Recognition of Handwritten Numerals Using a Combined Classifier with Hybrid Features

Kyoung Min Kim[1,4], Joong Jo Park[2], Young Gi Song[3], In Cheol Kim[1], and Ching Y. Suen[1]

[1] Centre for Pattern Recognition and Machine Intelligence (CENPARMI), Concordia University, 1455 de Maisonneuve Blvd. West, Suite GM606, Montreal, Canada H3G 1M8
{kkm,kiminc,suen}@cenparmi.concordia.ca

[2] Department of Control and Instrumentation Engineering, Gyeongsang National University, 900, Gazwa-dong, Chinju, Gyeongnam, 660-701, Korea

[3] Hyundai Information Technology Research Center, 1-8 San, Mabukri, Kusung Myun, Yongin City, Kyonggido 449-716, Korea

[4] Department of Electrical Engineering, Yosu National University, San 96-1, Dunduk-dong, Yeosu, Chonnam, 550-749, Korea

Abstract. Off-line handwritten numeral recognition is a very difficult task. It is hard to achieve high recognition results using a single set of features and a single classifier, since handwritten numerals contain many pattern variations which mostly depend upon individual writing styles. In this paper, we propose a recognition system using hybrid features and a combined classifier. To improve recognition rate, we select mutually beneficial features such as directional features, crossing point features and mesh features, and create three new hybrid feature sets from them. These feature sets hold the local and global characteristics of input numeral images. We also implement a combined classifier from three neural network classifiers to achieve a high recognition rate, using fuzzy integral for multiple network fusion. In order to verify the performance of the proposed recognition system, experiments with the unconstrained handwritten numeral database of Concordia University, Canada were performed, producing a recognition rate of 97.85%.

1 Introduction

As a typical example of complex pattern recognition system, totally unconstrained handwritten numeral recognition is a real challenge. It is difficult to achieve high recognition results using a single feature set and a single classifier, since this system contains many pattern variations which mostly depend upon individual writing styles. Several methods have been proposed and implemented in a number of different ways in the practical applications such as zip code recognition, document analysis and factory automation and accuracy of 95% and above have been reported [1][11]. To improve the recognition rate, current researchers aim at two major tasks, i.e. selection of good feature candidates and realization of universal classifiers. First, the selection

A. Fred et al. (Eds.): SSPR&SPR 2004, LNCS 3138, pp. 992–1000, 2004.

of good feature candidates largely depends upon one's experience and their capabilities can only be evaluated by the recognition results. In this paper, we have selected feature extraction methods, which represent pattern variations effectively. Features that we extracted from input numerals include directional features, crossing point features, and zoning features. To improve the recognition rate, we combine mutually beneficial features and create three new feature sets by considering both the local and global characteristics of the input images. Second, the design of classifiers generally falls into three categories depending upon the feature vectors used, i.e., feature vector based classifier, syntactic/structural based classifier, and neural networks classifier [2][3][11]. But it is very difficult to achieve good performances with a single classifier, thus the current trend is to implement pattern multiple classifiers rather than a single classifier to achieve a highly reliable performance[4][11]. In this paper, we have used three neural networks as single classifiers, and used fuzzy integral to combine their outputs to obtain a higher recognition rate.

This paper is organized as follows. In section 2, we present a review of earlier works. Feature extraction methods are introduced in section 3. In section 4, we describe our classification methodology and fuzzy integral for the aggregation of output of the single neural networks. Recognition result of each single classifier and overall recognition results are included in section 5. Finally a brief summary and future study are discussed in the last section.

2 Review of Earlier Works

In this section, we review some representative methods for the recognition of totally unconstrained handwritten numerals. These approaches generally fall into two realms according to the attributes of the feature vectors. The first category includes techniques such as template matching, moments, characteristic loci and mathematical transforms which generally represent global characteristics of the input numerals. In the second category, efforts are aimed at extracting the shape characteristics of the input numerals from their skeletons or contour profiles. Such features include loops, endpoints, junctions, arcs, concavities and convexities that generally represent local characteristics of the input numerals.

