CHAPTER 5

GLOBAL AND LOCAL FEATURES FOR FACE RECOGNITION

5.1 INTRODUCTION

Face recognition, means checking for the presence of a face from a database that contains many faces and could be performed using the different features. The face images considered for recognition undergo large variations due to changes in the illumination conditions, viewing direction, facial expression and aging. The face images have similar geometrical features. Hence, it is a challenging task to discriminate one face from the other in the database.

Feature extraction, extracting the features from the image is an important step in face recognition, by which the recognition could be made more accurate and easier. Both global and local features are crucial for face representation and recognition.

Feature extraction can be done by two methods,

1. Global feature extraction
2. Local feature extraction

Global and local facial features play different roles in face perception. Therefore, it is necessary to combine them smartly together.
Naturally, local information is rooted 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 perspective of frequency analysis, global features should correspond to the lower frequencies, whereas local features should be of high frequency and dependent on position and orientation in the face image. The 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. It is known that the Gabor Wavelet is a Gaussian modulated Fourier Transform and hence can be adjusted to extract global (usually low frequency) features by increasing the bandwidth and the radius of its Gaussian modulator. However, doing this is not as computationally desirable as applying the Direct Fourier Transform. The global features should be compact and orientation-independent. Using Multiple Gabor wavelets, orientation-independent features could be obtained but results in computational complexity. Hence, Fourier transforms is adopted for global feature extraction, instead of tuned Gabor Wavelet.

Kim et al (2005) proposed a combined subspace method using both global and local features for face recognition. The global and local features were obtained by applying the LDA based method to the face image. The joint subspace is constructed with the projection vectors corresponding to large eigenvalues since they have more discriminating power.

Yu Su et al (2009) suggested that the two features such as global and local features are essential for face representation and recognition and Feature extraction can be done using these features. By keeping the low frequency coefficients of Fourier Transform, Global Features are extracted from the complete face images. Real and imaginary components in the low frequency band are concatenated into a single feature vector and the vector is
Global Fourier Feature Vector (GFFV). Local features are high frequency components and are 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 using correlation and recognition is done.

5.1.1 **Different Roles of Global and Local Features**

The different roles of global and local features are shown in Figure 5.1.

**Figure 5.1 Different roles of global and local features in face perception**
The figure shows how two different input images of the same person if interpreted by two different systems fail to give the correct recognition. The first system uses only the global features and the second system uses the local features and the analysis is explained in the present section.

The leftmost two input images are the images of the same person with the same components (eye, nose, and mouth) but different external contour, posture and accessories (glasses). Thus, they look globally very different in terms of the overall structural configuration, hair, glasses and face contour. Consequently, the classifier based on global features alone will report them as different persons. However, the classifier based on local features (eyes, nose and mouth) would report them as the same person, since their components are almost the same. Though the local classifiers perform better than the global, the combination of both the global and local classifiers give good improvement in the performance.

5.2 PROPOSED METHOD FOR FACE RECOGNITION USING LOCAL AND GLOBAL FEATURES

Global and local facial features play different roles in face perception. In the proposed technique, both global and local features are extracted. Global features are extracted using Discrete Fourier Transform (DFT) and Gabor Wavelet Transform (GWT) is used for extracting local features.

The flow chart of the proposed method for face recognition using both the global and local features is shown in Figure 5.2.
5.2.1 Global Feature Extraction using DFT

Global based face representation is one of the popular techniques for face recognition. Global features describe the general characteristics of the holistic face and they are often used for coarse representation.
In the proposed method, only the Fourier features in the low-frequency band are reserved as global features. Explicitly, for a face image, the real and imaginary components in the low-frequency band are combined into a single feature set, named Global Fourier Feature Set (GFFS) and the vector is the Global Fourier Feature Vector (GFFV).

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. 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 corresponding to some holistic characteristic of face. 2-D Discrete Fourier Transform (DFT) is adopted for global feature extraction formulated as shown in Equation (5.1).

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

(5.1)

where, \( F(u,v) \) and \( f(x,y) \) represents 2-D images of size \( M \times N \) pixels in the frequency domain and time domain respectively. (\( M \) and \( N \) represent the number of rows and columns of the image respectively). \( f(x,y) \) is the input image and \( F(u,v) \) is the image obtained after applying DFT. \( u \) and \( v \) are frequency domain variables, \( 0 \leq u \leq M-1 \) and \( 0 \leq v \leq N-1 \).

When the Fourier Transform is applied to a real function, its outputs are complex numbers,

\[
F(u, v) = R(u, v) + jI(u, v)
\]

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. GFFV is the Feature vector obtained using both the real and imaginary low frequency components as shown in Figure 5.3.

