CHAPTER 4
FACE IMAGE DIMENSION REDUCTION THROUGH
IMAGE RESOLUTION VARIATION

4.1 Introduction

Digital images are made up of pixels. Pixels are the small sections of color and/or tone that together form a digital image. Pixels form an image-like pieces of a mosaic. A digital image is a grid of pixels and if the pixels are viewed together in proper registration, the image is formed. When there are enough pixels and they are small enough so as not to be individually discernible, the digital image can achieve photo quality. A digital image as it is created by a digital camera or scanner and as it resides on digital storage medium like a hard drive or a flash card is simply an informational record of a grid of pixels - a map of specific tones at specific locations (bitmap). It is common to refer to computer records as files, and thus digital images in this state, and generally, are commonly referred to as "digital image files," "image files," or even just "files." The image has no physical size, but if stored in certain file formats the image may have a specific size designated as part of the informational record, this can easily be set or later adjusted to any size. When an image is displayed or output it takes on physical form as an image. Whether being displayed on a monitor or printed on paper, the physical image has spatial dimensions - width and height. The same digital image can assume many different physical sizes.

Resolution refers to the number of pixels in an image. Resolution is sometimes identified by the width and height of the image as well as the total number of pixels in the image. Image size, resolution and compression are interrelated terminologies. Sometimes an image cannot be reproduced at the size needed. It lacks the needed resolution or has too much for practical use. In these cases, the computer can be called upon to modify the resolution of an image, reducing its resolution or increasing it. Different techniques to vary the image resolution are employed which have effects not only on quality of image but also on image recognition. This is demonstrated by evaluating effects of varying image resolution using different techniques of face recognition.
4.2 Types of Image Resolution

Digital image resolution designations fall into two basic categories: Pixel Count Resolution and Spatial Resolution.

4.2.1 Pixel Count Resolution

Pixel count resolution is simply the amount of pixels a digital image contains, or made up of. It is expressed in either mega pixels or pixel dimensions. The term mega pixels simply mean millions of pixels. A six mega pixel image is an image made up of 6 million pixels, or some rough approximation thereof. Pixel dimensions represent another more descriptive method of designating pixel count resolution. By stating the pixel dimensions of a digital image, the pixel count is inferred. In addition, the aspect ratio of the image is revealed as are the exact dimensions. Pixel count resolution is a fixed property of an image. Unless the image is resampled or cropped, the image remains with the same number of pixels.

4.2.2 Spatial Resolution

It relates to the number of pixels in a spatial measurement of a physical image - pixels per inch. Spatial resolution does not apply to an image file (except as a temporary/variable specification thereof), only to a physical image. An image literally can not have a spatial resolution if it doesn't take a physical form - it can't have any given number of pixels per inch if it doesn't have physical dimensions. The spatial resolution of an image is commonly referred to in terms of dots per inch (dpi). What is being specified is pixels per inch, however "dots" per inch has gained a foothold in common terminology. Spatial Resolution is a variable property of an image file - it only becomes a fixed property of an image once it is output in permanent form (i.e. printed). As this resolution is conditional upon output, this resolution is commonly called output resolution or print resolution.
4.3 Image Resolution and Face Recognition

In case of face recognition, behavior of facial images changes with variations in image resolution. Resolution variation is directly connected with the information possesses by face images. This information is of two types (i.e. discriminative information and structure information). Discriminative information represents the information distinguishing from other face images whereas structure information represents the common information of all the face images under the same resolution, which can be roughly estimated from low resolution face image. Similarly difference of the discriminate information between higher resolution face image and the lower face image is called enhanced discriminative information and difference between structure information is called enhanced structure information. When observing an object, people will usually see it more clearly when moving nearer and see it most clearly when distance is smaller than one fixed value $L_o$. In case of face image, we recognize it when distance is another fixed value $L_1$ which is usually larger than $L_o$. This conjecture can be supported through Shanon sampling theory which describes that face image with one certain resolution can retain all the original image information [89] (i.e., the corresponding reconstruction image will not lose any information). Similarly, the central challenge in face recognition lies in understanding the role of different facial features play in varying image resolution. These features can be the internal (eyes, nose and mouth) and external (hair and jaw-line) features. Most important is the relative contributions of internal and external features change as a function of image resolution. One very interesting conclusion on issue is that reports from several researchers studying face recognition by neonates [90] suggest that infants initially depend more on external features than on internal ones for discriminating between individuals, external features are more useful in face recognition when image resolution varies through decimation. Visual system uses a highly non-linear cue-fusion strategy in combining internal and external features along the dimension of image resolution and that the configural cues that relate the two feature sets play an important role in judgments of facial identity.
4.4 Image Resolution Variation Through Interpolation

Image resolution variation is always required in image processing operations as sometimes image under consideration lacks the needed resolution or has too much for practical use like recognition purposes. In these cases, various techniques are employed to alter the image resolution. Interpolation and image decimation are two distinct methods to achieve the image resolution variation.

Interpolation is estimation of an image pixel value at a location in between image pixels. The process is called into force when there is not enough resolution in an image to accomplish the reproduction needed. Interpolation can be a daunting task for a computer, in which it carries out the analysis and averaging the new pixel values and colors that are inserted in an image. The process of interpolation requires the computer to measure adjacent pixels and "guess" a correct value for new pixels inserted in the data stream. In the process, it can make errors, create unwanted noise, and introduce abnormal pixels into areas of the image causing "artifacts." The process is good enough, though, to produce an acceptable image in most cases. The important thing to remember is that:

\[
\begin{cases}
\alpha > 2 & \text{Resultant image quality is unacceptable} \\
\alpha = 2 & \text{Resultant image quality is acceptable}
\end{cases}
\]  

(4.1)

where \( \alpha \) is the interpolating factor that assume positive discrete values only. \( \alpha = 1 \) implies no interpolation. \( \alpha = 2 \) reflects that size of interpolated image is double the size of original image.