Template feature extraction methods have the problem that they are very susceptible to small variations such as translation and rotation, thus the modified version such as affine transform is a good alternative [7]. Structural decomposition methods of feature extraction strive to represent structural properties of the input patterns. Series expansions are methods of representing a signal as a series of coefficients created by projecting the signal onto some basis, but Fourier descriptors can not be applied to fragmented characters, which usually happen in the handwritten numerals because this method extracts only one single closed contour. Furthermore the frequency information of Fourier transform is global, intuitively; a more localized frequency representation should be more effective such as wavelets transform [10]. Moments are invariant to size and rotation, and some moment invariants have invariant characteristics to skewed and mirrored images. Using Zernike moments, we can obtain size and rotation independent features. And KL transform has the best optimal rotation independent features. And KL transform has the best optimal information compactness in terms of mean square error, and is used for object recognition in several application domains, e.g., face recognition and the verification of slab number.

3 Feature Extraction

In the case of handwritten numerals, it seems natural to use some a priori knowledge about the recognition task in order to transform the low level information of the pixel image into a data representation at a higher level. A good feature should represent characteristics of a class ('0'-'9') that helps to distinguish it from other classes, while remaining invariant to pattern diversities within the class. Features also should avoid redundant information since this will lead to a more complex distribution in the feature space and therefore requires a more complex model. Before we extract feature vectors from the input image, we first extract generic data area in the input image by using projections on vertical and horizontal axes. After this segmentation, we undertake preprocessing steps, viz, median filtering for removing small holes inside black pixels which may cause problems especially when extracting crossing point feature and then we perform size normalization to reduce the effect of pattern variations. In this paper, three features are used. The first one consists of directional and global features. Line segments and their directions seem to be an adequate feature. For each location in the image, information about the presence of a line segment of a given direction (4-directions) is stored in a feature map. Among many first-order differential operators, the Kirsch edge detector has been known to detect four directional edges more accurately in comparison to other methods because the Kirsch edge detector considers all eight neighbors. Directional feature maps for horizontal (H), vertical (V), right-diagonal(R), and left-diagonal (L) directions are easily calculated by using Kirsch masks[5]. Some researchers used the four directional feature maps which are normalized to an 8×8 format [5], and others used the four directional feature maps normalized to 4×4 [4][6], and one global feature normalized to 4×4. In this paper, the size of 4×4 is used for directional and global feature maps. From normalized 16×16 input images, four directional images are obtained by Kirsch operation in which 10 is used as a threshold value. After that, these 16×16 directional images and input image are compressed to 4×4 feature maps by accumulating the pixels in each 4×4 subregion. This feature extraction process is depicted in Figure 1.

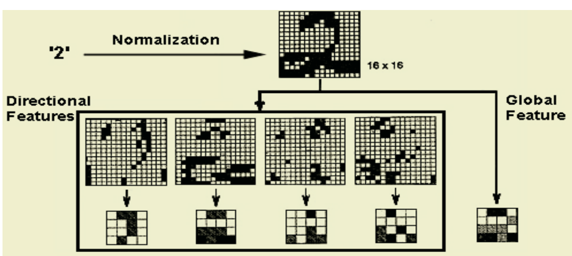

Fig. 1. Overview of feature extraction

The second feature is zoning feature that can be considered as a good candidate for global feature, since zoning feature is obtained by diminishing the image resolution, thus, small variations are easily expressed by the rate of scaling factor although it is susceptible to slant and rotation. In this paper, 10×10 are used as the size of zoning feature. While the first feature considers global and local characteristics of input numerals at the same time, zoning feature represents only global characteristics, which contain more detailed shape information. Finally, the third feature is a crossing point feature. Crossing point is defined as the point at which white pixel changes to black pixel at one scan line. And crossing point feature is obtained by summing the number of crossing points for each vertical and horizontal scan line. In this paper, from a normalized 20×20 input images, we obtain a 20-dimensional crossing point feature vector, 10-dimensions per axis. Each feature element of this feature vector is calculated as follows: one feature element is obtained by considering every two scan lines, i.e., one feature element is calculated by summing the number of crossing points on two adjacent scan lines and dividing it by the maximum number of occurrences in one scan line at each axis. We determine the maximum number of occurrences as 2 for a horizontal scan line and 4 for a vertical scan line. So, in our case, 4 and 8 are used respectively. The purpose of using two scan lines per feature element is to reduce the dimension of the feature vector and avoid confusions caused by shape variation of the numeral. Figure 2 shows this feature extraction process.

