Face Images Feature Extraction Analysis
for Recognition in Frequency Domain

M. ALMAS ANJUM and M. YOUNUS JAVED
Department of Computer Engineering
College of Electrical & Mechanical Engineering (NUST)
Peshawar Road, Rawalpindi 46000,
PAKISTAN

Abstract: In this paper a novel technique to extract facial features for recognition in frequency domain using Discrete Fourier Transform (DFT) is presented. In pre processing phase facial tilt and varying image background challenges have been addressed to improve the success rate. Varying facial expressions within class have been minimised by using decimation algorithm. Experiments on ORL and YALE datasets have been performed with success rate up to 99%.

1. Introduction
In recent years face recognition has received substantial attention from researchers in biometrics, pattern recognition, and computer vision communities [1,2,3]. There are three major research groups which propose three different approaches to the face recognition problem. The largest group [4,5,6] have dealt with facial characteristics which are used by human beings in recognizing individual faces. The second group [7,8,9,10,11] performs human face identification based on feature vectors extracted from profile silhouettes. The third group [12,13] uses feature vectors extracted from a frontal view of the face. Although there are three different approaches to the face recognition problem, there are two basic methods from which these three different approaches arise. The first method is based on the information theory concepts, in other words, on the principal component analysis methods. In this approach, the most relevant information that best describes a face is derived from the entire face image. Based on the Karhunen-Loeve expansion in pattern recognition, M. Kirby and L. Sirovich [4,5] have shown that any particular face could be economically represented in terms of a best coordinate system that they termed "eigenfaces". These are the eigenfunctions of the averaged covariance of the ensemble of faces. Later, M. Turk and A. Pentland have proposed a face recognition method [14] based on the eigenfaces approach. The second method is based on extracting feature vectors from the basic parts of a face such as eyes, nose, mouth, and chin. In this method, with the help of deformable templates and extensive mathematics, key information from the basic parts of a face is gathered and then converted into a feature vector. Frequency domain has been also explored for feature extraction required for recognition. In [15] it is found that information in lower frequency bands have a dominant role in face recognition as low-frequency components contribute to the global description, while the high-frequency components contribute to the finer details. In proposed recognition model DFT has been used to extract low frequency components through novel feature extraction method.
In pre-processing phase varying image background which contributes a lot in failure rate has been removed through image segmentation. At the same time facial tilt is addressed through geometric normalization. Experiments on two different datasets have been carried out which provide improved results.

## 2. Image Pre-Processing

### 2.1 RGB to Gray Scale Conversion

Color images being in three planes of Hue, Saturation and Value are computationally very extensive. To avoid color images handling they are converted to gray scale images by using expression:-

\[ Y = 0.3R + 0.59G + 0.11B \]

The weights are used to compute gray image because for equal amount of color eye is most sensitive to green, red and then blue [16, 17].

### 2.2 Scale Normalization

The scale normalization is performed to eliminate the unnecessary information except face from the image. Figure 1 shows the original image and the cropped image [19].

### 2.3 Image Background Removal

Gray scale image are converted from 255 intensity levels to 7 intensity levels and than Median filter of size \( n=5 \) is applied on the image. Median filter forces points with distinct gray levels to be more like their neighbors and isolated clusters of pixels that are light or dark with respect to their neighbors and whose area is less than \( n^2/2 \) are forced to median intensity. This segmentation isolates the image background with specific intensity values from rest of the image and change intensity value to white as shown in Figure 2. This process makes uniform image background of images in dataset which are reported failure with background and minimizes the varying image background effects on recognition.

![Fig 2: (From left to right)Original, segmentation and final result](image)

### 2.4 Facial Tilt Compensation

Eyes in face image are pivot point for tilt compensation. Pixel values near eyes change more rapidly as compared to rest of face image. This property of image is used to detect the general eye location in the face image. Iris localization in the rough region of eye is carried out through template matching. Let two points \((xl,yl)\) and \((xr,yr)\) be the center of right and the left eye respectively. These are then used to compute the tilt (slope, \(m\), and angle, \(\theta\)) in the image using:

\[
m = \frac{(yr - yl)}{(xr - xl)} \quad (2)
\]

\[
\theta = \arctan(m)
\]

Finally, the tilt compensation is applied [18] using the reverse rotation, i.e., rotating by \(-\theta\) as shown in Figure 2.
3. Image Decimation

Decimation algorithm [19] scans through lines of pixels or group of pixels according to decimation down scale factor (M). As a result Gaussian Pyramid of varying image resolution is obtained. Decimation process is shown in Figure 3.

\[ F(m,n) = C(n_1M, n_2M) \]

where M is decimation down scale factor and 

\[ 0 \leq m \leq (n_1/M), \ 0 \leq n \leq (n_2/M) \]

The resulting image is a reduced size mirror of the original image faithful in tonality to the original but smaller in size. Experiments on images with varying resolution are carried out and it is established that each dataset provides best recognition results at a specific resolution level. Moreover reduction in image resolution up to a certain level reduces the affect of changing facial expressions within a class.

