PERGAMON

Pattern Recognition 35 (2002) 2823–2832

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Combination of multiple classifiers using probabilistic dictionary and its application to postcode recognition

Yue Lu, Chew Lim Tan*

Department of Computer Science, School of Computing, National University of Singapore, Kent Ridge 117543, Singapore

Received 18 December 2000; accepted 3 December 2001

Abstract

Combination of multiple classifiers is regarded as an effective strategy for achieving a practical system of handwritten character recognition. A great deal of research on the methods of combining multiple classifiers has been reported to improve the recognition performance of single characters. However, in a practical application, the recognition performance of a group of characters (such as a postcode or a word) is more significant and more crucial. With the motivation of optimizing the recognition performance of postcode rather than that of single characters, this paper presents an approach to combine multiple classifiers in such a way that the combination decision is carried out at the postcode level rather than at the single character level, in which a probabilistic postcode dictionary is utilized as well to improve the postcode recognition ability. It can be seen from the experimental results that the proposed approach markedly improves the postcode recognition performance and outperforms the commonly used methods of combining multiple classifiers at the single character level. Furthermore, the sorting performance of some particular bins with respect to the postcodes with low frequency of occurrence can be improved significantly at the same time. © 2002 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Combination of multiple classifiers; Probabilistic dictionary; Numeral recognition; Postcode recognition; Single character level; Postcode level

1. Introduction

Recognition of unconstrained handwritten numerals is still a challenging problem encountered in many real world applications, such as postal mail sorting, bank cheque recognition. Several practical systems have been developed successfully to meet the requirements of the industry, in which automatic postal mail sorting is a typical application of character recognition. However, the success never stops researchers' efforts in improving the recognition performance farther [1,2]. A great deal of research has been focused on achieving a higher recognition rate and higher recognition reliability.

The performance of a recognition system can be evaluated by its recognition rate R_c, reject rate R_r and error rate R_e in general. Ideally, we want to achieve a high recognition rate, but in a practical system we are equally concerned with the issue of recognition reliability, which is commonly defined as

$$Rel = \frac{R_c}{1 - R_r} = \frac{R_c}{R_c + R_e}. \tag{1}$$

For a practical system such as automatic mail sorting machine, not only a high recognition rate is required, but also high recognition reliability is crucial. First, only when all of the characters in a postcode are correctly recognized, then the corresponding mail can be correctly sorted. A character recognition system with the recognition rate of R_c at the single character level can only achieve the recognition rate of about $(R_c)^N$ at the N-characters' postcode level in theory. Second, the users of the mail sorting machines are concerned much about the "sorting quality". High recognition reliability is another important performance measure for a practical sorting system. The erroneous recognition rate of

---

* Corresponding author. Tel.: +65-874-2900; fax: +65-779-4580.

E-mail address: tancl@comp.nus.edu.sg (C.L. Tan).

PII: S0031-3203(01)00237-0

R_e at the single character level will theoretically result in the mis-sorting rate of about $N \times R_e$ at the postcode level, because any one of the N characters in the postcode that is erroneously recognized leads to the mis-sorting of the corresponding mail. Maximum correct sorting and minimum mis-sorting are the two key requirements of an automatic sorting system. However, higher recognition rate generally means lower recognition reliability due to possibility of higher mistake rate. A trade-off between the recognition rate and the recognition reliability is a reasonable choice, under the precondition that the recognition reliability is acceptable in the view of recognition quality. Both of the recognition rate and recognition reliability, therefore, must be taken into account while we estimate the recognition performance of a system.

Moreover, in the view of mail sorting ability, the system's recognition performance at the postcode level is more significant rather than that at the single character level. Sometimes, a classifier optimized at the single character level does not necessarily result in the optimal recognition performance at the postcode level for a postcode recognition system.

The recognition problem is intrinsically so complicated that we often encounter many difficulties such as distortion of image due to high-speed sampling and broken strokes caused by various kinds of reasons. It is a fact that a single feature extraction method or a single classifier algorithm could not yield a satisfactory result. Many researchers pay their attention to the use of more complex structures involving several feature extractors and classifiers.

In 1990s, the concept of combining multiple classifiers (experts) was proposed for the development of highly reliable character recognition systems. This is motivated by the observation that different classifier designs potentially offered complementary information about the patterns to be classified, which could be harnessed to improve the overall performance of the recognition system. There are numerous possible strategies that may be utilized to combine individual classifiers [3–12]. All of these methods for combining multiple classifiers concentrate on improving the recognition performance of single characters.

We know that many word recognition systems benefit from the application of a word dictionary [13–15]. However, in the issue of postcode recognition, although the postcode dictionary sometimes is used in the post-processing to verify whether the results are valid or not, the knowledge such as the postcode dictionary and the frequency of occurrence of postcodes are ignored generally while combining multiple classifiers, to meet the object of designing a general-purpose recognizer. The approach of an algorithm to generalize its computation is now understood to be less than optimal in many circumstances. The principal motivation underlying this paper is to develop a special-purpose system instead of a general-purpose system, in which the system makes use of as much knowledge as it can in the special application circumstances.