Any interpolation is potentially harmful to an image and some interpolation may ruin it. If an image is interpolated to a small degree (up to a doubling of the image's original size), this is usually harmless. Beyond two times the original size, you will begin to see serious effects in most images. The types of interpolation are:

- **Nearest-Neighbor Interpolation.** Output pixel values are assigned the value of the pixel that the point falls within. No other pixels are considered.
- **Bilinear Interpolation.** Output pixel values are calculated from a weighted average of pixels in the nearest 2-by-2 neighborhood.
- **Bicubic Interpolation** Output pixel values are calculated from a weighted average of pixels in the nearest 4-by-4 neighborhood.

### 4.5 Image Gaussian Pyramid through Decimation

In most pattern recognition systems, image resolution reduction can be achieved through image decimation. Decimation algorithm is a novel form to obtain desired resolution where the sub-sampling under the best circumstances is harmless to image as it scans through lines of pixels, averaging together pair of pixels or group of pixels according to the value of Decimation factor (G). The resulting image is a reduced size mirror of the original image faithful in tonality to the original image but smaller in size. G can be calculated as:

\[ G = \frac{P}{Q} \]  

(4.2)

where

- \( P \) = Order of original image matrix
- \( Q \) = Order of desired decimated image matrix
- \( G \) = Arbitrary down scale decimation factor

The sliding mask shown in Figure 4.1 is applied on face images for the decimation. The simultaneous process of low pass filtering and decimation is performed where a low pass averaging filter is applied on face images and every next pixel according to order of 2D filter is decimated. The mathematical model of decimation process is as under where \( O(i,j) \) is original image and \( M,N \) is order of resulting image:
\[
I(M, N) = \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left[ \sum_{m=i}^{i+1} \sum_{n=j}^{j+1} O_{m,n} \right]}{G \times G}
\]

(4.3)

\[
\mu_w = \frac{W}{D_w} \quad \mu_h = \frac{H}{D_h}
\]

Where

W & H = Width and Height of the original image

\(D_w\) & \(D_h\) = Width and Height of desired output image

\(\mu_w\) = Ratio between actual width of image and desired width

\(\mu_h\) = Ratio between actual height of image and desired height

1/9

Figure 4.1. Decimation mask of 3 x 3

The sliding mask scans through the whole image and averages together the pixels under it. In scanning process, factor \(G\) determines the jump of the mask. The order of the mask determines the \(G\) and accordingly the resolution is varied. By changing the value of \(G\), a Gaussian Pyramid of varying image resolution as shown in Figure 4.2 is obtained. This decimation algorithm is applied on all the images of database by varying the value of \(G\) and recognition results are obtained against these images with varying resolution.

4.6 Effects of Varying Resolution on Template Matching Face Recognition Model

Nearest-Neighbor template matching is the most native method of face recognition [91,92], where individuals are represented by a set of complete images of
their face. An input face is compared with each image in the database. The image for which the template matching gives the highest response is considered to be the match. Proposed template matching model of face recognition has been modified where images were centered first and then through decimation process image resolution is varied. Proposed model of face recognition is shown in Figure 4.3.

![Gaussian pyramid](image)

Figure 4.2. Gaussian pyramid

4.7 Resolution and Dimension Reduction through Decimation Algorithm

Decimation algorithm is applied on each face image of database. By varying the value of $G$, the resolution of each image is changed and a Gaussian pyramid of each image is obtained. In learning process of proposed face recognition model, five images of each class at one particular resolution are taken and their mean $\bar{M}$ is calculated by:

$$\bar{M} = \frac{1}{T} \sum_{i=1}^{T} I_i$$  \hspace{1cm} (4.4)

Where:

$I_i$ is the $i$th image with specific Gaussian base value used for training of the system

$T$ varies from one to total number of images used for training
Images are centered by subtracting mean from each image. It produces a dataset whose mean is zero.

\[
C(T) = \left[ I(T) - \bar{M}_{\alpha \mu} \right]
\] (4.5)

The subspace of centered images is stored in the system as the reference against each individual image of training set.

![Figure 4.3. Model of the System](image)

In testing phase dimension of preprocessed test image is reduced as done in training of model and then it is centered by subtracting mean from it. For image classification, Euclidean distance is used as matching criterion. This process is repeated for next level of resolution of same database until Face Recognition Rate (FRR) on all the images of Gaussian pyramid is obtained.

4.8 Dissimilarity Space and Matching

In training of the model, five images of each subject are used and decimated image matrix of each image used in training process is obtained.
In matching process, a dissimilarity space $D(I_i, P)$ of test image with training images is obtained by using simple Euclidean distance. This dissimilarity space matrix:

$$E = D(I_i, P)$$

is converted to a vector and the vector with min difference is taken as matched image:

$$R = \text{arg min}[E]$$

where $I$ is training image and $i = 1, 2, \ldots, \text{total subjects}$ used in training of model, $R$ is the recognized image and $E$ is the difference matrix of test image with all training images. Here the $\text{arg min}$ which is accumulated minimum Euclidean distance provides the best match.

4.9 Implementation of Face Recognition Model

This model of face recognition has been implemented by developing following modules and data structures.

4.9.1 Scale Normalization Module

In this module, training images of database are loaded in an array. Dimension of the images is checked to distinguish between color and grayscale images. If image dimension is three (i.e. red, green and blue), images are converted to grayscale. Gaussian filter is applied on grayscale images to smooth the images. In next step, the edge detection through 2D image derivative is obtained. Upper and lower threshold values (i.e. 0.85 and 0.2) are used to obtain the thin outer curvature of face image. Resultant image is scanned in a loop from left to right and from top to bottom to find out expansion of binary image in up, down, left and right directions. Four outer most values are calculated in result of this scanning process. Finally the original input image is rescaled according to calculated values. The flowchart of scale normalization module is shown in Figure 4.4.
4.9.2 Geometric Normalization Module

Geometric deformation (mainly facial tilt) is calculated and removed in this module. Scale normalized grayscale image is loaded as input to this module. Following algorithm has been developed and implemented to remove the facial tilt.