4. Recognition in Frequency Domain

DFT is used to decompose an image into its sine and cosine components. The output of the transformation represents the image in the Fourier or frequency domain, while the input image is the spatial domain equivalent. In the Fourier domain image, each point represents a particular frequency contained in the spatial domain image.

The Fourier transform plays a critical role in a broad range of image processing applications, including enhancement, analysis, restoration and recognition. The Fourier Transform produces a complex number valued output image which can be displayed with two images, either with the real and imaginary part or with magnitude and phase. In image processing, often only the magnitude of the Fourier Transform is displayed, as it contains most of the information of the geometric structure of the spatial domain image. However, if we want to re-transform the Fourier image into the correct spatial domain after some processing in the frequency domain, we must make sure to preserve both magnitude and phase of the Fourier image.

For a square image \( f(x, y) \) of size \( M \times N \), its DFT is given by:

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

where \( f(x, y) \) is the image in the spatial domain and the exponential term is the basis function corresponding to each point \( F(u, v) \) in the Fourier space.

A Fourier spectrum of face image as shown in Figure 4 reflects that as we move away from center of image low frequency components start decreasing. In face recognition low frequencies have the most important information. The higher
frequencies which contribute towards finer details are less significant for recognition.

Fig 4: Fourier spectrum of Image

5. Low Frequency Coefficient Selection Methods

First DFT of face images is taken then logarithmic transform is applied for contrast enhancement of dark images. As it is very effective tool especially in case of DFT Spectrums as shown in Figure 5.

Fig 5. Original image:- range 0 to 21(Left) Log enhanced (Right)

For recognition purposes only real part of low frequency components which contain maximum information as compared to imaginary part are used directly. Two distinct methods have been used to extract low frequency coefficients which contribute maximum to facial features required for face recognition; first is square selection method and second is circular selection method.

5.1 Square Selection Method

In this method square of different dimensions as shown in figure 6 around the center of the spectra is selected and low frequency coefficients falling within this square are taken as feature vectors. In training of the model five images of each subject are used and feature vector comprises of low frequency coefficients fall within square of specific dimension is retained against each subject of dataset.

5.2 Circular Selection Method

In this method circles with different radii concentric with center of spectra are evaluated as shown in figure 6 and feature matrix of low frequency coefficients within circle is obtained.

Figure 6. Square and Circular Selection Methods

6. Experiments and Results

Due to the statistical nature of the problem most of face recognition techniques are dataset dependent. A technique might perform better under a given set of conditions and may perform poorly for another set of conditions. Experiments on ORL and Yale datasets are carried out.

6.1 ORL and YALE Dataset

The ORL dataset consists of 40 subjects each having 10 images. There are 400 images in total. All of the images are gray scale. They are front views and have a black background. These images were taken over a period of 2 and with variation in subject gestures and head orientation. The subject images have a tilt and rotational tolerance up to 20 degree. Images of two different subjects are shown in Figure 7.

Fig 7: Example of Images in ORL Database
Second dataset used is YALE which contains 165 gray scale images in GIF format of 15 individuals. There are 11 images per person one per different facial expression or configurations: centre-light, with or without glasses, sad, happy, sleepy, surprise and wink. Few Examples are shown in figure 8.

Fig 8: Example of Images in YALE Database

Before applying DFT image decimation was carried out and a Gaussian pyramid with varying image resolutions was obtained. Later on using circular and square methods the features around the center of image were extracted. Five images of each subject were used for training purposes and rest five randomly used for testing purposes. Results of these two dataset with varying image resolution are shown in Figure 9 and 10.

Fig 9: Results of ORL Dataset with varying Image resolution using both Feature Selection Methods

Fig 10: Results of YALE Dataset with varying Image resolution using both Feature Selection Methods

7. Discussion

In the proposed model of face recognition only real part of frequency coefficients which contribute maximum towards recognition of image have been used whereas imaginary part being little significant have not been considered. This dimension reduction has made the system computationally very efficient as ORL dataset took 23 seconds for training of complete dataset and only 0.14 seconds for recognition of each image. Training time for complete YALE dataset was 8.5 seconds and 0.09 seconds for recognition.

Experiments by using two distinct methods of square and circle for facial feature selection have been carried out which select the low frequency coefficients with different dimensions. This highlights the effect of number of coefficients on recognition. Image decimation has established the fact that images have more high frequency components are adversely affected by changing the image resolution. However small facial expression changes are also compensated through image decimation and each dataset at specific resolution level provides best recognition rate as shown in Figure 9 and 10. Five images of each class have been used for training and rest as test images. Results with maximum success rate of 99% at image decimation factor four have been achieved.
Conclusion

Real time security and surveillance due to certain limitations and restrictions (like tilt in faces and varying background, speed of system and its accuracy) have made this area of research more attractive and challenging for biometric researchers. In preprocessing of proposed model of face recognition, color images were converted to gray scale images. Scale normalization and compensation to facial tilt and varying background has been applied. Two feature selection methods, along with usage of only real part of frequency coefficients have not only provided considerable dimension reduction with enhanced computational speed but also improved results.

References: