Improvement of Matching and Evaluation in Handwritten Numeral Recognition Using Flexible Standard Patterns

Hirokazu MURAMATSU, Takashi KOBAYASHI, Takahiro SUGIYAMA, and Keiichi ABE
Graduate School of Information, Shizuoka University
3-5-1, Johoku, Hamamatsu-Shi, 432-8011, Japan
Contact at: abe@cs.inf.shizuoka.ac.jp

Abstract

The purpose of this study is to develop a flexible matching method for recognizing handwritten numerals based on the statistics of shapes and structures learned from learning samples. In the recognition method we reported before, there were problems in matching of the feature points and evaluation of matching. To solve them, we propose a new matching method supplementing contour orientations with convex/concave information and a new evaluation method considering the structure of strokes.

With these improvements the recognition rate rose to 96.0% from the earlier figure 91.9%. We also made a recognition experiment on samples from the ETL-1 database and obtained the recognition rate 95.2%.

1. Introduction

In off-line recognition of handwritten characters statistical or neural network methods based on various features extracted from input characters are widely used[1][2]. On the other hand, humans are supposed to recognize characters based on not such features but structures of characters. In this sense the current main stream of character recognition may differ from the way of human character understanding and recognition.

An alternative way of machine character recognition is structural approach[3][4][5][6], especially matching an input character elastically to standard patterns, or vice versa[7][8][9][10]. Character recognition or shape learning using deformable templates is also a closely related attempt[11][12][13]. Recognition by elastic matching compares the shapes and structures themselves between the input character and the standard pattern; thus it would be closer to the way of human recognition than feature extraction/statistical classification approach is.

We have reported our earlier work with the methods of learning the standard pattern and of flexible matching [14]. This paper describes mainly the recognition stage with the improvements made after the report.

2. Overview of the recognition system

A schematic diagram of the recognition system proposed is illustrated in Figure 1. The system consists of two parts, each dealing with the learning stage and the recognition stage. In the learning stage it learns the standard patterns and their ranges of variations from the given set of learning samples. We obtain standard line patterns for each subcategory of ten numeral categories, augmented by statistics of constituting strokes. This stage is briefly explained in the next section.

In the recognition stage it recognizes input unknown characters by matching them with the elastically deformable standard patterns. Recognition is performed in

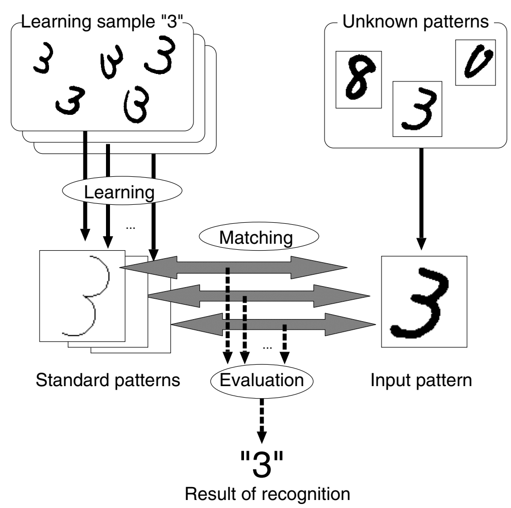

Figure 1. Overview of the recognition system

Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003)
0-7695-1960-1/03 $17.00

two substages, that is, matching of each standard pattern with the input character by deforming the standard pattern, and measurement of how well those two are matched. Deformation is done by using the orientations of contours in the input and of lines in the standard. After deformation converges, the standard line pattern is dilated into the (thick) input character and the extent of matching is evaluated based on the strain in deformation and the similarity between the input and the dilated after deformed standard. The recognition stage will be described in detail in Sections 4 and 5.

3. Method of learning the standard patterns

Construction of standard patterns is performed category by category. An outline of our method of learning the standard patterns is shown in Figure 2. Standard patterns, which describe the structure of a particular numeral character, are formed by line patterns of width one. In general it is necessary to provide more than one standard patterns per category because of difference of the structures. Accordingly we classify the learning samples of a category into several subcategories based on their structures and learn the ranges of variations in each subcategory.

