Look Inside the World of Parts of Handwritten Characters

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

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

***Abstract*—Part-based recognition is expected to be robust in difficult handwritten character recognition tasks. This is because part-based recognition is based on aggregation of independent recognition results at individual local parts without considering their global relations and thus is robust against various deformations, such as partial occlusion, overlap, broken stroke, etc. Since part-based recognition is a new approach, there are still several open problems toward its practical use. For example, compared with entire images, local parts are more ambiguous, i.e., less discriminative. For better recognition accuracy and less computations, we need to know the characteristics of local parts and then, for example, discard less discriminative parts. The purpose of this paper is to conduct some experiments in order to observe and analyze how the local parts of multiple classes are distributed in feature spaces. By handling parts appropriately based on the analysis, we will be able to enhance the usefulness of the part-based method.**

***Keywords*-handwritten character recognition, local features, part-based recognition, distribution**

I. Introduction

Part-based recognition of handwritten characters [1]–[4] is a method using their local parts. Roughly speaking, in part-based recognition, each character is first decomposed into a set of local parts. Each part is represented as a feature vector, called local descriptor. The simplest local descriptor is a small block (i.e., a patch). In [1]–[4], scale-invariant feature transform [5] (SIFT) or speeded-up robust features [6] (SURF) descriptors were employed, instead. Then a part-based recognition process is performed to get a recognition result of each part. Finally, the results of all local parts are aggregated to derive a final recognition result by, for example, by the use of majority voting.

One important point of part-based recognition is that it disregards global features, such as the global relation among local parts. That is, the original position of each part is not considered in part-wise recognition. Thus, it may happen that a top-right part of a “5” is recognized as a bottom-right part of a “2”. Despite the pessimistic expectation of a low performance due to the lack of global features, past trials [1]–[4] showed promising recognition accuracies.

The part-based recognition method obtains the following unique advantages over traditional methods:

- Since the part-based method does not rely on any global features, it has a great potential to realize recognizers robust against not only mild isomorphic deformations but also against more severe deformations, such as partial occlusion, broken strokes, touching and overlapped strokes.
- It is possible to realize preprocessing-free recognizers, by describing local parts in a scale-invariant and/or rotation-invariant manner.
- It is also possible to realize segmentation-free recognizers. This is because the detection and description of local parts, and the succeeding part-wise recognition do not require any pre-segmentation. A word-level recognition result would be obtained by aggregating the part-wise recognition results using clustering, bag-of-feature (i.e., histogram), or other techniques.

For the part-wise recognition process, a database of the local descriptors, hereafter called reference keypoint database, must be prepared from a training image set in advance. The part-wise recognition process is then simply realized as 1-nearest-neighbor search in the database. The final recognition result of a query image is determined by aggregating the 1NN recognition results of all of its local parts.

Since the part-based method is a very new approach for character recognition, there are still several open problems to be tackled. First, its computational cost for large-scale 1NN search on the reference keypoint database should be reduced. Second, although the part-based method has already shown promising recognition rates [4], it still needs to be improved. Compared with an entire character image, the local parts are more ambiguous, i.e., less discriminative. Hence we could discard less discriminative parts in order to improve the performance.

In this paper, we focus on the above open problems and try to discover possible solutions by observing and analyzing the distribution of local parts from multiple classes in their feature space. For this purpose, several recognition experiments are conducted. From their results we can investigate the distribution of the local parts in qualitative and quantitative ways. This investigation results in important suggestions of selection strategies, such as editing [7] and condensing [8] in order to reduce the size of the reference database or to improve the recognition rate.