Begin

1. Repeat till end of training images
2. **Carryout eye localization**
   - Load scale normalized grayscale image
   - Calculate gradient of image
   - Find out horizontal projection
   - Calculate horizontal line of eyes
   - Find vertical projection
   - Calculate two peaks
   - Determine size of eyes

3. **Carry out eye template selection**
   - Load eye templates of different sizes
   - Make selection of template according to face image eye size

4. **Obtain iris localization**
   - Carry out scanning of template image over face image
   - Obtain cross correlation
   - Locate iris of both eyes

5. **Carry out tilt compensation**
   - Compute slope and angle between two eyes
   - Apply reverse rotation

End
To remove the facial tilt, first rough region of both eyes is located through feature based method. Precise iris location of both eyes is determined by using template matching in areas of both eyes. Slope and angle between two eyes give amount of facial tilt present in the image which is finally compensated by applying reverse rotation.

4.9.3 Background Module

After scale and geometric normalization, background module has been employed to make the image background uniform by using grayscale morphological operations, where each pixel in output image is based on a comparison of the corresponding pixel in the input image with its neighbours. A disk type kernel (which conserves the foreground area that has the shape like it) is used to erode and dilate the image. It is termed as structuring element. First, image is eroded by scanning the structuring element over the image. The values of structuring element are subtracted from the image values and $\text{arg Min}$ is used to return the final value. In next step, eroded image is applied as input to the dilation process which is just a convolution like operation. Structuring element is scanned over the image and pixel by pixel sum with the structuring element is computed and $\text{arg Max}$ is used to return the value. This dilated image is subtracted from input image which gives uniform background as compared to original image.

4.9.4 Resolution Module

This module is devised to vary the image resolution according to the decimation down scale factor. Preprocessed images used for training of the system are provided to convolution sub module which scans a low pass averaging filter (shown in Figure 4.1) over the entire image. Convolution process is carried out according to Equation 4.9 which returns the averaged value of all pixels under the filter window at that particular instance.

\[
I(m,n) = O(i,j) * h(x,y) \tag{4.9}
\]

\[
I(m,n) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} O(i,j) h(x-i,y-j) \tag{4.10}
\]
Convolved image is applied to image decimator which throws away the every next pixel according to decimation down scale factor. Resultant image is a condensed size image with reduced resolution. By varying the decimation rate, images of varying resolution are obtained. An image with original resolution and one with reduced resolution through image decimation is shown in Figure 4.5.

Figure 4.5. Original and Decimated Image (left to right)

4.9.4 Training Module

Training part of the proposed face recognition model is implemented in this module where five preprocessed and decimated images of each class are loaded as input. First mean of these images is calculated and every image is centered by subtracting the mean from it. Decimated images provide dimension reduction and at the same time reduce the variation within class images. These images with reduced dimension are converted into a column vector. A subspace of dimension of $L \times T$ is attained whose each column represents an image. Where $L$ is total number of pixels of decimated image and $T$ is total images used for training of the model.

4.9.5 Recognition Module

Recognition module is last module of the proposed face recognition model where a test image is presented as input. Test image is first preprocessed and afterwards it is decimated with same value of down scale factor as used in training module. Then image is converted into a column vector with a dimension of $L \times 1$. In matching process, the Euclidean distance of this vector with each column of subspace gathered during training of model is acquired and finally a difference matrix of dimension $L \times T$ is obtained.
Euclidean distance of each image is accumulated, a matrix of L x T is converted into 1 x T dimension and arg Min is used to find out the best match.

4.10 Experiments and Results

Four different databases (ORL, YALE, FERET and CEME_NUST color database) as discussed in chapter 3 have been used to obtain the Face Recognition Rate (FRR) against varying image resolution. Figure 4.6 illustrates the results of ORL and YALE databases. Images in both the databases were acquired at resolution of 112 x 92 and tested on various resolution levels of 112 x 92, 56 x 46, 28 x 23, 18 x 15 and 14 x 11. Images of CEME_NUST color database and FERET have been acquired at resolution of 112x100 and 256x384 respectively. CEME_NUST color database has been tested on varying image resolutions (112x100, 56x50, 28x25, 18x16 and 14x12) and FERET images have been evaluated at different resolution levels (256x386, 128x193, 64x96, 42x64 and 32x48). Plot lines of Figure 4.6 and Figure 4.7 show that as the image resolution of face images is reduced FRR is improved and reaches at optimum value at a specific image resolution. Above or below this specific resolution FRR is influenced negatively due to inclusion of redundant data or lack of required facial information for recognition in image used for recognition.

Figure 4.6 FRR of ORL and Yale databases with varying image resolution
Figure 4.7 FRR of FERET database (upper) and CEME_NUST Color database (lower) with varying image resolution

The image resolution against optimum FRR varies from database to database as ORL database images provides FRR of 97.2% at resolution of 56 x 46, Yale database images provide FRR of 92% at resolution of 56x46, CEME_NUST color database provides FRR of 92.5% at resolution of 28 x 25 and FERET database showed optimum performance of 93.7% at resolution of 128x193.

Reduced image resolution not only provides improved FRR but also addresses the curse of dimensionality with improved speed. Resolution variation through image decimation presents a considerable image dimension reduction which ultimately reduces the database training time and image recognition time. Table 4.1 reflects the database dimensions against which best FRR is achieved and correspondingly the reduction in training and recognition time are encircled. As the image resolution is reduced accordingly image dimensions are reduced which improves the computation speed of the recognition model. Best FRR is obtained in ORL and YALE database at dimension reduction of 56x46 from 112x92, CEME_NUST Color database provides optimum FRR at image dimension of
28x25 which is four times image resolution reduction whereas FERET database offered best performance at 128x193 which is twice image resolution reduction.