As far as the problem of postcode recognition is concerned in this paper, a probabilistic postcode dictionary of a particular mail sorting site is constructed according to the frequency of occurrence of postcodes at the corresponding site. Then, the probabilistic postcode dictionary is applied to the procedure of the combination of multiple classifiers, in which the combination decision is carried out at the postcode level rather than at the single character level.

In our experiments, five individual classifiers are combined to recognize the postal numerals taken from Chinese postal mails. The five classifiers are chosen from tools available to us that represent a variety of sources, methodologies and output transformations. The outputs of two classifiers are at the abstract level, and that of the other three classifiers are at the measure level. Two classifiers are from commercial sources and the others are research tools based on neural networks and decision trees. All of the postcode images were collected from a particular mail sorting site. A probabilistic postcode dictionary of the corresponding site is utilized for combining the results of the five classifiers. The experiments have shown that an efficient approach for improving the postcode recognition ability by combining multiple classifiers at the postcode level using the probabilistic postcode dictionary can be achieved. The proposed approach outperforms the traditional approach with about 5% improvement of recognition ability at the postcode level. Furthermore, the sorting performance of some special bins with respect to the postcodes with low frequency of occurrence can be improved significantly, which obviously benefits from the application of probabilistic postcode dictionary.

The paper is organized as follows. Section 2 outlines five individual classifiers used in the experiments. In Section 3, we discuss the transformation of outputs of the five individual classifiers. Section 4 describes the approach for combining multiple classifiers at the postcode level using a probabilistic postcode dictionary, and Section 5 gives the experimental results of the proposed approach. Finally, Section 6 summarizes the results of the paper and offers concluding remarks.

2. Classifiers for isolated numeral characters

Five isolated handwritten numeral character classifiers are included in the experiments. Each of them is designed independently, in which different features and different classifier structures are utilized.

Two of the five classifiers are provided by another research group. As the two classifiers are commercial products, we do not know much details about them, except that the outcome of one classifier (CP1) is at the abstract level, in which the classifier assigns the input pattern a unique label to declare which class it belongs to; and the other classifier (CP2) gives a similarity measurement that evaluates each class with a measurement value to represent the degree that the input pattern is similar to the corresponding class.

The other three classifiers are the multiple-layer neural network classifier (MLP), the tree classifier based on topological features (TCTF) and the HAVNET classifier (HAVNET), respectively.

2.1. Multiple-layer perceptron (MLP)

One feed forward neural network (multiple-layer perceptron) is trained as a numeral classifier and the back-propagation algorithm is applied during the training procedure. The directional codes histogram features and gray-scale transformation features [4] are employed as the input to the neural network. A rectangular frame enclosing the contours of the numeral is divided into 4×6 rectangular grids. In each grid, a local histogram of the contour chain codes is calculated. These local histograms compose the directional codes histogram features. The contour direction is quantized into one of 4 possible values (H,V,L,R), thus the feature has 96 components for the image of one numeral character. The gray-scale transformation features were generated by applying an averaging filter to the binary image of numerals three times on a 3×3 window. The entire image is divided into 8×12 grids, and the normalized average value in each grid is chosen to be its feature. The network is structured as 192 nodes in the input layer corresponding to the feature space (96 nodes for the directional codes histogram features and 96 nodes for the gray-scale transformation features), 30 nodes in the hidden layer, and 10 nodes in the output layer corresponding to the 10 numerals.

2.2. Tree classifier based on topological features (TCTF)

In this method, a tree classifier based on topological features and contour information is employed for recognizing numerals, which is similar to [16]. A contour-following algorithm is carried out on the images of numeral characters, and then the numerals are divided into three groups according to their topological properties.

In the first group, the characters have only one contour, which include the numerals “1”,“2”,“3”,“4”,“5”,“6”,“7”, etc. The numerals in this group are recognized by their Fourier descriptors of their outer contour. In calculating the Fourier descriptors, the contour pixel with the furthest distance from the contour centroid is chosen as the starting point, and then 36 sampled points are obtained from the contour by an equal sampling interval. To avoid the effect of difference of character image size, the coordinates of the 36 samples are normalized to have the maximum sizes of $80(W) \times 96(H)$. The discrete Fourier transform is done based on the coordinates of the sample pixels, in which the coordinates are treated as a complex numbers. The complex coefficients including 72 parameters constitute the normalized Fourier descriptors of the contour.

