Face Recognition based on Radial Basis Function and Clustering Algorithm

Yuanfeng Gao

ECE 539 Project
Fall 2008
Abstract

This project consists of two parts. The first part is a general review of the previous and current research on human face recognition, including initial motivation, approaches, major problems and solutions, etc. The second part propose a new method for learning of radial basis function (RBF) neural networks which is based on subtractive clustering algorithm (SCA) and its application to face recognition. Experiments on face recognition using ORL database show feasibility of the method. Results present that RBF neural networks classifier using proposed algorithm is more precise and faster than corresponding one using general K-means clustering algorithm.
目录

1. Introduction ....................................................................................................... 4
2. Face localization methods .................................................................................. 5
   2.1 Knowledge based methods ........................................................................... 5
   2.2 Feature Based Methods ............................................................................... 6
       2.2.1 Moment Invariants .............................................................................. 6
       2.2.2 Differential invariants ........................................................................ 7
       2.2.3 Fourier Descriptor .............................................................................. 8
       2.2.4 Summation invariants .......................................................................... 8
       2.2.5 Dyadic wavelet invariants ................................................................. 9
3. Two-Dimensional and Three-Dimensional Face Recognition ....................... 10
   3.1 Overview of 3D Face Algorithms .............................................................. 10
   3.2 Major Challenges and Possible Solutions ................................................. 11
4. Image Resolution and Face Recognition ....................................................... 12
5. Neural network approaches to Face Recognition ............................................ 13
   5.1 RBF neural networks model ....................................................................... 13
   5.2 Subtractive Clustering Algorithm (SCA) .................................................. 14
   5.3 Applying SCA to RBF Neural Networks .................................................... 15
   5.4 Application in Face Image Recognition and Experiment ......................... 15
       5.4.1 Preprocessing and Feature extraction of Face images ....................... 15
       5.4.2 Networks Parameters .......................................................................... 16
       5.4.3 Experiments and Analysis ................................................................... 17
       5.4.4 Conclusion ............................................................................................ 18
1. Introduction

New information technology and media have always attracted the attention of the public and has changed the modern world greatly. Among them, more effective and friendly methods for Human-Computer interaction have been developed, and have distinguished themselves from traditional methods that rely on devices such as keyboards, mice and displays. Furthermore, with the decreasing video image acquisition cost that results in the ever-increasing performance/price ratio, computer vision systems have been deployed in desktop and embedded systems [1]. As a result, research of human face recognition by artificial intelligence that once been limited by restraints of apparatus, has now been rapidly expanding over the last decade. It has attracted much attention though this study has already been worked on for more than twenty years by neuroscientists, engineers, and psychophysicists. Face recognition has found a wide range of applications, and there has been a growing interest in machine recognition of faces due to potential commercial application, such as secure monitoring, person identification, law enforcement, object identification and tracking etc. The newest image processing software, “Iphoto 09 ™” by Apple®, could detect different faces from photos and organize photos by faces in the picture which made sorting photos much easier than before. Over no more than two decades, sales of identity verification products rocketed from $100 million in 1994 [2] to $500 million in 2007 [3]. Face recognition is one of the hottest research topics nowadays and numerous face recognition approaches have been presented so far.

The first step of any face recognition system is to detect the locations of different parts of face in the image. For instance, the positions of forehead, nose, cheek, mouth, chin, etc. However, this is a rather challenging task and has drawn a lot of attention. Reasons basically fall within two areas: 1. the environmental reason includes factors such as the variety of scale, location, illumination conditions, image resolutions, and angles. 2. The other reason is related to the human in the image, such as pose (frontal or profile), different expressions, and occlusion. After successfully overcoming the difficulties listed above, the next step usually involves data processing and storage. Depending on the approaches, different “name cards” for each person would be created, and could be used as identification of the same person in the future. Different approaches would result in different recognition rate, ability to reduce noise, and processing time.

However, evaluations such as the Face Recognition Vendor Test (FRVT) 2002 [4] also made it clear that the current state of the art in face recognition is not sufficient for the more demanding applications nowadays. The vast majority of face recognition research use normal greyscale intensity images of face, which are referred to as “2-D” images, in contrast to “3-D”. Combinations of “2-D” and “3-D” multi-models are been used in recent research and the term “multi-modal biometrics” refers to these kinds of multiple imaging modalities.
2. Face localization methods

In this section, we review several existing techniques to localize faces from a single intensity or colour image.