Table.4.1 Training and Recognition time against Varying Image Resolution

<table>
<thead>
<tr>
<th>Image Resolution</th>
<th>Training Time (Seconds)</th>
<th>Recognition Time per Image (Seconds)</th>
<th>ORL Database</th>
<th>Yale Database</th>
<th>FRR (%)</th>
<th>FRR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ORL Database</td>
<td>Yale Database</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>112x92</td>
<td>28.78</td>
<td>9.25</td>
<td>0.217</td>
<td>0.1048</td>
<td>94</td>
<td>90.0</td>
</tr>
<tr>
<td>56x46</td>
<td>11.47</td>
<td>3.03</td>
<td>0.112</td>
<td>0.053</td>
<td>97.25</td>
<td>92.12</td>
</tr>
<tr>
<td>28x23</td>
<td>5.337</td>
<td>1.79</td>
<td>0.072</td>
<td>0.033</td>
<td>96.0</td>
<td>92.00</td>
</tr>
<tr>
<td>18x15</td>
<td>3.585</td>
<td>1.4</td>
<td>0.064</td>
<td>0.029</td>
<td>95.75</td>
<td>91.50</td>
</tr>
<tr>
<td>14x11</td>
<td>3.175</td>
<td>1.32</td>
<td>0.060</td>
<td>0.026</td>
<td>95</td>
<td>89.00</td>
</tr>
</tbody>
</table>

4.11 Discussion of Results

Template matching is a very native technique of face recognition which has been implemented to evaluate the changes in recognition rate caused by image resolution variation. As reflected from Figures 4.6 and 4.7, for each database recognition rate changes with change in image resolution up to a certain level. This conjecture supports the phenomena that in each dataset up to certain image resolution, the image information is not significantly lost. Moreover, in decimation process by applying the averaging filter high frequency components present in the image are reduced while retaining the low frequency components intact which may reconstruct the image without any significant loss to the image quality. This reduction of undesired high frequency components for recognition not only minimize the redundant clutter of facial information but also improve the computation time of model and provides better matching as it is analytically proved in Tables 4.2 & 4.3. In Table 4.2 first test image is S12-6 of CEME_NUST database which means sixth image of S12 class. Once its resolution is set to 112x100 or 56x50, its best matching is with second image of S11 class with Euclidean distance of 3,1284 x 10^6, hence giving wrong recognition results. Once its resolution is further reduced to 28x25 Table 4.2 reflects that its best matching is with S12-1 having Euclidean
distance of $6.364 \times 10^5$ which is first image of its own class, so resulting in correct recognition. Similarly S6-9 and S8-7 images of same database have been also used as test images and results are illustrated in Table 4.2.

<table>
<thead>
<tr>
<th>Test Image</th>
<th>G</th>
<th>Best Euclidean Distance within class</th>
<th>Best Match Image within class</th>
<th>Best Euclidean Distance out of class</th>
<th>Best Match Image out of class</th>
</tr>
</thead>
<tbody>
<tr>
<td>S12-6</td>
<td>2</td>
<td>$3.174 \times 10^6$</td>
<td>S12-1</td>
<td>$3.1284 \times 10^6$</td>
<td>S11-2</td>
</tr>
<tr>
<td>S12-6</td>
<td>4</td>
<td>$6.364x 10^5$</td>
<td>Correct Match with own Class</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S6-9</td>
<td>4</td>
<td>$1.696x 10^6$</td>
<td>S6-3</td>
<td>Correct Match with own Class</td>
<td>-</td>
</tr>
<tr>
<td>S6-9</td>
<td>8</td>
<td>$3.762 \times 10^5$</td>
<td>S6-3</td>
<td>$3.362x 10^5$</td>
<td>S5-4</td>
</tr>
<tr>
<td>S8-7</td>
<td>4</td>
<td>$1.876 \times 10^6$</td>
<td>S8-1</td>
<td>Correct Match with own Class</td>
<td>-</td>
</tr>
<tr>
<td>S8-7</td>
<td>8</td>
<td>$4.117 \times 10^5$</td>
<td>S8-1</td>
<td>$3.4766 \times 10^5$</td>
<td>S9-3</td>
</tr>
</tbody>
</table>

The conjecture, “face recognition behavior of face image changes with change in image resolution” has also been analyzed by using images of ORL database at different resolutions. Table 4.3 and Figure 4.8 reflect that when S1-10 test image with resolution of 112x92 is presented to model it provides minimum matching difference with image of wrong class i.e. S17-2. Once same image is offered at reduced resolution of 56x46, it provide correct recognition with minimum matching difference with S1-2 of its own class. If resolution is further reduced Figure 4.8 reflects that again false recognition with S17-5 is obtained.

These results conclude that at certain resolution facial feature of face images become so promising to the facial feature of training images that it gives better recognition results as compared to results obtained at normal resolution. As each database images are acquired at different image resolution so every database provides improved FRR at different resolution. Table 4.4 reveals the best image resolution of each database against optimum FRR.
Table 4.3. Test image results of ORL database with different values of “G”

<table>
<thead>
<tr>
<th>Test Image</th>
<th>G</th>
<th>Best Euclidean Distance within class</th>
<th>Best Match Image within class</th>
<th>Best Euclidean Distance out of class</th>
<th>Best Match Image out of class</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1-10</td>
<td>2</td>
<td>$1.215 \times 10^6$</td>
<td>S1-5</td>
<td>Correct Match with own Class</td>
<td>-</td>
</tr>
<tr>
<td>S1-10</td>
<td>4</td>
<td>$2.971 \times 10^6$</td>
<td>S1-4</td>
<td>$2.954 \times 10^6$</td>
<td>S17-5</td>
</tr>
<tr>
<td>S3-8</td>
<td>2</td>
<td>$8.154 \times 10^6$</td>
<td>S3-1</td>
<td>Correct Match with own Class</td>
<td>-</td>
</tr>
<tr>
<td>S3-8</td>
<td>6</td>
<td>$7.453 \times 10^5$</td>
<td>S3-1</td>
<td>$7.302 \times 10^5$</td>
<td>S38-4</td>
</tr>
<tr>
<td>S27-8</td>
<td>2</td>
<td>$8.343 \times 10^6$</td>
<td>S27-5</td>
<td>Correct Match with own Class</td>
<td>-</td>
</tr>
<tr>
<td>S27-8</td>
<td>8</td>
<td>$4.321 \times 10^5$</td>
<td>S27-5</td>
<td>$4.030 \times 10^5$</td>
<td>S4-1</td>
</tr>
</tbody>
</table>

Figure 4.8. Effects of Changing Resolution on Recognition

Disparity of image resolution against best FRR among these databases is due to two factors:- (1) different database images have different original resolution as ORL and Yale
database images have resolution of 112 x 92, where as EME color database images have resolution of 112 x 100 and FERET database images are obtained at resolution of 128 x 192 and (2) strength of high frequency components present in images differ from database to database. These two factors contribute towards best FFR at different resolutions in different databases.