In the second group, the characters have two contours, which include the numerals “0”,“2”,“4”,“5”,“6”,“9”, etc. The numeral “5” is recognized first according to the topological relationship between the two contours as the first level of recognizing the numeral of this group. Because the two contours are disjunct only for the numeral “5” in this group, while for the other numerals in this group, one contour is always inside the other contour, the numeral “5” can be distinguished from this group easily. In the second level, the features including the mean and variance of the normalized distance function are utilized to discriminate the numeral “0” from the others. To compute the mean and variance of the normalized distance function, the Euclidean distance from the contour centroid to each contour pixel is calculated, and normalized by dividing it by the longest distance. In the third level, the relative position of the two contours’ centroids is used for discrimination of the numeral “4” and “9” from the numeral “2” and “6”, because the interior contour centroids of “4” and “9” are always in the upper part of their images while that of “2” and “6” are always in the lower part of their images. In the fourth level, the Fourier descriptor of the outer contour is applied to discriminate numeral “4” from numeral “9”. The numeral “2” and “6” are discriminated by the difference of the relative position of the outer and interior contours, because the interior contour centroid of numeral “2” is always at the left side of its outer contour centroid, while the interior contour centroid of numeral “6” is always at the right side of its outer contour centroid.

In the third group, basically there is only the numeral “8” that has three contours. However, some abnormal numerals with three contours are still classified into this group because of the complexity of writing styles. These numerals are discriminated by the comparison of outer contour Fourier descriptors of the input image with that of stored models to verify whether the input image is “8” or not.

2.3. Havnet

The third classifier is based on HAVNET [17,18], which is a neural network employing Hausdorff distance as a similarity metric. Two-dimensional binary images are used directly for training and recognition after preprocessing of noise elimination and scale normalization. HAVNET consists of three layers: the plastic layer, the Voronoi layer, and the Hausdorff layer. The input of HAVNET is a two-dimensional binary image. The numeral image is scaled to an image with the size of height = 12 and width = 10. Plastic layer weights are trained on samples first to represent all sub-class patterns. In the recognition procedure, the Voronoi layer and Hausdorff layer are used to compute the overall dissimilarity between the input image and the learned patterns in the format of Hausdorff distance. The output layer of the network consists 10 nodes, one representing each numeral category. The HAVNET neural network outputs the directed Hausdorff distance between the input image and the patterns stored in the channels of plastic layer which have been pre-trained prior to the recognition procedure.

3. Classifiers' output transformation

The five classifiers can be divided into two categories. The first type is that the outputs are at the abstract level (CP1 and TCTF), in which the classifier assigns a unique label to declare which class the input pattern belongs to. The second type is that the outputs are at the measure level (CP2, MLP and HAVNET), in which the classifier evaluates each class with a measurement value to address the degree that the input pattern belongs to the corresponding class.

The outputs of each classifier are transformed into the same range before combining them into the score of possible results.

3.1. Classifiers with abstract level outputs

A confusion matrix CM_k of a classifier e_k with abstract level outputs (i.e. CP1 and TCTF) gives statistical results on the basis of a large set of samples to demonstrate the classifier's discriminating ability by recognition rate and substitution rate. It is employed to estimate the recognition performance of the classifier e_k.

The definition of CM_k is the same as that in Ref. [7]. It is a $M \times (M+1)$ matrix, where M denotes the number of classes ($M = 10$ for the problem of numerals recognition), and class $(M+1)$ represents rejection. Each row of the matrix corresponds to a class and each column corresponds to a recognition result. Thus, its element $n_{mm'}^{(k)}$ denotes that:

- The correctly recognized number of samples from class C_m by the classifier e_k when $m = m'$,
- The erroneously recognized number at which classifier e_k assigns samples from class C_m to class C'_m when $m \neq m'$ and m' is not equal to $(M+1)$,
- The rejected number of samples from C_m by classifier e_k when $m' = (M+1)$.

If the number of samples used for generating the confusion matrix CM_k is sufficiently large, and the samples reflect the distribution of pattern space, the confusion matrix will reflect the discriminating ability of the classifier e_k. The confusion matrix CM_k can be regarded as a priori knowledge of classifier e_k during the combination procedure.

For e_k, the total number of samples recognized as $m' = e_k(x)$ is

$$n_{\cdot m'}^{(k)} = \sum_{m=1}^{M} n_{mm'}^{(k)} \quad m' = 1,2,\ldots,M+1. \tag{2}$$

For the samples whose recognition results are m', the probability of coming from class $C_m (m = 1,2,\ldots,M)$ can be indicated in the form of conditional probabilities [7]. We take it as the contributive score for the combination of multiple classifiers in the condition of $e_k(x) = m'$.

$$s^{(k)}(x \in C_m) = P(x \in C_m | e_k(x) = m') = \frac{n_{mm'}^{(k)}}{n_{\cdot m'}^{(k)}} \quad m = 1,2,\ldots,M. \tag{3}$$

3.2. Classifiers with measure level outputs

For the outputs of classifiers in the measurement format, a transformation of the outputs is required for the combination as well, because the output values of the individual classifiers may vary from one to another. For example, they may represent similarity measures or distance measures. Therefore, it is usually necessary to transform them to the same range.

Furthermore, many transformation schemes can be applied, such as linear, logarithmic, or exponential transformation to improve the discriminating ability of the measure score. The experiments by Achermann [19] and Yu [20] showed that the logarithmic transformation yields the best results in general.

