Exploring the Ability of Parts on Recognizing Handwriting Characters

Takafumi Matsuo, Wang Song, Yaokai Feng, Seiichi Uchida
Kyushu University, Fukuoka, 819-0395, Japan

Abstract. This paper shows an experimental trial on part-based online character recognition. The purpose of the trial is to evaluate the discrimination ability of parts of a handwriting character, that is, a temporal pattern showing a character. Each part is a temporal segment of the character and the character is decomposed into a set of parts. Those parts are then represented as a "bag-of-features", which is a histogram showing how many parts similar to a specific representative part exist in the character. This representation totally disregards the global structure of the character and thus invariant to any deformation in the global structure, such as stroke order variations. Our experimental result has revealed that, without any global structure information, such as position of each part, relative relationship between parts, and temporal order of parts, 50-60% accuracies can be still achieved for 80 elementary Chinese character classes.

1. Introduction

Handwriting character is represented as a temporal trajectory on two-dimensional space (i.e., x-y space) and the trajectory is considered to be representing the entire character shape in two layers; that is, local structure and global structure. The local structure is a part of a handwriting character. In other words, the local structure is a segment of the temporal trajectory. For example, character "A" is considered as a set of five parts, "/", "/", "\", "\", and "-." (Here, we assume parts shorter than a single stroke and thus one "/"-shaped stroke is decomposed into two "/"-shaped parts.) In contrast, the global structure is the temporal and geometric arrangement of the local structures, i.e., parts. For example, the rule of forming "A" from the above five temporal trajectory segments is the global structure of "A".

The main contribution of this paper is to evaluate the discrimination power of local parts. Specifically, we will observe the recognition accuracy of handwriting characters *after discarding their global structure*. In our trial, each character is represented as an un-ordered set of parts. In the above example, we can consider "A" as a set of five parts, {"/", "/", "\", "\", "-"}, or equivalently, {"/"×2, "\" ×2, "-" ×1}. It is interesting to note that by discarding the global structure, now we need not care about the stroke writing order. For example, even if "A" is drawn in an unusual writing order "-" →"\" →"\" →"/" →"/", it has the same part-based representation {"/"×2, "\" ×2, "-" ×1}. It is also property of the part-based representation that "∀" will also have the same representation.

As shown by those examples, we can expect a merit and a demerit of the part-based representation of handwriting characters. The merit is that the part-based representation is totally invariant to deformations in global structure. Global structure is largely deformed by a change of stroke writing order, which is a serious problem in recognition of multi-stroke characters such as Chinese characters. Global structure is also deformed by a change in relative position of strokes. Consequently, handwriting character recognition with the part-based representation will be robust to such large deformations. On the other hands, the demerit is that two characters comprised of the same set of parts (e.g., "A" and "∀") may be misrecognized under the part-based representation.

To understand the role of our trial from a wider viewpoint, let us refer to recent general image pattern recognition research (such as Caltech101 trials), where part-based image pattern representation is a common way to achieve a high recognition accuracy. With the Naive-Bayes assumption and the Nearest Neighbor classifier (Boiman, et al., 2008), the method was able to reach the state-of-the-art performance on Caltech101. This kind of part-based method was soon introduced to the text extraction task (Ahmed, et al., 2012) and even the handwritten character "image" recognition (e.g., Diem & Sablatnig, 2009, Uchida & Liwicki, 2010 and Wang et al., 2011) with impressive performance. One problem of the above part-based trials is that they discard not only the global structure but also the composition restrictions of the parts. For example, a bicycle may be recognized as a tricycle because each part of the bicycle can find similar part from the tricycle (tire to tire, handle to handle, etc.).

Consequently, in order to solve this problem of the part-based method, it is quite natural that "bag-of-features (BOF)" (Csurka, et al., 2004) is employed as the part-based representation. BOF is a histogram and each bin corresponds to a representative part, called *visual word*. Thus, each image pattern is decomposed into parts, each part is classified into one of M visual words, and then a histogram with M bins is created by counting the number of parts classified to each visual word. This paper introduces this BOF representation, which has been utilized for general image pattern recognition, into temporal pattern recognition, i.e., recognition of handwritings. In other words, this is the first trial of part-based online character recognition (while the above part-based trials all focused on the handwritten character "image" recognition).