Table 4.4 Image Resolution against optimum FRR

<table>
<thead>
<tr>
<th>Name of Database</th>
<th>Image Resolution against Best FRR</th>
<th>FRR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORL</td>
<td>56x46</td>
<td>97.2</td>
</tr>
<tr>
<td>YALE</td>
<td>56x46</td>
<td>92</td>
</tr>
<tr>
<td>FERET</td>
<td>128x193</td>
<td>93.7</td>
</tr>
<tr>
<td>CEME_NUST</td>
<td>28x25</td>
<td>92.5</td>
</tr>
</tbody>
</table>

These results also proved that change in image resolution by a certain factor in each dataset removes the unnecessary clutter of information (i.e high frequency components) from image which is not required for the face recognition and at the same time achieve considerable dimension reduction.

4.12 Principal Component Analysis

Principal Components Analysis (PCA) is a way of identifying patterns in data and expressing the data in such a way as to highlight their similarities and differences. Since in high dimension data it is hard to find patterns, where the luxury of graphical representation is not available PCA is a powerful tool for analyzing data. Once patterns have been extracted from the data, and one needs to compress the data (i.e. by reducing the number of dimensions) without much loss of information, PCA is a good choice for it. In terms of information theory the idea of using PCA is to extract the relevant information in a face image, encode it as efficiently as possible and compare test face encoding with a database of similarly encoded models. A simple approach to extract the information contained in an image of face is to somehow capture the variations in a collection of face images independent of judgment of features and use this information to
encode and compare individual faces [93]. In mathematical terms, the purpose of using PCA is to find the principal components of distribution of faces, or the eigenvectors of the covariance matrix of the set of face images, treating each image as a point (vector) in a very high dimensional space. The eigenvectors are ordered, each one accounting for a different amount of variation among the face images. These eigenvectors can be thought of as a set of features that together characterize the variation between face images. Each image location contributes more or less to each eigenvector. Each eigenvector appears as a ghostly face. Each individual face can be represented exactly in terms of a linear combination of the eigenfaces and can also be approximated using only the “best” eigenfaces – those that have the largest eigenvalues, and which therefore account for the most variance within the set of face images. If an image has dimensions of 112 x 92 and we have a subspace of 10304 dimensions then this image will be a point in the space as shown in Figure 4.9. Here $X_1, X_2, \ldots, X_M$ show the dimension of face space.

The main use of PCA is to reduce the dimensionality of a data set while retaining as much information as possible. It computes a compact and optimal description of the data set. The first principal component is the combination of variables that explains the greatest amount of variation. The second principal component defines the next largest amount of variation and is independent of the first principal component. A mathematical model of PCA used for face recognition is shown in Figure 4.10 whose input is a set of
training images \( I=(i_1,i_2,\ldots,i_n) \). The mean \( \bar{m} \) of this training set is calculated, each image is centered by subtracting mean from it. This produces a dataset whose mean is zero. In next step, two dimensional variance called covariance of this dataset is calculated. As covariance matrix is a square matrix, its eigen values and eigen vectors are calculated which provide the information about patterns in the data. These eigen-values are ordered from highest to lowest and similarly the corresponding eigenvectors which provides data components in order of significance. This arrangement of data allows to decide on ignoring the data of less significance.

In this way, we do lose some information, but if the eigenvalues are small, we don’t lose much and the final dataset will have lesser dimensions than the original. Finally, this reduced dimension data is transposed so that eigenvectors are in row, with most significant eigenvector at the top and multiplied by the transpose of centered image. This new data matrix is projection of face image in eigenface space.

### 4.13 Implementation of PCA Based Face Recognition Model

To assess the contribution of face image resolution on FRR, a PCA based recognition model is proposed where face images with varying resolution are applied as input to system. The algorithm developed for this purpose is as below:-

**Begin**

1. **Input all training images**

2. **Carry out image preprocessing**
   - Call scale normalization module
   - Call geometric normalization module

3. **Call Resolution module**
4. Calculate eigenvalues and eigenvectors
   - Calculate mean of training images
   - Carry out image centering
   - Find out Covariance matrix of centered images
   - Obtain eigenvalues and eigenvectors of covariance matrix

5. Arrange the eigenvalues and corresponding eigenvectors in ascending order

6. Carry out dimension reduction through selection of highest eigenvalues and eigenvectors

7. Made face image projection in PCA subspace

8. Carry out image Recognition
   - Load test image
   - Repeat the steps 2 to 7
   - Obtain Euclidean distance of test projection with training images projection
   - Find out closest match

9. Display image with closest match

End
Five out of ten images of ORL database are used to learn the model. Scale and geometric normalizations are applied to the training images. Image resolution is varied, dimension reduction through PCA is carried out and face images are projected in PCA subspace as illustrated in Figure 4.10. Once the test image is presented to model in recognition process, it is preprocessed, its resolution is varied according to resolution of training.
images and its feature vector obtained through PCA is projected in the PCA subspace. Euclidian distance criterion is used to get the best match. Training and recognition process is shown in Figure 4.11.

### 4.14 Experiments and Results

Experiments have been conducted by varying the image resolution through image decimation as well as on actual resolution on which image have been acquired. In all these tests, five out of ten images are used for training purposes and complete database images have been used as test images. The number of eigenvectors (against highest eigenvalues) were retained according to number of image classes used in the model. Results have also been gathered by varying the number of classes used in the model as shown in Table 4.5.