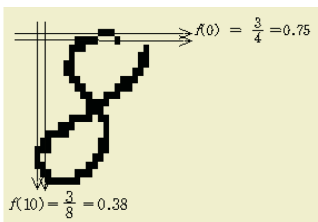

Fig. 2. Extraction of crossing point feature

By using the three features described above, we created three feature sets to improve the recognition rate; feature set 1(FS1) is 4×4 directional and global feature maps which is the first feature itself, feature set 2(FS2) is the combination of zoning feature and crossing point feature and feature set 3(FS3) is the combination of feature set 1 and crossing point feature. Each feature set is used as the input for each single classifier.

4 Multiple Classifications

In this paper, we combine three independent neural networks classifiers to recognize totally unconstrained handwritten numerals. Recently the concept of combining multiple networks has been actively studied for developing highly reliable neural networks system. One of the key issues of this approach is how to combine the results of the various networks to give the best estimate of the optimal result. The basic idea of the multiple network classifiers is to develop n independently trained classifiers with particular features, and to classify a given input pattern by obtaining a classification from each replica of the network and then using combination methods [8]. Each classifier uses an independent feature vector as input. The combiner acts as a final decision making processor for selecting the appropriate output class. There are two methods of combining multiple classifiers: loosely coupled and tightly coupled subnetworks. Ryu used tightly coupled subnetworks [2], based on the results of the intermediate layer as the inputs of the final classification stage. Loosely coupled subnetworks use final outputs of each subnetwork and they evaluate the overall classification results using only these values. There are several methods to achieve this goal, such as winner take all, majority voting, Dempster-shafer method and BKS (Behavior knowledge space) method etc. [9][11]. In this paper we use the concept of fuzzy integral to combine multiple classifiers. Fuzzy integral is a nonlinear functional that is defined with respect to a fuzzy measure, especially g_λ-fuzzy measure introduced by Sugeno and provides a useful way for aggregating information. To calculate fuzzy integral, we first select fuzzy measure satisfying boundness, monotonousness and continuity. Fuzzy measure can be determined in several ways. Sugeno introduced the g_λ-fuzzy measure satisfying the following property.

$$g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B) \tag{1}$$

For all $A, B \subset X, A \cap B = \Phi$ and for some $-1 < \lambda < \infty$.

Fuzzy integral is defined as follows; Let Y be a finite set and $h : Y \to [0,1]$ a fuzzy subset of Y. The fuzzy integral over Y of the function h with respect to a fuzzy measure g is defined by

$$h(x) \circ g(\bullet) = \max_{E \subseteq Y} \left[\min \left(\min_{y \in E} h(y), g(E) \right) \right] = \max_{a \in [0,1]} \left[\min \left(\alpha, g(F_a) \right) \right] \tag{2}$$

where $F_a = \{ y \mid h(y) \geq \alpha \}$.

The calculation of the fuzzy integral when Y is a finite set is easily given. Let $Y = \{ y_1, y_2, y_3, \ldots, y_n \}$ be a finite set and let $h : Y \to [0,1]$ be a function which is the output of each classifier. Suppose $h(y_1) \geq h(y_2) \geq \cdots \geq h(y_n)$, (if not, Y is rearranged so that the relation holds). Then fuzzy integral I, with respect to a fuzzy measure g over Y can be computed by

$$I = \max_{i=1}^{n} \left[\min(h(y_i), g(A_i)) \right] \tag{3}$$

where $A_i = \{ y_1, y_2, \cdots, y_n \}$.

Note that when g is a g_λ-fuzzy measure, the values of $g(A_i)$ can be determined recursively as

$$g(A_i) = g^i + g(A_{i-1}) + \lambda g^i g(A_{i-1}), \quad \text{for } 1 < i < n. \tag{4}$$

where λ is given by solving the following equation

$$\lambda + 1 = \prod_{i=1}^{n} (1 + \lambda g^i) \tag{5}$$

and $\lambda \in (-1, \infty)$ and $\lambda \neq 0$ and n is the number of subnetworks. This can be easily calculated by solving a $(n-1)$ th degree polynomial and finding the unique root greater than -1. The calculation of the fuzzy integral with respect to a g_λ -fuzzy measure would only require the knowledge of density function, where i -th density g^i is interpreted as the degree of importance of the source y_i toward the final evaluation. Fuzzy integral shows a good performance when the difference of outputs of subnetworks is minute because the value of g^i dramatically changes the performance compared to simple averaging or majority voting method. There is however, a key issue unsolved in the application of fuzzy integral, the determination of density values that determine the fuzzy measure used in the fusion process. In this paper, we determine these fixed densities from the knowledge of recognition rates of three neural networks, which also reflect the importance of each feature set they used.

