21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Part-Based Method on Handwritten Texts

Song Wang Seiichi Uchida
Kyushu University, Japan
wangsong@human.ait.kyushu-u.ac.jp, uchida@ait.kyushu-u.ac.jp

Marcus Liwicki
DFKI, Germany
Marcus.Liwicki@dfki.de

Abstract

This paper reports a trial of handwritten text recognition by a part-based method. The part-based method recognizes individual characters by their parts without considering their whole shape. This realizes great robustness to severe deformations. This robustness is also effective for text recognition. Especially, for handwritten texts whose segmentation into individual characters is very difficult by deep touching and heavy slant, the part-based method still can recognize them because it does not request segmentation results to provide their whole shapes. Experimental results using digit sequences proved this robustness.

1. Introduction

Segmentation and recognition of handwritten texts is a classic but still important task for practical handwritten character recognition. For this task, two general approaches have been studied: segmentation-then-recognition and segmentation-by-recognition [1]. In the first approach, the segmentation is limited to predefined rules, thus when the query string image does not fit the rule well, the segmentation will be failed. In the second approach, also called analytical approach [2], the recognition process is conducted for every possible segmented candidate and then it finds the optimal choice. Consequently, they also require a good segmentation candidates for recognition.

Recently, many researchers employed the local feature detection method on character images. By using the local feature to describe the character, a flexible matching and robustness against image degradation can be obtained. In [3] this kind of method was used for word spotting and in [4] it was used for handwritten character segmentation. In [5] this method was first used on single handwritten digit recognition, which was also called the part-based method. With a simple framework, the part-based method for single handwritten digit recognition achieved a promising recognition rate. However, above trials never tried the part-based method on handwritten texts directly, in which the segmentation and recognition of the handwritten words are both conducted.

In this paper, we will propose a part-based method for handwritten texts. Our method is supposed to have the following merits.

- Compared with the segmentation-based-on-recognition methods, the part-based method does not require accurate segmentation, such as [6]. This is the most important merit of the proposed method, that is, its robustness to difficult segmentation situations.
- The proposed method inherits the robustness of the part-based method. Therefore it can deal with severe character deformation, such as heavily touching, overlapping, slanted writing style, and rule line distortion. All of those deformations are common to see in the handwritten texts and which are difficult for the conventional method to deal with.

This paper will concentrate to show the above merits of the proposed method. Although the recognition performance of the part-based method is often not better than the above conventional methods *for well-written inputs*, we can find the advantage of the part-based method on severe cases of the handwritten texts.

2. Methodology

The basic idea of the part-based recognition is to decompose a query handwritten text into parts, i.e., local areas. Each part is represented as a feature vector, called *local feature*. Then, each part is recognized as a class. This step is later called *feature-level recognition*. Finally recognition results of all parts are aggregated to provide the text recognition result.

In our method, the speeded-up robust features (SURF) [7] is used to extract the local feature whereas

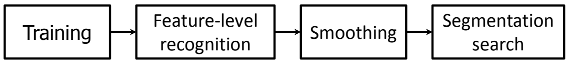

Figure 1. The framework of the part-based method on handwritten digit texts.

any local feature (e.g., SIFT [8]) can be a good alternative. SURF is a rotation-invariant and scale-invariant feature by automatically changing the size and the orientation of its local area where a feature vector is calculated. This is very good property especially for handwritten text recognition because the sizes and the rotation angles (i.e., slant) of individual characters are different. In the following however, we do not make SURF as a rotaton-invariant feature (by fixing the local area to be upright) because perfect rotation-invariance makes many confusion (e.g., "6" and "9"). A similar remedy can be found in [9], where not only rotation but also scale are fixed.

Figure 1 shows the overall framework of the part-based method on handwritten digit texts. Hereafter, we consider handwritten digit sequences as an example of handwritten text. The proposed method is comprised of four steps: training, feature-level recognition, smoothing, and segmentation search. Their details are described as follows.

