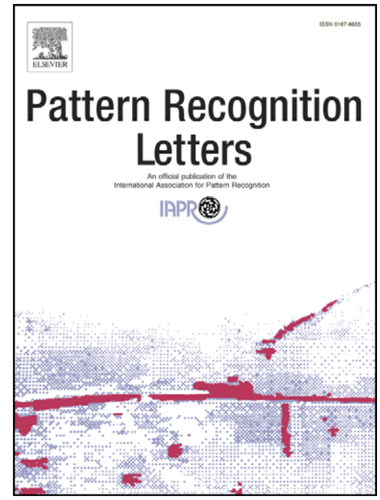
Accepted Manuscript

Global Feature for Online Character Recognition

Minoru Mori, Seiichi Uchida, Hitoshi Sakano

PII: S0167-8655(13)00141-4
DOI: http://dx.doi.org/10.1016/j.patrec.2013.03.036
Reference: PATREC 5683

To appear in: *Pattern Recognition Letters*



Please cite this article as: Mori, M., Uchida, S., Sakano, H., Global Feature for Online Character Recognition, *Pattern Recognition Letters* (2013), doi: http://dx.doi.org/10.1016/j.patrec.2013.03.036

Global Feature for Online Character Recognition

Minoru Mori[a,*], Seiichi Uchida[b], Hitoshi Sakano[c]

[a]*NTT Communication Science Laboratories, NTT Corporation,*
3-1 Morinosato-Wakamiya, Atsugi-shi, Kanagawa, 243–0198 Japan
[b]*Faculty of Information Science and Electrical Engineering, Kyushu University,*
Fukuoka-shi, Fukuoka, 819–0395 Japan
[c]*NTT Communication Science Laboratories, NTT Corporation,*
Seika-cho, Kyoto, Japan 619–0237 Japan

Abstract

This paper focuses on the importance of global features for online character recognition. Global features represent the relationship between two temporally distant points in a handwriting pattern. For example, it can be defined as the relative vector of two xy-coordinate features of two temporally separated points. Most existing online character recognition methods do not utilize global features, since their non-Markovian property prevents the use of the traditional recognition methodologies, such as dynamic time warping and hidden Markov models. However, we can understand the importance of, for example, the relationship between the starting and the ending points by attempting to discriminate "0" and "6". This relationship cannot be represented by local features defined at individual points but by global features. Since $O(N^2)$ global features can be extracted from a handwriting pattern with N points, selecting those that are truly discriminative is very impor-

[*]Corresponding author
Email addresses: `mori.minoru@lab.ntt.co.jp` (Minoru Mori), `uchida@ait.kyushu-u.ac.jp` (Seiichi Uchida), `sakano.hitoshi@lab.ntt.co.jp` (Hitoshi Sakano)

tant. In this paper, AdaBoost is employed for feature selection. Experiments prove that many global features are discriminative and the combined use of local and global features can improve the recognition accuracy.

Keywords: Online character recognition, Feature extraction, Global shape description, Feature selection

1. Introduction

In recent years, interest in online character recognition has increased because of the rapid adoption of smartphones and tablet PCs. Online character recognition is a better method than software keyboards for such small devices. An online character recognition method is needed that offers better recognition performance and is more robust against the variation in input environment.

In the long history of handwritten character recognition, feature extraction has been one of the most important topics. Any handwritten character is comprised of one or more strokes and thus has a peculiar structure unlike those of visual objects. Feature extraction as the representation of character strokes is, therefore, an important stage and affects the classification accuracy.

In this paper we tackle the unsolved problem of feature extraction; that is, how important are *global features* of character strokes for character recognition?[1] Prior to introducing the idea of global features, let us start with an

[1] This paper extends (Mori et al., 2012) with an improved algorithm for feature selection, additional empirical results, and an in-depth analysis of the performance of global features compared to local features.

explanation of *local features*. Since a character stroke is the trajectory (of a pen), its local features are defined as sequences of local parts of the trajectory. The most basic local feature is xy-coordinates at each point on the stroke (that is, the position of the pen-tip at each time step). Another popular local feature is the local direction feature (Tappert et al., 1990; Bahlmann, 2006), which is derived as the relative vector of two adjacent points. Moreover, dynamic features that consider handwriting movements such as the velocity (e.g. Plamondon et al., 1993) or other biomechanics (e.g. Plamondon and Guerfali, 1998) have been proposed (Bezine et al., 2003; Kherallah et al., 2008). These methods use Beta modeling and elliptical trajectory modeling to analyze handwriting, and use the parameters defined in each modeling method as features for online digit recognition. Therefore, they can be regarded as not only local features but also medium-sized shape descriptions.

In contrast, the global features examined in this paper capture the global structure of character strokes. The most basic definition of a global feature is a numerical representation of the relationship between two temporally separated points. An example is the positional relationship between the two end-points of "0".