II. The Part-Based Method

A. Decomposition of Character into Parts

The part-based method relies on the detection of *keypoints* in character images and the description of a local part around

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.161

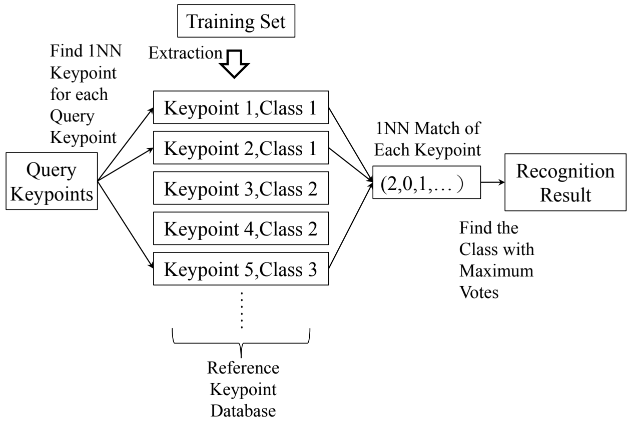

Figure 1. Process of the part-based method.

each keypoint. In this sense, one keypoint corresponds to a local part. Thus, hereafter, we will often use the term "keypoint" instead of "local part".

In this paper, SURF [6] is used as the keypoint detector and descriptor. SURF detects keypoints as local maxima of Hessian values in a scale space and then describes a local part around keypoints as a 128-demensional vector. The original SURF is scale and rotation invariant by adaptively controlling the size and direction of the local part. In contrast, our descriptor employs a local part with fixed size and direction, for simpler analysis. Thus, we will use a version of SURF, which is neither scale-invariant nor rotation-invariant.

B. Part-Based Recognition

As already outlined in Section I, and shown in Fig. 1, part-based recognition is applies the following steps. First, a reference keypoint database is created by extracting the keypoints from a training set of images. Each reference keypoint will be labeled by the class that it is extracted from. Second, the keypoints are extracted from a given query image. Third, as the part-wise recognition process, for each query keypoint the Euclidean 1NN reference keypoint is found. The recognition result of a query keypoint is the class label of its 1NN reference keypoint. Finally a voting process for aggregating the part-wise recognition results is applied. The recognition results of the query keypoints can be seen as votes of different classes. The class with the maximum votes wins and is taken as a final recognition result of the given query image.

III. Analysis of the Reference Keypoints

A. Good Keypoint and Bad Keypoint

The main viewpoint of our analysis is to know what is a good keypoint and what is a bad keypoint for handwritten character recognition. This is done by observing how the keypoint is discriminative. There are several possible ways to observe the discriminability. In this paper, we use the empirical class distribution of each keypoint. A keypoint is good if its class distribution shows a peak at its class of the original reference image. A keypoint is bad if its class distribution shows a peak at different class or if it is close to a uniform distribution.

The class distribution of each reference keypoint is estimated by considering how many times the keypoint is referred to as the 1NN of a query keypoint from each class (a verification set is used as the query images). For example, if a reference keypoint from class "1" is always referred by the query keypoints from class "1", the class distribution of the keypoint shows a strong peak at class "1". This reference keypoint is considered as a good keypoint. In contrast, if a reference keypoint is always referred by query keypoints from other classes, this reference keypoint may be seen as a bad keypoint.

B. Experiment for Estimating the Class Distribution

In this paper, the MNIST database was used for experiments and analysis. It contains isolated handwritten digits and thus its total class number is 10. For stable detection and description of keypoints, the images of MNIST were magnified four times after addition of 10-pixel margin (final size is 192×192). The size of fix-scaled SURF keypoint is 16×16. The average number of keypoints from a single image was 59.

An experiment for estimating class distribution of each reference keypoint was done. The training set of MNIST (about 6,000 images per class) was divided into two subsets; a small training set (50 images of each class) for extraction of a reference keypoint database, and a (rather large) verification set (5,000 images of each class). Then, by using every image from the verification set as a query image, a part-wise recognition by 1NN was performed and the referred time was counted for all reference keypoints.

