

**T.C.**  
**BAHÇEŞEHİR UNIVERSITY**  
**OCCLUSION ANALYSIS IN FACE FRONTALIZATION**

**Master's Thesis**

**ANIL ÇELİK**

**İSTANBUL, 2016**

**T.C.  
BAHÇEŞEHİR UNIVERSITY**

**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES  
COMPUTER ENGINEERING**

**OCCLUSION ANALYSIS IN FACE  
FRONTALIZATION**

**Master's Thesis**

**ANIL ÇELİK**

**Thesis Supervisor: Assoc. Prof. Nafiz ARICA**

**İSTANBUL, 2016**

**THE REPUBLIC OF TURKEY  
BAHCESEHIR UNIVERSITY**

**FACULTY OF ENGINEERING  
COMPUTER ENGINEERING**

Name of the thesis: Occlusion Analysis in Face Frontalization

Name/Last Name of the Student: Anıl Çelik

Date of the Defense of Thesis: 12.01.2016

The thesis has been approved by the Graduate School of \_\_\_\_\_.

Assoc. Prof., Nafiz ARICA  
Graduate School Director  
Signature

I certify that this thesis meets all the requirements as a thesis for the degree of Master of Arts.

Asst. Prof., Tarkan AYDIN  
Program Coordinator  
Signature

This is to certify that we have read this thesis and we find it fully adequate in scope, quality and content, as a thesis for the degree of Master of Arts.

Examining Comittee Members

Signature

\_\_\_\_\_

Thesis Supervisor  
Assoc. Prof., Nafiz ARICA

-----

Member  
Assoc. Prof., M. Alper TUNGA

-----

Member  
Assoc. Prof., Songül ALBAYRAK

-----

## **ACKNOWLEDGEMENTS**

With deep gratitude to my thesis advisor, Assoc. Prof. Nafiz ARICA, for his guidance, support and significant comments throughout the preparation of this thesis.

With deep gratitude to my family for their love and support in all stages of my education. I appreciate them for letting me free to take my own steps in life and dedicate this thesis to them.

İSTANBUL, 2016

ANIL ÇELİK

## ABSTRACT

### OCCLUSION ANALYSIS IN FACE FRONTALIZATION

Anıl Çelik

Computer Engineering

Thesis Supervisor: Assoc. Prof. Nafiz ARICA

January 2016, 57 Pages

Frontalization is the process of generating frontal faces from the posed ones appearing in unconstrained environments. Occluders occurring over the face region can make the frontalization approaches generate faulty results. In this thesis, we propose an approach, to address this problem by reducing negative effects of occluders on the frontalization operation. The proposed approach has the capability of choosing the most suitable frontalization procedure, to generate a more visually appealing output. After the posed face image is hard-frontalized, the possible occlusion occurrences over the face are analyzed using two different techniques; region and pixel based analysis. Finally, according to occlusion analysis output, the best approach for frontalization process is chosen to be applied. The experiments performed on Caltech Occluded Faces Database (COFW) show that the proposed algorithm produces satisfactory results in terms of accuracy and visual appearance.

**Keywords:** Face Normalization, Face Frontalization, Occlusion Analysis

## ÖZET

### OCCLUSION ANALYSIS IN FACE FRONTALIZATION

Anıl Çelik

Bilgisayar Mühendisliği

Tez Danışmanı: Doç. Dr. Nafiz ARICA

Ocak 2016, 57 Sayfa

Önleştirme, kısıtsız ortamlarda bulunan surat imgeleri üzerinden ön görünüme sahip olan surat imgeleri sentezlemek olarak tanımlanabilir. Yüz bölgesi üzerinde oluşan kapanmalar, önleştirme yaklaşımlarının hatalı sonuçlar üretmesine sebep olabilmektedir. Bu tezde, bahsi geçen probleme çözüm olarak, kapanmaların önleştirme işlemi üzerindeki olumsuz etkilerini azaltmak için bir yaklaşım önerilmektedir. Bahsedilen yaklaşım, daha tatmin edici görsel sonuçlar elde edebilmek için en uygun önleştirme yöntemini kendi başına seçme kabiliyetine sahiptir. Poz açısına sahip olan yüz, sert-önleştirme sürecinden geçtikten sonra, yüz bölgesi üzerindeki olası kapanmalar iki farklı analiz yöntemine tabi tutulur, bölge ve piksel tabanlı olmak üzere. Son olarak, kapanma analizi sonuçlarına dayandırılarak en uygun önleştirme yöntemi, uygulanmak üzere seçilir. Tezdeki yaklaşımın isabetli ve görsel olarak tatmin edici sonuçlar verdiğini göstermek için Caltech Occluded Faces Dataset üzerinde deneyler yapılmıştır.

**Anahtar Kelimeler:** Yüz normalizasyonu, Yüz önlemeştirme, Kapanma Analizi

## CONTENTS

<b>TABLES</b> .....	<b>viii</b>
<b>FIGURES</b> .....	<b>ix</b>
<b>ABBREVIATIONS</b> .....	<b>xii</b>
<b>SYMBOLS</b> .....	<b>xiii</b>
<b>1. INTRODUCTION</b> .....	<b>1</b>
<b>2. LITERATURE SURVEY</b> .....	<b>4</b>
<b>2.1 FRONTALIZATION</b> .....	<b>4</b>
<b>2.1.1 2D Frontalization Methods</b> .....	<b>4</b>
<b>2.1.1.1 Piece-Wise Warping</b> .....	<b>5</b>
<b>2.1.1.2 Patch-Wise Warping</b> .....	<b>8</b>
<b>2.1.1.3 Pixel-Wise Displacement</b> .....	<b>9</b>
<b>2.2 LINEAR REGRESSION MODELS</b> .....	<b>11</b>
<b>2.3 NON LINEAR REGRESSION MODELS</b> .....	<b>13</b>
<b>2.4 3D FRONTALIZATION METHODS</b> .....	<b>15</b>
<b>2.4.1 3D Frontalization From Single Image</b> .....	<b>15</b>
<b>2.4.2 3D Frontalization From Multiple Images</b> .....	<b>16</b>
<b>2.5 FACIAL OCCLUSION DETECTION</b> .....	<b>18</b>
<b>3. PROPOSED METHOD</b> .....	<b>23</b>
<b>3.1 SYSTEM OVERVIEW</b> .....	<b>23</b>
<b>3.2 HARD FRONTALIZATION</b> .....	<b>24</b>
<b>3.2.1 Facial Feature Detection</b> .....	<b>25</b>
<b>3.2.2 Facial Pose Estimation</b> .....	<b>26</b>
<b>3.2.3 Frontal Pose Generation</b> .....	<b>28</b>
<b>3.2.4 Visibility Estimation</b> .....	<b>28</b>
<b>3.2.5 Mirroring</b> .....	<b>30</b>
<b>3.3 OCCLUSION ANALYSIS</b> .....	<b>34</b>
<b>3.3.1 Region Based Occlusion Analysis</b> .....	<b>35</b>
<b>3.3.1.1 Learning</b> .....	<b>35</b>
<b>3.3.1.2 Detection</b> .....	<b>42</b>
<b>3.3.2 Pixel Based Occlusion Analysis</b> .....	<b>43</b>

<b>3.4 OCCLUSION BASED SOFT-FRONTALIZATION.....</b>	<b>46</b>
<b>3.4.1 Both Sides Occluded.....</b>	<b>47</b>
<b>3.4.2 Occlusion on Visible Side.....</b>	<b>48</b>
<b>3.4.3 Occlusion on Non-Visible Side.....</b>	<b>48</b>
<b>3.4.4 Un-Occluded on Both Sides.....</b>	<b>50</b>
<b>4. PERFORMANCE EVALUATION.....</b>	<b>51</b>
<b>4.1 Setup.....</b>	<b>51</b>
<b>4.2 Region Based Occlusion Analysis Accuracy.....</b>	<b>51</b>
<b>4.3 Pixel Based Occlusion Reconstruction Results.....</b>	<b>52</b>
<b>4.4 Occlusion Based Soft Frontalization Results.....</b>	<b>53</b>
<b>5. CONCLUSION.....</b>	<b>56</b>



## TABLES

Table 3.1: Pixel Based Occlusion Analysis Results.....	46
Table 4.1: Individual Occlusion State Accuracies.....	51
Table 4.2: Occlusion Correction Results.....	52
Table 4.3: Soft-Frontalization Results Comparison.....	54



Figure 3.5: Hard-Frontalized Face Image.....	28
Figure 3.6: Changes on Visibility Depending on Camera Location.....	29
Figure 3.7: Visibility Map.....	29
Figure 3.8: Side Selection on Visibility Map.....	30
Figure 3.9: Left flank.....	31
Figure 3.10: Emphasized Flank.....	31
Figure 3.11: De-Emphasized Flank.....	32
Figure 3.12: Initial Mirror.....	32
Figure 3.13: Extracted Eye Socket Regions.....	33
Figure 3.14: Initial Mirror without Eye Sockets.....	33
Figure 3.15: Final Mirror.....	34
Figure 3.16: Occlusion Occurrence Percentages in COFW.....	34
Figure 3.17: Extraction Mask.....	36
Figure 3.18: Extraction Result.....	36
Figure 3.19: Detected Facial Landmark Points with Akshay et al.(2013)'s approach.....	37
Figure 3.20: Occlusion Occurrence Regions.....	38
Figure 3.21: Label Image.....	39
Figure 3.22: Occlusion Labeling.....	40
Figure 3.23: Visual Representation of HOG Features .....	41
Figure 3.24: Pixel Neighbourhood Styles.....	41
Figure 3.25: Feature Weights for Every Occlusion Occurrence Region.....	43
Figure 3.26: Sample Region Based Occlusion Analysis Results.....	43
Figure 3.27: Mean Textured Face Image.....	44
Figure 3.28: Intensity Values From Hard-Frontalized Face and Region Values From Mean Textured Face.....	44
Figure 3.29: Occlusion Mask of a Hard-Frontalized image.....	45
Figure 3.30: Process of Occlusion Mask Generation.....	45
Figure 3.31: Frontalization Guidelines.....	47
Figure 3.32: Soft-Frontalization Results of a Face with Both Sides Occluded.....	47

Figure 3.33: Soft-Frontalization Results of a Face with Occlusion on Visible Side .....	48
Figure 3.34: Soft-Frontalization Results of a Face with Occlusion on Non-Visible Side.....	48
Figure 3.35: Pixel Based Occlusion Mask of a Hard-Frontalized Face.....	49
Figure 3.36: Comparison of a Mirrored face and a Corrected Face.....	49
Figure 3.37: Soft-Frontalization Results of a Face with No Occlusions.....	50

## ABBREVIATIONS

ATM	:	Ceritification Authority
CCA	:	Canonical Correlation Analysis
CNN	:	Convolutinal Neural Network
COFW	:	Caltech Occluded Faces Database
GEM	:	Generic Elastic Model
HOG	:	Histogram of Oriented Gradients
LBP	:	Local Binary Pattern
LLR	:	Local Linear Regression
MQVM	:	Multilevel Quadratic Variation Minimization
MVP	:	Multi View Perceptron
RCPR	:	Ceritification Authority
SVM	:	Support Vector Machine

## SYMBOLS

Camera Matrix	:	$C_M$
Facial Feature Points in Reference View	:	$p'_i$
Intrinsic Matrix	:	$A_M$
Landmark Points in Query Image	:	$p_i$
Number of Black Pixels	:	$P_B$
Number of White Pixels	:	$P_W$
Percentage of Black Pixels	:	$P_P$
Reference View	:	$I_R$
Rotation Matrix	:	$R_M$
Translation Matrix	:	$t_M$

## 1. INTRODUCTION

There has been a growing interest in the field of face based computer vision approaches such as facial expression analysis, face recognition and face detection. Computer vision methods working on such fields, rely working in controlled environments most of the time because of reduced pre-processing steps. However, methods that can work on the unconstrained environments are considered more viable, because of their ability to handle extended set of problems. Arguably, facial pose variation is one of the harder issues that exist in the unconstrained domain. The process of turning a posed face into a frontal one is called frontalization.

The frontalization approaches can be investigated under 2D and 3D domains. 2D approaches are simpler in comparison to the 3D methods, and they can produce results without a significant computational cost however, results are not exceptionally realistic. On the other hand, 3D methods shine in generating personalized results with impressive accuracy, however 3D methods have moderately big computational requirements. Frontalization approaches, by themselves, can handle facial pose variations, that is, until they encounter with another issue, also existing in unconstrained domain; occlusions appearing over the face region. They can occur naturally, in any part of the face and are enough to damage the frontalization process, due to the way that frontalization operates.

In this thesis, we propose a method to handle the undesirable effects of occlusions on frontalization operation by analyzing occlusion patterns and choosing the most suitable way to frontalize faces. We first apply a hard-frontalization process on the posed faces occluded in various ways. Hard-frontalized faces are then set to be analyzed for occlusion patterns with two different ways; region and pixel based.

In region based analysis, the aim is coarsely to detect the possible occluded face regions in a hard-frontalized face. After the facial landmark points are extracted, hard-frontalized face is divided into nine meaningful parts, where occlusions are most likely to occur. By using a neutral, reference face, the image data and facial landmark points are aligned to a

single reference face. Each region is represented by the combination of two different features, namely Histogram of Oriented Gradient(HOG) and Local Binary Pattern(LBP) features. Then two Support Vector Machine (SVM) classifiers are trained for each type of feature in every region.

In Pixel Based Occlusion Analysis, we try to establish an Occlusion Mask where the occlusion occurrences are plotted over the face image. In this step, instead of face regions in the previous step, each pixel on the hard-frontalized face is analyzed based on a mean textured reference face. After the aligning both images based on the landmark points, we employ a distance measure to compare two pixels at the same coordinate of the hard-frontalized and reference images. The distance measurement can handle the alignment problems originated from the hard-frontalization stage and the variations in the structures of two faces.

At the final step, depending on the outcome of occlusion analysis, the hard frontalized face with occlusion analysis results is soft-frontalized according to an occlusion map providing the most suitable frontalization solution. The experiments performed on Artizzu et al.(2013)'s Caltech Occluded Faces Database (COFW)show that the proposed method produces satisfactory results in terms of accuracy and visual appearance.

This thesis is organized under 4 main chapters;

### **Chapter 1: Literature Survey**

In this chapter, we take a survey on the literature, expanding our knowledge on Frontalization and Occlusion detection areas to build a better understanding.

### **Chapter 2: Proposed Method**

In this chapter, the overview of the proposed approach is presented in an understandable fashion under three main headlines;



**i. Hard Frontalization**

Hard-frontalization process is the very backbone of our approach, by back-projecting the appearance of the query photo to the reference coordinate system hard-frontalization process turns posed faces into frontal ones.

**ii. Occlusion Analysis**

Our main contribution in this thesis, born from the lack of self-awareness in the Frontalization approaches. Instead of simply accepting the hard-frontalized face as problem-free, we analyze occlusion patterns on it. Then, we pass the output of the Occlusion Analysis to the soft-frontalization step, to be processed.

**iii. Occlusion Based Soft Frontalization**

Soft-Frontalization is a visual improvement, more importantly a correction, acquired by modifying the hard-frontalized output. It is accomplished by selecting the best outcome for the Frontalization process, according to the Occlusion Analysis results.

**Chapter 3: Performance Evaluation**

In this chapter, we evaluate our performance and show results, both accuracy and visual appeal wise.

**Chapter 4: Conclusion**

In this chapter, we go over our ideas once again, explain what we did, and what are we planning to do in the future to improve the Frontalization process.

## 2. LITERATURE SURVEY

In this chapter, a survey on Frontalization and Occlusion Detection approaches are conducted to show the importance these approaches as well as the different types of them.

