Automatic recognition of serial numbers in bank notes

Bo-Yuan Feng\textsuperscript{a,\ast}, Mingwu Ren\textsuperscript{a}, Xu-Yao Zhang\textsuperscript{b}, Ching Y. Suen\textsuperscript{c}

\textsuperscript{a} Department of Computer Science, Nanjing University of Science and Technology, 200 Xiaolingwei, Nanjing 210094, PR China
\textsuperscript{b} National Laboratory of Pattern Recognition, Institute of Automation of Chinese Academy of Sciences, No. 95 Zhongguancun East Road, Beijing 100190, PR China
\textsuperscript{c} Centre for Pattern Recognition and Machine Intelligence, Concordia University, 1455 de Maisonneuve Blvd West, Montreal, Quebec, Canada H3G 1M8

A R T I C L E  I N F O

Article history:
Received 31 July 2013
Received in revised form 4 January 2014
Accepted 18 February 2014

Keywords:
Bank note serial number recognition
Cascade rejection
Synthetic training samples
Multiple classifier system

A B S T R A C T

This paper presents a new topic of automatic recognition of bank note serial numbers, which will not only facilitate the prevention of forgery crimes, but also have a positive impact on the economy. Among all the different currencies, we focus on the study of RMB (renminbi bank note, the paper currency used in China) serial numbers. For evaluation, a new database NUST-RMB2013 has been collected from scanned RMB images, which contains the serial numbers of 35 categories with 17,262 training samples and 7000 testing samples in total. We comprehensively implement and compare two classic and one newly merged feature extraction methods (namely gradient direction feature, Gabor feature, and CNN trainable feature), four different types of well-known classifiers (SVM, LDF, MQDF, and CNN), and five multiple classifier combination strategies (including a specially designed novel cascade method). To further improve the recognition accuracy, the enhancements of three different kinds of distortions have been tested. Since high reliability is more important than accuracy in financial applications, we introduce three rejection schemes of first rank measurement (FRM), first two ranks measurement (FTRM) and linear discriminant analysis based measurement (LDAM). All the classifiers and classifier combination schemes are combined with different rejection criteria. A novel cascade rejection measurement achieves 100% reliability with less rejection rate compared with the existing methods. Experimental results show that MQDF reaches the accuracy of 99.59% using the gradient direction feature trained with gray level normalized data; the cascade classifier combination achieves the best performance of 99.67%. The distortions have been proved to be very helpful because the performances of CNNs boost at least 0.5% by training with transformed samples. With the cascade rejection method, 100% reliability has been obtained by rejecting 1.01% test samples.

\copyright 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Research on character recognition including online/offline handwriting recognition and printed character recognition has received a lot of attention since 1960s. Nowadays it plays an important role in industrial applications and financial transactions. Many papers have reported promising results in the fields of bank cheque processing [1], ZIP code [2] and car licence plate recognition [3]. However, little study has focused on the automatic recognition of bank note serial numbers (e.g. [4,5]), which is beneficial in reducing financial crime and improving financial market stability and social security.

In this paper, we particularly investigate the RMB (renminbi, the paper currency used in China) serial numbers, which are printed at the lower left corner of the Chinese paper currency composed of 10 machine-printed alphanumeric characters. Fig. 1 illustrates the serial numbers in a scanned RMB image enclosed in a red rectangle. Being designed uniquely, they are used to identify RMB to provide essential information in economic activities. As China has become one of the fastest growing economies, RMB banknotes play an increasingly important role in the world economy. A high reliable RMB serial number recognition system can help to detect the fraudulent banknotes efficiently as well as manage the currency circulation. Nevertheless, the identification is a challenging problem due to the degradation of image quality caused by uneven illumination, background variation, smear, inaccurate character extraction, and so on. Specifically, the input patterns always contain some small circles and various backgrounds, which make the character close to multiple categories and seriously affect the recognition.

Despite a few previous works focused on the RMB character classification, none of them have achieved high accuracy so far. Zhao et al. [4] combined the genetic algorithm and BP (back
propagation) artificial neural network, reported an accuracy of 95%. A RMB serial number identification system using SVM proposed in [5] classified the input pattern after binarization. However, the segmentation was inaccurate because of the complex background among the scanned RMB serial number image, which led to inferior recognition results. In a related study, we extracted the RMB serial numbers directly from the RMB images by skew correction, orientation identification, and combined thresholding techniques [6].

In this study, we provide a complete solution to RMB serial number recognition including feature extraction, classification, multiple classifier combination, training data distortion and rejection. Extensive evaluations are conducted on a newly collected RMB serial number dataset, namely NUST-RMB2013, containing 35 categories with 17,262 training samples and 7000 testing samples extracted from scanned RMB images. Three of the most effective feature extraction methods such as gradient direction feature, Gabor feature and CNN trainable feature are implemented and compared using different classifiers. After that, we analyze the performances of four state-of-the-art recognition methods including SVM, LDF, MQDF and CNN on our dataset, and discuss the properties of each learning method.

To further improve the recognition accuracy, the multiple classifier system is introduced. Several cascade classifier ensemble methods have been proposed in the previous literatures [7,8]. Ref. [7] emphasizes the complementary property of the features in classification combination by training the classifiers with features of common difference principal components and difference principal components. Zhang et al. [8] proposed a complicated cascade classifier system which consists of three layers of classification, and each layer is composed of four levels of Multi-artificial neural networks (ANNs) and gating networks (GNs).

In this paper, besides the traditional combination schemes, e.g. sum-rule and max-rule, a novel cascade combination strategy is designed utilizing six classifiers trained with both original and gray level normalized samples. Although the general concepts of the cascade classifier systems are similar, our cascade combination and rejection technique are very different from [7,8]. Ref. [7] combined classifiers by adding the weighted Euclidean distance between different orders of difference principal components to the discriminant function. Nonetheless, we rejected the samples with relative low confidence and sent them to the higher level of classifiers for further analysis. Compared to our method, which trains each of the classifiers separately with the same original and gray level normalized data, [8] requires a large quantity of training samples. Since in each cascade layer, the ensemble classifiers from the second to the fourth levels are trained by the rejected characters of the previous level. In [8], the confidence values of three layers are aggregated using GNs optimized by genetic algorithms, and a reliability of 99.96% is achieved with a 99.19% recognition rate on the MNIST [9] database. However, they did not investigate the reject capability of their cascade system while we comprehensively analyzed the rejection results of different classifier combination models in the paper.

Since expanding the dataset through distortions has been proved very helpful in [10], we synthesize and expand the training samples by 9 or 10 times through translations and transformations of scaling and rotations, and obtain superior performances.