We define a global feature as the relative vector between arbitrary point pairs on the stroke. In spite of its simple definition, global features have high potential to represent various key characteristics of character strokes. In the above example of "0", its starting and ending points are spatially close to each other. This closeness is represented directly by the global feature between those points, an ability not offered by local features.

Most online character recognition methods have used only local features such as the xy-coordinate feature and the local direction feature to represent character strokes. This might be derived from the fact that online character recognition methods are often based on dynamic time warping (DTW) or hidden Markov models (HMM) using sequences of local features. Both of them require that the problem have the Markovian property. Unfortunately, the use of global features clearly violates this Markovian constraint because they deal with the relationship between distant points. Consequently, to the best of the authors' knowledge, global features have not fully been investigated and thus the superiority of global features over local features has not been confirmed.

The main contributions of this paper can be summarized as follows: first, we prove that global features are often more discriminative than local features. This is proved through an experiment on automatic feature selection using the AdaBoost-based machine learning framework. In the experiment, an online numeral dataset is utilized for training a classifier. We also observe the global features that are selected and show that the global features that link separated points are certainly important for character shape description. Second, through a recognition experiment on a test dataset, we prove that the use of global features yields better recognition accuracy.

We examine the performance of global features for the online character recognition task throughout this paper. An important note is that if global features are useful in the online classification task, they will also be useful in the offline classification task. In fact, nowadays, offline recognition methods can employ techniques created for online recognition. For example, in (Mori et al., 1992), offline recognition is performed by extracting local online features of strokes by image processing. More dramatically, it is possible to use some stroke recovery method to convert an offline pattern into an online pattern.

The rest of this paper is organized as follows: Section 2 reviews related works in feature extraction. Section 3 explains the global feature approach. Experiments and results are reported and discussed in Section 4. Section 5 derives conclusions and future works.

2. Related Work

Most online character recognition methods use local features, such as xy-coordinate features and local direction features, as noted in Section 1. Those local features are sequential and thus traditional Markovian methodologies such as DTW and HMM can be applied. However, this fact does not prove that the handwriting process is purely Markovian. While writing a character pattern, we are usually watching not only the pen-tip but also the stroke shape of the written part. This process is totally non-Markovian and helpful in avoiding confusion between “0” and “6”.

In spite of the expected merit of global features, only a few trials have been made on utilizing them in the online character recognition task. One study examined the relative stroke position feature (Shin et al., 1999; Ota et al., 2007). This feature represents inter-stroke relationship and thus does not represent the stroke shape. Another trial is the star feature (Mandalapu and Krishna, 2007) which is based on an eight-directional representation (i.e., a quantized representation) of the entire character stroke; it can be seen as a online version of the classical Sonde method (Johnson, 1956) for offline character recognition.

The trial by Izadi and Suen (Izadi and Suen, 2009) is the work closest to our study. They proposed a feature, called relational context, that computes the relative pairwise distances and angles between arbitrary point pairs. Their trial, however, was merely a preliminary evaluation of the usefulness of global features. They used online patterns, each of which was re-sampled to just 6 points, and all ${}_6C_2$ pairs were used for extracting ${}_6C_2$ global features. As shown by the feature selection experiment in Section 4, our method has no need to use all such pairs in extracting useful global features. In other words, we reveal that each character class has its own important global structure; this important fact is not examined in (Izadi and Suen, 2009).

In offline handwritten character recognition (OCR), we can find features representing the relationship *spatially* distant points, i.e., pixels. One classical example is the so-called crossing features (Johnson, 1956; Glucksman, 1967), which describe the relationship between a target pixel and spatially distant strokes. Features that extract relative angle and relative position from spatially adjacent strokes have also been proposed (Mori et al., 1998).

3. Global features

We denote a handwriting pattern as the sequence, $p_1, \ldots, p_n, \ldots, p_N$, where p_n is the xy-coordinate feature of the n-th trajectory point. We assume that the number of trajectory points, N, is fixed for all patterns by using some resampling procedure. The coordinate feature, p_n, is considered as a local feature. Another local feature candidate is the local direction feature, which is defined as $p_n - p_{n-1}$. The experiments in Section 4 uses them as

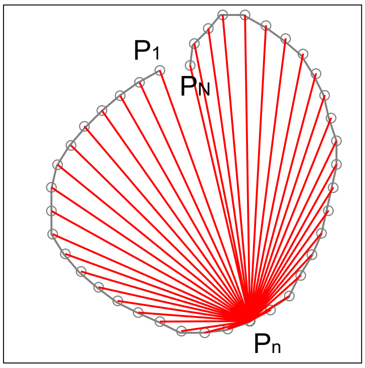


Figure 1: Global feature extraction.

typical local features.

The global feature used in this paper is simply defined as the relative vector between two temporally separated points n, n', that is, $p_n - p_{n'}$. Figure 1 shows the global features $p_n - p_{n'}$ from n to $n' \in \{1, \ldots, N\}$. It should be noted that the global features include the local direction features as a special case. Hereafter, we assume that $n > n'$ for choosing one of the two reciprocal features, $p_n - p_{n'}$ and $p_{n'} - p_n$. Thus, for a handwriting pattern with N trajectory points, there are ${}_N C_2$ possible global features.

