JinFeng GAO, Bilan Zhu, Masaki Nakagawa:Proc. 2nd China-Korea-Japan Joint Workshop on Pattern Recognition (CKJPR2010), Fukuoka, Japan, pp. 20-23 Nov.2010.

Effects of Feature Extraction and Dimensionality Reduction for Off-line Handwritten Japanese Character Recognition

JinFeng GAO†, Bilan ZHU††, and Masaki NAKAGAWA††

†, ††Tokyo University Agriculture and Technology Naka-cho 2-24-16,Koganei-si, Tokyo, Japan

E-mail : †gaojinfeng19790213@yahoo.co.jp, ††{ zhubilan, nakagawa}@cc.tuat.ac.jp

Abstract In this paper, we investigate the effects of feature extraction and dimensionality reduction for off-line handwritten Japanese character recognition. Two multi-dimensional feature vectors, corresponding to different sampling intervals of low-pass Gaussian filters, are extracted and concatenated for each character pattern, and the dimensionality is further reduced to 100 by Fisher linear discriminant analysis (FDA) or modified PCA (MPCA). We compare the performance of the reduced features with a baseline system using 100-dimensional original features without reduction. The experimental results demonstrate the effectiveness of our proposed method and the character recognition rate is improved by 2.1 point with FDA compared to the baseline system.

Key words: FDA, modified PCA, feature reduction, off-line character recognition

1. Introduction

With the development of handwriting recognition, although great achievements have been obtained since many online and offline character recognition systems have been put into applications, the basic research in handwriting recognition still calls much attention.

For on-line handwritten Japanese character recognition, we have reported a combined off-line and on-line recognition method which shows that an off-line feature vector with 100 dimensions gives the best performance while maitaining the memory size [1]. Higher feature dimensionality should bring better recognition accuracy, but the higher the feature dimensionality is, the more the parameters of a classifier are, so that more training samples are demanded to train the parameters of the classifier. However, we do not actually have enough training samples with the result that under the condition of a limited amount of training samples extracting more features and compressing to lower dimensionality makes better recognition accuracy than originally extracting them in low dimensionality. Many works have investigated the effects of dimensionality reduction [2, 3, 4] in which a higher dimensional feature vector extracted with a constant sampling interval is reduced to a lower dimensionality. In our research, we found that the combination of multi-scale feature extraction and dimensionality reduction could further improve the system performance.

In the paper, two-scale feature vectors with different sampling intervals of low-pass Gaussian filters are extracted and concatenated for each character pattern, whose dimensionality is further reduced to 100 by FDA or MPCA. We compare the performance of the reduced feature vector with the baseline system using 100-dimensional original features without reduction.

The rest of this paper is organized as follows: Section 2 gives an overview of our proposed system. Section 3 is an introduction of the modified PCA and FDA algorithms. Section 4 reports the experimental results and Section 5 provides the concluding remarks.

2. System Overview

Fig. 1 describes the flow chart of the experimental system. Because this paper aims at feature reduction using FDA and PCA, normalization and recognition parts are not mentioned.

As for the feature extraction part, in the paper, we extract the directional feature. Directional feature is accepted in both on-line and off-line handwriting recognition because of the stable recognition effects. Extracting the directional feature from character pattern image include two steps: all local edges of the image are assigned to quantized direction planes (in this paper, use four directions); then each plane is uniformly sampled with a sampling interval by a low-pass Gaussian filter, then these sampled values from the all planes are used to form

JinFeng GAO, Bilan Zhu, Masaki Nakagawa:Proc. 2nd China-Korea-Japan Joint Workshop on Pattern Recognition (CKJPR2010), Fukuoka, Japan, pp. 20-23 Nov.2010.

a feature vector of the character pattern image.

