We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK
CITATION
INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

A Classifier Fusion System with Verification Module for Improving Recognition Reliability

Ping Zhang
Department of Mathematics and Computer Science
Department of Advanced Technologies
Alcorn State University
USA

1. Introduction

Recognition reliability is a vital and sensitive issue in the pattern recognition applications. Recognition with a proper rejection option provides a means to reduce the error rate and to increase recognition reliability (Chow, 1970). It deals with how research projects can be developed into real applications. Relatively few research papers on this topic have been reported in literature (Frelicot & Mascarilla, 2002). In order to clearly explain the recognition reliability problem, the Recognition Rate (RR), Mis-Recognition Rate (MR), Rejection Rate (RJ), and Reliability (Re) are defined and their relationships are analyzed as follows:

The Recognition Rate (RR) is defined as:

$$RR = \frac{Number\ of\ correctly\ recognized\ objects}{Total\ number\ of\ testing\ objects} \tag{1}$$

The Misrecognition Rate (MR) is presented as:

$$MR = \frac{Number\ of\ misrecognized\ objects}{Total\ number\ of\ testing\ objects} \tag{2}$$

The Rejection Rate (RJ) is written as:

$$RJ = \frac{Number\ of\ rejected\ objects}{Total\ number\ of\ testing\ objects} \tag{3}$$

The Reliability (*RE*) can be denoted as:

$$RE = \frac{Total\ number\ of\ testing\ objects - Number\ of\ misrecognized\ objects}{Total\ number\ of\ testing\ objects} \quad (4)$$

The reliability can also be deduced as:

$$RE = 1 - MR = RJ + RR \quad (5)$$

Technically speaking, there is a tradeoff for a recognition system to pursue a very high recognition rate and a very low misrecognition rate at the same time given a specific set of feature set and a classifier or a combination of classifiers (Chow, 1970). It is common sense that setting a relatively low threshold for a recognition system can achieve a high recognition rate; however, it also introduces more misrecognitions.

In many pattern recognition systems, the goal is to seek the highest reliability and the highest recognition rate as possible at the same time. In other words, the misrecognition rate must be suppressed. When designing a sensitive object recognition system, it is preferred rejecting objects with low confidences over mistakenly recognizing the objects (Zimmermann, Bertolami & Bunke, 2002). For example, in an automatic bank check processing system, a very high recognition reliability is a vital criterion. The misrecognition is absolutely forbidden and a small percentage of rejection is allowed. The rejected checks can be sent for manually handling.

An automatic bank check processing system can be divided into the following aspects: 1) magnetic ink character recognition (MICR); 2) handwritten legal amount recognition (English character recognition, or other language character recognition); 3) handwritten courtesy amount recognition (handwritten digital recognition); 4) payer's signature verification or recognition; 5) the recognition of name and address of a payer, etc.

Among the above mentioned recognitions, MICR plays an important role in the automatic bank check processing system based on the following reasons:

a) Information in the MICR area includes the account number of a payer and bank identification number; both of which need to be firstly recognized in order to verify the payer's identification and the payer's bank number while transition is processed;

b) Individual character recognition rate in the MICR area is very high (over 99%), it is possible to use an automatic process.

However, some errors have been reported in the bank applications due to the following reasons: the mechanical deficiency of MICR scanners; the distortion of the printed characters in the MICR area, and others.

In this paper, we will propose a novel classifier fusion system with a verification module to improve system's reliability. The arrangement of the paper is as follows: In Section 2, the concept of MICR is briefly introduced, which includes one dimensional MICR waveform process and two dimensional image process. The flowchart of the new classifier fusion system with a verification module is proposed in the section 3; In Section 4, the basic concepts of classifiers: Artificial Neural Networks (ANNs), modified K-Nearest Neighbor (KNN) classifier and Support Vector Machines (SVMs) are reviewed. A gating network for congregating the outputs of ANN and KNN is applied to the classifier fusion system. Three

experiments conducted on the MICR character recognition are reported, and the recognition reliability is analyzed in the section 5. Finally, the conclusion ends this paper.

2. MICR Information Processing

Fig. 1 shows a blank bank check image. The MICR area is located at the bottom of the check. In North America, E13B font has been officially used as the printed fonts in the MICR area for almost commercial banks. E13B symbols include ten numerals and four symbols: On-Us symbol, Transit symbol, Amount symbol, and Dash symbol.
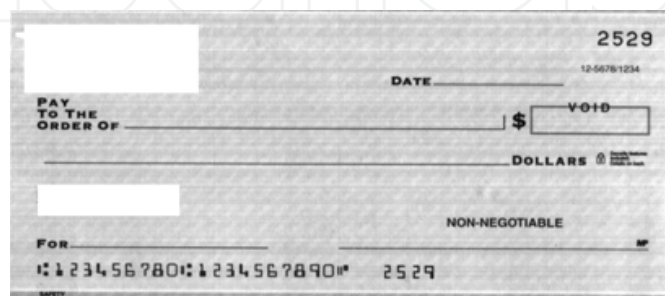

