Face Recognition Using DCT and Hierarchical RBF Model

Yuehui Chen and Yaou Zhao

School of Information Science and Engineering
Jinan University, Jinan 250022, P.R. China
yhchen@ujn.edu.cn

Abstract. This paper proposes a new face recognition approach by using the Discrete Cosine Transform (DCT) and Hierarchical Radial Basis Function Network (HRBF) classification model. The DCT is employed to extract the input features to build a face recognition system, and the HRBF is used to identify the faces. Based on the pre-defined instruction/operator sets, a HRBF model can be created and evolved. This framework allows input features selection. The HRBF structure is developed using Extended Compact Genetic Programming (ECGP) and the parameters are optimized by Differential Evolution (DE). Empirical results indicate that the proposed framework is efficient for face recognition.

1 Introduction

Face recognition has become a very active research area in recent years mainly due to increasing security demands and its potential commercial and law enforcement applications. Face recognition approaches on still images can be broadly grouped into geometric and template matching techniques. In the first case, geometric characteristics of faces to be matched, such as distances between different facial features, are compared. This technique provides limited results although it has been used extensively in the past. In the second case, face images represented as a two dimensional array of pixel intensity values are compared with a single or several templates representing the whole face. More successful template matching approaches use Principal Components Analysis (PCA) or Linear Discriminant Analysis (LDA) to perform dimensionality reduction achieving good performance at a reasonable computational complexity/time. Other template matching methods use neural network classification and deformable templates, such as Elastic Graph Matching (EGM). Recently, a set of approaches that use different techniques to correct perspective distortion are being proposed. These techniques are sometimes referred to as view-tolerant. For a complete review on the topic of face recognition the reader is referred to [1] and [2].

Neural networks have been widely applied in pattern recognition for the reason that neural-networks-based classifiers can incorporate both statistical and structural information and achieve better performance than the simple minimum distance classifiers [2]. Multilayered networks(MLNs), usually employing

E. Corchado et al. (Eds.): IDEAL 2006, LNCS 4224, pp. 355–362, 2006.

the backpropagation (BP) algorithm, are widely used in face recognition [3]. Recently, RBF neural networks have been applied in many engineering and scientific applications including face recognition [7]. HRBF networks consist of multiple RBF networks assembled in different level or cascade architecture in which a problem was divided and solved in more than one step. Mat Isa et al. used Hierarchical Radial Basis Function (HiRBF) to increase RBF performance in diagnosing cervical cancer [4]. Hierarchical RBF network has been proved effective in the reconstruction of smooth surfaces from sparse noisy data points [5]. In order to improve the model generalization performance, a selective combination of multiple neural networks by using Bayesian method was proposed in [6].

In this paper, an automatic method for constructing HRBF networks is proposed. Based on a pre-defined instruction/operator set, the HRBF network can be created and evolved. The HRBF network allows input variables selection. In our previous studies, in order to optimize the Flexible Neural Tree (FNT) and the hierarchical TS fuzzy model (H-TS-FS), the hierarchical structure of FNT and H-TS-FS was evolved using Probabilistic Incremental Program Evolution algorithm (PIPE) [11][12] and Ant Programming with specific instructions. In this research, the hierarchical structure is evolved using the Extended Compact Genetic Programming (ECGP). The fine tuning of the parameters encoded in the structure is accomplished using the DE algorithm. The novelty of this paper is in the usage of HRBF model for selecting the important features and for face recognition.

2 Discrete Cosine Transform

Like other transforms, the Discrete Cosine Transform (DCT) attempts to decorrelate the image data [8]. After decorrelation each transform coefficient can be encoded independently without losing compression efficiency. This section describes the DCT and some of its important properties.

The 2-D DCT is a direct extension of the 1-D case and is given by

$$C(u,v) = \alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y) cos\frac{\pi(2x+1)u}{2N} cos\frac{\pi(2y+1)}{2N} \tag{1}$$

for $u, v = 0, 1, 2, \ldots, N-1$ and $\alpha(u)$ and $\alpha(v)$ are defined as follows, $\alpha(u) = \sqrt{1/N}$ for $u = 0$, and $\alpha(u) = \sqrt{2/N}$ for $u \neq 0$. The inverse transform is defined as

$$f(x,y) = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} \alpha(u)\alpha(v)C(u,v) cos\frac{\pi(2x+1)u}{2N} cos\frac{\pi(2y+1)}{2N} \tag{2}$$

for $x, y = 0, 1, 2, \ldots, N-1$.

The DCT possess some fine properties, i.e., de-correlation, energy compaction, separability, symmetry and orthogonality. These attributes of the DCT have led to its widespread deployment in virtually every image/video processing standard of the last decade [8].

For an $N \times N$ image, an DCT coefficient matrix covering all the spatial frequency components of the image. The DCT coefficients with large magnitude are mainly located in the upper-left corner of the DCT matrix. Accordingly, we scan the DCT coefficient matrix in a zig-zag manner starting from the upper-left corner and subsequently convert it to a one-dimensional (1-D) vector. As a holistic feature extraction method, the DCT converts high-dimensional face images into low-dimensional spaces in which more significant facial features such as outline of hair and face, position of eyes, nose and mouth are maintained. These facial features are more stable than the variable high-frequency facial features. As a matter of fact, the human visual system is more sensitive to variations in the low-frequency band.

In this paper, we investigate the illumination invariant property of the DCT by discarding its several low-frequency coefficients. It is well-known that the first DCT coefficient represents the dc component of an image which is solely related to the brightness of the image. Therefore, it becomes DC free (i.e., zero mean) and invariant against uniform brightness change by simply removing the first DCT coefficient.

3 The RBF Network

An RBF network is a feed-forward neural network with one hidden layer of RBF units and a linear output layer. By an RBF unit we mean a neuron with multiple real inputs $\boldsymbol{x} = (x_1, \ldots, x_n)$ and one output y computed as:

$$y = \varphi(\xi); \quad \xi = \frac{\|\boldsymbol{x} - \boldsymbol{c}\|_C}{b} \tag{3}$$

where $\varphi : R \to R$ is a suitable activation function, let us consider Gaussian radial basis function $\varphi(z) = e^{-z^2}$. The center $\boldsymbol{c} \in R^n$, the width $b \in R$ and an $n \times n$ real matrix $\boldsymbol{C}$ are a unit's parameters, $|| \cdot ||_C$ denotes a weighted norm defined as $\|\boldsymbol{x}\|_C^2 = (\boldsymbol{Cx})^T(\boldsymbol{Cx}) = \boldsymbol{x}^T\boldsymbol{C}^T\boldsymbol{Cx}$.

Thus, the network represents the following real function $\boldsymbol{f} : R^n \to R^m$:

$$f_s(\boldsymbol{x}) = \sum_{j=1}^{h} w_{js} e^{-(\frac{\|\boldsymbol{x}-\boldsymbol{c}\|_C}{b})^2}, \qquad s = 1, \ldots, m, \tag{4}$$

where $w_{js} \in R$ are weights of s-th output unit and f_s is the s-th network output.

The goal of an RBF network learning is to find suitable values of RBF units' parameters and the output layer's weights, so that the RBF network function approximates a function given by a set of examples of inputs and desired outputs $T = \{\boldsymbol{x}(t), \boldsymbol{d}(t); t = 1, \ldots, k\}$, called a *training set*. The quality of the learned RBF network is measured by the *error function*:

$$E = \frac{1}{2} \sum_{t=1}^{k} \sum_{j=1}^{m} e_j^2(t), \qquad e_j(t) = d_j(t) - f_j(t). \tag{5}$$

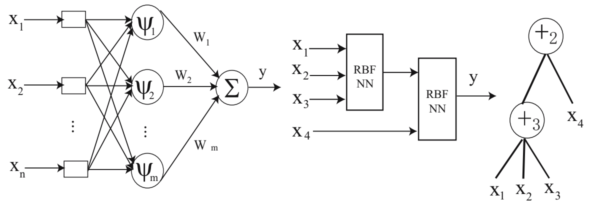

Fig. 1. A RBF neural network (left), an example of hierarchical RBF network (middle), and a tree-structural representation of the HRBF network (right)

4 The Hierarchical RBF Network

4.1 Encode and Calculation

A function set F and terminal instruction set T used for generating a HRBF network model are described as $S = F \bigcup T = \{+_2, +_3, \ldots, +_N\} \bigcup \{x_1, \ldots, x_n\}$, where $+_i (i = 2, 3, \ldots, N)$ denote non-leaf nodes' instructions and taking i arguments. x_1,x_2,...,x_n are leaf nodes' instructions and taking no arguments. The output of a non-leaf node is calculated as a HRBF network model (see Fig.1). In this research, Gaussian radial basis function is used and the number of radial basis functions used in hidden layer of the network is same with the number of inputs, that is, $m = n$.

In the creation process of HRBF network tree, if a nonterminal instruction, i.e., $+_i (i = 2, 3, 4, \ldots, N)$ is selected, i real values are randomly generated and used for representing the connection strength between the node $+_i$ and its children. In addition, $2 \times n^2$ adjustable parameters a_i and b_i are randomly created as radial basis function parameters. The output of the node $+_i$ can be calculated by using Eqn.(1) and Eqn.(2). The overall output of HRBF network tree can be computed from left to right by depth-first method, recursively.

4.2 Tree Structure Optimization by ECGP

Finding an optimal or near-optimal HRBF is formulated as a product of evolution. In this paper, the ECGP [13] is employed to find an optimal or near-optimal HRBF structure. ECGP is a direct extension of ECGA to the tree representation which is based on the PIPE prototype tree. In ECGA, Marginal Product Models (MPMs) are used to model the interaction among genes, represented as random variables, given a population of Genetic Algorithm individuals. MPMs are represented as measures of marginal distributions on partitions of random variables. ECGP is based on the PIPE prototype tree, and thus each node in the prototype tree is a random variable. ECGP decomposes or partitions the prototype tree into sub-trees, and the MPM factorises the joint probability of all nodes of the prototype tree, to a product of marginal distributions on a partition of its sub-trees. A greedy search heuristic is used to find an optimal MPM mode under the framework of minimum encoding inference. ECGP can represent the probability distribution for more than one node at a time. Thus, it extends PIPE in that the interactions among multiple nodes are considered.

4.3 Parameter Optimization with DE Algorithm

The DE algorithm was first introduced by Storn and Price in 1995 [9]. It resembles the structure of an evolutionary algorithm (EA), but differs from traditional EAs in its generation of new candidate solutions and by its use of a 'greedy' selection scheme. DE works as follows: First, all individuals are randomly initialized and evaluated using the fitness function provided. Afterwards, the following process will be executed as long as the termination condition is not fulfilled: For each individual in the population, an offspring is created using the weighted difference of parent solutions. The offspring replaces the parent if it is fitter. Otherwise, the parent survives and is passed on to the next iteration of the algorithm. In generation k, we denote the population members by x_1^k, x_2^k, …, x_N^k. The DE algorithm is given as follows [10]:

S1 Set $k = 0$, and randomly generate N points x_1^0, x_2^0, …, x_N^0 from search space to form an initial population;
S2 For each point $x_i^k (1 \leq i \leq N)$, execute the DE offspring generation scheme to generate an offspring $x_i^{(}k+1)$;
S3 If the given stop criteria is not met, set $k = k + 1$, goto step S2.