In real-life applications, it is unrealistic to create a new system with 100% recognition accuracy. However, even a small misrecognition in the banking system could cause a huge financial loss. As a result, a reliable system achieves 100% reliability by rejecting the input samples with low confidence values. Compared to the existing rejection methods such as first rank measurement (FRM) [11], first two ranks measurement (FTRM) [12] and linear discriminant analysis based measurement (LDAM) [13], we combine different classifier combination strategies with FTRM, and propose a new rejection scheme, which can reach high reliability with a much lower rejection rate.

The rest of this paper is organized as follows: Section 2 presents a new RMB serial number dataset. Section 3 describes the underlying feature extraction methods, the evaluated classifiers, multiple classifier combination strategies, distortions and rejection schemes. Experimental results are listed and discussed in Section 4. Finally, Section 5 draws concluding remarks.

2. Database

For evaluations of different techniques on RMB serial number recognition, we collected the serial numbers from daily-used RMB images, and release a database called NUST-RMB2013. Fig. 2 shows four scanned RMB images used in our database, they suffered from certain abrasion, crumple and contamination, which represents the image quality of RMB in the real-world.

To the best of our knowledge, this is the very first publicly available database related to bank note serial numbers, and can be utilized for design and evaluation of character recognition algorithms and classifiers. Provided by gray-scaled character images with gray level from 0 to 255, it consists of a training set of 500 samples for each of the 35 categories (numeral 0–9 and alphabet A–Z except V) and a testing set of 200 samples for each class. The database represents all the different classes of RMB characters. The training set and the test set are disjointed and independent from each other, therefore they can be used for fair comparison of different algorithms. Some character samples included in the NUST-RMB2013 database are presented in Fig. 3.

We collected the RMB database in three stages using two versions (1999 and 2005) of RMB with the denomination of 100 yuans. First, the contact image sensor (CIS) installed in the money counting machine scanned the RMB currencies with a resolution of 200 × 180 dpi, and produced RMB images. We then extracted the serial numbers from these RMB images. To make the database more challenging, two versions of RMB serial numbers were mixed together. A see-through image Fig. 4a and a reflection image Fig. 4b can be obtained from CIS at the same time. The see-through image includes the complex security texture from both.

![Fig. 1. Serial numbers in scanned RMB image.](image-url)
sides of the currency while the reflection image provides either the image side containing clean serial numbers without circles or the other side excluding serial numbers. In order to recognize the RMB serial number by scanning the paper currency only once without considering the input side, just the see-through image can fulfill our requirement. Therefore, the serial numbers were extracted directly from the see-through image.

However, the character extraction is a challenging problem due to uneven illumination, contrast variation, smear, various pattern background (including complex texture and little anti-counterfeiting circles), and so on. In this paper, we first detected the orientation of the scanned RMB image, and then located the RMB serial number region by the a priori knowledge of serial numbers’ size and location [6]. Finally, we used a semi-automatic method to extract and label the character samples. The character boundaries were located directly by variance contrast and processed with bi-moment normalization (BMN) [14] to the size of 36 × 60. Because the extraction process produced errors such as false alarms and missing characters, we manually chose the complete and human readable extraction results, labeled their categories and assigned them to the training and test datasets.

Fig. 5 shows more samples from the test data with abrasion and small circles touching the distinctive part of the character strokes, which makes them hard to be recognized and also increases the challenge of our database. We provide our database to the academic community freely,1 and the licensing information can be found at http://www.patternrecognition.cn/.

3. Recognition methods

The RMB serial number recognition system consists of four major components: feature extraction, classification, multiple classifier combination, and rejection. To improve the recognition performance and avoid binarizing the input patterns, we propose a recognition process directly on gray-scale images. In the following,

1 The database will be released once this paper is accepted.
we will outline three different feature extraction methods, four types of classifiers, five classifier combination methods, three distortions and rejection strategies.

3.1. Feature extraction

3.1.1. Gray-scale normalization

To overcome the gray-level variances and noises in the background caused by complex texture, uneven illumination and smear (Fig. 3), we propose a gray-scale normalization method to alleviate the intensity variation by piecewise linear stretch, which maximizes the contrast between character strokes and circles, while minimizing the gray-levels variances in the background.

As the gray-level histogram distribution of the character strokes is nearly invariable in each training pattern, therefore, the proportion of character strokes can be heuristically estimated as \( \alpha \% \), the circles and other noises hold \( \beta \% \), and all the remaining area is considered as background. To normalize the input image on gray scale, first, the gray level histogram is computed. Then, we stretch the gray level of the first \( \alpha \% \) pixels to \([0, 25]\) and enlarge the gray distribution for the pixels among \( \frac{1}{2} \alpha ; (\alpha + \beta)C_138 \% \) to \([25, 255]\). The pixels located in the last \([100 - (\alpha + \beta)] \% \) are compressed to \([235, 255]\). Given an input image \( g \) with gray level range \([\mu, \upsilon]/C_138 \), the piecewise linear normalization stretches it to the range of \([\mu_0, \upsilon_0]/C_0 \) by

\[
g_0 = \mu_0 + \frac{\upsilon_0 - \mu_0}{\upsilon - \mu} \times (g - \mu).
\]

In our database, we empirically set \( \alpha \) as 10\%, \( \beta \) as 20\%. Fig. 6 illustrates the stretch curve of the gray level normalization. Fig. 7 shows the normalization results of the samples presented in Fig. 3 in which the image quality of normalized samples greatly exceeds that of the original ones.

3.1.2. Gradient direction feature

Although numerous features have been proposed for character recognition, the gradient direction feature is widely used because of its efficiency and robustness, and it has been justified as one of the best features by previous experiments [15]. The feature extraction operating on the gray scale image consists of three main steps. First, \( 3 \times 3 \) Sobel operators get the horizontal and vertical gradients of each pixel. The range of gradient direction is then decomposed into eight chain code directions. If a gradient direction lies between two standard directions, we decompose it into two nearest standard directions by the parallelogram rule [16] as shown in Fig. 8. As a result, all pixels in the image are assigned to eight direction feature planes. To make the extracted feature less sensitive to noises, a \( 7 \times 7 \) Gaussian filter blurs all of the eight feature planes. After that, feature vectors centered at the sampling points are extracted by convolving the gradient feature planes with Gaussian masks.

In our study, after applying the direction decomposition, the eight corresponding gradient direction planes have the same size as the input pattern \((36 \times 60)\). Each gradient direction plane is divided into \(6 \times 10\) blocks. We uniformly extract the gradient magnitude of each block by the Gaussian blurring, and obtain \(480 (6 \times 10 \times 8)\) feature vectors for each sample. Before classification, the Linear Discriminant Analysis (LDA) [17] reduces the feature vector to a 34-dimensional (class number minus one) subspace.