2.1 Knowledge based methods

Human could recognize different faces based on the normal knowledge of what constitutes a typical face. These rule-based methods encode this human knowledge and capture the relationships between facial features. It’s easy to come up with some simple rules to detect faces in an image: faces often appear with two eyes that are symmetric to each other, a nose, and a mouth. Although these rules differ in different persons since the typical “landmarks” in different person would have different sizes, different distances in between, and might have different skin colours. Relative distances and positions are used to calculate the relationship between features. Before any further progress could be made, facial features in an input image are extracted first and face candidates are identified based on the rules. However, a verification process is usually applied to prevent false detections.

The major problem with this approach is that there is no certain universal rule to apply human knowledge. It came to a dilemma that if the rules were too general, different person’s faces would be recognized as same ones, and if the rules are too detailed it might fail to detect faces that should pass the rules. Therefore a hierarchical knowledge-based method to detect faces is derived by Yang and Huang [5]. Their system consists of three levels of rules. The lowest level is the most general rule and scans the window to look for all possible faces. After the lowest level is applied, the next level performs the local histogram equalization on the face candidates, followed by edge detection. The highest rule relies on the details of the facial features, and would be applied to the surviving candidates. Different resolution images are also used for each level. The lowest resolution image is used at the lowest level to search for face candidates and would be further processed at better resolutions if initial result satisfies the rules of the previous level. The results were rather acceptable at that time, with a test set of 60 images, it managed to locate 50 images and 22 of which succeeded in recognition. The correct rate is low but the interesting feature is that a “coarse-to-fine” and “focus-of-attention” strategy is used to reduce the computation. The idea of using multi-resolution images also benefits late face detection works.
2.2 Feature Based Methods

In contrast to the knowledge-based methods, research has been done to find invariant features of faces for detection. The idea came from that human could easily detect faces and objects even if the illumination condition and poses changed. So there must exist properties or features that remain unchanged over different situations, or, could change in a predictable way. Feature invariant approaches search for face structural features that are invariant to changes in pose, viewpoint, illumination, and expression. An example of largely adopted feature is the skin colour; several works [6], [7], [8] suggest modelling the skin colour distribution with a Gaussian mixture model. Other facial features such as forehead, eyebrows, eyes, nose, cheeks and mouth also extracted as invariant features using edge detectors. The face image is usually divided into small regions that contain the extracted invariant features and a statistical model is built. As a result, the feature of each region, together with relationships between these regions suggest different facial expressions, illumination condition, viewpoints, etc. [9] The drawback of this kind of method is that image features could be severely destroyed due to bad illumination condition, noise, and other occlusion, and the boundaries between features could be too weak to detect while the shadows could produce strong fake edges. Image invariants can be designed to fit the needs of specific systems. Some require only that it be non-discriminating to an object’s geometric pose or orientation. Others may be only interested in it being insensitive to the change of illumination. More complex systems however demand that it be insensitive to a combination of several environmental changes. Clearly the latter case is more difficult to achieve.

The next section introduces and compares some common invariant features.

2.2.1 Moment Invariants

Ming-Kuei Hu first presented the mathematical foundation of two-dimensional moment invariant and their applications for planar geometric figure in 1962 [10]. Moment invariants are useful features of a two-dimensional image since they are invariant to shifts, to change of scale and rotations, truly independent of position, size and orientation. Particularly, Hu defined seven values that are invariant to the position, scale and orientation of the object. It is defined globally, and the moment requires the whole shape to accomplish calculation. This causes occlusion problems and is the main drawback of moment invariants used for object recognition.

Moment invariants are properties of connected regions in binary images that are invariant to translation, rotation and scale. They are useful because they define a simply calculated set of region properties that can be used for shape classification and part recognition.
Given the intensity function of an image \( f(x,y) \), which is assumed to be piecewise continuous and with compact support, one could define the regular moments \( m_{pq} \) as
\[
m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q p(x, y) dx dy
\]
\[p, q = 0, 1, 2 \ldots \]
The density distribution function \( f(x,y) \) could have nonzero values only in the finite part of the \( x,y \) plane, then moment of all orders exist and the uniqueness theorem could be proved. The use of moment invariants makes possible the derivation of models, which automatically generate additional dimensions for the purpose of discrimination or eliminating noise.