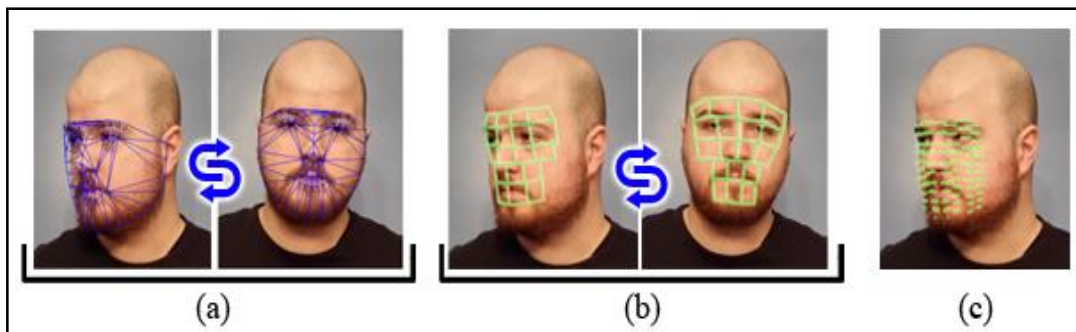
### 2.1 FRONTALIZATION

Facial pose Frontalization is a crucial, yet often ignored pre-process step, where a posed face is transformed to become one with a frontal view. Recent generation of face based approaches e.g. Face Recognition, Facial Expression Analysis, started to include the Frontalization process as a pre-processing step, because of the shift from controlled environments to the unconstrained ones. Whether, it is a Face recognition method in a social media website which tags faces in photos, or a sophisticated forensics application, which is capable of doing a cross-check on a database of known criminals with a frontalized face photo, the results will be significantly more accurate when faces are frontalized beforehand. Frontalization based methods can be classified into four major groups according to the techniques environment; 2D, Linear Regression, Non-Linear Regression and 3D.

#### 2.1.1 2D Frontalization Methods

2D Frontalization methods can be divided into three main groups ; Piece-wise warping based, Patch-wise warping based, Pixel-wise warping based (**Fig. 2.1**);

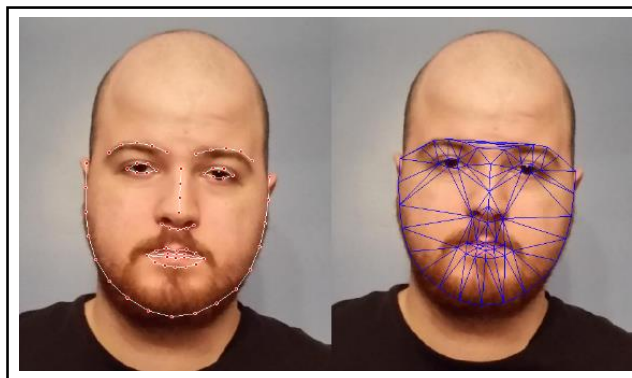
**Figure 2.1: (a) Piece-wise warping, (b) Patch-wise warping, (c) Pixel-wise displacement**



### 2.1.1.1 Piece-Wise Warping

The piecewise warping approaches transform the shape of the face image in a piecewise manner to another specified pose. Piece in the piece-wise based methods refers to a single triangle in the triangular mesh, generated by Delaunay triangulation of the landmark points (**Fig. 2.2**). Piece-wise based methods aim to warp the target image mesh to the original mesh using geometric transformations.

**Figure 2.2: Delaunay Triangulation of facial landmark points**



The aforementioned transformation can be affine warping(**Fig. 2.3**) (Gao et al 2009, pp. 2876-2896) or a thin-plate splines-based warping (Bookstein 1989, pp. 567-585).

**Figure 2.3: Affine warp of face image under different poses**

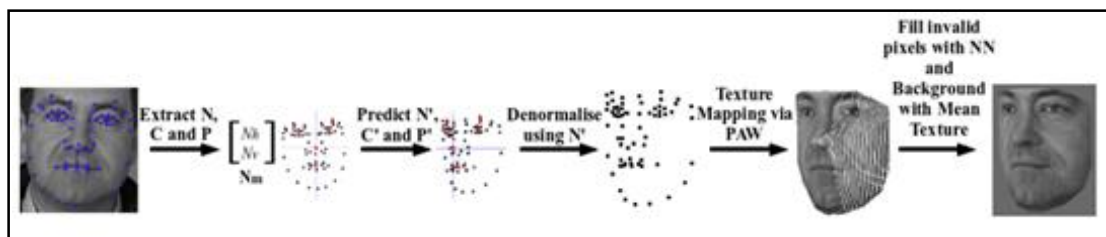


*Reference:* Gao et al., (2009) Pose Normalization for Local Appearance-Based Face Recognition

There are multiple Piece-wise based works that aim to convert face images with varied poses to a frontal view with neutral facial expression. Asthana et al.(2009) proposed the Gaussian Process Regression model to learn the correspondence between facial landmark

points in the target and frontal image. In the testing phase, given a frontal face image and its facial landmark points, the feature coordinates of non-frontal faces to be synthesized are predicted by Gaussian Process Regression method (**Fig. 2.4**), followed by warping the texture of the frontal image. In this way, frontal face dataset is expanded with synthesized non-frontal faces and a probe image can be compared with images in similar pose. However, as a downside, enlarging the gallery set reduces the efficiency of face recognition systems.

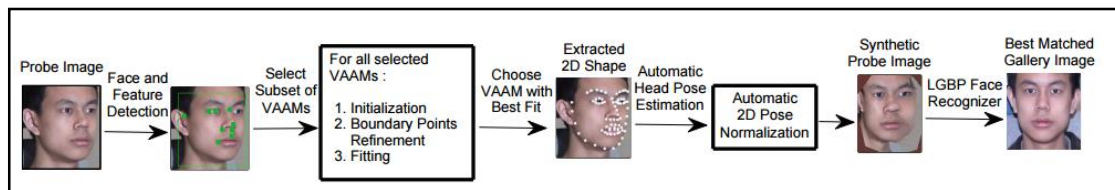
**Figure 2.4: GPR method overview**



*Reference:* Asthana et al., (2009) Region-based face synthesis for pose-robust recognition from single image

In another work Asthana et al.(2011), trained GPR to predict the facial landmark points of the virtual frontal pose from the landmark points of a non-frontal face (**Fig. 2.5**). With this way, all of the images are transformed to the frontal pose and are compared without expanding the dataset.

**Figure 2.5: GPR based Pose Normalization**



*Reference:* Asthana et al., (2011) Pose Normalization via Learned 2D Warping for Fully Automatic Face Recognition

Taijman et al.(2014) proposed the inference of the facial landmark point locations in the virtual frontal pose for each non-frontal face with the help of a generic 3D face model. First, the 3D face model is rotated to the pose of the 2D image by aligning the facial landmark points of both the model and the image. The residuals between the landmarks

of the 2D image and the projected locations of the 3D landmarks are then added to the uniform shape in the frontal pose as compensation which is assumed to reduce the distortion to the identity caused by pose normalization. In spite of the usage of 3D face model, pose normalization still happens in the 2D image domain.

Berg and Belhumeur(2012) proposed another method of saving the identity information. For each non-frontal image, triangulation is performed(**Fig. 2.6**) on generic landmark locations of all faces in this pose rather than the image's own facial landmarks. The image is then transformed to the uniform shape of the frontal pose by piecewise affine warping. In this way, a smaller distortion of the identity of the test image is achieved. In the experiment on Labeled Faces in the Wild dataset, the authors reported better performance with the identity-preserving warping strategy rather than the uniform warping method.

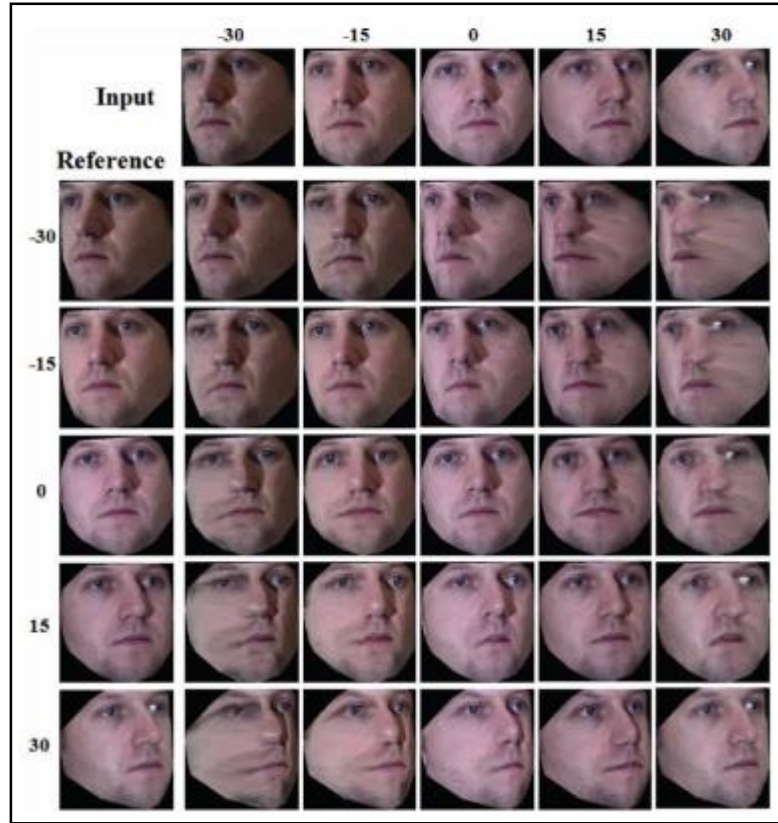
**Figure 2.6: Triangulation of inner and outer facial landmark points**



*Reference:* Berg and Belhumeur, (2012) Tom-vs-pete classifiers and identity-preserving alignment for face verification

Piece-wise warping methods are simple to use, which makes them prone to limited efficiency. Firstly, pose of the output cannot deviate too far from the original image. Heo and Savvides(2012) investigated the ability of this method to handle yaw angle variation. The conclusion they gave was that if the yaw difference exceeds  $15^\circ$ , the resulting warped outputs generate several stretchmarks, caused by over-sampling of the textures in relatively small regions (**Fig. 2.7**).

**Figure 2.7: Frontalization under variable pose angles**



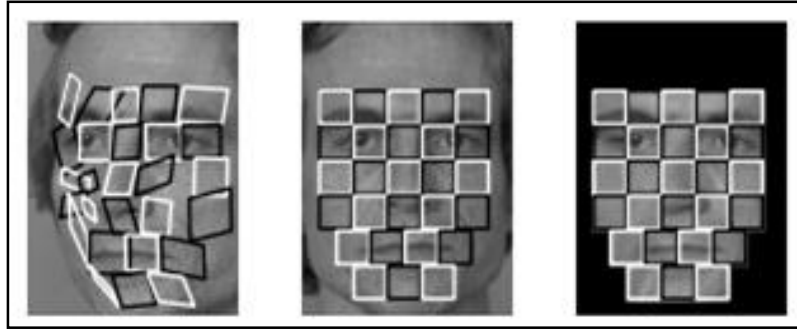
*Reference:* Heo and Savvides, (2012) 3-d generic elastic models for fast and texture preserving 2-d novel pose synthesis

Second, quality of the output images is very dependent on the accurate placement of facial landmark points, as the warping is determined only by the facial landmark points.

### 2.1.1.2 Patch-Wise Warping

Patch-wise Warping methods are known for their ability to cover a wider variety of poses, which is was one of the main problems with the Piece-wise warping methods. Ashraf et al.(2008) modeled the face image as a collection of patches and accomplished the reconstruction of the face image using a Patch-wise strategy, called "stack-flow". Proposed method synthesized frontal patches non-frontal pose patches. In the training phase, method learns the optimal affine warp using the Lucas-Kanade algorithm which aligns patches in the stack of the non-frontal faces to corresponding patches in the stack of frontal faces (**Fig. 2.8**). In the testing phase, each patch is warped to the frontal pose with the pre-learnt warp in previous step, which finally generates a frontal pose image.

**Figure 2.8: Warp Region With Stack-Flow Algorithm**



*Reference:* Ashraf et al., (2008) Region patch correspondences for improved viewpoint invariant face recognition

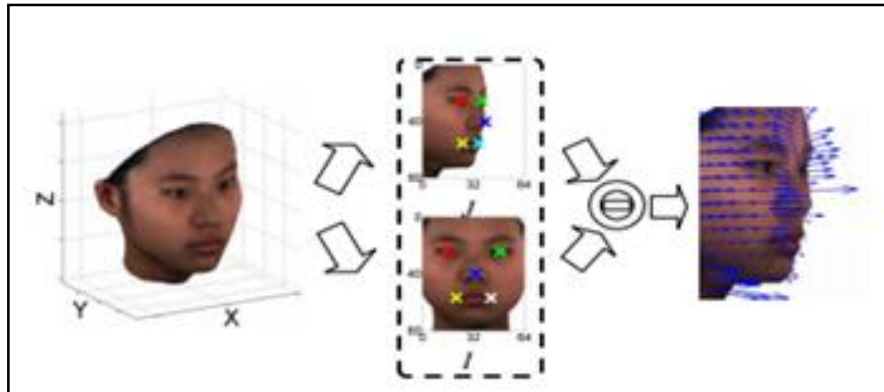
To learn the affine warp between the half-profile patches and the frontal patches, the authors also proposed a composite strategy that successively learns the warp parameters between intermediate poses. Due to the flexibility of the stack-flow approach, it is reported that it can cover a wider range of pose variations than piecewise warping approaches.

### **2.1.1.3 Pixel-Wise Displacement**

While patch and piece wise methods are definitely useful, they can't handle the local non-linear warps that appear in each piece or patch. Beymer and Poggio (1995) proposed the parallel deformation approach which predicts the pixel-wise displacement between two poses. First and the most challenging part is, establishing the dense pixel-wise semantic correspondence between images of different poses and subjects using the optical flow based method. The displacement fields containing dense pixel-wise displacement between two poses from a set of training subjects are recorded as the template displacement fields. Given a face image, the displacement field can be estimated by a linear combination of the template displacement fields. With the estimated displacement field, the face image can be deformed to a image under another pose. Li et al.(2012, 2014) proposed the generation of the template displacement fields using images synthesized by a set of 3D face models. The pixel-wise correspondence between the synthesized images can be easily inferred via the 3D model vertices (**Fig. 2.9**),



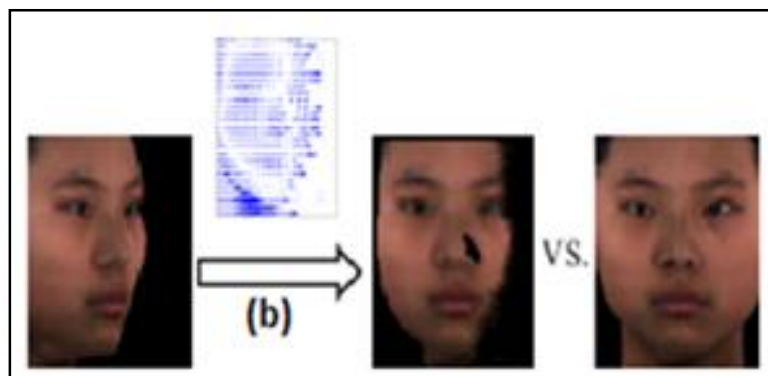
**Figure 2.9: Template Displacement Field**



*Reference:* Li et al., (2012) Morphable displacement field based image matching for face recognition across pose

therefore this approach implicitly utilizes 3D facial shape priors for pose normalization. The authors also proposed the implicit Morphable Displacement Field method (Li et al. 2012, pp. 102-115) and the Maximal Likelihood Correspondence Estimation method Li et al 2014, pp. 4587-4600) to effectively estimate the convex combination of the template displacement fields where the 3D models are discarded and only the template displacement fields are utilized for face synthesis (**Fig. 2.10**).