Note that, the training set is quite small. This is because for a statistically reliable class distribution, we need enough number of referred times and thus need enough verification images. Also the computation cost will be a problem if we use a large training set. Note that the small training set, unfortunately, degrades the absolute recognition rates, which will be shown in the later sections. (In [4], about 1,000 images per class were used for achieving 93.8%).

C. General Observation

Figure 2 shows three images from training set. Each circle in the image corresponds to a reference keypoint. The colors of the circle represent the percentage of referred times of each class. From the images we can see that many reference keypoints are most referred by their respective class, such as Keypoints 1 and 2. However, there are still some reference keypoints, such as Keypoints 3 and 4, which are most

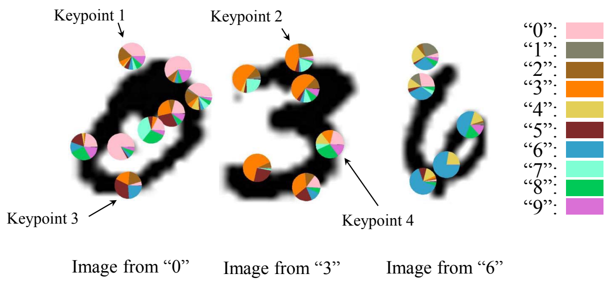

Figure 2. Distribution of referred times at reference keypoints. Note that for the sake of clarity, the number of keypoints were reduced.

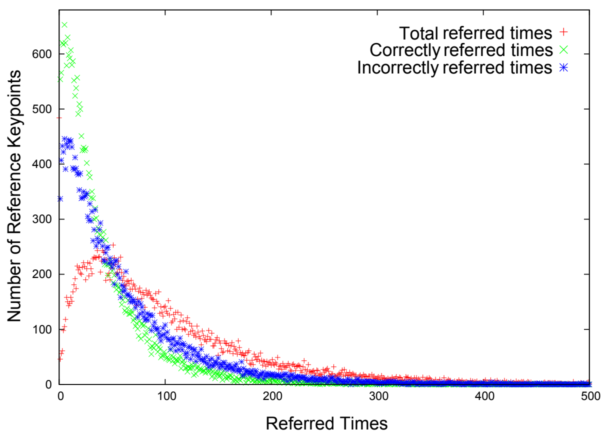

Figure 3. Distribution of reference keypoints as a function of referred times.

referred by other classes. In the remainder of this section we will analyze this fact more carefully.

Figure 3 shows the distribution of all reference keypoints by their total referred times. Total referred times is the referred times by all the classes of a reference keypoint. If this reference keypoint is referred by its class, it is correctly referred; if not, it is incorrectly referred. A correctly referred reference keypoint gives correct recognition result of a query keypoint. We can see that most reference keypoints are around 50 referred times. There are only a few keypoints with a total referred times over 500. The number of referred times indicates the importance of a reference keypoint during the recognition process.

In Fig. 3, the distributions of correctly referred times and incorrectly referred times are also shown. The two distributions have a similar shape, which means a reference keypoint has approximately the same probability to be referred correctly or incorrectly. In [4], the recognition rate of a single query keypoint was reported as about 50%, which can be seen as a proof of this observation.

D. Classification of Reference Keypoints

As shown in Fig. 4, the class distributions of the reference keypoints can be classified into eight types by using the top 2 classes with higher probabilities. The definitions of the eight types are the following:

- The distributions satisfying the 1st max class $>$ 30% and 2nd max class $<$ 10% are called *delta*. As can be seen in the Fig. 4, this distribution has a strong peak at a class. If the peak class is the class of the reference keypoint, the delta distribution is called *good-delta*; otherwise called *bad-delta*.
- The distributions satisfying the 1st max class $>$ 30% and 10%$\leq$2nd max class $\leq$ 30% are called *unimodal*. From Fig. 4, we can see that this kind of distributions have one mild peak. Like delta, there are *good-unimodal* and *bad-unimodal*.
- The distributions satisfying the 2nd max class $>$ 30% are called *multimodal*. Figure. 4 shows that the multimodal distribution has two or more peaks. If the correct class corresponds to the top, the distribution is called *good-multimodal*; otherwise, *bad-multimodal*.
- The distributions satisfying the condition 15% $<$ the 1st max class $\leq$ 30% are called *uniform*. From Fig. 4, we can observe that the probabilities of all classes are not so different.
- The distributions satisfying the 1st max class $\leq$ 15% are called *heavily uniform*. From the Fig. 4, we can see that probabilities of all classes are almost the same.

