5. Chapter 5

CHAPTER 5

RECOGNISING PARTIALLY OCCLUDED FACE USING MODIFIED EXEMPLAR INPAINTING

5. RECOGNISING PARTIALLY OCCLUDED FACES USING MODIFIED EXEMPLAR INPAINTING

5.1 Introduction

People are studied to magnified levels of precision, using video and photography, to understand the finer nuances that could highlight them as possible subjects of study. As a self-sufficient examiner, the video arrangement is able to identify potential subjects without human intervention. Face recognition highlights a very crucial element in PC vision. Since the client is not required to be ever intervene with the system, there may arise a situation where in the visage of the subject may not be completely captured, this highlights the need to detect the totality of the human face to bring forth a perfect frame of reference. The same occurs if a user is using facial accessories, this can likewise bring about impediment which makes it troublesome for the framework to recognize the face image. If a face is occluded either due to self-occlusion or inter object occlusion it limits the information in an image. To regain this occluded region and perform recognition to achieve higher recognition rate is a challenge.

5.2 Proposed Model

This work concentrates on an improvement in recognition rate when human faces are partially occluded in a video succession. Figure 5.1 shows the proposed model. In the proposed work, a preparation set is made with reproduced blocked appearances. The recreation of occluded region is performed using Inpainting strategy. Inpainting is a technique of recovering lost region in an image. There are different methods to recover lost pixels in an image. Exemplar inpainting is one of the best methods to recover the lost region. Modified Exemplar Inpainting is utilized as a part of remaking the occluded face region. Once the input video is introduced to the framework, the occluded area is recognized if any. Features from the face are extracted using Curvelet Transform.

