12-2013

Face recognition in the wild.

Mostafa A. Farag
University of Louisville

FACE RECOGNITION IN THE WILD

By

Mostafa A. Farag
B.S., University of Louisville, 2012

A Thesis
Submitted to the Faculty of the
University of Louisville
J. B. Speed School of Engineering
as Partial Fulfillment of the Requirements
for the Professional Degree

MASTER OF ENGINEERING

Department of Electrical and Computer Engineering

December 2013

FACE RECOGNITION IN THE WILD

Submitted by:_____
Mostafa Farag

A Thesis Approved On

_____
(Date)

by the Following Reading and Examination Committee:

_____
Dr. James Graham, Thesis Director

_____
Dr. Patricia Ralston

_____
Dr. Aly Farag

ACKNOWLEDGEMENTS

I wish to thank my mother, Salwa Elshazly, my father, Dr. Aly Farag, my sister, Dr. Amal Farag, my two brothers Ahmed and Ibrahim, and the rest of my family for their ongoing support throughout my education. I would like to thank Dr. James Graham for the opportunities he has made possible for me, as well as his patience for not giving up on me during stagnant times. I would also like to thank Ahmed Shalaby, Ahmed Elbarouky, and all other members of the CVIP Lab both past and present that I have shared this experience with. A special thank you to Dr. Mostafa Abdelrahman and Mr. Mike Miller who provided guidance and expertise during the research reported in this thesis. I would also like to thank the 21 men and women that participated as test subjects; without them, the results of this thesis would not have been possible. Lastly, I would like to thank my final committee member, Dr. Patricia Ralston. I am truly honored and blessed that she accepted my committee invite. She has always been a friendly guide in my college career, and it is only fitting that my first college professor is present as I move on to the next chapter of my life.

ABSTRACT

Research in face recognition deals with problems related to Age, Pose, Illumination and Expression (A-PIE), and seeks approaches that are invariant to these factors. Video images add a temporal aspect to the image acquisition process. Another degree of complexity, above and beyond A-PIE recognition, occurs when multiple pieces of information are known about people, which may be distorted, partially occluded, or disguised, and when the imaging conditions are totally unorthodox! A-PIE recognition in these circumstances becomes really "wild" and therefore, *Face Recognition in the Wild* has emerged as a field of research in the past few years. Its main purpose is to challenge constrained approaches of automatic face recognition, emulating some of the virtues of the Human Visual System (HVS) which is very tolerant to age, occlusion and distortions in the imaging process. HVS also integrates information about individuals and adds contexts together to recognize people within an activity or behavior. Machine vision has a very long road to emulate HVS, but face recognition in the wild, using the computer, is a road to perform face recognition in that path.

In this thesis, *Face Recognition in the Wild* is defined as unconstrained face recognition under A-PIE+; the (+) connotes any alterations to the design scenario of the face recognition system. This thesis evaluates the Biometric Optical Surveillance System (BOSS) developed at the CVIP Lab, using low resolution imaging sensors. Specifically, the thesis tests the BOSS using cell phone cameras, and examines the potential of facial biometrics on smart portable devices like iPhone, iPads, and Tablets.

For quantitative evaluation, the thesis focused on a specific testing scenario of BOSS software using iPhone 4 cell phones and a laptop. Testing was carried out indoor, at the CVIP Lab, using 21 subjects at distances of 5, 10 and 15 feet, with three poses, two expressions and two illumination levels. The three steps (detection, representation and matching) of the BOSS system were tested in this imaging scenario. False positives in facial detection increased with distances and with pose angles above $\pm 15°$. The overall identification rate (face detection at confidence levels above 80%) also degraded with distances, pose, and expressions. The indoor lighting added challenges also, by inducing shadows which affected the image quality and the overall performance of the system. While this limited number of subjects and somewhat constrained imaging environment does not fully support a "wild" imaging scenario, it did provide a deep insight on the issues with automatic face recognition. The recognition rate curves demonstrate the limits of low-resolution cameras for face recognition at a distance (FRAD), yet it also provides a plausible defense for possible A-PIE face recognition on portable devices.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. Introduction

Face recognition is an important field in behavioral and applied sciences. It deals with understanding the information content in the face, from physically looking at people, or through images of them. Face recognition may be performed in absolute (i.e., observing a face) or in relative terms (i.e., observing faces during an action). Under each scenario, recognizing a face means associating with it a known reference of it (e.g., a previous picture) and verification of it through subsequent steps to confirm that the recognized face is genuine. Human face recognition is an interesting multidisciplinary area in psychology, psychiatry, computer engineering, and related disciplines. Understanding how the human brain recognizes faces is a fascinating, and still non-conclusive, art and science (Ekman and Rosenberg, 2005 [1] contains a collection of views on what the face reveals).

Machine or computer (or automatic) face recognition is a maturing field dating back to the early 1970's (e.g., [2]). From an image or a video, faces are detected and a representation is generated for them, which is then compared with representation of people in a gallery (data base) in order to perform the recognition. The construction of the gallery, data structure and facial representations is performed a priori, and is done off-line. Once a match between a candidate face (probe) and the gallery (database) is obtained, a verification step follows to authenticate that

the recognized face is genuine. The steps of face recognition then are three: detection, representation and matching. A rich theory exists for each step, and great many algorithms have been developed in the past two decades, which enable fast face recognition, and will continue to improve with recurrent progress in sensors and computation.

As automatic face recognition starts from an image or video, the circumstances of acquisition of such images and videos may vary. In general, an image in the camera is an interaction of the individual, the lighting (imaging) condition and the camera itself. Natural, unconstrained, images are pose point instantiations of the people in the scene, which may be involved in a particular activity (e.g., working alone, interacting with a group, or in a sightseeing trip, etc.), and given lighting circumstances. Poses and expression are aspects of human behavior; illumination is an aspect of the lighting in the environment; the three characteristics: Pose, Illumination and Expression (PIE) are independent. Unconstrained face recognition is the methodology that addresses the PIE scenarios of imaging of an individual or a group. An additional factor dealing with imaging condition is that of Age (time of acquisition). Hence, the A-PIE recognition is the most general, and is the most applicable in current development of automatic face recognition. Researches in A-PIE face recognition seek approaches that tolerate (invariant to) age, pose, illumination and expression.

Another degree of complexity above and beyond A-PIE recognition is when multiple pieces of information are known about people, which may be distorted, partial, occluded, or disguised, and when the imaging conditions are totally unorthodox! A-PIE recognition in these circumstances becomes really "wild" and therefore, *Face Recognition in the Wild* has emerged as a field of research in the past few years. This thesis is on Face Recognition in the Wild! There is no specific definition as yet for this "wildness" in the literature; in this thesis it will be defined

as Unconstrained Face Recognition Under A-PIE+; the + will connote any alterations to the design scenario of the face recognition system. That may include alterations in the sensors, the imaging environment and intended applications. An automatic face recognition system based on high resolution CCD cameras may be asked to work on scenarios where cameras are low resolution. A system designed to work in homogenous lighting conditions may be asked to work on open environments such as stadium or shopping malls, or on a racing track! A system that is bulky and heavy in terms of sensors, computers and power sources may be tested on mundane devices such as a smart phone.

Perhaps, an image from the news outlets of the crowd attended the 2013 President Obama second term inauguration can provide a sense of variability of faces, and how an automatic face recognition system may be challenged (e.g., to perform law-enforcement or a public service function). Figure 1 is snap shot of some of the crowd that appeared by the US Capital to listen to Obama's inauguration speech on January 21, 2013.

FIGURE 1 - An image of a crowd in the open, illustrating the richness of faces and challenges for automatic face recognition in the wild (adopted from the New York Times archives, January 21, 2013)

Figure 2 from the same occasion as well, shows that even in a controlled seating, the crowd may be difficult to recognize; necessitating help of "find a person" by the New Times.

In this sense of "wildness", the problem is indeed fuzzy and cannot be defined. Yet, it is what it is, a "digression" of unconstrained facial biometrics under A-PIE into evolving or unintended domains of use, or when aspects of the face recognition process (e.g., sensors, representations, compute engines, power requirements, networking) change. To impose a degree of control to the problem, one needs to start with a system designed under A-PIE assumptions, and then modify some aspects of it beyond the design specifications.

FIGURE 2 - An image of a crowed in the open, but controlled and pre-assigned seating (adopted from the New York Times archives, January 21, 2013)

The Computer Vision and Image Processing Laboratory (CVIP Lab) designed, built and tested a facial Biometric Optical Surveillance System (BOSS) based on A-PIE constraints. This thesis will be based on evaluating BOSS using low resolution imaging sensors and multiple

subjects. As cell phones, portable, networked and "cloud" computing are part of modern era; this thesis will challenge BOSS as such.

A. Research Domain of the Thesis

Specifically, the thesis will test BOSS using low resolution cell phone cameras. The contribution of the thesis is on discovering portable face recognition, which may lead into "sensor networks" of facial biometrics units; which may be deployed in healthcare, law enforcement, and group activities such as camping and scouting.

The thesis is structured as describing the following: i) the face recognition problem; ii) A-PIE face recognition; iii) the BOSS facial biometric system; iv) describing BOSS using cell-phone in terms of sensors and portability. The next chapter will provide a concise discussion of these four issues.

B. Thesis Outline and Contributions

The thesis is arranged as follows: Chapter 2 will cover elements of the mathematical foundation related to "detection," "representation," and "matching" of faces. The chapter will also discuss invariance in A-PIE Facial Biometrics as well as give a summary of the BOSS project. Chapter 3 will discuss the performance of the BOSS system in its current form at the CVIP Lab. Chapter 4 will discuss performance evaluation of BOSS using the low resolution camera of the iPhone 4. Chapter 5 will discuss portability of a BOSS-like system on smart phones, and how sensor networks of cell phones may be used for practical applications in security, surveillance, disaster relief and healthcare. Chapter 6 will summarize the thesis

contributions and will put forth suggested extensions and postulates on possible future use of facial biometrics on the cloud.

II. BASICS OF FACIAL BIOMETRICS AND THE BOSS SYSTEM

A. Introduction

As stated in Chapter 1, facial biometrics aims at recognizing and authenticating faces. Figure 3 is a representation of the face recognition process, which is formed of three major components: detection, representation and recognition (also called classification or matching). This chapter will present the basic mathematical foundation for each of these three steps, with focus on the approaches used in the BOSS project.

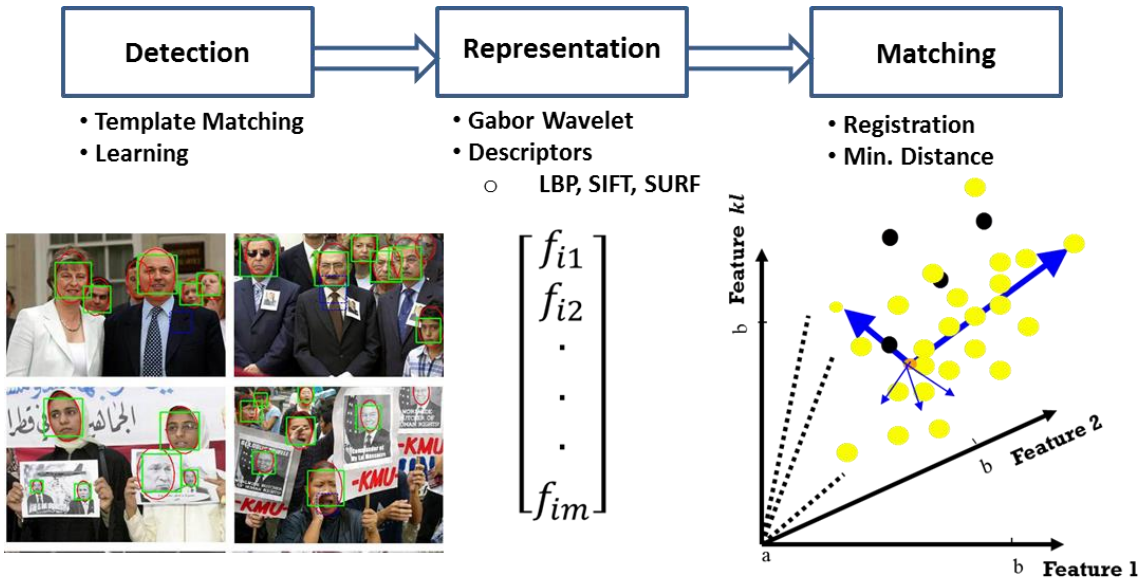


FIGURE 3 - Basic components of face recognition

From anthropometric point of view, over 70 features have been used to describe the anatomical points of interest in the human face, as shown in Figure 4 [32]. Figure 5 shows the NIST ISO/IEC code for defining the facial features [33]. From computational point of view, however, the information content in the face, vis-à-vis automatic face recognition systems, lies in the region from chin to above the eyebrows. The forehead does not carry discriminatory information.

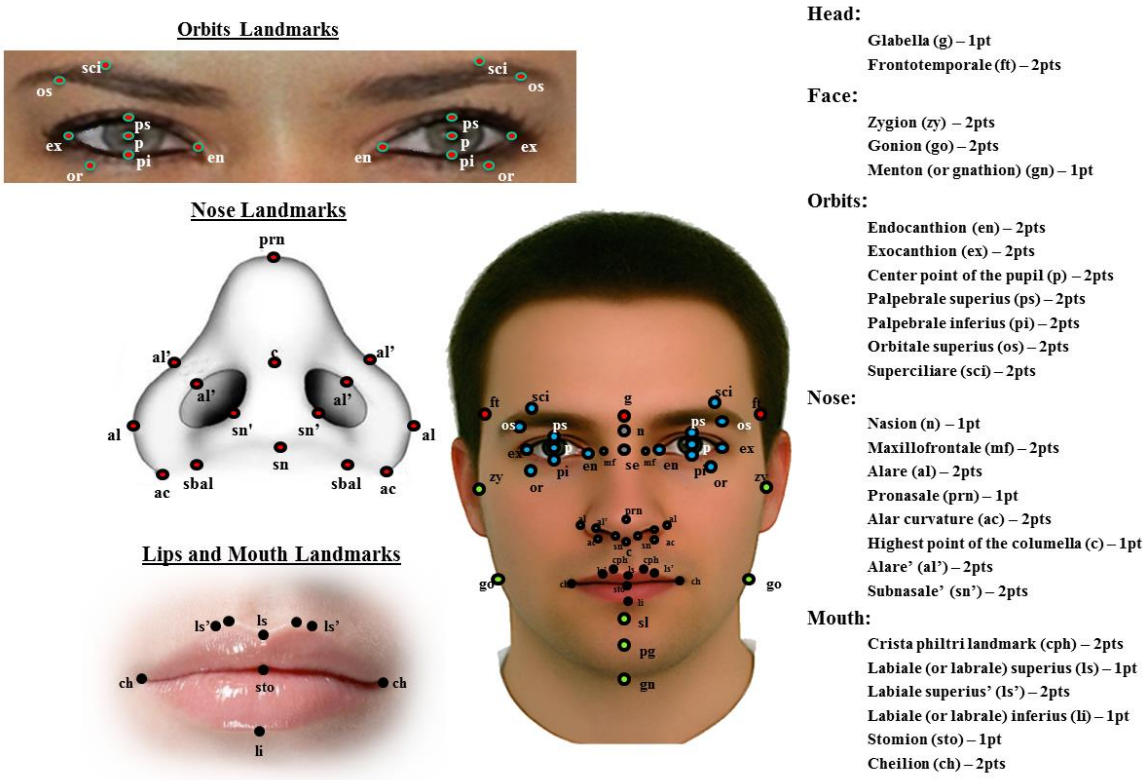


FIGURE 4 - Anthropometric features/landmarks of the face [32]

As automatic face recognition is to mimic human face recognition, in at least the ultimate goal of people identification, adhere is made, to the extent possible, to standard terminology used in human facial biometric and physiological literature.