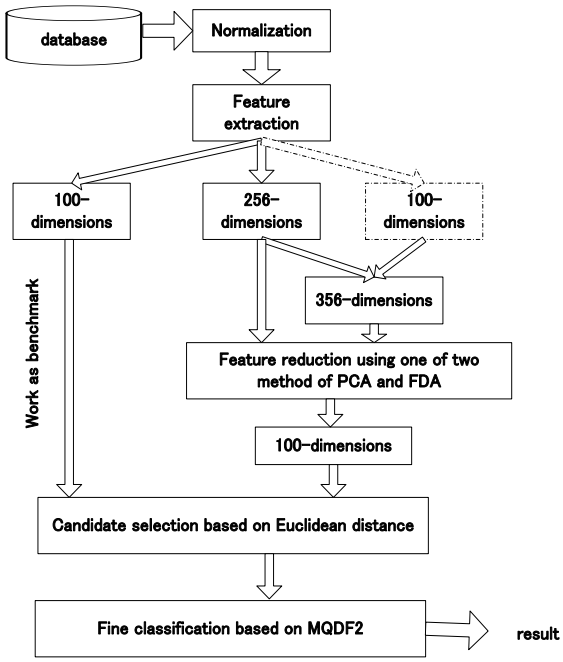

Fig. 1. Flow chart of the character recognition system

The feature vectors used in the paper include two types: original directional feature vectors and combined directional feature vectors. Original directional features are the directional features extracted by using the above approach. A combined feature vector is composed of original features extracted with different sampling intervals. Formally given n-dimensional original feature$\mathcal{F}_1(\mho_{11}, \mho_{12}, \dots \mho_{1n})$, $\mho_{1i}(i \le n)$ is the i-th component of the feature vector $\mathcal{F}_1$ and the m-dimensional original feature vector $\mathcal{F}_2(\mho_{21}, \mho_{22}, \dots \mho_{2m})$, then we can get the $(m+n)$-dimensional combined feature vector shown as $\mathcal{F}(\mho_{11}, \mho_{12}, \dots \mho_{1n}, \mho_{21}, \mho_{22}, \dots \mho_{2m})$, $\mho_j(j \le m+n)$ is the j-th component of the combined feature vector. Note that a combined feature vector used in this paper is composed of original features extracted from different scales but using the same method.

For testing the effect of dimension reduction using modified PCA and FDA, as shown in feature extraction stage in Fig. 1, we partition a character pattern into 5×5 zones with 4 directional planes to get 100-dimensional features and partition it into 8×8 zones with 4 directional planes to get 256-dimensional features. Then we reduce the 256-dimensional vector and 356-dimensional vector, which is combined from 100-dimensional features and 256-dimensional features, to 100-dimensions. We employ the Euclidean distance for candidate selection from all character categories and MQDF2 [5] for fine selection from the candidates.

3. Dimensionality Reduction Algorithms

In this section, we describe two methods by which we realize the feature space reduction. For better classification, we modify the criterion of the traditional PCA. In the realization of FDA literature, different definitions of between-class and within-class scatter covariance exist; for comparison in the paper we compare these two types.

3.1 Modified PCA

PCA technique also known as Karhunen-Loeve method, produces the optimal transformation matrix by maximizing the total scatter covariance of all samples for dimensionality reduction. It belongs to unspecified class linear projection [6].

Formally, given that we have a set of N feature vectors of handwritten pattern samples, noted as $\mathcal{X}(\mathcal{X}_1, \mathcal{X}_2, \dots \mathcal{X}_\mathrm{N})$, $\mathcal{X}_i(i \le \mathrm{N})$ is n-dimensions column vector. The total scatter covariance defined as the following Eq (1).

$$\mathcal{S}_\mathcal{T} = \sum_{i=1}^{\mathcal{N}} (\mathcal{X}_i - \mathcal{U})(\mathcal{X}_i - \mathcal{U})^\mathcal{T} \tag{1}$$

where $\mathcal{U}$ is the mean vector of all samples and $\mathcal{X}_i$ the i-th feature vector. We can obtain the optimal transformation matrix $\mathcal{W}_\mathrm{opt}$, according to the following formula (2).