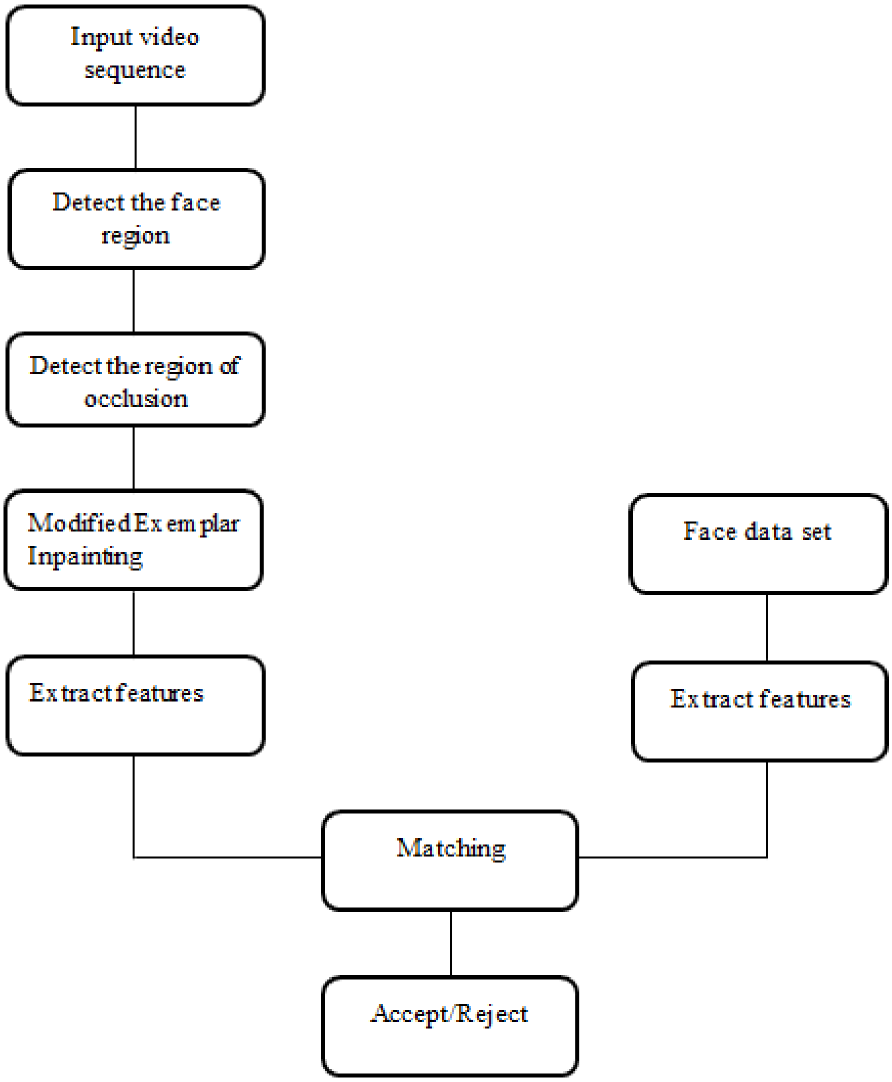

Figure 5.1: Proposed Model for Occluded Face Recognition

5.3 Face Detection

Face detection is an important step in face recognition technique. For real-time face detection, viola Jones algorithm gives a good detection rate. The work proposed by (Viola and Jones 2001) describes a machine learning approach for detecting objects which give high detection rate. According to the work carried out, it is a composition of three key features: (i) New image representation which authors represented as an Integral image. This helps in quickly computing the features used by the detector. (ii) Learning algorithm based on AdaBoost. This selects features from and yields efficient classifiers. (iii) Combines complex classifiers which discard the background region and does the computation of object-like regions. Based on this work, extended face detection for multi-view faces is proposed in (Jones and Viola 2003). In this work, authors used building different detectors for different views, (Rowley et al. 1998), and (Schneiderman and Kanade 2000). For this work, the authors extended their previous work for rotated and profile faces, as well as new sets of rectangle features, are defined for rotated face detection. The authors in their work used classifiers based on the weak classifier set from the Adaboost algorithm by (Freund and Schapire 1995).

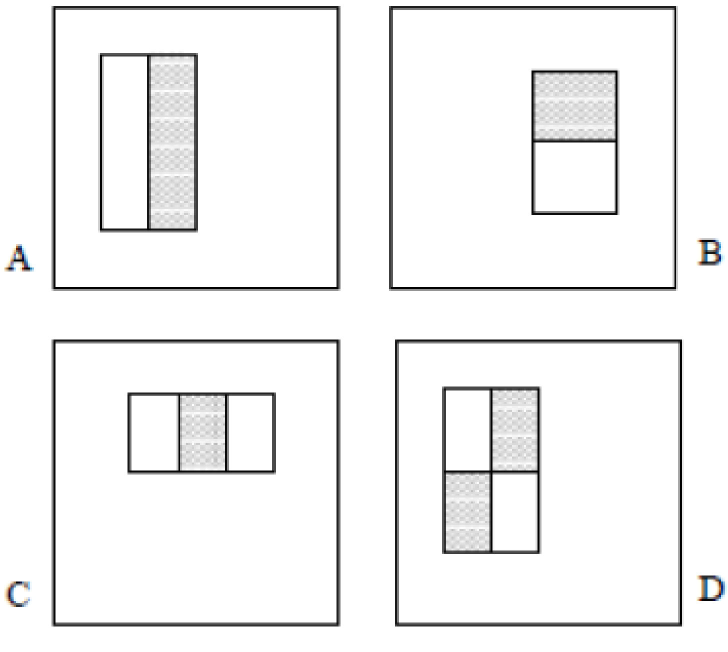

Figure 5.2: Haar features considered for Adaboost.

The features shown in Figure 5.2, are the important Haar features used to detect a face region.