![Diagram](image)

**Figure 5.3 Global fourier feature vector**

Therefore, after Fourier Transform a face image is represented by the real and imaginary components of all the frequencies. Generally, low frequencies reflect the whole and overall features of the input image. When the inverse transform of the image is observed with only the lower frequencies with approximately 30% of the energy, it contains the global structural configuration of the facial organs and the contour regions of the face as shown in Figure 5.4.

(i)  
(ii)  
(iii)  

**Figure 5.4**  (i) Input image(from ORL database)  (ii) Real & imaginary Parts of DFT  (iii) Real & imaginary components of low frequency
5.2.2 Local Feature Extraction using Gabor Wavelets

Face recognition using Gabor features has attracted considerable attention in computer vision, image processing and pattern recognition. Gabor Wavelets can be used for representation of local information (the high frequency region). 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 merely to a certain local region in the face, thus it encodes only the detailed traits within this specific area. Among the various local features, especially, Gabor Wavelets have been recognized 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, and 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 (hence, no need for manual annotations) as it captures the local structure corresponding to spatial frequency (scale), spatial localization, and orientation selectivity and is shown in Equation (4.9) in section 4.1.4.

Gabor Wavelets can take a variety of different forms with different scales and orientations. Figure 5.5 shows the real part of the 40 Gabor wavelets with 5 scales and 8 orientations.
Gabor Wavelets with a certain orientation respond to edges and bars along this direction, and with a certain scale extract the information in the corresponding frequency band. Thus, Gabor Wavelets can extract more details in some important facial areas such as eyes, nose and mouth, which are very useful for face representation.

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. Gabor Wavelets are similar to Fourier Transform in many ways but differ in a few ways. The main difference is that the Fourier Transform extracts the frequency information in the whole face region whereas, Gabor Wavelets 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

Figure 5.5  Real part of the 40, 2-D gabor wavelets with five scales and eight orientations
information near the centre of the Gaussian is taken and frequency information far away from the centre of the Gaussian has an insignificant effect.

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 a 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.

For local feature extraction, Gabor Wavelet Transform is used. 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 images called the Local Gabor Feature Vector (LGFV).

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. N LGFVs, corresponding to N non-overlapping local patches in the face image, are constructed. In the technique, four local patches are selected which forms four LGFVs and is shown in Figure 5.6.
After the above processes, 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.

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. That is, when a signal is convolved with the Gabor wavelet, the frequency information near the centre of the Gaussian is captured whereas the 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

![Figure 5.6 Local gabor feature vector](image-url)
and multi-orientation in these local areas. Gabor Wavelets can take a variety of different forms with different scales and orientations.

### 5.2.3 Fisher Linear Discriminant (FLD)

A linear classifier is used to identify which class the object belongs to, by making a classification decision based on the value of a linear combination of 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. A simple classifier is shown in Figure 5.7.

![Figure 5.7 A simple classifier function](image)

In Figure 5.7, H₁, H₂, H₃ are the linear classifiers. The solid and empty dots can be correctly classified by any number of linear classifiers. H₁ classifies them correctly, H₂ can be considered better and it is furthest from the both groups (H₁, H₃). Classifier H₃ fails to correctly classify the dots.

In computerised face recognition, each face is represented by a large number of pixel values. Linear Discriminant Analysis is primarily used to
reduce the number of features to a more manageable number before classification. Each of the new dimensions is a linear combination of pixel values, which form a template. The linear combinations obtained using Fisher's Linear Discriminant are called Fisher faces.

PCA based schemes are useful only with respect to data compression and decorrelation of low (second) order statistics. PCA does not take into account the discrimination aspect. Hence, a good classifier should give low dimensional patterns and also a high discrimination index, characteristic of separable low-dimensional patterns.

Belhumeur et al (1997) explained that FLD is a general discriminant criterion used in several face recognition methods which defines a projection that makes the within-class scatter small and the between-class scatter large. Hence, FLD derives compact and well-separated clusters. Belhumeur et al (1997) suggested that since the original image space is high dimensional, PCA should be initially applied for dimensionality reduction and FLD transformation is then used to build the Most Discriminating Features (MDF) space for classification. FLD attempts to model the difference between the classes of data, and can be used to minimize the Mean Square Error (MSE).

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 namely, local features and global features. Some previous global-feature-based face detectors classify frontal views of faces well, 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. Hence, local and global features are important features. When the non-zero
components of a feature are not many, it is a sparse feature. The number of non-zero components of a local feature is often smaller than the component number of the feature. Hence, it is also a sparse feature. When it becomes larger, it evolves into a global feature. Human face recognition mechanisms are,

- Both whole and local features are crucial for face recognition.
- Global description and dominant features have different contributions.
- Different facial features have different contributions to face recognition.

