Global and Local Classifiers for Face Recognition

S. Anila
Research Scholar, Government College of Technology, Coimbatore, India
E-mail: anilasatish@gmail.com

N. Devarajan
Professor, Government College of Technology, Coimbatore, India
E-mail: profdevarajan@yahoo.com

Abstract

Face recognition means checking for the presence of a face from a database that contains many faces. The face images considered for recognition undergo large variations due to changes in illumination conditions, viewing direction, facial expression and aging etc. The face images have similar geometrical features and hence discriminating one face from the other in the database is a challenging task. Global and Local features are crucial for face recognition. In the proposed method, both the global and local features are extracted from the input face images. Global features are the holistically structural configuration of facial organs, as well as facial contour. Global Features are extracted from whole face images by keeping the low frequency coefficients of Fourier transform. Real and imaginary components in the low frequency band are concatenated into a single feature vector named Global Fourier Feature Vector (GFFV). Local features are high frequency and dependent on position and orientation of the face images. Local features are extracted by Gabor wavelets. Gabor features are spatially grouped into a number of feature vectors named Local Gabor Feature Vector (LGFV). Fisher’s Linear Discriminant (FLD) is separately applied to the Global Fourier Features and each local patch of Gabor features. The resultant vectors are fused using region based fusion algorithm. The processed test face image is verified for a match with the faces in the database and recognition is done.

Keywords: Global features, local features, DFT (Discrete Fourier Transform), Gabor wavelets, Image fusion, Face Recognition

1. Introduction

Human face recognition is an important area of image processing. It has the important applications in bioinformatics. Face recognition although trivial task for the human brain has proved to be extremely difficult to imitate artificially. It is commonly used in applications such as human-machine interfaces (HCI) and automatic access control systems. Face recognition involves comparing an image with a database of stored faces in order to identify the individual of that input image. The related task of face detection has direct relevance to face recognition because images must be analyzed and identified, before they can be recognized. Detecting faces in an image can also help to focus the computational resources of the face recognition system, optimizing the system speed and performance. As one of the most successful applications of image analysis and understanding, face recognition has recently gained significant attention especially during the past several years.
Face recognition is used for two primary tasks: Verification (one-to-one matching): When presented with a face image of an unknown individual along with a claim of identity, ascertaining whether the individual is who he/she claims to be. Identification (one-to-many matching): When an image of an unknown individual is given, that person’s identity can be determined by comparing (possibly after encoding) that image with the database of (possibly encoded) images of known individuals.

2. Previous Research

W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld (2003) surveyed the large number of face recognition algorithms that have been proposed for the past two decades. In the literature of face recognition, there are various face representation methods based on global features, including a great number of sub-space based methods and some spatial frequency techniques. M. Turk and A. Pentland (1991) study the principal component analysis (PCA), P. Belhumeur, J. Hespanha, and D. Kriegman (1997) investigate the Fisher’s linear discriminant (FLD) and M. Bartlett, J. Movellan, and T. Sejnowski (2002) investigate the Independent component analysis (ICA), and they have been widely recognized as the dominant and successful face representation methods. These methods attempt to find a set of basis images from a training set and represent any face as a linear combination of these basis images. J. Lai, P. Yuen, G. Feng (2001) and W. Hwang, G. Park, J. Lee, and S. Kee (2006) proposed the Spatial-frequency technique of feature extraction using Fourier transform, and Z. Hafed, M. Levine, and M. Savvides (2001), J. Heo, R. Abiantun, C. Xie, and B. Kumar (2006) investigated the techniques using discrete cosine transform. In this method, face images are transformed to the frequency domain and only the coefficients in the low-frequency band are reserved for face representation. While Global based face representations were popular for face recognition, recently, more and more attempts are made to develop face recognition systems based on local features, which are believed more robust to the variations of facial expression, illumination and occlusion etc.,

Penve and Atick (1996) proposed Local Feature Analysis (LFA) to encode the local topological structure of face images. LFA is considered as a local method as it utilizes a set of kernels to implicitly detect the local structure such as eyes, nose and mouth. A. Timo et al. (2004) adopted the local binary pattern (LBP) that is originated from texture analysis for face representation. In this method, LBP operator is first applied and then the resulting LBP “image” is divided into small regions from which histogram features are extracted. The idea of dividing face image is also used in the component based methods, in which the face images are divided into some blocks by a certain rule. Then the image blocks may be taken as inputs of classifiers or given to next step for further feature extraction (e.g., PCA, FLD).