5 Experimental Results

In this paper, we use three different feature sets to recognize totally unconstrained handwritten numerals. Aforementioned features have certain deficiencies in representing the input numerals, e.g. directional features can be confused when the input numeral is slanted or connected among pixels, and zoning features always lose their local characteristics. To effectively manipulate global and local characteristics of input numerals, we combine two or three mutually complementary features and make three new feature sets. This can make up for the weakness of one feature by the strength of other features. And the network fusion algorithm is used to improve the overall recognition results. The overall recognition system block diagram of our scheme is shown in Fig. 3 where NN1, NN2, and NN3 are feedforward neural network classifiers with one hidden layer, for which backpropagation learning algorithm is used.

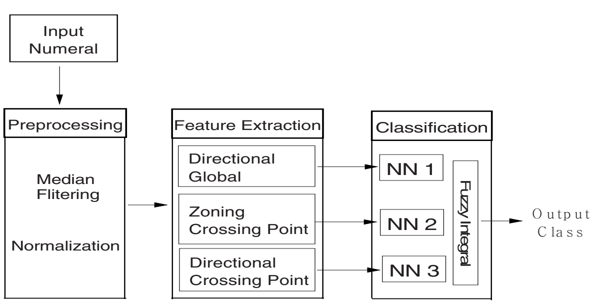

Fig. 3. The overall recognition system

5.1 Results from Single Classifiers

In this section, we use learning rate as 0.9, momentum as 0.7 for each neural network. The rejection criterion used in this paper is

$$RC = \frac{O_1 - O_2}{O_1 + O_2} \tag{6}$$

where O_1 is the highest output node value and O_2 is the second highest of a given input pattern. Rejection takes place if the RC falls below a threshold value that we set as 0.2. And the reliability in the table is computed according to the following equation:

$$Rel. = \frac{Rec.}{Rec. + Sub.} \times 100 \tag{7}$$

where *Rec.* = correct recognition rate, and *Sub.* = substitution error rate.

5.1.1 Feature Set 1

The first feature set (FS1) consists of directional features and global features. FS1 has been used in several studies by itself and in combination with other features. Directional features have robustness with small pattern variations. To extract directional feature, we first normalize the input image to 16×16 using bilinear interpolation. And we obtain the four directional feature maps normalized to a 4×4 and one global feature normalized to 4×4, in which we use the value 10 and 12 respectively for value normalization to the range [0,1]. In the training process, we used several different numbers of hidden layer nodes, and obtained satisfactory error rates (0.8%) when we used the number of hidden layer nodes as about 80% of the dimension of the input feature vector.

From the above recognition result, we see that feature set 1 has poor discriminative capability in number '8'. This is because of slanted and narrow inner spaces of the number '8'. And because we used only normalized 4×4 image as a global feature, this feature is inadequate to represent overall shape variations when distorted. But the overall recognition result is better than other relevant studies.

5.1.2 Feature Set 2

The second feature set (FS2) is composed of zoning features and crossing point features. The training result (0.7% error rate) is a little better than feature set 1. This feature set represents global characteristics very well, but it may lead to confusion among similar numerals. The recognition result of feature set 2 shows a little improvement in comparison with feature set 1. We can observe that this feature set shows lower recognition results for '2' and '3' than other numerals due to the weakness of the zoning feature.

5.1.3 Feature Set 3

To effectively manipulate the minute differences among input numerals, we make feature set 3 (FS3), which is composed of feature set 1, and crossing point feature which was derived in the same manner as used in the feature set 2. The training result (0.01%) is much better than those of previous two feature sets. From Table 1, we see that this feature set shows better recognition results than the previous two recognition results. This is mainly due to the FS3, which represents local and global characteristics of input numerals more effectively than the previous two individual feature sets. This recognition result alone has provided another possibility for practical use. But in numerals '2' and '3', there exist certain confusion in discerning one from the other. We will deal with an aggregation of three neural network outputs by using fuzzy integral to compensate for the weakness of a single classifier's limitations and improve the overall recognition result.