$$\mathcal{W}_\mathrm{opt} = \arg\max_\mathcal{W} |\mathcal{W}^\mathcal{T} \mathcal{S}_\mathcal{T} \mathcal{W}| = [\mathcal{W}_1\, \mathcal{W}_2 \dots \mathcal{W}_\mathrm{m}] \tag{2}$$

where $\{\mathcal{W}_i | i = 1,2,..,m\}$ are m n-dimensional eigenvectors of $\mathcal{S}_\mathcal{T}$ corresponding to the m largest eigenvalues. Then we can use the transformation matrix $\mathcal{W}_\mathrm{opt}$ to realize the linear transformation from n-dimensions to m-dimensions space according to the follow Eq (3):

$$\mathcal{Y}_i = \mathcal{W}^\mathcal{T} \mathcal{X}_i \; (\mathcal{Y}_i \in \mathbb{R}^m) \tag{3}$$

The optimal transformation matrix attained from PCA method maximizes not only the between-class scatter which is useful for classification but also within-class scatter covariance. But for classification, we expect the within-class scatter covariance be as smaller as possible. In our experiments, using the total scatter covariance $\mathcal{S}_\mathcal{T}$ brought lower recognition performance for Japanese characters recognition. So, in the paper, we take the between-class scatter covariance $\mathcal{S}_\mathcal{B}$ defined as following Eq (4) instead of the total scatter covariance $\mathcal{S}_\mathcal{T}$, and note the modified PCA as MPCA.

$$\mathcal{S}_\mathcal{B} = \sum_{i=1}^{\mathcal{C}} \mathcal{N}_i (\mathcal{U}_i - \mathcal{U})(\mathcal{U}_i - \mathcal{U})^\mathcal{T} \tag{4}$$

where $\mathcal{U}_i$ is a mean vector of the ith class; $\mathcal{U}$ is the mean vector of all training sample; $\mathcal{N}_i$ is the number of samples in ith class; $\mathcal{C}$ is the number of classes. In the same way we can

obtain the optimal transformation matrix $\mathcal{W}_{\text{opt}}$.

3.2 Fisher linear discriminant analysis

In the process of FDA, we need between-class scatter covariance $\mathcal{S}_{\mathcal{B}}$ and within-class scatter covariance $\mathcal{S}_{\mathcal{W}}$ of trianing samples. In the FDA realization literature, there are generally two types definition of $\mathcal{S}_{\mathcal{B}}$ and $\mathcal{S}_{\mathcal{W}}$. Previous works [6, 7] take the first type of definition; another previous work [8] employs the second type of definition. In our experiments, we test these two types of definitions and present the results in the following section.

The first type of definition of $\mathcal{S}_{\mathcal{B}}$ and $\mathcal{S}_{\mathcal{W}}$ is given in Eq (4) and (5).

$$\mathcal{S}_{\mathcal{W}} = \sum_{i=1}^{\mathcal{C}} \sum_{x_{\mathcal{K}} \in \mathcal{X}_i} (x_{\mathcal{K}} - \mathcal{U}_i)\,(x_{\mathcal{K}} - \mathcal{U}_i)^{\mathcal{T}} \qquad (5)$$

The second type of definition of $\mathcal{S}_{\mathcal{B}}$ and $\mathcal{S}_{\mathcal{W}}$ is given in Eq (6) and (7).

$$\mathcal{S}_{\mathcal{W}} = \sum_{i=1}^{\mathcal{C}} \mathcal{P}_i \, \mathrm{E}\{(\mathcal{X} - \mathcal{U}_i)(\mathcal{X} - \mathcal{U}_i)^{\mathcal{T}} | \omega_i\}$$

$$= \sum_{i=1}^{\mathcal{C}} \mathcal{P}_i \sum_i \qquad (6)$$

$$\mathcal{S}_{\mathcal{B}} = \sum_{i=1}^{\mathcal{C}} \mathcal{P}_i (\mathcal{U}_i - \mathcal{U})\,(\mathcal{U}_i - \mathcal{U})^{\mathcal{T}} \qquad (7)$$

In the above formulas, $\mathcal{N}_i$ and $\mathcal{U}_i$ are the number of samples and the mean vector within the ith class respectively; $x_{\mathcal{K}}$ is a feature vector of $\mathcal{K}$-th sample in sample sets of the ith class; $\mathcal{P}_i$ and Σ_i are probability of occurrence and a covariance matrix of i-th class, respectively.