### 2.2.2 Differential invariants

Differential Invariants is one of the most popular invariant features that had been studied and widely used. It is locally defined to accommodate occlusions and the large feature space improves discrimination performance. Since this method requires high order derivatives that amplify the effect of noise, several methods have been introduced to reduce the noise. The semi-differential Invariants developed by Van Gool et al. [11] combine coordinates and their derivatives with respect to some contour parameter at several points of the image contour, thus allowing an optimal trade-off between identification of points and the calculation of derivatives. A high order differential invariant could be approximated by a joint invariant depending on lower derivatives evaluated at several points on a curve. In such an approach, a calculation of high order derivatives could be avoided.

When differential invariants are applied to two dimension images, images are given in the form of sampled intensity values rather than in the form of closed formulas where derivative could be computed analytically. Gaussian derivatives are used to combine filtering with the computation of derivatives.

For example [12], Let \( I_1(x,y) \) and \( I_2(u,v) = I_2(\alpha x, \alpha y) \), which shows that these two images are related by a scaling factor \( \alpha \). According to Schmid and Mohr [13],
\[
\int_{-\infty}^{\infty} I_1(x, y) \otimes G_{i_1 \ldots i_n}(x, y; \sigma) dx dy = \alpha^n \int_{-\infty}^{\infty} I_2(u, v) \otimes G_{i_1 \ldots i_n}(u, v; \alpha \sigma) du dv
\]
Where the \( G_{i_1 \ldots i_n} \) are partial derivatives of the 2D Gaussian. Also, rotational invariance is a highly desirable property in most image retrieval tasks. However, while derivatives are translation invariant, the partial derivatives are not rotationally invariant. To solve this problem, there are also some
2.2.3 Fourier Descriptor

Fourier descriptor is used for shape description and is based on the boundary of an object. Therefore the boundary of the object must be determined and the input image segmented before calculation could be done. The image is first placed in a complex plane, and then an array of complex numbers would be used to present the boundary of the object. The numbers correspond to the pixels of the boundary of the image, and Fourier descriptors are calculated by combining Fourier transform coefficients of the complex array. To control the sensitivity to noise, some detail coefficients could be ignored and the number of the coefficients is determined experimentally. The drawback of Fourier descriptor is that Fourier transform is based on the whole boundary of the object and could not extract local characteristics in the space domain.

Any plane curve could be described in the real space or in the complex plane C. Under these two representations, an affine transform may be written as

\[ X = Ax^0 + b \quad \text{det} (A) \neq 0 \]

Where \( x \) and \( x^0 \) is in the real space and \( A \) is a \( 2 \times 2 \) matrix, \( b \) is a 2-vector and \( x \) is the affine-transformed version of \( x^0 \) or using the complex representation,

\[ X = ax^0 + bx^0* + c \quad aa* - bb* \neq 0 \]

Where * represents the complex conjugate operation. Some previous papers used complex representation under similarity transform, and it is easy to proof that \( a \) can be eliminated in the Fourier domain by simple normalizations. However, this method can’t extract local characteristics in the space domain since Fourier transformation requires the whole contour. As a result, the feature space would be much larger, and for 3-D planar object, an orthographic projection could be exactly described by an affine transform [14].

2.2.4 Summation invariants

The Summation Invariants introduced by Weiyang Lin [15], are obtained by the method of moving frames, and are based on the summation operation of discrete data. It is similar with integral invariants in the way of systematic production of invariants and low noise sensitivity. In this approach, Euclidean and affine transformation groups are used in illustrating surfaces and curves for both 2-D and 3-D space.

Using FRGC 3D image database, the results using summation invariants obtained higher recognition performance than FRGC 3D baseline algorithm.
[16], and could be obtained using the nose region of the face rather than the whole face region.