**Figure 2.10: Synthesized Face With Template Displacement Field**



*Reference:* Li et al., (2014) Maximal likelihood correspondence estimation for face recognition across pose



## 2.2 LINEAR REGRESSION MODELS

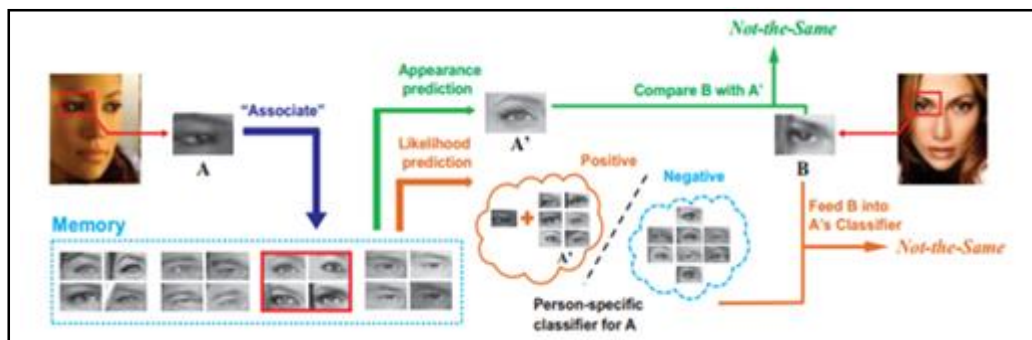
The earliest author that used face synthesis as a base for linear regression problem was by Beymer and Poggio(1995), under the assumption of orthogonal projection and a condition of constant illumination, the holistic face image is represented as a linear combination of training faces in the same pose, and the same combination coefficients are employed for face synthesis under another pose. However, as aforementioned, the approach requires dense pixel-wise correspondence between face images, which is very hard to achieve in practice. Later works conducted face synthesis in a patch-wise strategy (Chai et al. 2007, pp. 1716-1725), reducing the difficulty in alignment. Another advantage of the patch-based methods is that each patch can be regarded as a simple planar surface, thus the transformation of the patches across poses can be approximated by linear regression models.

Inspired by Beymer and Poggio (1995), Chai et al (2007) proposed Local Linear Regression(LLR) method for face synthesis. LLR works on the patch level, based on the key assumption that the manifold structure of a local patch stays the same across poses. Formally, suppose there exist two training matrices whose columns are composed of the vectorized local patches in the frontal pose and non-frontal pose, respectively. Note that the corresponding patches are of the same subject. In the testing time, given an image patch in a pose, the first step of LLR is to predict image patch from the linear combination of columns in, where the combination coefficients are computed by the least square algorithm. Then frontal patch is synthesized. By repeating the previous step for all patches in test image, the final virtual frontal pose image can be generated by overlapping all the predicted frontal patches. In spite of its simplicity, LLR approach suffers from the overfitting problem, which makes the manifold structure learnt in a pose may not faithfully represent the structure in another pose.

To relieve this problem, several improved approaches have been proposed, these solutions include the lasso regularization (Annan et al. 2012, pp. 305-315), (Zhang et al. 2013, pp. 1511-1521), local similarity regularization (Hao and Qi 2015, pp. 559-563), and neighborhood consistency regularization (Hao and Qi 2015, pp. 559-563).

Based on a similar assumption, Yin et al.(2011) proposed the Associate-Predict model (APM) for frontal face synthesis. The APM model has two steps for estimating the frontal pose patch from a non-frontal pose patch. In the associate step, non-frontal pose patch is associated with the most similar patch. In the “Predict” step, the associated patch's corresponding patch in is directly utilized as a prediction for the frontal pose patch, and the predicted frontal patch is employed for the purpose of face matching (Fig. 2.11).

**Figure 2.11:Associate-Predict Model**



Reference: Yin et al., (2011) An associate-predict model for face recognition

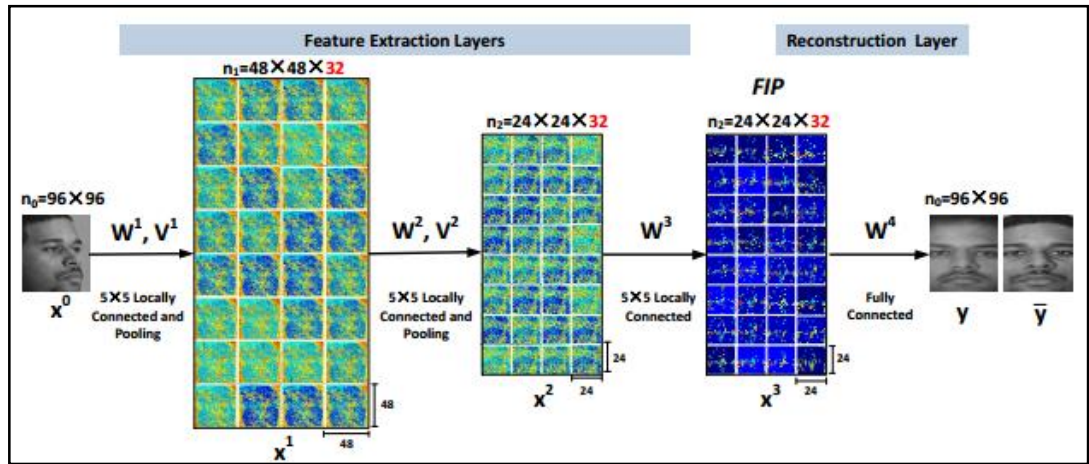
Unlike many least square regression-based methods for face synthesis, Li et al (2009) proposed the formulation of Canonical Correlation Analysis(CCA) as a regressor for frontal face reconstruction. In the training phase, CCA is employed to build the correlation-maximized subspace shared by the frontal and non-frontal patches. For frontal face synthesis, the non-frontal patches are first projected into the correlation-maximized subspace. Then, ridge regression is employed to regress them into the frontal face space. Overlapping all the synthesized frontal patches forms the required virtual frontal face.

As Li et al.(2009)'s approach is not based on such delicate assumption as that in Chai et al.(2007), they report both higher synthesized image quality and higher recognition performance on the synthesized faces than Chai et al.(2007). In summary, linear regression-based methods are simple, they typically require a certain amount of multi-pose training data. However, regressed face images also suffer from the blurring effect and lose critical fine textures for recognition.

### 2.3 NON LINEAR REGRESSION MODELS

The appearance variation of the face images across poses is intrinsically nonlinear, as a result of the substantial occlusion and nonlinear warp. To synthesize face images of higher quality, nonlinear regression models have recently been introduced. Zhu et al.(2013) and Yim et al.(2015) adopted Convolutional Neural Networks(CNN) to extract pose-robust features, to synthesize faces with frontal pose(Fig. 2.12).

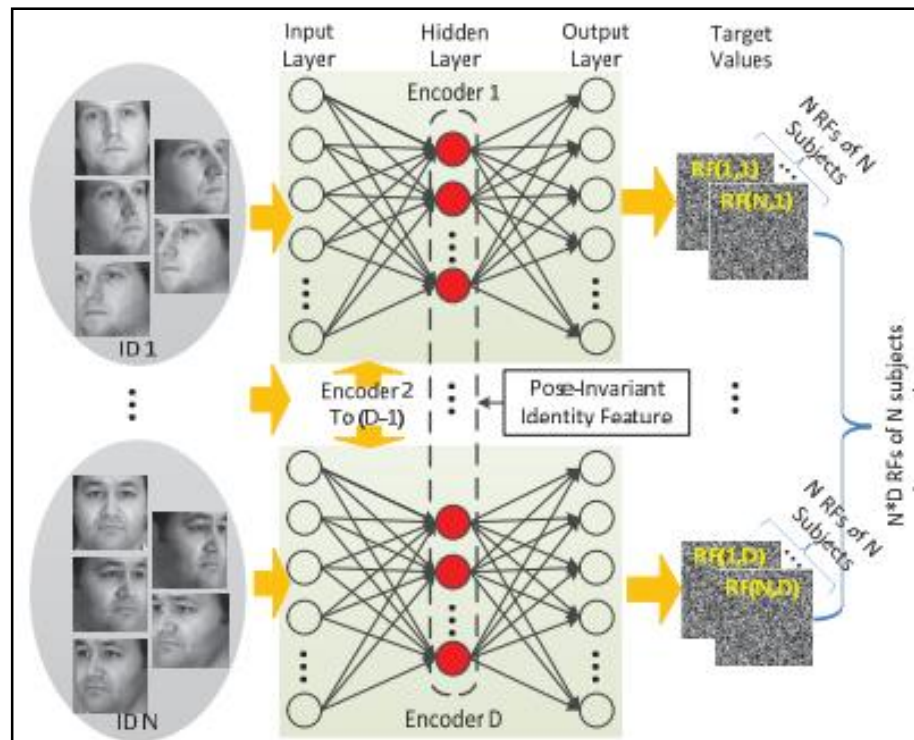
**Figure 2.12: Feature Extraction and Reconstruction Layers**



Reference: Zhu et al., (2013) Deep region identity-preserving face space

Zhang et al.(2013a) employed single-hidden-layer auto-encoders for frontal posed face synthesis(Fig. 2.13).

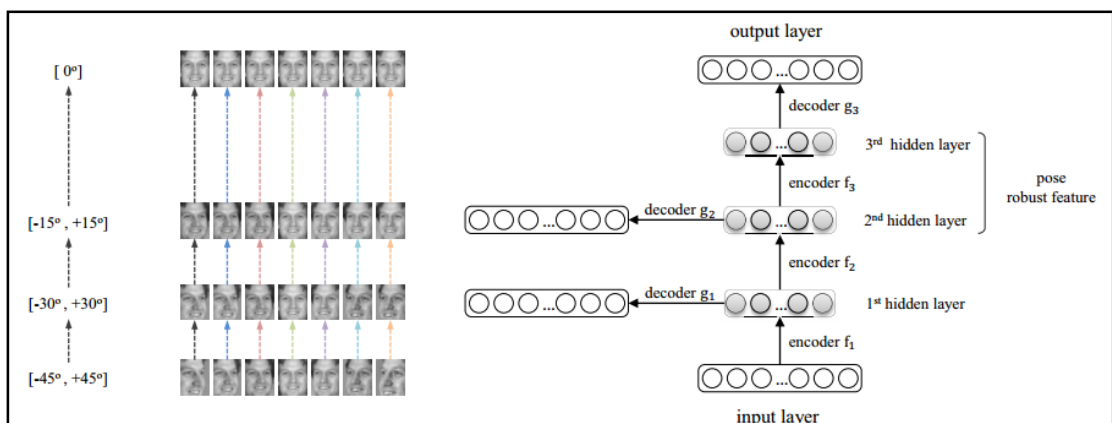
**Figure 2.13: Single Hidden Layer Autoencoder Network**



*Reference:* Zhang et al., (2013a) Random Faces Guided Sparse Many-to-One Encoder for Pose-Invariant Face Recognition

Kan et al.(2014) utilized stacked auto-encoders to recover the frontal face image from non-frontal input faces in a progressive manner, thus reducing the difficulty of face synthesis for each auto-encoder (**Fig. 2.14**).

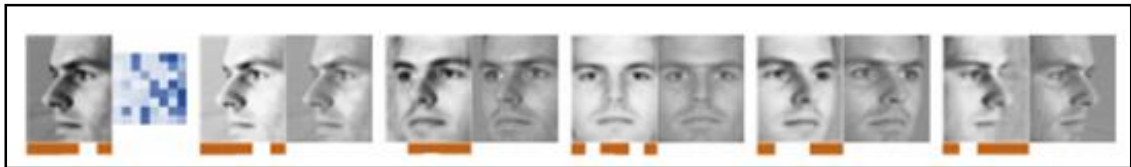
**Figure 2.14: Progressive Reduction of Pose Variations**



*Reference:* Kan et al., (2014) Stacked progressive auto-encoders (spae) for face recognition across poses

Because of their deep structure, Zhu et al.(2013), Kan et al.(2014), Yim et al.(2015) generated synthesized faces with high quality. The common limitation of Zhu et al (2013), Zhang et al.(2013a) and Kan et al.(2014) is that they are all deterministic networks that recover face images of a fixed pose. In comparison Zhu et al.(2014), Yim et al.(2015) designed deep neural networks that can synthesize face images of varied poses. For example, Zhu et al.(2014) proposed the Multi-View Perceptron(MVP) approach. Given a single input 2D face, approach was able to generate wide range of pose-varied face images from the same identity (**Fig. 2.15**).

**Figure 2.15: Synthesized posed images of the same figure**



*Reference: Zhu et al., (2014) Multi-view perceptron: A deep model for region face identity and view representations*

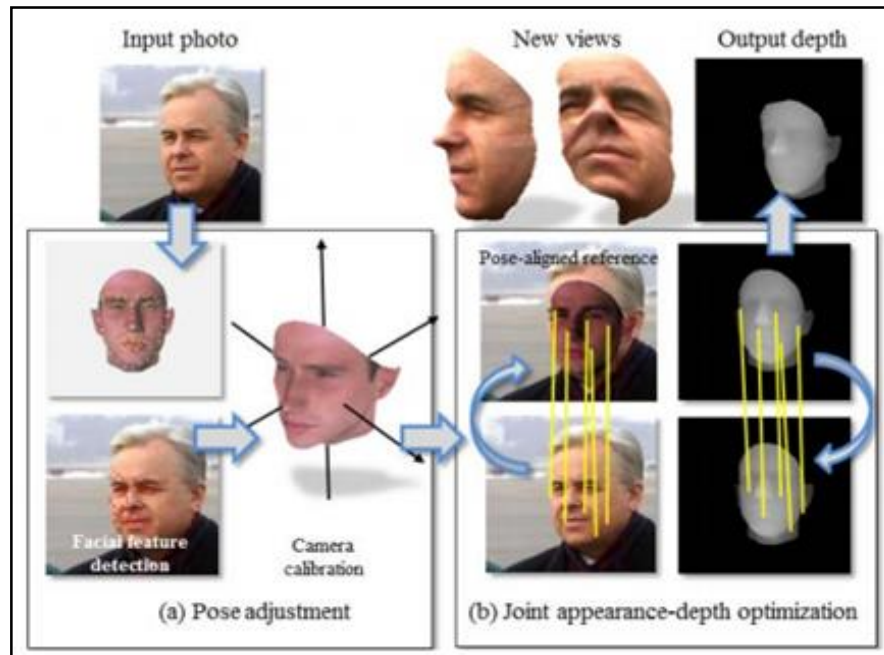
## 2.4 3D FRONTALIZATION METHODS

The human head is a complex non-planar 3D structure rotating in 3D space while the face image lies in the 2D domain. The lack of one degree of freedom makes it difficult to conduct face synthesis using 2D techniques alone. 3D based methods build a 3D model of the human head to conduct frontal posed face synthesis. The 3D methods can be classified into two subcategories; 3D Frontalization from single image and multiple images.

### 2.4.1 3D Frontalization From Single Image

The 3D pose normalization approaches employ the 3D facial shape model as a tool to correct the nonlinear warp of facial textures appearing in the 2D images. The general principle is that, the 2D face image is first aligned with a 3D face model, typically with the help of facial landmark(Hassner 2013, pp. 3607-3614). Then, the texture of the 2D image is mapped to the 3D model. Lastly, the textured 3D model is rotated to a desired pose and a new 2D image in that pose is rendered (**Fig. 2.16**).

**Figure 2.16: 3D Face Frontalization**



*Reference:* Hassner, (2013) Viewing real-world faces in 3d.