2.1 Training

In this step, a reference keypoint database is created by extracting and storing SURF keypoints from training samples. Each training sample is a digit image. For each reference keypoint, its digit class label is attached. The reference keypoints are used for recognizing query digit string images as described later.

2.2 Feature-Level Recognition

In this step, first the SURF keypoints are extracted from the query digit string image. Then for each query keypoint, its first nearest neighbor (1NN) in the SURF feature space is searched from the reference keypoint database by Euclidean distance, as shown in Fig. 2.

During the 1NN search, the 1NN of each individual class is also recorded for the latter smoothing step. Consequently, each query keypoint has ten 1NN reference keypoints of ten individual digit classes.

2.3 Smoothing

As shown in Fig. 3, before smoothing, the feature-level recognition results of each individual digit were mixed of many classes. It is difficult to find the correct

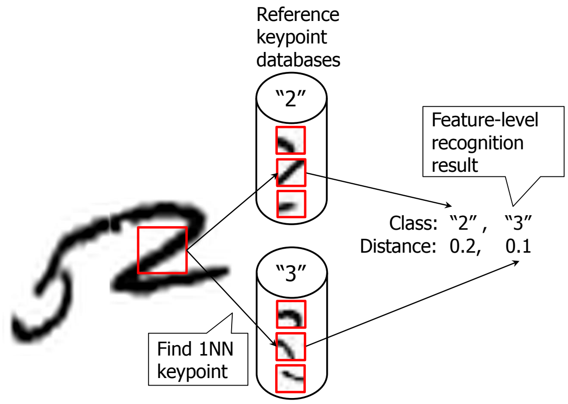

Figure 2. Feature-level recognition.

segmentation by using those results directly. We have to reveal the main class of a certain image area, and thus a smoothing process is employed. The smoothing tries to estimate a query keypoint by its nearest neighbors of geometric distance on the image. After smoothing the query keypoints from the same area of the image will have the same class label.

As shown in Fig. 4, first, for each query keypoint, its nearest neighbors of geometric distance on the image are found. This query keypoint is also called the target keypoint. Next, we simply add all the 1NN distances of the keypoints (including the target keypoint) together class by class. The class with the minimum total 1NN distance is the estimation result. Then the feature-level recognition result of the target keypoint is changed to the new estimation result. This process is similar to the part-based method of class distance in [9].

Figure 3 shows the smoothing result of a query image. After the smoothing, we can have keypoints of pure class in each digit. The following segmentation search is based on the smoothing result and the keypoints of pure class will reduce the incorrect segmentations.

2.4 Segmentation Search for Sequence-Level Recognition

A simple dynamic programming (DP)-based optimal segmentation search is conducted on the horizontal dimension. After the segmentation search, we estimate the class of each segmented area in the same way with the smoothing. It should be emphasized again that this is a very basic DP segmentation and in which only the vertical lines are used as the boundaries. Even with such a simple segmentation, our method is capable of recognizing severely connected, overlapping, and slanted characters.

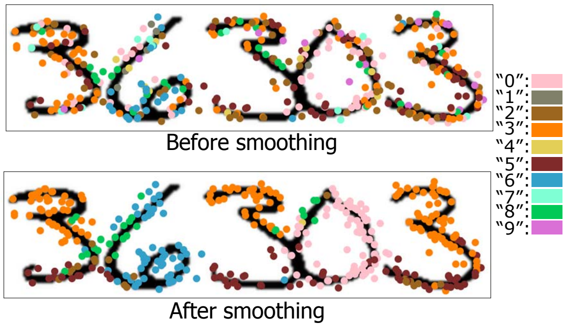

Figure 3. The effect of the smoothing step. The circle and the color shows the query keypoint and its feature-level recognition result. Only half of the query keypoints were showed.

