A Holistic Approach to Handwritten Numeral Pair Recognition Based on Generative Models of Numeral Pairs

Yanjie Wang[1] *Xiabi Liu*[1,2,*] *Yunde Jia*[1]

[1]Beijing Laboratory of Intelligent Information Technology, School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China

[2]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China

{wangyanjie,liuxiabi,jiayunde}@bit.edu.cn

Abstract

This paper proposes a model based holistic approach to recognition of handwritten numeral pairs. The models of numeral pairs are generated as the combinations of two corresponding numerals. Each numeral is modeled as a set of polygonal lines. We find out two bounding boxes of numerals in the skeleton of the input image and use them to determine the appearance of each numeral in the generated model of numeral pair. We then match the generated models with the input image to get the recognition result. The experiments on touching numeral pairs from NIST SD19 database confirm the effectiveness of the proposed approach. Furthermore, the segmentation can be completed based on the recognition result using our approach. We show this feature in the experiments.

Keywords: Character segmentation; Character recognition; Holistic approach; Recognition based segmentation

1. Introduction

Recognition of handwritten numeral strings is important for many applications, such as automatic processing of envelopes, cheques, and etc. The existing approaches to this problem can be classified into three categories: segmentation-then-recognition [1-3], segmentation-based-recognition [4-6], and holistic recognition [7-9]. In first two categories of approaches, numerals in the image of numeral string are segmented and recognized as one of ten classes. The recognition suffers from unstable segmentation for numeral strings written freely. Oppositely, holistic approaches recognize the numeral string as a whole, which do not care the classes of isolated numerals. Therefore, the segmentation is avoided in holistic approaches. Several holistic methods of handwritten numeral string recognition have been presented recently, such as data-driven error correcting output coding (DECOC) [7], the multi-experts approach [8], and the modular neural network [9].

In this paper, we propose a new holistic approach to numeral pair recognition based on generative models of numeral pairs. The input image is classified as one of 100 classes which are from '00' to '99'. Firstly, the models of these 100 classes are generated according to the models of single numerals and the normalized skeleton of the input image. Then the recognition result is obtained through matching the generated models with the normalized skeleton of the input image. The model generation of numeral pairs is a key problem in our approach, which is solved by considering a numeral pair as the combination of two corresponding numerals. Each of single numerals from '0'-'9' is modeled as a set of strokes which are represented as polygonal lines. Given an input image, we find out two rectangle boxes bounding the numerals in the normalized skeleton of the input image through searching a 2-dimensional discrete parameter space. By translating and scaling two models of single numerals to fit the detected rectangle boxes, the model of corresponding numeral pair is generated.

In order to test the proposed approach, we conducted the recognition experiments on touching numeral pairs from NIST SD19 database. The recognition rate of 93.6% was achieved. Compared with other holistic approaches, our method has the advantage of obtaining the segmentation result based on the recognition result, which is demonstrated in our experiments. This feature is useful for the verification of recognition results and the fuse of segmentation and recognition. To our knowledge it is not explored in the literature on the segmentation and recognition of numeral strings.

The rest of this paper is organized as follows. Section 2 introduces the generation method of the models of numeral pairs. Section 3 describes the classification method based on generated models of numeral pairs. The experimental

* Corresponding author. Tel.: +86 10 68913447, Fax: +86 10 86343158

results are discussed in Section 4, and the conclusions are given in Section 5.

2. Generative models of numeral pairs

The models of numeral pairs are generated according to the models of single numerals and the input images. That is to say, the models of single numerals are fixed, but those of numeral pairs change with input images.

Firstly, numerals from ‘0’ to ‘9’ are modeled as sets of polygonal lines. We write each numeral on a graphic’s tablet and capture each pen trajectory between pen-down and pen-up as a stroke which is represented as a polygonal line.

Then each model of numeral pairs from ‘00’ to ‘99’ is generated by determining the appearances of two corresponding numerals according to the normalized skeleton of the input image.

In this paper, the normalized skeleton of the input image is obtained in three steps: (1) the slant of the input image is corrected by using the method of Yamaguchi et al. [10]; (2) the image is skeletonized by the principle curve based algorithm of Liu and Jia [11]; (3) the size of the image skeleton is normalized in order that its width is 30 pixels. Fig. 1 illustrates this process through an example. Fig. 1a shows an original image. Fig. 1b-d respectively shows the result after the slant correction, skeletonization, and normalization.

(a) (b) (c) (d)

Figure 1. An illustration of the process of obtaining normalized skeleton of the input image: (a) the original image, (b) the image after slant correction, (c) the skeleton image of Fig. 1b, (d) the normalized skeleton.