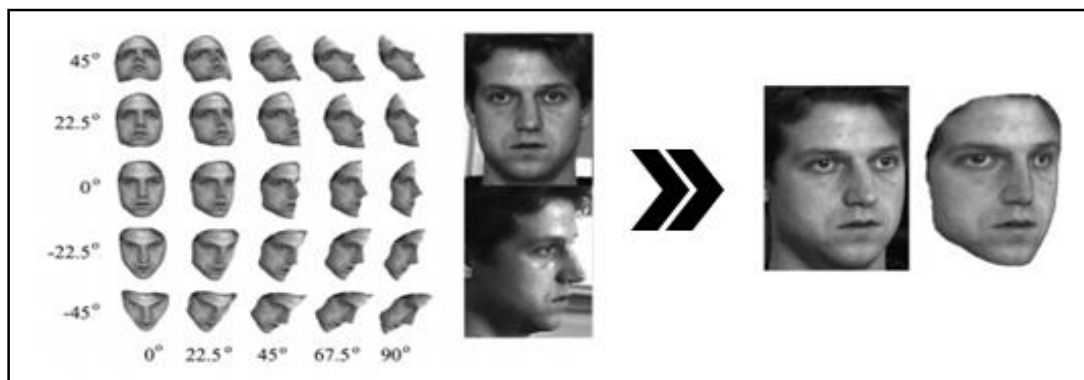
Early approaches utilized simple 3D models, like the cylinder model (Gao et al 2001, pp. 248-253), the wire frame model (Lee and Ranganath 2003, pp. 1835-1846), (Zhang et al 2006, pp. 1151-1154), and the ellipsoid model (Liu and Chen 2005, pp. 502-509), to roughly model the 3D structure of the human head, whereas newer approaches strive to build accurate 3D facial shape models, most notable examples being Hassner et al. (2015) and Zhu et al.(2015). Hassner et al.(2015), frontalizes the face by finding the correspondence between 2D facial landmark points in query image and 3D facial landmarks in the 3D model. With the projection matrix acquired from the correspondence the pixels in the query image are back projected to generate a frontalized face. Zhu et al.(2015) took the frontalization with a different approach, by using a morphable face model(Pascal et al. 2009, pp. 296-301) instead of a generic one.

#### **2.4.2 3D Frontalization From Multiple Images**

3D modeling from a single input image, is the most common setting for real-life face recognition applications, yet there is a shortcoming; the personalized 3D shape parameters cannot be precisely approximated, since 3D modeling from a single face image is essentially an ill-posed problem. However, in some areas like law enforcement, multi-

view images are available for each subject during enrollment(Zhang et al. 2008, pp. 684-697). In that case, it is desirable to utilize the multiple images together to build a highly accurate 3D face model. Zhang et al (2008) proposed the Multilevel Quadratic Variation Minimization (MQVM) approach to reconstruct an accurate 3D structure from a pair of frontal and profile face images of a subject (**Fig 2.17**).

**Figure 2.17: 3D face structure construction**

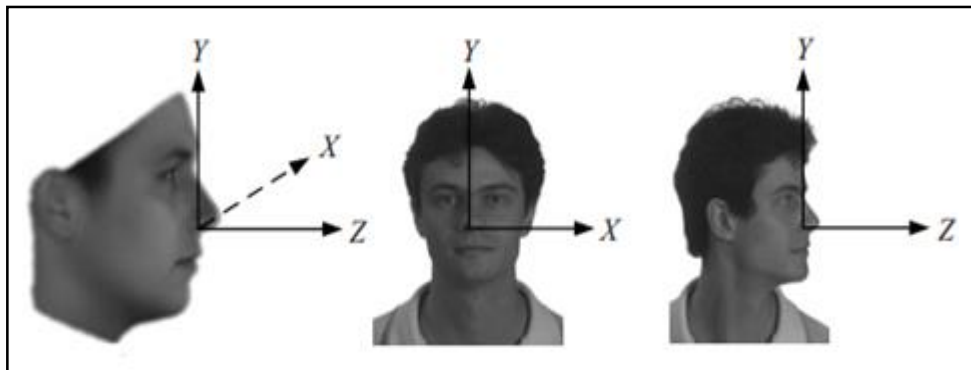


*Reference:* Zhang et al., (2008) Recognizing rotated faces from frontal and side views: An approach toward effective use of mugshot databases

Heo and Savvides(2011) employed the depth information in the profile face image to modify the generic depth map in Generic Elastic Model(GEM). In brief, a sparse 3D face shape is constructed by aligning the facial landmarks of the frontal and profile face images. The depth map of the sparse face shape is merged with the generic depth map. The combined depth map replaces the generic depth map in GEM for face synthesis. It is argued that the combined depth map is more accurate and thus more realistic face images can be synthesized. However, there are no experiments to validate the effectiveness of the proposed approach for face recognition. Similarly, Han and Jain (2012) extended the GEM approach to utilize the complementary information incorporated in the frontal and profile image pair. In the proposed approach 2, 3D models are derived, one is from the frontal face and the other is from the profile face (**Fig. 2.18**).



**Figure 2.18: Frontal face information in X-Y plane and Profile Face Information in Y-Z plane**



*Reference: Han and Jain, (2012) 3d face texture modeling from uncalibrated frontal and profile images*

frontal face model, preserves the 2D shape information of the frontal face, while profile face model depicts more accurate depth information. The information in frontal and profile face models are complementary since they contain accurate 2D shape information and 3D depth information, respectively. The final 3D shape model for the subject is obtained by replacing the depth value of 3D vertices. The 3D model is textured by mapping the pixel value in the frontal face image to the 3D model.

## **2.5 FACIAL OCCLUSION DETECTION**

Facial occlusion detection is less known, compared the Frontalization process itself, although this doesn't make it any less important as it gains a whole lot more importance when its used in the context of the Frontalization, reason being that, it enhances the Frontalization by giving it self awareness it lacks when used correctly. Due to the fact that facial occlusion detection methods are a sub-genre of occlusion detection, an unexplored area, selected approaches in this part will be examined individually, rather than in groups.

Lin and Liu(2006), inspired by the automatic alarm systems in Automated Teller Machines(ATM) and the part of human error plays in manual video surveillance systems proposed an approach, focused on effective detection of facial occlusion(**Fig. 2.19**) to assist security personnel in surveillance by providing both valuable information for further video indexing applications and important clues for investigating a crime.



**Figure 2.19: Occluded faces of criminals**

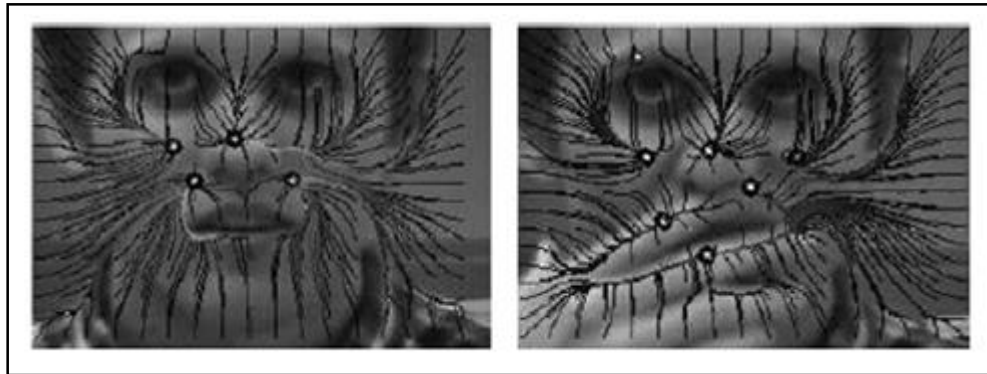


*Reference: Lin and Liu, (2006) Face Occlusion Detection for Automated Teller Machine Surveillance*

Firstly, a multiple of methods focused on identifying and segmenting moving objects are formed. Then, moving edges are captured using change detection of the interframe difference and the Sobel operator. Next, a Straight Line Fitting algorithm is developed to merge the splitting blobs. Additionally, a mechanism involving moving forward or backward justification is used to determine whether an individual is approaching a camera. Moreover, the lower boundary of a head is computed, followed by use of an elliptical head tracker to match the head region. Finally, skin area ratio is calculated to determine whether the face is occluded or not.

Smith et al.(2006) focused on the the specific issue of a hand occluding over the face. In their approach, they presented a method to segment the hand over the face where the color and texture indifference between the hand and face make the approach particularly interesting. The approach is built on the underlying concept of an “image force field”, where change over the face is measured through how particles move through the field. Using this assumption, the occluders are detected when partices move around the area **(Fig. 2.20)**.

**Figure 2.20: Image force field changes in case of an occluder**



*Reference:* Smith et al., (2006) Resolving hand over face occlusion

Based on the principles of liberating the user from marking the occlusion area, Lin and Tang(2007) proposed an approach for accurate occlusion detection and recovery. Firstly a Bayesian formulation unifying the occlusion detection and recovery stages is derived. Then a quality assessment model is established for detection and recovery phases. Built on aforementioned formulation, GraphCut-based Detection and Confidence-Oriented Sampling approaches are proposed to attain optimal detection and recovery respectively (Fig. 2.21).

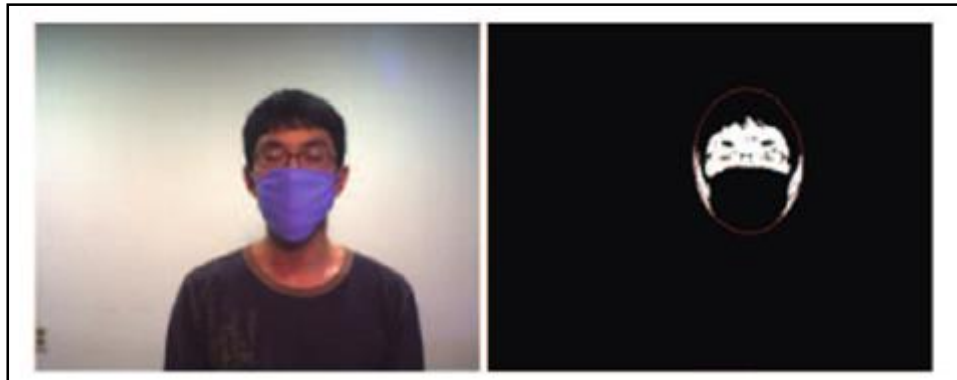
**Figure 2.21: Facial recovery from different occluders**



*Reference:* Lin and Tang, (2007) Quality-Driven Face Occlusion Detection and Recovery

Kim et al.(2010) proposed an occlusion detection approach fusing shape-based face detection and skin color area ratio based occlusion detection. First, frame difference image is established with an appropriate time interval calculated by kurtosis-based frame selection. Then, moving edges are detected by combining a distance transformation of frame difference image and edge detection result of current frame. Exact face area is calculated, by fitting an ellipse to points on head shape of B-spline active contour estimated in the first phase. Finally, skin color area ratio is calculated focusing on face area, eye area and mouth area, and then face occlusion detection is performed by using the aforementioned ratios (**Fig. 2.22**).

**Figure 2.22: Occlusion map of a masked face**



*Reference:* Kim et al., (2010) Face Occlusion Detection by using B-spline Active Contour and Skin Color Information

Artizzu et al.(2013), shaped their approach, aiming to improve Cascaded Pose Regression approach where it lacks, by adding it ability to detect occlusions. Cascaded Pose Regression approach is capable of approximating shapes using any parametrized variation of the object's appearance, which made it a particularly effective way to use when estimating facial landmark points landmarks. As their first contribution, they proposed Robust Cascaded Pose Regression (RCPR), which robust to bad initializations and occlusions. Additionally, RCPR is the first landmark based occlusion detection approach (**Fig. 2.23**). Finally, they introduced of a dataset; Caltech Occluded Faces in the Wild which includes heavy occlusions and large shape variations.

**Figure 2.23: 25 Landmark based Occlusion detection**



*Reference: Artizzu et al., (2013) Robust face landmark estimation under occlusion*

Ghiasi and Fowles(2014), built their approach on the basis that, occluder should not be inferred from the lack of object features, but positive evidence for the occluding object that explains away the lack of object features. They created a hierarchical deformable part model, which divides face into multiple meaningful parts, specifically adjusted to localize landmarks and detect possible occlusions happening over them (**Fig. 2.24**).

**Figure 2.24: 68 Landmark based Occlusion Detection**



*Reference: Ghiasi and Fowles, (2014) Occlusion Coherence: Localizing Occluded Faces with a Hierarchical Deformable Part Model*

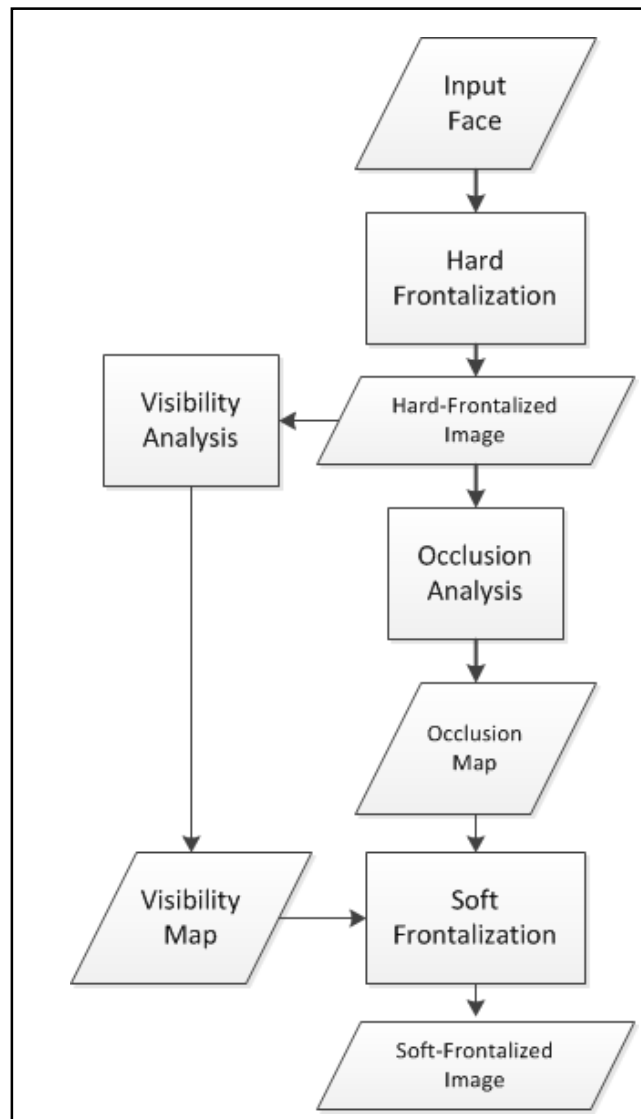
### 3. PROPOSED METHOD

In this chapter, we explain the main pillars of our approach, Hard-Frontalization, Occlusion Analysis, and Soft-Frontalization. Each step is explained in detail, while backed with distinctive visual illustrations.

#### 3.1 SYSTEM OVERVIEW

Before going into detail, the workflow of our whole system can be examined in **Figure 3.1**.

**Figure 3.1: System Overview**



First, the input face gets fed into the Hard-Frontalization process, which outputs the hard-frontalized Image. The hard-frontalized Image becomes, the input for two different operations, Visibility Analysis and Occlusion Analysis. In both sides, the hard-frontalized image gets processed in different ways to create the baseline for the Soft-Frontalization process. Visibility Analysis, outputs the Visibility map, showing regions over the face where visibility is poor pixel-wise. Occlusion Analysis outputs Occlusion Map, in two different forms, one established by a region based approach, the other established by a pixel based one. Finally, both Visibility and Occlusions maps are fused, to establish a guideline, to result in the most suitable soft-frontalized image.

We explain our methodology under three main headlines;

**i. Hard-Frontalization**

In here, we explain the back-bone of our approach, supplied by Hassner et al.(2015)'s novel 3D model based approach, to generate the foundations for Occlusion Analysis step to work on.

**ii. Occlusion Analysis**

In here, we take the hard-frontalized face Image, and process it under our region and pixel based Occlusion Analysis approaches, to guide Soft-Frontalization.