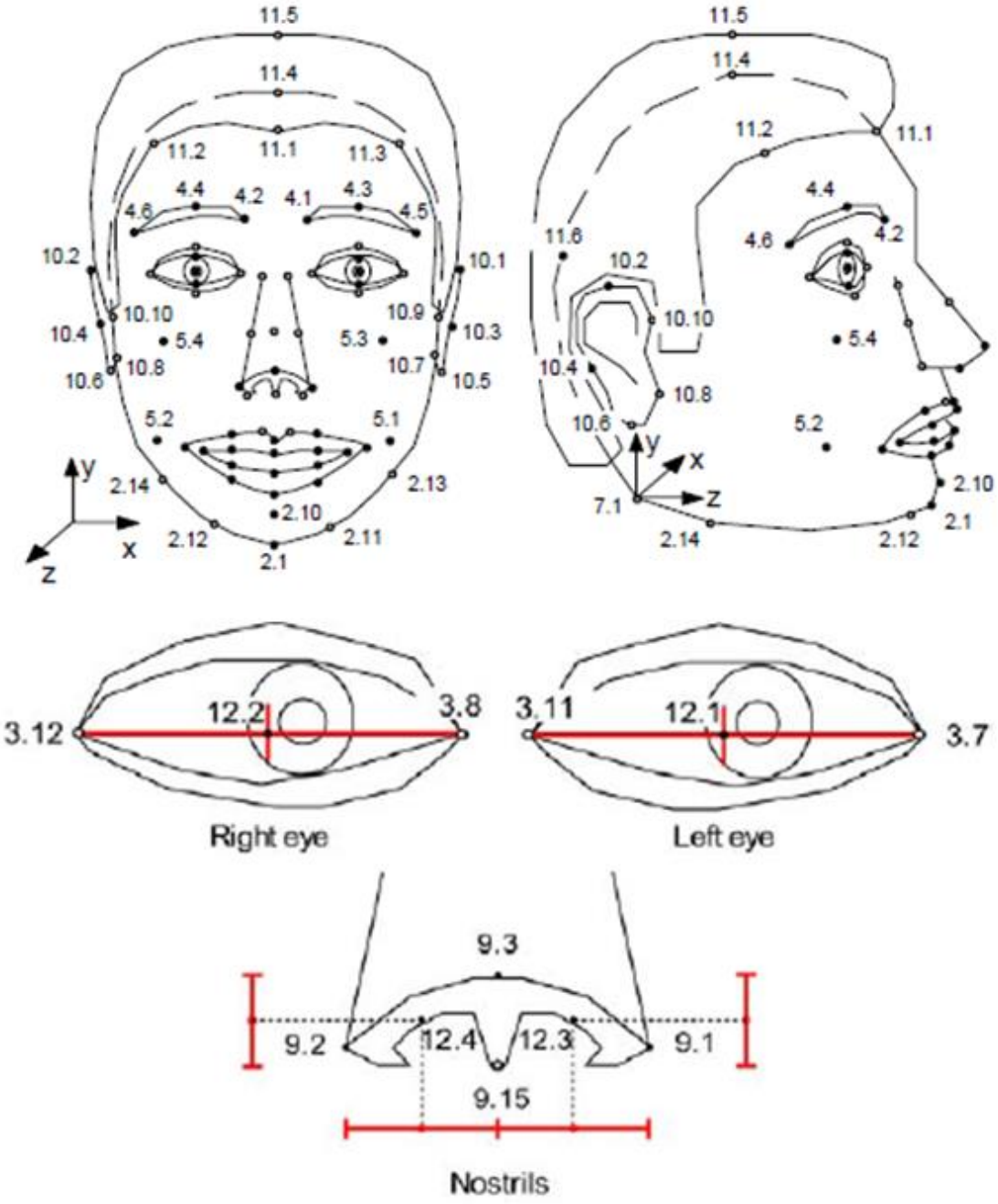


FIGURE 5 - ISO/IEC 14492-2 code for facial feature points [33]

Indeed, the anthropometric landmarks are used to guide the modeling process to generate the mesh of the “cropped” facial region which will be used in automatic face recognition.

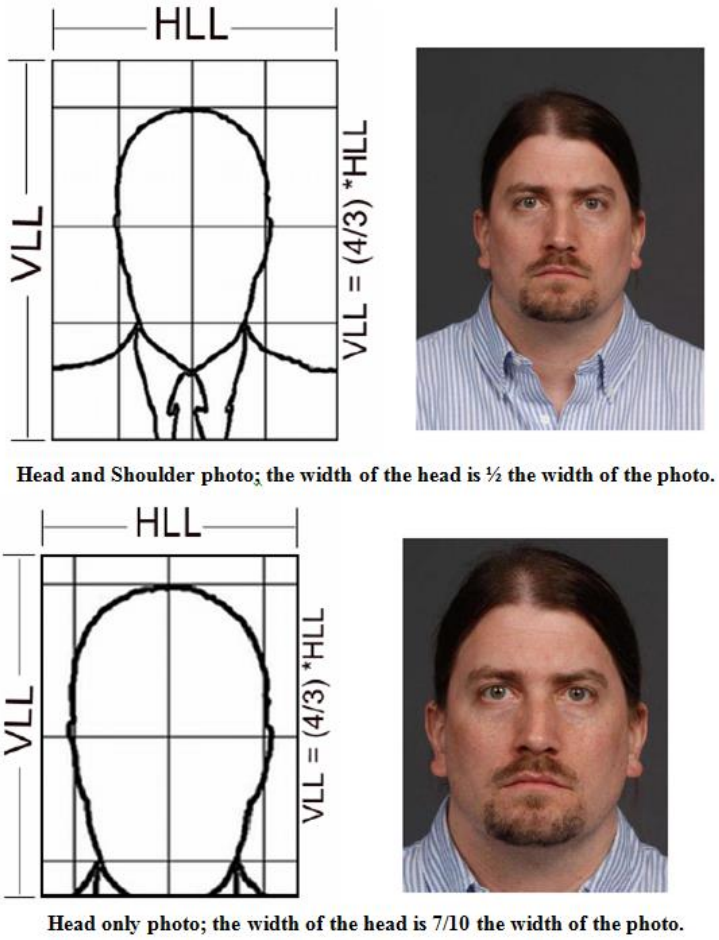


FIGURE 6 - ISO/IEC standard for head and shoulder and head only photos [33]

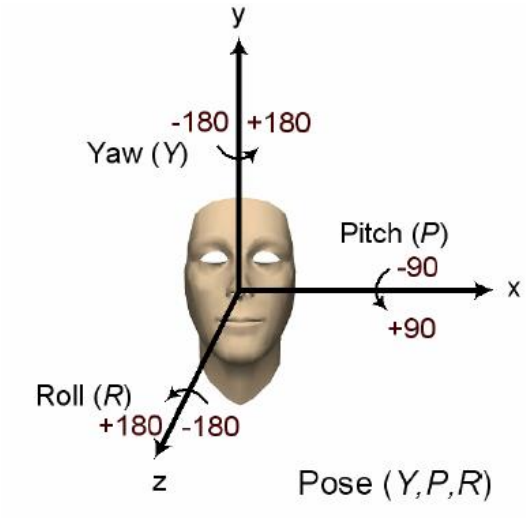


FIGURE 7 - The definition of pose with respect to frontal view [33]

B. Face Detection

Given an image of a scene, the purpose of the face detection step is to identify the facial region(s) in the image. Various approaches in the literature have been proposed. A technique that is natural to apply is the template matching method [4], which creates a template for faces, and sweeps it through the image in a raster fashion, and calculates *similarity* with corresponding segments in the image. A face is detected if the similarity exceeds a certain threshold. Among the similarity measures that are common is the cross-correlation, registration using mutual information and other methods. As can be expected, such approach will be expensive computationally.

Another approach is to use learning approaches. The Viola-Jones [5] method is very popular in face detection. Its main idea is the following: a) feature extraction of facial parts; b) train a classifier with various facial parts; and c) use a search approach to match the facial model with portions of the image, and mark those with high similarity value. The Adaboost algorithm is used to perform the training of the face detector, and a search method is used in execution of the detection. This chapter, highlights the components of the Viola-Jones algorithm, and refers to some of the modifications and enhancements that are being pursued, in order to improve the efficiency of the algorithm, especially, in the face recognition in the wild.

1. Viola-Jones Algorithm

The main idea is to scan a small window, reminiscent of a template, across the image, and analyze the content of the template using a series of primitive features that are sensitive to facial parts; e.g., eyes, nose, and lips. In image processing/analysis, usually window-based operations are performed at fixed template (window) and on multiple scales of the input image;

e.g., in wavelet analysis. The Viola-Jones algorithm does the opposite; i.e., changes the window size to multiple scales and rescan the input image. In each scale change, the size of the primitive features change accordingly (base template is 24×24 and gets enlarged to 30×30, 36×36, 40×40 then 48×48, etc., in scale of 1.25). To reduce time in calculating the features, they transformed the input image into a representation called the integral image, which makes the scanning invariant to scale; i.e., scans are performed at same number of operations.

Below is demonstration of the integral image and the primitive features within a template.

- **Integral image:**

The original $M \times N$ image $I(r,c), r \in [1,M], c \in [1,N]$ would be transformed to $I_i(r,c), r \in [1,M], c \in [1,N]$) such that a pixel at locations $(r,c)_i$ in the integral image will be sum of all pixels to the left and above of it. This "causal" representation codes the original information in the image in a suitable form for window-type computation of the primitive features, which will simply calculate difference between regions within the template, at different scales.

E**xample:**

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20

Original Image $I(r,c)$; $(r,c) \in [1,5]$

1	3	6	10
6	14	24	36
15	33	44	78
28	60	97	143
45	95	150	210

Integral Image $I_i(r,c)(r,c) \in [1,5]$

- **Primitive templates:**

Viola-Jones use templates reminiscent of those use in the Haar Transform, known as Haar features, or Haar-like features, which are sensitive to transitions in the image (i.e., nearly estimate the gradient). Five types are illustrated below:

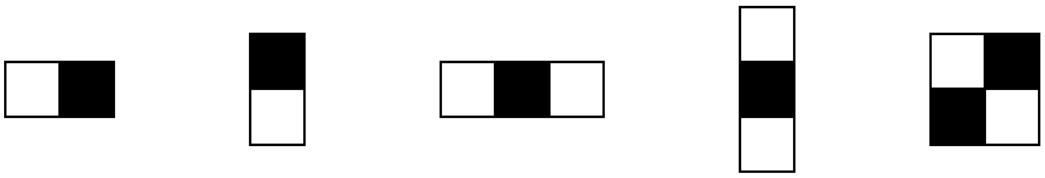


FIGURE 8 - Haar feature types computed within a template, at different scales, as it sweeps through an image. Computation is performed on the integral image

These features are calculated, at a given scale, as the difference between all pixels under the white region and the black region. The output of these computations is used to train a classifier.

Viola-Jones empirically selected base template of size 24×24 $(\text{i.e.}, 576 \text{ pixels})$. For each feature type, at all positions and scales, within this template, the numbers of features were calculated empirically to be around 160,000 (a lot more than the 576 region of support of the template).

- **Calculation of the primitive templates from the integral image**

The features can be computed by the integral image as follows:

$$\sum_{(r,c)\in ABDC} I(i,j) = I_i(D) + I_i(A) - I_i(B) - I_i(C) \qquad (2.1)$$

Therefore, the primitive features, at a given scale, will be calculated from the integral image by a series of computations from the corner values, which have been calcualted only once in the integral image (Figure 9).

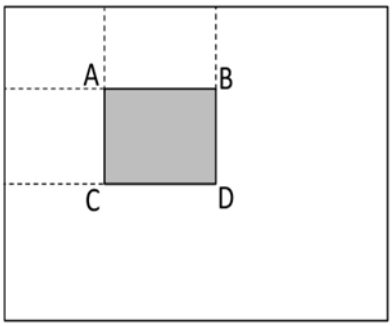


FIGURE 9 - Illustration of computation of areas of "causal" regions from the integral image. The shaded triangle will be equal to $\boldsymbol{D - (B + C) + A}$

Using the integral image, the calculations of the features is very straight forward and can be programmed efficiently. Each calculation results with a value which is compared to a threshold, and based on that, the pixel is declared "face" or "none face." When the template is over a face region, these features are expected to provide large values (because the eye, nose, lip regions carry discriminations). An approach is needed to get the "face" features fast from among the many features to be calculated (again, about 160,000 features within a 24×24 region – with increasing scales, the same number of calculations remains, thanks to the integral image). The classification approach that will decipher "face" features quickly is a modification of the Adaboost algorithm, which is described next.

- **Adaboost classification**

Viola-Jones approach for classification is as follows: Let $X = x(r, c) \in [1,24]$ represent that base region or scaled version of it. For each pixel in X, calculate the features as defined before.

Let the decision of these classifications be y; hence, pairs of decisions can be generated as follows: At each pixel $x(r, c)$ and for a feature f (one of five types above) one can obtain a decision y; which will conclude that the regions under the template is a face candidate or not. This can be written as:

$$h(X, f, p, \theta) = \begin{cases} 1 & if\ pf(X) > p\theta \\ 0 & Otherwise \end{cases} \quad (2.2)$$

where f is the applied feature, p is the polarity and θ is a threshold. As expected, so many decisions will be performed, majority would be inconclusive (weak), and the Adaboost algorithm consolidates these "weak" classifiers. The Adaboost algorithm "Adaptive Boosting," is a machine learning algorithm formulated by Yoav Freund and Robert Schapire, 1995 (see their 1997 publication [34]). Its optimality has been studied in the machine learning literature, and requires good training dataset. Viola and Jones adapted the algorithm for face detection. It is highlighted in Table I on the next page (see Viola-Jones, [5]).

As stated before, Viola and Jones run the basic classifier using templates of larger scales than the base scale of 24×24, each time they enhance the quality of the decision by eliminating non-faces. The overall decision approach is known as "Cascaded Classifier". The Viola-Jones approach is trained over thousands of "face" and non-face images. In the literatures, there are considerable numbers of cropped face and non-face images suitable for training the algorithm, in order to obtain the optimum set of features, parity (p) and thresholds (θ)weights. In summary, the Viola-Jones algorithm performs face detection using sliding window of region at different scales, and in each scale a process of no-face elimination is performed. As the number of computations is huge (fixed per scale), the weights of the Adaboost classifier are obtained off-line over tens of thousands of faces and no-face images (usually the number of no-face images is

much higher than the face images). Better training results when the face images contain lots of varieties, including pose and intensity variations. An implementation of Viola-Jones exists on OpenCV, and has been adapted in the BOSS project [19].

TABLE I: THE MODIFIED ADABOOST ALGORITHM

1. Given example images and decisions: $(X_1, y_1), (X_2, y_2), \cdots (X_n, y_n)$, where $X_k, k \in [1, n]$ are the regions under the sliding template. Let $y_1 = 1 \; or \; 0$, corresponding to the decision of face/no face.

2. Initialize the weights $w_{1k} = \begin{cases} 1/2m & if \; y_1 = 1 \\ 1/2l & if \; y_1 = 0 \end{cases}$, where m and l are the number of positive (face) and negative (no face) decisions.

3. For $t \in [1, T]$ do:
   a. Normalize the weight $\frac{w_{tk}}{\sum_{j=1}^{n} w_{tj}} \rightarrow w_{tk}$
   b. Select the best weak classifier with respect to the weighted error
   $$\varepsilon_t = min_{\{f,p,\theta\}} \sum_k w_k |h(X_k, f, p, \theta) - y_k|$$
   c. Define $h_t(X) = h(X, f_t, p_t, \theta_t)$, where f_t, p_t, θ_t are minimizers of ε_t.
   d. Update the weights: $w_{t+1,k} = w_{tk} {\beta_t}^{1-e_k}$, where
   $$e_k = \begin{cases} 0 & if \; X_k \text{ is correct (face)} \\ 1 & if \; X_k \text{ is not correct} \end{cases}$$
   and $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$

4. The final – strong – classifier is
$$C(X) = \begin{cases} 1 & if \sum_{t=1}^{T} \alpha_t h_t(X) \geq 0.5 \sum_{t=1}^{T} \alpha_t \\ 0 & Otherwise \end{cases}$$

Where $\alpha_t = \log \frac{1}{\beta_t}$.■

The output of the Viola-Jones algorithm is facial regions, which need to be cropped further to highlight the region that carries the most discriminatory information in the face (region between the chin and eyebrows).

C. Face Representation

The faces, output of the detection stage, may be represented by various methods. As an image, a full representation may be through the gray level values. However, this is not robust due to size and the degree of redundancy in the facial information. Various descriptors have been proposed to describe the feature of the face image; especially those belonging to the nose, eye, lips regions, which carry the most of the discriminatory information in the face. Among these descriptors the Linear Binary Patterns (LBP) [9], Scale Invariant Feature Transform (SIFT) [10] and the Speed Up Robust Features (SURF) [11] descriptors. Image matching algorithms consist of three major parts: feature detector, feature descriptor, and feature matching.

This section describes some of the feature detectors and descriptors common in image analysis. Figure 10 shows a test image used to evaluate these object descriptors.

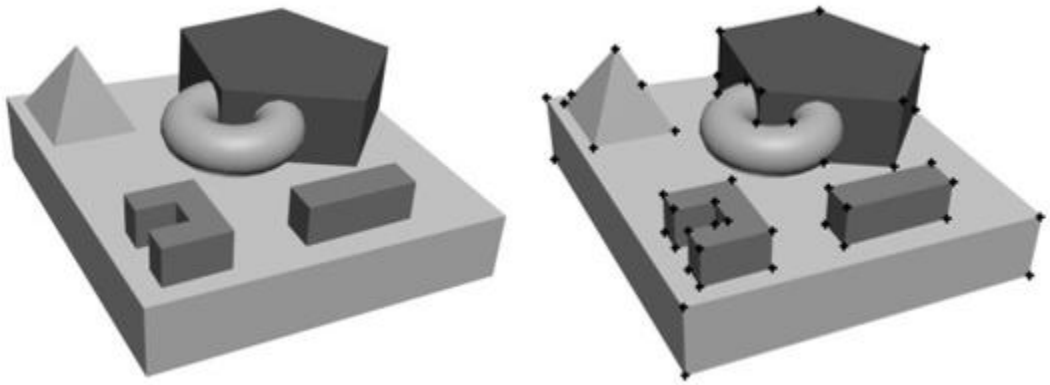


FIGURE 10 - Test image and keypoints, used to test object descriptors

1. Multi-Resolution Local Binary Pattern (LBP)

The Local Binary Pattern (Ojala et al., 2002 [9]) is an operator invariant to monotonic changes in grayscale and can resist illumination variations as long as the absolute gray-level

value differences are not badly affected. The original operator labeled the pixels of an image by thresholding the 3×3 neighborhood of each pixel with the center value and considered the result as a binary number. At a given pixel position (x_c, y_c), the decimal form of the resulting 8-bit word is given by the following equation: $LBP(x_c, y_c) = \sum_{i=0}^{7} s(I_i - I_c)2^i$; where, I_c corresponds to the center pixel (x_c, y_c), I_i to gray level values of the eight surrounding pixels and function $s(\cdot)$ is a unit-step function.