Table I shows the total number of reference keypoints of each type. From this table, the following facts are revealed:

- The most prominent type is the good-unimodal with about 38% occurrence.
- Keypoints with good-delta distributions are 10%. This indicates that some parts are stably surrounded by the parts from the same category, and, hopefully, discriminative.
- The total of stable keypoints is around 48%(= good-delta 10.35% + good-unimordal 37.99%). Even if we consider the keypoints from good-multimodal are also stable, the total is around 56%. The remaining 44% reference keypoints are still unstable and thus cause misrecognition by reference from a query keypoint of a different category. This fact also indicates that many local parts are ambiguous and less discriminative.
- About 15% of the keypoints have uniform or heavily uniform distributions and thus are less discriminative. This proves that in the feature space, there are some very confusing areas (where three or more classes are overlapping).
- Surprisingly, there are 21% bad-unimodal keypoints. This indicates that the world of parts is very unstable; in the feature space, these keypoints are surrounded by a "strong enemy" class being larger than the correct class.

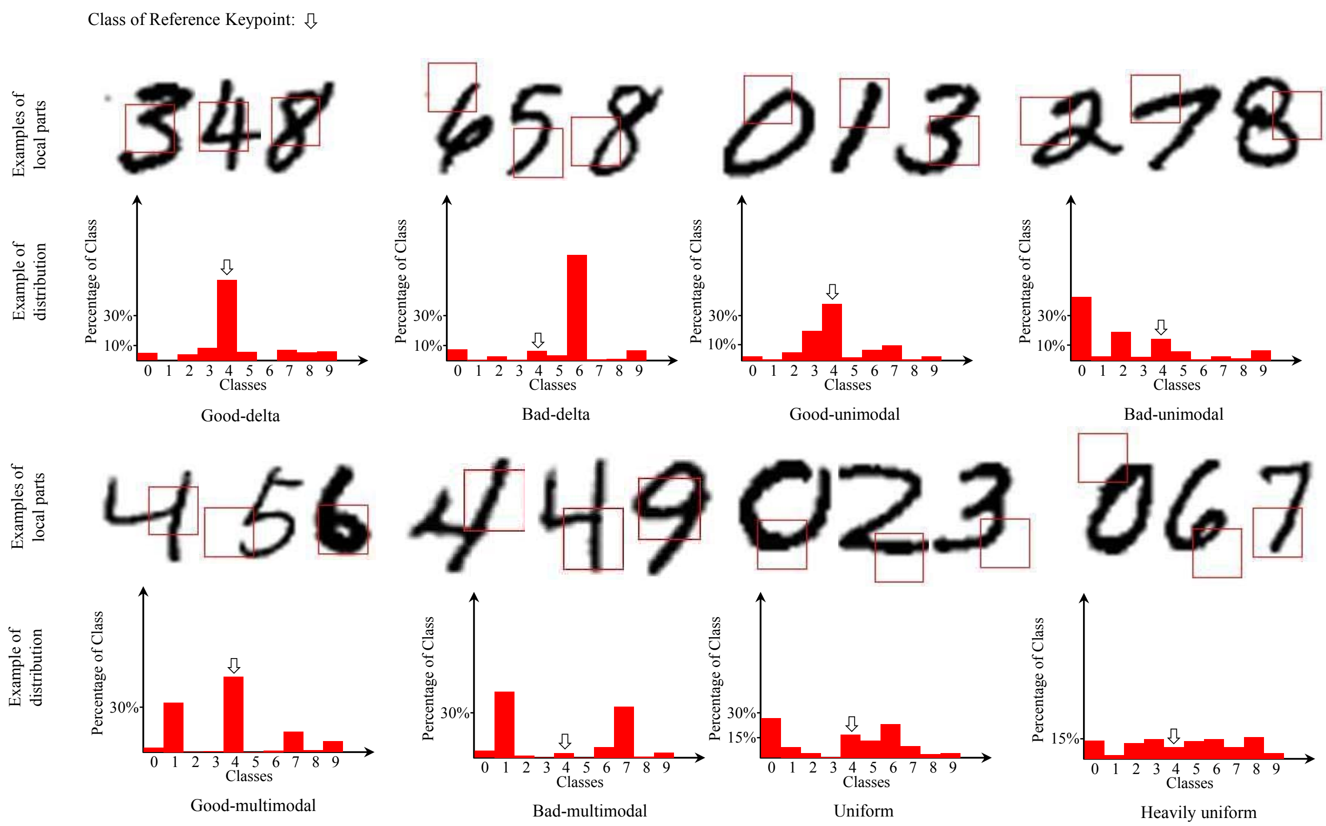

Figure 4. Eight types of reference keypoints. All of the examples are from the true class distributions.

Table I
TYPES OF REFERENCE KEYPOINTS.

Classification criteria	Type of reference keypoints	Total number of keypoints	Percentage in all	Total referred times	Percentage in all
1^{st} max class>30%	Good-delta	3,072	10.35%	274,084	9.28%
2^{nd} max class<10%	Bad-delta	198	0.67%	12,524	0.42%
1^{st} max class>30%	Good-unimodal	11,275	37.99%	1,139,730	38.59%
10%≤2^{nd} max class≤30%	Bad-unimodal	6090	20.52%	636,280	21.54%
2^{nd} max class>30%	Good-multimodal	2358	7.94%	227,043	7.69%
	Bad-multimodal	2102	7.08%	197,580	6.69%
15%<1^{st} max class≤30%	Uniform	4,098	13.81%	465,874	15.77%
1^{st} max class≤15%	Heavily uniform	487	1.64%	158	0.01%
Selected keypoints (referred times>10)					16,211
Unselected keypoints					13,469