Fig. 1. MICR area at the bottom of check

MICR characters are printed with a magnetic ink or toner. A specially designed MICR scanner can read not only the E13B character waveforms (one dimensional signals), but also the characters' images with different resolutions (two dimensional images). The standard font images and their waveforms of the fourteen characters are shown in Fig. 2.
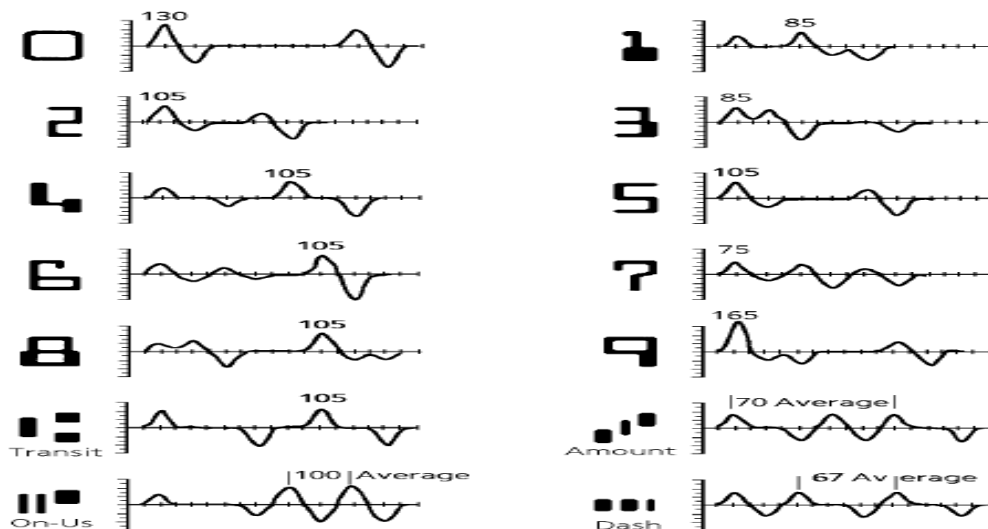

Fig. 2. E13B MICR fonts and their waveforms

2.1 One Dimensional MICR Waveform
One dimensional MICR waveform processing and recognition has been extensively researched for a long time. The very high recognition rate was reported under an ideal condition. However, there are still some recognition errors and rejections reported in the commercial applications due to unstable paper feeding mechanism of scanners, the

distoration of printed MICR characters on the checks and other factors, which lead to waveform distortions and noise on MICR images. Fig. 3 shows two MICR waveforms scanned from two personal bank checks: one has two sub-MICR areas; another has four sub-MICR areas.

A waveform segmentation algorithm can segment each sub-MICR waveform into individual waveforms, each representing one character. The detail algorithm for waveform segmentation is beyond the scope of this paper.

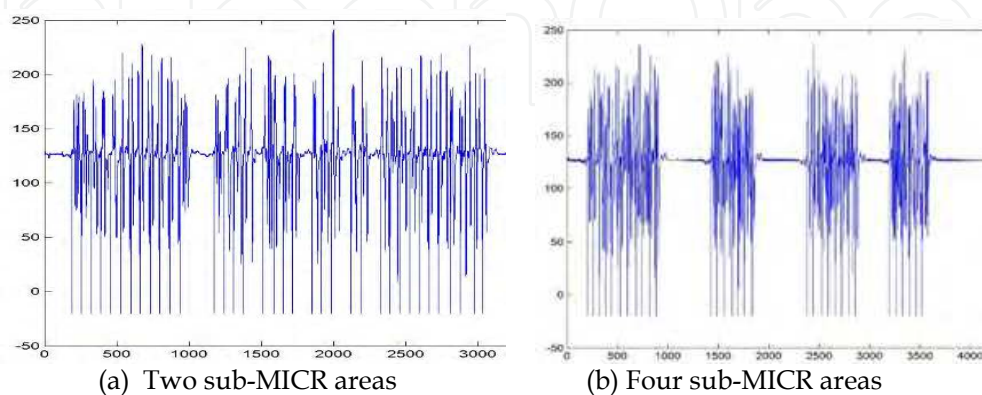(a) Two sub-MICR areas   (b) Four sub-MICR areas

Fig. 3. MICR waveform

2.2 Image-based Character Segmentation

For the recognition of image-based MICR characters, the key issue is how to deal with image segmentation, noise removal, image enhancement, feature extraction, and the design of classifiers for recognition. The image-based character image segmentation in the MICR area can be divided into two steps: 1) locating top and bottom lines of the characters; 2) segmenting each character from the vertical direction.

In order to accurately locate the top and the bottom lines of the MICR characters in the check images, a fast algorithm is proposed as follows: scanning each line in the horizontal direction, counting the number of zero-crossing points in each horizontal line as NC; If the number of characters in the MICR area is N, then following conditions apply to locate the top or the bottom lines of MICR characters:

Condition 1: If $NC >= 2*N$, then the line likely belongs to MICR character area.

Condition 2: Beginning at the Lth line of the MICR image, then moving downwards, if there are a few consecutive lines satisfy condition 1, then the Lth line is the top line of MICR character area. The same method is applied to locate the bottom line of MICR character area.

Condition 3: if a check is scanned with certain angle w, the line seeking algorithm presented in Condition 2 is also traced with this angle.

As soon as the top line and the bottom line of the MICR character area are located, the characters are segmented based on character's connectivity and vertical profiles. As to four special symbols, each symbol consists of three black blocks. The following criteria are used to combine three black blocks as a symbol:

a) The distance between two adjacent black blocks is shorter than the distance between two characters;

b) The length and the width of any black block of the four symbols are shorter than that of the ten numerals.