The LBP operator was extended to a circular neighborhood of different radius size to overcome the limitation of the small original 3×3 neighborhood size failing to capture large-scale structures. Each instance is denoted as (P, R), where P refers to the equally spaced pixels on a circle of radius R. The parameter P controls the quantization of the angular space and R determines the spatial resolution of the operator. An LBP pattern is considered uniform if it contains at most two bitwise transitions from 0 to 1 and vice-versa, when the binary string is circular. The reason for using uniform patterns is that they contain most of the texture information and mainly represent texture primitives. The operator is derived on a circularly symmetric neighbor set of P members on a circle of radius R denoting the operator as LBP_{PR}^{u2} . In the multi-resolution analysis the responses of multiple operators realized with different (P, R) are combined together and an aggregate dissimilarity is defined as the sum of individual log-likelihoods computed from the responses of individual operators. The notation LBP_{PR}^{u2} used here refers to the extended LBP operator in a (P, R) neighborhood, with only uniform patterns considered.

Figure 11 shows readings of the LBP for some keypoints on the test image in Figure 10 under the effect of blur, noise, rotation and scale. In general, the LBP descriptor works well when the neighborhood around the keypoints have reasonable texture content.

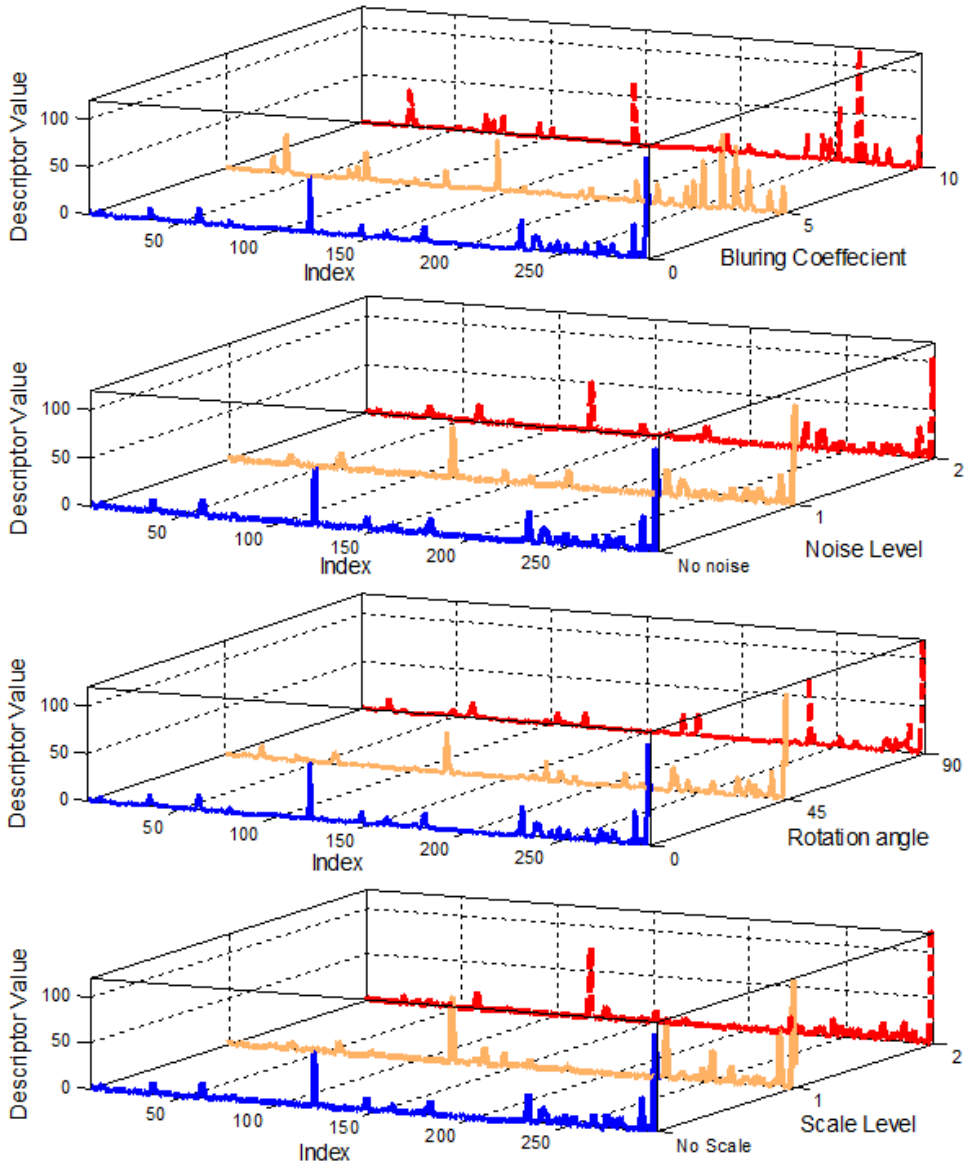


FIGURE 11 - The plot of the LBP descriptor performance on the test image under different blur, noise, rotation, and scale levels at the same selected point on transformed images

2. Scale Invariant Feature Transform (SIFT)

As detailed in Lowe, 2004 [10], SIFT consists of four main steps: (1) scale-space peak selection, (2) keypoint localization, (3) orientation assignment, (4) keypoint descriptor.

- **Scale space selection:**

The scale space $\mathbf{L}(\mathbf{x}, \sigma_s)$ is constructed by the linear convolution of the image $\mathbf{I}(\mathbf{x})$ with a cylindrical Gaussian kernel $\mathbf{G}(\mathbf{x}, \sigma_s)$ which can be viewed as a stack of 2D Gaussians one for each band. The scale is discretized as $\sigma_s \in \{k^s\}$ where $k = 2^{1/3}$ and $s = \left\{-1, 0,1,2, \dots, \frac{\log(s_{max})}{1/3log2}\right\}$. Scale-space extrema detection is perfromed through searching over all scales σ_s and image locations $\mathbf{x} = \{(x, y)\}$, in order to identify potential interest points which are invariant to scale and orientation. This can be efficiently implements using Difference-of-Gaussians $\mathbf{D}(\mathbf{x}, \sigma_s)$ which takes the difference between consecutive scales, i.e. $\mathbf{D}(\mathbf{x}, \sigma_s) = \mathbf{L}(\mathbf{x}, \sigma_s) - \mathbf{L}(\mathbf{x}, \sigma_{s-1})$, where for a spectral band b , a point $\mathbf{x}$ is selected to be a candidate interest point if it is larger or smaller than its $3 \times 3 \times 3$ neighborhood system defined on $\{D(\mathbf{x}, \sigma_{s-1}; b), D(\mathbf{x}, \sigma_s; b), D(\mathbf{x}, \sigma_{s+1}; b)\}$, where σ_s is marked to be the scale of the point $\mathbf{x}$.

This process leads to too many points some of which are unstable (sensitive to noise); hence removal of points with low contrast and points that are localized along edges is accomplished.

- **Keypoint localization**:

In order to obtain a point descriptor which is invariant to orientation, a consistent orientation should be assigned to each detected interest point based on the gradient of its local image patch. Considering a small window surrounding x, the gradient magnitude and orientation

can be computed using finite differences. Local image patch orientation is then weighted by the corresponding magnitude and Gaussian window. Eventually the orientation is selected to be the peak of the weighted orientation histogram.

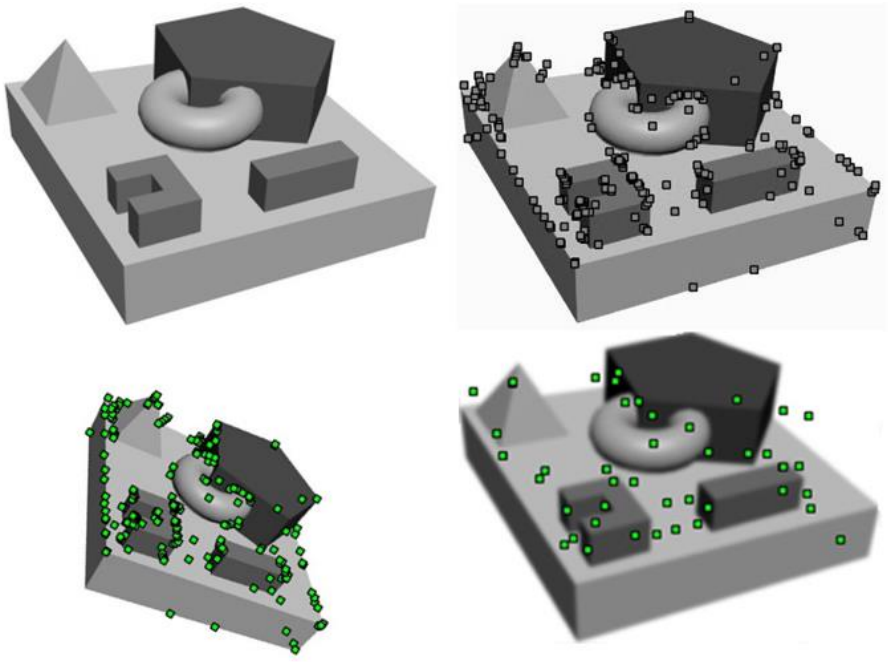


FIGURE 12 - Examples of keypoints detection using the SIFT detector with moderate rotation and blurring. The original image is upper left

- **Building a point descriptor:**

The process of building a descriptor around a key point is similar to orientation assignment. A 16×16 image window surrounding the interest point x is divided into sixteen 4×4 sub-window, an 8-bin weighted orientation histogram is computed for each sub-window, ending up with $16 \times 8 = 128$ descriptors for each interest point. Thus each detected interest point can now be defined at location, specific scale , certain orientation θ and a descriptor vector as $x = \{x, y, \sigma, \theta, d\}$.

Figure 13 shows the plot of the 128 values of the SIFT descriptor under different blur, noise, rotation and scale levels at the same selected point on transformed images in Figure 12.

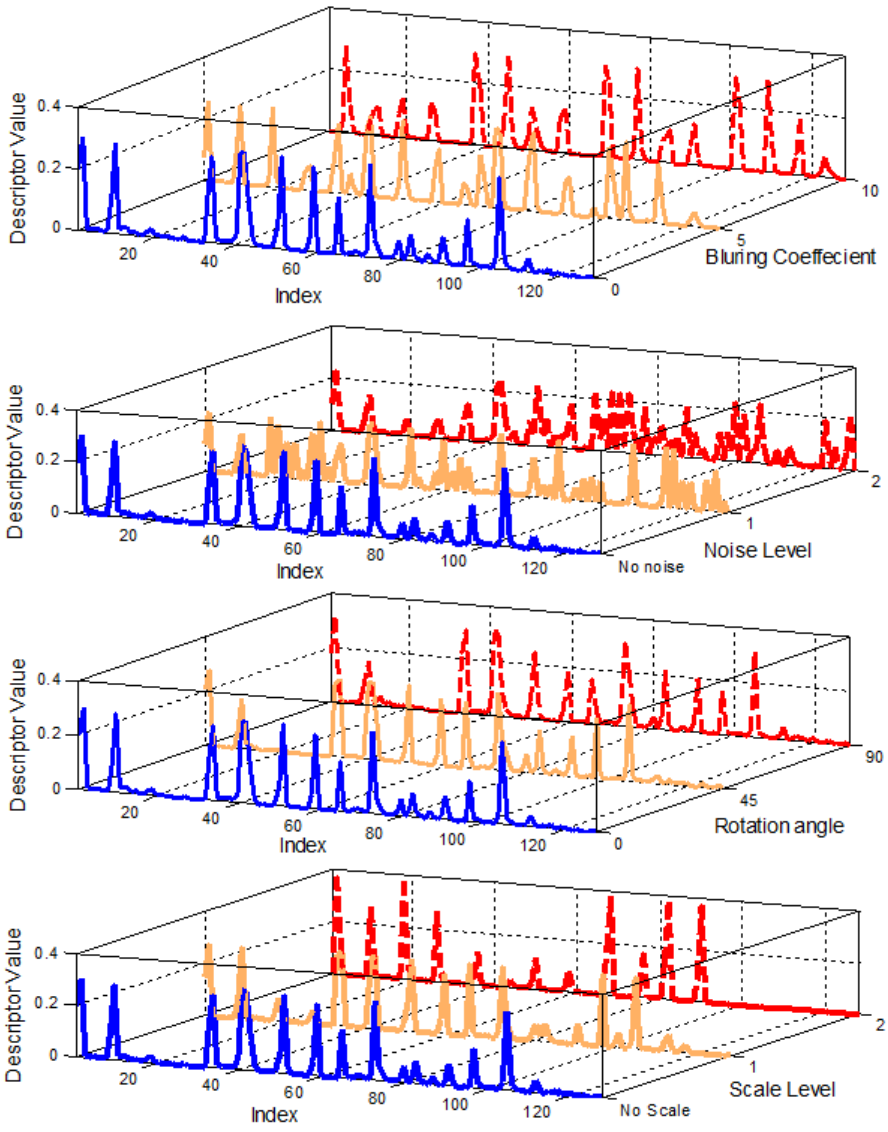


FIGURE 13 - SIFT descriptor under different blur, noise, rotation and scale levels at same selected point on transformed images

- **Interest point matching**:

Interest point matching is performed to provide correspondences between the given images. Two points $\mathbf{x}_i^t$ and $\mathbf{x}_j^{t+1}$ with SIFT descriptors $\mathbf{d}_i^t$ and $\mathbf{d}_j^{t+1}$ are said to be in correspondence, if:

$$d_{L_2}\left(\mathbf{x}_i^t, \mathbf{x}_j^{t+1}\right) = \sqrt{\left\|\mathbf{d}_i^t - \mathbf{d}_j^{t+1}\right\|^2} \text{ is minimum.}$$

This measure is computed as by:

$$d_{L_2}\left(\mathbf{x}_i^t, \mathbf{x}_j^{t+1}\right) = \left(\sum_{k=1}^{128} \left|d_{ik}^t - d_{jk}^{t+1}\right|^2\right)^{1/2}.$$

3. The Speeded-Up Robust Features (SURF)

The (SURF) descriptor (Bay et al., 2008 [11]) is a distribution of Haar-wavelet responses within the neighborhood of interest. The SURF descriptor consists of several steps; a square region is constructed around the interest point and oriented either in a rotation invariant method, where the Haar-wavelet response in the x − and y− directions are computed and weighted with a Gaussian centered at the interest point, or a non-rotation invariant method. The wavelet responses in both directions are then summed-up over each sub-region. The total number of descriptors for each point is 64. SURF uses mainly the texture information concentrated around interest points. Principle component analysis (PCA) and linear discriminate analysis (LDA) are used to project the extracted SURF descriptors to a low-dimensional subspace where noise is filtered out.

A plot of the 64 values of the SURF descriptor under different blur, noise, rotation and scale levels at the same selected point on transformed images is shown in Figure 14.