For the classifier with outputs of similarity measure (CP2) and the classifier of feed forward neural network (MLP), under a logarithmic transformation, outputs $y_m^{(k)} (m = 1,2,\ldots,M)$ of the classifier e_k are mapped into:

$$z_m^{(k)} = \log(1 + y_m^{(k)}) \tag{4}$$

then $s^{(k)}(x \in C_m)$ is calculated by

$$s^{(k)}(x \in C_m) = \frac{z_m^{(k)}}{\sum_{m'=1}^{M} z_{m'}^{(k)}}, \quad m = 1,2,\ldots,M. \tag{5}$$

For the classifier with outputs of distance measure (HAVNET), the output value in $I_{y'} = [y_{\min}'^{(k)}, y_{\max}'^{(k)}]$ are first linearly mapped inversely into $I_y = [0.0, 1.0]$ by interpolation:

$$y_m^{(k)} = 1 - \frac{y'^{(k)} - y_{\min}'^{(k)}}{y_{\max}'^{(k)} - y_{\min}'^{(k)}}. \tag{6}$$

Then, the transformation is computed as that for the classifier with outputs of similarity measure.

4. Combining the classifiers at the postcode level using probabilistic postcode dictionary

High recognition rate is certainly required for a practical system. On the other hand, high recognition reliability [2,12,16] is crucial too. Maximum correct sorting and minimum mis-sorting are the key requirements of an automatic sorting system. Undoubtedly, the postcode dictionary and the frequency of occurrence of the postcodes are beneficial knowledge for improving the postcode recognition ability. With this motivation, the present approach applies the knowledge such as the postcode dictionary and the

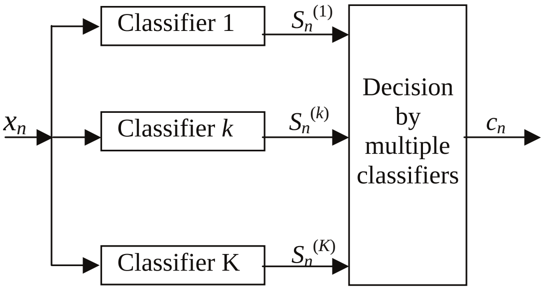

Fig. 1. Combination of multiple classifiers at the single character level.

Fig. 2. Dictionary used for verification in the post-processing.

occurrence frequency of each postcode to the procedure of multiple classifier combination, in which the five classifiers are combined at the postcode level using a probabilistic postcode dictionary.

Many postal numeral recognition systems combine the results of several classifiers at the single character level as illustrated in Fig. 1. The postcode dictionary is only applied in the post-processing procedure for verifying whether the numeral series is valid or not in the postcode dictionary, as shown in Fig. 2. In our approach, a probabilistic postcode dictionary of a particular mail sorting site is constructed according to the frequency of occurrence of postcodes at the corresponding site. Then, the probabilistic postcode dictionary is applied to the procedure of the combination of the five classifiers, in which the combination decision is carried out at the postcode level rather than at the single character level, as illustrated in Fig. 3.

Table 1
Postcodes vs. their frequency of occurrence

Postcode	Frequency of occurrence (%)
310027	0.05134
310013	0.04683
⋮	⋮
310428	0.04377
⋮	⋮
316428	0.00238
⋮	⋮
201501	0.00095
100081	0.00094
⋮	⋮
831911	0.00005
853402	0.00001
817200	0.00001

4.1. Probabilistic postcode dictionary

Our experimental data came from a Chinese postal mail sorting machine. It is found by investigating the China postcode dictionary that only about 125,000 unique postcodes are valid for the Chinese six-digit postcode system. A significant redundancy exists although there is no check digit in the postcode system. Furthermore, the distribution of postal destinations differs among various mail processing centers in China. Thus for a specified mail sorting machine at a particular mail processing center, the occurrence frequency of each postcode is surely not the same. For example, the frequency of occurrence of some postcodes given in Table 1 is the statistical results of all the postal mails processed by the automatic mail sorting machine in

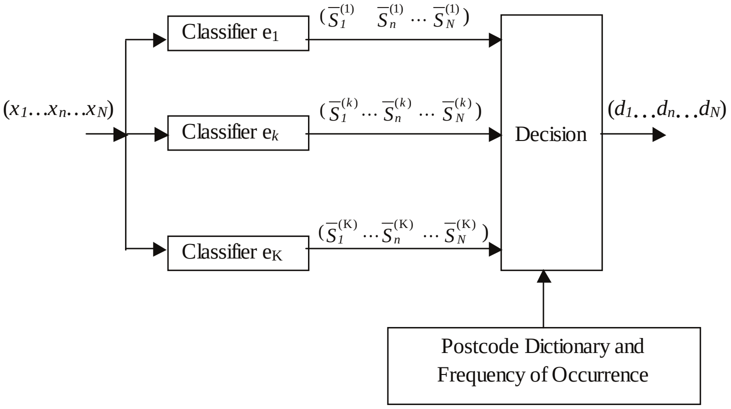

Fig. 3. Combination of multiple classifiers at the postcode level.

Hangzhou Postal Mail Processing Center over a long period of time.