In some application of recognition of similar objects, a global recognition result is not required and it won’t give accurate recognition results either, so semi-local summation invariant is defined as follows:

$$\beta[m] = (M(x_1 y_1 - x_0 y_0) + P_x (y_1 - y_0) - P_y (x_1 - x_0))^2$$

$$P_x = \sum_{n=m}^{m+M-1} x[\text{mod}(n, N)]$$

$$P_y = \sum_{n=m}^{m+M-1} y[\text{mod}(n, N)]$$

$$x_0 = x[m]$$

$$x_1 = x[\text{mod}(m + M - 1, N)]$$

$$y_0 = y[m]$$

$$y_1 = y[\text{mod}(m + M - 1, N)]$$

We could use the denominator of $\alpha$ to define a semi-local summation invariant. Normalized cross-correlation would cancel the scaling factor:

$$\rho = \frac{\sum_{n=0}^{N-1} \beta_1[n] \beta_2[n]}{\sqrt{\sum_{n=0}^{N-1} \beta_1^2[n] \sum_{m=0}^{N-1} \beta_2^2[m]}}$$

### 2.2.5 Dyadic wavelet invariants

Wavelet is a well-established tool in the field of image analysis. In particular, wavelet orthogonal and biorthogonal bases have been successfully used in many applications. It has been accepted that lack of translation invariance could cause severe drawback for pattern recognition. Dyadic wavelet invariants introduced by Mallat and Zhong [17] are used to derive an affine invariant function. It was proved very efficient when analyzing multiscale features. Dyadic wavelet invariants represent a signal in multiple resolution levels and the noise sensitivity could be adjusted by changing the numbers of resolution levels. However, this is based on approximation and will lose part of the data. The invariant function is normally based on analyzing the object boundary using the dyadic wavelet transform.

Let $f$ be a square integrable 2-D signal, and its 2-D Fourier transform defined by

$$f(\omega) = \int_{\mathbb{R}^2} d^2 x f(x) e^{-i\omega x}$$
And \( \otimes \) represents the 2-D convolution product, it could also be written as

\[
(f \otimes g)(x) = \int d^2y f(x-y)g(y)
\]

2-D wavelet integrable or square integrable complex-valued function of the 2-D real variable \( x \) that has \( N \) vanishing moments:

\[
\iint x^p y^q \psi(x,y) dx dy = 0, 0 \leq p, q \leq N
\]

Wavelets are thus orthogonal to function of the form:

\[
P(x, y) = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} r_{k,l} x^k y^l
\]

Which shows that they are immune to polynomials of order up to \( N-1 \) in the \( x \) and \( y \) variables.

3. **Two-Dimensional and Three-Dimensional Face Recognition**

Recently, most of the research of face recognition has focused on the use of two-dimensional intensity images. However, experiments show that 3D images show more useful aspects and are invariant to condition change. Therefore 3D face recognition is attracting more and more researchers nowadays. This section gives an overview of 3D face recognition and the major challenges and possible solutions to them.

3.1 **Overview of 3D Face Algorithms**

Cartoux et al [18] found a plane of bilateral symmetry through the face image and this plane is used for normalization for pose. The method is based on principal curvature and they search for matching between the face surface and the plane of symmetry.

Gaussian curvature is also popularly applied in algorithms. Lee and Milios [19] created an EGI (Extended Gaussian Image) for each convex region. A probe image is defined as the target image which is to be recognized and a gallery image is defined as the image enrolled in the system. A match between a region in a probe image and a gallery image is done by correlating EGI. Convex regions should change less than the other regions when face expression appears. This also suggests some ability to cope with face expressions in the future. The drawback of this method is that it is not sensitive to the size of the object so it would be possible that two faces that have similar shape but different sizes would be not easy to distinguish via this method.

Another method using Gaussian curvature is suggested by Gordon [20], which used the nose region and ridge and valley lines to register the image to a standard pose. Then the difference between probe image and gallery image is
computed by comparing the volume differences. It solves the problem from the method by Lee and Milios, and could recognize faces that are similar in shape but different in size. However, it needs extension to cope with changes in facial expression.

A method to produce 2D images from 3D images is suggested by Chang et al. [21]. Under different illumination conditions, they used 3D images of a person and weighted sums of the three normal components of the 3D images. The correlation could be performed with a 2D probe image to achieve illumination insensitive recognition. The PCA-based recognition experiments performed using a dataset of 2D and 3D images from two hundred persons. Several experiments were performed using 2D alone, 3D alone and 2D+3D multi-modal. The result shows that the highest recognition ranking (performance) was the approximately 99% of the multi-modal and 94% for 3D, while the 2D experiment was the lowest with 89%.

3.2 Major Challenges and Possible Solutions

The lack of sensitivity to size variation has been one of the limitations to existing 3D face recognition methods. Approaches that use Gaussian curvatures are not able to distinguish between faces that have similar shapes but different sizes. And approaches that could cope with size variation, such as PCA-type algorithm could handle size differences, but the change of facial expression between probe image and gallery image made it hard to distinguish.

How to cope with facial expressions still remains as one of the biggest challenge in face recognition research. An experiment was conducted by W. Bowyer and K. Chang [22] to show the variation in facial expression between target image and gallery images. The experiment focuses on the effect of expression changes. The research was done with the PCA-based 2D and 3D algorithms. Under the condition that no expression change occurred between target and gallery images, both 2D and 3D result in rank-one-recognition rate of over 90%. However, there is a significant drop in performance when expression change is occurred, and the rate is 73% for 2D and 55% for 3D. The reason is that 3D recognition algorithm effectively assumes the face as a rigid shape and the facial expression actually has greater impact on 3D algorithms. The facial expression is a major cause of performance degradation, but research that copes with this issue is starting to be addressed in the next generation of algorithms.

Another obstacle to the implementation of 3D face recognition algorithms is the lack of 3D database. First of all, not many research institutes have 3D scanning device and secondly, obtaining a qualified dataset includes the following aspects: (1) the images from a person should be taken more than once, ideally taken at repeated intervals of time, for instance, once a week over a whole year. This allows the variation of age, occlusion, and color of skin etc. to be included in the dataset, therefore provide more applicability; (2) a large
number of persons show up in the data acquire process to provide variation in face patterns; (3) images of a person with different facial expression. Some datasets include facial expressions made by hired professional actors; (4) high spatial resolution; (5) low frequency of sensor-specific artifacts in the data.

The 3D face recognition is an area that expands greatly and has lots of important potential applications. The idea of breaking the whole face region into small regions and calculating the relationships between would be a possible solution to improve the 3D algorithms.

4. Image Resolution and Face Recognition

In the past, it is difficult to obtain images that have good qualities of recordings of faces. But nowadays, rapidly expanding digital technology allows us to obtain images at high resolutions, and good quality images could even be captured in video surveillance systems. Therefore it is popular accepted that higher resolution image would lead to better recognition rate since more information could be extracted from the image. However, we decide to investigate the lowest resolution at which a face recognition system could still achieve acceptable performance.

Currently available face recognition systems usually require face images with more than 50 pixels between the eyes. Zhao et al [23] use a combination of PCA and LDA for face recognition on a resolution of 24 × 21 pixels and claim that their approach will even give good results on 19 × 17 pixels. Kukharev et al [24] report that the images should be larger than 28 × 23 pixels using PCA and LDA. By using simple downsampling or taking the mean pixel value, the image would still contain high frequency components. According to the result by B. J. Boom and G. M. Beumer, their registration algorithm performs best on the upcaled image with a resolution of 32*32 pixels; the landmark finding wasn’t improved by training on the same resolution as used for testing. Other registration methods may behave differently under various resolutions. This confirms that accurate registration is of vital importance for face recognition.

Real time security and surveillance due to certain limitations and restrictions (like constrained environment, speed of system and its accuracy) have made this area of research more attractive and challenging for biometric researchers. Proposed system of face recognition has been developed to analyze the effects of varying resolution on recognition which in result provides better success rate besides enhancing speed. Results have been obtained by using five images as a standard for training purposes. Variations in number of training images affect both the success rate and speed of system.
5. Neural network approaches to Face Recognition

Neural network approaches are one of the most important methods to solve face recognition problems. All elements of neural network can work simultaneously, and this parallel functioning is very useful which helps to solve image processing tasks effectively. Neural network approaches have advantage over other methods since it could recessively express many rules for face recognition and has much stronger adaptability by training different networks.