Please cite this article as: B.-Y. Feng, et al., Automatic recognition of serial numbers in bank notes, Pattern Recognition (2014), http://dx.doi.org/10.1016/j.patcog.2014.02.011
3.1.3. Gabor feature

Gabor filter feature is considered as a candidate competitor to gradient direction feature [18]; it has been applied to character recognition since 1994 [19]. Liu et al. [20] proved that the Gabor feature performs comparably or better than the gradient feature on three well-known databases. As shown in Eq. (2), two-dimensional Gabor filter is a sinusoidal wave modulated by Gaussian function with the response in the spatial domain:

\[
g(x, y; \lambda, \theta, \sigma, \gamma) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \exp\left(\frac{2\pi x'\lambda}{\lambda}\right),
\]

where

\[
x' = x \cos \theta + y \sin \theta,
\]

\[
y' = -x \sin \theta + y \cos \theta,
\]

\[
b = \sigma / \lambda,
\]

\[
\sigma \text{ is the standard deviation of Gaussian envelope, } b \text{ represents the bandwidth, and } \gamma \text{ specifies the ellipticity of the support of the Gabor function. For } \gamma = 1 \text{, the support is circular. For } \gamma < 1, \text{ the support is elongated in orientation of the parallel stripes of the function, } \theta \text{ and } \lambda \text{ respectively denote the oscillation orientation and wavelength of a Gabor filter. Wang et al. [21] testified that the Gabor filtering achieves the best performance when the wavelength is twice as long as the width of the character stroke.}

To extract Gabor feature, we apply eight-orientation Gabor filters with two scales, and use the magnitude response to obtain the response planes. The orientation angle \(\theta_k\) is given by \(\theta_k = \pi(k-1)/8, \quad k = 1, 2, ..., 8\), and the wavelengths \(\{6, 12\}\) are selected based on the width (near six pixels) of the most strokes among serial character numbers. For feature extraction, we first partition the filtering response planes into \(5 \times 10\) blocks and \(6 \times 11\) blocks, then get the feature values of each block by dividing the number of pixels of salient signal (pixels with the response values above the mean value of the corresponding response plane) in the block by the total number of pixels of salient signal in the whole response plane [22]. In total, 928 \((5 \times 10 \times 8 + 6 \times 11 \times 8)\) feature measures can be obtained. Finally, the LDA reduces the dimension of the Gabor feature to 34, which is the same as the gradient feature.

3.1.4. CNN trainable feature

Traditional hand-crafted feature extraction methods are complicated, and they also require good knowledge of the data. However, some cases require an automatic feature extractor, which can retrieve discriminating features directly from the raw images, while the traditional features provide vague characteristics among different classes.

Lauer et al. [23] introduced a trainable feature extractor (TFE) based on the LeNet5 convolutional network [24] architecture, which can automatically retrieve features from raw images. They consider the feature extractor as a black box model, and the relevant features can be extracted through the training of convolutional neural network (CNN) [10] with no prior knowledge about the data.

For our problem, we replace the last two layers of LeNet5 by a single linear output layer with 35 units corresponding to 35 categories of the RMB databases, and then train the weights of each CNN layer by a back-propagation algorithm via convolutional filtering and down sampling. As a result, 100 outputs of the last hidden layer are linearly separable and can be fed as features for any classifier.

3.2. Classifiers

3.2.1. Kernel classifier

In the last few years, Support Vector Machine (SVM) [25,26] based on a kernel learning technique has received increasing attention and shown superior performance in the field of character recognition. An SVM is known as a binary classifier with discriminant function in a high kernel feature space.

The SVM model chooses the nonlinear \(\Phi\) functions, which map the input pattern to a higher-dimensional space, and then finds the optimal hyperplane that separates the data from two categories with the largest margin. The samples located within the region of the maximum margin are called support vectors. The SVM classifiers are always applied to the datasets of small category due to the high training cost [27].

Given a training dataset \((x_i, y_i), \quad i = 1, 2, ..., \epsilon\), where the input pattern \(x_i \in \mathbb{R}^n\) belongs to label \(y_i \in \{-1, 1\}\), the SVM uses quadratic programming (QP) to solve the following optimization problem:

\[
\text{minimize} \quad \frac{1}{2}\|\omega\|^2 + C \sum_{i=1}^{\epsilon} \xi_i,
\]

subject to \( y_i \cdot (\omega^T \Phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, ..., \epsilon, \)

where \(C\) is the cost parameter, \(\alpha\) specifies the weight vector, and \(\xi_i\) denote the constant and slack variable respectively. \(\Phi\) is the mapping function, which is not required to be explicitly defined. Only the kernel function \(k(x_1, x_2) = \Phi^T(x_1)\Phi(x_2)\) (the inner product function in the higher dimensional space of \(\Phi\)) needs to be defined and used for final decision-making. More details on kernel SVM can be found in [26].

For multi-class classification, the final recognition is determined by one-against-all [28] combination strategy. The SVM can use a wide option of kernels. In this paper, we selected both linear and radial basis function (RBF) kernels, which belong to the most popular kernel types, and conducted the experiments via the libSVM software [29] with the parameters \((C, \gamma)\) tuned by cross validation on the training dataset. The output of SVM classifier is a 35-dimensional probability vector (i.e. the probability output of the libSVM).

3.2.2. Statistical classifier

By assigning the input sample to the category with maximum posteriori probability, statistical methods based on the Bayes decision rule are widely used in pattern recognition problems [30]. Two types of statistical classifiers are named as non-parametric and parametric [17]. The non-parametric methods do not assume any functional form for the conditional distributions, including the Parzen window method and the k-NN rule. On the other hand, the parametric classifiers such as the linear discriminant function (LDF) and the quadratic discriminant function (QDF) hypothesize the multivariate Gaussian density of each class [31].

The negative of QDF can be viewed as a distance measure, and test sample is assigned to the class of minimum distance. Given an input pattern \(x\) and a priori probabilities \(P(o_i)\), the QDF distance to class \(i\) is

\[
g_{QDF} = -\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) - \frac{1}{2} \log |\Sigma_i| + \log P(o_i),
\]

where \(\mu_i\) and \(\Sigma_i\) denote the mean vector and covariance matrix of class \(o_i\) respectively. The LDF is obtained by assuming all the classes that share a common covariance matrix of Gaussian density \(\Sigma = \Sigma_1 = \Sigma_2 = \cdots = \Sigma\) and omitting the common terms in (4):

\[
g_{LDF} = \mu_i^T \Sigma^{-1} x - \frac{1}{2} \mu_i^T \Sigma_i^{-1} \mu_i + \log P(o_i).
\]