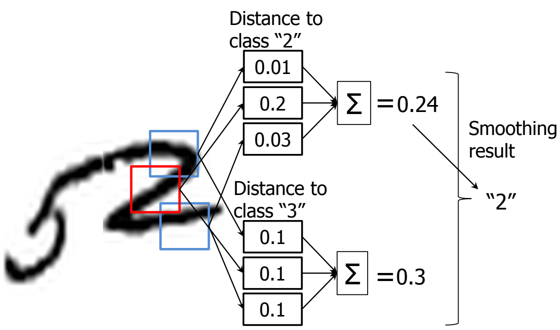

Figure 4. The process of smoothing. The red box on the query image represents the target query keypoint and the blue boxes represents the query keypoints neighboring to it. Originally, the target keypoint is recognized as "3" with the 1NN distance 0.1, and finally as "2" by considering the neighboring keypoints.

3. Experiment and Analysis

In the experiment, we used the BR sub-dataset of the CEDAR database. The segmented digit images in this sub-dataset were used for training. For the following analysis, we used the training dataset of 250 isolated digit images/class. We then selected 100 digit string images which are very severe for segmentation as our test set. Since all the test strings contain 5 digits, we set 5 as the maximum segmentation number in the segmentation search step. A character-wise accuracy is used as the final recognition rate.

3.1 Robustness

Compared with the conventional methods, the part-based method is robust on severely connected and degraded string images. Several processed images from the experiment are showed in Fig. 5. In the result (1), the second digit "0" missed the upper part and the third digit "5" was broken into upper part and lower part. Moreover, these two digits were totally connected. However, the part-based method can still have the correct result by using the parts of the digit. The results (2) and (3) were also recognized correctly, in which the digits were connected and slanted. In the results (4) to (6), the segmentation was not so promising, sometimes the boundary even went through the middle of the digit, but those images were all correctly recognized by the part-based method. This proves that the accurate segmentation is not necessary for the part-based method.

In the result (7), the image was distorted by a rule line and some other noise. However, the part-based method still made the correct segmentations and recognized four digits. In the result (8), there was severe overlapping between the first digit "4" and second digit "8". For this situation, most of the conventional methods may failed. However, our method was able to make correct segmentation and recognition. The similar thing happened to the result (9), the overlapped "9" and "4" were correctly segmented and recognized. The last three images of Fig. 5 all had a misrecognized digit, but this problem may be fixed by the following strategy of improving the part-based method.

3.2 Influence of Training Dataset

In the part-based method, with different training dataset, the feature-level recognition has different performance, and with a higher feature-level recognition rate the part-based method can achieve a higher final recognition rate. Usually, the larger training dataset used the higher recognition rate we can have. However, according to [10], for the part-based method, some reference keypoint may have negative influence to the recognition.

Table 1 shows the final recognition rate under the training dataset of different size. Because of the negative reference keypoints, the recognition rate dropped when the larger training dataset was used. If we remove those negative reference keypoints when using the larger training dataset, we can improve the recognition rate significantly and have less computation cost.

Again, it should be emphasized that the current part-based text recognition is not expected to outperform other conventional methods in the cases where individual characters/digits are well separated and well written. This is because the part-based method disregards whole character shape information, which is often important for recognizing characters. However, as shown in the