Each learning sample is first converted into its skeleton and then compared with the reference models. It is classified into the subset corresponding to the best matched reference model. After all learning samples are processed, samples corresponding to the same reference model constitute a cluster in Figure 2. Finally the statistics of variations are calculated within each cluster and the standard patterns are constructed.

Statistics are calculated for strokes composing characters as well as the spatial relations between two strokes. Here strokes are defined as the line sections partitioned by end points, corners, branching points, and crossing points. Orientation, relative effective length, and ratio of height to effective length are adopted as the statistics for each stroke. The vector connecting end points of two strokes, difference of stroke orientations, and the relative position of the crossing point of two strokes are adopted for the statistics describing relations of strokes. The definition of these features are in [14].

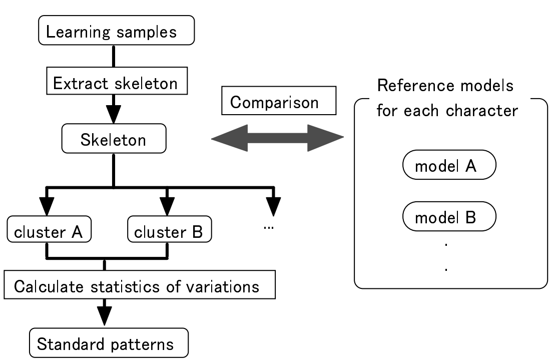

Figure 2. Procedure of learning standard patterns

Standard patterns are then constructed from the modes (most frequent values) of these statistics. They are composed of several strokes, accompanied with their statistics (Figure 3).

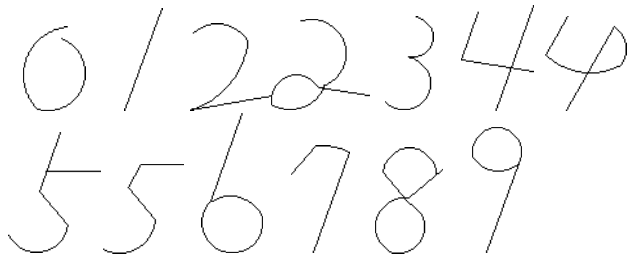

Figure 3. Derived standard patterns

4. Overall process of recognition

We describe here the overall process of recognition briefly, leaving the details in the next section.

Recognition is made by deforming every standard pattern to match with the input character as much as possible. Though standard patterns are line patterns, the input character is not thinned and dealt with as it is. The movement of the standard pattern is driven by a force $\vec{F}_{int}$ to keep its original shape and by the other force $\vec{F}_{image}$ to bring closer to the input character (see Figure 4). $\vec{F}_{int}$ is determined by the statistics learned in the learning phase. The details of deforming the standard pattern are in [14].

Such a movement is iterated and terminates when the standard pattern makes no more move, that is, gets into an equilibrium. When the standard pattern stops moving, the part of the standard pattern within the input character is dilated to fill the input character.

The cost of matching is then calculated based on both how heavily the standard pattern is deformed and how well

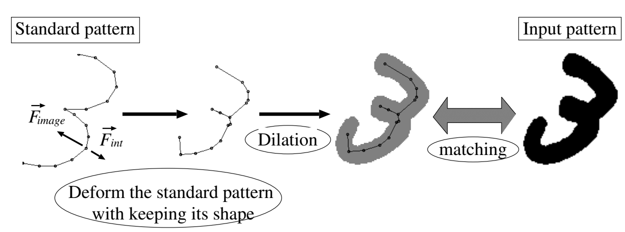

Figure 4. An illustration of matching

Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003)
0-7695-1960-1/03 $17.00 © 2003 IEEE

the dilated standard pattern matches with the input. Then the standard pattern matched the input with the minimum cost is identified, and its category is chosen as the category which the input character belongs to.

5. Improvements in the recognition stage

In our earlier paper, there were two main problems in the recognition stage. These two problems and the improvements of recognition method to solve them will be described in the following subsections.

5.1. Matching of the feather points

Deformation of the standard pattern was done by using the orientations of contours in the input and of lines in the standard. However, a feature point of the standard pattern sometimes failed in catching that of the input character. An example of failure is shown in Figure 5(a). In this case, the reason of failure in catching the crossing point is because the contours around the crossing point and around an end point of the input character have similar orientations (Figure 6). Thus the crossing point on the standard pattern was pulled from both the crossing point and the end point of the input character and located finally almost midway between them. The crossing and branching points often fail to match correctly in this reason.