In this sense, they are in a more severe condition than uniform and heavily uniform. (Note that the second max is not always the correct class.) In addition, these reference keypoints are surrounded by keypoints from a certain different classes, that is, they are outliers.

Figure 4 also shows three examples of reference keypoints for each type. The red square shows the area where a keypoint is described by SURF. Note that a Gaussian weight was imposed at keypoint description and thus the importance of the outer area in the square is lower than the center area. In other words, the effective area is smaller than the square.

In Fig.4, the keypoint from good-delta corresponds reasonably to a specific part of a class; for example, a "X"-shaped part is specific for "8". Such specific shapes cannot be observed from uniform and heavily uniform.

Table I also shows the total referred times of reference keypoints of each type. It can be observed that the referred times and the frequency of each type are almost the same. This indicates that the overall distribution of the reference keypoints and that of the keypoints from verification images

Table II
EXPERIMENT RESULTS.

	Reference database	Recognition rate
No selection	All keypoints	86.11%
Selection strategy 1	Selected keypoints	84.37%
	Unselected keypoints	52.92%
Selection Strategy 2	Selected keypoints have votes	85.62%
	Unselected keypoints have votes	39.19%

are similar.

IV. KEYPOINT SELECTION

Based on the classification of the reference keypoints, some selection strategies may be employed. Through the selection strategies we may find a way to reduce the size of reference database, hopefully, to improve the recognition rate. Intuitively, it seems reasonable to select the good-delta keypoints, the good-unimodal keypoints, and the good-multimodal keypoints. In order to ensure the reliability of the distributions, the keypoints of above types which satisfy total referred times ≤ 10 were discarded. Table I shows the selection result. (Note that from 16,705 reference keypoints, 484 were discarded by the above reliability condition.)