The modified QDF (MQDF) proposed by Kimura et al. [32] reduces both computational complexity and memory by replacing the minor eigenvalues \((i > k)\) with a constant \(\delta\), which stabilizes...
the generalization performance and improves the classification accuracy. The distance from test sample \( x \) to the class \( \omega_i \) is represented by

\[
S_{MQDF} = \frac{k}{2} \sum_{j=1}^{k} \log \lambda_{ij} + (d-k) \log \delta_i + \frac{1}{2} \sum_{j=k+1}^{d} [\mathbf{x} - \mu_i]^T \Sigma_i^{-1} (\mathbf{x} - \mu_i) + \frac{1}{2} \log |\Sigma_i|.
\]

where the mean vector \( \mu_i \) and covariance matrix \( \Sigma_i \) are estimated by maximum likelihood (ML), \( \lambda_{ij} \) and \( \phi_{ij} \), \( j = 1, \ldots, d \), denote the eigenvalues (sorted in decreasing order) and eigenvectors of \( \Sigma_i \), which are computed by Karhunen–Loeve (K–L) transform. We select the number of retained principal eigenvectors \( k \) as the average of minor eigenvalues [33]. The distance-based outputs of MQDF classifier are transformed to class probabilities in the form of sigmoid function [34,35] with the confidence parameters optimized by minimizing the cross entropy (CE) loss function on a validation dataset [36].

\[ y_i = \frac{1}{\sum_{j=1}^{M} p_j} p_i \]

where \( p_i, \ i = 1, 2, \ldots, N, \ j = 1, 2, \ldots, M \), denotes the confidence output of classifier \( i \) on class \( j \).

The test sample \( x \) is classified according to

\[ x = \underset{j=1}{\overset{M}{\arg \max}} y_j \]

Compared with the individual classifiers, the elementary rules can achieve outstanding recognition results with low computing complexity. Suen and Lam [38] found that the sum-rule is the most robust scheme against other combination rules.

3.3.2. Linear combination schemes

To further improve the combination performance, the multi-classifier combination can be considered as a linear problem in which the confidence measures \( p = [p_1, p_2, \ldots, p_K] \), \( K = N \times M \), obtained from \( N \) classifiers are combined by

\[ y_j = \sum_{i=1}^{K} a_i p_i, \quad j = 1 \ldots M \]

where \( w_j = [a_{i1}, a_{i2}, \ldots, a_{ik}] \) represents the weights for category \( j \). The input pattern will be assigned to the category with the maximum combined measurement.

Liu [35] used \( M \) binary SVMs to combine the confidence values of all classifiers. In this paper, we use both Linear SVM and LDF to find the optimal weights for all classifiers of each category. The SVM and LDF models are trained on a validation dataset with new synthetic features (\( N \times M \) dimensions) composed of the confidence values of each classifier. The classification can be divided into two steps. First, the confidence measurements of the test sample

**Fig. 9.** Simplified CNN architecture.
produced by individual classifiers are combined to generate a new confidence feature. The trained linear combination model then classifies the input data based on the integrated feature. By training this meta-classifier, we automatically get the optimized weights for the linear combination.

3.3.3. Cascade combination scheme

Noticing the complementarity among the classifiers described in the above section, we propose a specific cascade classifier combination scheme with rejection options to achieve higher accuracy of recognition in which the subsequent classifier processes the samples rejected by previous classifiers. Fig. 10 shows the seven stages of the recognition framework. The first six layers are connected in series, and each of them corresponds to a specific classifier. The last layer consists of a sum-rule committee.

Six classifiers (we name the classifiers SVM, MQDF, CNN that trained with gray level normalized data as GN-SVM, GN-MQDF, GN-CNN respectively) are trained separately with RMB dataset. In the test procedure, each classifier produces confidence measurements of its input sample. The current classifier will reject the sample if with a top-2 measures gap smaller than the threshold T, and pass it to the next classifier or the committee (if \( p_1(x) - p_2(x) < T \), then we reject x). Otherwise, the input pattern will be sorted to the category with the maximum confidence value. Taking the samples rejected by the sixth layer, the last layer (committee) uses the sum-rule to combine all the above classifiers on measurement level.

The threshold \( T \) should be large enough to reject all ambiguous samples and ensure the accuracy of recognition. Wu et al. [40] estimated \( T \) on the training dataset by

\[
T = \alpha \cdot \max_{i=1}^{N} |p_1(x_i) - p_2(x_i)|
\]

where \( p_1(x_i) \) and \( p_2(x_i) \) denote the confidence values of the first two candidates for training sample \( x_i \), the size of training dataset is \( N \), and \( \alpha \in [0, 1] \).

3.4. Synthetic training samples

Simard et al. [10] proved that a higher recognition rate can be achieved by generating additional synthetic training data, especially when the number of training samples is deficient. The best results on MNIST were obtained by elastic deformations [10]. Different types of distortions can greatly enlarge the size of the training set, and improve the generalization of the classifiers.

In RMB serial number recognition, we assume that the input samples contain variations such as rotations, scaling, and translations. As the serial numbers belong to printed characters, the elastic distortions imitating the transformations originated from human writing styles are unsuitable for our database.

Given a scale parameter \( \alpha \), the scaling result of position \((x, y)\) is \((\alpha x, \alpha y)\). As \( \alpha x \) and \( \alpha y \) could be non-integers, bilinear interpolation is needed. For rotations, each pixel \((x, y)\) moves to a new target location \((x + \Delta x, y + \Delta y)\), where \((\Delta x, \Delta y)\) is computed with respect to the rotation angle \( \theta \). In this paper, the distortions of scaling and rotations are applied simultaneously to the input pattern with \( \alpha \) and \( \theta \) randomly generated in a predefined range (in this paper, they are stochastically chosen in the ranges of \([0, 15]\) and \([0, 15]\) respectively). The size of dataset could be enlarged 10 times by synthesizing nine new data for each training sample.

We conduct two types of translations by shifting the input pattern one or two pixels towards eight directions to expand the training dataset by nine times. Eq. (12) shows the eight directions with \( d = \{1, 2 \} \). Here, we also considered the \( 3 \times 3 \) and \( 4 \times 4 \) translations during the implementations, however, they show inferior recognition results compared to the translation of two pixels. Since the \( 1 \times 1 \) and \( 2 \times 2 \) translations provide closer simulation of the variations between the data samples, they are adopted in the subsequent experiments:

\[
\begin{bmatrix}
  d \\
  0 \\
  0 \\
  d \\
\end{bmatrix},
\begin{bmatrix}
  0 \\
  d \\
  -d \\
  d \\
\end{bmatrix},
\begin{bmatrix}
  0 \\
  -d \\
  d \\
  -d \\
\end{bmatrix},
\begin{bmatrix}
  0 \\
  0 \\
  d \\
  d \\
\end{bmatrix},
\begin{bmatrix}
  0 \\
  d \\
  -d \\
  -d \\
\end{bmatrix},
\begin{bmatrix}
  0 \\
  -d \\
  d \\
  -d \\
\end{bmatrix},
\begin{bmatrix}
  -d \\
  d \\
  -d \\
  d \\
\end{bmatrix},
\begin{bmatrix}
  d \\
  0 \\
  0 \\
  -d \\
\end{bmatrix},
\begin{bmatrix}
  0 \\
  -d \\
  d \\
  -d \\
\end{bmatrix}
\]

Fig. 10. Framework of cascade combination.