The process of working out the transformation matrix is finding out the optimal ratio which makes the $\mathcal{S}_{\mathcal{B}}$ as large as possible while making the $\mathcal{S}_{\mathcal{W}}$ as small as possible, which is described in the following Eq (8) [9, 10, 11].

$$\mathcal{W}_{\text{opt}} = \arg\max_{\mathcal{W}} \left| \frac{\mathcal{W}^{\mathcal{T}} \mathcal{S}_{\mathcal{B}} \mathcal{W}}{\mathcal{W}^{\mathcal{T}} \mathcal{S}_{\mathcal{W}} \mathcal{W}} \right| = [\mathcal{W}_1\, \mathcal{W}_2 \ldots \mathcal{W}_{\text{m}}] \qquad (8)$$

where $\{\mathcal{W}_i | i = 1,2,..,m\}$ are m n-dimensional eigenvectors of $\mathcal{S}_{\mathcal{W}}^{-1}\mathcal{S}_{\mathcal{B}}$ corresponding to the m largest eigenvalues. The $\mathcal{W}_{\text{opt}}$ is the $n \times m$ matrix composed of the m n-dimensions eigenvectors.

4. Experimental Results

To evaluate the performance of using FDA and MPCA for feature reduction and combined features, we have experimented on the TUAT HANDS on-line database kuchibue and nakayosi. Nakayosi database contains 1,695,689 handwritten character patterns covering 4,438 classes; kuchibue database contains 1,435,440 handwritten character patterns covering 3,345 classes [12]. We transformed the on-line patterns of kuchibue and nakayosi into off-line patterns by interpolating pen points and thickening it to 3 pixel width for experiments. In order to test on all classes of 4,438, we integrated the kuchibue into nakayosi database, then, we divided the integrated database into two parts: 3/5 of every class in the integrated database works as training samples to train the character recognizer, the other part is used as testing samples to evaluate the performance.

Table 1. Recognition rates of the original and the three feature reduced methods from 256 to 100 dimensions.

4,438 classes	Eucli-dean	Number of eigenvectors for MQDF				
		10	20	30	40	50
D100	73.72	86.65	87.94	88.10	88.04	88.00
MPCA	73.38	85.93	87.42	87.66	87.75	87.93
FDA1	79.43	88.25	89.21	89.46	89.55	89.65
FDA2	81.89	88.66	89.33	89.46	89.51	89.65

Table 2. Recognition rates of the original and the three feature reduced methods from 356 to 100 dimensions

4,438 classes	Euclid-ean	Number of eigenvectors for MQDF				
		10	20	30	40	50
D100	73.72	86.65	87.94	88.10	88.04	88.00
MPCA	75.21	87.20	88.39	88.58	88.55	88.64
FDA1	79.55	88.46	89.44	89.67	89.78	89.85
FDA2	81.98	88.78	89.49	89.62	89.69	89.84

For comparison, we consider the recognition rate of the original 100-dimensional features as benchmark. First, we reduce the original features from 256-dimensions to 100-dimensions using FDA and MPCA. As for FDA method, we note the first type of definition as FDA1, the second type as FDA2, because of having two types of definition. The results for testing samples are shown in Table 1. From Table 1, compared with the benchmark, we can conclude the recognition rate is improved about 2.0 point using FDA, but low as little as 0.7 point using MPCA. Second, we do the similar experiment to the first one using the combined 356-dimensions features which consist of 100-dimensional original features and 256-dimensional features. The result is given in Table 2. From Table 2, we can see both MPCA and FDA worked positive for improving the recognition accuracy compared with the benchmark and the effect of using FDA is better than MPCA.

From Table 1 and Table 2, we can see (1) the accuracy of candidate selection based on the Euclidean distance is improved almost 7 point at the best condition, (2) the effect of the feature reduction using FDA is better than using MPCA, (3) the accuracy of combined features after feature reduction using MPCA and FDA is higher than the original feature.

In order to investigate the effects of reduction to different dimensions using MPCA and FDA, we repeat experiments on 256-dimensional and 356-dimensional features reduced to dimensions from 80 to 160 with step 20 employing FDA1, FDA2 and MPCA. The results is given in Fig 2. From the results, we can see 356-dimensional features in the process of dimension reduction produce the optimal value after reduced to about 120-dimensions; 256-dimensions features produce that after reduced to about 100-dimensions. We can also see that extracting more features and compressing to lower dimensionality can produce better recognition accuracy under the condition that a limited amount of training samples are avaiable.