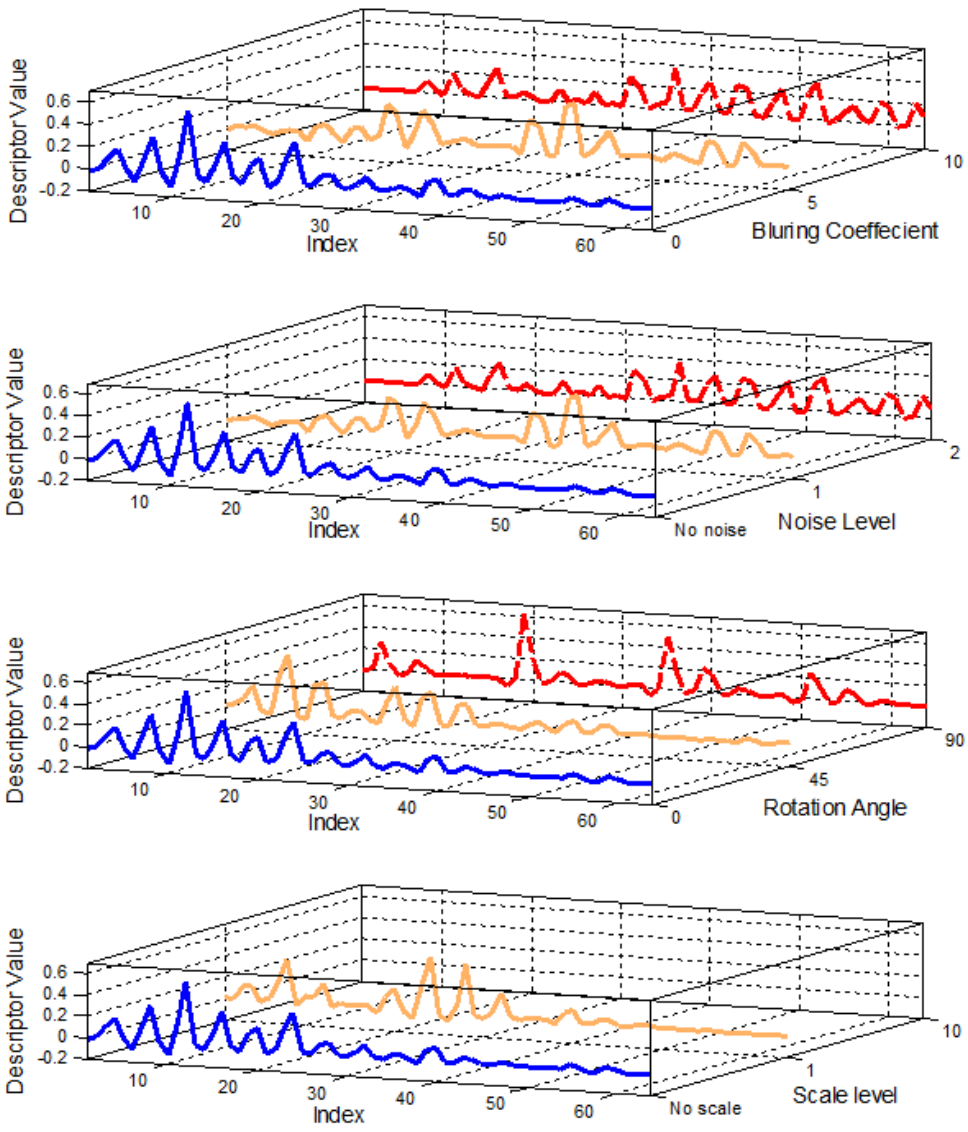


FIGURE 14 - The plot of the SURF descriptor under different blur, noise, rotation and scale levels at the same selected point on transformed images

The comparison of the three descriptors is shown in the following subsections. In general, the LBP works better with high textural contents, whereas the SIFT provides better performance with robust definition of keypoints.

4. Performance Evaluation on Test Images

The test image in Figure 12 is used to test the performace of the three descriptor (SIFT, SURF, and LBP). 46 points keypoints were selected manualy from the original image. The location ground trouth location of these points are calculated on every transformed image based on the transformation applied to generate this image.The descriptors are clculated at these points for all the images. The number of correct matched points are used as an evaluation criteria. The results are shown in Figure 15. The LBP showed a more robust performance with respect to noise, while the SIFT was more robust to rotation.

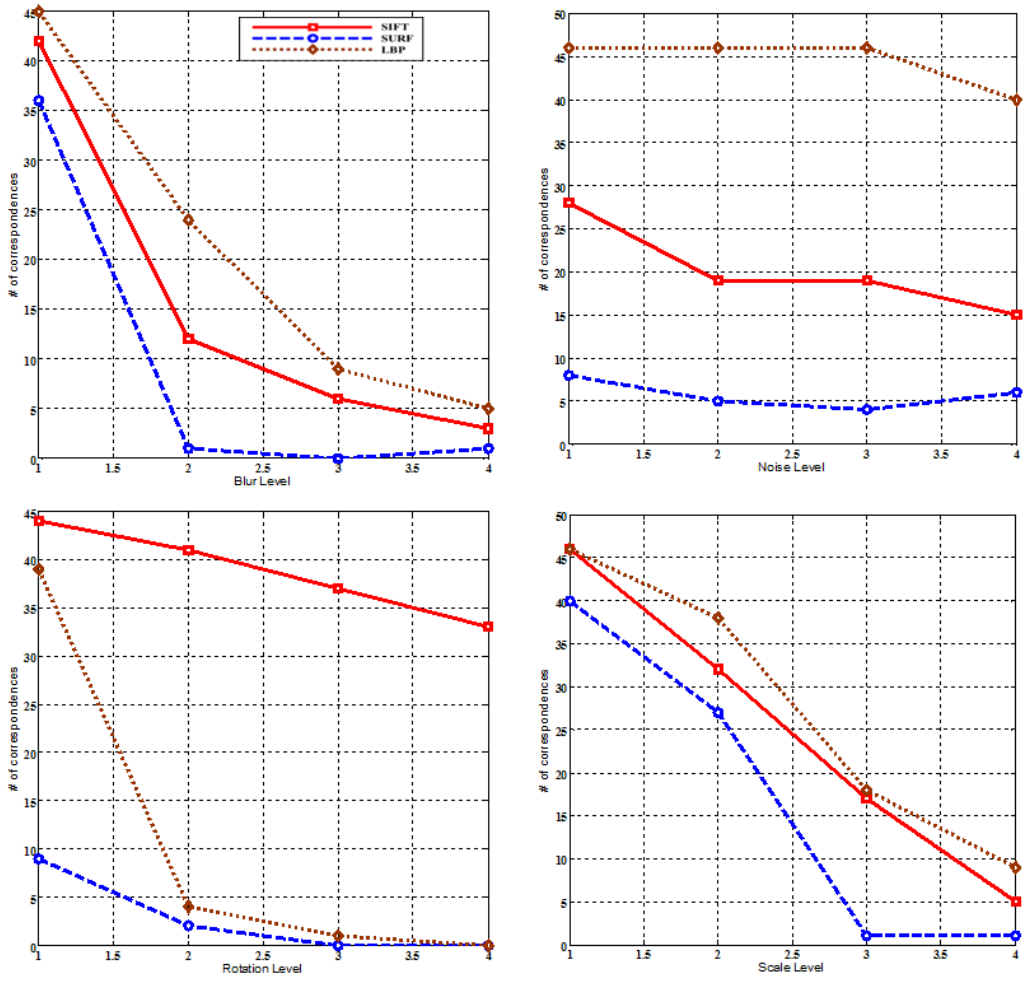


FIGURE 15 - The number of correspondences under different blur (upper left), noise (upper right), rotation (lower left) and scale levels (lower right) for the SIFT (solid curves), SURF (dashed curves), and LBP (dotted curves) descriptor

The above descriptors have been used for face recognition research by a great many researchers. For example, Ahonen et al., 2006 [35] used the LBP descriptor for face recognition, Bicego et al., 2006 [36], used the SIFT algorithm, while Dreuw et al., 2009 [37] used the SURF descriptor. In the BOSS project, both the LBP and a version of the SIFT descriptor are used (e.g., [38][20]).

D. Face Recognition

The gallery is stored in the representation of choice in a data structure that is efficient for search. This is performed offline. Given a candidate face (probe), detected by the face detector, its representation is computed. The recognition process becomes a simple comparison between the representation of the probe and the gallery. The recognition is declared based on ranking scores in the matching algorithm. Various efficient search methods are used to expedite search, especially when the galley is large.

Ignoring the age issue for a moment, the representation of data for a typical face recognition system will have pose, illumination, and expression (PIE) variations.

Given a gallery (database) of subjects, a pictorial representation may be as shown in Figure 16. A typical recognition strategy, as stated above, is formed of three steps: face detection, facial information representation, and matching, as illustrated in Figure 17. We briefly discuss each step below.

Figure 16 highlights two presentations for the images forming a gallery. The upper part of the figure shows that images will be represented by row (or column) concatenation. The bottom is an illustration of a general PIE representation, where the function $f_s(p, i, e)$ represents

the pose, illumination and expression for each subject. Such representation can be used in various computational scenarios to exploit the redundancies involved in the facial information (all faces have two eyes, one nose, lips, and two cheeks) and a specific number of features have been shown to hold the major discriminatory power (around the tip of the nose, corner of the lips and eyes regions). We will examine the issues of facial feature extraction and recognition later on in the thesis.

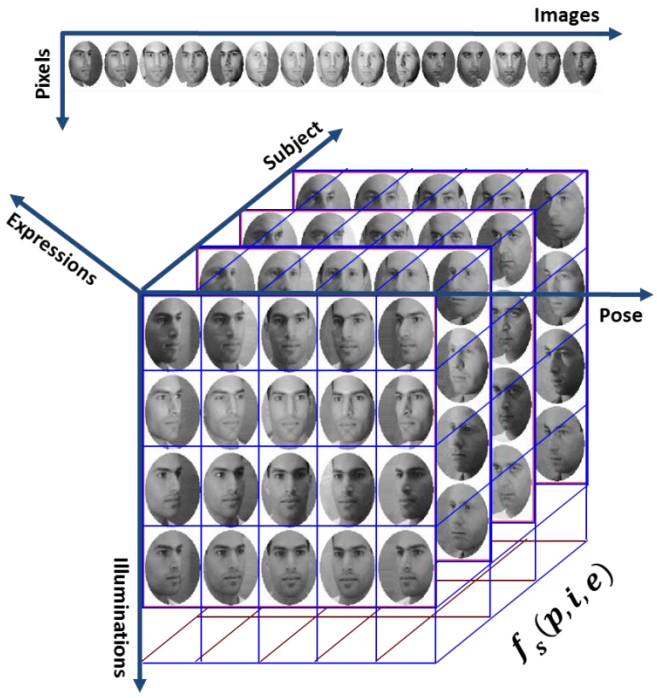


FIGURE 16 - Gallery arrangement with pose, illumination and expression (PIE) variations (adapted from Terzopoulos [3])

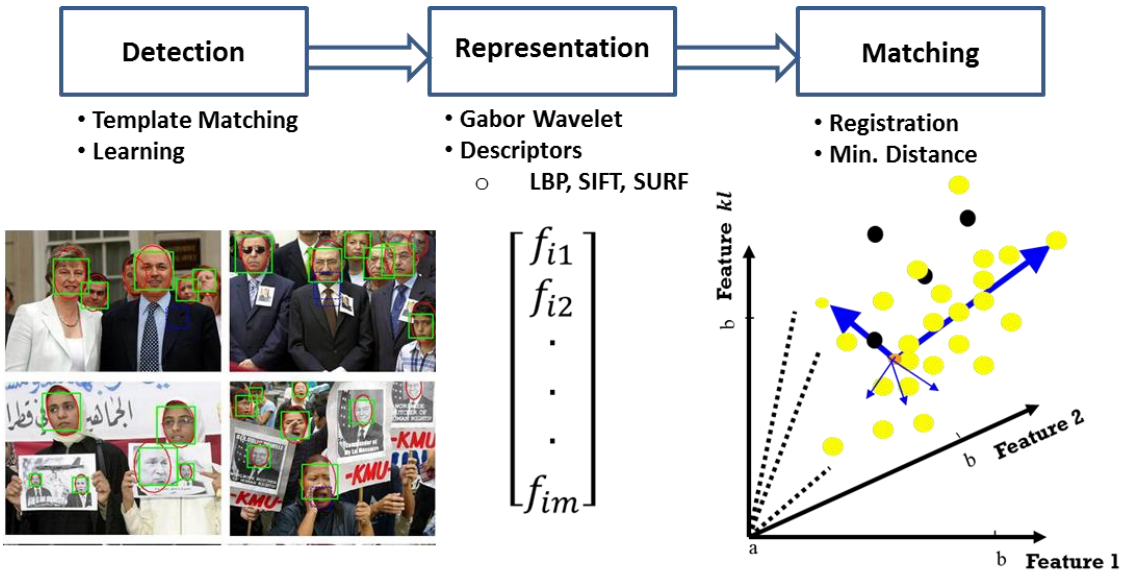


FIGURE 17 – Basic components of face recognition

E. A-PIE Face Recognition

There is no approach that fits all scenarios of image acquisition, and among the vast literature that exists on face recognition, we cannot pick a method that is optimal. Hence, a vast and comprehensive listing of methods would be a futile effort. Instead, we refer to sample literature that covers the A-PIE facial biometrics; many of which focuses on only a particular aspect of the problem.

1. Age Models

Aging affects the human face, in size, texture and overall appearance. Therefore, it is expected that aging will affect the accuracy of automated face recognition, especially is pictures were taken years apart. This is the conclusion of Ling et al., 2007 [13]. However, they also point out that "experiments show that, although the aging process adds difficulty to the recognition task, it does not surpass illumination or expression as a confounding factor." Wang et al., [14]

simulated the effects of aging on face recognition, while Sue et al., 2006 [15] discuss a model to simulate the aging. Park, et al., 2010 [16] proposed a 3D aging modeling technique to compensate for the age variations to improve the face recognition performance. The technique adapts view-invariant 3D face models to a given 2D face aging database, on which recognition is performed.

2. Pose-Invariance

Dislodging the effects of illumination and expression, pose-invariant face recognition has been shown to be possible, even in systems performing face recognition at a distance. For example, Mostafa et al., [2012] have developed two approaches for pose-invariant face recognition at a distance. The first one is called dynamic weighting of facial features [17]. In this approach, the similarity measure between the face signature of the probe image (query image) and face signature of gallery images is the sum of similarity measures of feature vectors of the patches around facial feature points. Since some facial feature can be partially occluded with head pose angle, a dynamic weight for these facial features was proposed. Dynamic weights are assigned for each facial feature at each pose based on the overlapping scores which is based on the number of pixels in the patch in the frontal gallery image and captured pose image that are corresponded to the same vertices in the 3D of the person.

The second approach is a hybrid 2D-3D, where a 3D shape from the single frontal gallery face image is constructed for each person [18]. The 3D shape and the texture from gallery image for each person are used to synthesis other face images at different poses. The gallery in this approach consists of multiple images for the person at different poses that are generated from

single frontal pose face image. Then, the face image is represented by the appearance in patches around facial feature points. Therefore, we have multiple face signatures for each person.

Combinations of these two approaches were used for stereo-based face recognition with the BOSS project at the CVIP Lab [19]. Instead of using one training sample per person at frontal pose in the gallery, two images are used at frontal pose per person. The two images are captured simultaneously with stereo camera to enable us to construct 3D shape for the face using geometric stereo algorithms. This 3D shape with the texture from gallery face images are used to synthesis other face images at different poses to solve the pose problem [20]. This approach is similar to Hybrid 2D-3D approach. The difference is that 3D shape is constructed from single image. A comparison between the stereo-system for pose invariant face recognition and other proposed approaches from single is done to study the importance of geometric stereo face recognition.

One of the challenges in geometric stereo imaging, the two cameras should be pointed to the person at the same time. Once one of the cameras is decided to capture one subject, the other camera should be pointed to the same subject. This problem is called camera steering in camera network. A solution of this problem is proposed based on using human face biometric measures to infer an approximate estimate of the subject's distance to the first camera that can be used to steer the other camera [21].

3. Illumination Modeling

Illumination research is very popular in the computer graphics and the computer vision literature. In terms of face recognition, the pioneering work of Kriegman and Belhumeur [22] is a good building block. They addressed the following question: what is the set of images of an

object under all lighting conditions and pose? For the set of images of an object under variable illumination, including multiple, extended light sources and shadows, they proved that the set of n-pixel images of a convex object with a Lambertian reflectance function, illuminated by an arbitrary number of point light sources at infinity, forms a convex polyhedral cone in illumination domain (i.e., for column or row-concatenated representation of n pixels, the cone will be in R^n space). They also showed that the dimension of this illumination cone equals the number of distinct surface normals. Furthermore, the illumination cone can be constructed from as few as three images. In addition, the set of n-pixel images of an object of any shape and with a more general reflectance function, seen under all possible illumination conditions, still forms a convex cone in R^n.

Great many studies since then focused on simulation of the illumination cone, and various mathematical models were introduced for illumination, through rendering and synthesis (computer graphics perspective) [23] or image formation (computer vision perspective) [24].

Modeling the image formation process addresses the object surface characteristics, the camera and the light source. Elhabian and Farag, 2013 [25] developed an analytic formulation for low-dimensional subspace construction in which shading cues lie while preserving the natural structure of an image sample. Using the frequency space representation of the image irradiance equation, the process of finding such subspace is cast as establishing a relation between its principal components and that of a deterministic set of basis functions, termed as irradiance harmonics. Representing images as matrices further lessen the number of parameters to be estimated to define a bilinear projection, which maps the image sample to a lower dimensional bilinear subspace. This approach links the illumination model to irradiance; thus from a given

image we can synthesis multiple illuminations. Logical expansion of this work is to expand it into multiple poses.

4. Expression Modeling

This is by far the toughest part of facial biometrics. As the face contains scores of muscles, one would expect that the number of expressions would be too many; some are related to cultural and ethnic upbringing (e.g., [1]). The facial action coding system (FACS [27]), Izard, et. al., 1983, is a human-based system designed to detect such subtle changes in isolated facial features through viewing a videotaped facial behavior in slow motion and manually recording the FACS code of all possible facial changes which are referred to as actions units. FACS consists of 44 action units, where thirty are related to the contraction of a specific set of facial muscles and the other 14 are referred to as miscellaneous since their anatomic basis is not specified. Ekman and Friesen [26] proposed that specific combinations of FACS action units represent prototypic expressions of emotion, however, emotion-specific expressions are not part of FACS, they have a separate coding system such as the emotional facial action system (EMFACS [28]). Henceforth, FACS is purely descriptive coding system where there is no inferential information provided such as joy or anger.

The study of Tian, Kanade and Cohn, 2001 [29] performed analysis of facial expressions based on both permanent facial features (brows, eyes, mouth) and transient facial features (deepening of facial furrows) in a nearly frontal-view face image sequence. The system recognizes fine-grained changes in facial expression into action units (AU) of the Facial Action Coding System (FACS), instead of a few prototypic expressions. The authors used multistate face and facial component models for tracking and modeling the various facial features,

including lips, eyes, brows, cheeks, and furrows. During tracking, detailed parametric descriptions of the facial features were extracted. With these parameters as the inputs, a group of action units (neutral expression, six upper face AU and 10 lower face AU) are recognized whether they occur alone or in combinations. The system has achieved average recognition rates of 96.4 percent (95.4 percent if neutral expressions are excluded) for upper face AU and 96.7 percent (95.6 percent with neutral expressions excluded) for lower face AU.

Various computational studies for modeling of expression and for performing face recognition under expression variation have been introduced in the past decade. Some of these models are based on morphable (e.g., Blanz [30]) models active appearance (e.g, Theobald, et a;., 2007 [31]). The BOSS system of the University of Louisville enables group face recognition, and would be convenient prototype for facial expression analysis of a group [19].

F. The BOSS System

This section will discuss the BOSS system in terms of design, modes of operation, and software. Understanding the system's components is crucial to devise an evaluation procedure, which will be the subject of Chapter 3. The literature about BOSS exists in the forms of technical reports, evaluation meetings, conference proceedings, and other communiques that required major efforts to describe in a short concise format. The section will not dwell into years of efforts of many researchers and engineers involved in the BOSS system; rather, it will include only glimpses of these efforts as pertaining to the overall purpose of this thesis.

Digging through the design papers of BOSS revealed considerable number of documents that have been exchanged by the CVIP Lab, EWA Government Systems, and the Government

evaluators. In this section, specific technical reports that dealt with design of the BOSS will be referred to. The technical reports will be referred to by the dates they were published. Again, only bare minimum of essential details about the system will be described in this chapter.

1. Overall System Components

The BOSS (Biometric Optical Surveillance System) project builds on past developments at the CVIP Lab in the domain of biometrics and computer vision systems. At the heart of the BOSS project is a trilogy: image acquisition, leading to capturing objects in the field of view of the sensor; reconstruction, leading to mapping the captured objects into a form suitable for the final recognition step; which identifies the detected objects by correspondence with a dynamic database.

Specifically, 1) the acquisition step is based on parallel skin detection and multichannel tracking, in order to enable unambiguous facial detection of a group. 2) The reconstruction step will involve simultaneous statistical modeling of multichannel information in order to enable parallel sparse reconstruction of facial features for recognition. 3) The identification will employ parallel networks of search algorithms and state-of-the-art methods of database access using proper representation of facial information. The hardware include special lenses to allow maximum possible range of identifiable pictures of a group, a range sensor for calibration and focusing, an IR sensor to add additional biometric information to enhance the sensitivity and specificity of BOSS performance as measured by improvements in acquisition and identification. The system will allow imaging of humans under changes of lighting and various environmental conditions. In addition, the system will allow intelligent capturing and discrimination of subjects,

within a group, under dynamical conditions of typical activities. Figure 18 illustrates the main components of BOSS.

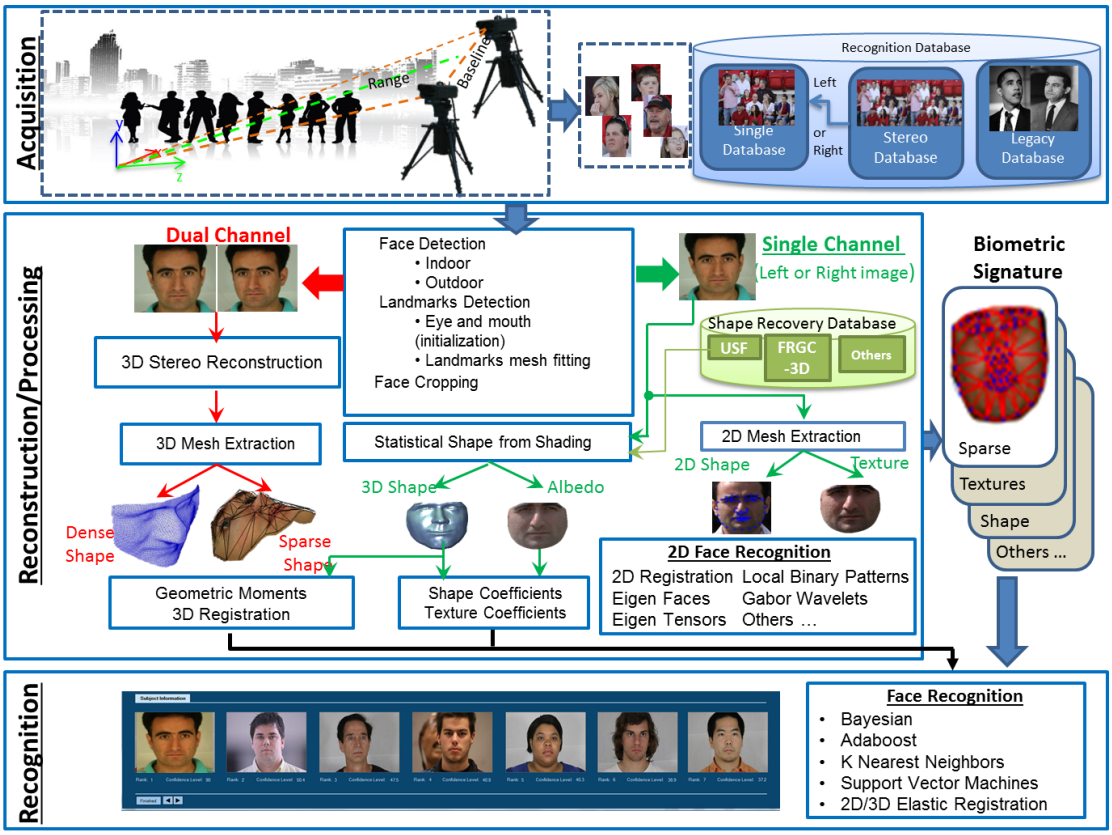


FIGURE 18 - The BOSS system components for performing face recognition at a distance. The system works on image-based (single channel) or stereo-based (dual channel) modes

BOSS is essentially three components as shown below; the three steps, not inherently serial: Acquisition, Processing and Recognition.

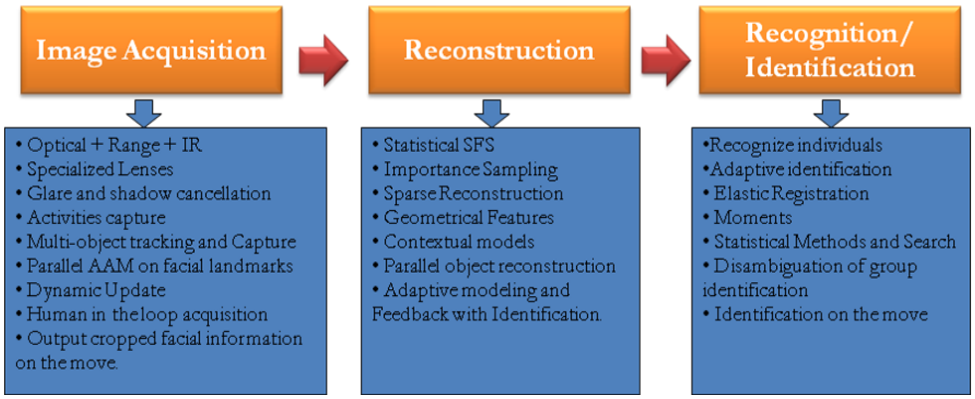


FIGURE 19 - Three components to the BOSS

- **Acquisition**:

Images may be captured by cameras or obtained from a database. Databases may be 2D images or 3D representations.

- **Processing:**

Cropped facial region is produced and two modes of processing may be conducted; Single channel (one image) or Dual channels (two or more images of the face, related to each other; e.g., a stereo pair).

A. Processing of single channel may be performed in two scenarios: 2D and 3D. *The 2D processing* produces shape and texture images, and the approaches for 2D Face Recognition may be the classical Eigen faces, Eigen tensors and various similarity measures that compare the processed image to a database of similar attributes. *The 3D processing* provides a 2D to 3D mapping using *a priori* information (e.g., a database of shape and texture information such as the University of South Florida database), which results in estimates of the shape and albedo using Statistical Methods (e.g., statistical shape from shading, spherical harmonics, etc.).

B. Processing of dual channel mode provides a direct metric shape estimate; e.g., depth from stereo followed by a 3D mesh fit to the surface, which can be sparse or dense. The features for recognition using shape information (obtained directly from dual channel processing or by estimation from the single channel processing) may provide various features for recognition, including geometric moments. The output of the Processing step provides the Biometric Signature, which includes texture, shape, and spares shape meshes.

- **Recognition:**

Is preformed based on the features extracted. Algorithms for recognition measure an optimum similarity between the biometric signatures and the reference databases.

2. Hardware

The BOSS System is used to collect images for identifying individuals. It consists of two sub-systems: the BCU (Biometric Collection Unit) and the REPS (Remote Processing System). A BCU consists of a high-resolution digital camera and a pan/tilt unit that is mounted to an adjustable tripod and is capable of subject tracking. The BOSS system is typically deployed in a stereo configuration of two BCU's (Figure 20). The full hardware equipment list is provided in Appendix A.

Each BCU collects a digital image of a subject and transmits that data to the REPS via fiber optic cable. The REPS is a high-end computer built from off-the-shelf components which runs software that transforms the image data into a 3-D biometrical signature of the subject. The signature can then be stored in a database and/or compared against existing signatures to return results for the closest matches.

The system can accept or acquire either a stereo-pair or a single digital image. The BOSS system requires a high-resolution imaging sensor with an available SDK (Software Development Kit). Other necessary features include live-view functionality, auto/manual focus capability, and an interchangeable lens mount. For a given focal length a greater resolution makes identification at longer ranges possible. An interchangeable lens mount ensures the system can be outfitted for a versatile range of standoff distances.

FIGURE 20 - BOSS System at a testing site, February 2012

3. System Modes of Operation

BOSS has two modes of operation: (1) Offline mode: where an offline database is constructed by BOSS setup and processed offline at the CVIP Lab. This database is divided into probes and gallery in order to assess the overall performance of the BOSS system. (2) Online mode: where a probe is captured and processed online then matched against a pre-determined gallery.

The following subsections discuss the process of data acquisition, database construction and a brief discussion of the 2D and 3D recognition strategies used.

4. Data Collection

Figure 21 illustrates a generic dual-channel (with one individual at a time) data acquisition setup. At a given roll, pitch and yaw angles, single and stereo-pair image sets are collected. This protocol can be employed in online and offline operating modes.

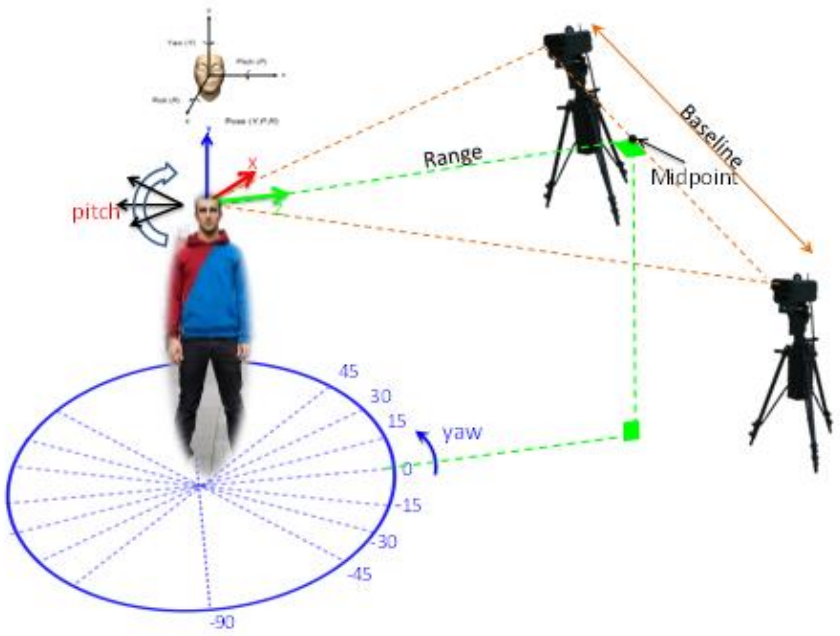


FIGURE 21 - A Dual-Channel data collection setup

TABLE II: RELATION BETWEEN A DISTANCE AND ITS CORRESPONDING BASELINE RANGE IN METERS

Distance	20	40	60	80	100
Baseline range	0.94 - 1.26	3.76 - 5.02	8.46 - 11.28	15.03 - 20.04	23.48 - 31.31

Since the scenarios intended for evaluating BOSS are intended to be flexible and real world in nature, hence, the algorithms need to be able to function as such. While it is virtually

impossible to perceive every possible scenario for pose, illumination and expression variations, and a representative ensemble of the "real world face recognition" may be generated by data acquisition at multiple yaw, pitch and roll angles, at various environmental conditions. Ideally, large number of subjects with variations in skin color, age, pose, illumination and expressions should be collected in order to establish the design thresholds for the system. The data structure used is similar to Figure 22.

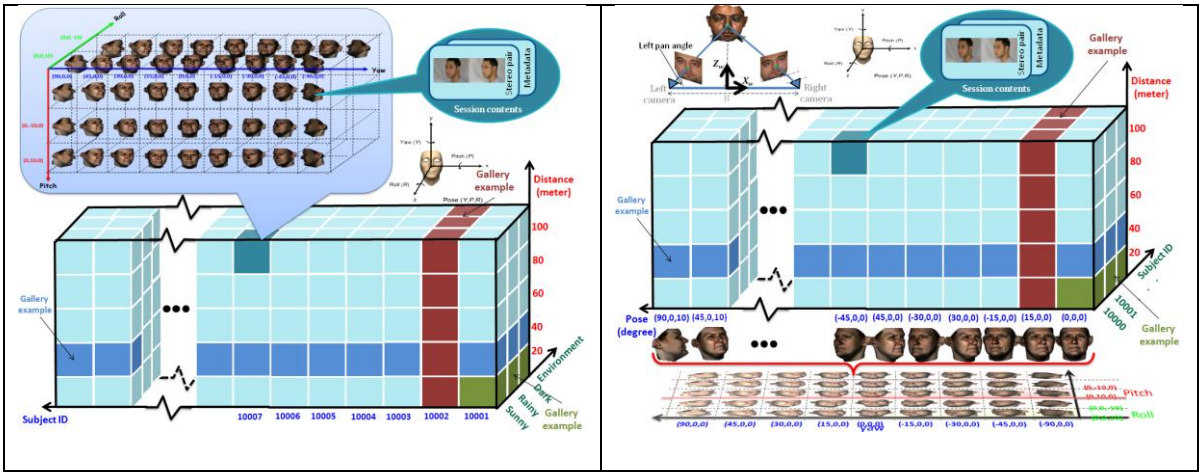


FIGURE 22 - Database arrangement for BOSS data collection used in system design

Metadata includes:

- the subject ID
- geometry of stereo setup (baseline B, pan and tilt angles for left and right units)
- distance
- pose
- Environment (e.g., outdoor status: cloudy, sunny or rainy, temperature, etc.)
- Biographic (e.g., name, gender, ethnicity, age, etc.)

The database is divided into a gallery and probes. As indicated in the BOSS terminology document, **gallery** refers to the collection of biometric representations of enrolled individuals in the database, whereas the **probe** is a biometric representation of an individual to be compared against the gallery.

According to the training and performance requirement, the gallery is chosen as follows:

- The set of the subject sessions for all participants at certain pose and certain distance (e.g., at distance 20 meter, and pose (0, 0, 0), as illustrated in green column in Figure 22).
- The set of the subject sessions for all participants at certain pose and for all distances (e.g., at distances 20-100 meter, and pose (15, 0, 0), as illustrated in brown slab in Figure 22).
- The set of the subject sessions for all participants at certain distance and some poses (e.g., at distance 40 meter, and all pose, as illustrated in blue slab in Figure 22).

While using all these combinations gives a more comprehensive database, a smaller database may be constructed and can constitute the shell of the comprehensive database.

Figure 23 shows a pictorial illustration of data collection. The figure illustrates the dilemma of data collection for design and testing of biometric system; the number of images is very large per individual! The robustness of the system comes in place if only few poses, per subject, are needed, and if the system is tolerant to small expressions and minor changes in illuminations.

Designing a system that is "invariant" to all A-PIE circumstances is simply unreachable; hence, the deployment scenarios should be exploited while designing the system.

