MVA '98 IAPR Workshop on Machine Vision Applications, Nov. 17-19, 1998, Makuhari, Chiba, Japan

3—23
Optimum Design Parameters of Classifiers Used for Omni-font Machine-printed Numeral Recognition based on the Minimum Classification Error Criterion

Katsumi Marukawa, Kazuki Nakashima, Hiroshi Shinjo, Yoshihiro Shima, and Hiromichi Fujisawa
Central Research Laboratory, Hitachi, Ltd.

Abstract

The optimal design parameters of classifiers for omni-font machine-printed numeral recognition based on the minimum classification error (MCE) criterion are determined experimentally. The design parameters that influence the accuracy of an optical character reader (OCR) are: similarity measure (or distance measure), kinds of features, dimension of the feature vector, method of training, number of templates per category, and the size of a training sample set. It was found that the optimum design parameters were simple similarity, four templates per category, and 576 dimensions (i.e., four directional feature planes of 12 x 12 blocks). The directional feature classifier with these design parameters gave the best performance and had the smallest memory size and computational cost of all the classifiers.

1. Introduction

An optical character reader (OCR) converts paper documents into an electronic form. In such OCR applications, numeral recognition is one of the most important methods and requires the highest accuracy. Furthermore, a large variety of type faces (fonts) must be recognized in order to widen the application fields.

There are two character recognition approaches based on pattern matching. One is multi-template pattern matching based on the nearest-neighbor decision rule [1]. The other is the statistical approach based on the statistical decision rule [2]. In the statistical approach, there are basic problems such as error in estimating distributions or distribution parameters. In several papers, methods to solve these problems were proposed [3],[4]. The multi-template pattern matching approach does not suffer these problems. However, this approach can't use information about the distributions. It is known that this approach requires many more training samples than the statistical approach does. A common objective in both approaches is that the classifier with the optimum design parameters is found [5]. This is fundamentally important in the design of highly accurate classifiers.

This study adopts the multi-template pattern matching approach rather than the statistical one due to its flexibility in training. The design parameters and factors that influence the accuracy are: similarity measure (or distance measure), kinds of features, dimension of the feature vector, method of training, number of templates per category, and the size of a training sample set. In this paper, a classifier with optimum design parameters is investigated experimentally. And the interrelationships among the accuracy, amount of computation, and memory size for the templates are determined.

Address: 1-280, Higashi-koigakubo, Kokubunji, Tokyo 185-0014, Japan. E-mail: marukawa@crl.hitachi.co.jp

2. Recognition Method

2.1 Pre-processing

Pre-processing includes three procedures: extraction of contours from a character image, noise reduction, and linear size normalization. The contour of character is represented with eight directional codes. It is normalized to a size of 64 x 64.

2.2 Feature extraction

Two kinds of features were investigated. One of them is a directional feature which composes four separate feature planes, each of which represents directional elements of the pattern that corresponds to one of the four directions: vertical, right-diagonal, horizontal, and left-diagonal directions. The directional features are gray-scale image patterns. The other feature studied here is a simple, blurred image.

The normalized data is divided into N (M horizontal x M vertical) blocks. And each feature is extracted from the blocks with a Gaussian filter.

2.3 Discriminant function

Three kinds of discriminant functions were investigated as similarity measures: 1) simple similarity, which is a normalized cosine of two feature vectors; 2) directional simple similarity, which is an average of four simple similarities calculated for the four directional planes; and 3) Euclid distance.

2.4 Training

For training the classifier to optimize the distribution of the multiple templates, we used a generalized version of the learning vector quantization (GLVQ [6]). This is one of the learning methods based on the minimum classification error (MCE) criterion [7],[8]. This criterion is a mathematically-derived optimization algorithm which minimize the loss function of the misclassification measure in an iterative manner.

The training procedure consists of two steps. First, an LBG clustering algorithm [9] is activated in order to make an initial dictionary. Then, in the second step, GLVQ is activated eight times over.

3. Classifiers

We defined four classifiers by combining the features and discriminant functions described above.