**iii. Soft-Frontalization**

In here, we form a map, to finalize the Frontalization process, then we shape the final result, depending on the most suitable way that the guideline provides.

### **3.2 HARD FRONTALIZATION**

Hard-Frontalization is, the process of back-projecting the appearance of the query photo to the reference coordinate system using the 3D surface as a proxy. This thesis has chosen Hassner et al.(2015)'s novel work on Frontalization, based on using a non-personalized, pre-established 3D model to approximate the matrix, which was used to capture query photo. With the approximated camera matrix, the query posed face can be re-shaped into a frontal one, using bilinear interpolation.

Hassner et al.(2015).’s Hard-Frontalization approach is explained under 5 steps;

**i. Facial Feature Detection**

In this step, facial landmark points on the query face is located, using a state-of-the-art facial feature detection approach.

**ii. Facial Pose Estimation**

In this step, camera matrix used to capture the query image is approximated, using the correspondences between the 2D query face and the 3D model.

**iii. Frontal Pose Generation**

In this step, pixels of the query image are back projected, creating the frontal posed face image.

**iv. Visibility Estimation**

In this step, visibility of the hard-frontalized face will be estimated, by counting the overlapping projected pixels to create, a “Visibility Map”.

**v. Mirroring**

In this step, with the help of the Visibility Map acquired from the Visibility Estimation step, hard-frontalized face will be mirrored with the more visible side, for outcome of a more visually satisfying result.

### **3.2.1 Facial Feature Detection**

Hassner et al.(2015)’s approach is compatible with different options for the Face Detection and Facial Landmark Point Localization, in this thesis, Ramanan et al.(2012)’s Deformable Part Model based approach with 68 landmarks is chosen because of its ability to detect faces and facial landmark points under challenging poses (**Fig. 3.2**).

**Figure 3.2: Face and Landmark Detections of Zhu and Ramanan(2012)**



*Reference: Zhu and Ramanan, (2012) Face Detection, Pose Estimation and Landmark Localization in the Wild*

### 3.2.2 Facial Pose Estimation

Before the pose estimation process, with the base 3D model, a synthesized reference face (Fig. 3.3) to use as the base of Frontalization process needs to be established. The reference face will guide the operation as a beacon for the frontal pose.

**Figure 3.3: Reference Face**



First, a reference projection matrix (Eq. 3.1) is specified;

$$C_M = A_M [R_M t_M] \tag{3.1}$$



Where,  $A_M$  is the intrinsic matrix, and  $[R_M t_M]$  are the extrinsic matrix parameters, consisting of rotation matrix  $R_M$  and translation vector  $t_M$ . Rotation and translation parameters are pre-defined, to produce a frontal view of the model  $I_R$ , which, in the end serves as the frontal posed reference coordinate system.

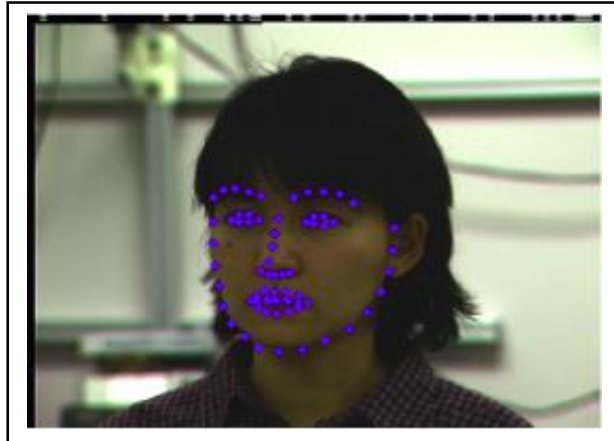
When reference view,  $I_R$  is generated its pixels  $p'$ , are stored in the 3D point coordinate space, the process is explained in **Equation 3.2**, where  $P = (X, Y, Z)^T$  represents the 3D points of pixels.

$$p' \sim C_M P \quad (3.2)$$

Let facial landmark points detected in the query image be,  $p_i = (x_i, y_i)^T$  and let  $p'_i = (x'_i, y'_i)^T$  be the facial feature points detected in reference view.

From the **Equation 3.1** we have the coordinates  $P_i = (X_i, Y_i, Z_i)^T$  of the point surface on the model, projected onto  $p'$  (**Fig. 3.2**).

**Figure 3.4: Projected facial landmark points**



This provides the correspondence  $(p_i^T, P_i^T) = (x_i, y_i, X_i, Y_i, Z_i)$ , which allows estimation of the projection matrix on query matrix with **Equation 3.3**;

$$M_Q = A_Q [R_Q t_Q] \quad (3.3)$$

### 3.2.3 Frontal Pose Generation

The initial hard-frontalized face image is produced by back-projecting query facial features back onto the reference coordinate system using the reference 3D model.

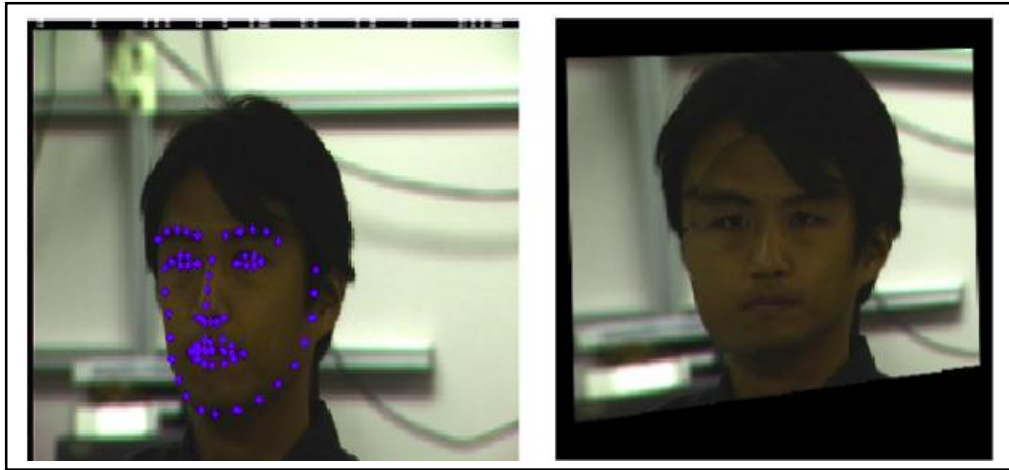
For every pixel coordinate  $q' = (x', y')^T$  in the 2D reference image, from **Equation 3.3** we have the 3D location  $P = (X, Y, Z)^T$  on the surface of the reference image which was projected onto  $q_0$  by projection matrix  $C_M$ .

**Equation 3.4** is used to estimate location  $p = (x, y)^T$  in  $I_Q$  for the same facial feature;

$$p \sim C_Q P \quad (3.4)$$

With the help of bi-linear interpolation intensities of the query photo are sampled, and then, the sampled colors are assigned to the pixel points in the new frontalized image to produce the hard-frontalized Result (**Fig. 3.5**).

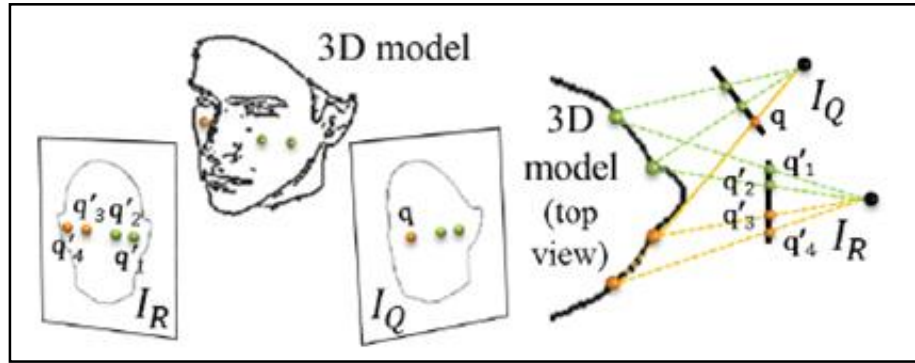
**Figure 3.5: Hard-Frontalized Face Image**



### 3.2.4 Visibility Estimation

By counting the number of times query pixels are accessed while generating the hard-frontalized image, visibility is estimated. As the face rotates away from the camera, the angle between its less visible features and the camera plane increases, it consequently increases the number of surface points projected onto the same pixel in the photo (**Fig. 3.6**).

**Figure 3.6: Changes on Visibility Depending on Camera Location**



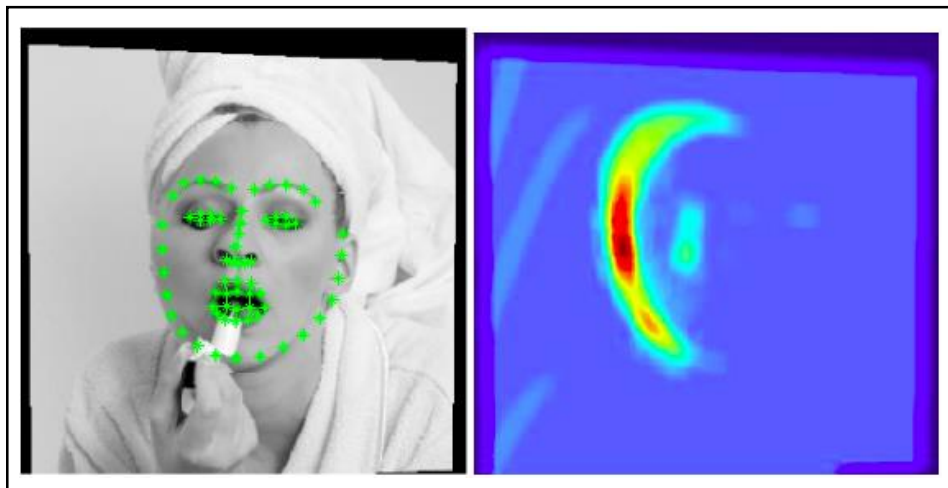
Reference: Hassner et al., (2015) Effective face frontalization in unconstrained images

In **Equation 3.4**, for each pixel  $q_0$  in the reference view  $I_R$ , location in the query photo of its corresponding pixel  $q$  is stored. Visibility score is determined for each pixel in the hard-frontalized image with the help of **Equation 3.5**.

$$v(q') = 1 - \exp(-\#q) \quad (3.5)$$

Where  $\#q$  is the number of times query pixel  $q$  overlapped with a pixel  $p_0$ . With the help of **Equation 3.5** Visibility Map is established (**Fig. 3.7**).

**Figure 3.7: Visibility Map**



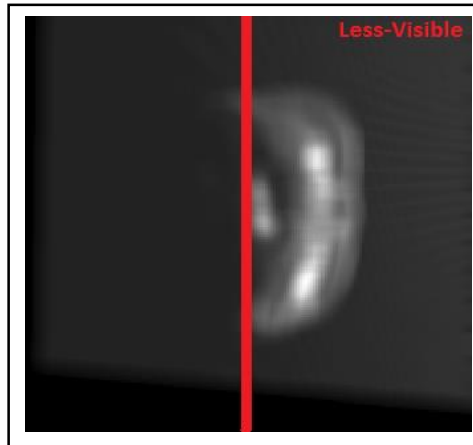
Finally, intensities of poorly visible pixels are replaced by a mean of their intensities and the intensities of their corresponding symmetric pixels, weighted by the visibility scores.

### 3.2.5 Mirroring

This step is the continuation of Hard-Frontalization from Hassner et al.(2015)'s work, however, we stop at this point and not process the hard-frontalized face until occlusions are analyzed. The process is explained however here, to show how Hassner et al.(2015)'s approach would finish it, without supervision.

In this step, it is assumed that a hard-frontalized face image with a proper visibility map is established. First, pre-generated visibility map from the Visibility Estimation step is divided into two from the vertical center of the image. By counting the low-visibility pixels on it, which side of the face is less visible is determined (**Fig. 3.8**).

**Figure 3.8: Side Selection on Visibility Map**



Then, the hard-frontalized face is divided into two from the vertical center of the image, however this time, with a more sophisticated approach is followed in thereafter. hard-frontalized image is processed under 3 parts;

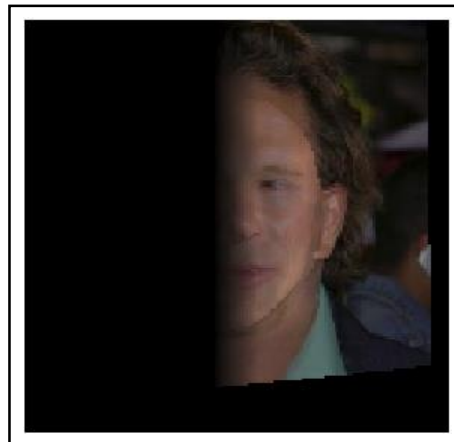
- i. Left flank of hard-frontalized face (**Fig. 3.9**)

**Figure 3.9: Left flank**



- ii. Mirror of left flank of hard-frontalized face, multiplied with weighted visibility map which emphasizes it (**Fig. 3.10**)

**Figure 3.10: Emphasized flank**



- iii. Right flank of hard-frontalized face multiplied with weighted visibility map which de-emphasizes it (**Fig. 3.11**).

**Figure 3.11: De-Emphasized flank**



With the combination of the 3 aforementioned image parts, the initial mirrored image is generated (**Fig. 3.12**).

**Figure 3.12: Initial Mirror**



However, the Mirroring step is yet to be completed at this point. To generate the final result, eye sockets must be corrected. The process is done by extracting eyes from the hard-frontalized face (**Fig 3.13**) with a mask that is specifically located under the eye socket regions.

**Figure 3.13: Extracted Eye Socket Regions**



Same process is done to the initial mirrored face, although this time, it is done to remove eye socket regions (**Figure 3.14**) to place corrected eyes from hard-frontalized face image.

**Figure 3.14: Initial Mirror without Eye Sockets**



Finally mirrored image with no eyes and eyes removed from hard-frontalized image is combined to generate the final result (**Fig. 3.15**).

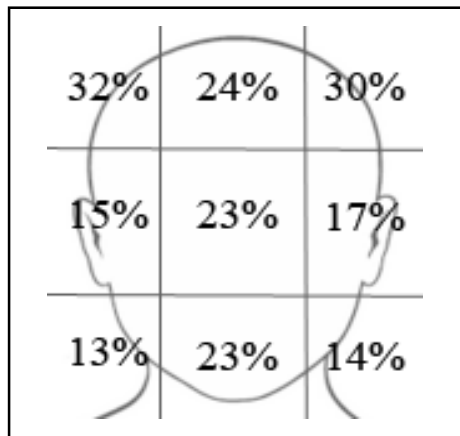
**Figure 3.15: Final Mirror**



### 3.3 OCCLUSION ANALYSIS

In this thesis, we analyze possible occlusion patterns, to create a guideline, for the Frontalization process to follow, thus allowing it to find its way around to the most suitable outcome. Otherwise, the Frontalization process will simply mirror itself, after Hard-Frontalization step, as explained before, resulting in a less desirable output when an occlusion is present in the scene. For the Occlusion Analysis step, we have chosen, COFW as the main dataset, which is specifically designed to present occluded faces in unconstrained real-world conditions. The faces are occluded with large variations in the occluder type. COFW dataset has an average occlusion coverage of over 23%. Individual occlusion percentages of over the face region in the database can be examined in **Figure 3.16**.

**Figure 3.16: Occlusion Occurrence Percentages in COFW**





In our approach, we take on the Occlusion Analysis under two different distinct approaches;

**i. Region Based Analysis**

Here, we take on the Occlusion Analysis process in an old fashioned, yet reliable way. First, we divide the face into nine regions, where occlusions are mostly likely to happen, then we extract features from the image patches and feed them to a classifier, to detect possible occlusions occurring over the face region.