Based on the statistics, a probabilistic postcode dictionary is constructed to recode each valid postcode and its corresponding frequency of occurrence.

4.2. Combining the classifiers using the probabilistic postcode dictionary

Based on the contributive score of each classifier to each character, the likelihood of each possible postcode is calculated to obtain the likelihood of the corresponding postcode, in which the probabilistic postcode dictionary is utilized for combining the five classifiers. Then, the postcode with the highest likelihood score, whose likelihood is absolutely high or sufficiently higher than the next best, is selected as the final postcode recognition result. The details are described as follows:

Let's suppose $X=(x_1,x_2,\dots,x_N)$ is the input of a postcode image with N numerals, and $D=(d_1,d_2,\dots,d_N)$ is a valid postcode in the postcode dictionary Ω, where $N=6$ for the Chinese postcode system, and $d_n \in \{0,1,\dots,9\}$.

K independent classifiers recognize the N numerals in parallel, and the score is computed as described in Section 3. For the input pattern x_n, the score with respect to all classes by the classifier e_k is represented as

$$\bar{S}_n^{(k)} = [s^{(k)}(x_n \in C_1),\dots,s^{(k)}(x_n \in C_m),\dots,s^{(k)}(x_n \in C_M)]. \tag{7}$$

The score of X coming from D is estimated as

$$S(D|X) = \prod_{n=1}^{N} s(d_n|x_n), \tag{8}$$

where

$$s(d_n|x_n) = \frac{1}{K}\sum_{k=1}^{K} s^{(k)}(x_n \in C_{dn}). \tag{9}$$

On the assumption that the occurrence frequency of postcode D for the specified site is expressed as $f_0(D)$, the score is computed as:

$$p(D|X) = e^{f_0(D)} S(D|X). \tag{10}$$

Our investigation has found that the occurrence frequency of each postcode varies significantly, as listed in Table 1. To avoid the situation when postcodes with low occurrence frequencies are never identified, the exponential function here is empirically set to regulate the contribution of the occurrence frequency of postcode dictionary to the decision score $p(D|X)$. Then, the following rules are employed to decide which postcode is optimal result for the input pattern X.

Rule 1:
If $\exists D \in \Omega$, $p(D|X) = \max_{D\in\Omega} p(D|X)$ and $p(D|X) > \alpha$.
Then $X = D$,
where α is a threshold for making the trade-off between the rejection and substitution.

Rule 2:
If $\exists D \in \Omega$, $p(D|X) = \max_{D\in\Omega} p(D|X)$
$\exists D' \in \Omega$, $p(D'|X) = \max_{D'\in\Omega-D} p(D'|X)$ and
$p(D|X) - p(D'|X) > \beta$
Then $X = D$,
where β is constant.

The result returned by each classifier consists of M ranked confidence levels for each digit position. As a result, the score of all M^N (1,000,000 for Chinese postal system) possible results should be computed and ranked in this strategy. But as a matter of fact, valid postcodes in Ω are far less than M^N and most of the scores are so low that they can be eliminated from further consideration. In order to save the computing time, only the top three ranked score levels $s(d_n|x_n)$ of each postcode numeral were utilized to hypothesize and rank the postcode. From this, the first 3 ranked choices at each numeral position were considered, and thus a total of 729 possible postcodes was taken into account and ranked.

5. Experimental results

Experiments have been conducted to combine the five individual classifiers described earlier. For the experimental images, 113,467 mail pieces containing 680,802 handwritten postal numerals were captured from live Chinese mail pieces at the site of Hangzhou Postal Mail Processing Center while the mails were passing the automatic mail sorting machine. The resolution of digitization is 200DPI. Some samples used in the experiment are shown in Fig. 4. Six numeral images segmented from each mail piece image make up a group. 370,026 numeral images coming from 61,671 mail pieces are selected to estimate the recognition ability of the classifiers whose outputs are at the abstract level, i.e. the classifier TCTP and CP1. The remainders are utilized to test the performance of the multiple classifier combination. The samples that were used to design the classifiers are not included in the above images.

As described earlier, to evaluate the recognition performance of a practical system, both the correct recognition rate and the recognition reliability are crucial. A tradeoff between them is necessary and unavoidable. In other word, rejecting some recognition results with low likelihood outputs is a reasonable way to make the system reach an acceptable recognition reliability in the view of recognition quality. Understandably, with the reject rate increasing gradually, the recognition reliability increases generally, but the correct recognition rate decreases simultaneously.