Please cite this article as: B.-Y. Feng, et al., Automatic recognition of serial numbers in bank notes, Pattern Recognition (2014), http://dx.doi.org/10.1016/j.patcog.2014.02.011
The implementations of distortion techniques are distinguished between classifiers. We apply the distortions according to the peculiar architecture of each classifier. For instance, the kernel and statistical classifiers need expanding the dataset before training while the neural classifier alternately uses distorted data and original data in the forward propagation.

3.5. Cascade rejection

A successful application of RMB serial number recognition requires high recognition accuracy and reliability at the same time, because even a tiny error could cause a great amount of economic loss in processing of banking documents. We would rather reject the patterns with low confidence values than receive misrecognitions. Therefore, the objective of rejection is to find a proper criterion to detect the confusing samples and achieve high reliability. The rejection rate and reliability [41] are defined by

\[
\text{Rejection rate} = \frac{\text{Number of rejected samples}}{\text{Size of test dataset}} \times 100\% \\
\text{Reliability} = \frac{\text{Number of correct recognitions in accepted samples}}{\text{Number of accepted samples}} \times 100\%.
\]

He et al. [13] considered the confidence outputs of classifiers as feature vectors for the rejection process. The ideal outputs have high first rank (most likely category) confidence values that are quite distinct from the other classes. We first introduce two basic rejection measurements (First Rank Measurement and First Two Ranks Measurement), as well as a newly emerged rejection method named Linear Discriminant Analysis based Measurement, and then propose a new delicate rejection scheme which can achieve 100% reliability with less rejection samples.

3.5.1. Basic rejection schemes

The First Rank Measurement (FRM) [11] rejects the input pattern \(x\) based on the first rank output \(p_1(x)\) of classifier and a constant threshold \(T\). If \(p_1(x) \leq T\), then \(x\) will be rejected, and labeled as a rejected pattern. A human operator may handle all the rejected patterns later. Although the FRM can improve the reliability of the classifier, it is unable to reject the misrecognition samples with high first rank confidence values [42].

To overcome the disadvantage of FRM, the First Two Ranks Measurement (FTRM) is proposed by Suen et al. [12]. It rejects the input patterns by measuring the gap between the top-2 outputs \(p_1(x)\) and \(p_2(x)\) of the classifiers. For example, if \(\|p_1(x)-p_2(x)\| \leq T\), then the classifier will reject the input pattern and leave it for post-processing. The value of threshold \(T\) is determined based on the training set. However, in some correctly recognizable cases, the distance between the first two ranks is lower than threshold \(T\), which means in order to achieve a high reliability, the classifier has to reject extra samples, which could be classified exactly.

3.5.2. LDA rejection schemes

The Linear Discriminant Analysis based Measurement (LDAM) [13] considers rejection as a two-class problem, which applies one-dimensional application of the Fisher Linear Discriminant Analysis (LDA) to rejection. For a problem with \(M\) classes, the output vector of the classifier is presented by \(P(x) = \{p_1(x), p_2(x), \ldots, p_M(x)\}\) in descending order. Given two sets \(G_1(x) = \{p_1(x), \ldots, p_{r1}(x)\}\) and \(G_2(x) = \{p_{r2}(x), p_{r3}(x), \ldots, p_M(x)\}\), the LDA criterion is defined by

\[
J(w) = \frac{S_B}{S_W} = \frac{\mu_1 - \mu_2}{\sum_{i=1}^{M} \|p_i(x) - p_j(x)\|} = \frac{1}{(M-1)\sum_{i=2}^{M} \|p_i(x) - p_j(x)\|}
\]

where

\[
\mu_1 = p_1(x), \\
\mu_2 = \frac{1}{M-1} \sum_{i=2}^{M} p_i(x), \\
\Sigma_1 = (p_1(x) - \mu_1)^2, \\
\Sigma_2 = \frac{1}{M-1} \sum_{i=2}^{M} (p_i(x) - \mu_2)^2, \\
\Sigma_{12} = \frac{1}{2} \Sigma_1.
\]

The test sample \(x\) will be accepted if \(J(w) \geq T\), where \(T\) is a constant threshold derived from the training set. He et al. [13] tried that the LDAM outperforms the FRM and the FTRM on the CENPARMI Arabic [43] and NIST Numerals Database [44] with SVM classifier.

3.5.3. Cascade rejection schemes

In recent years, the cascade models have been long pursued for improving the accuracy of character recognition. However, none of them has been applied to the field of rejection. In order to achieve 100% reliability while rejecting as fewer test samples as possible, we design a new cascade measurement for rejection.

The structure of cascade rejection system is similar to the cascade combination scheme introduced in Section 3.3.3. We combine the sum-rule committee in the last layer with FTRM to obtain the rejected samples. To be specific, the rejection of a given sample \(x\) which has been rejected by the sixth layer is measured by FTRM with the output vector of the sum-rule committee.

Compared with the existing rejection measurements [11–13;41,42], the proposed cascade rejection method takes full advantage of the rejection capability of each classifier; it is able to provide high reliability with minimum rejection rate without importing another binary classifier ([41] trained SVM as criteria to reject the patterns). For further comparison, we also evaluate the rejection capability of the other classifier combination schemes proposed in Section 3.3 such as sum-rule, max-rule, SVM-rule and LDF-rule with FTRM.

Fig. 11. Similar character samples in different versions of RMB: (a) “1” in 1999; (b) “1” in 2000; (c) “0” in 2000; and (d) “0” in 2000.
4. Experiments

We evaluated the performances of feature extraction, classification, multiple classifier combination, distortion and rejection methods on the NUST-RMB2013 database, which contains 35 classes, 17,262 training samples (about 500 samples per class) and 7000 test samples (200 samples per class). Because the database mixes two different versions of RMB serial numbers, the characters “1” (Fig. 11a) in the 1999 version are similar to the “1” (Fig. 11b) in the 2000 version and the digits “0” (Fig. 11c) close to “O” (Fig. 11d) in both 1999 and 2000 versions of RMB. As a result, the misrecognitions among “1” and “I”, “0” and “O” are ignored (in practical applications, these samples can be recognized by means of the orders they appear in the serial numbers). The experiments are conducted on a PC with Intel(R) Core(TM) i7-2630QM CPU (2.00 GHz) and 8 GB memory.

