How Important Is Global Structure for Characters?

Minoru Mori[*], Seiichi Uchida[†] and Hitoshi Sakano[‡]

[*]*NTT Communication Science Laboratories, NTT Corporation, Atsugi-shi, Kanagawa, 243–0198 Japan*
E-mail: mori.minoru@lab.ntt.co.jp

[†]*Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka-shi, Fukuoka, 819–0395 Japan*
E-mail: uchida@ait.kyushu-u.ac.jp

[‡]*NTT Communication Science Laboratories, NTT Corporation, Seika-cho, Kyoto, Japan 619–0237 Japan*
E-mail: sakano.hitoshi@lab.ntt.co.jp

Abstract—This paper studies the importance of the features that represent the global structure of character strokes to character recognition. Most existing character recognition methods based on character stroke features utilize a set or a sequence of local features such as xy-coordinates and local direction of strokes. This is natural from the viewpoint that each stroke is a trajectory and thus can be represented as a sequence of local features. This viewpoint, however, has a clear limitation in that local features cannot deal with global structure directly. For example, the sequence of local features cannot deal with the fact that the two end points of character "0" should be close to each other. In this paper we propose a simple and novel global feature that describes the global structure of the character shape of each class. We prove the importance of the global feature through a feature selection experiment. Specifically, we show that the global features are more often selected than local features to enhance classification accuracy under the AdaBoost-based machine learning framework. Recognition experiments using online numeral data show also that the use of global features improves recognition accuracy.

Keywords-feature extraction, feature selection, online character recognition, global shape description

I. INTRODUCTION

Feature extraction has been one of the most important topics in the long history of handwritten character recognition. Any handwritten character is comprised of one or more strokes and thus has a peculiar structure unlike those of visual objects. Consequently, it seems natural to represent the character stroke in some way for extracting good features. In this paper we tackle the unsolved problem of feature extraction; that is, how important are *global features* of character strokes for character recognition? Prior to introducing the idea of global features, let us start with *local features*. Since a character stroke is the trajectory (of a pen movement), its local features will be defined by local parts of the trajectory. The most fundamental local feature is xy-coordinates at each point on the stroke (that is, the position of the pen-tip at each timing). In fact, handwritten characters are often represented by a sequence or a set of xy-coordinate features. Another popular local feature is local direction feature [1], [2], which is derived as the relative vector of two adjacent points.

In contrast, the global features examined in this paper capture global structure of character strokes. Specifically, we define global structure 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. For example, one of the most important characteristics of a handwritten numeral "0" is that its starting and ending points are close to each other. This closeness is represented directly by the global feature between those points, an ability not offered by local features.

As briefly reviewed in Section II, most character recognition methods, especially online character recognition methods, use only local features, such as the xy-coordinate feature and the local direction feature, to represent character strokes. This might be a consequence of 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 and both of them require that problem have the Markovian property. In fact, use of global features clearly violates this Markovian constraint. Consequently, to the best of the authors' knowledge, global features haves not fully been investigated and thus the superiority of global features over local features has not been confirmed.

The main contributions of this paper are twofold.

- First, through an experiment on automatic feature selection within the AdaBoost-based machine learning framework, we prove that global features are often selected as being more important than local features with regard to improving classification performance, i.e., better character representation. An online numeral dataset is utilized for training the classifier. We also observe the selected global features and show that, for example, the global feature linking the starting and ending points is surely important for the character "0".
- Second, through a recognition experiment on a test dataset, we prove that use of global features yields better recognition accuracy.

We examine the performance of global features for the

IEEE
computer
society

online character recognition task throughout this paper. An important note is that if global features are useful in this task, they will also be useful in the offline classification task. In fact, nowadays, online recognition methods employ the idea of offline recognition. Conversely, offline recognition methods can employ the idea of online recognition. For example, in [3], 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 II reviews related works in feature extraction. Section III explains the global feature proposal. The experiments and results are reported and discussed in Section IV. Section V derives conclusions and future works.