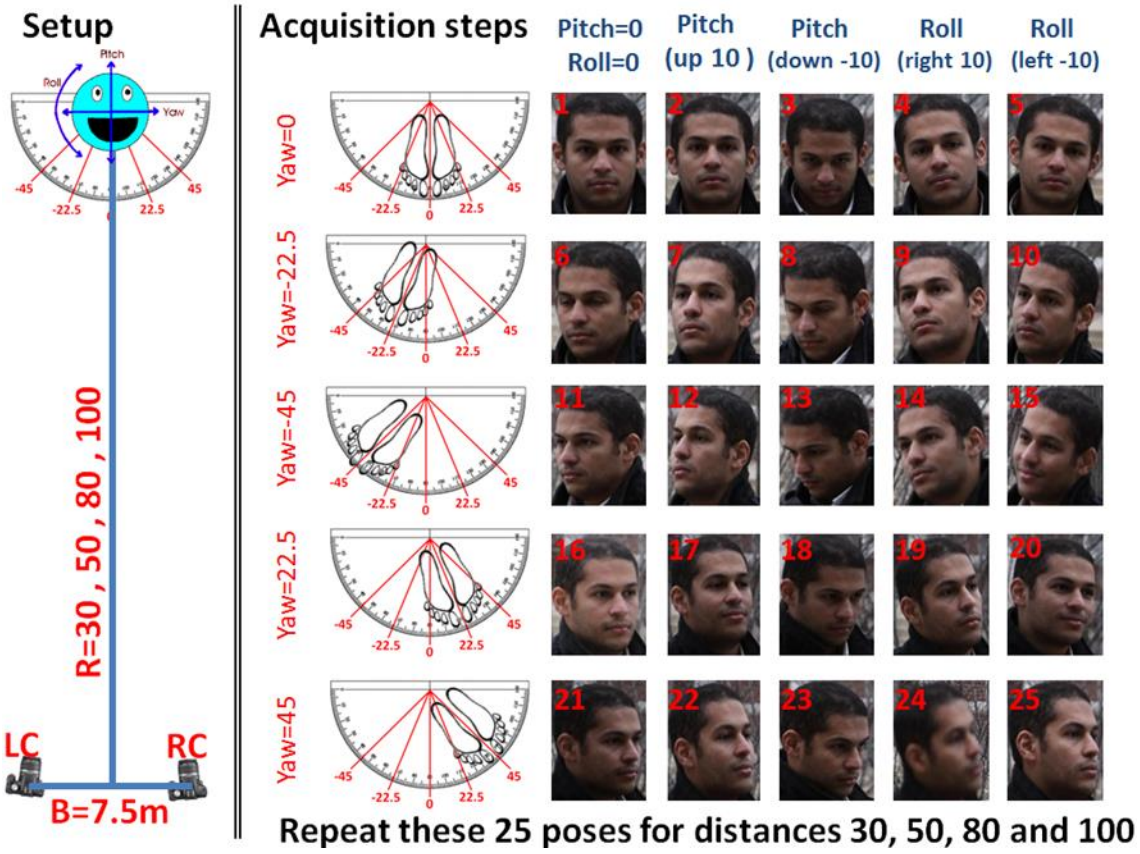


FIGURE 23 - Illustration of data collection per individual in design phase of BOSS

5. BOSS Algorithms

a. Face Detection and Cropping. This is performed by adapting the Viola-Jones algorithm to detect face candidates; each candidate is then cropped using a mesh generated by the active appearance modeling approach [19][42]. System starts by detection an individual in the field of view (FOV) of the camera, take an image, by the two cameras (right and left images). After the image acquisition (has faces and non faces), the Algorithm starts the face detection phase. The face detection phase uses the Viola-Jones algorithm in order to find the faces within the captured image and to output an image with these detected faces encased in a square. The

face detection phase uses the Viola-Jones algorithm for face detection which has a pre-processing step and a post processing step.

The algorithm works on analyzing an image using different scaling factors in order to detect possible faces. Each input image is subject to different scaling factors, a window is then scanned over the scaled image in order to specify an area where a possible face lies. The image is then scaled again in order to search for other possible face areas, as well as refine the previous found face area. By this, the face detection algorithm is capable to detect all the faces in the image, even if there are different face sizes in the image; such as when a subject is closer to the sensor than a subject who is further away. The left side of Figure 24 depicts this change in scaling factors. These ratios are needed to be adjusted based on the input image resolution. The original setting for the BOSS uses these scaling factors for the Canon EOS 7D images (1/6, 1/7, and 1/8 of original image); and these scaling factors for mug shot images (1, 1/2, and 1/3 of original image).

This algorithm is also used as a facial feature detector, such as an eye detector as well as a nose and mouth detector.

Once the faces and facial features have been found in the two images; the system passes these candidate faces to the next step which is used to reject the false positive samples. A scoring algorithm was developed in order to take the detected faces, from both the left and right image (captured by the stereo setup cameras) and rank their possibility of being an actual face. This ranking was achieved by adding the number of facial features (eyes and mouth) found in each candidate face in order to decide whether it in fact belonged to an actual face. If the candidate

face had a score of 2 or more (2 facial features found), than it is more probable that it is actually a face (Figure 24).

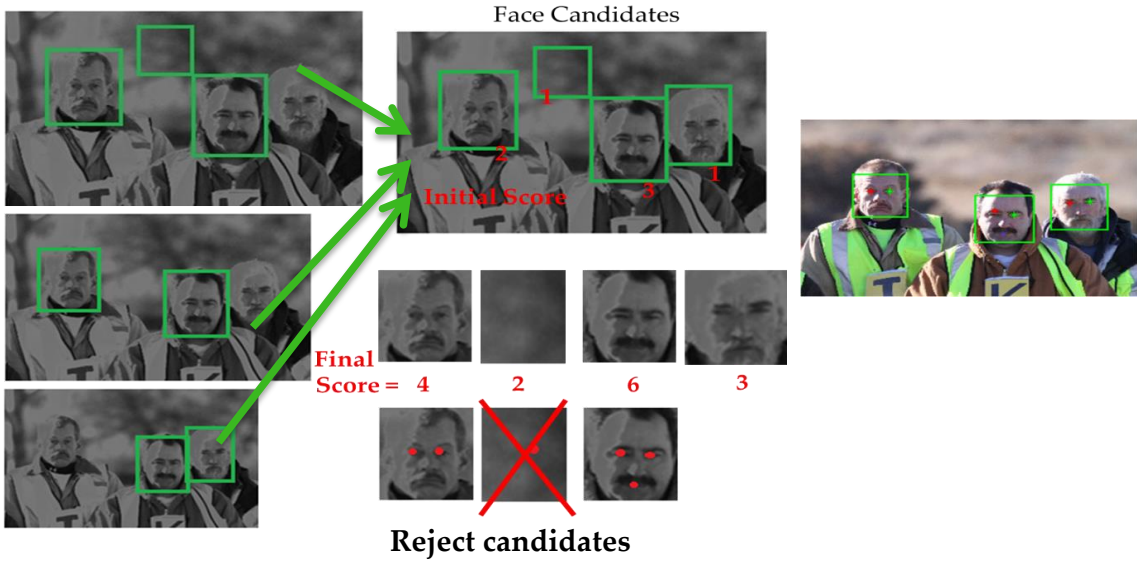


FIGURE 24 – Example of stereo setup ranking system in BOSS

Upon removal of the false positive candidates, the detected faces become an input to the facial feature detector, which will output facial feature points. Using the Adaboost classifier as a facial point detector, trained to find specific points, the system detects 9 points on the face along with some other steps for fitting of a global model to these points. The Active Shape Model (ASM) is then used to detect more points, making the initial 9 points into 68 points. These outputs of 68 facial feature points are consisted of such things as the eye corners, mouth corners, nose tip, and boundary points. There are various algorithmic details to carry out this step, including isolating the eye, nose and lips region (e.g., Farag et al, 2012 [19]).

b. Face Representation. The third step in the BOSS is to detect/reconstruct a signature around the previously found 68 points. Using the points found in previous step as an input, they are put through three different signature extraction techniques. These signature

extractors will be used to detect a feature vector, and this feature vector will be used to match between probe faces and enrolled faces. These three signature extraction techniques are composed of two 2D signature extractors along with one 3D signature extractor. The 2D signature extractors used are the Gabor wavelet signature extractor and the Local Binary Pattern (LBP) signature extractor; the 3D signature extractor is the sparse 3D points' reconstruction for 2D points. So the 68 points found previously are used to create a 3D reconstruction.

Another use of the Gabor signature found is in another false positive reduction step for the dual channel setup. Using the Gabor signatures found from the left and right images captured, the system is able to match candidate faces from the left image to faces in the right image in order to find which face in the left image corresponds to which image in the right image (Gabor signature from a face in the left image is used to compare with the Gabor signature from a face in the right image in order to find a match, which means they are the same face). The system is capable of removing false positives further by comparing a non-face that was detected in left image to the right image, if the right image did not detect this non-face as well, the system will discard it.

The BOSS has a database of enrolled subjects that stores the feature vectors from the three signature extractors described above. Once the three feature vectors are computed from each of the three signature extractors above, for the new probe subject input, it is passed to a minimum distance classifier to select the nearest neighbor for each subject. In other words, the system compares the feature vector for a probe subject with the feature vectors from the gallery (enrolled) pictures stored in the database.

c. Face Recognition. The current implementation of BOSS uses a minimum distance (k-NN) classifier. The feature vector per probe is compared to the entire database through a distance measure. The system subtracts the sparse 3D reconstruction feature vector of the probe from the feature vector of an individual in the database and sums the absolute value of the difference giving a distance measure (call the error/difference the "distance measure"). Similarly, the system also gets the distance measure for the Gabor feature vector and the LBP feature vector for the entire database.

The database is sorted based on the distance measure for each feature vector. Then combine the decision; get the decision from the Gabor +decision of LBP + decision of 3D points weighted with a vector. Weight for Gabor is 50%, LBP is 48%, and 3D is 2%. This will give you the final decision; based on this final decision you will sort the database from the closest subject to the probe to the furthest.

G. Summary

This chapter discussed some of the terminologies and standards of facial biometrics, and the major elements of the theory of face recognition, which is formed of a trilogy of steps: detection, representation and recognition. The common threads in the literature at the fundamental level are presented, without sinking into the details of the applications and the various competing algorithms. In particular, the chapter contains a concise description of the popular Viola-Jones algorithm for face detection, which produces candidate facial regions. A subsequent step is performed to crop the facial regions holding the discriminatory information. The later part of this chapter described the BOSS project developed by the CVIP lab in its entirety.

As unconstrained face recognition involves various uncertainties in the imaging process, the need for more accurate detection, representation and recognition will continue to persist.

As this thesis deals with evaluating an existing system by relaxing the sensors and the imaging scenarios, the immense theory and algorithms involved and the efforts to put them together into work, cannot go unnoticed. Even learning some of these methodologies and describing them in this thesis is an extremely difficult undertaking. In the subsequent chapters, the thesis will focus on the performance evaluation of BOSS.

III. BOSS EVALUATION

The previous chapter discussed the BOSS project as well as described the mathematical foundations and algorithms it uses. In this chapter, the performance of the aforementioned BOSS project will be discussed.

A. Performance Evaluation

The BOSS system was officially evaluated at a number of settings by a third party. The results of one setting are described in this section. The purpose is to study the evaluation process that BOSS was evaluated against in an open environment, face recognition at a distance practical scenario. Understanding this process will guide the evaluation procedure to be used when the BOSS software will be evaluated using low-resolution cameras, as described in Chapter 4.

The total number of images received from the baseline data taken in Washington State by PNNL was 178 stereo pair; 47 of them were excluded due to either the lack of ID cards or blurred images; i.e., 131 useful stereo pairs containing 11 different subjects were used in the testing. The header file of each image included the subject ID and the range which is the distance between the cameras and the subject. Different probes were created from these stereo pairs

categorized by the range. The ranges were 30, 50, 80, 100 and 150 meters. Table 3 illustrates the number of stereo pairs we have for every range. Outside of this table, 16 stereo pairs of groups of subjects and 8 images for one subject with different yaw angles starting from -90 to 90 through -45, -30, -15, 15, 30 and 45.

TABLE III: NUMBER OF STEREO PAIRS AT EACH RANGE

Range	**Number of stereo pairs**
30	21
50	20
80	24
100	24
150	18
Total	**107**

1. Component-wise Performance Evaluation

In the following subsections, we present the system's results based on three main processes: (1) Face Detection, (2) Face Cropping, and (3) Face Recognition.

a. Face Detection. Given an arbitrary image, the goal of face detection is to determine whether there are any faces in the image and, if present, return the image location and extent of each face. Up to this point, we are dealing with face localization, which aims to determine the image position of a single face; this is a simplified detection problem with the assumption that an input image contains only one face. After detecting faces, the system detects the two eyes and the mouth. These face features are used in following stages to complete the recognition process.

Table 4 summarizes the results of the face detection rates. For every range, the number of stereo pairs is multiplied by two, because for the face detection step these stereo pairs are two separate images. The face detection rate is then calculated as the ratio between the correctly

detected images and the total number of images for that range. After that, the face features detection rates are calculated similarly but with respect to the correctly detected images not the total number of images for that range.

TABLE IV: FACE AND FACE FEATURES DETECTION RATE

Range	Number of Stereo pairs	Number of images	Face detection rate	Left eye detection rate	Right eye detection rate	Mouth detection rate
30 m	21	42	40/42=95.24%	39/40=97.5%	40/40=100%	38/40=95%
50 m	20	40	38/40=95%	38/38=100%	38/38=100%	38/38 =100%
80 m	24	48	40/48=83.33%	40/40=100%	38/40=95%	34/40=85%
100 m	24	48	32/48=66.67%	32/32=100%	31/32=96.88%	28/32=87.5%
150 m	18	36	17/36=47.22%	16/17=94.12%	16/17=94.12%	15/17 =88.24%

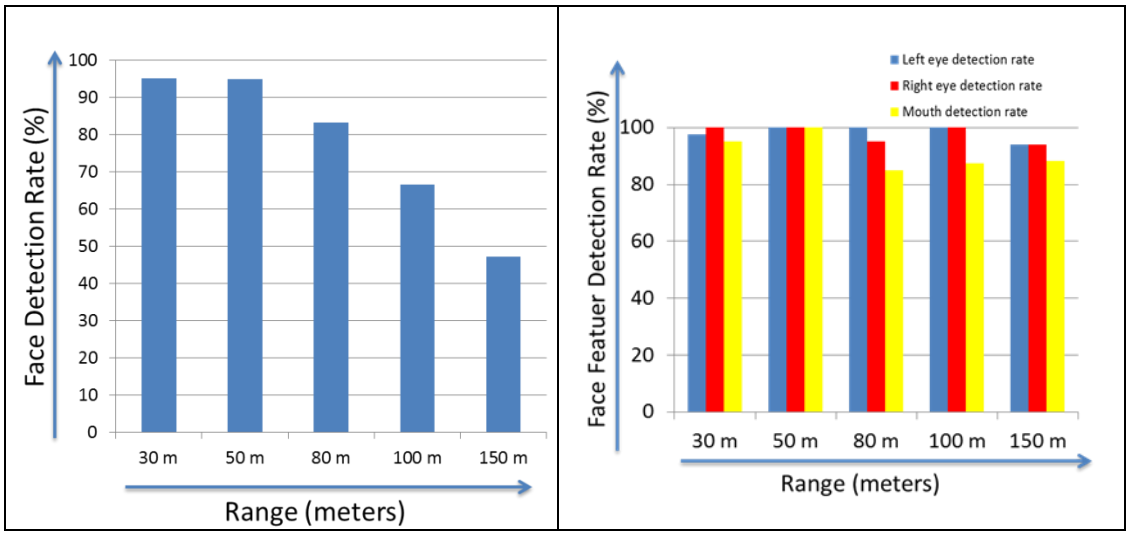


FIGURE 25 - Face detection (left) and Facial Part (right) success rates as a function of distance from the camera

The system succeeded in detecting the faces in many challenging stereo pairs. Most of these images were challenging because of the sun effect. Some of them had occlusions like subjects wearing caps and sun glasses. Also some of the subjects had moustaches, beards or even a strand of hair hiding part of the face (Farag, et al., 2012 [19]). Figure 26 shows sample results.

