1998 International Computer Symposium
Workshop on Image Processing and Character Recognition
December 17-19, 1998, N.C.K.U., Tainan, Taiwan, R.O.C.

Feature Reduction for Face Recognition by PDBNN *

L.J. Shen, Y.P. Lee, and H.C. Fu
Department of Computer Engineering,
National Chiao Tung University, Hsin-Chu, Taiwan, R.O.C
ljshen@pington.csie.nctu.edu.tw

ABSTRACT

Selecting proper features for efficient face recognition is an essential task in neural network design. Most of recognition or classification algorithms using features in uniform manner for each class. We believe this constraint could be relaxed to achieve better recognition performance. In this paper, we present a PDBNN based feature reduction algorithm that deletes some feature vectors which contribute the least among the whole feature set. The deletion is performed on each facial basis. By applying the proposed algorithm, we performed some face recognition experiments on an 151 people facial database. The experimental results show the recognition accuracy improved from the original 87.91% (2500 features) to 91.22% by using only 500 features.

1 Introduction

Neural networks are very popular as classifiers due to their fast parameter estimation ability, but a serious disadvantage of neural networks is the curse of dimensionality. One way to tackle this problem is the feature reduction. The problem of feature reduction has long been an active research topic in pattern recognition. Feature reduction techniques generally based on a criterion function and a search strategy. The criterion function has been distinguished into wrappers and filters approaches: Wrapper approaches use the accuracy of a classifier to decide the superiority of a feature subset whereas filter approaches use a criterion function which is independent of classifier accuracy. There exist a lot of search strategies for the determination of an optimal feature subset. The search strategy can be grouped into two categories: exhaustive or heuristic search. Given N features we have $2^N - 1$ possible feature subsets thus an exhaustive search is computationally too expensive if N is large. There are many heuristic search methods out of this exponential computation using genetic algorithm[1], decision tree algorithm[2], random algorithm[3] and neural networks[4] etc.

Face recognition usually consumes considerable computation time because the inherence of high dimensional face images. In this research, we propose a feature reduction algorithm that reduces feature spaces according to each person's own unique facial characteristics. Each facial class is represented by a probabilistic distribution of its specific feature set. By using a PDBNN classifier, the input feature values are computed into posteriori probability that relates the input image to a facial class. With a dedicated feature set for each facial class, the face recognition experiments on our database show that it is possible to achieve better recognition accuracy with much less number and distinct of class features.

The rest of this paper are arranged as follows: In section 2, the probabilistic decision based neural networks are introduced. In section 3, explains the feature reduction concept and algorithm. In section 4, presents the experiment results on face recognition. In last section, concluding remarks are given.

2 Probabilistic DBNN

The Probabilistic DBNN proposed by [6] is a modular and hierarchical architecture neural network. It adopts a competitive credit-assignment scheme to decide which subnets should be trained or used. To training a PDBNN only a subnet is reinforced

*This research was supported by National Science Council under the contract of NSC 87-2213-E009-040.

and usually only few subnets need to be anti-reinforced. Thus, this model has been very effective for many signal/image classification applications.

The discriminant function of PDBNN is in a form of probability density. Figure 1 shows the PDBNN structure. Each subnet of the PDBNN is designed to model a mixture of Gaussian distributions, i.e., the class likelihood function is a linear combination of R cluster likelihood probabilities,

$$p(\mathbf{f} \mid c_i) = \sum_{r=1}^{R} P(\Theta_r \mid c_i) p(\mathbf{f} \mid c_i, \Theta_r) \quad (1)$$

$$p(\mathbf{f} \mid c_i, \Theta_r) = \frac{1}{(2\pi)^{\frac{D}{2}} \prod_{d=1}^{D} \sigma_{rd}} exp(-\frac{1}{2} \sum_{d=1}^{D} \frac{(f_d - \mu_{rd})^2}{\sigma_{rd}^2}). \quad (2)$$

where $P(\Theta_r \mid c_i)$ is cluster prior probability and $p(\mathrm{f} \mid c_i, \Theta_r)$ is a D-dimensional Gaussian distribution. By the Bayesian decision rule, for input feature vector f, the class c_i score is

$$P(c_i \mid \mathbf{f}) = \frac{P(c_i) p(\mathbf{f} \mid c_i)}{\sum_{j=1}^{N} P(c_j) p(\mathbf{f} \mid c_j)}. \quad (3)$$