II. RELATED WORK

As noted before, most online character recognition methods use xy-coordinate features or local direction features (e.g., [1], [2]). As the exceptions, several online character recognition methods exploit some kinds of global features; One example is the relative stroke position feature for representing inter-stroke relationship. That, unfortunately, is used as just a supplement to a local feature [4], [5]. Another example is the star feature [6] 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 [7] for offline recognition. The trial by Izadi and Suen [8] 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, is 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 IV, 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 [8] at all.

Other kinds of global features have been utilized in offline handwritten character recognition (i.e., OCR). A greater variety in features have been considered than in the online case mentioned above because they are not restricted by the Markovian constraint that prevents online character recognition from using global features. The Sonde method [7] and Glucksman's characteristic loci [9] are classical methods in terms of extracting some global features of strokes in several quantized directions. The method proposed in [10] accumulates features by projecting local direction and projects peripheral information of strokes. Features that extract relative angle and relative position from adjacent strokes have also been proposed [11]. Gabor feature [12]
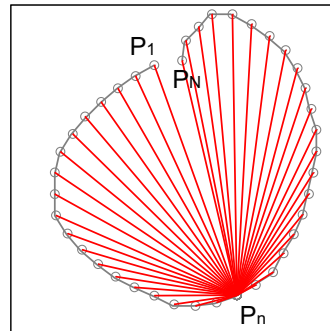

Figure 1. Global feature extraction.

extracts local directions of strokes around each point of interest by scanning in several directions. None of those methods directly addresses our problem, that is, the importance of the global features of character strokes. For example, they fail to indicate how the relationship between the starting and ending points should be treated.

III. GLOBAL FEATURES

We define global structure as a simple relative vector between an arbitrary point pair on the stroke. Details of its generation are as follows: First a character pattern is scaled linearly and re-sampled holding N trajectory points. N is should be large enough to represent the character shape[1]. In this paper, N was fixed at 40, since this is enough to represent numeral patterns smoothly. Next, for each of $_NC_2$ point pairs, relative vector (dx, dy) is calculated as the global feature. Note that the set of the $_NC_2$ global features includes the conventional local direction features as a subset. Figure 1 depicts an example of our global feature extraction. Global features are defined and extracted between an interested point p_n ($n = 1, \ldots, N$) and other points for the interested point p_n.

Our global feature is simple but has high potential to describe the unique characteristics of each class. Figure 2 illustrates this potential. Figure 2 (a) and (b) show that the local features around point p_n have no power to discriminate between classes "2" and "3". However, the global features between the point p_n and other points such as p_N differ for "2" and "3". This indicates that we can introduce new discriminative features for "2" and "3" in addition to the local features around their ending parts. Figure 2 (c) illustrates that global features can regulate the relationship between two points widely separated on the character stroke. Especially for the online recognition task, it is possible to say that the global feature can represent the non-Markovian characteristics of handwriting. Although the non-Markovian

[1]The overall performance is not sensitive to parameter N because we will select M important features from $_NC_2$ global features. In other words, the performance will not be affected by N if it is large enough to provide M important features.
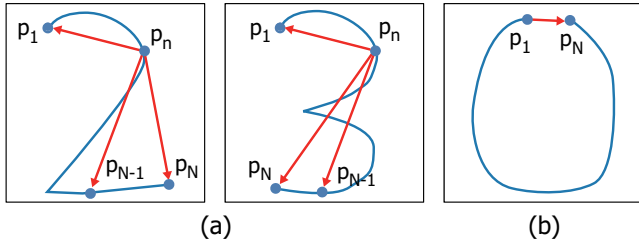
Figure 2. Potential of global features. (a): The local features around p_n do not have any discriminative power but the global features related to p_n do. (b): The global features can regulate the relationship between two widely separated points such as the starting and ending points.

Figure 5. Distribution of the global feature vector selected at the first iteration for "0"-vs.-others classifier.

characteristics are neglected in DTW and HMM, our actual handwriting process clearly exhibits non-Markovian characteristics. We usually watch the stroke already created in order to control the entire character shape when writing a character. In contrast, we must observe only the previous point under the Markovian assumption. Thus, there is a significant risk that "0" will become "6". Our global feature allows the recognition process to reflect our writing process.