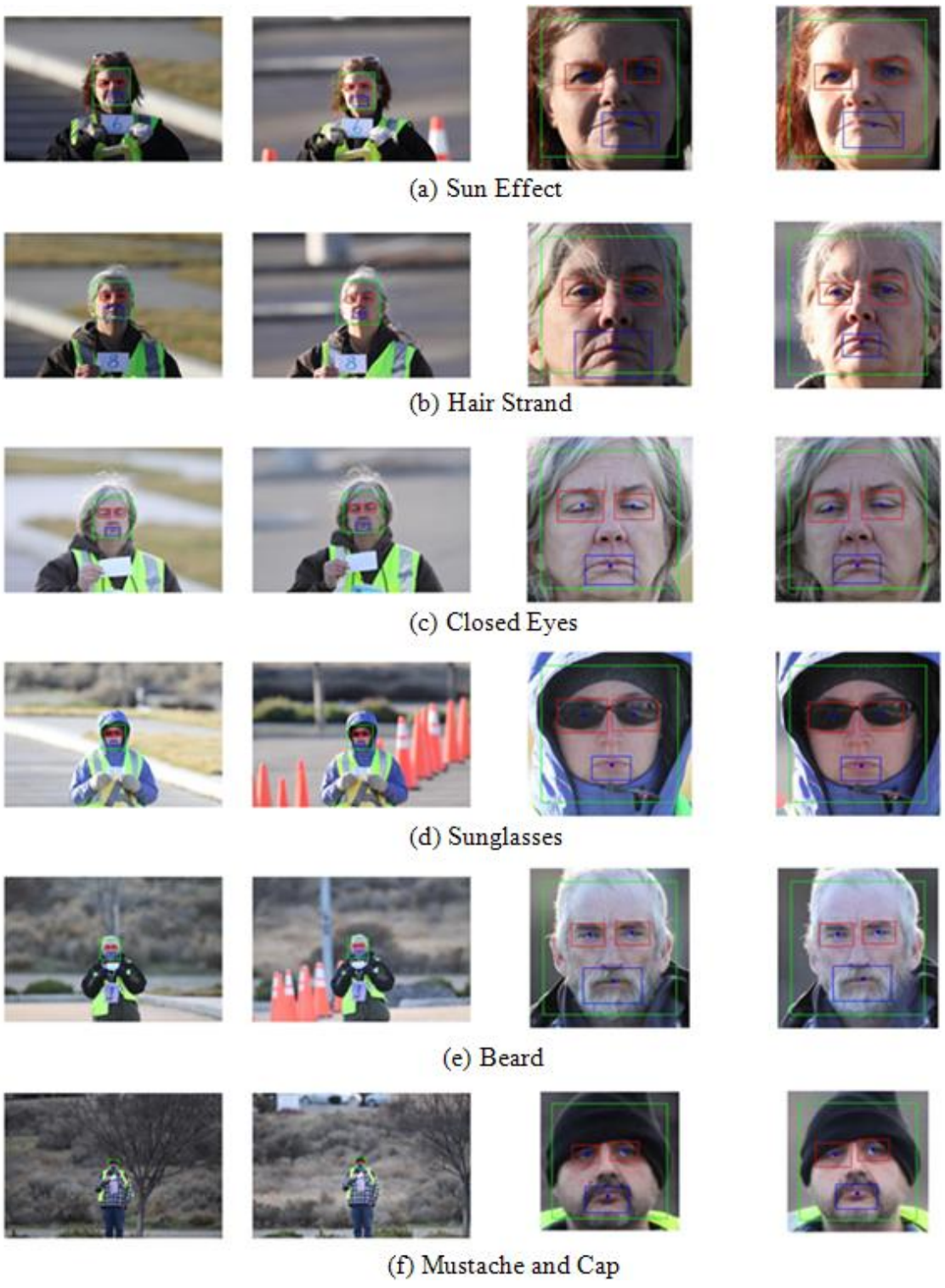


FIGURE 26 - Face detection challenges (a) sun effect (b) hair strand (c) closed eyes (d) cap and sunglasses (e) beard (f) moustache and cap. The first two columns show the left and right images, respectively, with face detection results overlaid on them. The last two columns show a zoomed in view for the detection results

Combining these difficulties for different subjects on different ranges led to some errors that will be illustrated in the next subsection. In addition, the effect of the range will be illustrated. However, there are face and facial features detection have encountered some challenging problems.

For both the 30 meters and the 50 meters images, the face detection failed in only one stereo pair in each group. The failure was for the same reason in both of them; the subject was wearing a cap and eye glasses that combined with the effect of the sun resulting in that failure. For the 30 meters case, part of the face was detected in the right image but failed with detecting the face features and for the left image the face wasn't detected at all. For the 50 meters case, the face was not detected in both the right and the left images. The results are shown in Figure 27(a) and (b) for 30 and 50 meters respectively. The same subject without eyeglasses and reversing the cap was detected successfully and the result is shown in Figure 27(c).

For the 80, 100 and 150 meters the errors in face detection took two forms either the system failed to detect the face at all or the system detected the face in a wrong place. This is illustrated in Figure 28. For some images, the face was not detected in both stereo pair images but for most of them, the face was detected correctly in the left image but failed in the right image. Figure 29 shows other face detection errors that are due to occlusion either with sunglasses and cap or with hair strands.

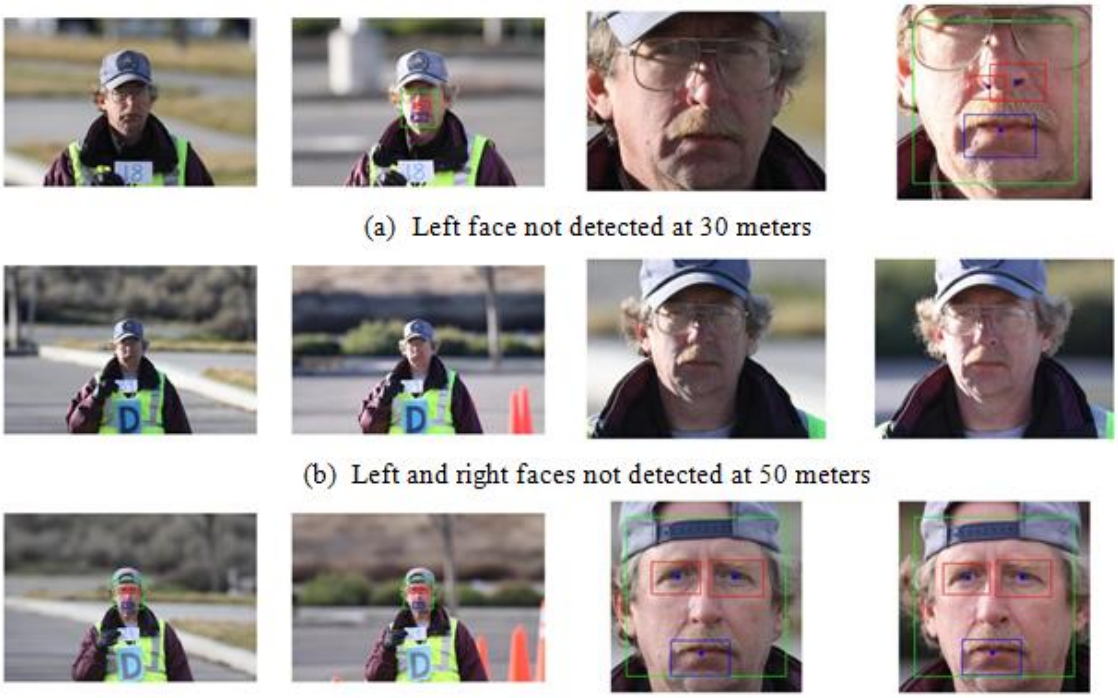


FIGURE 27 - (a) and (b) Face detection failures at 30 and 50 meters respectively (c) same subject detected correctly after taking off the eyeglasses and reversing the cap. The first two columns show the left and right images, respectively, with face detection results overlaid on them. The last two columns show a zoomed in view for the detection results

FIGURE 28 - Face detection errors at 80 and 100 and 150 meters

FIGURE 29 - Other face detection errors due to sunglasses, cap and hair strands. The first two columns show the left and right images, respectively, with face detection results overlaid on them. The last two columns show a zoomed in view for the detection results

For face features, there were some errors in detecting the position of the eyes and the mouth. As illustrated previously, the mouth detection rates were lower than eyes detection rates. Figure 30 shows some of the errors in detecting face features. Errors in detecting eyes were due to sunglasses or hair strands on eyes combined with sun effect. Errors in mouth were mostly due to moustache and beard.

b. Facial Cropping. Facial cropping starts with facial features that have been detected in the face detection module. These features are used to initialize the facial mesh used for cropping, which is fitted based on the active appearance model trained by samples drawn from the CVIP-EWA database. Figure 31 and Figure 32 shows the initial and final cropping of two sessions in the PNNL database. The first session has acceptable final face cropping, while the latter one has inaccurate cropping due to occlusion presented by facial hair.

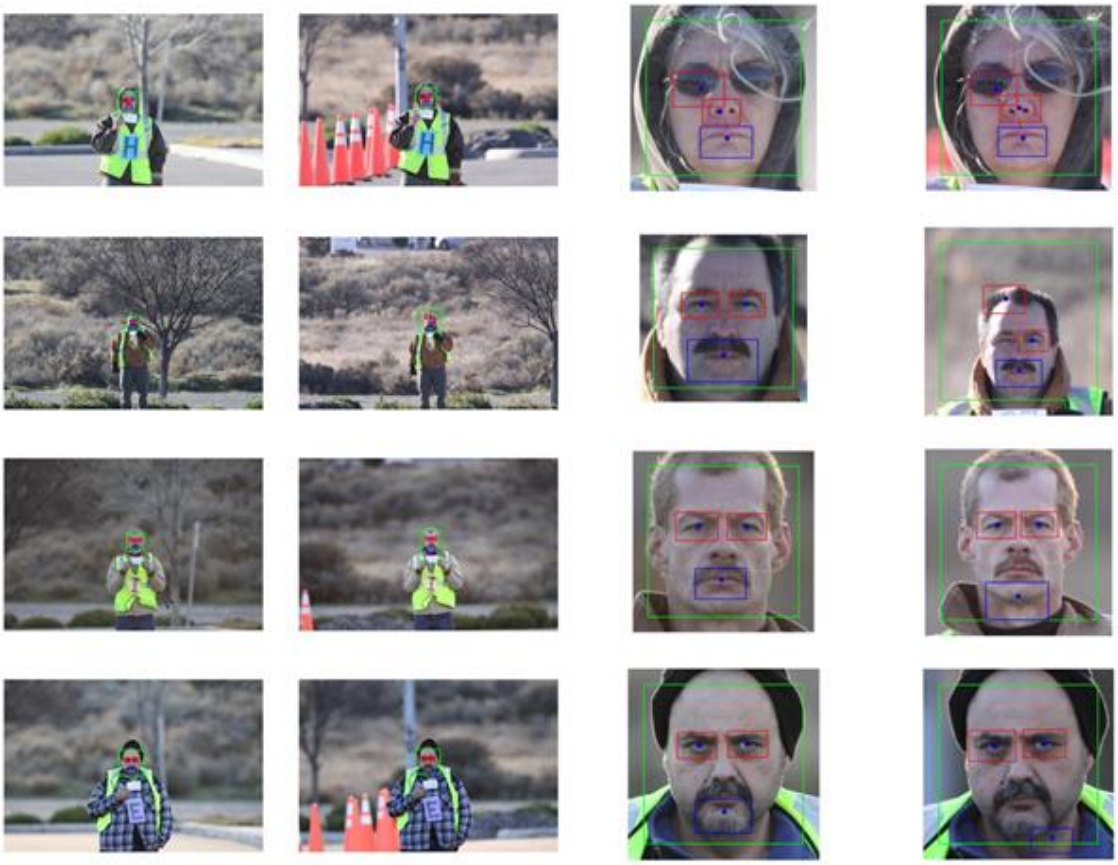


FIGURE 30 – Errors in detecting eyes and mouth. The first two columns show the left and right images, respectively, with face detection results overlaid on them. The last two columns show a zoomed in view for the detection results

FIGURE 31 - The output in each step in face cropping for a good candidate in initial and final face cropping

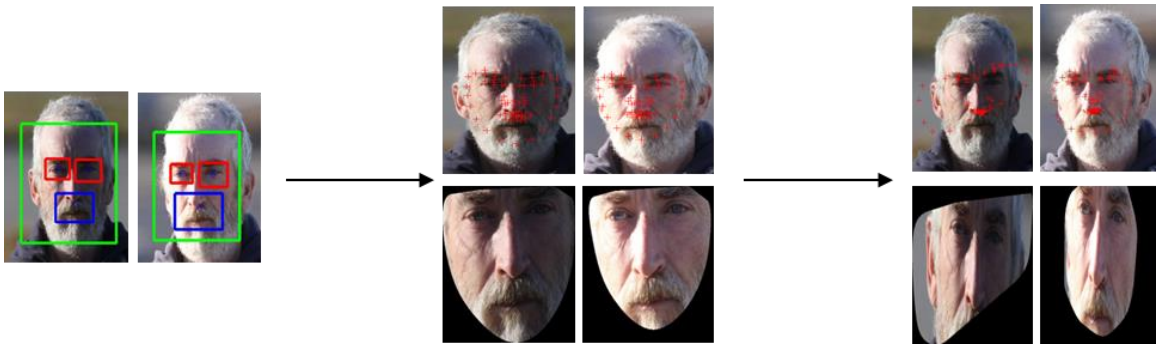


FIGURE 32 - The output in each step in face cropping for a good candidate in initial and bad in final face cropping

Table 5 shows percentages for acceptable facial cropping for the visual inspection viewpoint (Farag et al, 2012 [19]). It can be inferred that the result of final cropping is worse than the initial cropping at each distance, which means that the distance is not the issue. First, we explain what is initial and final cropping. Initial cropping is affine wrapping given three correspondence points; left and right eye and mouth. Therefore, the initial cropping will fail if one of these point correspondence has been detected wrong. The final cropping is applying Active Appearance model (AAM) on initial cropping [42]. The reason that the results of final cropping are worse than initial cropping is the algorithm diverges. The divergence is due to the AAM algorithm depends on the training data.

Among the possible enhancements for BOSS are the following: (1) Training AAM on uncontrolled environment database. (2) Investigating 3D-based mesh fitting algorithms, such as the work by Kanade [46] which propose a real-time combined 2D+3D active appearance models to solve the problem of pose and occlusion. (3) Investigating improvement in AAM algorithm to improve speed and robustness against illumination [47].

TABLE V: PERCENTAGES OF ACCEPTABLE FACIAL CROPPING (BASED ON VISUAL INSPECTION) AT DIFFERENT DISTANCES

Distance	No. of face detected session	Acceptable initial cropping percentage	Acceptable final cropping percentage
30 meter	18 sessions (36 images)	94.44%	50%
50 meter	18 sessions(36 images)	94.44%	88.89%
80 meter	19 sessions(38 images)	78.95%	50%
100 meter	12 sessions(24 images)	83.33%	66.67%
150 meter	5 sessions(10 images)	80%	60%

c. Recognition. A new gallery is constructed from the 11 subjects (10 from the 30 meter data and 1 from the 50 meter data). Four probe sets are also constructed at 30, 50, 80, 100, and 150 meters using sessions with faces successfully detected and cropped. For 30-meter probe, we have 18 sessions for 11 subjects. Figure 33 shows the recognition performance multi-classifier approach versus using each classifier alone. The recognition rate is 72.22 % at rank 1 from the three classifiers, 66.66 % at rank 1 from the dense classifier, 55.55 % at rank 1 from the sparse classifier, and 27.77 % at rank 1 from the texture classifier.

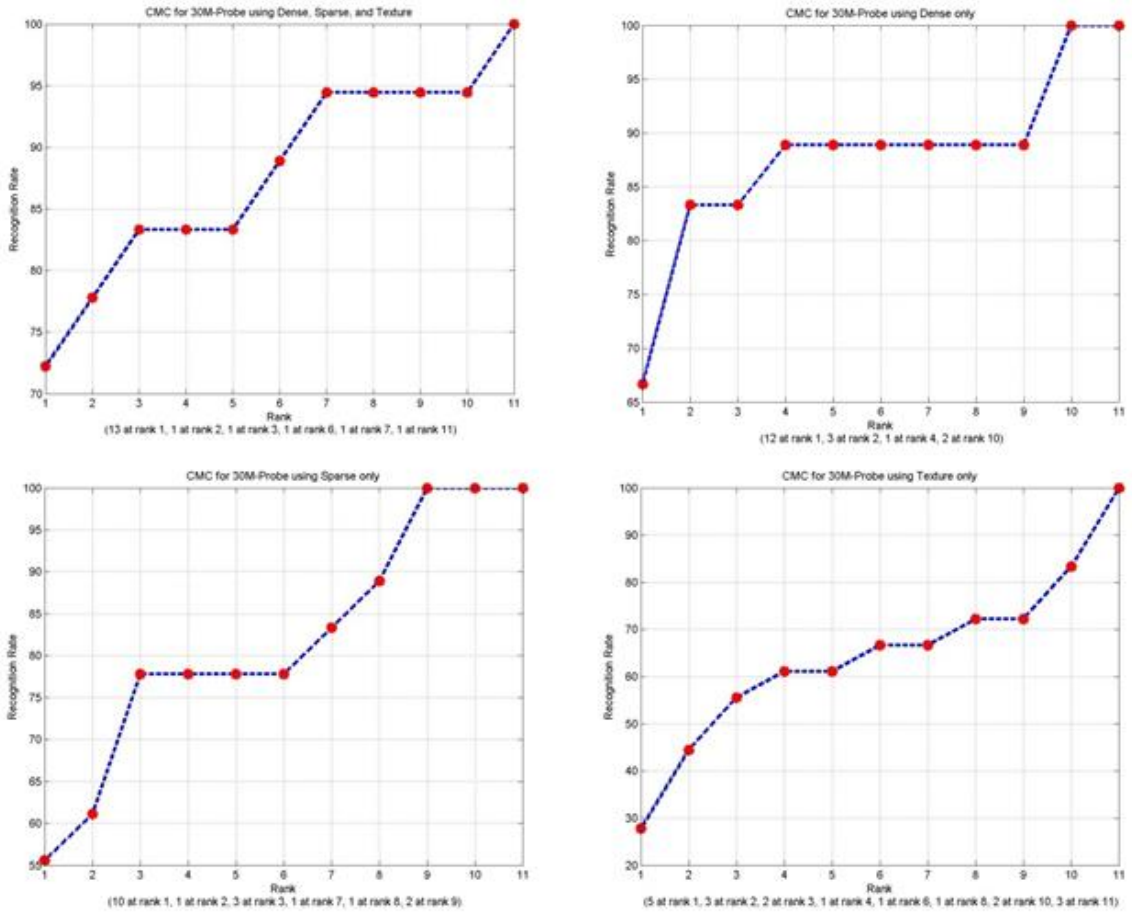


FIGURE 33 - Cumulative matching curves for 30-meter probe

For 50-meter probe, we have 17 sessions for 11 subjects. Figure 34 shows the recognition performance multi-classifier approach versus using each classifier alone. The recognition rate is 82.26 % at rank 1 from the three classifiers, 52.94 % at rank 1 from the dense classifier, 47.05 % at rank 1 from the sparse classifier, and 82.26 % at rank 1 from the texture classifier.