We define an evaluation function to estimate the recognition performance of different methods so that they can be compared in terms of a single rating, in which both the correct recognition rate R_c and the recognition reliability Rel are concerned. The evaluation function that gives equal weights

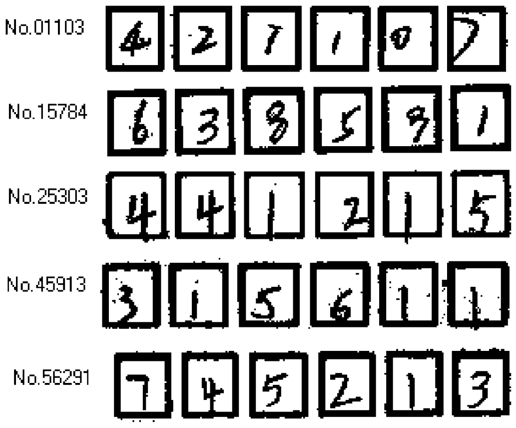

Fig. 4. Postcode images captured from live mail pieces (each mail includes 6 postal numerals).

to R_c and Rel, is defined as

$$\mu(R_c, Rel) = g(R_c)h(Rel), \quad (11)$$

where $g(R_c) = R_c$, and

$$h(Rel) = \begin{cases} Rel & \text{if } Rel > \delta, \\ 0 & \text{else.} \end{cases} \quad (12)$$

Here δ expresses the pre-defined acceptable recognition reliability.

5.1. Recognition performance of individual classifiers

In order to examine the recognition ability of individual classifier, a reasonable threshold is set for the classifier whose output is given in the format of measurement to reject the input characters if their measurement outputs have low likelihood, so that the classifier can achieve a maximum value of evaluation function $\mu(R_c, Rel)$. Table 2 tabulates the observed recognition performance at both the single character level and the postcode level of each classifier.

In the procedure of estimating the recognition ability of each classifier at the single character level, all of the correctly recognized characters are counted in, no matter the postcode composed by this character is legal or illegal in the postcode dictionary. This is the same as the commonly adopted approach for evaluating the single character recognition ability of classifiers. On the other hand, a postcode dictionary (without probabilistic property of each postcode) was invoked in the experiments while estimating the postcode recognition performance of each classifier. It means that the postcode is treated as being rejected if the postcode composed by the six numerals is illegal in the postcode dictionary, which may be caused by one or more erroneous recognition within the six numerals.

It can be found from the performance of the five individual classifiers that there are no significant differences in recognition performance among different classifiers. Generally speaking, a classifier achieves higher recognition rate at the expense of higher error rate. The reverse is also true. Furthermore, the classifier optimizing at the single character level does not necessarily result in the optimal performance at the postcode level. For example, the classifier TCTF achieves the highest value of evaluation function at the single character level (0.95082); while the classifier CP2 achieves the highest value of evaluation function at the postcode level (0.71557).

5.2. Comparison of the proposed approach with the other common methods

To evaluate the effect of the prior knowledge of postcodes on the recognition ability, the performance of the proposed approach combining the five classifiers at the postcode level and using the postcode dictionary as a priori is tested. Two experiments are carried out according to the different property of postcodes in the postcode dictionary. In the first method, the postcode dictionary without the probabilistic property of postcodes is utilized in the five classifier combination procedure (BPD method). Here, the postcode dictionary provides the $f_0(D)$ of each postcode D with 1 or 0 for valid or invalid postcode, respectively. In the second method, the postcode dictionary with the probabilistic property of postcodes is utilized in the five classifier combination procedure (PPD method). In this case, the $f_0(D)$ of each postcode D is its frequency of occurrence. The other two approaches of combining multiple classifiers, namely major voting method and Bayes method [7], are experimented as

Table 2
Performance of five independent classifiers

Classifier	Single character level					Postcode level				
	R_c (%)	R_r (%)	R_e (%)	Rel	$\mu(R_c, Rel)$	R_c (%)	R_r (%)	R_e (%)	Rel	$\mu(R_c, Rel)$
MLP	95.36	3.78	0.86	0.99104	0.94506	72.44	26.56	1.00	0.98638	0.71453
TCTF	96.41	2.24	1.35	0.98623	0.95082	72.16	26.60	1.24	0.98316	0.70945
HAVNET	94.76	4.25	0.99	0.98962	0.93776	71.87	26.96	1.17	0.98402	0.70722
CP1	93.28	6.28	0.43	0.99538	0.92849	71.65	27.16	1.19	0.98365	0.70479
CP2	95.83	2.54	1.63	0.98326	0.94226	73.06	25.41	1.53	0.97943	0.71557

Table 3
Performance comparison of combination approaches

Approach	Single character level					Postcode level				
	R_c (%)	R_r (%)	R_e (%)	Rel	$\mu(R_c, Rel)$	R_c (%)	R_r (%)	R_e (%)	Rel	$\mu(R_c, Rel)$
Major voting	95.73	4.10	0.17	0.99823	0.95561	73.82	25.18	0.99	0.98673	0.72840
Bayes method	96.91	2.86	0.23	0.99764	0.96681	74.58	24.08	1.33	0.98246	0.73272
BPD method	95.62	4.22	0.16	0.99825	0.95453	74.89	24.12	0.99	0.98691	0.73685
PPD method	96.78	3.04	0.18	0.99816	0.96602	79.73	19.15	1.12	0.98616	0.78627

comparison. The recognition performance achieved by each combination approach is tabulated in Table 3.