Figure 2 (a) shows an example where the global feature $p_N - p_n$ is discriminative for two classes, "2" and "3". It is also shown that the other global feature, $p_n - p_1$, is almost the same in both classes and thus less discriminative; however, global features between more distant points, say $p_N - p_1$, become more discriminative and thus we can expect a set of global features to yield correct discrimination between "2" and "3". Figure 2 (b) shows the global feature $p_N - p_1$ of "0". We can easily expect that this global feature can well discriminate "0" from "6".

As indicated by Fig. 2 (a), different global features have different dis-

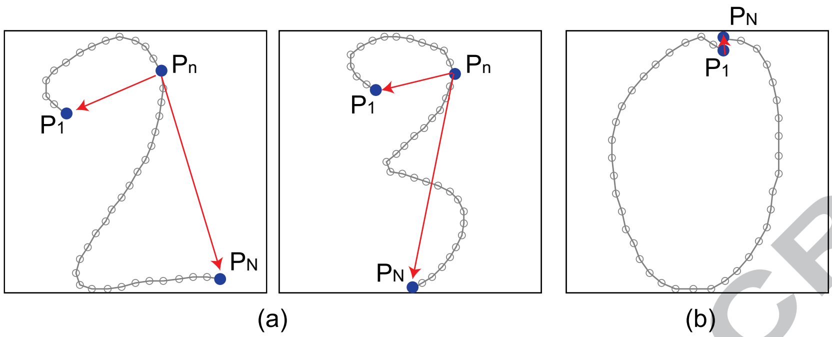


Figure 2: Potential of global features. (a): The local features around p_n do not have any discriminative power unlike the global features related to p_n. (b): The global features can regulate the relationship between two temporally separated points such as the starting and ending points.

criminative power. In other words, we do not need to use all of the ${}_NC_2$ global features. This observation suggests the necessity of some appropriate feature selection procedure. Section 4 adopts AdaBoost for feature selection (Freund and Schapire, 1997); AdaBoost has been utilized as not only a machine-learning method for training a classifier but also as a feature selection method. Thus, by using AdaBoost, we can select the global features that offer better classification and obtain a strong classifier for online character recognition.

4. Experimental evaluation of the importance of global features

In this section, the importance of global features is examined through two experiments; the first experiment addresses feature selection and the other recognition. They are detailed below.

4.1. Experimental setup

Online numeral samples from the UNIPEN database (Guyon et al., 1994) were used in our experiment. The database contains 31,386 samples of the 10 numeral classes ("0"-"9"). 90% of the samples (28,248) were used for training and the remaining 10% samples (3,138) were used for testing. In preprocessing, each sample was linearly normalized to 128×128 while keeping its original aspect ratio and then re-sampled at $N = 40$. Finally, N local features of xy-coordinates and ${}_NC_2$ global features including local direction features were extracted from each sample.

We set subclasses for each class to deal with changes in writing order and direction. We exploited the k-means algorithm to cluster the training samples. Finally, we used 22 subclasses for the 10 numeral classes; $2 \sim 3$ subclasses were prepared for each class. Therefore, we set $P = 22$ one-vs.-others classifiers in this paper.

4.2. Feature selection experiment

The first experiment uses AdaBoost (Freund and Schapire, 1997) to determine which features are most useful among all local features and global features in terms of discrimination. AdaBoost is an iterative learning scheme and iteratively selects a classifier, called a weak learner. After M iterations, we have M weak learners. The final result, the so-called strong classifier, is realized by weighting and summing the outputs of the M weak classifiers. In the experiment, the m-th weak classifier was designed to use only the single feature selected in the m-th iteration to provide a classification result. Thus, this iterative learning scheme can be considered as a feature selection process. A feature, as a weak learner, that is selected at an earlier step can be regarded as a more important feature for classification. Thus, under the condition that both global and local features are provided to AdaBoost, the selection of more global features than local features in the first few iterations means that the global features are experimentally confirmed to be more important than local features. The order of selection also reveals the importance of features. We, therefore, can observe which global features are important for each class from the selection result. The basic algorithm of AdaBoost is described in the Appendix.