Fig. 4 (a) shows an MICR character image; Fig. 4 (b) shows the segmentation result of Fig. 4 (a). MICR area image with background is shown in Fig. 5 (a) and the segmented image is shown in Fig. 5 (b)
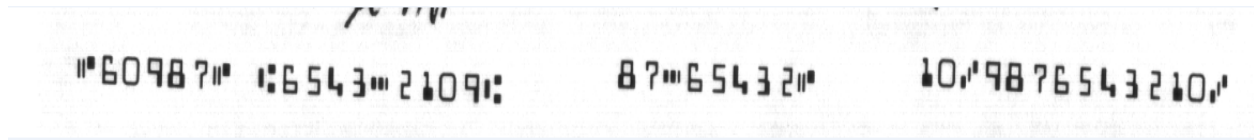
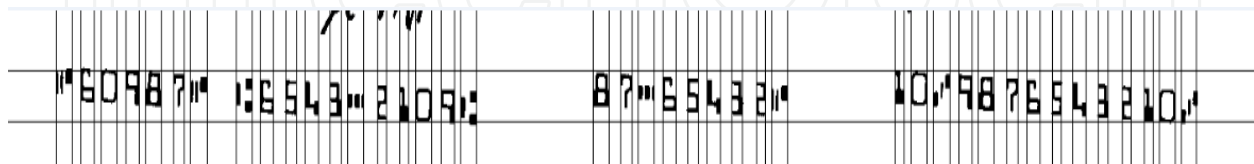Fig. 4 (a) Original MICR character image
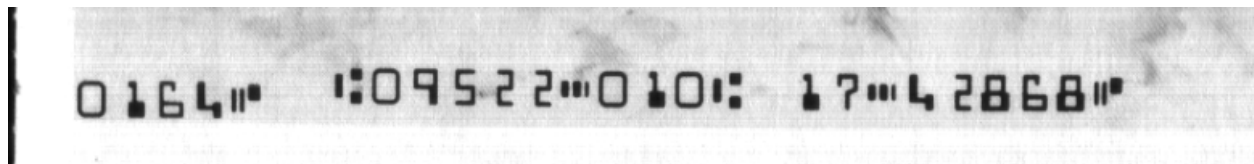
Fig. 4. (b) Segmentation result of Fig. 4 (a)
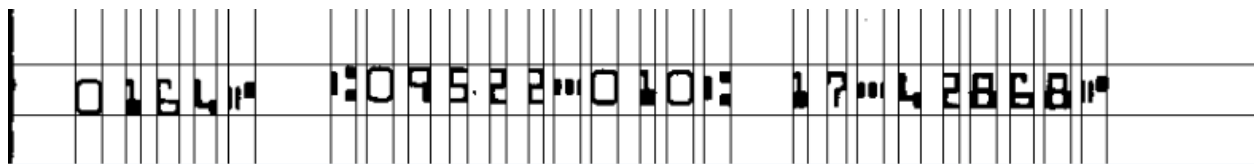
Fig. 5. (a) MICR character image with background

Fig. 5. (b) Segmented image of Fig. 5 (a)

3. Classifier Fusion System with SVM Verification Module

It is difficult for a single classifier to obtain a very high reliability and recognition rate at the same time for a complex pattern recognition system. Some theoretical advancements have been proposed in literature (Kuncheva, 2002; Kittler, Hatef, Duin & Matas, 1998). There are a few possible solutions to help reduce the number of errors. One solution is to employ a verification module. Another solution is to use a combination of multiple classifiers (Suen & Tan, 2005). The different features extracted by different means, which are inputted to different ensemble classifiers for classification, have different merits for recognition because some of the features are complementary (Zhang, Bui. & Suen, 2007). It is reasonable to combine two classifiers to produce a higher reliability and at the same time to seek the lowest misrecognition rate. A classifier fusion system with SVM verification module is proposed and it is shown in Fig. 6.

In the proposed recognition and verification scheme, a classifier fusion system consists of an ANN classifier and a KNN classifier, which are trained by two sets of feature vectors respectively. As the two sets of feature vectors may be complementary, the trained ANN and KNN as recognizers have their merits on character recognition. Experiments will prove that the combination of two classifiers can achieve a higher recognition rate.

$$G=\{g_0, g_1, \ldots, g_{13}\}$$

```
         Verification
           Module
              ↑
              |
        Gating Network
         ↑    ↑    ↑
   W1,0~13    W2,0~13
      ↑          ↑
   ┌──────┐  ┌──────┐  ┌──────┐
   │ ANN  │  │ KNN  │  │ SVM  │
   │Class.I│  │Class.II│ │Class.III│
   └──────┘  └──────┘  └──────┘
```
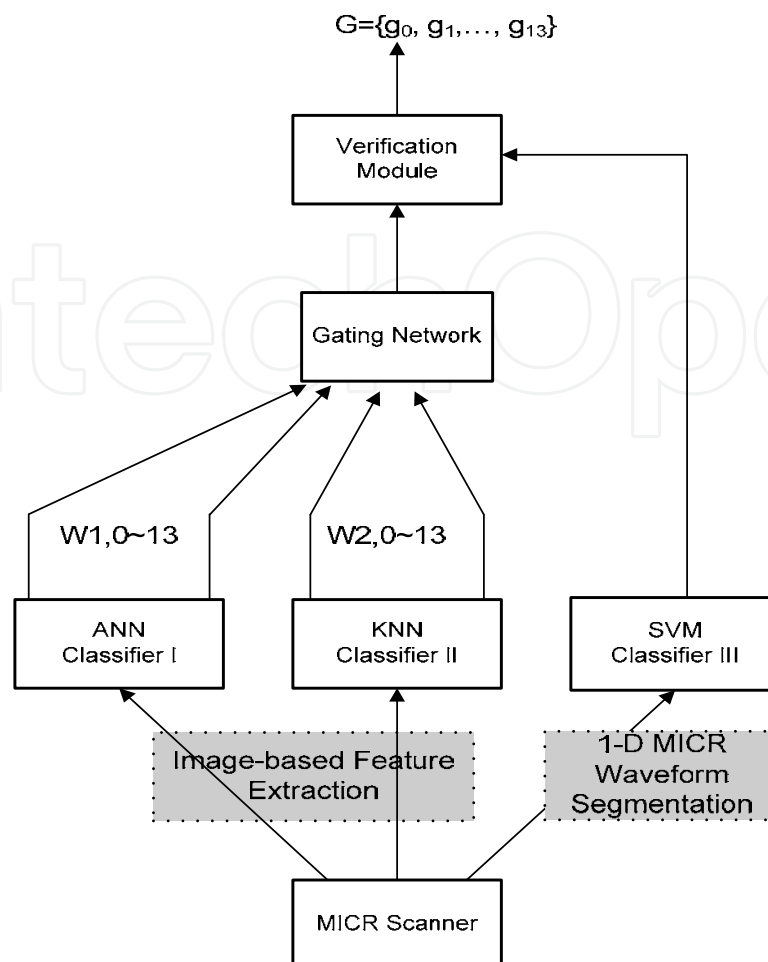
Fig. 6. Classifier Fusion system with SVM verification module

There are fourteen characters in the E13B character set. For the verification purpose, fourteen two-class SVMs are applied to classify the MICR waveforms. the result of SVMs is used to verify the image-based recognition result. The detail recognition and verification process will be elaborated in Section 5.