IV. EXPERIMENTAL EVALUATION OF THE IMPORTANCE OF GLOBAL FEATURES

A. Purpose of two experiments

In this section, the importance of global features is examined through two experiments. The first experiment applies the AdaBoost [13] framework for observing which features (local vs. global) are important. The details of this experiment are as follows.

- AdaBoost is an iterative learning scheme and can select features one-by-one while setting a classifier called the weak learner [14], [15]. Thus, after M iterations, we have M weak learners. The m-th weak classifier uses only the feature selected in the m-th iteration to provide a classification result. The final result is provided by weighting and summing the outputs of the M weak classifiers.
- A feature 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, if more global features are selected than local features in the first few iterations, we can say the global features were 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.
- Since AdaBoost is a two-class classifier, we train P "one-vs.-others" classifiers, where P denotes the number of classes; we set $P = 10$ in this paper.
- As the weak learner, we used a linear classifier based on the nearest-neighbor principle to one prototype for
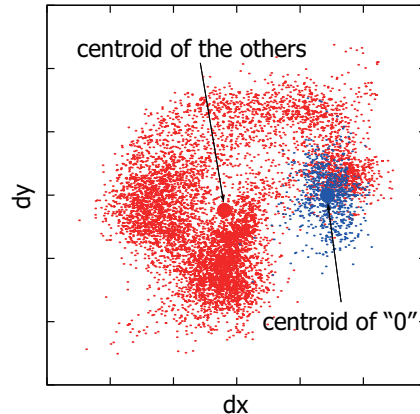
each of two classes (i.e., the target class and the "others" class). The prototype is a weighted average sample whose weight is determined by the AdaBoost framework. Several random perturbations were applied to the prototypes to enhance weak learner performance.

The second experiment is a recognition experiment. The P one-vs.-others classifiers were trained by AdaBoost using only global features, only local features, or both features together. Next, each test sample was subjected to those P classifiers and was classified into the class with highest score. If the classifiers trained using the global features outperform those using the local features, more experimental proof of the importance of global features can be posited.

B. Dataset

Online numeral samples from the UNIPEN database [16] 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, xy-coordinates, and $_NC_2$ global features were extracted from each sample.

C. Result of feature selection experiment

Figure 3 shows the classification accuracies of each one-vs.-others classifier with different features. The global features outperformed the local features, and their combination provided better accuracy. Note that the training iteration stopped when AdaBoost could not find any other weak learner that could improve overall accuracy.