In order to solve this problem, we supplement orientations with convex/concave information of the contour of the

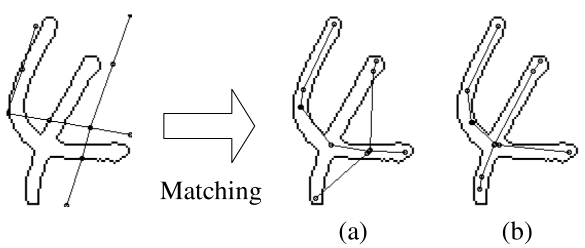

Figure 5. An example of failure in matching (a) and its improvement (b)

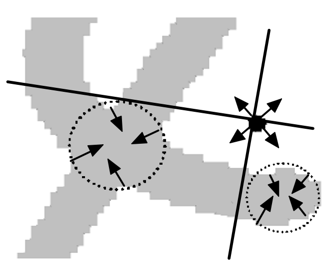

Figure 6. Orientations around the crossing point and the end point of the input character

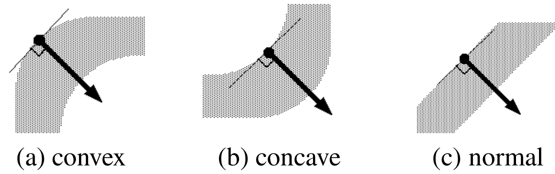

Figure 7. Orientations with convex/concave attribute

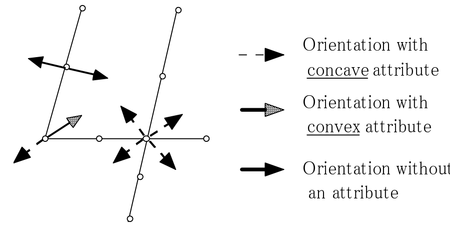

Figure 8. Orientations with convex/concave attributes on the standard pattern

input character (as in Figure 7) for matching of the feature points, such as corner points, branching points, and crossing points. The feature points on the standard pattern also have orientations with a convex/concave attribute (Figure 8). These points are pulled only from the contour points of the input character having the normal of a similar orientation and the same convex/concave attribute. For example, the contour points around the crossing point are surely concave. Therefore, discrimination of convex/concave contour points will improve the matching process so that the crossing point on the standard pattern is pulled only from the contour points of the input character having the normal of a similar orientation and the same convex/concave attribute.

A result of improvement is shown in Figure 5(b).

5.2. Evaluation of matching

In the previous report, evaluation of how well the deformed standard pattern and the input character coincide with each other is done by thickening the standard pattern with the average width of input and measuring both the distance from the thickened standard pattern to the input character and the distance vice versa. However, this evaluation method may fail to measure the difference correctly. Because this method evaluates only the shape of the thickened deformed standard pattern, the cost becomes small if its shape is similar to the shape of the input character. In other words, if an incorrect standard pattern is deformed close in shape to the input pattern, this evaluation method mistakes.

Therefore, it is necessary to consider the structure of the standard pattern in evaluation of matching. We improve the evaluation method so that it takes account of the structure

Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003)
0-7695-1960-1/03 $17.00 © 2003 **IEEE**

of the strokes. After deforming the standard pattern, we dilate it stroke by stroke inside the input character until it coincides the input. Then the result of dilation is used for evaluation of matching. The dilation and the evaluation will be described in the following two subsections.

5.3. Dilation of the deformed standard pattern

In order to measure how well the standard pattern R deformed by flexible matching coincides with the input character, considering the correspondences of strokes between them, the strokes of the standard pattern lying inside of the input character are dilated until they fill the whole input character (Figure 9).

We would like to check which stroke of the standard pattern is dilated into which part of the input character. For this purpose, a different label is assigned for every stroke of the standard pattern. These labels of strokes are spread synchronized with distance transform from the standard pattern.

In this study we utilize an unisotropic dilation. For this purpose, orientation attributes are set on both the standard pattern and the input character. They are one of the followings: horizontal(—), vertical (|), and two diagonals(/ and \ ). The orientation attributes on the strokes of the standard pattern are given for every line segment between its points on the standard pattern (Figure 10(a)).