4. Classifier Design and Feature Extraction

4.1 Artificial Neural Network Classifier

An ANN is an interconnected group of artificial neurons (Duda, Hart & Stork, 2000). ANN refers to electrical, mechanical, or computational simulations or models of biological neural networks. One of the most popular methods for training a multilayer network is based on the gradient descent principle using the back-propagation algorithm or generalized delta rule. The principle is a natural extension of the Least Mean Squares (LMS) algorithm because it is powerful, useful, and relatively easy to understand and implement.

An ANN classifier consists of input units, hidden units, and output units. In terms of classifying fourteen numerals and symbols in this research, the ANN will have fourteen output units. The signal from each output unit is the discriminant function $g_k(x)$. The discriminant function can be expressed as:

$$g_k(x) \equiv z_k = f\left(\sum_{j=1}^{r} w_{kj} f\left(\sum_{i=1}^{d} w_{ji} x_i + w_{j0}\right) + w_{k0}\right) \qquad (6)$$

where x_i is a feature component; w_{ji} is a weight between the input layer and the hidden layer; w_{kj} is a weight between the hidden layer and the output layer; $i=1,…,d$, and d is the number of nodes in the input node; $j=1,…, r$, and r is the number of nodes in the hidden layer; $k=0,1,2,…, m$, which represents the number of nodes in the outputs layer. For example, fourteen nodes of outputs represent ten digits and four special symbols used in this paper.

Thus, the discriminant function can be implemented by a three-layer neural network. The configuration of the three-layer neural network for the recognition is drawn in Fig. 7.

We now turn to the crucial problem of setting the weights based on training patterns and the desired output. Backpropagation algorithm is used to train the classifier. Some of considerations in the training and testing procedures are listed as follows:

Target Values: The target value (the desired output) of the output category is chosen as +1, while others are set equal to 0.0.

Number of Nodes in the ANN: According to a convenient rule of thumb, the total number of weights in the net is roughly chosen as $n/10 \sim n/4$. Here n is the number of training samples.

Initializing Weights: Random data are generated for all weights in the range of -1.0< all weights <+1.0.

Learning Rates: In general, the learning rate is small enough to ensure convergence. A learning rate of (0.1-0.4) is often adequate as a first choice.

Training Different Patterns: We used the following strategies to train the classifiers: our training procedure concentrates on the "difficult" patterns. Firstly, an ANN classifier is trained on all training samples, then the same set of training samples are fed into the ANN for testing. Those "difficult" patterns, which are not correctly recognized, are copied several times and randomly put into the training set for training again. As more "difficult" patterns are in the training set, the ANN can adaptively learn how to correctly recognize those "difficult" patterns without losing its generality.
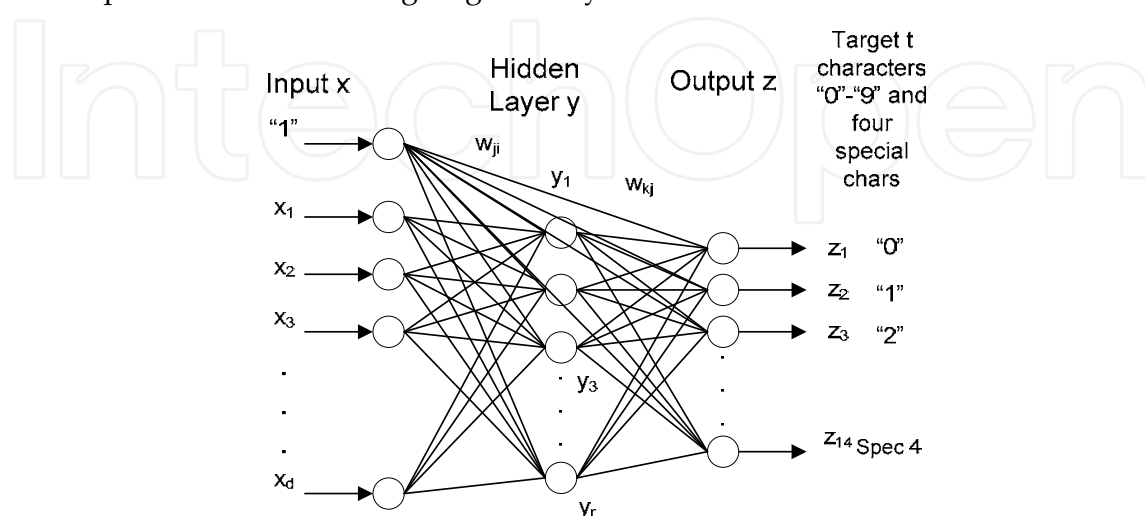

Fig. 7. Configuration of three-layer neural networks

4.2 KNN Classifier

In a KNN classifier, for each testing sample, the Euclidean distance between the testing sample and all the training samples are calculated. Let the testing sample x_i be represented by the feature vector $[x_1^i, x_2^i, x_3^i, \ldots, x_N^i]^T$, where x_k^i denotes the value of the *kth* feature component in the *ith* sample. The distance between x_i and x_j can be calculated by

$$d(x_i - x_j) = \sqrt{\sum_{k=1}^{N}(x_k^i - x_k^j)^2} \tag{7}$$

If the number of training data is N, then N distances will be identified as neighbors. If $K=1$, then the class label of the testing sample is equal to the closest training data. If $K>1$, then the class label of the testing sample is equal to the class label that most of the neighbors belong. The output of the KNN algorithm can be interpreted as a posteriori probability. Hence, instead of labeling the output class label equal to the class label that most of the neighbors have, we assign the following class confidence values of x:

$$p_c(x) = (no. \ of \ neighbors \ with \ class \ label \ c)/K \tag{8}$$

Here, p_c is the a posteriori probability that x belongs to the class c; K denotes the number of nearest neighbors. We can assign class label j to the testing sample x when

$$p_j(x) = \max_{1,2,\ldots,M}\{p_c(x)\} \tag{9}$$

Here M is the total number of classes.