<table>
<thead>
<tr>
<th>Resolution Down Scale Factor (G)</th>
<th>Image Resolution</th>
<th>Face Recognition Rate (%)</th>
<th>Improvement against normal PCA at normal resolution (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>112x92</td>
<td>84</td>
<td>Nil</td>
</tr>
<tr>
<td>2</td>
<td>56x46</td>
<td>87</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>28x23</td>
<td>85</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>18x15</td>
<td>78</td>
<td>-6</td>
</tr>
<tr>
<td>8</td>
<td>14x11</td>
<td>73</td>
<td>-11</td>
</tr>
</tbody>
</table>

Results shown in Table 4.5 conclude that when image resolution is reduced, FRR is changed and attain its best value at a particular resolution. Beyond this value FRR becomes very poor. ORL database images are acquired at resolution of 112 x 92, Table 4.5 reflects that at this resolution FRR is 84%. As we decrease the resolution the FRR is improved and resolution of 56 x 46 (at value of G=2) gives best FRR of 87%. A further decrease in image resolution starts causing adverse effects on FRR. It proves that in ORL database a resolution of 56 x 46 retains the required image information for recognition whereas below or above this resolution either image loses the information required for recognition or
presents unnecessary clutter of information giving wrong matching. Table 4.5 also represents the comparison of PCA based FRR obtained through varying image resolution with FRR gathered at normal resolution.

![Diagram of Face Recognition and Training Process using PCA](image)

**Figure 4.11** Face Recognition and Training Process using PCA

Afterwards, experiments were also conducted by keeping this reduced image resolution constant and varying number of classes used for experiments. A comparison of these results with normal PCA is carried out and is shown in Table 4.6 and Figure 4.12. This comparison demonstrates that as the numbers of classes are increased i.e. number of face representation in PCA subspace is increased, the FRR is reduced due to increase in the image clusters in the PCA domain and close placement of discriminating faces in PCA space.
Table 4.6. Comparison of FRR of ORL Database between normal and reduced image resolution

<table>
<thead>
<tr>
<th>No of subjects used</th>
<th>FRR (%) with Normal Image resolution (112x92)</th>
<th>FRR (%) with Reduced Image resolution (56x46)</th>
<th>Improvement with column 2 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>97</td>
<td>97</td>
<td>Nil</td>
</tr>
<tr>
<td>15</td>
<td>97</td>
<td>98.5</td>
<td>1.5</td>
</tr>
<tr>
<td>20</td>
<td>92</td>
<td>94.5</td>
<td>2.5</td>
</tr>
<tr>
<td>25</td>
<td>91</td>
<td>94</td>
<td>3</td>
</tr>
<tr>
<td>30</td>
<td>89</td>
<td>93</td>
<td>4</td>
</tr>
<tr>
<td>35</td>
<td>86</td>
<td>89.75</td>
<td>3.75</td>
</tr>
<tr>
<td>40</td>
<td>84</td>
<td>87</td>
<td>3</td>
</tr>
</tbody>
</table>

The best FRR of 98.5% is obtained when only 15 subjects are used in training and recognition process. As the numbers of classes are increased from 15 the FRR is reduced because above 15 classes the projection of face images in PCA subspace becomes closer and closer and chances of mismatched are more as compared to less number of classes used.

Figure 4.12. FRR with PCA for normal and reduced image resolution
4.15 Sub-Holistic PCA (SHPCA) Recognition Technique and Varying Image Resolution

The proposed technique is a simple offshoot of PCA where face image is split up into four parts as shown in Figure 4.13. Each of the four images formed are complete images and are a sub-set of the main image and called sub-images or sub-holons. Due to these sub-holons and mathematical nature of PCA, this technique is given name of sub-holistic PCA.

![Image and the four sub-Images](image_url)

Figure 4.13. Image and the four sub-Images

This hybrid approach is based on two-step face recognition. The first recognition stage reduces the number of test subjects to five and the second recognition stage searches the best match among these five subjects. This concept is a common practice in daily life as well as in intelligent systems. Instead of finding a single match to the face, a number of matches are found and from among these candidate matches the final result is sought.

4.16 Implementation of SHPCA

In this hybrid approach of face recognition first step is to reduce the image resolution up to level where best FRR is achieved. Table 4.6 reflects that in case of normal PCA, the best FRR of ORL database is gathered at resolution of 56x46 so the resolution of all the training images is reduced to this level. This reduction in image resolution causes blur in the face images which is reduced by applying a high pass filter. Complete image is used to create one database and the four sub-images obtained from the single image are used to create the other four face spaces. The five face spaces can be differentiated as top left corner face space, top right corner face space, bottom left corner face space, bottom right corner face space and full image face space. As the name
suggests the images formed by taking the top left corner portion of the images are used in
the creation of the top left corner face space and so on as shown in Figures 4.14 and 4.15.
Since the sub-face images formed are from within the larger image, they can be thought of as sub sets of the original image.

Training in SHPCA consists of two phases rather than a single phase as was the case in
PCA. Each face space is created, with its respective data, similar to the creation of face
space in PCA. Figure 4.14 describes the implementation of this proposed model of face
recognition. The first step is to reduce the image resolution through image decimation as
discussed earlier. Resultant image is divided into four sub images. The mean vectors are
calculated by using expressions given in Equation 4.10 \( \bar{E}_{f_{s1}} \) is calculated for the top left
corner face space. \( \bar{E}_{f_{s2}} \) is calculated for the top right corner face space. \( \bar{E}_{f_{s3}} \) is calculated
for the bottom left corner face space. \( \bar{E}_{f_{s4}} \) is calculated for the bottom right corner face
space. \( \bar{E}_{f_{s5}} \) is calculated for the full image face space.

\[
\begin{align*}
\bar{E}_{f_{s1}} &= \frac{1}{n} \sum_{j=1}^{n} E_{f_{s1}} \\
\bar{E}_{f_{s2}} &= \frac{1}{n} \sum_{j=1}^{n} E_{f_{s2}} \\
\bar{E}_{f_{s3}} &= \frac{1}{n} \sum_{j=1}^{n} E_{f_{s3}} \\
\bar{E}_{f_{s4}} &= \frac{1}{n} \sum_{j=1}^{n} E_{f_{s4}} \\
\bar{E}_{f_{s5}} &= \frac{1}{n} \sum_{j=1}^{n} E_{f_{s5}} \\
\end{align*}
\]  

(4.11)

An example of mean calculation of few sub-images is shown in Figure 4.15. In next step,
each sub-image is translated by using Equation 4.12 so that mean of complete dataset is
zero.
The resultant covariance matrices, $C_{f_1}, C_{f_2}, C_{f_3}, C_{f_4} \& C_{f_5}$, are used to calculate the eigenvectors $\lambda_{f_1i}, \lambda_{f_2i}, \lambda_{f_3i}, \lambda_{f_4i}$ and eigenvalues $\Phi_{f_1i}, \Phi_{f_2i}, \Phi_{f_3i}, \Phi_{f_4i} \& \Phi_{f_5i}$. The eigenvectors obtained are reordered in a descending eigenvalue order. The eigenvectors corresponding to the largest eigenvalues are retained because only they account for the most variance within the set of face images. Once the images have been projected into the five face spaces the system is considered as trained and it can be used for testing.