Each numeral in the normalized skeleton of the input image can be thought to be bounded in a rectangle box. The examples of this kind of rectangle boxes are shown in Fig. 2. If we know the bounding box corresponding with each numeral in the normalized skeleton of the input image, we can translate and scale the model of single numeral to fit the corresponding rectangle box. Thus the instance of each numeral in the numeral pair is determined and the corresponding model of numeral pair is generated.

In order to find out two rectangle boxes in normalized skeleton of the input image, we consider two parameters: (1) the ratio of the width of the first numeral to the width of the numeral pair, let it be α ; (2) the horizontal distance between the first numeral and the second numeral, let it be β . Given the values of α and β , two rectangle boxes in the normalized skeleton of the input image are decided. As shown in Fig. 2, the width of the rectangle box corresponding with the first numeral ‘4’ is $w_1 = \alpha \cdot 30$. And the height of this rectangle box can be estimated as the height of the corresponding part of the input image, which is h_1 in Fig. 2. As for the rectangle box corresponding with the second numeral ‘6’, its width is $w_2 = 30 - \alpha \cdot 30 - \beta$, and its height, i.e. h_2 , can be estimated using the same method above.

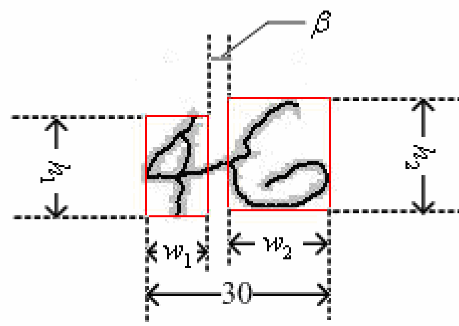


Figure 2. The examples of the rectangle boxes bounding the numerals in the image of numeral pair.

Consequently, we search a 2-dimensional parameter space constructed by all possible values of α and β to decide two rectangle boxes bounding two numerals in a given numeral pair. This 2-dimensional space is discretized for feasible computation. In fact, five values of α are used, which are $[0.3,0.4,0.5,0.6,0.7]$, and five values of β are used, which are $[-6,-3,0,3,6]$. Here the negative values of β are considered since two rectangle boxes overlap in the cases that two numerals are touched with each other. To sum up, we will select an optimal result from total 25 combinations of α and β .

For each combination of α and β , we compute two corresponding rectangle boxes according to which the models of single numerals are translated and scaled. The two transformed numerals are then combined to generate a candidate model of numeral pair. This means 25 candidate models will be generated for each of numeral pairs from ‘00’ to ‘99’.

Let $\mathbf{P_M} = \{\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_n\}$ be the point set in a candidate model, $\mathbf{P_I} = \{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_m\}$ be the point set in the normalized skeleton of the input image. We compute the following distance between each of 25 candidate models and the normalized skeleton of the input image:

$$D = \frac{1}{2}\left(\frac{1}{m}\sum_{i=0}^{m}\min_{j=1,\cdots,n}\left\|\mathbf{a}_i - \mathbf{b}_j\right\|^2 + \frac{1}{n}\sum_{j=0}^{n}\min_{i=1,\cdots,m}\left\|\mathbf{a}_i - \mathbf{b}_j\right\|^2\right) \quad (1)$$

The candidate model with the minimum distance is taken as the resultant model.

Fig. 3-4 illustrate the model generation by taking an input image of numeral pair '46' as an example. Fig. 3a-b shows the input image and its normalized skeleton, respectively. 25 candidate models of numeral pair '46' generated based on the skeleton are shown in Fig. 3c where the resultant model is indicated by the arrow. Fig. 4 shows resultant models of several other numeral pairs, which are generated for the same input image shown in Fig. 3a.

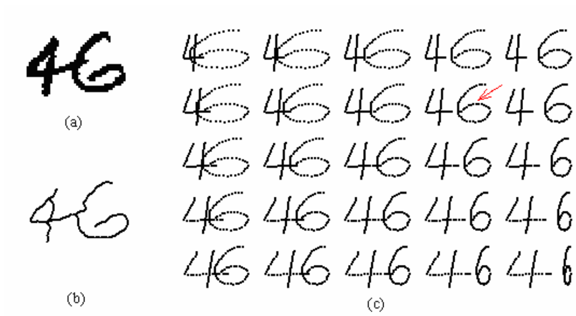


Figure 3. An illustration of numeral pair models generation: (a) the input image; (b) the skeleton of the input image; (c) the candidate models of numeral pair "46", where the resultant model is indicated by the arrow.

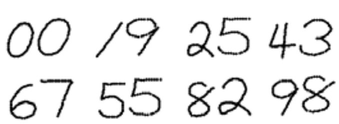


Figure 4. Generated resultant models of several other numeral pairs for the input image shown in Fig. 3a.