1) Classifier: #1

The directional feature is used. The discriminant function is directional simple similarity and is represented as follows,

$$f_k(i),\quad i=1,\dots,N;\quad k=1,\dots,4 \tag{1}$$

$$g_{jk}(i),\quad i=1,\dots,N;\quad k=1,\dots,4 \tag{2}$$

$$S1(f, g_j) = \sum_{k}^{4} \frac{\sum_{i}^{N} f_k(i) \bullet g_{jk}(i)}{\sqrt{\sum_{i}^{N} f_k^2(i) \sum_{i}^{N} g_{jk}^2(i)}} \Big/ 4, \tag{3}$$

where k is the number of the direction plane, N is the dimension of the directional feature for each plane, $f_k(i)$ is the feature vector of the input pattern, and $g_{jk}(i)$ is the jth template.

2) Classifier: #2

The directional feature is used. The discriminant function is simple similarity and is represented as follows,

$$f(i),\quad i=1,\dots,N \tag{4}$$

$$g_j(i),\quad i=1,\dots,N \tag{5}$$

$$S2(f, g_j) = \frac{\sum_{i}^{N} f(i) \bullet g_j(i)}{\sqrt{\sum_{i}^{N} f^2(i) \sum_{i}^{N} g_j^2(i)}}, \tag{6}$$

where N is the dimension of the directional feature.

3) Classifier: #3

The blurred image feature is used. The discriminant function is simple similarity and is represented as follows,

$$f(i),\quad i=1,\dots,N \tag{7}$$

$$g_j(i),\quad i=1,\dots,N \tag{8}$$

$$S3(f, g_j) = \frac{\sum_{i}^{N} f(i) \bullet g_j(i)}{\sqrt{\sum_{i}^{N} f^2(i) \sum_{i}^{N} g_j^2(i)}}, \tag{9}$$

where N is the dimension of the blurred image feature.

4) Classifier: #4

The blurred image feature is used. The discriminant function is the Euclid distance and is represented as follows,

$$f(i),\quad i=1,\dots,N \tag{10}$$

$$g_j(i),\quad i=1,\dots,N \tag{11}$$

$$D(f, g_j) = \sqrt{\sum_{i}^{N} (f(i) - g_j(i))^2}. \tag{12}$$

4. Experiment

4.1 Experimental method

We analyzed the performances of the classifiers, in terms of an error rate under forced recognition, by conducting recognition experiments.

Other parameters are dimensions of features, the number of templates for each category, and the size of the training sample set. First, three kinds of blocks divided into normalized data were investigated. The block sizes are 8 x 8, 12 x 12, and 16 x 16. Dimensions of the directional feature are thus 256, 576, and 1,024, respectively. Dimensions of the blurred image feature are 64, 144, and 256. Second, numbers of template for each category were 4, 8, and 16. Finally, one eighth, half, and all of the database for training were used as the training sample set.

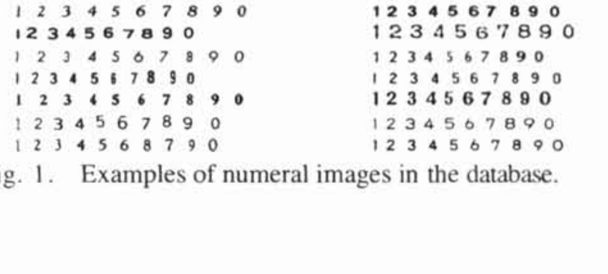


Fig. 1. Examples of numeral images in the database.

4.2 Database

We used three numeral pattern databases of the characters 0 to 9 printed in about 130 different fonts. The databases contain images with normal print quality and two kinds of poor print quality; i.e., blurred images and faint images. The font size was 10 points and the scan resolution was 200 dpi. The printing condition was not good enough to get high accuracy, but it was deliberately set to amplify the differences of the classifiers. Figure 1 shows examples of numeral images in the databases.

The number of samples included in the four kinds of databases is shown as follows.

1) Database for training: 213,504.
2) Normal quality database for testing: 266,968.
3) Database with blurred images for testing: 9,478.
4) Database with faint images for testing: 15,501.

4.3 Experimental Results

4.3.1 Effectiveness of training

For 213,503 training samples, the effectiveness of training is shown in Figure 2. The training process using GLVQ converged rather rapidly and reached the error rate of zero by increasing the number of iterations for the same training set.

