CHAPTER 6

ARTIFICIALLY OCCLUDED IMAGE MODEL

6.1 INTRODUCTION

In the security area and surveillance face recognition is a vital part. 3D image is considered for face recognition as 3D imaging systems are promising. The greatest challenge in face recognition is occlusion. Occlusion is the hinderance in identifying a face like scarf, coolers or lighting effect, the pose of the person etc. This work presents a 3D detection technique for removing occlusion and to restore the partly occluded face with any unexpected object [97]. Normalization of the face image orients the image for recognition of the face. In the technique proposed, an effective way of detecting occlusions is done to indicate the missing information from the occluded image. Then a restoration technique is applied to remove the occlusion and deliver a restored image of the face. The information given by the non-occluded element of the face is to restore the actual facial image. After this, appropriate face recognition method is applied to the restored faces. Since prior information about the occluding object is not required only restoration strategy can be applied. Missing pixels caused due to inaccuracies during acquisition and noise is removed through this strategy. Best accuracy in removing occlusion is guaranteed in the proposed Face Recognition Model based on Artificially Occluded Image (FRMAOI) model. The restoration technique does not dependent on the face recognition system.
6.2 FRMAOI MODEL

Figure 6.1 depicts the FRMAOI design uses efficient methods to provide better results with respect to the current issues in face recognition and its pre-processing stages. An efficient normalization is applied prior to restoration. Restoration module is carried out to retain the features of the facial image not available due to occlusion. Finally, an effective and efficient recognition is performed.

![Figure 6.1 Architecture of FRMAOI](image)

6.2.1 3D Acquisition

It involves acquiring a 3D face image using a 3D scanner. Similarly occluding 3D object image is captured separately using the same scanner and is applied over the 3D captured facial image. Acquiring three-dimensional images through a scanner is a complicated process. Hence, it uses UND 3D dataset, which comprise of acquisitions 3D face images of different types of occlusions. UMB-DB dataset contains 3D faces, which are occluded by artificial occlusions such as glass,
scarf etc., and this dataset are used for projecting the image in 3D plane and for further process. Since it uses 3D images, projection of the image as a range image is a preliminary process.

**Range Image** - Range image is a specialised set of digital images [82]. They encode to locate the surface directly, so that the shape of a 3D image could be computed easily. Every pixel in a range image conveys the distance between a recognized reference frame and a visible point in the scene. Since the 3D dataset is used in the range image, a 3D structure of the face can be reproduced.

**Viola Jones** - Using the projected image, face detection is done to find out whether the image contains a face region and if so, it locates the exact facial region. This has to be done for the test and the training images. Face detection provides localization of the facial region from the projected image [90]. A new face detection method built up by Viola based on an adaboost and cascade algorithm is used and it also uses the rectangular Haar features. The Haar chosen masks will characterize particular facial features effectively. Face normalization is performed after detecting the faces in a given image [22].

### 6.2.2 Face Localization using Eigen Face

The 3D face model is obtained by the eigenface approach via Principal Component Analysis (PCA). Principal components or the eigenvectors of the covariance matrix of a group of face images treating an image as a point or vector in high dimensional space. Each face is characterized by linear combination of a set of images called eigenfaces [69].
\[ x + e = m + \sum_{i=1}^{N} y_i v_i \]

Where \( x \) is the input range image (encoded as a \( d \)-dimensional vector), \( e \) is the approximation error, \( m \) is the range image of the mean face, \( v_i \) are the \( N \) eigenfaces considered and \( y_i \) are the coefficients of the linear combination. Each coefficient \( y_i \) is attained by projecting the vector \( x-m \) onto the respective eigenface \([2, 28]\). Every face can also be approximated via simply the top eigenfaces, which have the leading Eigen values that account for the majority variance within the set of face images. The PCA uses every three spatial coordinates to encode the faces and it requires dense point to point connection between each face and the reference face.

The eigenface approach to face detection \([88]\) is summarized as following:

- A group of characteristic face images is collected of known persons. This set includes a numeral of images for every individual with a slight variation in look and lighting.
- Calculate the matrix that finds its eigenvectors and Eigen values and choose with the maximum associated Eigen values.
- The normalized training set of images is combined to create the eigenfaces.
- For every known individual, the class vector is calculated by averaging the eigenface pattern vectors that is computed from the
actual image of the person. A threshold is chosen that describes the highest distance permissible from any face class.