To estimate the recognition performance of BPD and PPD, the threshold α in Rule 1 is adjusted to make sure that the system can achieve a maximum value of evaluation function $\mu(R_c, Rel)$. Because the BPD method and PPD output the postcode results directly, they are not the character recognition results as "semi-finished product". In order to compare their recognition performance at the single character level with the other two methods, we measure the performance by counting in every single character in all of the recognized postcodes.

Comparing the experimental results in Tables 2 and 3, we can find that the recognition performance achieved by the combination of multiple classifiers is better than that achieved by a single classifier, at either the single character level or the postcode level. Any of combination methods achieves a higher value of evaluation function than each of individual classifiers.

Furthermore, it can be seen from the experimental results that the recognition performance at the single character level achieved by each method is not much different from each other. The recognition performance at the postcode level achieved by BPD method improves a little in comparison with the major voting method and Bayes method, but not much. However, the PPD method outperforms all of them with about 5% improvement of recognition rate at the postcode level.

5.3. *Sorting performance of particular postcodes*

Furthermore, we test the sorting performance of particular bins with respect to the postcodes with low frequency of occurrence, which can be improved by the proposed approach PPD.

The observations discover that the mis-sorting rate of each bin with respect to a different postcode is not the same. Although the overall mis-sorting rate of the system is acceptable, the mis-sorting rates of a few bins reach so high that the "sorted" mails in these bins fall short of the target for correct sorting. In general, these bins belong to a few postcodes that have quite low frequencies of occurrence.

Table 4
Recognition performance of particular postcodes by the BPD method

	Postcode 310428	Postcode 316428	Overall
Total	2713	103	2816
Correct recognition	1965(72.43%)	74(71.84%)	2039(72.05%)
Erroneous recognition	20(0.74%)	2(1.94%)	22(0.78%)
Reliability	0.9899	0.9737	0.9893
Reason of mistake	310428 ⇒ 316428	316428 ⇒ 310428	

Table 5
Sorting performance of particular bins

	Bin of "310428"	Bin of "316428"	Overall
Correct sorting	1965	74	2039
Mis-sorting	2	20	22
Sorting reliability	0.9989	0.7872	0.9893
Reason of mistake	316428 ⇒ 310428	310428 ⇒ 316428	

Table 4 shows the recognition results of postcode "310428" and "316428" by the BPD method without the application of probabilistic postcode dictionary, in which we can find that the recognition performance of each postcode is acceptable. However, the 20 erroneously recognized mail pieces with postcode "310428" will be mis-sorted to the bin with respect to the postcode "316428", as shown in Table 5 which is a re-arrangement of Table 4 from the point of view of the sorting performance with respect to bins rather than the recognition performance with respect to postcodes. As a result, the mis-sorting rate for the bin with respect to the postcode "316428" is so high and the sorting reliability is so low that sorting result is absolutely unacceptable. It is caused by the fact that the volume of the processed mails with postcode "310428" is more than that of postcode "316428", as the postcode "310428" is more common in the mail-stream, although the error rates of misinterpreting "0" as "6" and vice versa are about the same.

By applying the probabilistic postcode dictionary (i.e. PPD method), the sorting performance of the bins with

Table 6
Improved sorting performance of particular postcode bins

	Bin of “310428”	Bin of “316428”	Overall
Correct sorting	1974	56	2030
Mis-sorting	3	2	5
Sorting reliability	0.9985	0.9655	0.9975
Reason of mistake	316428 ⇒ 310428	310428 ⇒ 316428	

respect to the postcode “310428” and the postcode “316428” is tabulated in Table 6. It can be found from the table that the mis-sorting rate of the bin of “316428” has significantly decreased and the corresponding sorting reliability has greatly increased, although the sorting reliability of the bin with respect to the postcode “310428” has decreased a little. This improvement is obviously the consequence of the application of the probabilistic postcode dictionary in the combination of multiple classifiers.

We have raised the sorting reliability of the bin of “316428” at the expense of lowering its sorting rate. However, because the mail pieces with the postcode “316428” are small in number, the rejected pieces can be later sorted manually instead of mis-directing these mails to wrong bins.

6. Conclusions

It is known that a priori knowledge plays a very important role in the issue of developing a practical character recognition system. The knowledge, such as the postcode dictionary and the frequency of occurrence of postcodes, has been ignored generally in the system design in the task of postcode recognition. With the motivation of optimizing the recognition performance of postcode rather than that of single characters, we have presented an approach to combine multiple classifiers in such a way that the combination decision is carried out at the postcode level rather than at the single character level, in which a probabilistic postcode dictionary is utilized as well to improve the postcode recognition ability. Its application to Chinese postal numeral recognition demonstrated that the proposed approach is effective and practicable for optimizing the performance of postcode recognition, which produced an overall performance improvement in comparison with the commonly used methods that combines the multiple classifiers at the single character level. Furthermore, the sorting performance of some special bins with respect to the postcodes with low frequency of occurrence can be improved significantly, which obviously benefits from the application of probabilistic postcode dictionary.

Although the proposed approach has been applied to a specified task of handwritten postcode recognition, the underlying concepts can be completely generalized and should be applicable to many other related task domains.