4.3.2 Normal quality database

For the normal quality database of 266,968 samples, the relations among the error rate and the sample size with the different parameters of feature dimensions are shown in Figure 3. The error rate decreased by increasing the size of the training set. Regarding the size of the training set, the classifier using the blurred image feature required about 110,000 samples in order to obtain zero

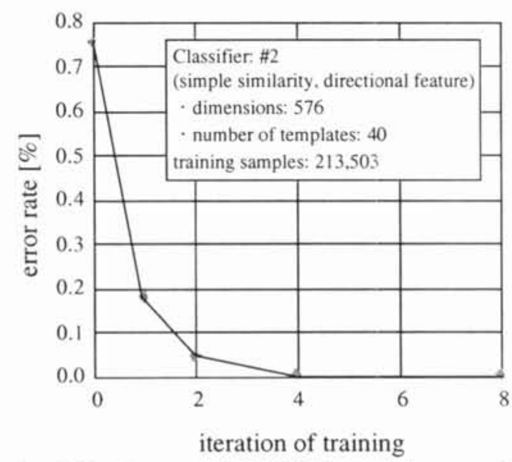
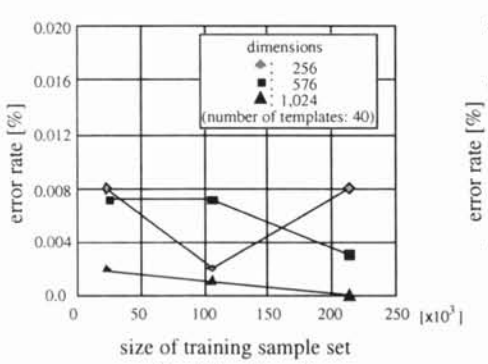
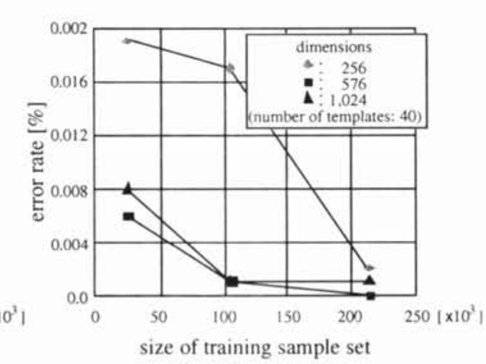


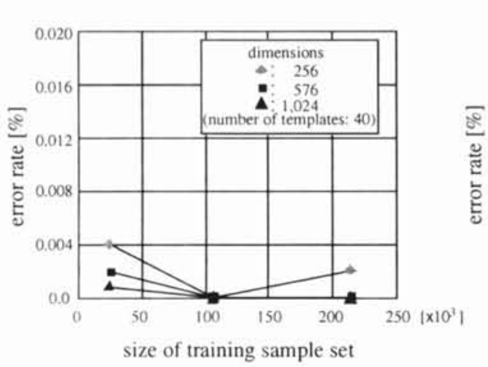
Fig. 2. Effectiveness of GLVQ for training sample set.



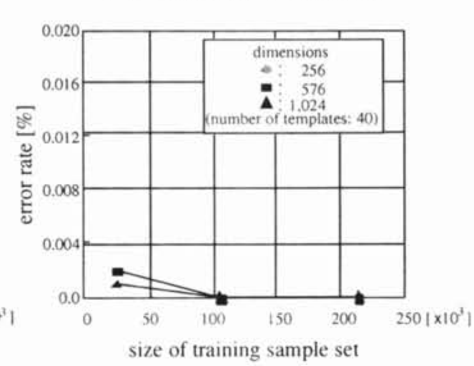
(a) Classifier: #1
(directional smiple similarity, directional feature)



(b) Classifier: #2
(simple similarity, directional feature)



(c) Classifier: #3
(simple similarity, blurred image feature)



(d) Classifier: #4
(Euclid distance, blurred image feature)

Fig. 3. Relations between error rate and size of training sample set with different parameters of feature dimensions.