There are two strategies of using the selected keypoints. Strategy 1 uses only the selected keypoints, and all the unselected keypoints were discarded completely in the 1NN search for the part-wise recognition. In contrast, Strategy 2 still uses all the keypoints in the 1NN search. Its difference from the original one is that the unselected keypoints have no vote in the voting process. This means that even the unselected keypoint was referred as a 1NN by query keypoint, the class of this unselected keypoint would not get a vote.

Table II shows the result of the experiments where MNIST test dataset was used for query images. (About 1,000 samples per category.) In Strategy 1, although only half of the reference keypoints were used (selected keypoints) the recognition rate was almost the same as with no selection. This result clearly proves that good keypoints are important for the part-based recognition and if we use only the good keypoints, we can reduce the computations effectively. The third row in Table II, shows the performance of Strategy 1 if only unselected keypoints were used. Table II indicates that there is no large difference between the number of selected and unselected keypoints comparison between 84% and 53% shows how the quality of keypoints affect the recognition accuracy. The 30% difference in their accuracy shows the goodness of selected, i.e., good keypoints[1].

[1]On the other hand, we still can say that the 53% is reasonable. This phenomenon is due to the fact that by discarding good keypoints, the distribution of reference keypoints (or precisely, the balance of power defined by the Voronoi diagram) is changed and then several bad keypoints become good keypoints.

The recognition rate by the strategy 2 was also similar to the original recognition rate. The last row in Table II, also shows the performance of Strategy 2 if only unselected keypoints had votes. In strategy 2, there is no change in their power of balance because no keypoint is discarded. (See, Footnote 1.) Thus, the number of votes from bad keypoints are simply decreased. Actually we can use good keypoints from another training set instead of all the bad keypoints to make a combination set, this set may have a chance to beat the original set.

V. CONCLUSION

The purpose of this paper is to observe and analyze local parts extracted from handwritten characters in their feature space. The key idea of the observation is to classify the local parts into several types by their class distributions. That is, we did not consider that local parts do not distribute stably but distribute under some severe confusion due to their ambiguity. Through the observation of different types, it was shown that about 50% local parts lie in a confusing area, that is, are surrounded by local parts from different categories. Based on the observation, we employed two selection strategies in order to improve the part-based method. In experiments it was shown that almost the same recognition rate can be achieved when using only half of the reference local parts.

ACKNOWLEDGMENT

This research was partially supported by JST, CREST, JSPS and the Research Grant (No. 23300072) of The Ministry of Education, Culture, Sports, Science and Technology in Japan.

REFERENCES

[1] M. Diem and R. Sablatnig, "Are Characters Objects?," Proc. ICFHR, 2010.
[2] M. Diem and R. Sablatnig, "Recognition of Degraded Handwritten Characters Using Local Features," Proc. ICDAR, 2009.
[3] A. Garz, M. Diem, and R. Sablatnig, "Detecting Text Areas and Decorative Elements in Ancient Manuscripts," Proc. ICFHR, 2010.
[4] S. Uchida and M. Liwicki, "Part-Based Recognition of Handwritten Characters," Proc. ICFHR, pp. 545-550, 2010.
[5] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," IJCV, vol. 60, no. 2 pp. 91-100, 2004.
[6] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded Up Robus Features," Proc. ECCV, 2006.
[7] D. L. Wilson, "Asymptotic Properties of Nearest Neighbor Rules Using Edited Data," Systems, Man and Cybernetics, IEEE Transactions on , vol.2, no.3, pp.408-421, 1972.
[8] B. V. Dasarathy, J. S. Sanchez, S. Townsend, "Nearest neighbour editing and condensing tools-synergy exploitation," Patter Anal Appl, vol. 3, no.1 pp. 19-30, 2000.