The orientation attributes on the input character are given in two processes. First, one of four kinds of the orientation attribute is allotted to each pixel on the contour of input character using the contour orientation. Next, the orientation attributes on the contour are propagated into the interior synchronized with 3-4 distance transform (Figure 10(b)).

By using these two kinds of orientation attributes, dilation processing is controlled. The dilation is fast, where the orientation attribute of the standard pattern is the same as that of the input character. The dilation is slow, if the two orientation attributes are different. With this technique, the structures of the input character and the standard pattern tend to match appropriately. The speed control of dilation is realized by switching among 3-4, 5-7, and 7-10 distance transforms.

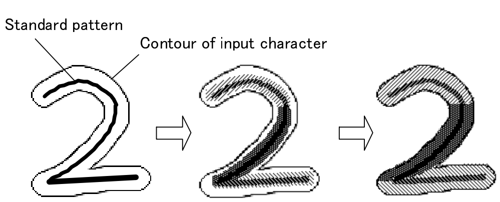

Figure 9. Dilation of the deformed standard pattern

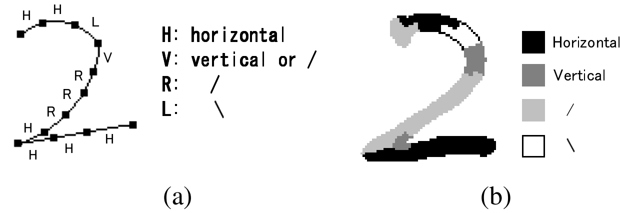

Figure 10. Orientation attributes on a standard pattern (a) and an input character (b)

5.4. New evaluation method

The cost of matching for evaluation is also changed reflecting the improvement described so far. It is calculated using the following three functions.

- Strain of the deformed standard pattern: $E(R)$
- Cost of dilation: $D(R)$
- Correspondence between the strokes after dilation: $C(R)$

$E(R)$ is the internal energy $E(R')$ which the standard pattern R has in its deformed shape R'. If it matches the input character with small deformation, then $E(R')$ is small.

Cost of dilation $D(R)$ is evaluated in three terms. The first is the sum $Dist(R)$ of each pixel got by the distance transform in the unisotropic dilation; the second is the sum $Del(R)$ of distances from the pixels of the standard pattern outside of the input character to the nearest pixel of the input character. If the standard pattern matches the input appropriately, then $Dist(R)$ and $Del(R)$ should be small. Moreover, for it is desirable that the distance values in each stroke are uniform, their standard deviation $\sigma(R)$ is also added. $D(R)$ is defined as the weighted sum of these three terms:

$$D(R) = w_{di} Dist(R) + w_{de} Del(R) + w_s \sigma(R) \quad (1)$$

Correspondence between strokes $C(R)$ is evaluated in two terms. One is the connection relation of the strokes after dilation, the other is correspondence of the strokes between the standard pattern and the input. If a stroke connection relation of the dilated standard differs from that of its original, a penalty $P(R)$ is given to $C(R)$. If the correspondence of the strokes between the standard and the input is one-to-many, another penalty $L(R)$ is given. Then, $C(R)$ is defined as:

$$C(R) = w_p P(R) + w_l L(R) \quad (2)$$

where w_p, w_l are the weights to be determined empirically.

From those three functions described above, the final cost function $Cost(R)$ for matching is

$$Cost(R) = w_e E(R) + D(R) + C(R) \quad (3)$$

where w_e as well as all the other weights were determined through a preliminary experiment.

The standard pattern for which this cost becomes the minimum is chosen as the recognition result for the input character.

6. Recognition experiments

We implemented the improvements described in the previous section. We also relearned the standard patterns and their statistics using 5116 samples collected from the IPTP handwritten ZIP code database [15] in total.

We made recognition experiments in order to check whether our improvements work well. First we tested on samples of 3000 characters, 300 per category chosen from the IPTP database but not used in the learning. We did not allow rejections. The recognition rate improved from 91.9% in our former report [14] to 96.0%.