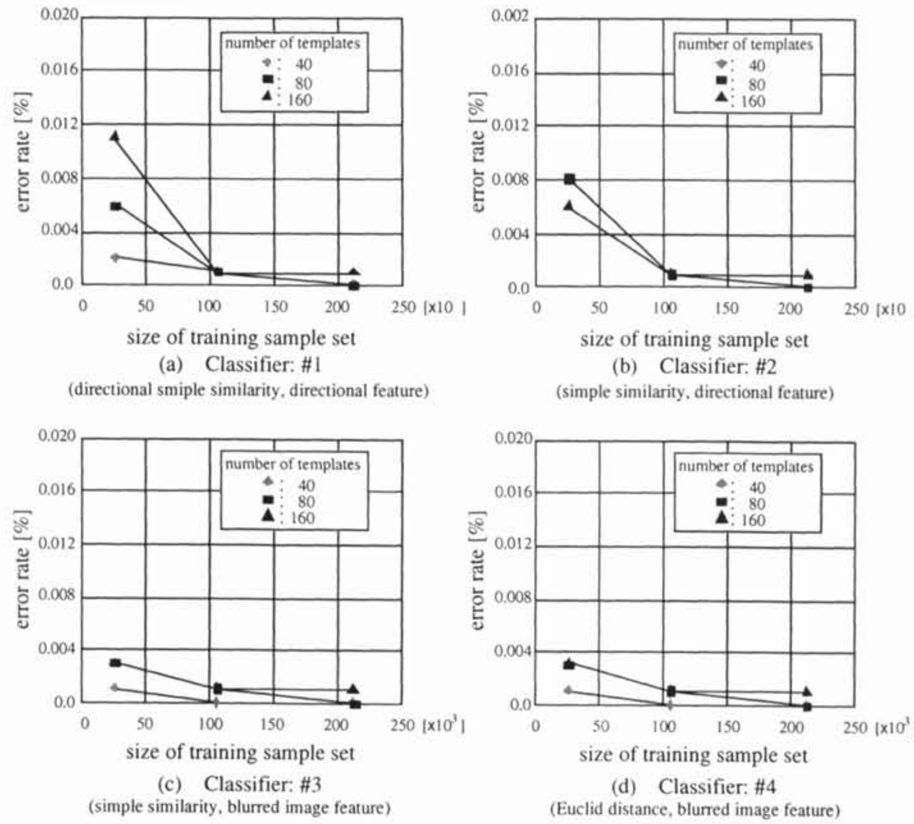


Fig. 4. Relations between error rate and size of training sample set with different number of templates in 16 x 16 blocks.

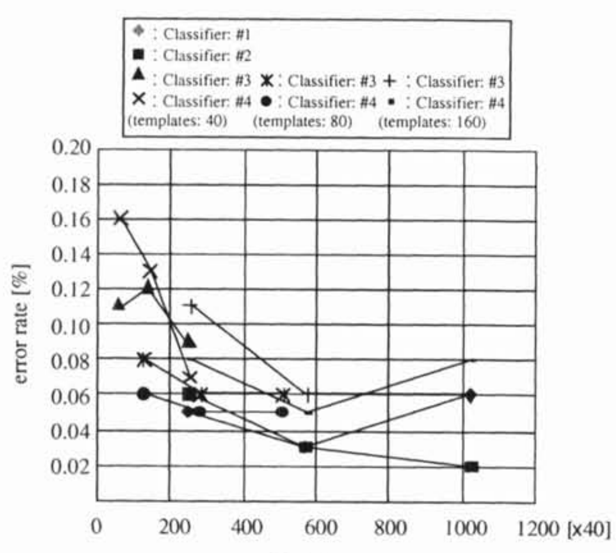


Fig. 5. Performance of each classifier with different magnified dimensions (for a database with blurred images).

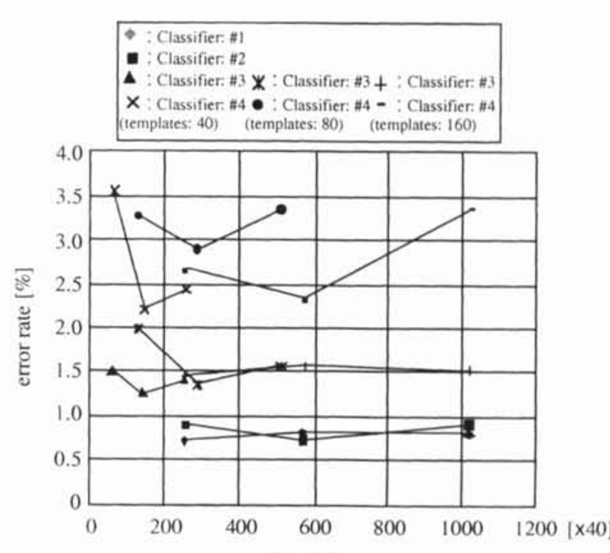


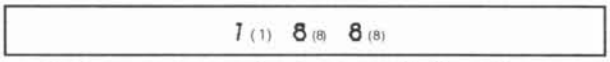
Fig. 6. Performance of each classifier with different magnified dimensions (for a database with faint images).

error rate, while the classifier using the directional feature required about twice as many, or 220,000 samples, to get the same performance.