3. Numeral Pair Classification

Given an input image of numeral pair, 100 models of numeral pairs from '00' to '99' are generated using the method described in Section 2. Then the input image is classified into one of 100 classes through matching the points in each generated model of numeral pair with those in the normalized skeleton of the input image.

In order to improve the computation efficiency, we uniformly sample points in the model of numeral pair and the skeleton of the input image to get two subsets of points: $\mathbf{F_I} = \{\mathbf{a}'_1, \mathbf{a}'_2, \cdots, \mathbf{a}'_{m'}\}$ and $\mathbf{F_M} = \{\mathbf{b}'_1, \mathbf{b}'_2, \cdots, \mathbf{b}'_{n'}\}$. Then the matching between $\mathbf{F_I}$ and $\mathbf{F_M}$ is solved using the point matching algorithm based on shape context patterns [12]. Let $\{(\mathbf{a}'_{ik}, \mathbf{b}'_{jk})\}, k = 1, \cdots, K$ be the computed matching, where $K = \min(m', n')$. We get the least-square solution of the affine transformation from the model of numeral pair to the skeleton of the input image. Let T be the arbitrary affine transformation, T^* be the computed affine transformation, then

$$T^* = \arg\min_T \sum \left[T(\mathbf{a}'_{ik}) - \mathbf{b}'_{jk}\right]^2 \qquad (2)$$

After matching, the distance between the model of numeral pair and the normalized skeleton of the input image is computed as the weighted sum of three terms: matching distance, shape distance, and matching cost.

We measure the matching distance as the Hausdorff distance between two shape context patterns, i.e.

$$D_h = \frac{1}{K}\sum_{k=1}^{K} H\left(T(\mathbf{a}'_{ik}), \mathbf{b}'_{jk}\right), \qquad (3)$$

where H denotes Hausdorff distance[13].

The shape distance between the transformed model and the normalized skeleton of the input image is estimated as

$$D_n = \frac{1}{2}\left(\frac{1}{m}\sum_{i=0}^{m} \min_{j=1,\cdots,n} \left\|T(\mathbf{a}_i) - \mathbf{b}_j\right\|^2 + \frac{1}{n}\sum_{j=0}^{n} \min_{i=1,\cdots,m} \left\|T(\mathbf{a}_i) - \mathbf{b}_j\right\|^2\right); \quad (4)$$

The matching cost reflects the 'amount' of affine transformation, which is

$$D_d = \frac{1}{n}\sum_{i=1}^{n} \left\|T(\mathbf{a}_i) - \mathbf{a}_i\right\|^2 . \qquad (5)$$

To sum up, the function for computing the distance between the model and the skeleton is

$$D = \varpi_1 D_s + \varpi_2 D_n + \varpi_3 D_d , \qquad (6)$$

where ϖ_1, ϖ_2, and ϖ_3 are weighs, and $\sum_{i=1}^{3} \varpi_i = 1$.

Using Eq. 6, the distances between the models of all numeral pairs and the normalized skeleton of the input image are computed. The numeral pair with the minimum distance is taken as the recognition result.

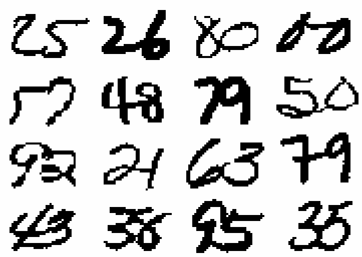
Based on the recognition result, two numerals in the input image can be segmented by assigning the points in the input image to one of generated numerals. We measure the distance between each point in the input image and each of generated numerals as the minimum Euclidean distance between this point and all points in this generated numeral. Each point will be assigned to one of two generated numerals according to the criterion that the distance between the point and the corresponding numeral is less than that for another numeral. In this way, the input image is segmented into two parts corresponding with two numerals. Similarly, the skeleton of the input image can also be segmented. We show some examples of segmentation results on input image and its skeleton in Fig. 7. This feature of 'recognition-then-segmentation' is useful for the verification of recognition result and the fuse of segmentation and recognition.

4. Experiments and discussions

The experiments of numeral pair recognition were implemented on the NIST SD19 database from which we manually collected 1000 test images of touching numeral pairs. Some examples of test images are shown in Fig. 5.

The values of weights in Eq. 3 are determined through careful experiments. The corresponding best recognition rate of 93.6% was achieved on these 1000 test images. Some of recognition errors come from the essential ambiguities which can not be solved reliably by even human, as illustrated in Fig. 6.

We also conducted recognition-then-segmentation experiments using the method described in the last paragraph of Section 3. Some segmentation results are shown in Fig. 7. Fig. 7a and Fig. 7c show normalized input images and corresponding skeletons, respectively. Fig. 7e shows the generated models corresponding with the recognition results and the skeletons shown in Fig. 7c by using the method described in Section 2-3. Based on the generated models, the segmentation results are obtained for input images and image skeletons, which are shown in Fig. 7b and Fig. 7d, respectively. In Fig. 7b and Fig. 7d, each of segmented parts in an input image is drawn in a specific grey level.