Since AdaBoost is a two-class classifier, we train P "one-vs.-others" classifiers. In the following experiments of numeral recognition, P was fixed not at 10 but at 22. This is because several numeral classes had a large diversity due to changes in writing order and direction and so were divided into 2 or 3 subclasses. As the weak learner, we used a linear classifier based on the principal of nearest-neighbor to the weighted centroid for each of two classes (i.e., the target class and the "others" class). In our AdaBoost implementation, the best weak learner was selected from ${}_NC_2$ weak learner candidates at each iteration. Moreover, for creating better candidates, a random perturbation process was introduced for each feature as a weak learner. Specifically, the position of the weighted centroid was perturbed by adding a small random vector. By using 10 different random vectors, each weak learner candidate were evaluated 10 times and then one with the minimum error rate among ${}_NC_2$ weak learners was selected as the weak learner at that iteration. Here we mention the reason to introduce the random perturbation and its effectiveness. In our initial experiment, when the random perturbation was not used, the iteration for selecting features as weak leaners terminated at earlier steps in the learning stage. The reason is the simplicity of the weak classifier used. For exmaples, the learning on the xy-coordinate feature as the lower dimensional feature terminates at less than 30 steps for some classes. This gives the lower recognition accuracy for the xy-coordinate feature, and we could not fairly evaluate the effectiveness of our proposal. Therefore, we needed to obtain more features for stronger classifiers and exploited the random perturbation for more fair comparison among features and higher recognition results. The superiority among features based on the learning with no random perturbation was similar to that with the random perturbation.

Figure 3 shows the features selected in the first 20 iterations of the "0"-vs.-others classifiers. In Figure 3 (a), selected xy-coordinate features do not appear to express the characteristic shape of "0". In Figure 3 (b), local direction features accounted for about half the total number of features selected. This means that local direction features are not clearly superior to xy-coordinate features. In contrast, from Figure 3 (c), the global features representing the relationship between distantly-positioned point pairs were selected as the most important features in the top 3. The feature defined between the point near the start and the point near the end was also selected as one of the top 5 features. Among the 10 numerals, "0" has a circular shape and the start and end points are close to each other. This means that these relationships are important for characterizing the shape of "0" and our feature describes this global information effectively. Figure 3 (c) also shows that global features were more frequently selected for "0" within the first 20 iterations. This tendency confirms that most selected global features represent the relationship between pairs of separated points; that is, they

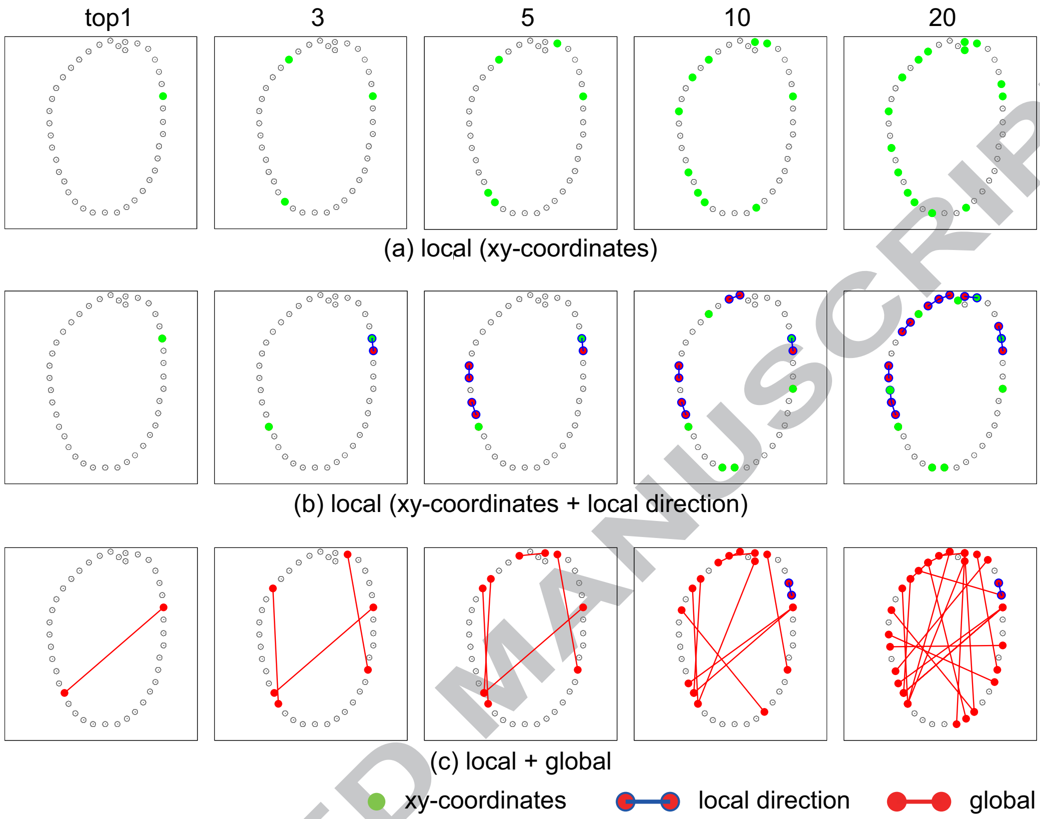


Figure 3: Selected features for "0".

represent that "0" is a circular pattern and has a large empty internal space.

Figure 4 visualizes the distributions of the global features selected in the first and second iteration under the AdaBoost framework. It seems that the two classes, "0" and the "others" class, are well discriminated by a simple linear classifier which validates the use of AdaBoost for feature selection.