- The pattern vector is calculated by measuring the distance from every known class and the distance to face space. This is done for every new face image to be detected. Thus the face image is categorized according to the pattern vector and their distances.

- Once the fresh image is identified as a known person, it is added to the original database of known face images. Then the eigenface value is recomputed.

### 6.2.3 3D Face Normalization by Wavelet Transform

The scope of the model is for face recognition, it is essential that frontal face to be projected. Wavelet transform is used to rotate the 3D shape of the image if necessary (if pose change is present) and delivers a frontal facial image. Since 3D images are trained, the images have to be rotated to frontal position even if the pose is towards right or left angle. The normalization procedure finding the location of the fiducially points is employed to put the face in a benchmark position. It also helps to perform face image pre-processing, smoothing and small holes interpolation. In order to keep the middle portion of the face, the occluded image is either revolved or translated or cropped.

Edges are identified with the high frequency components of images and computing the high frequency components of an image by convolving a kernel with the image. It symbolizes a signal as a set of basis functions at diverse scales. The
joint spatial-frequency resolution attained by the wavelet transform creates it as a pattern for the extraction of details and approximation of images.

In the two band multi-resolution wavelet transform [76], the signal is described by wavelet and scaling basis functions are as follows

\[
f(x) = \sum_k a_{0,k} \varphi_{0,k}(x) + \sum_j \sum_k d_{j,k} \psi_{j,k}(x)
\]

Where \( \varphi_{j,k} \) are scaling functions at scale \( j \) and \( \psi_{j,k} \) are wavelet functions at scale \( j \), \( a_{j,k}, d_{j,k} \) are scaling coefficients. For the 2D discrete wavelet transform, the approximation coefficients (low frequency components) and detail coefficients (high frequency components) can be easily computed using a 2D filter bank has low pass and high pass filters. The wavelet decomposition is used to obtain diverse band information of face images and diverse band coefficients are processed independently.

### 6.2.4 Facial Feature Extraction

The Dual Tree Discrete Wavelet Transform (DT-DWT) is employed [87] and the sub-bands are processed to derive the edges of the feature points. After feature point extraction, skin color area and skin modeling has been done. The DT-DWT has improved directionality, condensed shift sensitivity and approximately orientation invariant. It uses two real wavelet transforms in parallel, where the wavelets of single branch are the Hilbert transforms of the wavelets in the other. Several input image can be breakdown into six directional sub-bands with complex
coefficients in diverse directionalities. It can be rotated to any direction and scaled to several resolutions.

6.2.5 Face Restoration

Face restoration is the process by which the occluded regions are located from the given test image and the occlusion is removed to generate original facial image. To achieve this, Gappy Principal Component Analysis (GPCA) face restoration [20] method is used.

3D Occlusion Detection

Detection strategy for occlusion has two important steps. Initial, the most obvious occlusions are noticed and developed occlusion mask is determined by application to reinstate the occluded faces [19].

Initial 3D Occlusion Detection

The initial mask is calculated by computing Distance from Feature Space (DFFS), by thresholding vector \( e \). This results in the preliminary mask computation \( B \). In first 3D detection of occlusions, the three-dimensional faces are encoded as range images, that is, images whose pixels are considered with the coordinates of a point in 3D space [7]. The range images are attained by orthographic projection of the obtained scenes on a 3D plane. An ensemble masks with the value of each either zero or unity depending on whether the pixel is masked or not masked respectively [29]. The masks are randomly generated and it is characterized by the fraction of unmasked pixels. It proceeds iteratively as follows:
• Repair each ensemble marred face by filling in missing pixels by the average values of those locations.

• Employ Karhunen-Loeve (K-L) procedure [31] to generate Eigen function as a complete orthonormal system.

• Obtain repaired ensemble marred face by fitting each ensemble marred face of the original ensemble by a superposition of Eigen functions.

• Determine the repaired ensemble marred face by minimizing the criterion function over the pixels for which data are available.