The rest of this paper is organized as follows. In the next section, we will describe our part-based representation of handwriting character. In fact, the representation is a straightforward extension of BOF representation for temporal patterns. Then, we will observe experimental results of the part-based recognition for 80 classes of elementary Chinese characters.

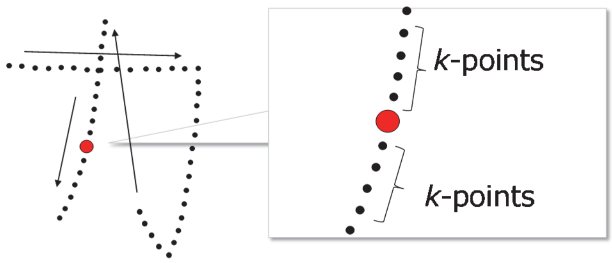

Figure 1. A part of a handwriting character.

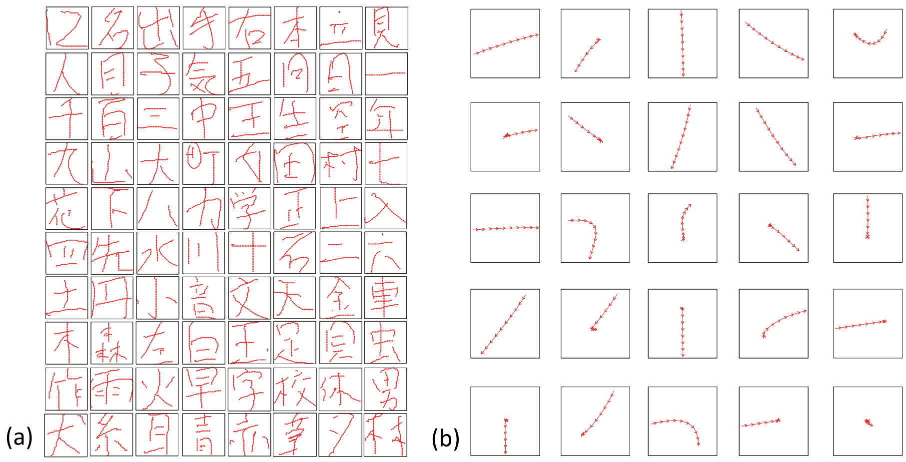

Figure 2. (a) 80 elementary Chinese character classes used in our experiment and (b) their 25 representative parts (*k*=5, *T*=50).

2. Part-Based Representation of a Handwriting Pattern

In our part-based representation, a handwriting pattern *P* is first decomposed into a set of parts. Each part is a short segment of the handwriting trajectory showing the entire character pattern. Figure 1 shows the definition of the part. Each part is represented as a segment comprised of (2*k*+1) consecutive points in the handwriting trajectory, where *k* is the radius of the part. Since a part can be defined at every point of the trajectory, a handwriting pattern with *T* points will be decomposed into *T* (overlapping) parts. Since each point *t* is defined by a two-dimensional coordinate (i.e., x_t and y_t), a part is represented as a 2(2*k*+1)-dimensional vector. (Note that points around the starting point or the ending point of each stroke need a special treatment since it is impossible to have *k* preceding or succeeding points. In this case, the starting (or ending) points are used repeatedly. For example, when *k*=5, the part centered at the third point of a stroke uses the starting point 4 times.)

Then, a set of *T* parts of the handwriting character pattern *P* are converted into a BOF, i.e., a histogram with *M* bins. Here, each bin corresponds to a representative part. Specifically, *M* representative parts are predetermined by clustering parts from many handwriting patterns from a training dataset. Since our part has the fixed dimensionality (2(2*k*+1)), we can easily determine representative parts by, for example, the orthodox k-means clustering method. BOF for *P* is created by finding the nearest neighbor representative part for each of *T* parts. If 10 among *T* parts have the same nearest neighbor representative part $m \in M$, the *m*th bin of BOF is set at 10. Clearly, the sum of entries at all the *M* bins is equal to *T*.

Figure 2 shows 25 representative parts learned from 80 elementary Chinese character classes, which will be used in the later experiment. Specifically, those 25 parts were selected by k-means clustering for parts from a training sample set of those 80 classes. In addition to straight line segments, corner segments are included as representative parts. Note that since they are temporal segments, they have their direction. Some similar parts, therefore, may have different writing directions. Also note that there is no “crossing” part since they are temporal segments and no contiguous temporal segment can form a cross.