One improvement to the KNN algorithm is to weigh the contribution of each of the K neighbors based on its distance to the testing sample. The closest neighbor should receive the highest weight. It can be represented by modifying the equation into following:

$$p_c(x) = \sum_{i=1}^{K}\left(\frac{\frac{1}{d(x,x_i)^2}}{\sum_{j=1}^{K}\frac{1}{d(x,x_j)^2}}\right) \tag{10}$$

The equation can be normalized as:

$$\sum_{c=1}^{M} p_c(x) = 1$$

The KNN algorithm with this refinement is also known as the fuzzy K-nearest neighbor algorithm (Keller, Gray and Givens, 1985). The normalized $p_c(x)$ can be used as input of gating network indicated in Fig. 6.

4.3 SVM Classifier

SVMs rely on the preprocessing the data to represent patterns in a higher dimension by an appropriate nonlinear mapping $\varphi(.)$. Data from two categories can always be separated by a

hyperplane. The detail theory can be referred to the references (Decoste & Scholkopf, 2002; Heisele, Serre, Prentice & Poggio, 2003).
In this research, the kernel function is a Gaussian radial basis kernel:

$$K(x,z) = \exp(- \| x - z \|^2 / \sigma^2)$$

(11)

Training a support vector machine for the pattern recognition problem leads to the following quadratic optimization problem:

$$W(\alpha) = -\sum_{i=1}^{l} \alpha_i + \tfrac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j k(x_i, x_j)$$

(12)

and subject to:

$$\sum_{i=1}^{l} y_i \alpha_i = 0$$
$$\forall i : 0 \leq \alpha_i \leq C$$

(13)

The number of training examples is denoted by l, α is a vector of l variables. Each component α_i corresponds to a training examples (x_i, y_i). The solution is the vector α for which (12) is minimized and the constraints (13) are fulfilled.
SVM is a two-class classifier. For each testing sample, we compare the coefficients of fourteen classifiers and assign a class with maximum coefficient as overall recognition output. If the output matches the character's label, it means that the testing character is correctly recognized. Otherwise, this character is misrecognized. For example, SVM0 classifier can be employed to distinguish character 0 from all other characters. The overall recognition is congregated from all SVM classifiers.
Fig. 8 shows the flowchart of MICR waveform recognition using fourteen SVM classifiers.
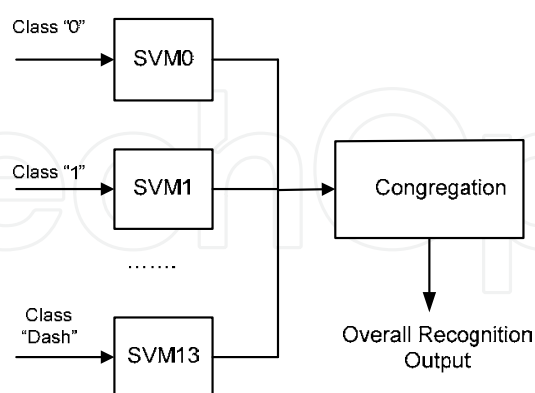

Fig. 8. Flowchart of MICR waveform recognition using SVMs

4.4 Feature Extraction
Feature extraction is a very important step for the image-based character recognition in the MICR area. Three feature extraction methods (Zhang, Bui & Suen, 2005) are used to construct two hybrid feature sets. The feature sets include: Directional-based Wavelet

Feature Set, Medial Axial Transformation (MAT) Gradient Feature Set, and Geometrical Feature Set.

4.5 Genetic Algorithm Used to Envolve Gating Network

A new combination scheme of classifiers is proposed in order to achieve the lowest error rate while pursuing the highest recognition rate for the recognition of E13B characters. The schematic diagram is shown in Fig. 6. The output confidence values of the ANN are weighted by $w_{1,0} \sim w_{1,13}$ and the output confidence values of the KNN classifier are weighted by $w_{2,0} \sim w_{2,13}$.

A genetic algorithm is used to evolve the optimal weights for the gating network from the confidence values of ANN classifier and KNN classifier.

Suppose the outputs of two classifiers are represented as: $\{c_{1,0}, c_{1,1}, \ldots, c_{1,13}\}$, $\{c_{2,0}, c_{2,1}, \ldots, c_{2,13}\}$.

The weighted outputs of the two classifiers' confidence values can be calculated as follows:

$$X_i = W_i^T . C_i \tag{14}$$

where $W_i = [w_{i,0}, w_{i,1}, \ldots, w_{i,13}]$, $C_i = [c_{i,0}, c_{i,1}, \ldots, c_{i,13}]$ $i=1,2$.

Add two weighted confidence values into a Y vector:

$$Y = \sum_{i=1}^{2} X_i \tag{15}$$

$$Y = [y_0, y_1, \ldots y_{13}]$$

In order to generalize the output, the j-th output g_j of the gating network is the "softmax" function of y_j as follows (Friedman, 1997):

$$g_j = \frac{e^{y_j}}{\sum_k e^{y_k}} \tag{16}$$

$$G = [g_0, g_1, \ldots, g_{13}]^T$$

G is the output of the gating network.