**ii. Pixel Based Analysis**

Here, we try to understand the occlusions occurring in a more fine-grained way, compared to Region Based Analysis. For each pixel in the hard-frontalized face, we extract an image region around it, and compare it pixel wise, to a face with mean texture and neutral illumination. From the comparison results, we select a value, depending on the setting and assign it to the original pixel, to create a visual representation.

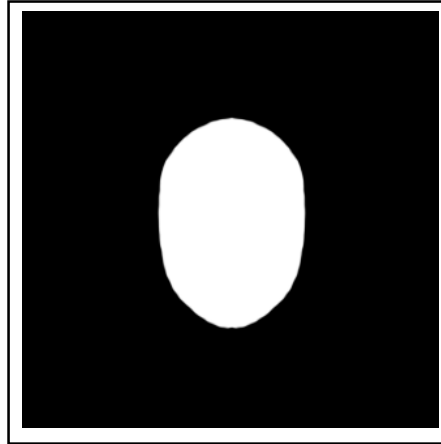
**3.3.1 Region Based Occlusion Analysis**

The aim here is to be able to take an unknown hard-frontalized face and detect possible occlusion occurrences over it. Region based approach is explained in two steps, Learning and Detection. Learning explains the supervised training process and Detecting explains the process of identifying possible occlusion occurrences over an unknown face.

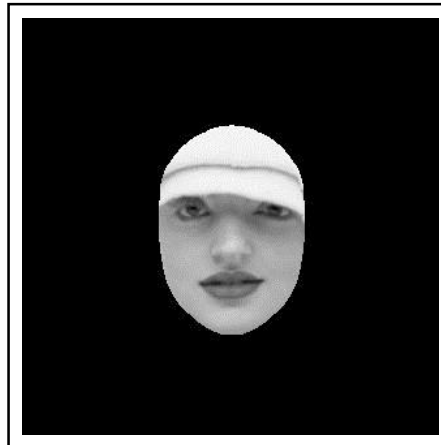
**3.3.1.1 Learning**

We selected 384 unique occluded images with different poses from the COFW dataset for training. Although, before the actual training process, the COFW dataset required significant normalization to become ready for further processing. First, each image was hard-frontalized and saved. To make the data more meaningful for feature extraction process, each image was cropped (**Fig. 3.18**) with the help of a mask (**Fig. 3.17**).

**Figure 3.17: Extraction Mask**

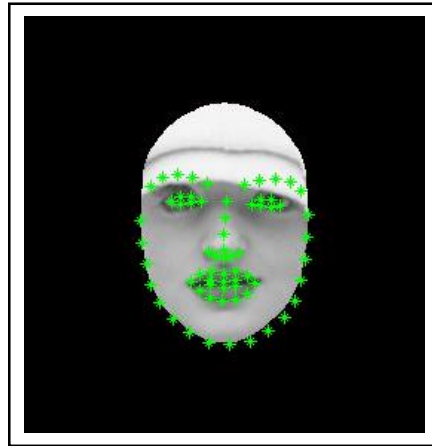


**Figure 3.18: Extraction Result**



Then, faulty images due to the poor facial landmark point localization were removed. For each remaining image, 66 facial landmark points were localized (**Fig. 3.19**) with Akshay et al.(2013)'s novel work on facial feature detection.

**Figure 3.19: Detected Facial Landmark Points with Akshay et al.(2013)'s approach**

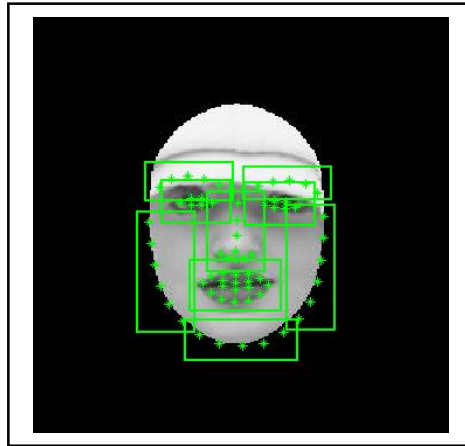


With the help of localized facial landmark points, nine different face regions with high possible occlusion occurrence possibility are chosen;

- i.** Left Eye
- ii.** Right Eye
- iii.** Mouth
- iv.** Nose
- v.** Left Flank of Face
- vi.** Right Flank of Face
- vii.** Bottom Flank of Face
- viii.** Left Brow
- ix.** Right Brow

Using pre-located facial landmark points, occlusion occurrence region boundaries are established (**Fig. 3.20**). With this way, no matter where landmarks are localized, the resulting occlusion regions will cover a good portion of the area due to the usage of moving facial landmark points.

**Figure 3.20: Occlusion Occurrence Regions**



Currently, there exists a pre-calculated binary mask set for COFW dataset available online and the masks are satisfyingly accurate. However, in our approach binary masks also required Hard-Frontalization as their image counterparts did. Frontalizing the binary occlusion masks however, resulted in faulty, corrupted and very poorly aligned results. The risk of using faulty binary occlusion mask images was too steep for us to take, since an occlusion mask needs to be as accurate as much as possible to provide best foundations for Region Based Occlusion Analysis. Because of this, we have created our own binary mask set, manually. From each pre-established occlusion occurrence region, single channel image data with its pre-located facial landmark points are collected. Facial landmark points are then normalized to be inside the cropped regions coordinate system to properly fit. Up until this point, no labeling on the training data exists. To decide where the occlusions lie in the training images, white and black pixels are counted. To avoid counting black pixels beyond the face region, irrelevant areas is colored with a darker tone of gray (**Fig. 3.21**).

**Figure 3.21: Label Image**

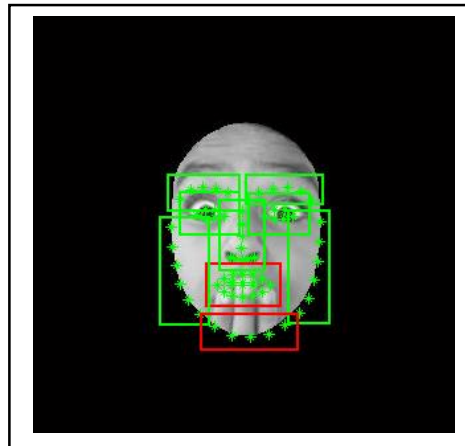


Using the white and black pixel numbers, each region is separately labeled as occluded or un-occluded, according to the percentage of black pixels in the region, calculated with the **Equation 3.6**.

$$P_p = \frac{100 * P_b}{P_w + P_b} \quad (3.6)$$

Where  $P_b$  is the number of the black pixels and  $P_w$  is the number of white pixels in the region. If the percentage of the black pixels in region,  $P_p$  is above a certain threshold, which is held different for each region, due to the difference in possible occlusion patterns, the region is labeled occluded. Otherwise region is labeled un-occluded. A sample of the process can be visually examined in **Figure 3.22** where occluded regions are marked red and un-occluded regions are marked green.

**Figure 3.22: Occlusion Labeling**



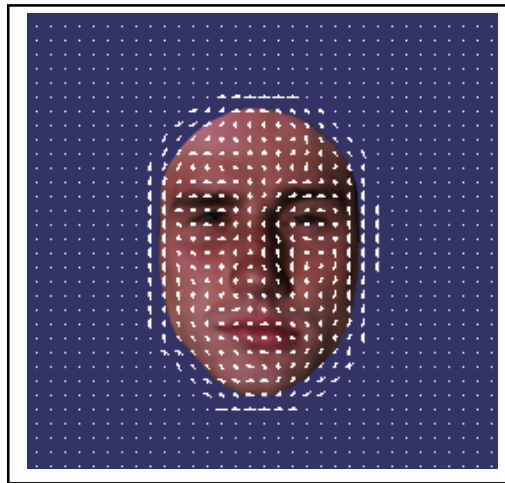
After the labeling is done, a sub-dataset, which consists of normalized facial landmark points and cropped regions with proper labels is established. Due to the positioning of the landmarks each cropped region may get a different width, height and orientation sometimes, this is the cost of the elasticity we get from landmark based region establishment. However, this cost can be compensated, by aligning all cropped images and facial landmark points to the reference face in each respective regions with the help of geometric affine transformation. For feature extraction, a combination of HOG features and LBP features, are chosen, because of the HOG's ability to handle shapes, and LBP's well-rounded ability to handle textures.

Before going any further, how HOG and LBP features are extracted will be explained;

The way HOG features are extracted as follows;

- i.** Normalization of color and gamma values(Optional)
- ii.** Gradients are computed
- iii.** Vote is weighted into spatial and orientation cells
- iv.** Overlapping spatial blocks are normalized under contrast
- v.** Establish Histogram of Oriented Gradients (**Fig. 3.23**)

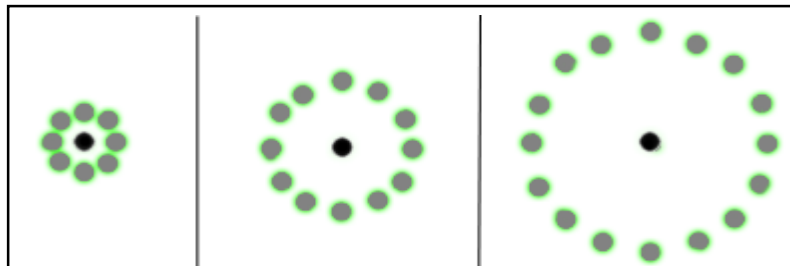
**Figure 3.23: Visual Representation of HOG Features**



The way LBP features are extracted as follows;

- i. Examination region is divided into  $n$  by  $n$  cells.
- ii. For each pixel in a cell, pixels are compared to its neighbors, then pixels are followed along a circle (**Fig. 3.24**).

**Figure 3.24: Pixel Neighbourhood Styles**



- iii. Where the center pixel's value is greater than the neighbor's value, write one, otherwise, write zero.
- iv. Compute a histogram, over the cell, of the frequency of each number occurring
- v. Histograms of all cells are concatenated. This establishes the final feature vector for the selected region.

SVM is chosen as the main classifier. SVM is a supervised learning model with associated learning algorithm that analyzes data and recognize patterns, most commonly used for classification. Given a labeled dataset, where each single data piece belongs to one of the

two groups, the SVM training algorithm will establish a model that assigns new examples into one of the groups, which makes SVM a binary classifier.

SVM model is a representation of how the data pieces are represented as points in space, mapped so that the data pieces of the separate groups are divided by a clear gap that is as wide as possible. New data pieces are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. SVMs was chosen because of its ease of deployment and its compatibility with the problem at hand.

For each region in the face, SVMs are deployed for each feature type, totaling 18 active SVM's at the end of the training session. SVM training times for both features are comparable, when combined, takes less than 10 minutes to finish, thanks to the small size of face regions at each part, which are less than 100 by 100 at any part.

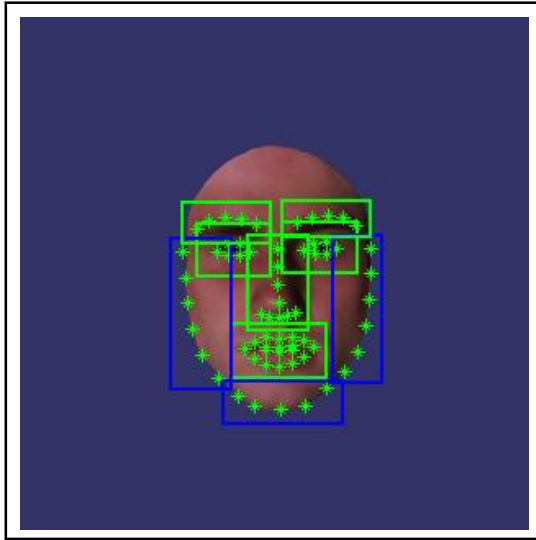
### **3.3.1.2 Detection**

First, the input image is first hard-frontalized, then 66 facial landmark points are localized, finally from each facial Occlusion region, image data and facial landmark points are extracted. Extracted data is then normalized using the reference face, with geometric affine transform. From the normalized image data LBP and HOG features are extracted and fed into their respective SVM's. The final classification from the SVM's decide if an occlusion is present in a region or not, decided by classifications from the both classifiers for a region. In each region classifiers show different occlusion classification performances. This is compensated by taking the vote of a classifier over another, when their occlusion detection votes are different from each other.

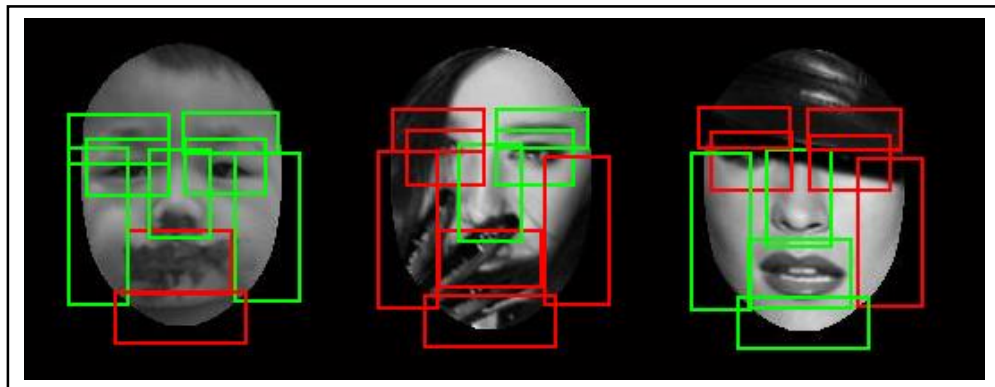
Each feature classifier has a weighted vote over another in a different region. Vote weights where, HOG weighted regions are painted green and LBP weighted regions are painted blue can be seen in **Figure 3.25**. Finally, few examples of how Region Based Occlusion Analysis results look on COFW dataset can be seen in **Figure 3.26**.



**Figure 3.25: Feature Weights for Every Occlusion Occurrence Region**



**Figure 3.26: Sample Region Based Occlusion Analysis Results**

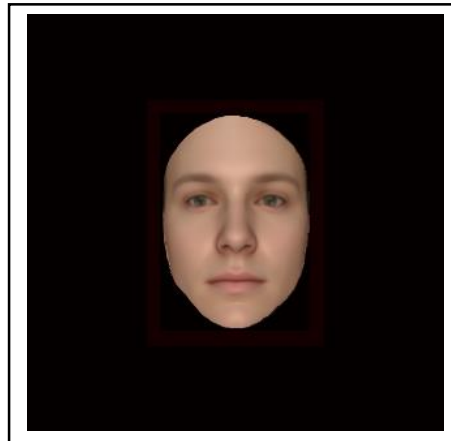


### 3.3.2 Pixel Based Occlusion Analysis

An alternative to the Region Based Occlusion Analysis, we also established a pixel based approach, to create an Occlusion Mask, aiming to capture occlusion occurrences in a more precise manner.

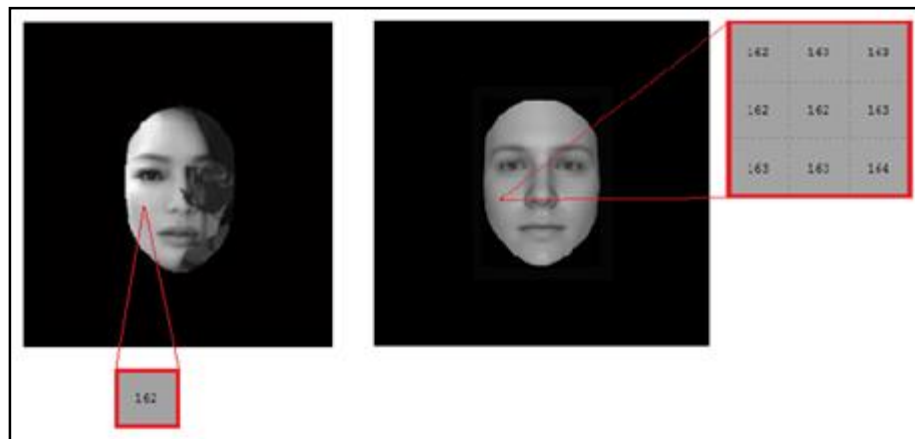
Before starting, we select a mean textured face, to take as the baseline (**Fig. 3.27**).