Figure 4 shows the selected features within the first 30 iterations of the "0"-vs.-others classifiers. In Figure 4 (b), the global feature representing the relationship between the starting and ending points was selected as the most important
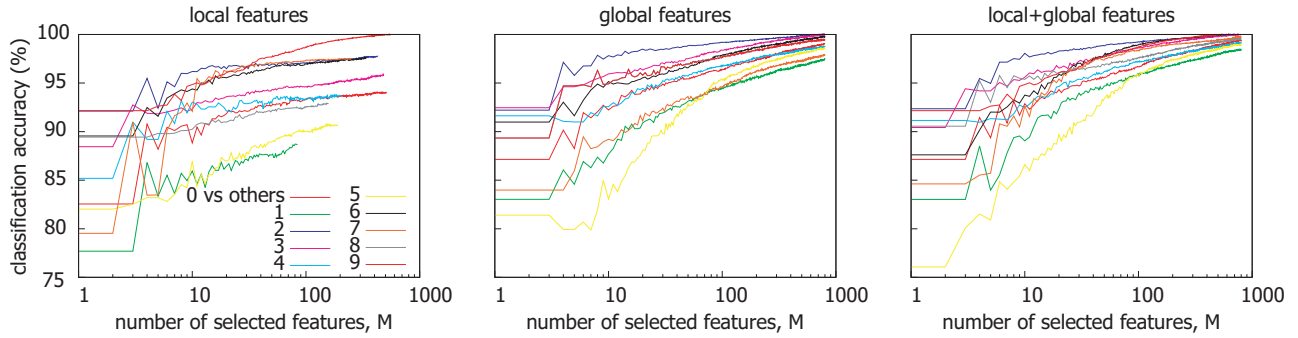
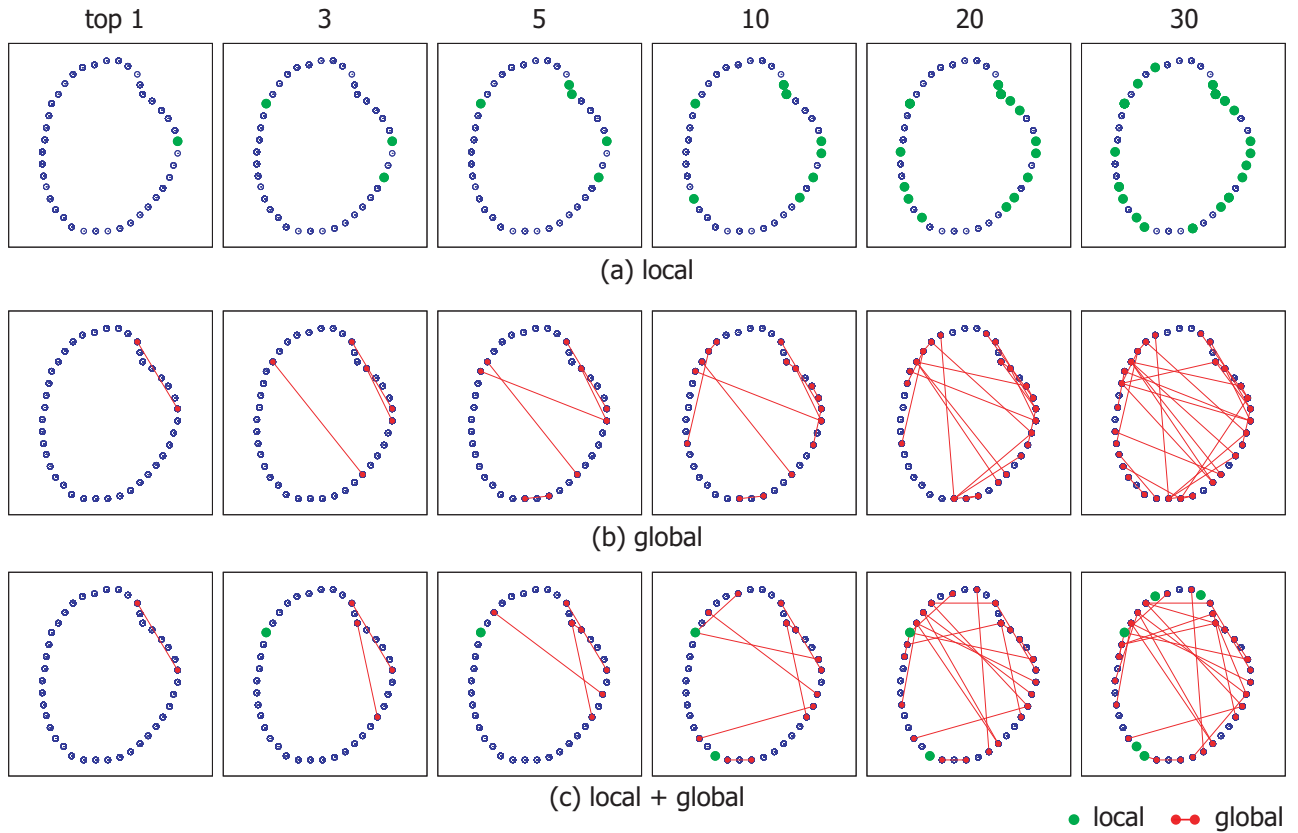Figure 3. Classification accuracy for each one-vs.-others classifier.

Figure 4. Selected features for "0".

feature. Among the 10 numerals, only "0" has those points close to each other. This selection was not changed if global and local features were available for selection as shown in Figure 4 (c). Figure 5 visualizes the distribution of this global feature. It seems that the two classes are reasonably separable by a simple linear classifier and thus this feature selection by AdaBoost is validated. Figure 4 (c) also shows that global features were more frequently selected for "0" within the 30 iterations. One interesting point is that most selected global features represent the relationship between pairs of distant points; that is, they represent that "0" is a circular pattern and has an big empty area inside it.