Our goal is to pursue a lowest misrecognition rate and at the same time to seek the highest recognition performance. We can create a vector O_{target} with fourteen elements, which represent ten numerals and four symbols of the E13B fonts. In the vector, the value of the corresponding label is set equal to 1.0, while others are set equal to 0.0. A fitness function f is chosen to minimize the difference between the output G and the corresponding training sample vector O_{target} as follows:

$$f = | G - O_{target} |^2 \tag{17}$$

By minimizing the equation (17) through a genetic evolution, the weights tend to be optimal. Then, the recognition criterion is set as follows:

A recognition result is accepted if one of the following three conditions is satisfied:

1) ANN classifier and KNN classifier vote for the same character at the same time, where the sum of the confidence values is equal to or larger than 1.6;

2) The gating network votes for a character, where the confidence value of the gating network is larger than 0.85;

3) The confidence value of any classifier is larger than 0.95 and the gating network votes for the same character;

Otherwise, the character is rejected.

4.6 Genetic Algorithms for Training Gating Network

Genetic Algorithms have been developed based on Darwinian evolution and natural selection for solving optimization problems. GA applies evolution-based optimization techniques of selection, mutation, and crossover to a population for computing an optimal solution (Siedlecki & Sklansky, 1989). The problem of the weight selection in the gating network is well suited to the evolution by GAs.

In the ANN training procedure, the most difficult problem is to find a reasonable fitness function for a large set of training samples. Ideally, the recognition rate can be used as a fitness criterion for training a classifier. However, using the recognition rate this way is unfeasible for some pattern recognition problems because it requires huge computations for each generation of learning.

In this paper, we use GAs to train the gating network. When equation (17) is used as the fitness function, the GAs pursue the smallest difference between the gating network's outputs and the target label vector O_{target}. The following is a description of the steps used for our genetic algorithms.

Chromosome Representation

There are an ANN classifier, a KNN classifier in the system. Each classifier's outputs have fourteen nodes. A chromosome is a vector consisting of 28 weights. A chromosome is presented as:

$[\ w_{1,0} \quad w_{1,1} \quad \ldots \quad w_{1,13} \quad w_{1,14} \quad w_{1,15} \ldots \quad w_{1,27} \quad w_{1,28}]$
|--14 weights for ANN--| |--14 weights for KNN--|

Population Initialization

The initial chromosomes P (48 populations), are randomly created (0.0~1.0).

Selection

The best 24 chromosomes with minimum fitness values, taken from 48 populations in each generation, are chosen to go into the mating pool.

Fitness Computation

Equations (17) are used to calculate the fitness function.

Crossover

Crossover occurs when information is exchanged between two parent chromosomes and the new information is introduced to child chromosomes. A single offspring parameter value, w_{new}, comes from a combination of the two corresponding parent parameter values. The

crossover begins by randomly selecting a parameter *a* in a pair of parents, which is a crossover point. The crossover is calculated as follows:

$$a = roundup\{random(M-1)\}$$

$$parent\ 1(mother\ ) = [w_{m0}, w_{m1}, w_{m2}, ... w_{ma}, ..., w_{mM-1}]$$

$$parent\ 2(father\ ) = [w_{d0}, w_{d1}, w_{d2}, ... w_{da}, ..., w_{dM-1}]$$

(18)

where *M* is the length of the weight vector. The subscripts *m* and *d* in the weight parameters (w_{mi}, w_{di}) represent the mother and the father in the mating pool. Then, the selected parameters are combined to form new parameters. Two new weights are calculated as follows:

$$w_{new\ 1} = w_{ma} - \beta[w_{ma} - w_{da}]$$

$$w_{new\ 2} = w_{da} + \beta[w_{ma} - w_{da}]$$

(19)

where β is a random value between 0.0 and 1.0. The next step is to exchange the right parts of two parents, consisting of the crossover point to the end for each parent.

$$offsprings\ 1 = [w_{m0,}w_{m1,}w_{m2,...,}w_{new1}, ..., w_{dM-1}]$$

$$offsprings\ 2 = [w_{d0,}w_{d1,}w_{d2,...,}w_{new2}, ..., w_{mM-1}]$$

(20)

Mutation

In our experiments, the mutation rate is set at 0.01. According to the mutation rate, we randomly replace w_{mi} (w_{di}) with a new weight element, which is produced by multiplying the old weight value with a new uniform random number (0.0-1.0).

Termination Criteria

Termination occurs when either the number of iterations reaches its defined number or the fitness value converges so that the weights in the chromosome pool are stable.

5. Experiments

In order to see how the proposed system can improve system's reliability, we conducted the following three experiments. In all of the experiments, the E13B characters extracted from 250 personal checks (6250 characters and symbols) are used as training samples; another set of 6250 characters and symbols is used as testing samples. The training samples and testing samples are separated.

Experiment One

In this experiment, two classifiers: ANN classifier and KNN classifier are individually used to test the recognition performance on E13B image-based characters. Two hybrid feature sets are used to train the two classifiers, respectively. Table I lists the rejection rate, recognition rates, and reliability results conducted on ANN classifier.

Feature Set	Rejection Rate (%)	Recognition Rate (%)	Reliability (%)
Hybrid Feature Set I	0.00	98.50	98.50
Hybrid Feature Set II	0.00	98.69	98.69

Table 1. Recognition performance of ANN classifier trained by two hybrid feature sets

Note: Hybrid Feature Set I: Directional-based Wavelet Feature+20 Geometrical Features
     Hybrid Feature Set II: MAT-based Gradient Feature + 20 Geometrical Features
We use the same training samples and testing samples for KNN classifier. The test result is shown in Table II.

Feature Set	Rejection rate (%)	Recognition Rate (%)	Reliability (%)
Hybrid Feature Set I	0.00	98.40	98.40
Hybrid Feature Set II	0.00	98.59	98.59

Table 2. Recognition performance of KNN classifier trained by two hybrid feature sets

From above tests, it can be concluded as follows:
1) There is no rejection rate since an individul classifier sets a threshold to either correctly recognize a testing character or mistakenly recognize it;
2) Two classifiers have a similiar recogntion rate trained by two feature sets;
3) Reliability is relatively low.