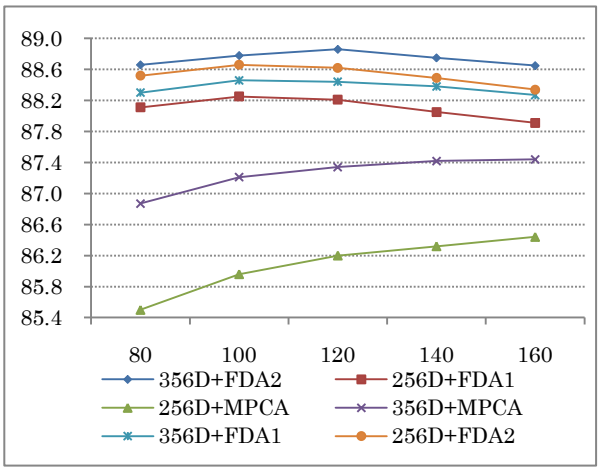

Fig 2 Results of reduction to different dimensions.

5. Conclusion and references

In this paper, we investigated the effects of feature extraction and dimensionality reduction for off-line handwritten Japanese character recognition. We extract two-scale feature vectors from each character pattern with different sampling intervals, and further reduce the dimensionality by MPCA and FDA. We compare the performance of the two-scale features and the dimensionality reduction algorithms (MPCA and FDA). Experimental results demonstrated the superiority of the combined features and feature reduction.

Future work includes investigating multi-scale features and other dimensionality reduction algorithms.

Acknowledgement

This work is being supported by the R&D fund for "development of pen & paper based user interaction" under Japan Science and Technology Agency.

References

[1] H. Oda, B. Zhu, J. Tokuno, M. Onuma, A. Kitadai and M. Nakagawa, A Compact On-line and Off-line Combined Recognizer, Proc. IWFHR-10, La Baule, France, pp.133-138, 2006.

[2] Cheng-Lin Liu and Xiang-Dong Zhou, Online Japanese Character Recognition Using Trajectory-based Normalization and Direction Feature Extraction, Proc.10th International Workshop on Frontiers in Handwriting Recognition, 2006.

[3] Tian-Fu Gao, Cheng-Lin Liu, High Accuracy Handwritten Chinese Character Recognition Using LDA-based Compound Distances, Pattern Recognition, PP3442-3451, 2008.

[4] K. C. Leung, C. H. Leung, Recognition of Handwritten Chinese Characters by Critical Region Analysis, Pattern Recognition, 43, PP949–961, 2010.

[5] Fumitaka Kimura et al., Modified Quadratic Discriminant Functions and the Application to Chinese Character Recognition, IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol.PAMI-9, No.1, 1987.

[6] Kresimir Delac and Mislav Grgic, Independent Comparative Study of PCA, ICA, and LDA On the FERET Data Set, Wiley Periodicals, 2006.

[7] Wankou Yang and Jianguo Wang, Feature Extraction Using Fuzzy Inverse FDA, Neurocomputing, PP3384-3390, 2009.

[8] Keinosuke Fukunaga, Introduction to Statistical Pattern Recognition, 2nd edition, Academic Press,1990.

[9] Peter N. Belhumeur, Eigenfaces vs Fisherfaces Recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence vol.19, No.7, 1997.

[10] R. P. W. Duin R.Haeb-Umbach, Multiclass Linear Dimension Reduction by Weighted Pairwise Fisher Criteria, IEEE Transactions On Pattern Analysis And Machine Intelligence, VOL.23, NO.7, 2001.

[11] Li Wen-shu and Zhou Chang-le, A novel face recognition method with feature combination, Journal of Zhejiang University Science, 2005 6A(5), PP454-459.

[12] Hideto Oda, Akihito kitadai, Bilan Zhu, A Scalable On-line and Off-line Combined Recognition System, IECE Technical Report TL2005-90, 2006.