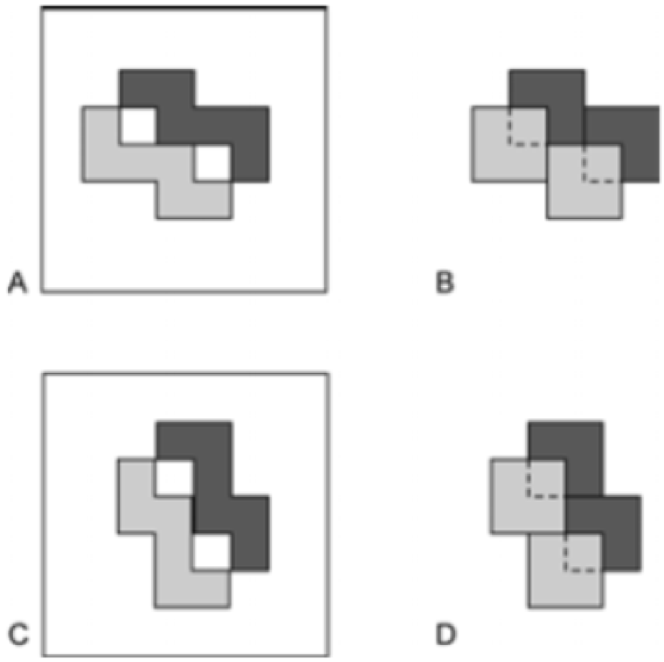

Figure 5.3: Diagonal Haar Features Designed for Face Detection

Figure 5.3 shows the diagonal filters used by the authors that are used to detect non-upright faces and non-frontal faces giving highly accurate results. Authors used a sequence of the complex classifier to improve the efficiency in computation. Input window is evaluated on the first classifier and if it returns false then computation ends on that window if it returns true, next classifier in the cascade evaluates the window. In a similar way, window passes through all the classifiers and if true, the detector will return a true value for that window. In order to detect the face with different poses, the poses are divided into various classes and different detectors are trained for each class. (Meynet et al. 2003) described the algorithm for AdaBoost in their work in detail and is summarised below. The working of AdaBoost algorithm includes one weak classifier is selected at each step. A weight is assigned to the data at each step and a weak learner is constructed based on the weight, (Schapire 2013). To estimate how the data is classified, the weak learner produces an empirical error. The weak classifier with the best classifier with the lowest error is selected and weights are updated accordingly. The iteration loop stops if the empirical error calculated is 0 or >= ½.

Table 5.1: AdaBoost Algorithm for face Detection

Input: Set $(x_1,y_1),\ldots.(x_m,y_m)$ where $y_i \epsilon \{-1,+1\}$

Set $D_1(i)=1/m$ for $i=1,\ldots.,m$.

For $t=1,\ldots.,T$:

- Normalize the weights to get the probability distribution D_t as $D_t(i) = \frac{w_{t,i}}{\sum w_{t,n}}$
- A weak classifier h_j is generated for each feature f_j.
- Determine the error ε_j of h_j with respect to D_t:

  $$\varepsilon_j = \sum_{n=1}^{N} w_{t,n}|h_j(x_n) - y_n|$$

- Choose classifier h_j with lowest error ε_j and assign $(h_t, \varepsilon_t) = (hj, \varepsilon_j)$.
- Update the weights $w_{t+1,n} = w_{t,n}\beta_t{}^{1-e_n}$ where $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$ and e_n =0 if x_n is classified correctly or else 0.
- The final strong classifier:

  $$h(x) = 1 \; if \; \sum_{t=1}^{T} \log\frac{1}{\beta_t} h_t(x) \geq \frac{1}{2}\sum_{t=1}^{T} \log\frac{1}{\beta_t}$$

  $$h(x) = 0 \; otherwise$$

5.4 Occlusion Detection

Once the face is detected from the sequence of video, next is to check if this face image suffers from occlusion. (Zitnick and Kanade 2000) proposed a method using stereo algorithm that gives in disparity maps that can detect occlusion. In their approach, a 3D array is constructed with elements (r,c,d) with (r,c) for each pixel in the reference image and d signifying the range of disparity. Authors have assumed that the disparity maps possess a unique value per pixel and are continuous everywhere. In this method, normalized correlation or squared differences are used to set the intensity values. With each pixel in the image, (r,c,d) element with the maximum match value is found if the match value is found to be lower than a threshold, the image is classified as occluded. (Chen et al. 2011) proposed a method of dividing the face into patches and matching each patch with the gallery images to detect occlusion. (Pan et al. 2010) proposed a method to track objects based on particle filter. To execute this method the pixels in an object were classified as foreground and background pixels using background subtraction method. Upon extracting the object from background, the object is considered to be the region of interest. Ellipse is used to represent the image and occlusion is detected by merging and splitting the image. SVM based approach is used for detecting occlusion in face.

5.5 Support Vector Machine for classification

Support Vector Machine (SVM) is a pattern classification algorithm developed by (Cortes and Vapnik 1995). The idea used in SVM classification is structural risk minimisation. Training a SVM is equivalent in training a quadratic programming problem and hence it is difficult when the number of data is large.