Figrue 5. The examples of the test images of touching numeral pairs.

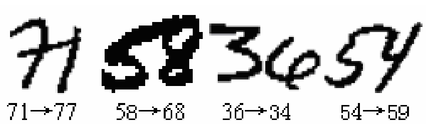


Figure 6. The illustration of essential ambiguities in misrecognized test images.

(a) (b) (c) (d) (e)

Figure 7. Examples of segmentation results: (a) normalized input images; (b) the segmentation results on the input image; (c) normalized skeletons of input images; (d) segmentation results on the skeletons; (e) numeral pair models generated for the recognition result.

5. Conclusions

This paper has proposed a novel holistic approach to numeral pair recognition based on the generative models of numeral pairs. The main contribution is the generation method of the numeral pair models. In our method, the models of numeral pairs are not fixed, but change with the input image. After the models of numeral pairs are generated according to the models of single numeral and the normalized skeleton of the input image, the recognition result is obtained through matching the generated models with the normalized skeleton of the input image. We conducted experiments of recognition on touching numeral pairs from NIST SD19 database. The experimental results show that the proposed approach is promising and effective.

Using our approach, the input image can be segmented into numerals based on the recognition result. We show this feature of recognition-then-segmentation in the experiments. To our knowledge this feature is not explored in the literature on the segmentation and recognition of numeral strings. We think that the ‘recognition-then-segmentation’ is useful for the verification of recognition results and the fuse of segmentation and recognition, which will be investigated in our future work.

Acknowledgments

This work was partially supported by the Grant 2006CB303103 from 973 Program of China and the Grant 2006Y1202 from Excellent Young Scholars Research Fund of Beijing Institute of Technology. The authors would like to thank the referees for their valuable comments and suggestions.

References

[1] Y.K. Chen, and J.F. Wang, "Segmentation of single- or multiple-touching handwritten numeral string using background and foreground analysis", *IEEE Trans. on PAMI*, 22:1304-1317, 2000.

[2] U. Pal, A. Belaid, and C. Choisy, "Touching numeral segmentation using water reservoir concept", *Pattern Recognition Letters*, 24:261-272, 2003.

[3] H. Fujisawa, Y. Nakano, and K. Kurino, "Segmentation Methods for Character Recognition: From Segmentation to Document Structure Analysis", *Proceedings of IEEE*, 80: 1079-1092, 1992.

[4] L.S. Oliveira, R. Sabourin, F. Bortolozzi, and C.Y. Suen, "Automatic recognition of handwritten numerical strings: a recognition and verification strategy", *IEEE Trans. on PAMI*, 24:1438-1454, 2002.

[5] C.L. Liu, H. Sako, and H. Fujisawa, "Effects of classifier structures and training regimes on integrated segmentation and recognition of hanwrtten numeral strings", *IEEE Trans. on PAMI*, 26:1395-1407, 2004.

[6] Y. Lei, C.S. Liu, X.Q. Ding, and Q. Fu, "A recognition based system for segmentation of touching handwritten numeral strings", *in Proc. IWFHR*, pp.294-299, 2004.

[7] J. Zhou, and C.Y. Suen, "Unconstrained numeral pair recognition using enhanced error correcting output coding: a holistic approach", *in Proc. ICDAR*, pp.484-499, 2005.

[8] X. Wang, V. Govindaraju, and S. Srihari, "Multi-experts for touching digit string recognition", *in Proc. ICDAR* , pp. 800-803, 1999.

[9] S.M. Choi, and I.S. Oh, "A segmentation-free recognition of handwritten touching numeral pairs", *Intl Journal of Pattern Recognition and Arificial Intelligence*, 15:949-966, 2001.

[10] T. Yamaguchi, Y. Nakano, M. maruyama, H. miyao, and T. Hananoi, "Digit classification on signboards for telephone number recognition", *in Proc. ICDAR*, pp.359-363, 2003.

[11] X. Liu, and Y. Jia, "A bottom-up algorithm for finding principal curves with applications to image skeletonization", *Pattern Recognition*, 38:1079-1085, 2005.

[12] X. Liu, Y. Jia, and Y. Wang, "An Eigenvector Approach Based on Shape Context Patterns for Point Matching", in *Proc. International Symposium on Communications and Information Technologies*, pp.455-458, 2006.

[13] D.P. Huttenlocher, G.A. Klanderman, and W.J. Ruchlidge, "Comparing images using the Hausdorff distance", *IEEE Trans. PAMI*, 15: 850-863, 1993.