In Figure 1 the **MAXNET** function selects the class of larger score as winner. When a training image is misclassified, the reinforced or anti-reinforced learning technique [6] is applied to adjust class means and variances:

$$\mu_{rd}^{(t+1)} = \mu_{rd}^{(t)} \pm \eta_\mu h_r^{(t)} \beta_{rd}^{(t)} (f_d^{(t)} - \mu_{rd}^{(t)}) \quad (4)$$

$$\beta_{rd}^{(t+1)} = \beta_{rd}^{(t)} \pm \eta_\sigma h_r^{(t)} \frac{1}{2} [\frac{1}{\beta_{rd}^{(t)}} - (f_d^{(t)} - \mu_{rd}^{(t)})^2] \quad (5)$$

$$\sigma_{rd}^2 = \frac{1}{\beta_{rd}} \quad (6)$$

where

$$h_r^{(t)} = \frac{P(\Theta_r \mid c_i) p(\mathbf{f}^{(t)} \mid c_i, \Theta_r)}{\sum_{k=1}^{R} P(\Theta_k \mid c_i) p(\mathbf{f}^{(t)} \mid c_i, \Theta_k)}. \quad (7)$$

3 Feature Reduction

A good feature selection algorithm will reduce the input space to a lower dimensionality while still maintaining a large portion of the information contained in the original pattern space. An appropriate choice of features can help in eliminating redundant and irrelevant information from the large feature set, thereby reducing some overhead (computing time and space) for the classifier. Almost all of the classifiers usually recognize patterns by using the same feature set, no matter whether the feature is critical or irrelevant to a particular class. We propose a feature reduction algorithm to improve these drawbacks. In our approach, different classes are not necessary to have the uniform feature set. Instead, each class has a dedicated set of feature vector such that it can precisely capture the most distinguishable characteristics for recognition purposes. The feature reduction algorithm is described as follows.

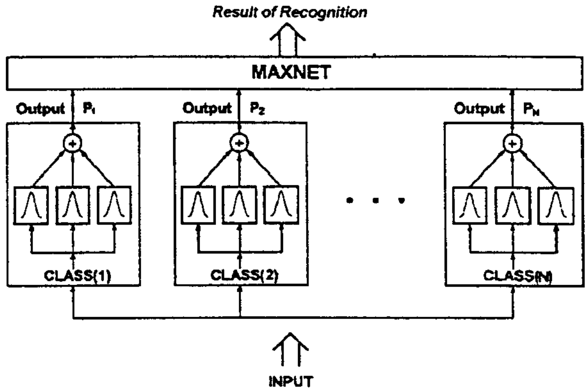

Figure 1: The PDBNN architecture.

Feature Reduction Algorithm

Given a set of D-dimensional training patterns with L classes $\mathbf{X}^+ = \{\mathbf{X_1}, \mathbf{X_2}, ..., \mathbf{X_L}\} \in \mathbf{R}^D$, for each class i, the training set $\mathbf{X_i} = \{X_i(t); t = 1, 2, ..., n_i\} \in \mathbf{R}^D$.

Step 1: Computes the average mean vector **M** by

$$\mathbf{M} = \frac{\sum_{i=1}^{L} \sum_{t=1}^{n_i} X_i(t)}{\sum_{i=1}^{L} n_i} \quad (8)$$

Step 2: For each training class i,

1. uses the training patterns $\mathbf{X_i}$ to compute the class mean vector $\mathbf{M_i} \in R^D$ and variance vector $\mathbf{V_i} \in R^D$ by

$$\mathbf{M_i} = \sum_{t=1}^{n_i} X_i(t)/n_i \quad (9)$$

$$\mathbf{V_i} = \sum_{t=1}^{n_i} (X_i(t) - \mathbf{M_i})^2/n_i \quad (10)$$

2. computes the vector $F_i = (f_{i1}, f_{i2}, ..., f_{iD})$

1998 International Computer Symposium
Workshop on Image Processing and Character Recognition
December 17-19, 1998, N.C.K.U., Tainan, Taiwan, R.O.C.

of feature j,

$$f_{ij} = \frac{(m_{ij} - m_j)^2}{1 + \sigma_{ij}^2} \qquad (11)$$

,where the value $m_{ij}, \overline{m_j}$ denote the jth element in mean vector $\mathbf{M_i}$,$\mathbf{M}$, and the value σ_{ij}^2 denotes the jth element in variance vector $\mathbf{V_i}$.

Step 3: For all class i, perform scree test on f_{ij} for class i to determing the number of significant features k_i.

Step 4: Select features that corresponding to the first k_i large f_{ij} as the final features of class i, $k_i \leq D$;

In above algorithm the constant k_i is the number of features to be used in the class i, which can be determined by scree test. The contribution of feature j to class i for classification and recognition purposes is denoted by f_{ij} in (11). In (11), if the distance $(m_{ij} - m_j)^2$ between mean of class i and mean of all classes in feature j is large, and the variance σ_{ij}^2 of class i is small, then feature j is a significant feature to class i.