A preliminary mask $M$ of occlusions can be acquired by threshold vector $e$,

$$M_i = \begin{cases} 
1 & \text{if } e^i > \tau \\
0 & \text{if } e^i < \tau 
\end{cases}$$

$\tau$ is a threshold that takes into account the resolution of the imaging device, the acquisition noise, and the exactness in face detection and pose normalization. Occluding objects remote from the face are detected on the basis of the differentiation among the occluded face $x$ and the mean (non-occluded) face $m$. Components of $m-x$ which are positive and great adequate can be considered part of an occlusion. A mask $B$ of the occlusion is acquired as follows.

$$B_i = \begin{cases} 
1 & \text{if } m_i - x_i > \nu \\
0 & \text{if } m_i - x_i < \nu 
\end{cases}$$

Here $\nu$ is a threshold that must be tolerant with respect to face changeability in the data set to be processed.
The GPCA is a variation of Principal Component Analysis and it is applied for more accurate estimation of occlusion $M$ by not including the pixels chosen in mask $B$ from the calculation of the reconstruction error $e$. Given a 3D face $x$, and the equivalent mask $B$, the coefficients in 1 are determined once more excluding the components $x_i$ for which $B_i=1$. We let $(u, v)_b$ be the gappy inner product, depends on the occlusion mask.

$$(u, v)_b = \sum_{i=1}^{d} u_i v_i (1 - B_i)$$

The coefficients which encode the processed face in the face space are selected to reduce the reconstruction errors $e$ on the non-occluded parts as given below.

$$\left\| e \right\|^2_B = \left( \left\| -x + m + \sum_{i=1}^{N} y_i v_i \right\|_B \right)^2$$

$$\left\| e \right\|^2_B = \left( \left\| x - m \right\|_B \right)^2 - 2 \sum_{i=1}^{N} y_i (x - m, v_i)_B + \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j (v_i, v_j)_B$$

Differentiating with respect to every and needed that the partial derivatives be null and get a system of $N$ linear equations, where $N$ is the numeral of eigen faces retained using the following equation.

$$\frac{\partial \left\| e \right\|^2_B}{\partial y_i} = -2(x - m, v_i)_B + 2 \sum_{j=1}^{N} y_j (v_j, v_i)_B = 0$$
The coefficients obtained are solved and used to determine a more accurate reconstruction error. On the basis of the new reconstruction error $e$ is gained using the coefficients estimated by solving the system and the occlusion mask $M$ can now be determined more exactly. While the reconstruction error is huge, the equivalent pixel is measured as occluded.

$$M_j = \begin{cases} 
1 & \text{if } (e_j > \tau) \quad B_j = 1, H_j = 1 \\
0 & \text{otherwise}
\end{cases}$$

If the resultant mask $M$ is blank, the face is measured as non-occluded and can be directly processed for recognition, if not $M$ restores $B$.

$$\frac{\delta \| \delta \|^2}{\delta y_i} = -2(x - m, v_i)_B + 2 \sum_{j=1}^{N} (v_j, y_j)_M = 0$$

### 6.2.6 Face Recognition by Average Regional Model

The restored faces can be used for recognition from a set of training image. This can be achieved by any holistic or feature based method. The input face is annotated manually taking into account the position of non-occluded landmarks. The face is aligned approximately by analyzing the landmarks. The Iterative Closest Point (ICP) algorithm registers the mean face. The mean face is subdivided into eight regions on the basis of a predefined division. Every region of the mean face is registered to the aligned test face defining a corresponding region on the test face. Every region of the test face is matched by an approximated volumetric difference with the corresponding region of gallery faces. The regions corresponding to a bad ICP registration (defined by a suitable threshold) are considered as occluded and are
discarded. At this point, the non-occluded regions match to the nearest neighbor classifiers is joint using fusion methods. This method is based on a preliminary physical selection of landmarks. So, the recognition performances are not affected by potential errors that introduced in automatic face detection and normalization procedure. Average Regional Model (ARM) finds the correct match from the given dataset of images.

A part-based 3D face-recognition method is able to perform robustly under expression variations and in the presence of significant amount of occlusions. This method is adopted from Average Face Model-based (AFM) Average Region Models (ARMs) [7] used for face registration. The facial area is separated into numerous meaningful components like eye, mouth, cheek and chin regions. Registration of faces is passed out by separate dense alignments to relative ARMs. The dissimilarities between the gallery and test faces gained for individual regions are combined to decide the final dissimilarity score. Under the extreme variations like occlusions, the combination method can automatically decide noisy regions and are discarded. Thus, the aim of the method is to discover regional association among any two face. It has following steps, i) coarse and dense ARM-based registration and ii) local matching and classifier fusion. The technique uses a two-phase approach are global coarse and local dense registration.