Radial Basis Function (RBF) is introduced in our ECE 539 course and after learning it, it gave me the motivation to apply it to face recognition research due to its computational simplicity and robust generalization. Conventionally, K-means clustering algorithm could be applied to find RBF centres which are the most important parameters. The algorithm which randomly choice initial centres makes RBF Networks easily converge to local minimum. The goal of this project is to implement an improved method that subtractive clustering method is used to obtain the centre parameters of RBF networks. Using the algorithm, number of data points and calculation is irrelevant to the dimension of the considered information. Also, the algorithm works well on high-dimensional face patterns since it could effectively pick up training speed. Compared with K-means clustering algorithm, the algorithm could automatically terminate clustering criterion to ascertain RBF numbers and improve classification accuracy.

5.1 RBF neural networks model

RBF neural networks have a feed forward architecture with an input layer, a hidden layer, and an output layer. The nodes of hidden layer are composed of radial function which performs the nonlinear mapping from input nodes to hidden nodes. The structure of RBFNN model is shown in Fig. 1.

![Fig.1. Structure of RBF neural networks model](image-url)
The \( j \) th neuron activity of output layer is given by the following equation:

\[
y_j = w_{0j} + \sum_{i=1}^{J} w_{ij} \psi(||X - c_i||)
\]

\( j=1,2,...,J \)

Where \( X = (x_1, x_2, \ldots, x_M)^T \) is an input sample, and \( w_{ij} \) weights between hidden layer and input layer. The radial function would have various forms, but as stated in the previous sections, Gaussian function is the most general function:

\[
\psi = \exp\left(-\frac{1}{2\sigma^2} \sum_{i} ||X - c_i||^2 \right)
\]

In this way, each hidden node is defined by two parameters: its centre \( c_i \) and the width of the radial function.

### 5.2 Subtractive Clustering Algorithm (SCA)

For an \( n \)-dimensional real space, the data points would be in the form of \((x_1, x_2, \ldots, x_m)\), and the definition of each component \( x_i \) is expressed by the following formula:

\[
D_i = \sum_{j=1}^{m} \exp\left[-\frac{||x_i - x_j||^2}{(\gamma_a / 2)^2} \right]
\]

According to clustering theory, \( \gamma_a \) is a proper positive constant that decides a neighborhood of \( x_i \). The data points outside radius have little effect on the density of each point and if the data point has maximum density, it should have a lot of adjacent data points too.

The first step here is to calculate the density of each data point. The first clustering center that has maximum density is been selected and if \( x_{c1} \) is the selected point and \( D_{c1} \) is its density, then \( D_i \) is revised by the following formula:

\[
D_i = D_i - D_{c1} \exp\left[-\frac{||x_i - x_{c1}||^2}{(\gamma_b / 2)^2} \right]
\]

And \( \gamma_b \) decides a neighborhood where density is greatly reducing, and to avoid getting too close to clustering centers, we define that \( \gamma_a \) is smaller than \( \gamma_b \). Clearly, the data points that are close to the first clustering center \( x_{c1} \) would be almost impossible to become the next clustering center.
The second step gets the next data point \( x_{c2} \) in the same way and also revises the density. When the latest clustering center includes very few data points, the clustering centre could be omitted and the clustering should be terminated. Therefore the operation is repeated until termination criterion of the clustering as follows:

\[
D_{\text{max}} / D_{c1} \leq \xi
\]

### 5.3 Applying SCA to RBF Neural Networks

The center parameters and width parameters are essential in designing RBF neural networks. The weights between the hidden layer and input layer could be obtained by linear equations after the two groups of parameters have been determined.

With the selected face image samples for training networks, we could determine the first clustering center. The density of each sample is calculated and \( D_i \) is saved in the set \( A(c) \) for next iteration. The data point that has the maximum density would be selected as the first clustering center \( x_{c1} \). Set \( H=1 \).

The formula is now used to revise the density of each data point and then to find the maximum density \( D_{\text{max}} \).

\[
D_i = D_i - D_{cH} \exp[-\frac{\|x_i - x_{cH}\|^2}{(\gamma_{cH} / 2)^2}]
\]

After this, the termination rule is applied here to see whether \( D_{\text{max}} / D_{c1} \leq \xi \).