For classifying pattern in non-uniform class feature spaces, the Probabilistic DBNN is adopted here as pattern recognition model due to its class probability estimation ability. For a L-class recognition problem, a Probabilistic DBNN recognizer consists of L different subnets. Utilize the feature reduction algorithm for associating feature set on each PDBNN subnet. When a training pattern is misclassified, the reinforced and anti-reinforced learning rule is applied to the feature spaces of winner/correct classes to fine tune the class probabilities. Thus recognizing a pattern in distinct class feature spaces is feasible.

4 Experimental Results

In order to demonstrate the efficiency and accuracy of our feature reduction method, some experiments on face recognition are conducted. The database we used contained 151 people and totally 151 × 8 frontal face images. Experiments were performed with 4 training images and 4 test images per person. There was no overlap between the training and test sets. Some face images in our database are shown in Figure 2.

In our experiments, a *facial region* is defined as a face portion which includes the eyes, nose, and mouth.

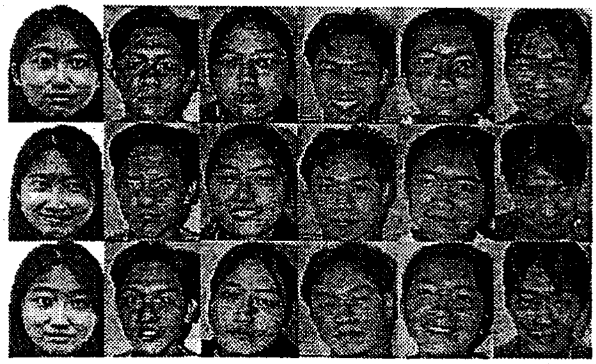

Figure 2: The examples of some images in the facial database

Figure 3 shows the face recognition diagram in our system. In Figure 3, a facial extractor is used to segment a 50×50 pixels of image portion out of each face image. Then, apply the proposed feature reduction algorithm to determine the proper feature sets of all face classes. Finally, A PDBNN is used for recognition by devoting one of its subnets to the representation of a particular person. In our system, each subnet of PDBNN received a dedicated k ($k < 2500$) features. Thus the feature dimensions are greatly reduced. Table 1 shows results of these experiments for the case of varying feature dimension k with above 97% training rate.

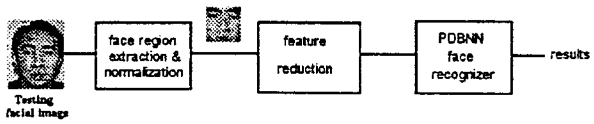

Figure 3: The face recognition system diagram

Table 1: The face recognition rate of 151 people

Feature dimension k	Training rate	Recognition rate
200	97.19%	86.09%
500	97.52%	91.22%
1000	97.35%	91.55%
1500	97.02%	90.19%
2500	97.19%	87.91%

In Table 1, the best recognition case is occurred at the number of features $k = 1000$ rather than the original $d = 2500$ features. By our feature reduction algorithm, some of the irrelevant and large variation features are eliminated. Our experiments show that it is possible to achieve better recognition accuracy with much less number and distinct of class features.

5 Concluding remarks

Feature reduction is a very important aspect of solving the problem of pattern recognition especially for high dimension features. Many of feature reduction algorithms reduce the feature space uniformly for each class. We have proposed a feature reduction algorithm that reduces and selects distinct features for each class respectively. Using PDBNN classifier all of class features are reformed to class probabilities. The position of PDBNN clusters will be updated when misclassification occurs. Thus the class probabilities are adjusted in separate class feature spaces. The face recognition experiments on our database show that it is possible to achieve better recognition accuracy with much less number and distinct of class features.

References

[1] Yang, J., and Honavar V., "Feature subset selection using a genetic algorithm." *IEEE Expert.* In press.

[2] Cherkauer, K.J., and Shavlik, J.W., "Rapidly estimating the quality of input representations for neural networks" *In Proceedings of the IJCAI-95 Workshop on Data Engineering for Inductive Learning*, pp99-108 Montreal, Canada. 1995.

[3] Skalak,D., "Prototype and feature selection by sampling and random mutation hill climbing algorithms." *Proc. Of the Eleventh International Machine Learning Conference*, pp293-301,1994, New Brunswick, NJ: Morgan Kauffmann.

[4] Setiono, R., and Liu, H., "Neural network feature selector", *IEEE Trans. on Neural Networks*, vol 8, May, pp654-662, 1997.

[5] S.Y.Kung and J.S.Taur. "Decision-Based Neural Networks with Signal/Image Classification Application," *IEEE Trans. on Neural Networks*, 6(1):170-181, 1995.

[6] S.H. Lin and S.Y. Kung, "Face recognition/detection by Probabilistic Decision-Based Neural Network," *IEEE Trans. on Neural Networks*, Vol. 8, No.1, January 1997.

[7] S.Y. Kung, M. Fang, S.P. Liou, and J.S. Taur. "Decision-based neural network for face recognition system," *In ICIP'95.*