Feature extraction means extracting the features from the image so that recognition is made accurate and easy. Both global and local features are crucial for face representation and recognition as suggested by Yu Su, Shiguang Shan, Xilin Chen and Wen Gao (2009). Feature extraction can be done by two methods; Global feature extraction and Local feature extraction.

Global and local facial features play different roles in face perception. Therefore, it is necessary to combine them together smartly. Intuitively, local information is embedded in the detailed local variations of facial appearance, while global information means the holistically structural configuration of facial organs, as well as facial contour. Thus, from the viewpoint of frequency analysis, global features should mainly correspond to the lower frequencies, while local features should be of high frequency and dependent on position and orientation in the face image. Considering that, global information is represented as the Fourier coefficients in low frequency band, and local information is encoded as the responses of multiscale and multiorientation Gabor wavelets. However, doing like this is not as computationally desirable as using Fourier transform directly. Specifically, we hope the global features should be compact and orientation-independent. Multiple Gabor wavelets are applied to achieve orientation-independent, thereby computational burden of global feature extraction will increase significantly. In addition, the high dimensionality of Gabor features also brings the problem of
“curse of dimensionality” and makes the following process much computationally expensive. That is the reason why Fourier transforms rather than tuned Gabor Wavelet is adopted to extract global features. Rabia Jafri and Hamid R. Arabnia (2009) surveyed the various Face recognition techniques and there are three categories: 1. Methods that operate on intensity images. 2. Methods that deal with video sequences. 3. Methods that require the sensory data such as 3D information or infra-red imagery. Among various local features, especially, Gabor wavelets have been recognized as one of the most successful local feature extraction methods for face representation due to their biological relevance. Typically face recognition methods based on Gabor features include the Elastic Bunch Graph Matching (EBGM), Gabor Fisher Classifier (GFC), AdaBoost based Gabor feature selection and local Gabor binary pattern (LGBP).

3. Proposed Technique
Global and local facial features play different roles in face perception. In the proposed technique, both local and global features are found and extracted and the flow chart is shown in Fig.1. Local information is embodied in the detailed local variations of facial appearance while global information means the holistically structural configuration of facial organs as well as facial contour.

**Figure 1:** Proposed Technique

```
Input Image (Test Image)

Discrete Fourier Transform (DFT)

Global Fourier Feature Vector (GFFV)

W

Gabor Wavelet Transform (GWT)

Local Gabor Feature Vector (LGFV)

W1, W2, W3, W4

Image Fusion

Test Image is not recognized

No

Matching Algorithm

Database

Test image is recognized

Yes
```
Global and Local Classifiers for Face Recognition

Global features are extracted from the whole face images by 2D Discrete Fourier Transform. Then, the real and imaginary components of the low frequency band are concatenated to form a single feature vector, called Global Fourier Feature Vector (GFFV).

For local feature extraction, Gabor wavelet transform is exploited. Gabor wavelets are used to extract local features at every position of the face image. These features are spatially grouped into a number of feature vectors, each corresponding to a local patch of the face image and called as Local Gabor Feature Vector (LGFV). After above process a face image can be represented by one GFFV and multiple LGFVs. These feature vectors encode diverse discriminatory information: GFFV contains global discriminatory information and each LGFV embodies discriminatory information within certain region.

FLD is applied to the vectors of GFFV and LGFVs. The current statistical features used to distinguish faces and non-faces can be divided into two categories: local features and global features. Some previous global-feature-based face detectors work very well for classifying frontal views of faces, but they are highly sensitive to translation and rotation of the face.

Local-feature-based face detectors can avoid this problem by independently detecting parts of the face. For instance, the changes in the parts of the face are small compared to the changes in the whole face pattern for small rotations. So local and global features are both important features. When the non-zero components of a feature are not too much, we call it a sparse feature. The number of non-zero components of a local feature is often largely smaller than the component number of the feature, so it is also a sparse feature. And when the number of non-zero components of a feature becomes larger, it evolves into a global feature.