Acknowledgements

The authors would like to thank the anonymous reviewers for their constructive and valuable comments and suggestions that have been of great use in the revision process.

References

[1] S. Mori, C.Y. Suen, K. Yamamoto, Historical review of OCR research and development, Proc. IEEE 80 (7) (1992) 1029–1058.

[2] C.Y. Suen, J. Kim, K. Kim, Q. Xu, L. Lam, Handwriting recognition—the last fouriers, Proceedings of the International Conference on Pattern Recognition, Barcelona, Spain, 2000, Vol. 4, pp. 1–10.

[3] P.D. Gadar, Pipelined system for recognition of handwritten digits in USPS zip codes, Proceedings of the U.S. Postal Service Advanced Technology Conference, Washington, DC, 1992, pp. 539–547.

[4] J. Cao, M. Ahmadi, M. Shridhar, Recognition of handwritten numerals with multiple feature and multistage classifier, Pattern Recognition 28 (2) (1995) 153–160.

[5] Y.S. Huang, K. Liu, C.Y. Suen, The combination of multiple classifiers by a neural network approach, Int. J. Pattern Recognition Artif. Intell. 19 (3) (1995) 579–597.

[6] L. Lam, C.Y. Suen, Application of majority voting to pattern recognition: an analysis of its behavior and performance, IEEE Trans. Systems Man Cybern. Part A 27 (5) (1997) 553–568.

[7] L. Xu, A. Krzyzak, C.Y. Suen, Methods of combining multiple classifiers and their applications to handwriting recognition, IEEE Trans. Systems Man Cybern. 22 (3) (1992) 418–435.

[8] A.F.R. Rahman, M.C. Fairhurst, An evaluation of multi-expert configurations for the recognition of handwritten numerals, Pattern Recognition 31 (9) (1998) 1255–1273.

[9] J. Kittler, M. Hatef, R. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.

[10] T.K. Ho, J.J. Hull, S.N. Srihari, Decision combination in multiple classifier systems, IEEE Trans. Pattern Anal. Mach. Intell. 16 (1) (1994) 66–75.

[11] C.Y. Suen, C. Nadal, R. Legault, T.A. Mai, L. Lam, Computer recognition of unconstrained handwritten numerals, Proc. IEEE 80 (7) (1992) 1162–1180.

[12] Y.S. Huang, C.Y. Suen, A method of combining multiple experts for the recognition of unconstrained handwritten numerals, IEEE Trans. Pattern Anal. Mach. Intell. 17 (1) (1995) 90–94.

[13] K. Gyeonghwan, V. Govindaraju, A lexicon driven approach to handwritten word recognition for real-time applications, IEEE Trans. Pattern Anal. Mach. Intell. 19 (4) (1997) 366–379.

[14] S.H. Kim, S. Jeong, C.Y. Suen, A lexicon driven approach for off-line recognition of unconstrained handwritten Korean words, Proceedings of the Fifth International Conference on Document Analysis and Recognition, Bangalore, India, 1999, pp. 507–510.

[15] A. Kundu, Handwritten word recognition using hidden Markov model, in: H. Bunke, P.S.P. Wang (Eds.), Handbook of Character Recognition and Document Image Analysis, World Scientific, Singapore, 1997, pp. 157–182.

[16] D. Cheng, H. Yan, Recognition of handwritten digits based on contour information, Pattern Recognition 31 (3) (1998) 235–255.

[17] R.G. Rosandich, HAVNET: a new neural network architecture for pattern recognition, Neural Networks 10 (1) (1997) 139–151.

[18] X. Wu, P. Shi, Unconstrained handwritten numeral recognition using Hausdorff distance and multi-layer neural network classifier, Proceedings of the Fifth International Conference on Document Analysis and Recognition, Bangalore, India, 1999, pp. 249–252.

[19] B. Achermann, H. Bunke, Combination of face classifiers for person identification, Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria, Vol. C, 1996, pp. 416–420.

[20] K. Yu, X. Jiang, H. Bunke, Lipreading: a classifier combination approach, Pattern Recognition Lett. 18 (1997) 1421–1426.

About the Author—YUE LU is a research fellow at the Department of Computer Science, National University of Singapore. He received his B.E and M.E degree in Telecommunications and Electronic Engineering from Zhejiang University, Zhejiang, China in 1990, 1993, respectively, and his Ph.D degree in Pattern Recognition and Intelligence System from Shanghai Jiao Tong University, Shanghai, China in 2000. His research interests include document image processing, character recognition, intelligence system and computer vision.

About the Author—CHEW LIM TAN is an associate professor in computer science at the School of Computing, National University of Singapore. His research interests are computer vision, document image analysis, intelligent text processing and neural networks. He obtained a B.Sc. (Hons) degree in Physics in 1971 from the University of Singapore, an M.Sc. degree in Radiation Studies in 1973 from the University of Surrey in UK, and a Ph.D. degree in Computer Science in 1986 from the University of Virginia, USA.