Figure 5 shows the features selected in the first 10 iterations in each class. Global features were frequently selected not only for "0" but also the other classes. This fact is shown more clearly in Figure 6 which depicts the order

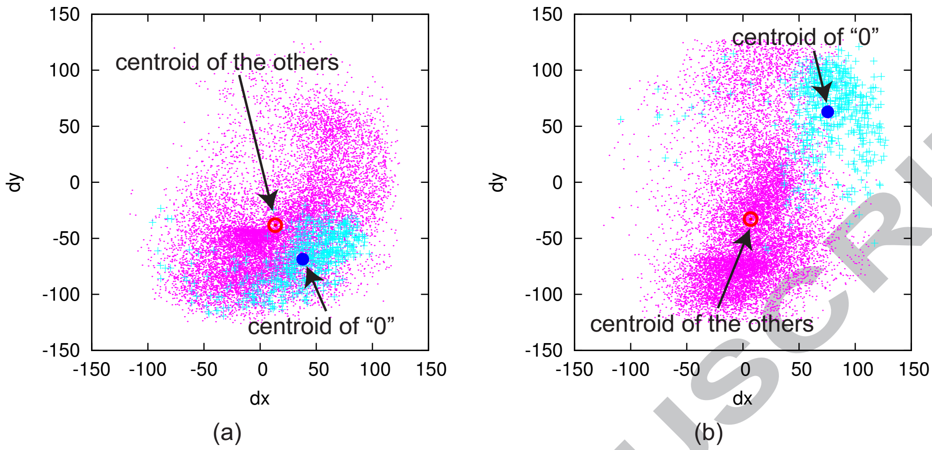


Figure 4: Distribution of the global feature vector. (a) Feature selected at the first iteration (leftmost feature of Fig. 3 (c)) for "0"-vs.-others classifier. (b) Feature selected at the second iteration.

of features selected in the first 20 iterations. Here the global feature excludes local direction. In most classes, global features are selected more frequently in the first 20 iterations than local ones. These results prove the stability and the discrimination power of global features.

4.3. Recognition experiment

In this experiment the P one-vs.-others classifiers are trained by AdaBoost using only global features, only local features, or both features together. Next, each test sample is subjected to the P classifiers and then classified into the class indicated by the subclass with the highest score. If the classifiers trained using the global features outperform those trained using the local features, an experimental proof of the importance of global features can be posited.

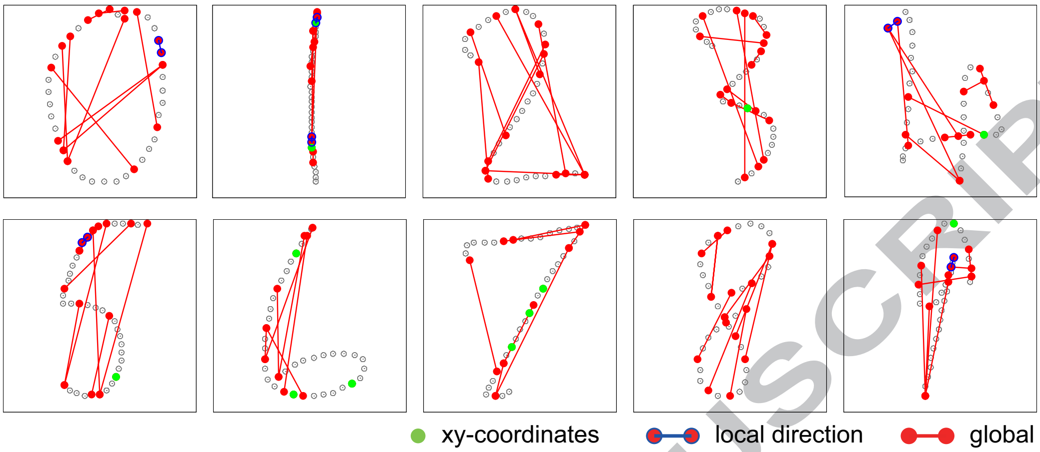


Figure 5: The first 10 features selected from among global and local features.

Figure 7 shows the classification accuracies during feature selection, that is, AdaBoost training. Global features outperformed the local features, and their combination yields better accuracy. These results suggest that the global features extract more important information of each class than the information captured by the local features. Here, note that the training iteration stopped when AdaBoost could not find any other weak learner that could improve overall accuracy. Also note that each feature could be selected repetitively. Therefore, the number of selected feature, M, can exceed the number of original features.

Figure 8 shows the recognition accuracy of the test samples when subjected to 10-class numeral classification. As expected from the accuracy of the training step (Figure 7), the combination of local and global features achieves the best accuracy (97.3%) with the use of just global features a close second. These results prove that our global features are more effective for online character recognition and have more discrimination power than local features. We note that our current result (97.3% accuracy) is 1.6% lower than the best past trial (98.9%) (Ratzlaff, 2003). The main purpose of this paper is to clearly prove the relative superiority of global features over local features. Thus, our experimental setup has a lot room for improvement. For example, we can introduce more subclasses as used in (Ratzlaff, 2003). Many post-processing techniques, such as 1-vs.-1 classification between an ambiguous class pair, also can be introduced.