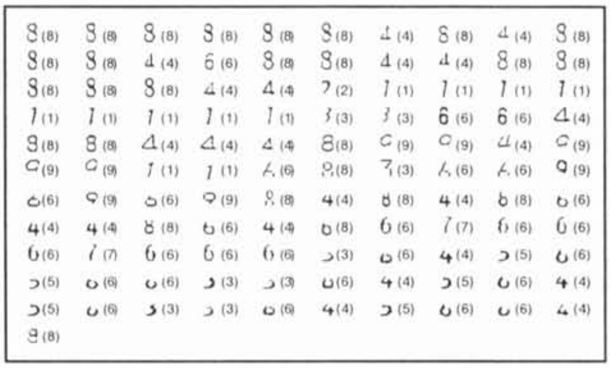
The relations among the error rate and the sample size are shown in Figure 4. Here, the parameters are the number of templates in a 16 x 16 block. That is, the feature space constructed by using about 100 fonts is well represented by using four templates per category.

4.3.3 Poor quality databases

The amount of computation and the memory size are proportional to feature dimension multiplied by the



(1) Erroneous patterns in a database with blurred images



(2) Erroneous patterns in a database with faint images

Fig. 7. Erroneous patterns in poor quality databases.

number of templates. Here, we define this as "magnified dimension". The performances of classifiers with the different magnified dimensions for the databases of blurred images and faint images are shown in Figures 5 and 6, respectively. Erroneous patterns in poor quality databases are shown in Figure 7.

It turned out that the directional feature classifier is more robust than the blurred image feature classifier. In other words, although they gave almost the same performance for the normal quality samples, the directional feature classifier attained a better performance for the poor quality samples; that is, it made half as many errors as the classifier using the blurred image feature. This result shows that a structural feature such as the directional feature can give robustness.

It was found that the optimum design parameters were simple similarity, four templates per category, and 576 dimensions (i.e., four directional feature planes of 12 x 12 blocks). The directional feature classifier with these design parameters gave the best performance and had the smallest memory size and computational cost of all the classifiers.

5. Summary

We studied the interrelationships among the accuracy, the amount of computation, and the memory size for omni-font machine-printed numeral recognition based on the minimum classification error criterion. It was found that the optimum design parameters were simple similarity, four templates per category, and four directional feature planes of 12 x 12 blocks. The directional feature classifier gave the best performance and had the smallest memory size and computational cost.

Moreover, we will investigate the optimum design parameters for a statistical classifiers and compare these results with those presented in this paper. This should enable us to find the best classifiers for numeral recognition.

REFERENCES

[1] R. O. Duda and P. E. Hart: "Pattern Classification and Scene Analysis," Willy-Interscience Publication (1973).

[2] K. Fukunaga: "Introduction to Statistical Pattern Recognition," 2nd edition Academic Press, New-York (1990).

[3] T. Wakabayashi, S. Tsuruoka, F. Kimura, and Y. Miyake: "Study on Feature Selection in Handwritten Numeral Recognition," *Trans. IEICE Japan*, J78-D-II, 11, pp. 1627-1638 (1995) (in Japanese).

[4] N. Kato, M. Abe, and Y. Nemoto: "A Handwritten Character Recognition System Using Modified Mahalanobis Distance," *Trans. IEICE Japan*, J79-D-II, 1, pp. 45-52 (1996) (in Japanese).

[5] T. Matsunaga and H. Kida: "An Experimental Study of Learning Curve on Linear Classifiers," *Technical Report of IEICE Japan*, PRU 94-56, pp. 67-74 (1994) (in Japanese).

[6] A. Sato and K. Yamada: "Character Recognition using Generalized Learning Vector Quantization," *Technical Report of IEICE Japan*, PRU 95-219, pp. 23-30 (1996) (in Japanese).

[7] S. Amari: "A Theory of Adaptive Pattern Classifiers," *IEEE Trans.*, EC-16, 3, pp. 299-307 (1967).

[8] B. H. Juang and S. Katagiri: "Discriminative Learning for Minimum Error Classification," *IEEE Trans.*, SP-40, 12, pp. 3043-3054 (1992).

[9] Y. Linde, A. Buzo, and R. M. Gray: "An Algorithm for Vector Quantizer Design," *IEEE Trans.*, COM-28, 1, pp. 84-95 (1980).