After feature extraction, N+1 feature vectors are obtained, that is, one global Fourier Feature Vector (GFFV) and four Local Gabor Feature Vectors (LGFVs). Five classifiers can then 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 a certain degree of error diversity. In other words, these component classifiers might agree or disagree with each other when making a decision.

Considering that the Ensemble Classifier is generally superior to the Single Classifier in situations where 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. Hierarchical Ensemble Method consist of two layers of ensemble, the ensemble of all the Local Component Classifiers (LCC), and the ensemble of Local Classifier and Global Classifier (GC). In the first layer,
Local Ensemble Classifier (LEC) is obtained by combining $N$ Local Component Classifiers (LCC), $C_{Li}$ ($i = 1 \ldots N$) each trained on an LGFV, $L_i$ ($1 \ldots N$) with the number of selected patches ($N$ is the number of selected patches and in the proposed method since four patches are selected, $N = 4$). It is formulated as given in Equation (5.2).

$$C_L = \sum_{i=1}^{4} w_{Li} C_{Li} \tag{5.2}$$

where, $C_L$ is the Local classifier,

and $w_{Li}$ is the weight of the $i^{th}$ LCC.

In the second layer, the LCC obtained in the first layer is combined with the GC trained on the GFFV to form the hierarchical ensemble classifier (HEC) and is given by Equation (5.3),

$$C_H = W_G C_G + (1 - W_G) C_L \tag{5.3}$$

Where,

$C_H$ is the ensemble classifier

$C_G$ is the global classifier (GC)

$W_G$ is the weight of $C_G$.

As discussed in the previous sections, global and local features play different roles in face perception. While the global features capture the universal characteristics of the face, the local features encode more details in local face areas. Therefore, global and local features are used for coarse and finer representation respectively. Considering that, in the proposed method, the input face image is normalized differently for global and local feature extraction. The Global Fourier features are extracted from the face image of
lower resolution, but covering both the 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 for 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.

In Figure 5.8, Hierarchical Ensemble Method consisting of two layers of the ensemble is shown.

![Model of ensemble classifier](image)

**Figure 5.8** Model of ensemble classifier

### 5.2.4 Image Fusion

Vector fusion related to the same image or object becomes more and more vital in remote sensing applications. 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. Spatial domain approaches of image fusion produces spatial distortion in the fused image and error due to misregistration is prominent. Hence, region based algorithm is used that has many advantages as listed below,

- Less sensitive to noise
- Better contrast
- Less affected by misregistration

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 is functionally compared in the scatter-plots at each vector’s unique and specific phase angle. The shape of the locus is the composite functional relationship that exists for all m rows of the n dimensional vector.

### 5.2.5 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. Correlation defines that when one variable changes, the other seems to change in a predictable way. 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. 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 and is a negative relationship.

Uncorrelation has to be determined, which means the features should not be dependent from each other in order to provide discriminant information.

Uncorrelation is measured with the possible pairs of feature combinations, class to class, based on the covariance matrix:

$$C_i = \begin{pmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & & \vdots \\ C_{n1} & \cdots & C_{nn} \end{pmatrix}$$

Correlation Coefficient of two features $x_i, x_j$ is given by Equation (5.4)

$$\gamma_{ij} = \frac{c_{ij}}{\sqrt{c_{ii} c_{jj}}} , \quad -1 < \gamma_{ij} < +1$$

$$\gamma = \text{corrcoef}(X)$$

These variables will be more independent to each other when the correlation coefficient is close to zero. The correlation coefficient represents the normalized measure of the strength of linear relationship between variables.

The mean Correlation Coefficient is given by Equation (5.5)
\[ \gamma_{ij} = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} \]  

(5.5)

If \( \gamma \) is the correlation coefficient, then the relationship is obtained as follows:

- If \( \gamma = 1.00 \) there is a perfect positive relationship between the two variables.
- If \( \gamma \) is greater than 0.00 but less than 1.00, there is a positive relationship between the two variables.
- If \( \gamma = 0.00 \) there is no relationship between the two variables.
- If \( \gamma \) is between 0.00 and -0.50, there is a negative relationship between the two variables.
- If \( \gamma = -1.00 \) there is a perfect negative relationship between the two variables.

5.3 PERFORMANCE EVALUATION

Face images from ORL database and FRGC database are chosen for testing. FRGC database has 12,776 images with varying lighting conditions and expressions. ORL database contains 400 images of 40 subjects i.e. 10 faces per subject with variations in pose and expressions. The image Resolution is 112 × 92. Sample images from the ORL and FRGC database are shown in Figure 5.9.
Figure 5.9 Images from ORL and FRGC database

Global, gabor, local features and Graphical User Interface (GUI) developed for the system are shown in Figure 5.10.
5.3.1 Comparison of Recognition Rate for Various Features

Comparison is made by using local, global features separately and the combination of local and global features. It is observed that the usage of both the local and global feature combinations lead to higher recognition rate.
Table 5.1 shows the performance comparison of the three different features considered namely, local, global and the combination of local and global features. The recognition rate of the proposed system for ORL database is shown in Figure 5.11. The recognition rate is higher when the combination of local and global features is used than when either local or global features are considered separately.