Experiment Two
In Experiment Two, a classifier fusion scheme, which consists of an ANN classifier, a KNN classifier, and a gating network to congregate the two classifiers, is tested without SVM verification module. Different feature sets are applied to train two classifiers. There are four options:
Combo I: ANN trained by Hybrid Feature Set I+KNN trained by Hybrid Feature Set I+gating network
Combo II: ANN trained by Hybrid Feature Set I+KNN trained by Hybrid Feature Set II+gating network
Combo III: ANN trained by Hybrid Feature Set II+KNN trained by Hybrid Feature Set I+gating network
Combo IV: ANN trained by Hybrid Feature Set II+KNN trained by Hybrid Feature Set II+gating network
The recognition rate, rejection rate, and reliability are conducted on the four fusion schemes. The test results are listed in Table III. Since the classifier fusion system introduces a rejection option, some characters with a relative low confidence value are rejected. The rule to reject characters in the classifier fusion system was described in the last part of Section 4.5.

Classifier Fusion Scheme	Rejection Rate (%)	Recognition Rate (%)	Reliability (%)
Combo I	0.69	98.60	99.29
Combo II	0.72	98.70	99.42
Combo III	0.80	98.67	99.47
Combo IV	0.72	98.64	99.36

Table 3. Recognition performance of classifier fusion system, consisting of ANN, KNN and gating network, traned by different hybrid feature sets

It is observed that some checks have been severely folded. As a result, character images are deteriorated and noises are added on the check images, which affected recognition rate.

Although a rejection strategy is introduced in the tests and the reliability is increased; however, some recognition errors still remain.

For example, in the Combo III experiment, the recognition rate is 98.67%. The rejection rate is 0.80% and the reliability is as high as 99.47%. As such, there are still 33 misrecognized characters, which is unacceptable in an automatic bank check processing system.

Experiment Three

In order to pursue an execellent reliability in the system, we propose a verification module. Firstly, SVM classifers are used to recognize the segmented one dimensional MICR waveforms. Then, the recognition results are used to verify the recognition results of the classifier fusion system. The recognition rate of waveform-based MICR is 99.52%, which means that 30 characters out of 6,250 testing samples were misrecognized. The reliability of the SVMs is also 99.52%.

The verification rule is explained as follows: if the classifier fusion system votes a character and SVMs also vote the same character, the recognition is confirmed; otherwise, the testing sample is rejected.

Since two classifications use entirely different input sginals (the classifier fusion system uses image-based OCR method, whereas the SVM classifier uses one dimensional MICR waveform), the overall reliability is significately incresed. Table VI shows the overall recognition rate, rejection rate and reliability of the classifier fusion system with SVM verification module.

Classifier Fusion Scheme + SVM Verification Module	Rejection Rate (%)	Recognition Rate (%)	Reliability (%)
Combo I+ SVM Module	1.80	98.19	99.99
Combo II+ SVM Module	1.66	98.34	100
Combo III+ SVM Module	1.70	98.30	100
Combo IV+ SVM Module	1.80	98.15	99.95

Table 4. Recognition performance of classifier fusion system with SVMs verification module

Comparing Table IV with Table III, it can be concluded that the system's reliability has been improved significantly. Both Combo II+SVM Module and Combo III +SVM Module achieve 100% reliability and have recognition rates over 98.30%. The remaining characters will be rejected and will be processed manually.

There are a few reasons behind the better recognition performance:

1) ANN classifier and KNN clasifier are trained using different feature sets, which makes the two classifiers in the fusion system complementary;

2) Gating network can enhance recognition rate and reliability;

3) SVM verification module is trained by different signal input, which ensures that the overall system reliability will increase.

Fig. 9 shows the reliability comparsion between the classifier fusion system and the classifier fusion system with SVM verification module.

Fig. 10 shows the reliability improvement from individul classifier to classifier fusion system (including an ANN, a KNN, and a gating network), and to the fusion classifier system with

a SVM verification module. Experiments demonstrated that the reliability increases from 98.5% to nearly 100%.
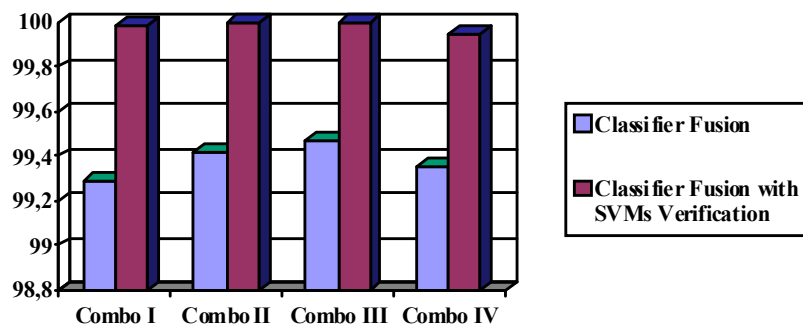

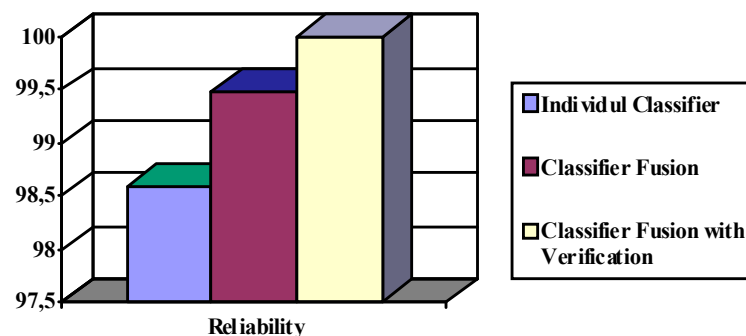Fig. 9. Reliability comparsion of classifier fusion system and the system with SVM verification module

Fig. 10. Reliability improvement from individul classifier to classifier fusion system with SVM verification module

6. Conclusions

In this paper, we proposed a novel classifier fusion system to congregate the recognition results of an ANN classifier and a modified KNN classifier. The recognition results are verified by the recognition results of SVM. As two entirely different classification techniques (image-based OCR and 1-D digital signal SVM classification) are applied to the system, experiments have demonstrated that the proposed classifier fusion system with SVM verification module can significantly increase the system's recognition reliability and can suppress misrecognition rate at the same time.

