wang, D. Tao, Z. Li, Entropy controlled Laplacian regularization for least square regression, Signal Process. 90 (6) (2010) 2043–2049.
[48] D. Xu, S. Yan, D. Tao, S. Lin, H.-J. Zhang, Marginal Fisher analysis and its variants for human gait recognition and content-based image retrieval, IEEE Trans. Image Process. 16 (2007) 2811–2821.
[49] J. Yu, D. Liu, D. Tao, H. Seah, Complex object correspondence construction in two-dimensional animation, IEEE Trans. Image Process. 20 (11) (2011) 3257–3269.
[50] J. Yu, D. Liu, D. Tao, H. Seah, On combining multiple features for cartoon character retrieval and clip synthesis, IEEE Trans. Syst., Man, Cybern. B 42 (5) (2012) 1413–1427.
[51] J. Yu, M. Wang, D. Tao, Semisupervised multiview distance metric learning for cartoon synthesis, IEEE Trans. Image Process. 21 (11) (2012) 4636–4648.
[52] J. Yu, D. Tao, M. Wang, Adaptive hypergraph learning and its application in image classification, IEEE Trans. Image Process. 21 (7) (2012) 3262–3272.
[53] J. Yu, D. Tao, Y. Rui, J. Cheng, Pairwise constraints based multiview features fusion for scene classification, Pattern Recognition 46 (2) (2013) 483–496.
[54] T. Zhang, D. Tao, X. Li, J. Yang, Patch alignment for dimensionality reduction, IEEE Trans. Knowl. Data Eng. 21 (9) (2009) 1299–1313.
[55] F. Zheng, L. Shao, Z. Song, X. Chen, Action recognition using graph embedding and the co-occurrence matrices descriptor, Int. J. Comput. Math. 88 (18) (2011) 3896–3914.
[56] Y. Zhu, T. Tan, Y. Wang, Font Recognition Based on Global Texture analysis, IEEE Trans. Pattern Anal. Mach. Intell., 23 (10) (2001) 1192–1200.
[57] H. Zou, The adaptive lasso and its oracle properties, J. Am. Stat. Assoc. 101 (476) (2006) 1418–1429.
[58] H. Zou, T. Hastie, R. Tibshirani, Sparse principal component analysis, J. Comput. Graphical Stat. 15 (2) (2006) 262–286.
[59] A. Zramdini, R. Ingold, Optical font recognition using typographical features, IEEE Trans. Pattern Anal. Mach. Intell. 20 (8) (1998) 877–882.

Dapeng Tao received the B.S. Degree in Electronics and Information Engineering from Northwestern Polytechnical University, Xi'an, China. He is currently a Ph.D. Candidate in Information and Communication Engineering at South China University of Technology, Guangzhou, China. His research interests include Machine Learning, Computer Vision and Cloud Computing.

Lianwen Jin received a B.S. Degree from the University of Science and Technology of China and a Ph.D. Degree from South China University of Technology in 1991 and 1996, respectively. He is now a Professor with the School of Electronic and Information Engineering, South China University of Technology. He is the author of more than 100 scientific papers. He is member of IEEE Signal Processing Society, IEEE Communication Society, and IEEE Computer Society, China Image and Graphics Society, the Cloud Computing Experts Committee of China Institute of Communications. He received the award of New Century Excellent Talent Program of MOE in 2006, the Guangdong Pearl River Distinguished Professor award in 2011. He served as Program Committee member for a number of international conferences, including ICMLC2007–2011, ICFHR2008–2012, ICDAR2009, ICDAR2013, ICPR2010, ICPR2012, ICMLA2012, etc. His research interests include Image Processing, Handwriting Analysis and Recognition, Machine Learning, Cloud Computing, and Intelligent Systems.

Shuye Zhang received the B.S Degree from School of Information Science and Technology, Sun Yat-sen University, Guangzhou, China. He is currently pursuing Ph. D. Degree in Information and Communication Engineering at South China University of Technology, Guangzhou, China. His research interests include Machine Learning and Handwritten Chinese Character Recognition.

Zhao Yang received the B.E. Degree in Communication Engineering from Hubei University, China, in 2008. He joined the School of Electronic and Information Engineering, South China University of Technology as a Master student in 2009 and is pursuing the Ph.D. Degree in Machine Learning and computer Vision.



Yongfei Wang received the B.S. Degree in Faculty Information Engineering from Guangdong University of Technology, Guangzhou, China. He is currently a Master's Degree candidate in Information and Communication Engineering at South China University of Technology, Guangzhou, China. His research interests include Machine Learning and Computer Vision.