Figure 9 shows examples that were recognized erroneously by the local features but recognized correctly by the global features. These examples show that global features can effectively extract crucial information for discriminating the focused class from the other classes, and provide robustness against the shape distortion common in handwriting. In contrast, local features lost their power of discrimination in the presence of shape distortion. For example, the local direction features in the upper part of the misrecognized "3" have similar directions to local direction features in the upper part of "5". On the other hand, our method describes the global structure of the upper part of "3" and identifies the global features that can effectively discriminate "3" from "5".

Figures 10 are examples of erroneous recognition by global features. The causes of errors are (a) different stroke order or different stroke direction that are not contained in the training samples used for each subclass, (b) similar shape to another class, (c) strong shape distortion, and (d) mislabeling. To deal with cause (a), more subclasses for each category are needed. A classifier that can discriminate ambiguous class pairs could offset cause (b). Cause (c) could be tackled by non-linear normalization techniques as pre-processing.

5. Conclusion

We have described the importance of the global features of character strokes and confirmed it in two experiments. A global feature is defined as a relative vector between arbitrary separated pairs of two points on a character stroke. Even though our proposal is very simple, it can represent various characteristics that can never be represented by local features. For example, a global feature can represent the very small separation between the start and end points of "0". Conventional online recognition frameworks, such as DTW and HMM, ignore or neglect "non-Markovian" features, i.e. global features, because their algorithmic principle is overly constrained. Thus, one contribution of this paper is to recommend that researchers re-consider the usefulness of "non-Markovian" features. Our experiments used an online character recognition task with numeral samples from the UNIPEN dataset. The feature selection experiment with the AdaBoost-based machine learning framework revealed that global features were more important in terms of discrimination power than local features; global features are crucial for identifying the characteristics of each class. The recognition experiment also proved that global features yield better classification accuracy not only for training samples but also for test ones.

Future works include the theoretical extension of our global feature. the global feature proposed in this paper is just a relative vector, i.e., a difference vector, between two xy-coordinates of a point pair; this can be extended to a difference vector of another local feature. Our feature can also be extended to deal with the relationship between three or multiple points. For the online recognition setup, introduction of the time warping function (other than

DTW) is also promising. For higher recognition accuracy, it is straightforward to prepare several classifiers for each class in order to deal with stroke order. To observe the important global features of different character sets such as Latin alphabets and Chinese characters is also interesting.

References

Bahlmann, C., 2006. Directional feature in online handwriting recognition. Pattern Recognition 39 (1), 115–125.

Bezine, H., Alimi, M., Derbel, N., 2003. Handwriting trajectory movements controlled by a beta-elliptic model. In: ICDAR'03. pp. 1228–1232.

Freund, Y., Schapire, R. E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55 (1), 119–139.

Glucksman, B. A., 1967. Classification of mixed-font alphabetics by characteristic loci. In: Dig. IEEE Comput. Conf. pp. 138–141.

Guyon, I., Schomaker, L., Plamondon, R., Liberman, M., Janet, S., 1994. Unipen project of on-line data exchange and recognizer benchmarks. In: ICPR'94. pp. 29–33.

Izadi, S., Suen, C. Y., 2009. Integration of contextual information in online handwriting representation. In: ICIAP'09. pp. 132–142.

Johnson, R. B., 1956. Indicia controlled record perforating machine. U.S. Patent 2741312.

Kherallah, M., Haddad, L., Alimi, A., Mitiche, A., 2008. On-line handwritten digit recognition based on trajectory and velocity modeling. Pattern Recognition Letters 29, 580–594.

Mandalapu, D., Krishna, S. M., 2007. A feature based on encoding the relative position of a point in the character for online handwritten character recognition. In: ICDAR'07. Vol. 2. pp. 1014–1017.

Mori, M., Uchida, S., Sakano, H., 2012. How important is global structure for characters? In: DAS'12. pp. 255–260.

Mori, M., Wakahara, T., Ogura, K., 1998. Measures for structural and global shape description in handwritten kanji character recognition. In: Document Recognition V. Vol. 3305. pp. 81–89.

Mori, S., Suen, C. Y., Yamamoto, K., 1992. Historical review of ocr research and development. Proc. IEEE 80 (7), 1029–1058.

Ota, I., Yamamoto, R., Sako, S., Sagayama, S., 2007. Online handwritten kanji recognition based on inter-stroke grammar. In: ICDAR'07. Vol. 2. pp. 1188–1192.

Plamondon, R., Alimi, M., Yergeau, P., Leclerc, F., 1993. Modelling velocity profiles of rapid movements: a comparative study. Biological Cybernetics 69, 119–128.

Plamondon, R., Guerfali, W., 1998. The generation of handwriting with delta-lognormal synergies. Biological Cybernetics 78, 119–132.

Ratzlaff, E. H., 2003. Methods, report and survey for the comparison of diverse isolated character recognition results on the unipen database. In: ICDAR'03. Vol. 1. pp. 623–628.