4.1. Comparison of feature extraction methods

The performances of three feature extraction methods are evaluated by various classifiers trained with both original and gray level normalized data. We experimented two scales of Gabor feature with $\gamma = 0.5$, $\beta = 0.56$ and $\lambda = (6, 12)$, 480 gradient direction features (GDF) and 928 Gabor features (GF) can be extracted from the input pattern. After dimensionality reduction by LDA, both of them contain 34 feature vectors. The CNN features were extracted from the third layer of CNN trained without distortions. Since the dimension of the CNN features is much lower (only 100) than GDF and GF, we fed them directly to the classifiers for better performance. The hyperparameters of SVM classifiers with linear kernel or RBF kernel were selected by cross-validation on the training data.

Table 1 demonstrates the test accuracies of feature extraction methods on the NUST-RMB2013 database. The gradient direction feature slightly outperforms the Gabor feature with SVM and MQDF classifiers, yet inferior with LDF classifier on the original data. We surprisingly found that both GDF and GF yield higher accuracies than CNN. The best result of 99.55% is given by gradient direction feature performed on the gray level normalized dataset. We also implemented the Gabor feature with only one scale ($\lambda = 12$), and the recognition accuracies are lower than using two scales. The gray level normalization produces a better recognition rate (raised around 0.3%) for hand-crafted features such as GDF and GF, but it fails on the automatically extracted CNN features.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Original data</th>
<th>Normalized data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SVM</td>
<td>LDF</td>
</tr>
<tr>
<td>CNN</td>
<td>98.89</td>
<td>98.66</td>
</tr>
</tbody>
</table>

4.2. Performances of classification models

Using the best feature extraction methods (gradient direction feature), four classifiers SVM, LDF, MQDF and CNN have been tested on our database trained with original and gray level normalized data. Before classification, the features were reduced to 34 dimensions by LDA. The parameters of Linear and RBF kernel SVM were selected via 5-fold cross-validation. For MQDF, the number of principal eigenvectors ($k$) was heuristics empirically set as 17 according to 35 categories. We tested three CNNs with different feature maps (6, 15, 25) in the first layer. The result shows that additional features maps consume much longer training time and hardly provide any improvement on the recognition accuracy, therefore, 6 feature maps are adequate for the first layer. Despite the CNN generalizes better when trained with distorted samples, in this section, we insisted training the CNN without any transformations for fair competition with other classifiers. Before training, the initial weights were set randomly. We chose an initial learning rate of 0.001, which decayed 20% every 12 $\times$ 17,262 epochs until reaching the minimum learning rate of 0.00005. Normally, the mean squared error (MSE) of the CNN would stop decreasing after around 200 epochs.

Table 2 and 3 compare the recognition accuracies and computation times of five classifiers on NUST-RMB2013. MQDF is the most competitive model, which achieves the highest classification accuracy on both original and gray level normalized data with minimal training time. The accuracy of SVM model is lower than MQDF. Apart from its low complexity, LDF reaches comparable recognition rate on gray level normalized dataset with MQDF and SVM. We can see that the CNN performs poorly when trained without any distortions. The reason why CNN is the only classifier which decreases the performance after the gray level normalization lies in the interaction of the inaccurate gray-scale normalization and the unique mechanism of CNN. Due to the imprecise normalization method, the gray level of the small circles and background could also be stretched, as a result, they may become even more obvious after normalization. Since the CNN retrieves discriminative features directly from the raw images, it would consider the enhanced circles and the texture of background along with the character strokes as relevant features, which causes misrecognition of the input samples. However, the gradient direction features and Gabor features densely sample on a regular grid disregarding the gray level distribution of the image, which helps to minimize the influence from the ill-stretched circles and background, and provides superior results.

As listed in Table 3, the statistical classifiers process faster than the other classifiers in both training and test procedures. Due to the huge calculations, it really takes an enormous amount of time to train the neural network.

4.3. Recognition results on multiple classifier system

We compared the performances of different classifier combination schemes on RMB dataset. The outputs of three selected classifiers including SVM with RBF kernel, MQDF and CNN (we
Table 4
Test accuracies (%) of classifier combination methods on NUST-RMB2013.

<table>
<thead>
<tr>
<th>Combination</th>
<th>Original data</th>
<th>Normalized data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum</td>
<td>99.60</td>
<td>99.63</td>
</tr>
<tr>
<td>Max</td>
<td>99.57</td>
<td>99.57</td>
</tr>
<tr>
<td>SVM-rule</td>
<td>99.61</td>
<td>99.64</td>
</tr>
<tr>
<td>LDF-rule</td>
<td>99.44</td>
<td>99.53</td>
</tr>
<tr>
<td>Cascade</td>
<td>99.67</td>
<td></td>
</tr>
</tbody>
</table>

Table 5
Comparison among distortion methods on NUST-RMB2013.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Distortion</th>
<th>Recognition rate</th>
<th>Original data</th>
<th>Normalized data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear SVM</td>
<td>None</td>
<td>98.90</td>
<td>99.53</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scaling &amp; rotations</td>
<td>99.21</td>
<td>99.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 × 1 Translation</td>
<td>99.29</td>
<td>99.57</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 × 2 Translation</td>
<td>99.51</td>
<td>99.61</td>
<td></td>
</tr>
<tr>
<td>RBF SVM</td>
<td>None</td>
<td>99.31</td>
<td>99.46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scaling &amp; rotations</td>
<td>99.30</td>
<td>99.47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 × 1 Translation</td>
<td>99.40</td>
<td>99.61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 × 2 Translation</td>
<td>99.59</td>
<td>99.64</td>
<td></td>
</tr>
<tr>
<td>MQDF</td>
<td>None</td>
<td>99.37</td>
<td>99.59</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scaling &amp; rotations</td>
<td>99.30</td>
<td>99.53</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 × 1 Translation</td>
<td>99.43</td>
<td>99.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 × 2 Translation</td>
<td>99.56</td>
<td>99.56</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>None</td>
<td>99.60</td>
<td>98.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scaling &amp; rotations</td>
<td>99.26</td>
<td>99.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 × 1 Translation</td>
<td>99.47</td>
<td>99.40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 × 2 Translation</td>
<td>99.09</td>
<td>98.54</td>
<td></td>
</tr>
</tbody>
</table>

Table 6
Test accuracies (%) of distorted CNN features on NUST-RMB2013.

<table>
<thead>
<tr>
<th>Distortion</th>
<th>CNN + SVM</th>
<th>CNN + MQDF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original</td>
<td>Normalized</td>
</tr>
<tr>
<td>None</td>
<td>98.89</td>
<td>98.46</td>
</tr>
<tr>
<td>Scaling &amp; rotations</td>
<td>99.31</td>
<td>98.57</td>
</tr>
<tr>
<td>1 × 1 Translation</td>
<td><strong>99.37</strong></td>
<td>99.27</td>
</tr>
<tr>
<td>2 × 2 Translation</td>
<td>99.36</td>
<td>99.07</td>
</tr>
</tbody>
</table>

exclude LDF classifier from the classifier combination due to its inferior recognition accuracy) were transformed to confidence values, and combined at the measurement level with parameters selected in Section 4.2. Table 2 lists the performances of the individual classifiers. Before combination, all of the classifiers were trained separately with both original and gray level normalized data. For a higher recognition rate, the CNN model was trained with samples distorted by 1 × 1 eight direction translations.

The elementary schemes combine confidence measurements with sum-rule and max-rule. The weights $\omega$ of the linear combination schemes were estimated by Linear SVM and LDF trained on the validation dataset. The confidence outputs of the test sample were fed to the meta-classifier to get the classification result. The cascade combination scheme—composed of seven layers—was arranged in the order of SVM, MQDF, CNN, GN-SVM, GN-MQDF, GN-CNN (“GN-Classifier” means the classifier is trained with gray level normalized data). The first six layers sequentially rejected the suspicious samples according to the threshold $T$, and the last layer combined the results of the above six classifiers through a sum-rule committee. The value of $T$ should be large enough to reject all confusing data in each layer. Because the confidence outputs of the six classifiers have the same value range $[0, 1]$, we set $T$ as 0.75 to ensure a high confidence [40].

The classification accuracies on the test dataset are shown in Table 4. Despite its simplicity, the sum-rule combination is not sensitive to the noises and performs comparable with the other complicated models. The Linear SVM method of combining individual classifiers with trained weights achieves higher accuracies than elementary rules. However, due to insufficient validation samples, the meta-LDF classifier is under-generalized, which leads to inferior combination performances.

The cascade combination scheme gives the highest accuracy of 99.67% by utilizing the classifiers trained on both original and gray level normalized datasets with the recognition rates of 99.31%, 99.37%, 99.47%, 99.53%, 99.59%, 99.40%. Setting the threshold $T=0.75$, the first six layers only rejected 1.64% test samples and fed them to the sum-rule committee.

4.4. The benefits of distortions

To justify the benefits of different types of distortions described in Section 3.4, we conducted experiments to compare the recognition results with and without expanding the dataset using the gradient direction feature with dimensionality reduction by LDA. Each classifier was trained by means of three types of distortion methods. The translation methods shifting the input data one or two pixels in eight directions multiplied the size of training dataset by nine. For distortions of scaling and rotations, the transformed scales and rotation angles were stochastically chosen in the ranges of $[0, 15]$ and $[0, 15^\circ]$; nine new samples were generated for each training data.

According to the specific training techniques of different classifiers, we expanded the input samples before feeding them to SVM (with linear and RBF kernel) and MQDF. The distorted and original samples were alternately used in the neural network, e.g. eight translated samples were sequentially applied to the forward propagation after the original one. The parameters of SVM were selected by cross-validation on the training dataset, and $k$ of MQDF was set to 17. The learning rate of CNN decreased every 9 × 17,262 epochs if trained with translated samples, and 10 × 17,262 epochs for affine distortions.

Table 5 shows the recognition results. All of the distortion methods lead to approving results, especially the performances of CNNs on the original dataset boosted at least 0.5% by training with transformed samples. The distortions are even more beneficial to the classifiers trained with original samples than the gray level normalized ones. Since the skew of serial numbers is corrected during the extraction process and the scales of each data sample are almost the same (all the RMB of 100 yuan have the same size and scanned with the same contact image sensor), compared to the distortion of scaling and rotations, it appears that the variances simulated by the translations are closer to the transformations in the test samples, as a result, both of the translation methods outperform the distortions of scaling and rotations on the test dataset. The SVM with RBF kernel trained by $2 \times 2$ translated dataset achieves the best recognition rate of 99.64%.

To further improve the recognition rate, we tested the translations, which expanded the training dataset 17 times by using both $1 \times 1$ and $2 \times 2$ translations, and also tried multiplying the size of the training dataset by 20 with the distortions of scaling and rotations. However, both of them barely give any improvement while taking much longer training time, the results are omitted in this paper. More distinct CNN features can also be obtained by distortion techniques. Table 6 reports the superior performance of distorted CNN features.

4.5. Rejection analysis

We investigated the rejection performance of six classifiers used in Section 3.3.3, namely SVM, MQDF, CNN, GN-SVM, GN-MQDF,
GN-CNN ("GN-Classifier" means the classifier is trained with gray level normalized data) with the rejection criteria of FRM, FTRM and LDAM. The measurement thresholds are related to the rejection rates, normally a lower threshold produces less rejection samples. Choosing the thresholds of the rejection measurements is a tradeoff between reliability and rejection rate.

The tradeoff curves with different rejection rates for FRM, FTRM and LDAM are shown in Fig. 12, 13 and 14 respectively. The reliability grows rapidly with the rejection rate under 2%, and the curves become flat when approaching 1. Since FTRM is more distinct than FRM, the reliability performances of the classifiers trained with normalized data using FTRM exceed those of the FTW with a steeper increase among the tradeoff curves. The results indicate that, although LDAM is proved superior to FRM and FTRM on the CENPARMI Arabic Database [13], it performs slightly lower than FTRM in our experiment. As the LDAM considers the rejection problem as a two-class problem by maximizing the separation between the first rank confidence value and the others, it fails to provide proper measurement when the second rank confidence value is as high as the first rank while the other confidence values are low. Table 7 lists the rejection rates of each classifier for 99.9% and 100% reliability with the criteria of FRM, FTRM and LDAM.

We conducted the cascade rejection scheme with majority voting and sum committees in the last layer. However, as all of the six classifiers could possibly give the same misclassification results for the input sample, the majority voting cannot guarantee 100% reliability of the system. By combining the sum-rule committee in the last layer with FTRM, we consider the rejection capability of the six different classifiers, as well as the statistical advantage of the sum committee. Since the classifier combination strategies (sum-rule, max-rule, SVM-rule and LDF-rule) proposed in Section 3.3 likewise have certain advantages compared to the

Please cite this article as: B.-Y. Feng, et al., Automatic recognition of serial numbers in bank notes, Pattern Recognition (2014), http://dx.doi.org/10.1016/j.patcog.2014.02.011
single classifier, their rejection capability is evaluated by FTRM along with the cascade scheme (Figs. 15 and 16).

Table 8 shows the reliability of the multiple classifier system with FTRM by rejecting 1% of the test samples. The cascade strategy provides superior results to the classifier combination strategies on both original and normalized data, which achieve the reliability of 100% with a rejection rate of 1.01%. The multiple classifier system of sum-rule, max-rule and SVM-rule trained on the gray-level normalized data benefited from the characteristics of different classifiers and the model's structural robustness obtains approving results as well, which reaches 100% reliability by rejecting 1.36%, 1.63% and 1.67% of the test data.