Support Vector machine can be used as a regression technique and classification problem. According to (Cortes and Vapnik 1995), SVM in its simplest linear form is a hyper plane that separates a set of positive examples from a set of negative examples with maximum margin. This is depicted in Figure 5.4. The figure shows the positive data in squares and the negative data in circles. There can be ‘n’ number of lines that can differentiate the positive and negative data. In SVM the line that separates the data with maximum distance is chosen.

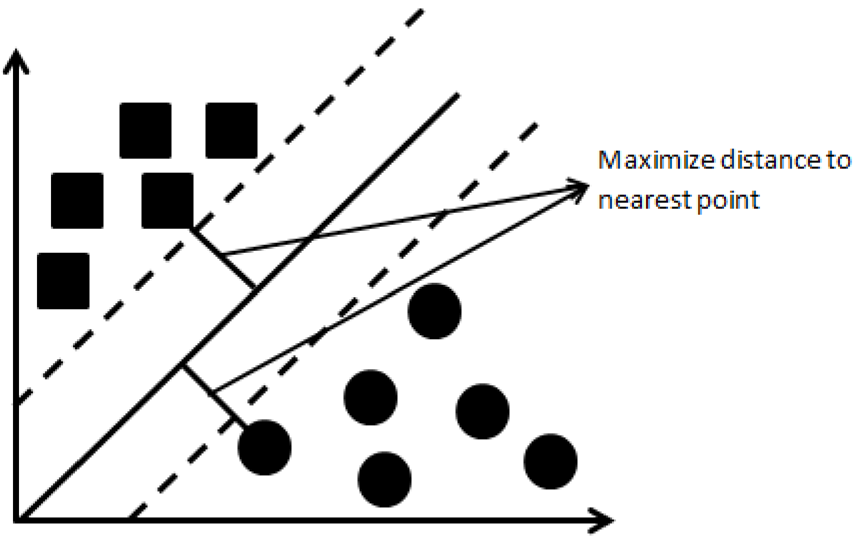

Figure 5.4: Classification of data using support vector machine.

Support vector machine (SVM) works well with variety of problems under different domains like hand written character recognition, face detection, pedestrian detection and text categorization. Even then, there exist minor disadvantage with SVM as the training algorithm for SVM is slow for large problems. SVM training algorithms suffer with a minor disadvantage because they are complex, subtle and difficult to implement. SVM in its linear form, the maximum margin is defined by the distance of the hyper plane to the nearest of the positive and negative examples, (Wang 2005).

The algorithm can be described below.

Considering a linear form of data, the output of a linear SVM can be defined as

$u = \bar{w}\bar{x} - b$ (5.1)

Where, w is the normal vector to the hyper plane and x is the input vector.

The separating hyper plane is the plane $u = 0$

The nearest point lies on the planes $u = \pm 1$

The margin m can be given as m= $\frac{1}{\|w\|_2}$ (5.2)

Maximising the margin can be expressed as

$\min_{\vec{w},b} \frac{1}{2}\|\vec{w}\|^2$ subject to $y_i(\vec{w}\vec{x_i} - b) \geq 1 \; \forall i$ (5.3)

Here

x_i is the ith training example

y_i is the correct output of SVM for ith training example

y_i is +1 for positive examples in a class and -1 for negative examples.

Converting this optimisation problem in QP problem using Langrangian,

$\min_{\vec{a}} \varphi(\vec{a}) = \min_{\vec{a}} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \left(\vec{x}_i \vec{x}_j\right) a_i a_j - \sum_{i=1}^{N} a_i \, , a_i \geq 0,$

$\forall i \; \sum_{i=1}^{N} y_i a_i = 0$ (5.4)

In equation 5.4, the objective function is solely dependent on a set of langrange multipliers a_i

Once langrange multiplers are determined, the normal vector and threshold can be derived from langrange multipliers.

5.6 Feature Extraction

Curvelets are delivered remembering the finished objective to vanquish the controls that are gone up against by wavelets. Curvelets have various orientations and positions at each length scale and at fine scales they possess needle formed parts. The basic advantage with respect to discrete curvelet transforms are (i) They provide optimally sparse representation that displays smoothness along the curve. (ii) Curvelets are the very important tool for analysing and computing Partial Differential Equations (PDE). Curvelets can behave both like waves and particles. (iii) In severely ill-posed problems, curvelets can be optimally used in image reconstruction.