The test phase of the scheme can be divided into four phases. During the first phase, the image to be tested is passed through the same pre-processing procedure as was adopted for the images of the training phase. This results in a reduced size and sharpened image with four sub-images. The complete face image and the four sub-images are projected in their respective face spaces through PCA by pursuing the following steps:

The covariance matrices of each subspace are calculated by:

$$
\begin{cases}
E_{f_{1i}} = E_{f_{i}} - \overline{E}_{f_{1}} \\
E_{f_{2i}} = E_{f_{i}} - \overline{E}_{f_{2}} \\
E_{f_{3i}} = E_{f_{i}} - \overline{E}_{f_{3}} \\
E_{f_{4i}} = E_{f_{i}} - \overline{E}_{f_{4}} \\
E_{f_{5i}} = E_{f_{i}} - \overline{E}_{f_{5}} \\
\end{cases}
$$

(4.12)

$$
\begin{cases}
C_{f_{1i}} = (E_{f_{1i}} - \overline{E}_{f_{1i}})(E_{f_{1i}} - \overline{E}_{f_{1i}})^T \\
C_{f_{2i}} = (E_{f_{2i}} - \overline{E}_{f_{2i}})(E_{f_{2i}} - \overline{E}_{f_{2i}})^T \\
C_{f_{3i}} = (E_{f_{3i}} - \overline{E}_{f_{3i}})(E_{f_{3i}} - \overline{E}_{f_{3i}})^T \\
C_{f_{4i}} = (E_{f_{4i}} - \overline{E}_{f_{4i}})(E_{f_{4i}} - \overline{E}_{f_{4i}})^T \\
C_{f_{5i}} = (E_{f_{5i}} - \overline{E}_{f_{5i}})(E_{f_{5i}} - \overline{E}_{f_{5i}})^T \\
\end{cases}
$$

(4.13)
Figure 4.14. Representation of Proposed Face Recognition Model
The mean face image is subtracted from the test image. This results in the translated test image. Mathematically:

\[
\begin{align*}
T_{f_1} &= T_{f_1} - \overline{E}_{f_1} \\
T_{f_2} &= T_{f_2} - \overline{E}_{f_2} \\
T_{f_3} &= T_{f_3} - \overline{E}_{f_3} \\
T_{f_4} &= T_{f_4} - \overline{E}_{f_4} \\
T_{f_5} &= T_{f_5} - \overline{E}_{f_5}
\end{align*}
\]  

(4.14)

4.16.2 The mean face image for the top left corner face space is subtracted from the top left corner of the test face image, the mean face of the top right corner face space is subtracted from the top right corner of the test face image and so as mentioned in Equation 4.14. Finally, the mean face of the full image face space is subtracted from the test face image. The resultant normalized or translated images are shown in Figure 4.16. The images on the left are the test images and the images on their right are the translated versions.

Figure 4.16. Test Image before and after translation
4.16.3 The translated images are then projected into their respective face space making use of the eigenvectors calculated during training phase. The minimum Euclidean distance is calculated between the projected test image and the training images already present in the face spaces. For each face space, the class/subject that returns the minimum difference is selected as the best match for that face space.

For the test image of Figure 4.17 the following matches were obtained:

- Top Left Corner Face Space  Class / Subject 1
- Top Left Corner Face Space  Class / Subject 1
- Top Left Corner Face Space  Class / Subject 2
- Top Left Corner Face Space  Class / Subject 8
- Top Left Corner Face Space  Class / Subject 2

Thus, the five face spaces suggest that the test image could belong to the following three classes or subjects.

![Test Image and the Results](image)

Test Image                     Subject 1     Subject 1     Subject 2    Subject 8     Subject 2

FaceSpace1 FaceSpace2 FaceSpace3 FaceSpace4 FaceSpace5

Figure 4.17    Test Image and the Results

Thus, the second phase results in five possible matches of the test image, one from each of the five face spaces. If four of the face spaces produce the same subject as the best match then the next two phases of the testing process are skipped and the result of the four face spaces is considered to be the best match. Otherwise, the five resultant subjects/classes are used as training set for the DCT based classification as described in flow chart 4.1.
Flow Chart No 4.1

In DCT based classification, (Flow Chart 1) the 2-dimensional DCT of the training set images is calculated and the coefficients are compared with the DCT coefficients of the test image. This is done by calculating the Euclidean Distance between the test image coefficients and the coefficients of the training set images. Minimum Euclidean distance is considered as the best result and the corresponding subject is considered to be the best match.
4.17 Experiments and Results

The proposed face recognition technique of SHPCA was tested and compared with PCA under two different set of conditions. In the first testing phase, the numbers of images of each class were fixed to five and the system was tested after training it with different number of subject classes. The number of classes were increased in steps of 5 from 10 to 40. The resultant comparison between standard PCA and SHPCA is shown in Table 4.7 and Figure 4.18.

Table 4.7. Comparison between standard PCA and Sub-Holistic PCA

<table>
<thead>
<tr>
<th>No of Classes</th>
<th>FRR for Normal PCA</th>
<th>FRR for SHPCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>97 %</td>
<td>99 %</td>
</tr>
<tr>
<td>15</td>
<td>97 %</td>
<td>99 %</td>
</tr>
<tr>
<td>20</td>
<td>92 %</td>
<td>98 %</td>
</tr>
<tr>
<td>25</td>
<td>91 %</td>
<td>98 %</td>
</tr>
<tr>
<td>30</td>
<td>89 %</td>
<td>97 %</td>
</tr>
<tr>
<td>35</td>
<td>86 %</td>
<td>96 %</td>
</tr>
<tr>
<td>40</td>
<td>84 %</td>
<td>94 %</td>
</tr>
</tbody>
</table>

Figure 4.18. Comparison of PCA and SHPCA Recognition Rate for 5 training images of each class
In second testing phase, number of training images of each class is reduced from five to three. Proposed face recognition model was tested by increasing the number of classes used from 10 to 40 in a step of 5. Results are shown in Figure 4.19.