The DE Offspring Generation approach used is given as follows,

S1 Choose one point x_d randomly such that $f(x_d) \quad f(x_i^k)$, another two points x_b, x_c randomly from the current population and a subset $S = \{j_1, \ldots, j_m\}$ of the index set $\{1, \ldots, n\}$, while $m < n$ and all j_i mutually different;
S2 Generate a trial point $u = (u_1, u_2, \ldots, u_n)$ as follows:
DE Mutation. Generate a temporary point z as follows,

$$z = (F + 0.5)x_d + (F - 0.5)x_i + F(x_b - x_c); \tag{6}$$

Where F is a give control parameter;
DE Crossover. for $j \in S$, u_j is chosen to be z_j; otherwise u_j is chosen a to be $(x_i^k)_j$;
S3 If $f(u) \leq f(x_i^k)$, set $x_i^{k+1} = u$; otherwise, set $x_i^{k+1} = x_i^k$.

4.4 Procedure of the General Learning Algorithm

The general learning procedure for constructing the HRBF network can be described as follows.

S1 Create an initial population randomly (HRBF network trees and its corresponding parameters);
S2 Structure optimization is achieved by using ECGP algorithm;
S3 If a better structure is found, then go to step S4, otherwise go to step S2;
S4 Parameter optimization is achieved by DE algorithm. In this stage, the architecture of HRBF network model is fixed, and it is the best tree developed during the end of run of the structure search;

S5 If the maximum number of local search is reached, or no better parameter vector is found for a significantly long time then go to step S6; otherwise go to step S4;

S6 If satisfactory solution is found, then the algorithm is stopped; otherwise go to step S2.

5 Face Recognition Using HRBF Paradigm

We performed extensive experiments on two benchmark face datasets, namely the ORL and the Yale face database. In all the experiments, the background is cut out, and the images are resized to 92×112. No other preprocessing is done. Besides our method, the PCA based method, LDA-based method, neural networks etc. were also tested for comparisons.

5.1 The Face Database

For ORL face dateset, 40 persons with variations in facial expression and . All images were taken under a dark background, and the subjects were in an upright frontal position, with tilting and rotation tolerance up to 20 degree, and tolerance of up to about 10%. Fig. 2(left) shows 12 images of one subject from the selected dataset. The Yale face database contains 165 images of 15 subjects. There are 11 images per subject with different facial expressions or lightings. Fig. 2(right) shows the 11 images of one subject. For each experiment, 5 images are generated randomly to form the training data set and the remaining were chosen as test data set. This process was repeated to 20 times for each experiment.

5.2 Experiments on ORL and Yale Face Database

For this simulation, the DCT is employed to training and testing data sets, respectively. The extracted 60 input features are used for constructing a HRBF model. A HRBF classifier was constructed using the training data and then the classifier was used on the test data set to classify the data as an face ID or not. The instruction sets used to create an optimal HRBF classifier is

Fig. 2. Example in ORL face dataset (left), and example in YALE data set (right)

Table 1. Comparison of different approaches for ORL face recognition (test)

Method	Recognition rate
PCA+RBF [14]	94.5%
LDA+RBF [14]	94.0%
FS+RBF [14]	92.0%
NN [15]	94.64%
PCA [15]	88.31%
LDA [15]	88.87%
DCT+HRBF (this paper)	97.68%

Table 2. Comparison of different approaches for Yale face recognition (test)

Method	Recognition rate
NN [14]	83.51%
PCA [14]	81.13%
LDA [14]	98.69%
DCT+HRBF (this paper)	98.95%

$S = \{+_2, +_3, \ldots, +_6, x_0, x_1, \ldots, x_{59}\}$. Where $x_i (i = 0, 1, \ldots, 59)$ denotes the 60 features extracted by DCT.

A comparison of different feature extraction methods and different face classification methods for ORL face dataset (average recognition rate for 20 independent runs) is shown in Table 1. Table 2 depicts the face recognition performance of the HRBF by using the 60 features for Yale data set. The HRBF method helps to reduce the features from 60 to 6-15. For each experiment, the true positive rate (fp), false positive rate (fp) were also computed. For save space, they're not shown here.