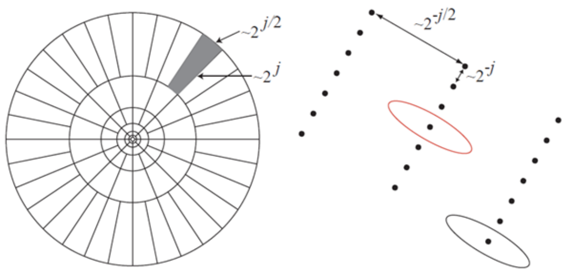

Figure 5.5: Curvelets

The major disadvantage with respect to wavelet is that they will denoise the image. To overcome this disadvantage curvelet transform were developed.

5.7 Inpainting

Modified Inpainting approach is used in filling in the occluded region of the face image. This method is modifies Criminisi approach in identifying the priority of a pixel to start the texture filling in procedure. In this approach, the data term is determined by the edge detection by Sobel operator and further gradient calculation.

Table 5.2: Modified Exemplar algorithm

I: The whole image
Ω: The portion of the image I that is occluded
Ψ_p: the patch formed for a search
Step 1: *If* p ϵ I
C(p)=0;
else
C(p)=1;
Step 2: *For* all pixels on the fill front
{
$C(p) = \frac{\sum q \in \Psi_p \cap \emptyset\, C(q)}{\lvert\Psi_p\rvert}, 0 \leq C \leq 1$
Calculate gradient magnitude $D(p) = \sqrt{{G_x}^2 + {G_y}^2}$
Calculate C(p).D(p)
} *Continue*
Step 3: Form a patch Ψ_p with p as the pixel at center.
Step 4: *For* all p ϵ I- Ω
Find the most similar patch Ψ_p'
Step 5: Replace the similar patch with Ψ_p
Step 6: Repeat step2 to step 6 until Ω is filled.

5.8 Proposed Algorithm

The proposed work described in Section 5.6 is consolidated in Table 5.3 and Table 5.4. The input is a video sequence with frames consisting of face with occlusion. To recover the lost data a modified exemplar-based inpainting that is described in *Chapter 4* is incorporated.

Table 5.3 : Proposed algorithm for occlusion recovery

```
V: Defines the video sequence as the input.
I: Frame that is chosen from the video.
G: Set of gallery images.
N: Number of gallery subjects.

for all frames in the video
    Apply Viola Jones face detector
    Detect face and represent using a bounded box
    for all detected face in a frame
        Apply SVM classifier
        Detect occlusion if any of the face region is occluded
        if (occluded == true)
        {
            Apply modified exemplar-based inpainting to recover the lost region.
            Extract features of the recovered face using Curvelet Transform.
Match function();
End for
End for
```

Table 5.4:Proposed algorithm for Occlusion recovery to match the face images.

Match function()
Match the face with the face images stored in the database.
If (Match == True)
Accept
Else
Reject
}
Else
{
Extract features of the recovered face using Curvelet Transform.
Match the face with the face images stored in the database.
If (Match == True)
Accept
Else
Reject
}

5.9 Performance Analysis

In order to test the performance of the method, five different occluded areas are considered on the face as illustrated in Figure 5.5.

Figure 5.6: Patches included on face image for inpainting

As the first step patches are placed on the right side of the face. Recognition rate with respect to each patch applying the proposed method is given in Table 5.5.

Table 5.5: Recognition Rate with respect to Different Patches(Right side of the Face)

	P1	**P2**	**P3**	**P4**	**P5**
Proposed method	85%	94%	98%	96%	98%

This experiment is repeated for left side of the face with same size of the patch. The result after the proposed method is applied is given in Table 5.6.

Table 5.6: Recognition Rate with respect to Different Patches (Left side of the Face)

	P1	**P2**	**P3**	**P4**	**P5**
Proposed method	90%	94%	98%	96%	98%

For the above experiment, occlusion on part of a particular feature from the face is considered. The results show that if the occluded region does not completely cover a particular feature from the face, it can be recovered with the proposed method with a good recognition rate.

To check the efficiency of the algorithm for a larger occluded region, following patches are considered.