**Figure 3.27: Mean Textured Face**



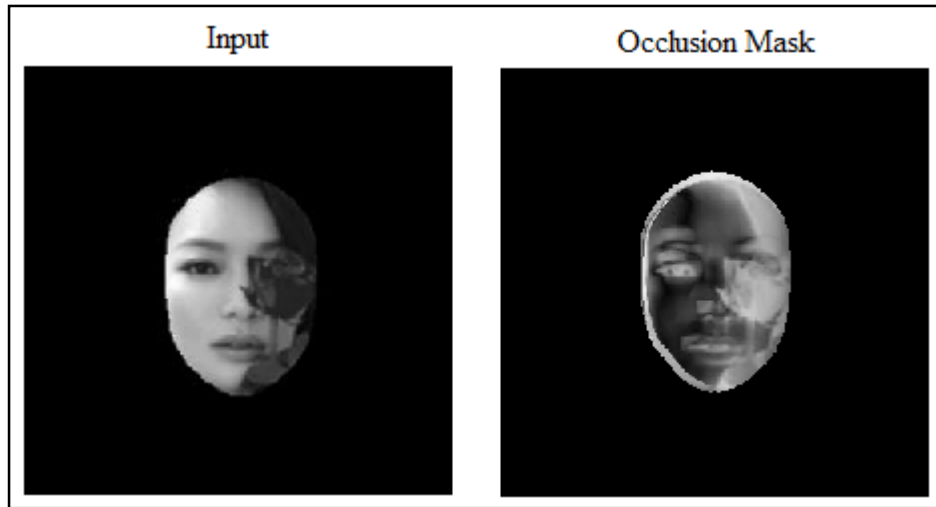
First, we select a query, hard-frontalized face image, and align it with the mean textured face image. Then for each pixel in the query image, we extract an  $N \times N$  sized region, centered around the same pixel selected in the query image (**Figure 3.28**).

**Figure 3.28: Intensity Values from Hard-Frontalized Face and Region Values From Mean Textured Face**



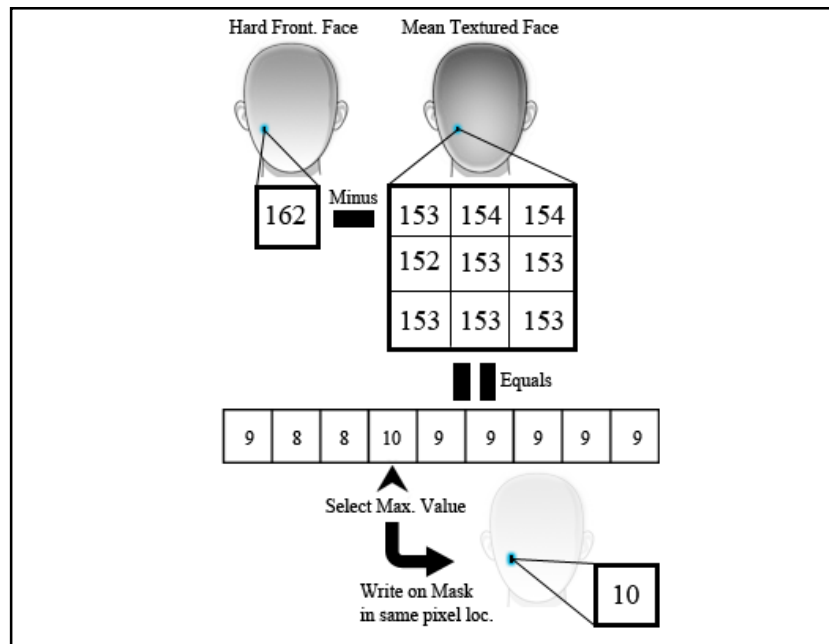
For each pixel in the extracted  $N \times N$  sized region, we calculate the absolute differences with the pixel selected from the query image and store them. From the stored values, we take the maximum value, and place it in the same location in a new image, effectively generating an Occlusion Mask (**Fig. 3.29**).

**Figure 3.29: Occlusion Mask of a Hard-Frontalized image**



The way our pixel based approach works to create the Occlusion Mask can be examined in illustration on **Figure 3.30**.

**Figure 3.30: Process of Occlusion Mask Generation**


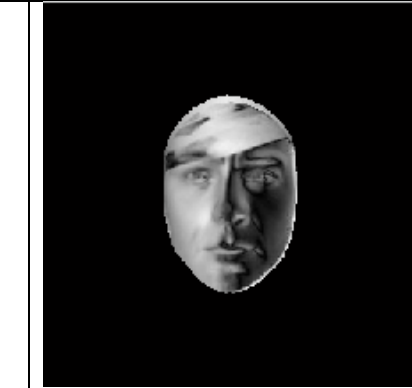

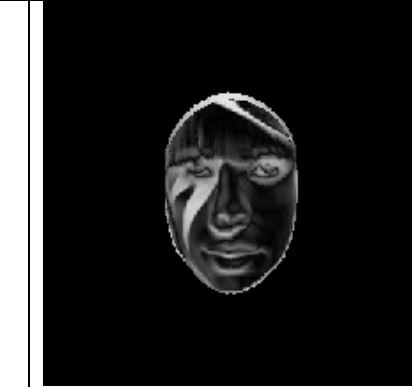


In the Occlusion Mask, regions with high intensity represent where occlusion occurrence possibility is the highest and the regions with low intensity where the occlusion occurrence possibility is the lowest.

Although less-sophisticated as an implementation, Pixel Based Occlusion Analysis provides a more lucid picture of the occlusions occurring over the face region in a more precise fashion. The shortcoming of this Analysis method is, the fact that its intensity based, therefore, Pixel Based Occlusion Analysis will only detect occlusions that have significant intensity changes compared to the mean textured face, also, this means that Eye regions are especially prone to experience these effects.

Sample results of Pixel Based Occlusion Analysis can be examined in **Table 3.1**.

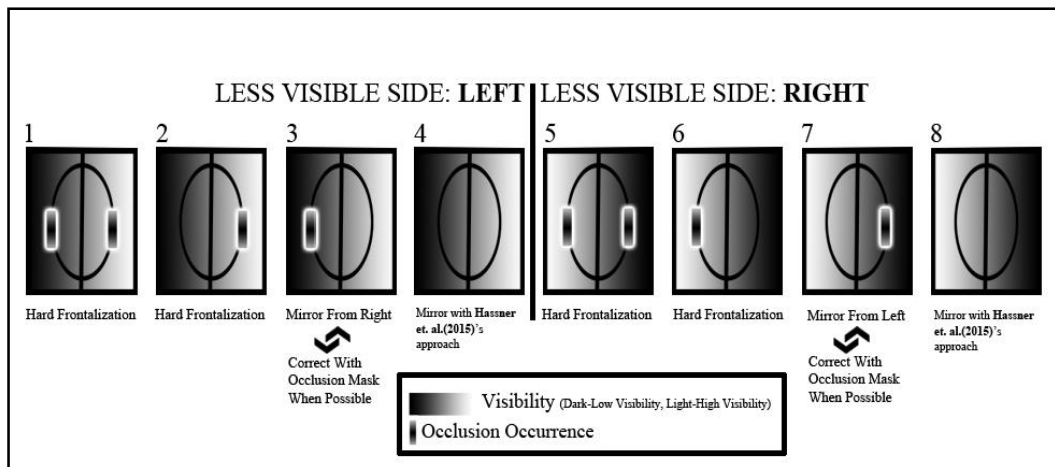
**Table 3.1: Pixel Based Occlusion Analysis Results**

Hard-Frontalized Face	Occlusion Mask
	
	

### 3.4 Occlusion Based Soft-Frontalization

Soft-Frontalization is the process of correcting a hard-frontalized face image, to achieve higher visual quality. To find the most suitable Soft-Frontalization approach, we have established Frontalization Guidelines. With Frontalization Guidelines, each occlusion occurrence pattern in the face region has a suggested approach to finalize it (**Fig 3.31**).

**Figure 3.31: Frontalization Guidelines**



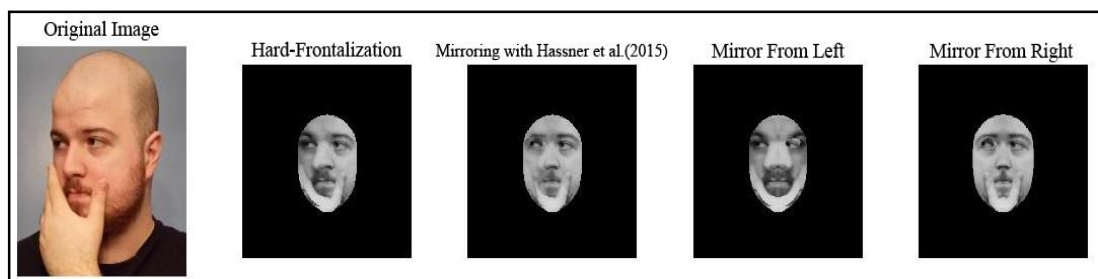
We established our guidelines, with the help of the Occlusion Analysis results and the Visibility Map, established after the Hard-Frontalization process.

There are total eight cases, and eight total possible solutions with two interchanging ones, totaling ten possible ways to finalize the Frontalization Process. In Soft-Frontalization part, we will explain our guidelines and their respective solutions under four main cases; Both Sides Occluded, Occlusion on Visible Side, Occlusion on Non-Visible Side, Un-Occluded on Both Sides.

### 3.4.1 Both Sides Occluded

In this case, it is very apparent that, finalizing the Frontalization process at Hard-Frontalization step is the most suitable choice, considering any type of mirroring at this environment is unfit and will yield faulty results (**Fig. 3.32**).

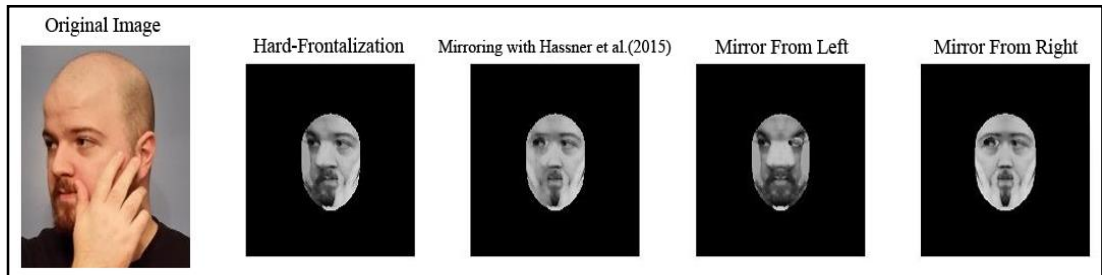
**Figure 3.32: Soft-Frontalization Results of a Face with Both Sides Occluded**



### 3.4.2 Occlusion on Visible Side

Once again, most suitable decision to finalize the Frontalization process here is to finish it after Hard-Frontalization step, otherwise, same problems risen on previous case will make themselves apparent (Fig. 3.33).

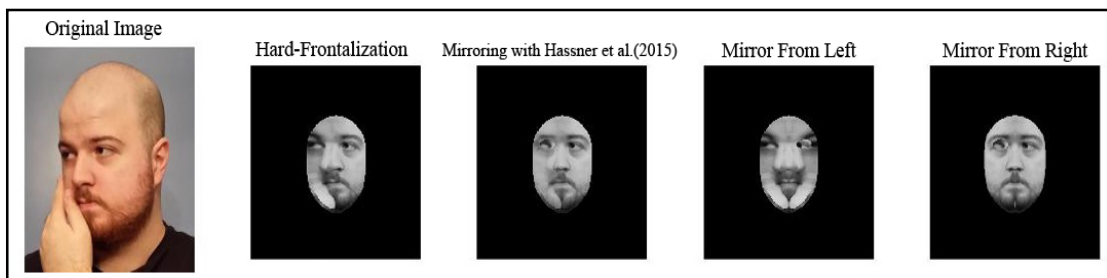
**Figure 3.33: Soft-Frontalization Results of a Face with Occlusion on Visible Side**



### 3.4.3 Occlusion on Non-Visible Side

In this case, we can take the mirror from the Un-Occluded, Visible side of the face, which provides better results, compared to finalizing the process at Hard-Frontalization step (Fig. 3.34).

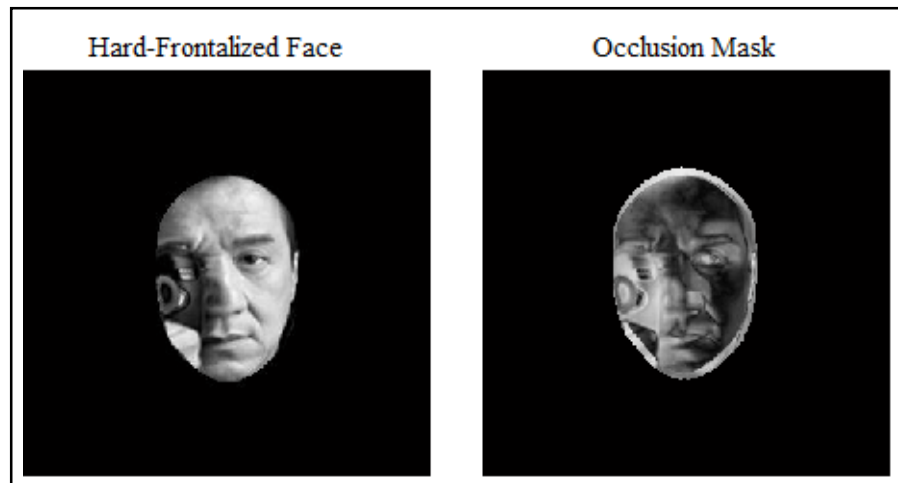
**Figure 3.34: Soft-Frontalization Results of a Face with Occlusion on Non-Visible Side**



This case, is however interchangeable with another solution, tracking back to our Pixel Based Occlusion Analysis method, depending on the occluder type. If, there exists an occluder over the face region, with overall intensity distribution, dissimilar to human skin, we can say, that the occluder is correctable, by extending our Pixel Based Occlusion Analysis method.

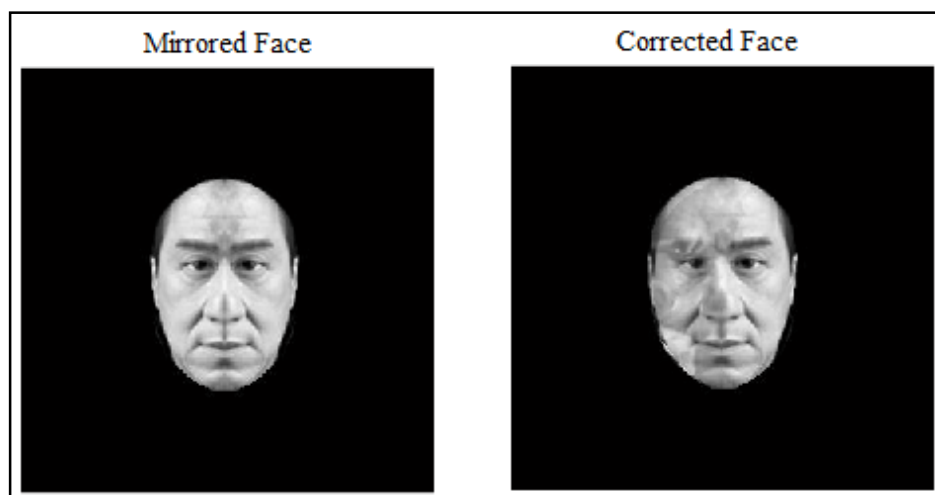
Simply put, we can use the Occlusion Mask generated during the Pixel Based Occlusion Analysis, to reconstruct the occluded region of the face, from the on-occluded side, by using the Occlusion Mask as weights (Fig. 3.35 ).