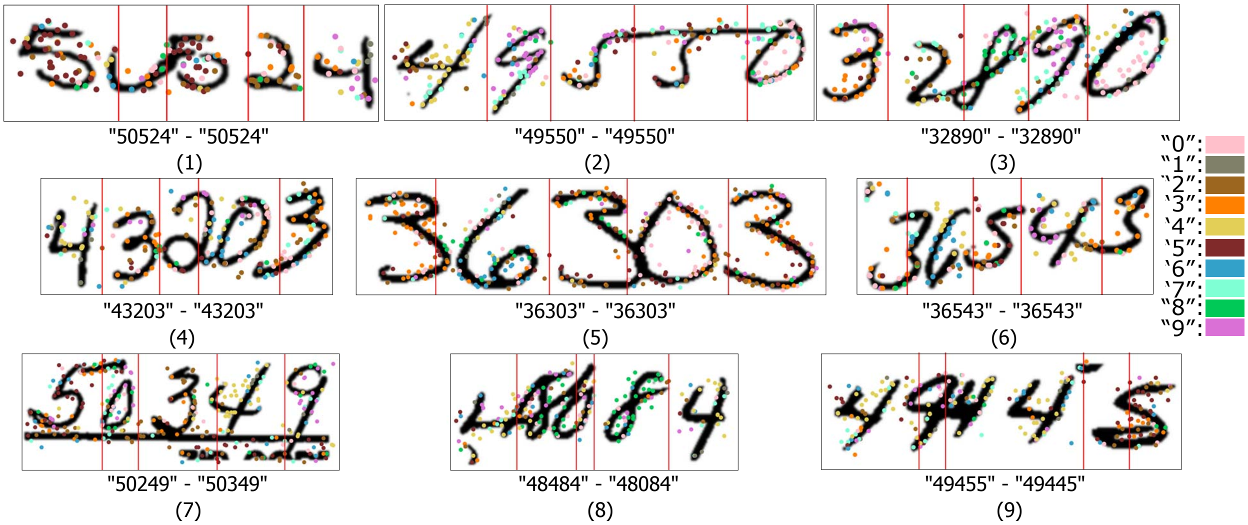

Figure 5. The processed images of the part-based method. The circle and the color shows the query keypoint and its feature-level recognition result. Only half of the query keypoints were showed. The first digit string below the image is the recognition result and second is the ground truth.

Table 1. Recognition Rate (%) under dataset different size (images/class).

Size by image number	100	250	500	1000
Recognition rate	64.8	69.0	67.4	64.8

above examples, the part-based method can be a good remedy for the case where the conventional method fails due to severe deformations. (In addition, recall the recognition rates of Table 1 are derived for very difficult 100 test sets selected manually.)

4. Conclusion and Future Work

The experiment above proved the merits of the part-based method on handwritten texts. Moreover, the part-based method has a potential to be improved by using better reference keypoint database. In the future, the performance of the reference keypoint will be studied and then some selection strategy may be employed for better recognition rate. More feature extraction methods will be examined in order to design a better local feature for the part-based method.

References

[1] L.S. Oliveira, R. Sabourin, F. Bortolozzi and C.Y. Suen, "Automatic recognition of handwritten numerical strings: a recognition and verification strategy," PAMI, vol. 24, no. 11, pp. 1438–1454, 2007.

[2] S. Madhvanath and V. Govindaraju, "The role of holistic paradigms in handwritten word recognition," PAMI, vol. 23, no. 2, pp. 149–164, 2001.

[3] Ravi. Shekhar and C.V. Jawahar, "Word Image Retrieval using Bag of Visual Words," Proc. DAS, pp. 297-301, 2012.

[4] L. Kang and D. Doermann, "Template based Segmentation of Touching Components in Handwritten Text Lines," Proc. ICDAR, pp. 569-573, 2011.

[5] S. Uchida and M. Liwicki, "Part-Based Recognition of Handwritten Characters," Proc. ICFHR, pp. 545-550, 2010.

[6] Javad. Sadri, Ching Y. Suen and Tien D. Bui, "A genetic framework using contextual knowledge for segmentation and recognition of handwritten numeral strings," Pattern Recognition., vol. 40, no. 3, pp. 898–919, 2007.

[7] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded Up Robus Features," Proc. ECCV, 2006.

[8] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," IJCV, vol. 60, no. 2 pp. 91-100, 2004.

[9] S. Wang, S. Uchida and M. Liwicki, "Comparative Study of Part-Based Handwritten Character Recognition Methods," Proc. ICDAR, pp. 814-818, 2011.

[10] S. Wang, S. Uchida and M. Liwicki, "Look Inside the World of Parts of Handwritten Characters," Proc. ICDAR, pp. 784-788, 2011.