We also tested on ETL-1 handwritten character database[16] with the standard patterns and their statistics learned from IPTP database as stated above. The result of experiments is summarized in Table 1.

Average recognition time per character was 5.6sec by using DELL Dimension 4500C computer (Pentium IV 2.2GHz used).

From these results we see the improvements are effective. Especially, the fact that for the unlearned ETL-1 database the recognition rate dropped only 0.8% shows adequacy of our learning method.

The recognition rate is still low for practical use, however, suggesting that we look into the causes of the misclassfications and make further improvements.

7. Conclusion and further studies

We improved our earlier method of recognition of handwritten numerals in two points and obtained better recognition results. Further studies needed is mainly in four points:

1. Possibility of using the dilation processing not only in matching evaluation but in deformation of the standard pattern.
2. The energy function derived from the stroke statistics, and consequently, the forces to deform the standard pattern.
3. An examination on the stroke features for more suitable standard patterns.
4. Test on other character databases.

Table 1. Result of recognition experiments

	#samples	#correct	#error	correct rate
IPTP	3000	2880	120	96.0%
ETL-1	10000	9518	482	95.2%

Acknowledgement

We thank Institute for Post and Telecommunications Policy, Japan and Electrotechnical Laboratory, Japan, for permitting us to use their databases.

References

[1] R. G. Casey and E. Lecolinet, “A survey of methods and strategies in character recognition,” IEEE Trans. on PAMI, 18(7):690–706, 1996.

[2] T. Matsui, T. Noumi, I. Yamashita, T. Wakahara, and M. Yoshimuro, “State of the art of handwritten numeral recognition in japan - the result of the first iptp character recognition competition -,” Proc. 2nd ICDAR, pages 391–396, 1993.

[3] L. Heutte, T. Paquet, J. V. Moreau, Y. Lecourtier, and C. Olovier, “A structural/statistical feature based vector for handwritten character recognition”, Pattern Recognition Letters, 19(5):629–641, 1998.

[4] J. Hu and H. Yan, “Structural primitive extraction and coding for hand-written numeral recognition”, Pattern Recognition, 31(5):493–509, 1998.

[5] H. Nishida, “newblock Automatic construction of structural models incorporating discontinuous transformations,” IEEE Trans. on PAMI, 18(4):400–411, 1996.

[6] J. Rocha and T. Pavlidis, “A shape analysis model with application to a character recognition system,” IEEE Trans. on PAMI, 16(4):393–404, 1994.

[7] T. Nagasaki, T. Yamamoto, and M. Nakagawa, “The behavior of dynamic relaxation in an elastic stroke model for character recognition,” Proc. 4th ICDAR, 1:16–22, 1997.

[8] S. Uchida and H. Sakoe, “A monotonic and continuous two-dimensional warping based on dynamic programming,” Proc. 14th ICPR, 1:521–524, 1998.

[9] S. Uchida and H. Sakoe, “An approximation algorithm for two-dimensional warping,” IEICE Trans. Inf. & Syst., E83-D(1):109–111, 2000.

[10] R. G. Webster and M. Nakagawa, “An interface-oriented approach to character recognition based on a dynamic model,” Pattern Recognition, 31(2):193–203, 1998.

[11] T. F. Cootes, C. J. Tailor, D. H. Cooper, and J. Graham, “Active shape models - their training and application,” Computer Vision and Image Understanding, 61(1):38–59, 1995.

[12] A. K. Jain and D. Zongken, “Representation and recognition of handwritten digits using deformable templates,” IEEE Trans. on PAMI, 19(12):1386–1391, 1997.

[13] M. D. Revow, C. K. I. Williams, and G. E. Hinton, “Using generative models for handwritten digit recognition,” IEEE Trans. on PAMI, 18(6):592–606, 1996.

[14] K.Kobayashi, K.Nakamura, H.Muramatsu, T.Sugiyama, and K.Abe, “Handwritten numeral recognition using flexible matching based on learning of stroke statistics,” Proc. 6th ICDAR, pp612–616, 2001.

[15] IPTP, IPTP CDROM1 : Database of Handwritten 3-digit Postal Code, IPTP Japan, 1994.

[16] ETL Character Database (ETL-1), ICDAR’93 Program Committee, 1993.