\( A(c) \) rejects the data point with \( D_{\text{max}} \). And after this step, the clustering is ended. Otherwise, \( A(c) \) accepts the data point with \( D_{\text{max}} \) and the point is the \( H \) th clustering center \( x_{cH} \) and then \( D_i \) is saved. Set \( D_{cH} = D_{\text{max}} \), and then revise the density again until termination rule is satisfied.

### 5.4 Application in Face Image Recognition and Experiment

#### 5.4.1 Preprocessing and Feature extraction of Face images

As mentioned above in the previous section, input images are preprocessed and the features extracted to remove redundant information and reduce the complexity of networks structure and also improve the efficiency of training networks. Interpolation algorithm are applied to face image compression and the grey value of compressed image is normalized to \([0,1]\).

In this experiment, the ORL database of faces is used. The database was used in the context of a face recognition project carried out in collaboration with the Speech, Vision and Robotics Group of the Cambridge University Engineering Department. There are ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details.
(glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The files are in PGM format, and can conveniently be viewed on UNIX (TM) systems using the 'xv' program. The size of each image is 92x112 pixels, with 256 grey levels per pixel. The images are organized in 40 directories (one for each subject), which have names of the form sX, where X indicates the subject number (between 1 and 40). In each of these directories, there are ten different images of that subject, which have names of the form Y.pgm, where Y is the image number for that subject (between 1 and 10). The database can be retrieved from:

http://www.cl.cam.ac.uk/Research/DTG/attarchive:pub/data/att_faces.zip

Some example images from the ORL database is shown in the following figure:

After preprocessing, the face images are shrunk to 32*32 pixels and principal components extraction neural networks extracts 49 principal components from each image which can get primary information and effectively reduce the following networks calculation.

5.4.2 Networks Parameters

The input pattern vectors are 49-dimension face feature images, therefore the input layer of RBF networks consists of 49 nodes. And the center parameters are set to $\gamma_a = 0.6, \gamma_b = 0.8$. Four sets of training samples are used and we could obtain four set of clustering center values and width values, so there are different numbers of hidden layer nodes in different cases, and the output layer of RBF networks consists of 40 nodes. Width parameters of networks are obtained in terms of experiential formula and weights are learned by Least Square Method (LSM):

$$\sigma^2 = \frac{1}{I\sqrt{2}} \sum_{k=1}^{I} \frac{1}{L} \sum_{i=1}^{L} \left| c_i - c_k \right|, i = 1, 2, \ldots I$$
5.4.3 Experiments and Analysis

Four experiments are conducted to apply K-means clustering algorithm and subtractive clustering algorithm to learn RBF neural networks. Each experiment select different numbers of images from the ORL network to train the networks and the rest for testing.

<table>
<thead>
<tr>
<th>Sample Groups</th>
<th>Training Images</th>
<th>K-Means Algorithm</th>
<th>Subtractive Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>First image</td>
<td>49.79%</td>
<td>49.92%</td>
</tr>
<tr>
<td>2</td>
<td>First three</td>
<td>86.25%</td>
<td>90.76%</td>
</tr>
<tr>
<td>3</td>
<td>First five</td>
<td>90.13%</td>
<td>95.87%</td>
</tr>
<tr>
<td>4</td>
<td>Last three</td>
<td>91.03%</td>
<td>96.94%</td>
</tr>
</tbody>
</table>

Table 1 Recognition precision results of 4 sets of experiment

(1) Error curve of training networks in K-Means clustering algorithm

(2) Error curve of training networks in subtractive clustering algorithm
5.4.4 Conclusion

This part of project presents the method that RBF model neural networks using subtractive clustering algorithm, and has been applied to face recognition. Compared with traditional algorithm (K-means clustering algorithm) used in RBF neural networks, by experiment results the method shows improved recognition precision and speed. However, the proposed method just is proved effectively in ORL Database of Face in which images are normal, when expression and gesture of face and illumination changes greatly, the method cannot ensure high recognition precision.


22. An evaluation of multimodal 2D+3D face biometrics
   Chang, K.I.; Bowyer, K.W.; Flynn, P.J.; Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 27, Issue 4, April 2005 Page(s):619 - 624