Figure 6 shows the features selected in the first 10 iterations. Global features were selected not only for "0" but also the other classes. This fact is shown more clearly in Figure 7. In most classes, global features are selected more frequently in the first 20 iterations. These results prove the stability and the dissemination power of global features.
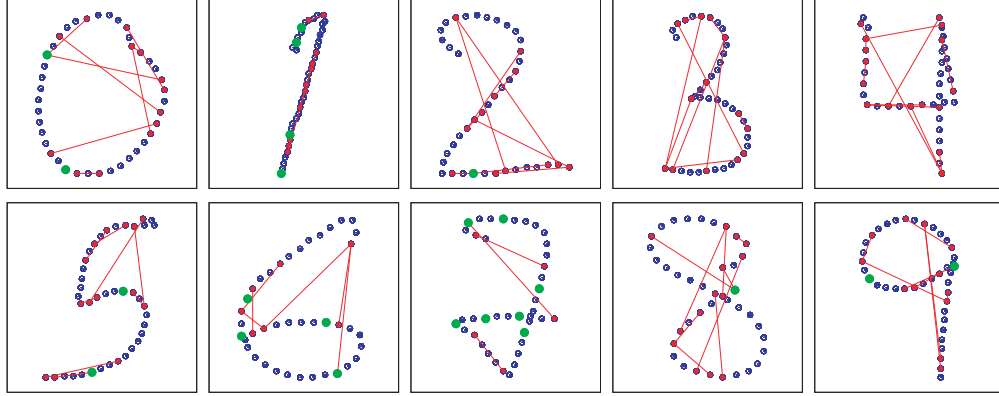
Figure 6. The first 10 features selected from global and local features for each class.
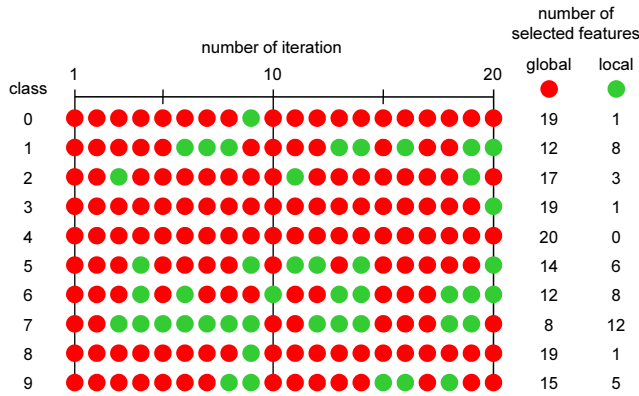

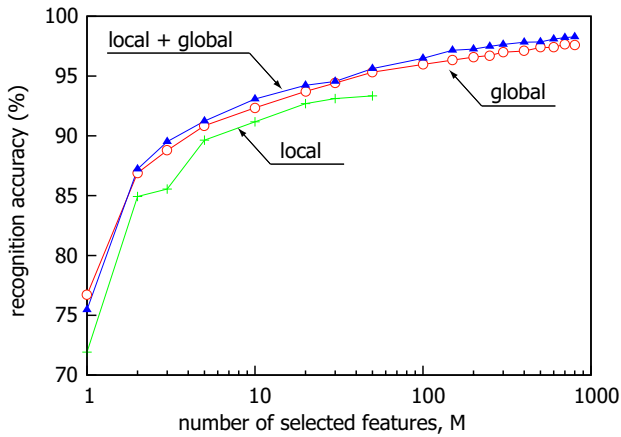Figure 7. The features selected in the first 20 iterations.

Figure 8. Recognition accuracy for test data.