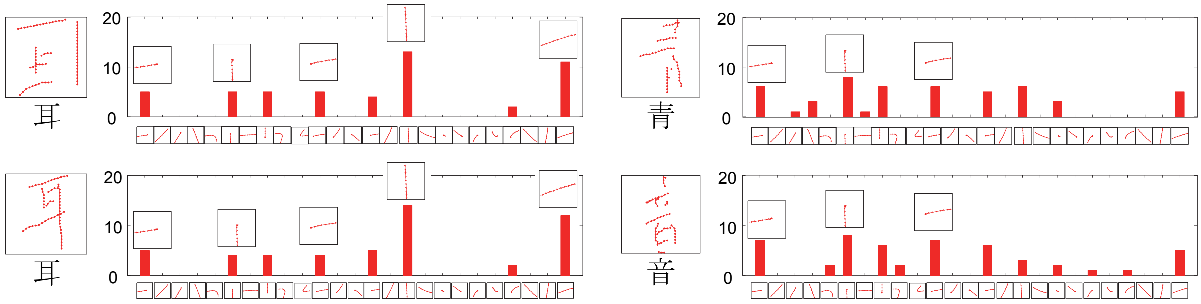

Figure 3. Examples of BOF (Bag-of-Features) representation.

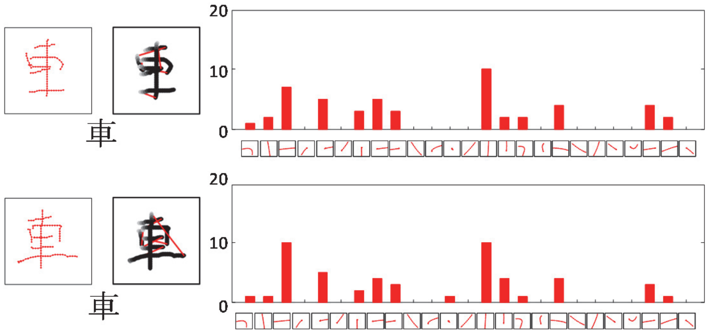

Figure 4. BOFs for patterns in different writing orders.

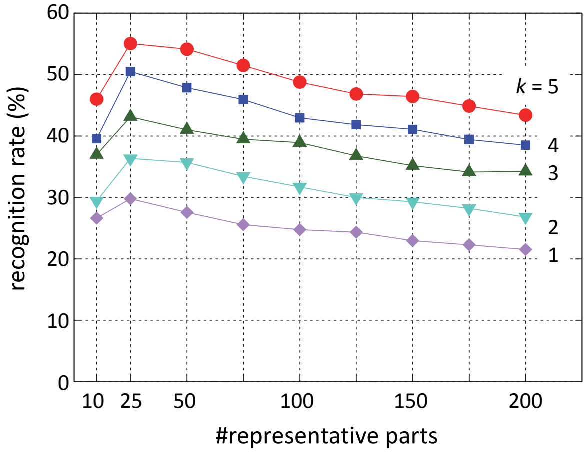

Figure 5. Recognition rate of 80 Chinese character classes under the part-based representation.

3. Experimental Analysis of Discriminative Power of Parts

Experimental setup. A recognition experiment was conducted for analyzing discriminative power of parts. 80 classes of elementary Chinese characters were used in the experiment. For each class, 120 samples by different writers were collected from well-known Kuchibue Chinese character database. The experiment was performed in leave-one-out manner; 80 samples (1 sample for each class) by a single writer were selected as the test sample set and the remaining 80×119 samples were used as the training samples for creating M representative patterns and also 80×119 reference BOFs. (That is, 199 reference BOFs were created for each of 80 classes.) The creation of the representative parts was done by k-means clustering for $50\times80\times119=476{,}000$ parts from the training sample set. This procedure was repeated 120 times by changing the writer for the test sample set and recognition results was averaged over those 120 trials.

Note that our recognition task is not so easy. Since the structure of the elementary Chinese character is rather simple, they are often confusing, such as "石" (stone) ↔ "右"(right) and "犬"(dog)↔ "大"(big). In addition, handwriting patterns in Kuchibue database were often not written in their "ordinary" writing order. Furthermore, their consecutive strokes are often connected with some ligatures and thus the number of strokes is not constant even for characters from the same class.