5.2 Results for Using Multiple Classifiers

As introduced in the previous section, the network fusion using fuzzy integral computes an overall recognition result by using fuzzy integral and the results of all the subnetworks. We use $g^1 = 0.31$, $g^2 = 0.32$, and $g^3 = 0.33$ as density values for NN1, NN2 and NN3 according to their recognition rates. The final recognition result using combined multiple classifiers is shown in Table 1. And some examples of misrecognized numerals are shown in Fig. 4.

Comparatively, when we use simple averaging method in which the used weight values are the same as the above density values and majority voting techniques as a network fusion algorithm, we obtained final recognition results of 97.4% and 96.5%, respectively. From these results, we see that network fusion using fuzzy integral outperforms other methods, but we also need to pay more attention to selecting proper feature sets for each subnetwork.

Table 1. Final recognition result using feature sets and fuzzy integral

Methods	Rec.	Sub.	Rej.	Rel.
Feature Set 1	95.15	4.10	0.75	95.88
Feature Set 2	95.65	3.90	0.45	96.09
Feature Set 3	96.95	3.00	0.05	96.97
Multiple Classifier (Fuzzy Integral)	97.85	2.15	0.0	97.85

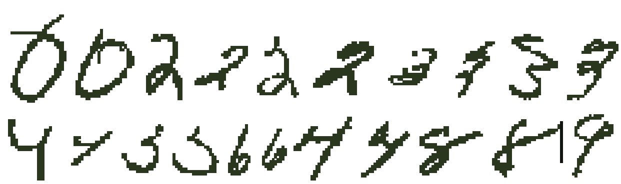

Fig. 4. Examples of misrecognized numerals

6 Conclusion

In this paper, we presented three feature sets, which consisted of several feature vectors to represent input numerals effectively. The recognition result shows that the proposed feature set effectively represents pattern variations such as slant, size and thickness etc. We used multilayer perceptron neural networks (MLP) as each single classifier, and used fuzzy integral to combine the outputs of three single classifiers. Combining of multiple classifiers using fuzzy integral considers the importance of input feature sets and a higher recognition rate was obtained in comparison with other proposed recognition systems. Further studies should be made to design classifiers which have more generalization capabilities and feature extraction methods which are mutually helpful for the recognition of unconstrained handwritten numerals.

References

1. Anil K. Jain and Torfinn Taxt: Feature Extraction Methods for Character Recognition -A Survey. Pattern Recognition 29(4) (1996) 641-662
2. Ryu Gang Soo, Chien Sung Il: Recognition of Handwritten Numerals by Relearning Hidden Layer Outputs of Modular Neural Nertworks. Korea Information Science Society 23(9) (1996) 931-939
3. Cho S. B.: Neural network classifiers for recognizing totally unconstrained handwritten numerals. IEEE Trans. on Neural Networks 8(1) (1997) 43-53
4. Mai T. A. and Suen C. Y.: A Generalized Knowledge Based System for The Recognition of Unconstrained Handwritten Numerals. IEEE Trans. on Systems, Man, and Cybernetics 20(4) (1980) 835-848
5. Knerr S. and Personnaz L. and Dreyfus G.: Handwritten Digit Recognition by Neural Networks with Singlelayer Training. IEEE Trans. Neural Networks 3(6) (1992) 962-968
6. Mohiuddin K. M. and Mao J.: A Comparative Study of Different Classifiers for Handprinted Character Recognition. Pattern Recognition in Practice IV (1994) 437-448
7. Wang J.: Resolving Multifont Character Confusion with Neural Networks. Pattern Recognition. 26(1) (1993) 175-187
8. Huang Y. S. and Suen C. Y.: A Method of Combining Multiple Experts for The Recognition of Unconstrained Handwritten Numerals. IEEE Trans. PAMI 17(1) (1995)
9. Cho S. B. and Kim J. H.: Multiple Network Fusion Using Fuzzy Logic IEEE Trans. Neural Networks 6(2) (1995) 487-501
10. Lee S. W. and Kim Y. J.: Multiresolution Recognition of Handwritten Numerals with Wavelet Transform and Multilayer Cluster Neural Network. 3rd international conference on document analysis and recognition (1995) 1010-1031
11. Suen Ching Y. et al.: Computer Recognition of Unconstrained Handwritten Numerals. Proceeding of the IEEE 80(7) (1992) 1162-1180