Shin, J.-P., Ali, M. M., Katayama, Y., Sakoe, H., 1999. Stroke order free on-line character recognition algorithm using inter-stroke information. IEICE Trans. Inf. & Syst. J82-D-II (3), 382–389.

Tappert, C. C., Suen, C. Y., Wakahara, T., 1990. The state of the art in on-line handwriting recognition. IEEE Trans. PAMI 12 (8), 787–808.

Appendix

The algorithm of AdaBoost (Freund and Schapire, 1997) is described as follows:

Given:

- Training set: $(x_1, y_1), \ldots, (x_s, y_s); x_i \in \chi, y_i \in -1, 1$
- Number of iterations: M

Initialize weights: $D_1(i) = \frac{1}{s},\ i = 1, \ldots, s$

For $t = 1, \ldots, M$:

1. Find the weak learner $h_t : \chi \rightarrow \{1, -1\}$ with minimum error rate $\epsilon_t$with respect to the distribution D_t. Here $\epsilon_t = \sum_{i=1}^{s} D_t(i) I(y_i \neq h_t(x_i))$ and I is the indicator function.
2. if $|0.5 - \epsilon_t| \leq \beta$ then stop. Here β is a threshold.

3. Choose $\alpha_t \in \Re$, here $\alpha_t = \frac{1}{2} \ln \frac{1-\epsilon_t}{\epsilon_t}$
4. Update $D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$. Here Z_t is the normalization factor $\sum_i D_t(i) \exp(-\alpha_t y_i h_t(x_i))$ and D_{t+1} is a distribution.

Output the strong classifier: $H(x) = sign(\sum_{t=1}^{M} \alpha_t h_t(x))$

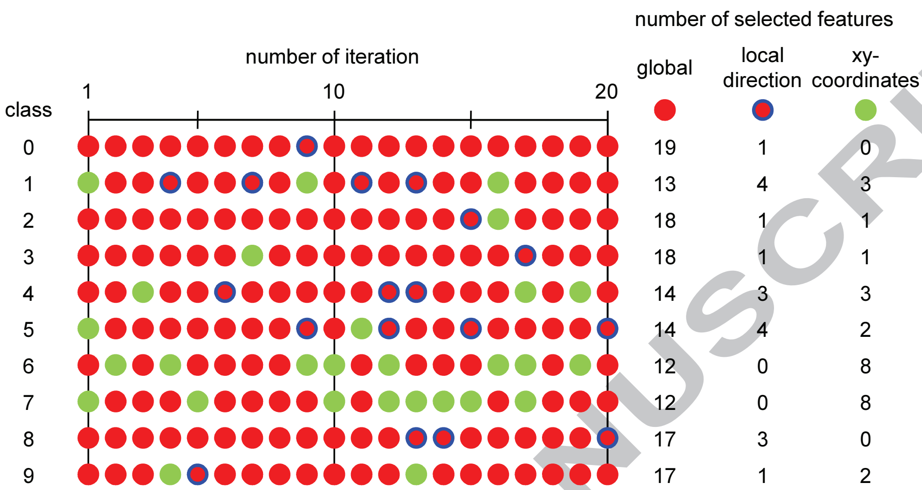
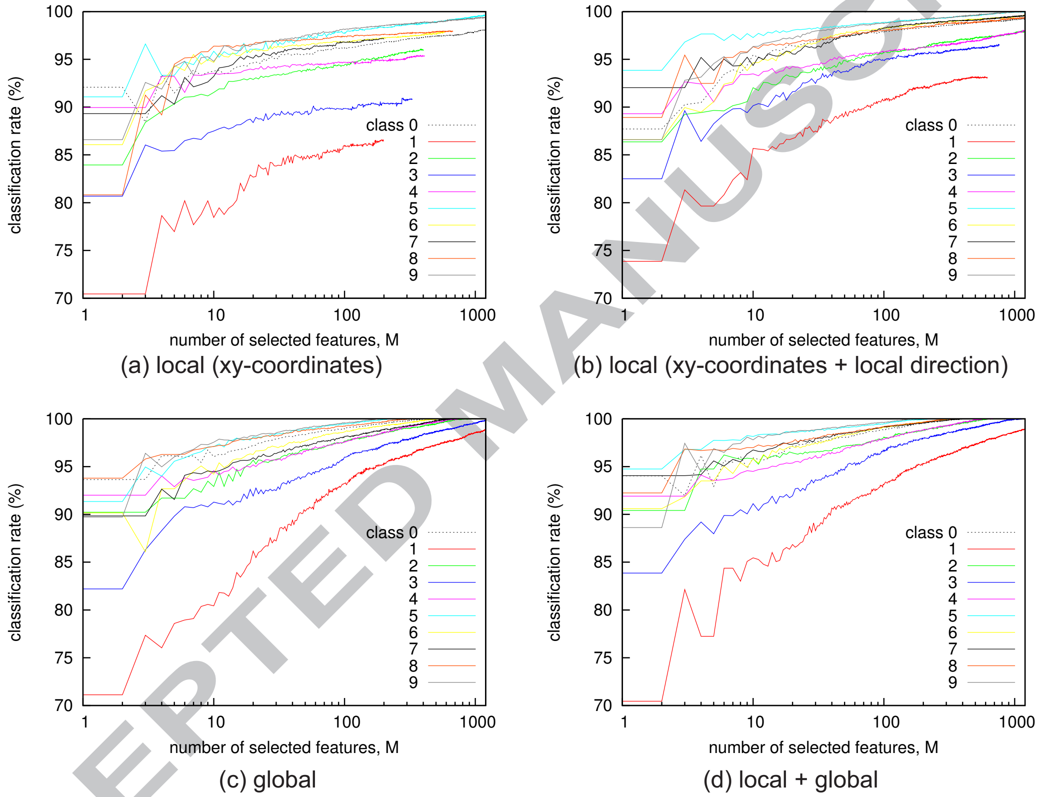