Table 5.1  **Comparison of recognition rate for local and global features separately and local + global features**

<table>
<thead>
<tr>
<th>Recognition rate</th>
<th>FAR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Global features only</strong></td>
<td><strong>Local features only</strong></td>
</tr>
<tr>
<td>0.5</td>
<td>0.7</td>
</tr>
<tr>
<td>0.55</td>
<td>0.75</td>
</tr>
<tr>
<td>0.6</td>
<td>0.8</td>
</tr>
<tr>
<td>0.65</td>
<td>0.81</td>
</tr>
<tr>
<td>0.69</td>
<td>0.82</td>
</tr>
<tr>
<td>0.7</td>
<td>0.83</td>
</tr>
<tr>
<td>0.7</td>
<td>0.83</td>
</tr>
<tr>
<td>0.7</td>
<td>0.83</td>
</tr>
<tr>
<td>0.7</td>
<td>0.83</td>
</tr>
</tbody>
</table>

An average recognition rate of 64% using global features, 80% using local features and 92% using both global and local features is observed from the experiments. Hence, an increase of 28% than global features and an increase of 12% than local features has been observed when both the local and global features are used.
5.3.2 Performance of Hierarchical Ensemble Classifier (HEC)

GC and LEC are combined to form an Ensemble Classifier as explained in section 5.2.3 so that full use of both the global and local discriminant information could be made. In Equation (5.4), it could be understood that the GC, $W_G$ can actually balance the importance of global and local information. This is necessary because it is inferred that the performances of GC and LEC are quite different, as can be observed from the Table 5.1. The performance of GC is relatively poorer than LEC. Hence, it is sufficient to assign a smaller weight for GC. Sample weight vector $W_G$ obtained by combining the four local feature vectors $W_1, W_2, W_3$ and $W_4$ is shown in Equation (5.6).

To check the influence of $W_G$ on the performance of the HEC, experiments are conducted. The performance is checked for different values of $W_G$. Table 5.2 shows the recognition rate for different values of the weight of $W_G$. 

![Figure 5.11 Recognition rate versus FAR](image)
The performance changes with the varying $W_G$ and is shown in Figure 5.12. It could be found that the best performance occurs when $W_G = 0.2$ for ORL database.

**Table 5.2 Effect of $W_G$ on the recognition rate**

<table>
<thead>
<tr>
<th>Weight of Global Classifiers ($W_G$)</th>
<th>Recognition Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.82</td>
</tr>
<tr>
<td>0.2</td>
<td>0.96</td>
</tr>
<tr>
<td>0.4</td>
<td>0.82</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>1</td>
<td>0.56</td>
</tr>
</tbody>
</table>
Figure 5.12 The effect of different $W_G$ on the performance of the ensemble classifier on ORL database

It could be found that the local features are more important than the global features and the combination of both yield good results.

5.3.3 Comparison of Recognition Rate on Different Databases

The algorithm has been tested with the two databases, ORL and FRGC. Table 5.3 gives the comparative performance of the recognition rate on different databases. Figure 5.13 shows the comparison of the recognition rate, using the combination of both the local and global classifiers.

Table 5.3 Comparative performance on different databases

<table>
<thead>
<tr>
<th>Databases</th>
<th>Recognition Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Global features only</td>
</tr>
<tr>
<td>ORL database</td>
<td>0.7</td>
</tr>
<tr>
<td>FRGC database</td>
<td>0.6</td>
</tr>
</tbody>
</table>
Figure 5.13 Comparison of the recognition rate on two different datasets

The recognition rate is lower for the FRGC database, since the faces have been captured under varying illumination conditions. The faces have been taken in groups and tested at five different instances.

5.4 SUMMARY

Global and local facial features are important for face recognition. 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 are better than global features for face recognition. In the proposed method, both the global and local features are extracted by Discrete Fourier transform (DFT) and Gabor Wavelet Transform (GWT) respectively. The classification errors are reduced using Fisher Linear Discriminant (FLD). Ensemble Classifier designed using the combination GFFV and LGFV. The Fusion of the vectors is done by Region Based Image Fusion. Correlation Coefficient is used for
matching the test image with the database. ORL and FRGC are used for testing. Comparisons are made using only local features, only global features and then the combination of both. An average recognition rate of 64% using global features, 80% using local features and 92% using both global and local features is observed from the experiments. Hence, an increase of 28% when compared to global features and an increase of 12% when compared to local features has been observed when both the local and global features are used.