In the future, the theory foundation of classifier fusion with rejection strategies will be further investigated. It is expected that the theory will be employed to solve more complex pattern recognition problems.

7. References

Chow, C. K. (1970), On Optimum Recognition Error and Reject Tradeoff. *IEEE Transactions on Information Theory*, Vol-16, No. 1, pp. 40-46.

Decoste, D. & Scholkopf, B. (2002), Training Invariant Support Vector Machines. *Machine Learning*, Vol. 46, No. 1-3, pp.160-190.

Duda, R. O.; Hart, P. E. & Stork, D. G. (2000). *Pattern Classification*, John Wiley & Sons, Inc., Wiley-Interscience, Second Edition.

Frelicot, C. & Mascarilla, L. (2002). Reject Strategies Driven Combination of Pattern Classifiers, *Pattern Analysis and Applications*, Vol. 5, No. 2, pp. 234-243.

Friedman, J. H. (1997). On Bias, Variance, 0/1-loss and the Curse-of-dimensionality, *Data Mining and Knowledge Discovery*, Vol. 1, No. 1, pp.55-77.

Giusti, N.; Masuli, F. & Sperduti, A. (2002). Theoretical and Experimental Analysis of A Two-stage System for Classification, *IEEE Transactions on PAMI*, Vol-24, No. 7, pp. 893-904.

Heisele, B.; Serre, T.; Prentice, S. & Poggio, T. (2003). Hierarchical Classification and Feature Reduction for Fast Detection with Support Vector Machines, *Pattern Recognition*, Vol. 36, No. 9, pp.2007-2017.

http://en.wikipedia.org/wiki/Magnetic_ink_character_recognition

Keller, J. M., Gray, M. R. & Givens Jr.; J. A. (1985). A fuzzy K-Nearest Neighbor Algorithm, IEEE Trans. on SMC, Vol.SMC-15, No.4, pp. 580-585.

Kittler, J; Hatef, J; Duin, R. P. & Matas, J. (1998). On Combining Classifier, *IEEE Transactions on PAMI*, Vol. 20, No. 3, pp. 226-239.

Kuncheva, L. I. (2002). A Theoretical Study on Six Classifier Fusion Strategies, *IEEE Transactions on PAMI*, Vol. 24, No. 2, pp. 281-286.

Liu, C. L.; Nakashima, K.; Sako, H. & Fujisawa, H. (2004). Handwritten Digit Recognition: Investigation of Normalization and Feature Extraction Techniques, *Pattern Recognition*, Vol. 37, No. 2, pp.265-279.

Siedlecki, W. & Sklansky, J. (1989). A Note on Genetic Algorithm for Large-scale Feature Selection, *Pattern Recognition Letters*, Vol. 10, No. 5, pp.335-34.

Suen, C. Y. & Tan, J. (2005). Analysis of Errors of Handwritten Digits Made by A Multitude of Classifiers, *Pattern Recognition Letters*, Vol. 26, No. 1, pp. 369-379.

Zhang, P. ; Bui, T. D. & Suen, C.Y. (2007). A Novel Cascade Ensemble Classifier System with A High Recognition Performance on Handwritten Digits, *Pattern Recognition*, Vol. 27, No. 12, pp. 3415-3429.

Zhang, P. ; Bui, T. D. & Suen, C.Y. (2005). Hybrid Feature Extraction and Forest Feature Selection for Increasing Recognition Accuracy of Handwritten Numerals, *in the Proceedings of 8th International Conference on Document Analysis and Recognition (ICDAR)*,

Zimmermann, M.; Bertolami, R. & Bunke, H. (2002). Rejection Strategies for Offline Handwritten Sentence Recognition, *in the Proceedings of the 17th International Conference on Pattern Recognition (ICPR2002)*, Vol. 2, Quebec, Canada, pp. 550-553.

Machine Learning

Edited by Yagang Zhang

Machine learning techniques have the potential of alleviating the complexity of knowledge acquisition. This book presents today's state and development tendencies of machine learning. It is a multi-author book. Taking into account the large amount of knowledge about machine learning and practice presented in the book, it is divided into three major parts: Introduction, Machine Learning Theory and Applications. Part I focuses on the introduction to machine learning. The author also attempts to promote a new design of thinking machines and development philosophy. Considering the growing complexity and serious difficulties of information processing in machine learning, in Part II of the book, the theoretical foundations of machine learning are considered, and they mainly include self-organizing maps (SOMs), clustering, artificial neural networks, nonlinear control, fuzzy system and knowledge-based system (KBS). Part III contains selected applications of various machine learning approaches, from flight delays, network intrusion, immune system, ship design to CT and RNA target prediction. The book will be of interest to industrial engineers and scientists as well as academics who wish to pursue machine learning. The book is intended for both graduate and postgraduate students in fields such as computer science, cybernetics, system sciences, engineering, statistics, and social sciences, and as a reference for software professionals and practitioners.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Ping Zhang (2010). A Classifier Fusion System with Verification Module for Improving Recognition Reliability, Machine Learning, Yagang Zhang (Ed.), ISBN: 978-953-307-033-9, InTech, Available from: http://www.intechopen.com/books/machine-learning/a-classifier-fusion-system-with-verification-module-for-improving-recognition-reliability

INTECH
open science | open minds