D. Recognition selection experiment

Figure 8 shows the results of our second experiment, the recognition accuracy of 10-class numeral classification of the test samples. As expected from Figure 3, the combination of local and global features has achieved the best accuracy with the use of just global features a close second. One important point is that our current experimental setup made it difficult to manage stroke order because only a single AdaBoost classifier was prepared for each class. Some misrecognition occurred due to variation in stroke order, which drastically perturbed not only local features but also global features from their usual distribution. In other words, global features are effective because, even under such a difficult setup, our result (98.3%) showed a penalty of less than 0.6% from past trials with multiple prototypes or multiple classifiers (98.9%) [17].

V. CONCLUSION

We have described the importance of the global features of character strokes and confirmed it in two experiments. Our global feature is defined as a relative vector between arbitrary 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, the global feature can represent that the separation between the starting point and the ending point of "0" is small. 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" feature. Our experimental validations used a online character recognition task using numeral samples from the UNIPEN dataset. The feature selection experiment with the AdaBoost-based machine learning framework revealed that global features were more frequently selected as important features than local features. This means that 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 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 be 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, the preparation of several classifiers for each class in order to deal with stroke order is straightforward. To observe the important global features of different character sets such as Latin alphabets and Chinese characters is also interesting.

REFERENCES

[1] C. C. Tappert, C. Y. Suen, and T. Wakahara, "The state of the art in on-line handwriting recognition," *IEEE Trans. PAMI*, vol. 12, no. 8, pp. 787–808, 1990.

[2] C. Bahlmann, "Directional feature in online handwriting recognition," *Pattern Recognition*, vol. 39, no. 1, pp. 115–125, 2006.

[3] S. Mori, C. Y. Suen, and K. Yamamoto, "Historical review of ocr research and development," *Proc. IEEE*, vol. 80, no. 7, pp. 1029–1058, 1992.

[4] J.-P. Shin, M. M. Ali, Y. Katayama, and H. Sakoe, "Stroke order free on-line character recognition algorithm using inter-stroke information," *IEICE Trans. Inf. & Syst.*, vol.J82-D-II, no. 3, pp. 382–389, 1999.

[5] I. Ota, R. Yamamoto, S. Sako, and S. Sagayama, "Online handwritten kanji recognition based on inter-stroke grammar," in *ICDAR'07*, vol. 2, 2007, pp. 1188–1192.

[6] D. Mandalapu and S. M. Krishna, "A feature based on encoding the relative position of a point in the character for online handwritten character recognition," in *ICDAR'07*, vol. 2, 2007, pp. 1014–1017.

[7] R. B. Johnson, "Indicia controlled record perforating machine," U.S. Patent 2741312, 1956.

[8] S. Izadi and C. Y. Suen, "Integration of contextual information in online handwriting representation," in *ICIAP'09*, 2009, pp. 132–142.

[9] B. A. Glucksman, "Classification of mixed-font alphabetics by characteristic loci," in *Dig. IEEE Comput. Conf.*, 1967, pp. 138–141.

[10] T. Akiyama and N. Hagita, "Automated entry system for printed documents," *Pattern Recognition*, vol. 23, no. 11, pp. 1141–1154, 1990.

[11] M. Mori, T. Wakahara, and K. Ogura, "Measures for structural and global shape description in handwritten kanji character recognition," in *Document Recognition V*, vol. 3305, 1998, pp. 81–89.

[12] C.-L. Liu, M. Koga, and H. Fujisawa, "Gabor feature extraction for character recognition: Comparison with gradient feature," in *ICDAR'05*, vol. 1, 2005, pp. 121–125.

[13] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, 1997.

[14] P. Viola and M. J. Jones, "Robust real-time face detection," *IJCV*, vol. 57, no. 2, pp. 137–154, 2004.

[15] M. Tsuchiya and H. Fujiyoshi, "Evaluating feature importance for object classification in visual surveillance," in *ICPR'06*, 2006, pp. 978–981.

[16] I. Guyon, L. Schomaker, R. Plamondon, M. Liberman, and S. Janet, "Unipen project of on-line data exchange and recognizer benchmarks," in *ICPR'94*, 1994, pp. 29–33.

[17] E. H. Ratzlaff, "Methods, report and survey for the comparison of diverse isolated character recognition results on the unipen database," in *ICDAR'03*, vol. 1, 2003, pp. 623–628.