**Global Coarse Registration**

It is carried out to approximately align a specified 3D face image to the AFM. ARMs are constructed on the AFM by deciding the semantic regions
physically. The entire facial model is separated into four parts, i) eye/forehead, ii) nose, iii) cheeks and iv) mouth chin regions.

**Local Dense Registration**

This is carried out by aligning local regions with ARMs using the ICP algorithm. Each region over the test face is registered to its corresponding ARMs separately. Registered regions are then regularly re-sampled. Therefore, after local dense registration, facial components are automatically determined over the given facial surface.

**Local Matching and Classifier Fusion**

Following the registration phase, every part of the face is labeled and it is likely to compute regional difference scores. Each region to be measured as individual classifier and the 1-nearest neighbor algorithm is used as a pattern classifier. The regional similarity or these classifiers are employed in fusion setting.

Many fusion techniques are used in different levels like plurality voting at the abstract level, Borda count used at the rank point and summation rule, product rule and logistic regression at the score level. In plurality voting, every expert give the class label of the adjacent gallery subject and the class label with maximum vote is allocated as the final label. In the Borda count method, the individual classifiers deliver a ranked list of class labels and the class having minimum count is selected. The score level fusion techniques employ sum and product rules. The similarity measures produced by individual experts are joined together with easy arithmetic rules. The min-max normalization method is employed to normalize the scores. The
scores calculated by every classifier are regarded as random variables and logistic regression calculates the optimal weights for the estimation of unknown class label.

Because each region is aligned, this has chosen to employ an approximation of the volumetric difference among local surface pairs with the aid of point set differences. Euclidian distance is calculated for the facial regions separately for the test and training images, which is then compared to find out the test image corresponds to the training image. If occlusions are present in faces, then the benefit of the projected ARM-based 3D face recognizer is efficient, either by fusing all independent area or by routinely detecting or removing occluded area. It is likely to get best classification rate, hence the resultant face is appropriate for face recognition.

6.3 IMPLEMENTATION

In FRMAOI design is tested using UMB-DB image dataset [21]. The 3D facial information can give a hopeful approach to comprehend the quality of the person face in 3D domain and has potential chance to get better the performance of the recognition system. Since the 3D cameras are not as general as 2D cameras, it is costly to construct a community 3D face dataset, which gets the complexity to validate the schemes planned in an identical platform. The University of Milano Bicocca 3D face dataset is a group of multimodal (3D + 2D) colour images) facial acquisitions. The traditional dataset called UMB-DB [21] has been with a particular focus on facial occlusions, i.e. scarves, caps, hands, spectacles and other types of occlusion, which can happen in real-world situations. Figure 6.2 shows the sample faces of unbiased, with expression and occluded.
Figure 6.2 Sample Faces (Neutral, Expression and Occluded)
In FRMAOI model have three steps, Step 1 takes an input face image, which is an occluded image and it is processed the face portion detected in step 1. The face portion is retrieved from the output of step 1 by using step 2. This step also removes the occluded object and produces a non-occluded image as output. Step 3 contemplates the output of step 3 image to 3D image in 3D coordinate plan. Figure 6.3, 6.4 and 6.5 shows the output in each step for recognizing a 3D image.

6.4 EXPERIMENTAL RESULTS AND ANALYSIS

The results are verified using UMB-DB dataset consisting of 3D images captured using 3D scanner. The training dataset consists of 150 images. Each image is classified with various occlusions like cap, glass, pen, scarf, eyeglass, hair, hand etc. The dataset not only classifies under the occlusion category but it also provides sufficient images classified under different types of expression. The system is fully implemented with the proposed approaches using UND dataset [41].

(a) Normalized Frontal Image (b) Face Detection from Test Restoration Image

Figure 6.3 Implementation-Step 1
The image is normalized and from this, the mask (occluded region) is annotated. The restored face is used to compute the Eigen average face from which face is recognized using the samples in the test dataset. Test and training dataset are classified separately. The input to the model would be an image from test dataset. The image follows the subsequent steps and finally the recognition of face is performed using the image in the training dataset. Every image is tested with the ODRM for different images handling the various occlusions. The result of the
execution (whether recognized or not) is calculated separately and tabulated. It is noted that the model handles occlusions like eyeglass and hat efficiently. Images with changes in expression show considerably low recognition. False recognitions are also calculated and highlighted. The model is focuses on reducing the false positive rate.