6 Conclusions

In this paper DCT based feature extraction method and HRBF classification model are proposed for face recognition. The ORL and Yale database images are used for conducting all the experiments. Facial features are first extracted by the DCT which greatly reduces dimensionality of the original face image as well as maintains the main facial features. Compared with the well-known PCA approach, the DCT has the advantages of data independency and fast computational speed. The presented HRBF model for face recognition with a focus on improving the face recognition performance by reducing the input features. Simulation results on ORL and Yale face database also show that the proposed method achieves high training and recognition speed, as well as high recognition rate. More importantly, it is insensitive to illumination variations.

Acknowledgment

This research was partially supported the Natural Science Foundation of China under contract number 60573065, and The Provincial Science and Technology Development Program of Shandong under contract number SDSP2004-0720-03.

References

1. Zhao, W., Chellappa R., Rosenfeld A. and Phillips P.J.: Face Recognition: A literature survey. Technical Report CART-TR-948. University of Maryland, Aug. 2002.
2. Chellappa R., Wilson, C.L., and Sirohey, S.: Human and machine recognition of faces: A survey, Proc. IEEE, 83(5): 705C740, 1995.
3. Valentin D., Abdi H., Toole, A. J. O., and Cottrell, G. W.: Connectionist models of face processing: A survey, Pattern Recognit., 27: 1209C1230, 1994.
4. N. A. Mat Isa, Mashor, M. Y., and Othman, N. H., "Diagnosis of Cervical Cancer using Hierarchical Radial Basis Function (HiRBF) Network," In Sazali Yaacob, R. Nagarajan, Ali Chekima (Eds.), Proceedings of the International Conference on Artificial Intelligence in Engineering and Technology, pp. 458-463, 2002.
5. S. Ferrari, I. Frosio, V. Piuri, and N. Alberto Borghese, "Automatic Multiscale Meshing Through HRBF Networks," IEEE Trans. on Instrumentation and Measurment, vol.54, no.4, pp. 1463-1470, 2005.
6. Z. Ahmad, J. Zhang, "Bayesian selective combination of multiple neural networks for improving long-range predictions in nonlinear process modelling," Neural Comput & Applic. Vol. 14. pp. 78C87, 2005.
7. Yang F. and Paindavoine M.: Implementation of an RBF neural network on embedded systems: Real-time face tracking and identity verification, IEEE Trans. Neural Netw., 14(5): 1162C1175, 2003.
8. Sorwar G., Abraham A. and Dooley L., Texture Classification Based on DCT and Soft Computing, The 10th IEEE International Conference on Fuzzy Systems, FUZZ-IEEE'01, IEEE Press, Volume 2, pp. 545 -548, 2001.
9. R. Storn, and K. Price, "Differential evolution - a simple and efficient adaptive scheme for global optimization over continuous spaces," Technical report, International Computer Science Institute, Berkley, 1995.
10. K. Price, "Differential Evolution vs. the Functions of the 2nd ICEO," In proceedings of 1997 IEEE International Conference on Evolutionary Computation (ICEC'97), Indianapolis, USA, pp. 153-157, 1997.
11. Chen, Y., Yang, B. and Dong, J.: Nonlinear System Modeling via Optimal Design of Neural Trees. International Journal of Neural Systems. 14 (2004) 125-137
12. Chen, Y., Yang, B., Dong, J. and Abraham, A.: Time-series Forecasting using Flexible Neural Tree Model. Information Science. 174 (2005) 219-235
13. Sastry, K. and Goldberg, D. E.: Probabilistic model building and competent genetic programming. In: R. L. Riolo and B. Worzel, editors, Genetic Programming Theory and Practise. (2003) 205-220
14. Su, H., Feng D., Zhao R.-C.: Face Recognition Using Multi-feature and Radial Basis Function Network, Proc. of the Pan-Sydney Area Workshop on Visual Information Processing (VIP2002), Sydney, Australia, 183-189, 2002.
15. Huang, R. and Pavlovic, V. and Metaxas, D.N: A hybrid face recognition method using markov random fields, ICPR04, pp. 157-160, 2004.