Figure 6: Features selected in the first 20 iterations.



(a) local (xy-coordinates)

(b) local (xy-coordinates + local direction)

(c) global

(d) local + global

Figure 7: Classification accuracy for training data by each one-vs.-others classifier.

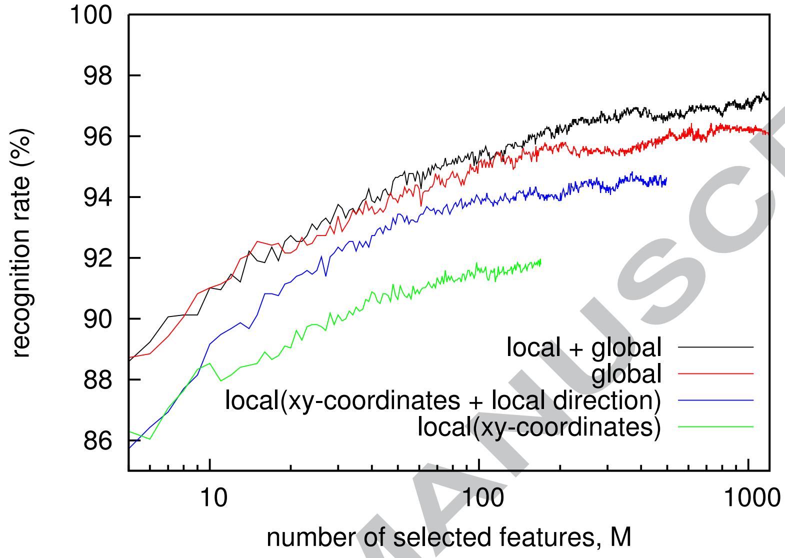


Figure 8: Recognition accuracy for test data.

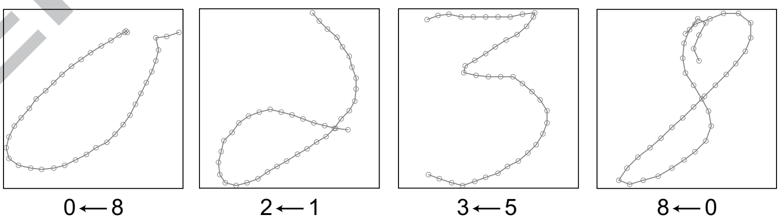


Figure 9: Patterns recognized erroneously and correctly by local and global features, respectively.

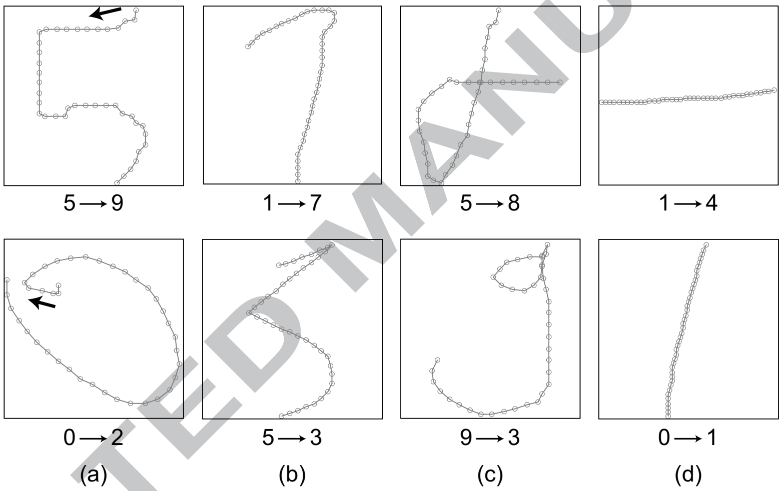


Figure 10: Patterns misrecognized by all four features.

We propose global features that extract global structure of online character pattern.

Global features represent the relationship between distantly-positioned points.

We validate global features through feature selection and recognition experiments.

Global features are more often selected as important classifiers than local features.

Global features achieve higher recognition rates than local features.