The results obtained after applying the proposed method is given in Table 5.7.

Table 5.7: Efficiency of the Proposed Method.

	Scarf	**Sunglass**	**Glass**	**hair**
Proposed method	50%	52%	80%	90%

From Table 5.7, it can be concluded that, if the feature is completely occluded, recognition rate reduces as for inpainting, similar patches are not available and is recovered with most similar patch which is not same as the feature.

Considering the feedback from the above experiment, ten videos from YouTube dataset is considered to test the recognition rate of faces. These input videos included occlusion due to hair and glasses. The results are given below in Table 5.8. The Table shows the recognition rate with different video sequences. A comparative study is performed with existing methods to analyze the performance.

Table 5.8: Recognition Rate on Occluded faces.

Video sequences	Proposed approach	Patch Based	SLNMF
1	93	92	90
2	92	91	89
3	92	90	88
4	92	90	88
5	91	89	86
6	90	88	85
7	89	86	83
8	89	86	83
9	86	85	82
10	85	83	80

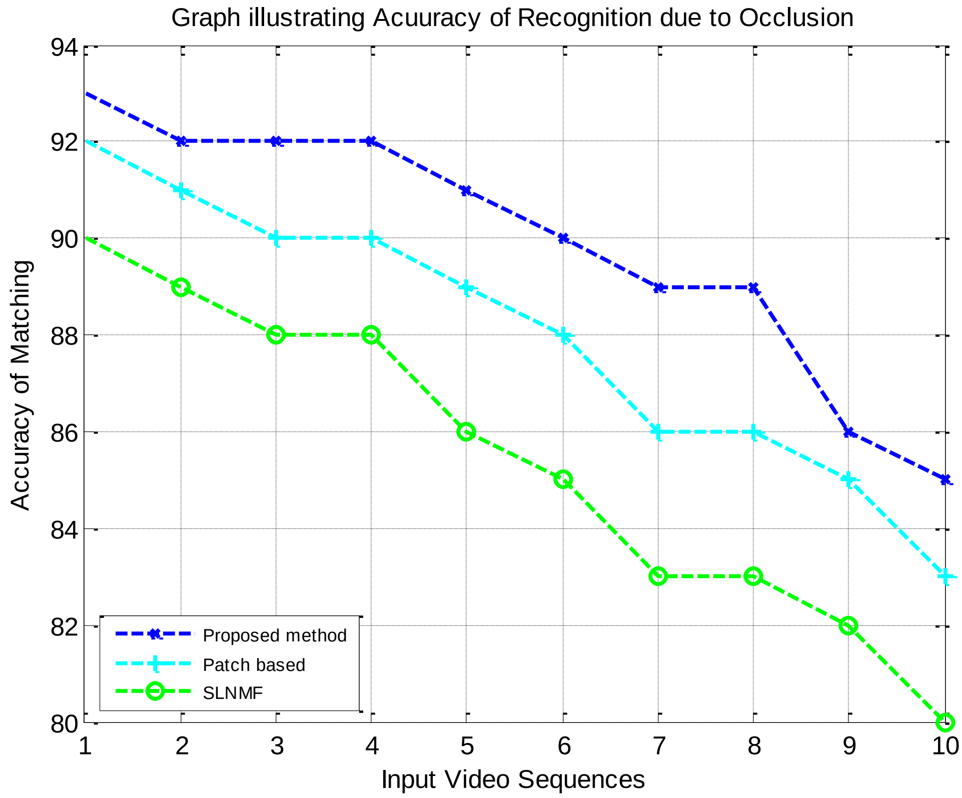

Figure 5.7: Graphical Representation for Recognition Rate

5.10 Summary

Recognizing face from video is an important biometric widely accepted by public. This biometric suffers from drawbacks like decrease in recognition rate due to illumination, occlusion and pose variation. In this work the problem of occlusion is attended. A modified exemplar inpainting is used in recovering the lost part of a face image. Exemplar is a texture synthesis method where a patch from known area replaces the occluded region depending on the priority of the pixels from the contour of the missing region. The result when this method is applied on videos of You Tube dataset is depicted in Table 5.8 and the result is analyzed in Figure 5.6. The result shows a better recognition rate when compared with two of the existing methods, patch based and using SLNMF.