Comparing with the rejection measurements of FRM, FTRM and LDAM, the classifier combination system significantly enhances the rejection behavior. For instance, the best result provided by FRM, FTRM and LDAM still needs to reject 3.46% test samples to reach 100% reliability while the rejection rate of the proposed cascade method is only 1.01%.

The cascade method takes the full advantage of the rejection capability of each classifier and the robust property of the sum committee to achieve promising results. The proposed six classifiers (SVM, MQDF, CNN, GN-SVM, GN-MQDF and GN-CNN) have certain internal complementarity among them, which mainly derives from two aspects. Firstly, the principles of different classifiers are quite distinct. To be more specific, the MQDF is a parametric classifier, which assumes a functional form for the class-conditional density of each class and estimates the parameters by maximum likelihood. Meanwhile both the SVM and the CNN are discriminative learning models, aim to maximize a geometric margin of hyper-plane and minimize a sum-of-squares error function, respectively. Secondly, even the classifiers involving the same classification model, if trained with original and gray-level normalized data, would have individual attributes too.

To this end, our cascade rejection method provides an effective way to utilize these complementarities and be able to reject the confusing samples with much lower threshold than each of the relevant classifier.

5. Conclusions

This paper investigates a new topic of serial numbers recognition in bank note images. A new database NUST-RMB2013 is introduced for the research of character recognition, which contains about 25,000 serial number character samples of 35 classes; it can be normalized on gray level by a piecewise linear stretch method based on the histogram distribution of the character strokes. We comprehensively compared the effect of recently emerged techniques used in character recognition, including three efficient feature extraction methods, four well-known classification methods, five classifier combination strategies, three types of distortions and eight rejection schemes. To the best of our knowledge, this is the first work that applies the cascade schemes to the context of rejection, which could dramatically reduce the number of rejected samples while achieving 100% reliability.

The experimental results in feature extraction reveal the advantage of gradient direction feature on printed characters. It outperforms both the Gabor feature and CNN trainable feature on our dataset. The MQDF model shows superior performance with the lowest memory space and minimum computation cost. It achieves the highest test accuracy of 99.59% among all the classifiers trained with gradient feature extracted from the gray level normalized dataset. The convolutional neural network performs inferiorly when trained without distortions. As for multiple classifier combination, the linear combination method using SVM performs slightly better than the sum-rule, which optimizes the weights for each classifier via training a meta-classifier. Benefiting from the complementarity among the classifiers, the best result of 99.67% on our dataset is achieved by the proposed cascade classifier combination strategy.

This paper describes three different types of distortions, namely translations of 1 or 2 pixels and transformations of scaling and rotations, which are applied to the training data either before or during the training process. The results demonstrate that almost all the classifiers and CNN trainable features could evidently benefit from the distortions, especially, recognition accuracy of CNN increases more than 0.5%. We also discover that the translations provide a closer simulation of the variances among the test dataset, and produce a higher recognition rate than scaling and rotations.

We evaluated three basic rejection schemes for six classifiers with different rejection rates. The result indicates that the FTRM leads to a higher reliability than FRM and LDAM when set to the same rejection rate. A desirable rejection strategy should give a high reliability as well as a low rejection rate. The GN-SVM (SVM trained with gray level normalized data) with FRM measurement achieves the best performance of 100% reliability at a 3.46% rejection rate. We also tested the classifier combination schemes with FTRM, and the sum-rule classifier ensemble system achieves 100% reliability with a rejection rate of 1.36%. Meanwhile, our novel cascade rejection method shows obvious advantages compared with other measurements—it gives 100% reliability by rejecting only 1.01% of the test samples.

In the future, we expect to yield higher recognition accuracies by designing a new classifier, which is more suitable to our problem. Further improvements will also focus on the classifiers used in the cascade rejection scheme; applying different classification models may produce better rejection performances. Moreover, we will also extend our study to the recognition of serial numbers in other types of banknotes, such as Euros, U.S. and Canadian dollars.

Conflict of interest

None declared.
Acknowledgments

This work was supported by the China Scholarship Council (CSC) under File no. 201206840018. We wish to thank the colleagues of Centre for Pattern Recognition and Machine Intelligence (CENPARMI) for their great support and help in our work. The anonymous reviewers who provided constructive and helpful comments are much appreciated.

References


Bo-Yuan Feng received the B.S. degree in computer science from the Department of Computer Science, Nanjing University of Science and Technology, China, in 2009. He researched at Centre for Pattern Recognition and Machine Intelligence (CENPARMI), Concordia University for 1 year since September 2012. He is currently a PhD Candidate in pattern recognition and intelligent systems at the Department of Computer Science, Nanjing University of Science and Technology, China. His research interests include image processing, pattern recognition, machine learning, printed and handwritten character recognition.
Mingwu Ren is a Professor at the School of Computer Science and Technology, Nanjing University of Science and Technology, China. He received M.S. and Ph.D. degrees in Pattern Recognition and Computer Image Processing from Nanjing University of Science and Technology, China. His research interests include computer vision, digital image processing, pattern recognition and artificial control.

Xu-Yao Zhang received the B.S. degree in computational mathematics from Wuhan University, Wuhan, China, in 2008, the Ph.D. degree in pattern recognition and intelligent systems from Institute of Automation of Chinese Academy of Sciences, Beijing, China, in 2013. He is currently an Assistant Professor at the National Laboratory of Pattern Recognition, Institute of Automation of Chinese Academy of Sciences, Beijing, China. His research interests include pattern recognition, machine learning, and especially large category classification, dimensionality reduction, classifier adaptation, handwritten character recognition, and image processing.

Ching Y. Suen is the Director of CENPARMI and the Concordia Chair on AI & Pattern Recognition. He received his Ph.D. degree from UBC (Vancouver) and his Master’s degree from the University of Hong Kong.

He has served as the Chairman of the Department of Computer Science and as the Associate Dean (Research) of the Faculty of Engineering and Computer Science of Concordia University. He has served at numerous national and international professional societies as President, Vice-President, Governor, and Director. He has given 40 invited/keynote papers at conferences and about 200 invited talks at various industries and academic institutions around the world. He has been the Principal Investigator or Consultant of 30 industrial projects. His research projects have been funded by the ENCS Faculty and the Distinguished Chair Programs at Concordia University, FCAR (Quebec), NSERC (Canada), the National Net-works of Centres of Excellence (Canada), the Canadian Foundation for Innovation, and the industrial sectors in various countries, including Canada, France, Japan, Italy, and the United States. Currently, he is the Editor-in-Chief of the journal of Pattern Recognition and an Adviser or Associate Editor of 5 journals. Actually he has held previous positions as Editor-in-Chief, or Associate Editor or Adviser of 5 other journals.

Please cite this article as: B.-Y. Feng, et al., Automatic recognition of serial numbers in bank notes, Pattern Recognition (2014), http://dx.doi.org/10.1016/j.patcog.2014.02.011