Table 6.1 shows the false positive rate for both the model and Figure 6.6 represents the plot using Table 6.1. The following graph represents the false-positive rate for various occlusions handled using UND dataset by the ODRM. From this graph, it shows that the false positive rates are high for mouth occlusion and changes in expression. The both ODRM and proposed FRMAOI model uses the same type of images and occlusions in UND dataset for efficiency comparison.

**Table 6.1 False Recognition Rate - ODRM Vs. FRMAOI Model**

<table>
<thead>
<tr>
<th>Type of Occlusion</th>
<th>False Positive Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ODRM</td>
</tr>
<tr>
<td>Eyeglass</td>
<td>0.005</td>
</tr>
<tr>
<td>Hat</td>
<td>0.0175</td>
</tr>
<tr>
<td>Hair</td>
<td>0.03</td>
</tr>
<tr>
<td>Eye</td>
<td>0.0575</td>
</tr>
<tr>
<td>Mouth</td>
<td>0.04</td>
</tr>
<tr>
<td>Angry</td>
<td>0.02</td>
</tr>
<tr>
<td>Smile</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Table 6.2 shows the comparison results for various types of occlusion for both the model for face recognition efficiency. Figure 6.7 shows the recognition efficiency of FRMAOI model which outperforms the ODR model.

**Table 6.2 Face Recognition Efficiency ODRM Vs. FRMAOI Model**

<table>
<thead>
<tr>
<th>Type of occlusion</th>
<th>ODRM</th>
<th>FRMAOI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eyeglass</td>
<td>97.84%</td>
<td>99.03%</td>
</tr>
<tr>
<td>Hat</td>
<td>95.56%</td>
<td>98.52%</td>
</tr>
<tr>
<td>Hair</td>
<td>90.56%</td>
<td>89.62%</td>
</tr>
<tr>
<td>Scarf</td>
<td>92.62%</td>
<td>95.39%</td>
</tr>
<tr>
<td>Mouth</td>
<td>92.85%</td>
<td>95.14%</td>
</tr>
<tr>
<td>Angry</td>
<td>88.59%</td>
<td>96.71%</td>
</tr>
<tr>
<td>Smile</td>
<td>93.19%</td>
<td>94.04%</td>
</tr>
</tbody>
</table>
Figure 6.7 Comparison of Face Recognition Efficiency

The comparison of false detection rate is provided in the Table 6.3. From the Figure 6.8 shows the decrease in false detection rate and increase in efficiency of face recognition. The recognition efficiency and false detection rates are considerably improved by FRMAOI model. The results of both the model are discussed and highlighted.

Table 6.3 False Detection Rate ODRM Vs. FRMAOI Model

<table>
<thead>
<tr>
<th>Type of occlusion</th>
<th>ODRM</th>
<th>FRMAOI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eyeglass</td>
<td>97.84%</td>
<td>99.03%</td>
</tr>
<tr>
<td>Hat</td>
<td>95.56%</td>
<td>98.52%</td>
</tr>
<tr>
<td>Hair</td>
<td>90.56%</td>
<td>89.62%</td>
</tr>
<tr>
<td>Scarf</td>
<td>92.62%</td>
<td>95.39%</td>
</tr>
<tr>
<td>Mouth</td>
<td>92.85%</td>
<td>95.14%</td>
</tr>
<tr>
<td>Angry</td>
<td>88.59%</td>
<td>96.71%</td>
</tr>
<tr>
<td>Smile</td>
<td>93.19%</td>
<td>94.04%</td>
</tr>
</tbody>
</table>
6.5 SUMMARY

The face recognition approach robust to artificial occlusions and other challenges like expressions and pose is developed. Face detection finds the face in the given image and normalization helps to bring the frontal view of the 3D image if the test image does not provide the frontal view. The ARM based Face restoration method is implemented in this model, which secures high efficiency, due to this removal of artificial occlusion and recognition rates are improved. The FRMAOI model is developed with the above techniques that handle three-dimensional images. This model is evaluated using 3D UMB-DB dataset. It provides efficient results for face recognition.