In the preprocessing step, each handwriting character pattern was resampled to have T=50 points. Thus, each pattern was converted into a set of 50 representative parts, i.e., a BOF. The number of representative parts, M, was changed for observing its effect in the recognition experiment. The radius of each part, k, was also changed, although it is expected that a smaller k (i.e., a shorter part) will make parts less discriminative and thus results in a lower recognition rate.

Recognition of each sample P was done simply by the nearest neighbor method; by searching 80×119 reference BOFs for the closest BOF with the minimum Euclidean distance to the BOF of P, the test sample P is recognized to the class of the closest BOF.

BOF representation. Figure 3 shows examples of BOF representation for M=25 and k=5. The two examples at the left side are the BOFs of two character samples from the same class "耳"(ear). It can be observed that both BOFs are almost the same in spite of large difference in their shapes. Since the large difference is a deformation in global structure (i.e., relative position between strokes), it does not affect the part-based representation by BOF. The two examples at the right side of Fig. 3 are the BOFs of two character samples from different classes, "青" (blue) and "音" (sound). It can be observed that those BOFs are similar. This similarity will cause misrecognitions between those classes.

Figure 4 shows BOFs of two patterns from the same class "車"(car). Those patterns are written in different stroke order (as depicted in the left side image where pen-up strokes are drawn in red lines) and thus recognition methods using global structure may fail to evaluate their similarity. Those BOFs show high similarity and thus show that the part-based representation by BOF is invariant to stroke order variations.

Recognition rates. Figure 5 shows recognition rates of 80 elementary Chinese handwriting characters under different numbers of representative patterns, M, and different radii of parts, k. First of all, it is somewhat surprising that we can recognize 80 different Chinese characters with 50% or more accuracies, "without their global structures"! Major misrecognitions were caused between ambiguous classes, such as "三" (three)↔"二" (two), "字" (character) ↔"学" (learn), "石" (stone) ↔ "右"(right), "犬"(dog)↔ "大"(big), and "王"(king) ↔ "土"(soil). Those class pairs have almost the same parts and thus are represented as similar BOFs, like the right example of Fig. 3.

As expected, a larger k (i.e., larger parts) achieves a better rate because larger parts can represent larger variations in their shapes and thus are more discriminative. In this experiment, the highest rate was achieved at M=25. That is, the representative parts of Fig. 2(b) are the most suitable in this task. Degradation of recognition rates by more representative parts might be caused by two reasons: (i) Existence of too similar representative parts makes BOF representation unstable. (ii) The curse of dimensionality.

4. Conclusion

This paper shows an experimental fact that it is not impossible to recognize temporal characters just by using their temporal segments, i.e., parts, without their relationships. In the experiment, each handwriting pattern was decomposed into parts and they are then represented as a histogram called Bag-of-Features. For the BOF representation, representative parts are determined in advance and each part is classified into one of those representative parts. Although BOF is a common method for image pattern recognition, it has been applied to temporal pattern, that is, handwriting pattern, in this paper and achieved around 50-60% recognition accuracy for 80 elementary Chinese character classes. This accuracy is still promising because it was achieved without any global structure of patterns. It was also promising that BOF representation is totally invariant to stroke order variations and other global structure deformations.

References

Boiman, O., Shechtman, E. & Irani, M. (2008). In Defense of Nearest-Neighbor Based Image Classification. Proc. Int. Conference on Computer Vision and Pattern Recognition – CVPR2008, pp.1-8.

Ahmed, S., Liwicki, M. & Dengel, A. (2012). Extraction of Text Touching Graphics Using SURF. Proc. Int. Workshop on Document Analysis Systems – DAS2012, pp. 349 – 353.

Diem, M. & Sablatnig, R. (2009). Recognition of Degraded Handwritten Characters Using Local Features, Proc. Int. Conf. Document Analysis and Recognition - ICDAR 2009, pp.221-225.

Uchida, S. & Liwicki, M. (2010). Part-Based Recognition of Handwritten Characters. Proc. Int. Conf. Frontiers in Handwriting Recognition - ICFHR2010, pp. 545-550.

Wang, S., Uchida, S. & Liwicki, M. (2011). Comparative Study of Part-Based Handwritten Character Recognition Methods. Proc. Int. Conf. Document Analysis and Recognition - ICDAR 2011, pp.814-818.

Csurka, G., Dance, C., Fan, L.X., Willamowski, J. & Bray C. (2004). Visual Categorization with Bags of Keypoints. Proc. ECCV Int. Workshop on Statistical Learning in Computer Vision.