**Figure 3.35: Pixel Based Occlusion Mask of a Hard-Frontalized Face**



We can emphasize the regions from the un-occluded side of the face over the Occluded side, without damaging the authenticity of the Occluded side, resulting in a more visually appealing result, compared to basic mirroring (Fig. 3.36).

**Figure 3.36: Comparison of a Mirrored face and a Corrected Face**

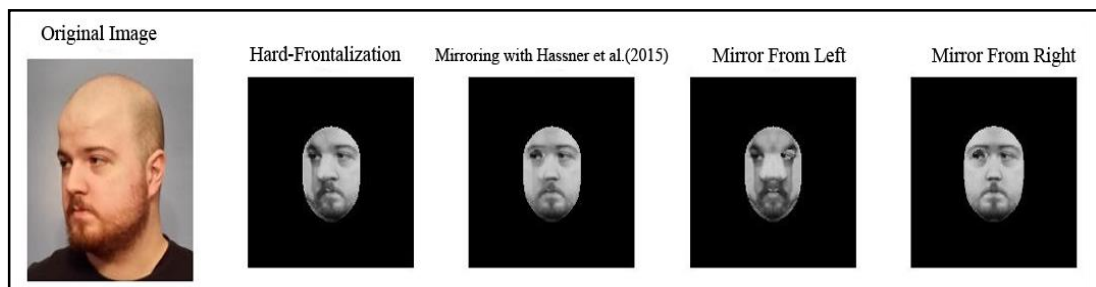


The advantage of this interchangeable solution scheme is that, when possible, we can push our luck to create a soft-frontalized face, which keeps its identity well preserved, compared to mirrored one.

### 3.4.4 Un-Occluded on Both Sides

In this case we use Hassner et al.(2015)'s mirroring approach on hard-frontalized face image, to achieve the most suitable result (**Fig. 3.37**).

**Figure 3.37: Soft-Frontalization Results of a Face with No Occlusions**



Advantage of this solution is that, during the mirroring process, Visibility is also taken into the account, with help of the Visibility Map, created during the Hard-Frontalization process. Regions, with the low visibility, are filled from the side with high visibility, essentially doing the aforementioned Occlusion Correction in the context of visibility.



## 4. PERFORMANCE EVALUATION

In this chapter, we will present our performance results, under two headlines, Region Based Occlusion Analysis Accuracy, Pixel Based Occlusion Reconstruction results.

### 4.1 SETUP

Software implementation of this thesis was done in MATLAB environment entirely, working on 2013b version with Visual Studio 2010 Compiler for Mex files. Setup PC is equipped with a 2.9 GHz i7 Processor and 12 GBs of RAM. Dataset used on experiments is COFW, and is readily available online, with both RGB and Grayscale versions.

### 4.2 REGION BASED OCCLUSION ANALYSIS ACCURACY

To evaluate our Region Based Occlusion Analysis accuracy we selected 278 test images from the test part of the COFW dataset and hard-frontalized them.

While calculating the accuracy of the analysis, we take occlusion state of the each individual pre-established occlusion region into consideration, comparing classification results to the the ground truth occlusion data. The results can be examined in **Table 4.1**.

**Table 4.1: Individual Occlusion State Accuracies**

	Left Eye	Right Eye	Mouth	Nose	Left Flank of Face	Right Flank of Face	Bottom Flank of Face	Left Brow	Right Brow
Accuracy	92.1%	91.4%	86.4%	92.9%	85.7%	81.3%	83.5%	83.1%	77.4%






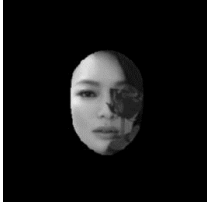
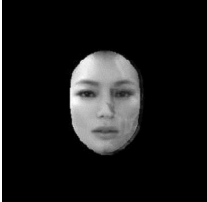
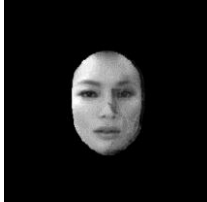

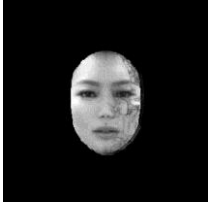

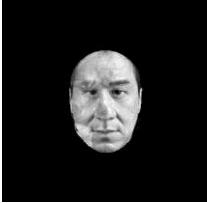

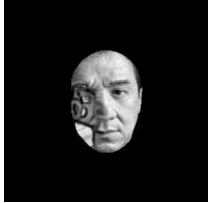

Left eye, right eye and nose region accuracies are particularly high, staying above 90%. Mouth and left flank of face region accuracies are, acceptably high hovering above 85%. Right flank of face, bottom flank of face and right brow accuracies are, still above expectations, however, could be higher, especially for the right brow. The lower percentages on some regions can be arguably explained by the lack of quality occlusion data on that respective region, considering the COFW dataset isn't specifically created for each individual region.

### 4.3 PIXEL BASED OCCLUSION RECONSTRUCTION RESULTS

In addition to a detection based accuracy measure, we also show how aesthetically pleasing the results of Pixel Based Occlusion Reconstruction Results are, under different reconstruction settings.

We show our results under four settings, Max, Min, Mean and Normalized Cross Correlation. Max, Min and Mean are indications of on what condition the absolute value is selected from the calculation matrix. Max is our dedicated Correction setting as aforementioned. Normalized Cross Correlation is a method, for measuring the similarity between two images and main comparison to our outputs. The results can be examined in **Table 4.2.**

**Table 4.2: Occlusion Correction Results**

Hard-Frontalized Image	Pixel Based Result(MAX)	Pixel Based Result(MEAN)	Pixel Based Result(MIN)	Normalized Cross Correlation Result
				
				
				



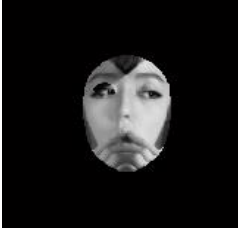
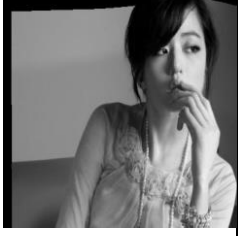





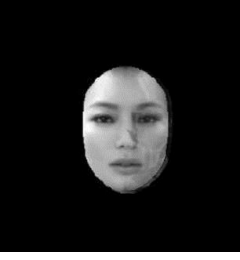
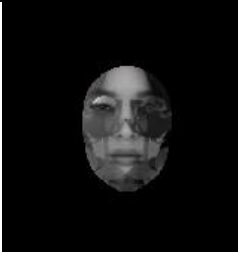





It is visually apparent that, MAX setting produces the most visually appealing result compared to other settings, without damaging the authenticity of the original image. MEAN setting is a close contender, although there are slight traces of occlusion remains over the face. MIN setting is, arguably the worst, barely doing anything in the context of Occlusion Correction. Normalized Cross Correlation is close to the MEAN setting, effectiveness wise, definitely providing results, although not acceptable because of the occluder remains.

Compared to others, MAX setting is clearly the most suitable choice, while not being perfect, it provides highly acceptable results, which can be used as a proper soft-frontalized output in the right case in Frontalization Guidelines.

#### **4.4 OCCLUSION BASED SOFT FRONTALIZATION RESULTS**

Finally, we compare our final Soft-Frontalization Results visually, with Hassner et al.(2015) and Zhu et al.(2015), which can be examined in the **Table 4.3**. We give examples for each case in our Frontalization Guidelines.

**Table 4.3: Soft-Frontalization Results Comparison**

	<b>Hard-Frontalized Image</b>	<b>Our Results</b>	<b>Hassner et al.(2015)</b>	<b>Zhu et al.(2015)</b>
<b>1 and 3</b>				
<b>2 and 6</b>				
<b>3 and 7</b>				
<b>4 and 8</b>				

In case 1-3, Our results are visually appealing, providing very close results with Zhu et al.(2015). Hassner et al.(2015) is the least visually acceptable result here, because of the mirroring based approaches not being robust against occlusions.

In case 2-6, Once again our result and Zhu et al.(2015) are close visual appeal wise. Hassner et al.(2015) is the least visually acceptable result here, because of the mirroring based approaches not being robust against occlusions.

In case 3-7, our result is arguably the most visually appealing, due to occlusion corrected Frontalization. Zhu et al.(2015)'s results are also acceptable, although occlude is still present. Hassner et al.(2015) once again, the least visually acceptable result.

In case 4-8, our result and Hassner et al.(2015)'s result is equal, due to Occlusion Analysis, choosing the Hassner et al.(2015)'s mirror based Soft-Frontalization as the most suitable approach. Zhu et al.(2015)'s results are also acceptable, however, there is a slight anomaly in the mouth region, making the expression look unnatural.

## 5. CONCLUSION

In this study, we investigated negative effects of occlusion on Frontalization approaches, and how can we establish a self-aware decision-maker approach to avoid and correct them as much as possible. Frontalization approaches are in the brink of becoming a dedicated pre-processing step, and in this thesis, we did our best to try to take it one step closer.

As our main contribution, we took a state-of-the-art Frontalization approach and added it the ability to detect, analyze and correct occlusion occurrences over the face region. With the new learned abilities, modified Frontalization approach gained a new stage of complexity, allowing it to simply, reason and act, thus making it the first Frontalization approach, taking occlusions into consideration when processing. We have tested our Region Based Occlusion Analysis method in COFW dataset reported between 77 percent and 92 percent individual occlusion detection accuracy over 278 unique occluded images. Also we showed our Pixel Based Occlusion Correction results on the same dataset, this time however, under visual satisfaction wise. We illustrated that our results under MAX setting is superior to the novel image similarity measurement approach Normalized Cross Correlation. For the next step in the evolution of the Frontalization approaches, we aim to contribute, by creating a more sophisticated occlusion model to understand occlusion occurrences at much higher precision and to establish our own Frontalization method to further complement it, all, at real time computational efficiency.

## REFERENCES

### *Other Publications*

- Akshay et al., 2013. Robust Discriminative Response Map Fitting with Constrained Local Models. *Computer Vision and Pattern Recognition*, pp.3444–3451
- Annan et al., 2012. bias-variance tradeoff for cross-pose face recognition. *IEEE Trans Image Process*, pp.305-315
- Artizzu et al., 2013. Robust face landmark estimation under occlusion. *International Conference on Computer Vision*, pp.1513-1520
- Ashraf et al., 2008. Region patch correspondences for improved viewpoint invariant face recognition. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1-8
- Asthana et al., 2009. Region-based face synthesis for pose-robust recognition from single image. *Proc. Brit. Mach. Vis. Conf.*, pp.1-10
- Asthana et al., 2011. Pose normalization via learned warping for fully automatic face recognition. *Proc. Brit. Mach. Vis. Conf.*, pp.1-11
- Bergand and Belhumeur, 2012. Tom-vs-pete classifiers and identity-preserving alignment for face verification. *Proc. Brit. Mach. Vis. Conf.*, pp.7
- Beymer and Poggio, 1995. Face recognition from one example review. *Proc. IEEE Int. Conf. Comput. Vis.*, pp.500-507
- Bookstein, 1989. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Analysis Mach. Intel*, pp.567-585
- Chaïetal., 2007. Locally linear regression for pose-invariant face recognition. *IEEE Trans Image Process*, pp.1716-1725
- Gao et al., 2001. Fast face identification under varying pose from a single 2-d model view. *IEEE Proceedings-Vision, Image and Signal Processing*, pp.248-253
- Gao et al., 2009. Face recognition across pose: A review. *Pattern Recognition*, pp.2876-2896
- Ghiasi and Fowles, 2014. Occlusion Coherence: Localizing Occluded Faces with a Hierarchical Deformable Part Model, *Computer Vision and Pattern Recognition*, pp.1899-1906
- Han and Jain, 2012. 3d face texture modeling from uncalibrated frontal and profile images. *Proc. IEEE Int. Conf. Biometrics Theory Appl. Syst.*, pp.223-230
- Hao and Qi., 2015. A unified regularization framework for virtual frontal face image synthesis. *IEEE Sig. Process Lett.*, pp.559-563
- Hassner, 2013. Viewing real-world faces in 3d. *Proc. IEEE Int. Conf. Comput. Vis.*, pp.3607-3614
- Hassner et al., 2015. Effective face frontalization in unconstrained images. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp.4295-4304
- Heo and Savvides, 2011. Rapid 3d face modeling using a frontal face and a profile face for accurate 2d pose synthesis. *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognit.*, pp.632-638
- Heo and Savvides, 2012. 3-d generic elastic models for fast and texture preserving 2-d novel pose synthesis. *IEEE Trans. Inf. Forensics Security*, pp.563-576
- Kan et al., 2014. Stacked progressive auto-encoders(spae) for face recognition across poses. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp.1883-1890
- Kim et al., 2010. Face Occlusion Detection by using B-spline Active Contour and Skin Color Information, *Control Automation Robotics&Vision*, pp.627-632

- Lee and Ranganath, 2003. Pose-invariant face recognition using a 3d deformable model. *Pattern Recognit*, pp.1835-1846
- Li et al., 2009. Maximizing intra-individual correlations for face recognition across pose differences. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp.605-611
- Li et al., 2012. Morphable displacement field based image matching for face recognition across pose. *Proc. Eur. Conf. Comput. Vis.*, pp.102-115
- Li et al., 2014. Maximal likelihood correspondence estimation for face recognition across pose. *IEEE Trans. Image Process.*, pp.4587-4600
- Lin and Liu, 2006. Face Occlusion Detection for Automated Teller Machine Surveillance, *Advances in Image and Video Technology*, pp.641-651
- Lin and Tang, 2007. Quality-Driven Face Occlusion Detection and Recovery, *Computer Vision and Pattern Recognition*, pp.1-7
- Liu and Chen, 2005. Pose-robust face recognition using geometry assisted probabilistic modeling. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp.502-509
- Pascal et al., 2009. A 3D Face Model for Pose and Illumination Invariant Face Recognition *Advanced Video and Signal Based Surveillance*, pp.296-301
- Smith et al., 2006. Resolving hand over face occlusion, *Computer Vision in Human-Computer Interaction*, pp.160-169
- Taigan et al., 2014. Deepface: Closing the gap to human-level performance in face verification. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp.1701-1708
- Yim et al., 2015. Rotating your face using multi-task deep neural network. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp.676-684
- Yin et al., 2011. An associate-predict model for face recognition. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp.497-504
- Zhang et al., 2006. Automatic texture synthesis for face recognition from single views. *Proc. IEEE Int. Conf. Pattern Recognit.*, pp.1151-1154
- Zhang et al., 2008. Recognizing rotated faces from frontal and side views: An approach toward effective use of mugshot databases. *IEEE Trans. Inf. Forensics Security*, pp.684-697
- Zhang et al., 2013. Pose-robust face recognition via sparse representation. *Pattern Recognition*, pp.1511-1521
- Zhang et al., 2013a Random Faces Guided Sparse Many-to-One Encoder for Pose-Invariant Face Recognition. *International Conference on Computer Vision*, pp.2416-2423
- Zhu et al., 2013. Deep region identity-preserving face space. *Proc. IEEE Int. Conf. Comput. Vis.*, pp.113-120
- Zhu et al., 2015. High-Fidelity Pose and Expression Normalization for Face Recognition in the Wild *Computer Vision and Pattern Recognition*, pp.787-796
- Zhu and Ramanan, 2012. Face Detection, Pose Estimation and Landmark Localization in the Wild, *Computer Vision and Pattern Recognition*, pp.2879-2886
- Zhu et al., 2014. Multi-viewperceptron: A deep model for region face identity and view representations. *Proc. Adv. Neural Inf. Process. Syst.*, pp.217-225