![Figure 4.19. Comparison of PCA and SHPCA Recognition Rate for 3 training images of each class](image)

**4.18 Discussion of PCA and SHPCA Results**

PCA is a set of variables that defines a projection which encapsulates the maximum amount of variation in a dataset and is orthogonal (and therefore uncorrelated) to the previous principal component of the same dataset. It is commonly used in microarray research as a cluster analysis tool. It is designed to capture the variance in a dataset in terms of principal components. In fact, one is trying to reduce the dimensionality of the data to summarize the most important (i.e. defining) parts while simultaneously filtering out noise. Same is also being used in this model for face recognition through reduction of dimension and projection of variances in face images.
Table 4.5 illustrates that image resolution has a significant effect on FRR. At a particular resolution, facial features within a class become promising to each other. Posture of certain image information changes with variation in image resolution and becomes more pledging with minimum difference within own class for recognition at a particular resolution. With a change in image resolution, the discriminative information possessed by face images are also changed whereas structure information remains same in all images at one particular resolution which gives best FRR. Table 4.5 describes that with the reduction in image resolution FRR also changes and attains its best FRR at a particular resolution (56x46 in case of ORL database). This phenomenon supports Shannon sampling theory which states that up to certain resolution a face image retains all its original information which depicts that if image is reconstructed from this resolution it does not lack any original information. Further reduction in image resolution beyond this value causes negative effect on FRR because below this particular image resolution requisite image information required for recognition is lost.

In Table 4.6, results were gathered by keeping the image resolution (56 x 46) against best FRR constant and number of image classes were varied. As the number of classes used are increased, FRR is reduced due to close representation of face images in PCA subspace which increases the chance of wrong matching and resulting reduction in FRR.

In mathematical model of PCA covariance matrix is the base of calculations of eigenvalues and eigenvectors. It can be defined as tendency of two datasets to vary together and a covariance matrix is merely collection of many co-variances:

\[
C = (I_i - \bar{m})(I_i - \bar{m})^T
\]  

Once image resolution is reduced up to certain factor it reduces the high frequency clutter from the image not required for face recognition. Variance of this reduced data matrix provides better spread of variant data by removing undesired values from data matrix. Reduction in face data matrix should be such that if image is reconstructed from this resolution it does not lack any original information. The spread plot of one image with different resolutions is shown in Figure 4.20.
Figure. 4.20. Covariance spread with different image resolutions 56 x 46 (upper) and with 112 x 92 (lower)

Similarly, reduction in unnecessary data from covariance matrix correspondingly provides better eigenvalues choice as shown in Figure 4.21.
It also depicts that eigenvalue graph of reduced image resolution provides smooth elbow curve with better choice of few eigenvalues with maximum variance in images. It results in enhancing matching and recognition results. This particular image resolution level where reconstruction of image is perfect without loss of much facial information provides better eigenvector choice without any undesired clutter information. Moreover better choice of eigenvectors against narrow spread of eigenvalues provides improved FRR. Experiments on different resolution levels have been carried out and it is found that in ORL database two times reduction in resolution provide best FRR of 87%. This image resolution was also tested against change in number of classes used and results are shown in Figure 4.12 and Table 4.6.

![Figure 4.21. Plot of Eigenvalues of with 112 x 92(right side) and with 56 x 46 (left side)](image)

SHPCA is a new methodology used to improve the recognition rate as compared to standard PCA. First, image resolution was reduced to level where classic PCA achieved best FRR as discussed in previous classic PCA approach. Once image resolution through
decimation was reduced it created blurring effect in the image which was removed through a sharpening filter shown in Figure 4.22.

Image was divided into four parts (Figure 4.15), a two step face recognition approach is used to improve the image classification. In first step, PCA was individually applied on these four parts of image and on complete image as well and recognition was performed by using all these five parts. If decision is not yet made then DCT is applied to get best recognized image.

Figure 4.22. Sharpening Filter (right) and its Frequency Response (down)
The creation of the five face spaces in SHPCA results in increasing the training time to 50 seconds, compared to PCA’s single face space training time of 30 seconds for all 40 subjects of the ORL dataset. Although two step testing approach has increased the average recognition time per image to 0.66 seconds compared to PCA’s average recognition time per image of 0.02 seconds for the complete ORL dataset. However the use of the proposed technique increases the correct detection percentage to 94% for 40 subjects, compared to standard PCA’s correct detection percentage of 84% for 40 subjects. However, time delay can be reduced by incorporating better hardware.

4.19 Comparison with other PCA Based Face Recognition Techniques

A comparison of proposed face recognition techniques with current existing PCA based recognition techniques is shown in Table 4.8.
Table 4.8. Comparison of results of Proposed Face Recognition Model with Existing PCA based Recognition Techniques

<table>
<thead>
<tr>
<th>Face Recognition Technique</th>
<th>FRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>A novel incremental principal component analysis and its application for Face recognition[94] (H.Zho August 2006)</td>
<td>92.50%</td>
</tr>
<tr>
<td>Weighted PCA space and its application in face recognition [95] (H.Y. Wang, August 2005)</td>
<td>88%</td>
</tr>
<tr>
<td>Face recognition using kernel principal component Analysis[96] (K.In.Kim 2002)</td>
<td>90.0%</td>
</tr>
<tr>
<td>Face recognition based on face –specific subspace[97] (S.Shan 2003)</td>
<td>88.4%</td>
</tr>
<tr>
<td>Two-dimensional Weighted PCA algorithm for Face Recognition[98] (V.D. M. Nhat, Jun, 2005)</td>
<td>95.06%</td>
</tr>
<tr>
<td>Proposed SHPCA face recognition model</td>
<td>94%</td>
</tr>
</tbody>
</table>