Human face recognition mechanisms are: 1) Both holistic and local features are crucial for face recognition; 2) Global description and dominant features have different contributions; 3) Different facial features have different contributions to face recognition.

The proposed algorithm has been tested with different standard databases (FERET and FRGC v 2.0) and a few of them are shown in Fig. 2. The recognition rate is found to be very high.

**Figure 2:** FERET data base
3.1. Global Feature Extraction using DFT

Global based face representation was popular for face recognition. Global features describe the general characteristics of the holistic face and they are often used for the coarse representation. Global information is represented as the Fourier coefficients in low frequency band. Thus, from the frequency point of view global features correspond to low frequency as shown in the flow chart, Fig.3. In global based face representation, each dimension of the feature vector contains the information embodied in every part (even each pixel) of the face image, thus corresponds to some holistic characteristic of face. 2-D Discrete Fourier Transform is adopted for global feature extraction.

![Figure 3: Global Fourier Feature Vector(GFFV)](chart)

2D - Discrete Fourier Transform that transforms the image to frequency domain is applied as in the following expression.

\[
F(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)}
\]

Where, F(u, v) represents a 2-D image of size M by N pixels, u and v are frequency variables. 0 ≤ u ≤ M-1 and 0 ≤ v ≤ N-1.

When the Fourier transform is applied to a real function, its outputs are complex numbers, F(u,v) = R(u,v) + j I(u,v)

Where, R(u, v) and I(u, v) are the real and imaginary components of F(u, v) respectively. Thus the Fourier transform is represented by the real and imaginary components as shown in Fig.4.

![Figure 4: Application of DFT to an image](chart)
3.2. Local Feature Extraction using Gabor Wavelets

Face recognition using Gabor features has attracted considerable attention in computer vision, image processing, pattern recognition, and so on. Local information (the high frequency region) is represented by using Gabor wavelets. For local feature extraction, Gabor wavelets are exploited considering their biological relevance. In contrast, for the local based face representation, each dimension of the feature vector corresponds to merely certain local region in the face, thus only encodes the detailed traits within this specific area. Among various local features, especially, Gabor wavelets have been recognised as one of the most successful local feature extraction methods for face representation. Local features reflect and encode more detailed variation within some local facial regions such as mouth, eyes, nose.

Face recognition is used for surveillance and security, telecommunication and digital libraries and human-computer intelligent interaction. The Gabor wavelet representation facilitates recognition without correspondence because it captures the local structure corresponding to spatial frequency (scale), spatial localization, and orientations selectivity.

Gabor wavelets are defined as follows:

\[ g_{\mu, \nu}(z) = \frac{k_{\mu, \nu}}{\sigma^2} e^{-k_{\mu, \nu} z^2 / 2\sigma^2} \left[ e^{i k_{\mu, \nu} z} - e^{-\sigma^2 / 2} \right] \]

\[ k_{\mu, \nu} = k_{\nu} e^{i \varphi_{\mu}} \]

Where, \( k_{\nu} = \frac{k_{\nu}^{\text{max}}}{f_{\nu}} \) gives the frequency and \( \varphi_{\mu} = \frac{\mu \pi}{8}, \ \varphi_{\mu} \in [0, \pi] \)

Where, \( \mu \) and \( \nu \) define the orientation and scale of the Gabor kernel \( z = (x, y) \). \( k_{\mu, \nu} \) is the wave vector \( K_{\text{max}} \) gives the maximum frequency and \( f \) is the spacing between the kernels in the frequency domain.

Gabor wavelet consists of a planar sinusoid multiplied by a two dimensional Gaussian. The sinusoid wave is activated by frequency information in the image. The Gaussian insures that the convolution is dominated by the region of the image close to the centre of the wavelet. Therefore, compared with Fourier transform which extracts the frequency information in the whole face region, Gabor wavelets only focus on some local areas of the face and extract information with multi-frequency and multi-orientation in these local areas. Gabor wavelets can take a variety of different forms with different scales and orientations. That is, when a signal is convolved with the Gabor wavelet, the frequency information near the centre of the Gaussian is captured and frequency information far away from the centre of the Gaussian has a negligible effect.

Therefore, compared with Fourier transform which extracts the frequency information in the whole face region, Gabor wavelets only focus on some local areas of the face and extract information with multi-frequency and multi-orientation in these local areas. Gabor wavelets can take a variety of different forms with different scales and orientations.

**Figure 5:** Gabor wavelets of 5 scales and 8 orientations
3.3. Local Patches from the Images

As Gabor features are calculated by convolving Gabor wavelets with the whole face image, it covers all the positions of the face image. Thus, the local information provided by the spatial locations of Gabor features is lost when they are integrated to form one single feature vector. In order to reserve more location information, Gabor features are spatially partitioned into a number of feature sets named Local Gabor Feature Set (LGFS), each of which corresponds to a local patch of the face image. In addition, since each LGFV is relatively low dimensional, this can greatly facilitate the sequent feature extraction and pattern classification.

Human Faces contain some components with fixed high-level semantics such as eyes, nose and mouth. Consequently, the locality information is very meaningful for face modelling. Gabor features are spatially grouped into number of feature vectors named Local Gabor Feature Vector (LGFV) each of which corresponds to a local patch of the face image, also called as patch based representation.

3.4. Fisher Linear Discriminant (FLD)

A linear classifier is used to identify which class (or group) the object belong by making a classification decision, based on the value of a linear combination of the characteristics. An object's characteristics are also known as feature values and are typically presented to the machine in a vector called a feature vector.

After feature extraction, N+1 feature vectors are obtained, that is, one global Fourier Feature Vector (GFFV) and N Local Gabor Feature Vectors (LGFVs). Then, N+1 classifiers can be trained by applying FLD to each feature vector. These classifiers are named as component classifiers, opposite to the forthcoming ensemble classifier, i.e., the combination of component classifiers. N+1 Feature vectors contain diverse discriminative information for face recognition. Thus, component classifiers trained on these feature vectors should have certain degree of error diversity. In other words, these component classifiers might agree or disagree with each other when making decision. Considering that the ensemble classifier is generally superior to the single classifier when the predictions of its
component classifiers have enough diversity, the component classifiers trained on all the feature vectors are combined into a hierarchical ensemble classifier to improve the recognition accuracy. In Fig. 8, hierarchical ensemble method consists of two layers of ensemble: the ensemble of all the local component classifiers, and the ensemble of local classifier and global classifier. In the first layer, local ensemble classifier (LEC) is obtained by combining $N$ local component classifiers (LCC), each trained on an LGFV, with the number of selected patches. It is formulated as follows:

$$C_L = \sum_{i=1}^{N} \omega_{l_i} \cdot c_{l_i}$$

Where $W_{l_i}$ is the weight of the $i$th LCC. In the second layer, the LCC obtained in the first layer is combined with the global classifier (GC) trained on the GFFV to form the hierarchical ensemble classifier (HEC). As mentioned previously, global and local features play different roles in face perception. While global features capture the holistic characteristics of the face, therefore, better for coarse representation; local features encode more details in local face areas, therefore, better for finer representation. Considering that, in the proposed method, the input face image is normalized differently for global and local feature extraction. As shown in Fig. 5.2, the global Fourier features are extracted from the face image of lower resolution, but covering both external and internal facial features, especially the face contour. On the contrary, the local Gabor features are extracted from the face image of higher resolution, which covers only the internal facial features, e.g., the facial organs. The reason using this strategy lies in the sensitivity of Gabor features to the possible background introduced along with the contour, to which the Fourier features are very robust.

**Figure 8: Ensemble classifier**

3.5. Image Fusion

Vector fusion related to a same image or a same object becomes more and more essential in remote sensing applications. It is often necessary to associate additional and/or redundant information, in order to reject, confirm or create a decision. A definition of vector fusion was formulated by Bloch and Maître: “vector fusion is the joint use of heterogeneous information for the assistance with the decision-making”.
Image fusion is the process of combining relevant information from two or more images into a single image. The resulting image will be more informative than any of the input images. The image fusion techniques allow the integration of different information sources. Region based image fusion has been utilised. In this fusion method, level decomposion $l$ is always 1. The feature extracted images represent a coarse representation of the original image and may have many inherited properties of image. The information can be calculated based on mean intensity or based on some prior information so as to compare the performance of the algorithm. The averaging fusion rule is applied as follows:

$$Y_S(I) = \frac{Y_A(I) + Y_B(I)}{2}$$

Where $Y_A(I)$ and $Y_B(I)$ is an approximation coefficient of image A and B respectively. The result of 0 means all information is lost and 1 means all information is preserved. The vector-sum uses, captures and represents the contributions (and properties) of each vectorized dimension because each vector represents the measures, units, and properties of each dimension. This property or ability of each vector arises because each vector’s contributions are always accumulated in row sequence and functionally compared in the scatter-plots at each vector’s unique and specific phase angle.

### 3.6. Matching using Correlation Coefficient

The term “correlation” refers to a process for establishing relationships between two variables. A Correlation coefficient measures the strength and direction of a linear association between two variables. One of them does not “causes” the other. Correlation defines that when one variable changes, the other seems to change in a predictable way. A correlation coefficient is a “ratio” not a percent. However it is very easy to translate the correlation coefficient into a percentage. “Square the correlation coefficient” which means that you multiply it by itself. So, if the symbol for a correlation coefficient is “r”, then the symbol for this new statistic is simply “$r^2$” which can be called “r squared”. If the dots on the scatter plot tend to go from the lower left to the upper right it means that as one variable goes up the other variable tends to go up also. This is a called a “positive relationship”. On the
other hand, if the dots on the scatter plot tend to go from the upper left corner to the lower right corner of the scatter plot, it means that as values on one variable go up values on the other variable go down. This is called a “negative relationship”.

\[ R = \text{corrcoef}(X) \]

returns a matrix \( R \) of correlation coefficients calculated from an input matrix \( X \) whose rows are observations and columns are variables. The matrix \( R = \text{corrcoef}(X) \) is related to the covariance matrix \( C = \text{cov}(X) \) by

\[ R(i, j) = \frac{C(i, j)}{\sqrt{C(i, i)C(j, j)}} \]

Corrcoef \((X)\) is the zeroth lag of the normalized covariance function, that is, the zeroth lag of \( \text{xcov}(x, \text{’coeff’}) \) packed into a square array.

\[ R = \text{corrcoef}(x, y) \] where \( x \) and \( y \) are column vectors is the same as \( \text{corrcoef}([x \ y]) \). If \( x \) and \( y \) are not column vectors, \( \text{corrcoef} \) converts them to column vectors. For example, in this case \( R=\text{corrcoef}(x,y) \) is equivalent to \( R=\text{corrcoef}([x(:) \ y(:)]) \).

### 4. Results & Discussions

The proposed method is evaluated using different databases. Thus the table below shows that the recognition rate of the proposed technique lies above 90% and it is observed that the system is efficient. The table also implies that the False Acceptance Rate (FAR) is very low giving high detection rate.

<table>
<thead>
<tr>
<th>S.no</th>
<th>Database images tested</th>
<th>Number of input faces tested</th>
<th>Number of faces recognized correctly</th>
<th>Recognition rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FRGC (Face Recognition Grand Challenge)</td>
<td>50</td>
<td>46</td>
<td>92%</td>
</tr>
<tr>
<td>2</td>
<td>Yale database</td>
<td>40</td>
<td>38</td>
<td>95%</td>
</tr>
<tr>
<td>3</td>
<td>SCface (Surveillance Camera face database)</td>
<td>50</td>
<td>45</td>
<td>90%</td>
</tr>
</tbody>
</table>

When the proposed technique, that is the technique using both local and global classifiers is compared with the technique utilising only one classifier, the proposed classifier performs better. The false recognition rate is very low and the detection rate is high since both the local and global features are considered. The average detection rate is 92.3%.

### 5. Conclusion

Human beings recognize faces by global and local facial features. Global features are extracted from the whole face images by using Fourier transform, and the local features are emphasized on some spatially divided face patches by using Gabor wavelets. The local features seem to be significantly better than global features for face recognition.

In this face recognition method global and local features are extracted by effective Discrete Fourier transform (DFT) and Gabor Wavelet Transform (GWT) respectively. When both the local and global features are utilised the recognition rate increases than by use single feature alone. The classification errors are reduced using Fisher Linear Discriminant (FLD). The Fusion of the vectors is done by “Region Based Image Fusion”. The method of correlation coefficient is used for matching the test image with the database. The proposed algorithm can be applied to real time images as well.
References