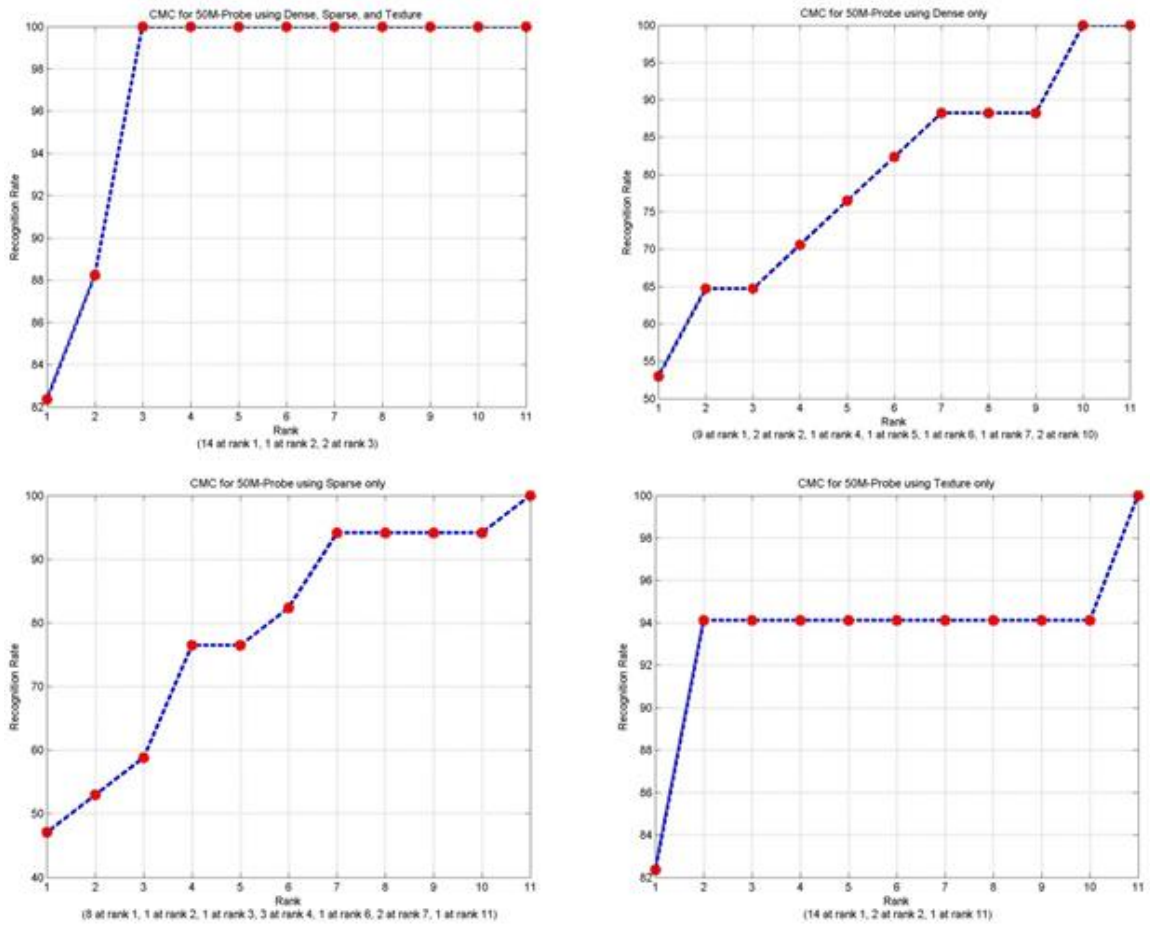


FIGURE 34 - Cumulative-matching curves for 50-meter probe

For 80-meter probe, we have 19 sessions for 11 subjects. Figure 35 shows the recognition performance multi-classifier approach versus using each classifier alone. The recognition rate is 52.63 % at rank 1 from the three classifiers, 21.05 % at rank 1 from the dense classifier, 31.57 % at rank 1 from the sparse classifier, and 57.89 % at rank 1 from the texture classifier.

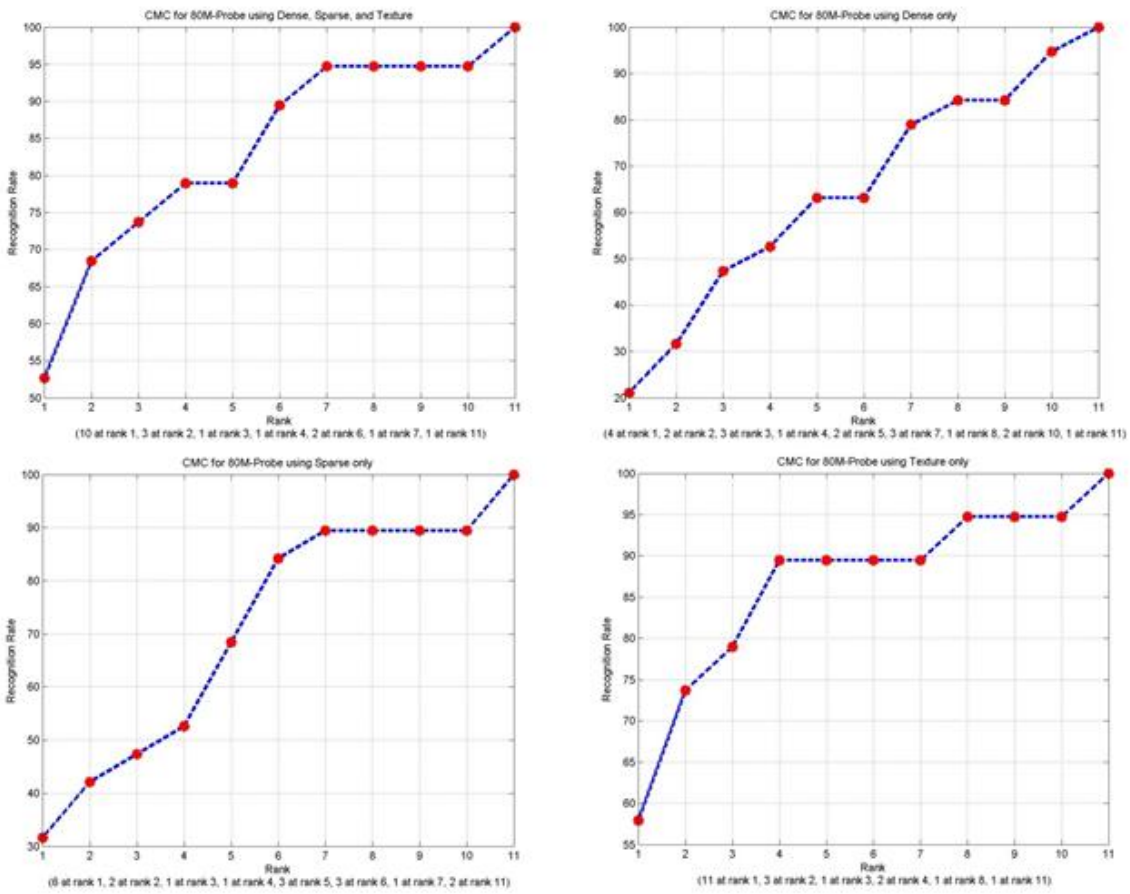


FIGURE 35 - Cumulative-matching curves for 80-meter probe

For 100-meter probe, we have 12 sessions for 11 subjects. Figure 36 shows the recognition performance multi-classifier approach versus using each classifier alone. The recognition rate is 83.33 % at rank 1 from the three classifiers, 16.66 % at rank 1 from the dense classifier, 25.00 % at rank 1 from the sparse classifier, and 100.00 % at rank 1 from the texture classifier.

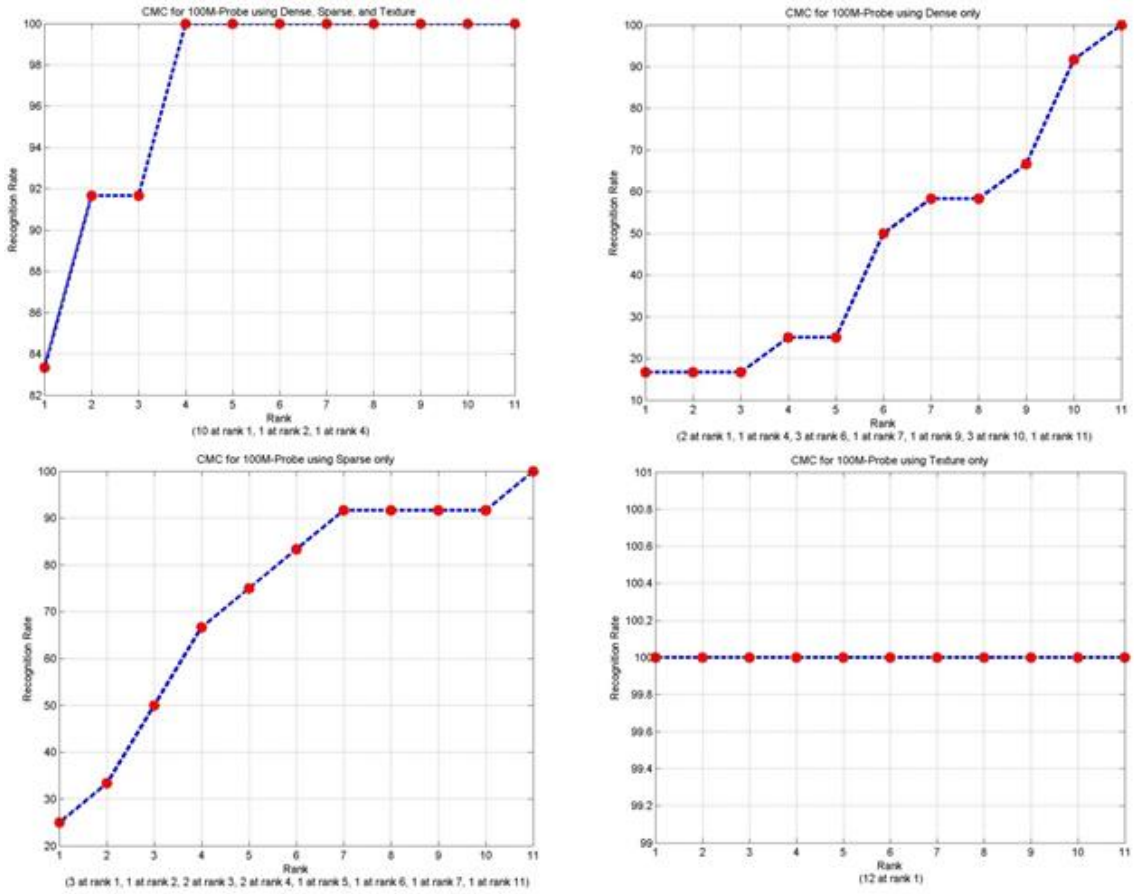


FIGURE 36 - Cumulative-matching curves for 100-meter probe

For 150 meter probe, we have 3 sessions for 11 subjects. Figure 37 shows the recognition performance multi-classifier approach versus using each classifier alone. The recognition rate is 66.67 % at rank 1 from the three classifiers, 33.33 % at rank 1 from the dense classifier, 33.33 % at rank 1 from the sparse classifier, and 33.33 % at rank 1 from the texture classifier.

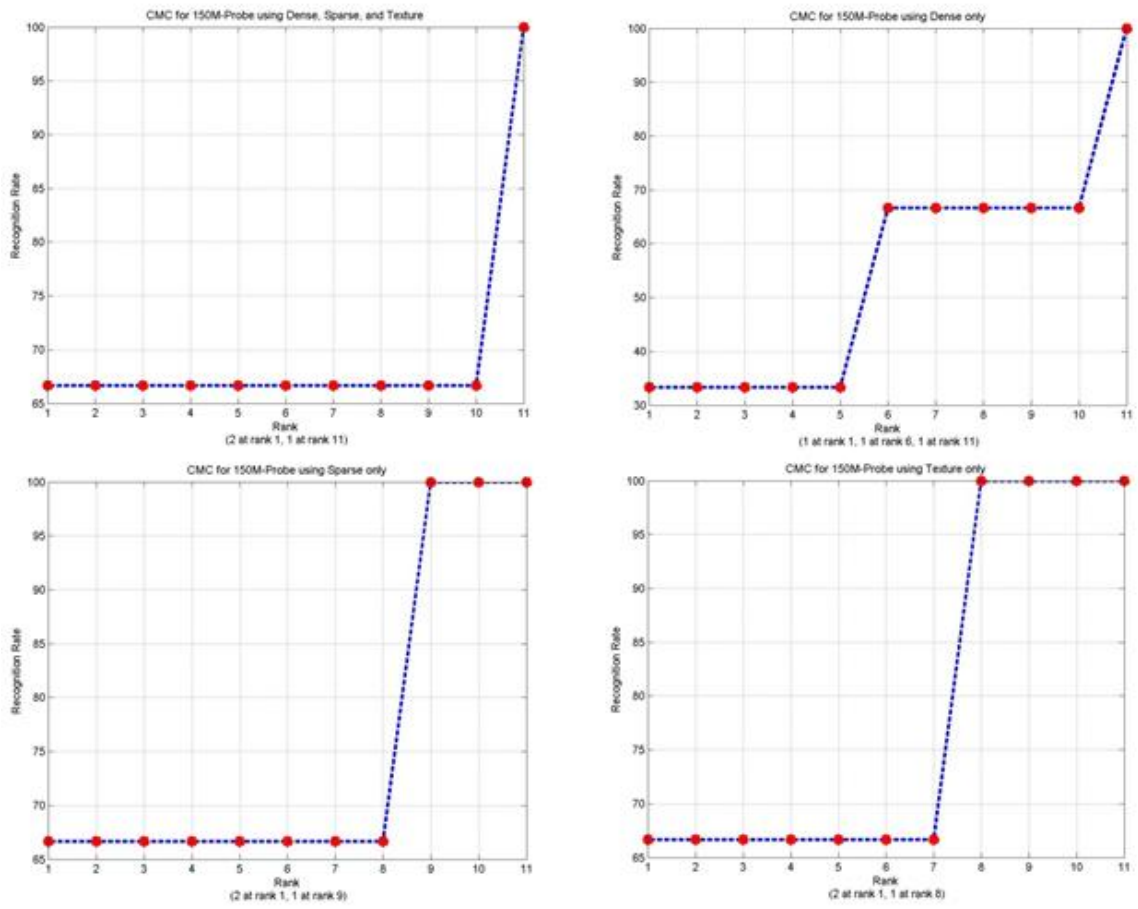


FIGURE 37 - Cumulative-matching curves for 150-meter probe

As expected, the performance deteriorates with severe imaging conditions and as distance increases.

2. Holistic/Overall System Performance

The above details about the database construction and testing mechanisms provides a glimpse of what it is involved in design, test and evaluate a facial biometric system designed for conduct face recognition at a distance in an open environment. An overall evaluation could be a "binary" decision or a ranked one. In addition, if the person is not in the gallery, a decision would be two-stage: Imposter/Genuine, then Binary or Ranked if the person is enrolled into the

gallery. Of course, evaluations can also be performed by adding uncertainties to the probe; e.g., altering the image quality, adding occlusion, etc.

B. Summary

This chapter examined testing strategies, as well as an evaluation, of the BOSS system in a dual-channel setup using high resolution cameras. In the following chapter the BOSS evaluation will be performed using low-resolution cameras. The next chapter will also discuss differences using the BOSS between a dual-channel setup compared to a single channel setup. This thesis will consider face recognition in low resolution images that have different poses, illuminations, distances, and expressions. Farag et al., 2012 [19] and various other literature listed in this document detail the system performance and implementation details.

IV. IMPLEMENTATION (BOSS LOW RESOLUTION CAMERA/TESTING)

The previous chapter studied the BOSS facial biometric system in use with high resolution cameras. The main issues in design, test and evaluation of facial biometric systems were discussed. In this chapter, the BOSS will be evaluated using low resolution cameras, specifically the iPhone 4 camera. The system will be evaluated for its performance against varying poses, illuminations, distances, and expressions.

A. Motivation and Challenges

Resolution, when pertaining to cameras, is what is considered to be the most important aspect when talking about crispness of an image. It corresponds with the amount of detail that can be seen in an image captured by a camera.

> **Resolution** - a measure of the sharpness of an image or of the fineness with which a device (as a video display, printer, or scanner) can produce or record such an image, usually expressed as the total number or density of pixels in the image.

A common way to describe resolution is through the number of pixels an image contains, usually seen as a megapixel rating. A megapixel rating describes how many pixels in a photo.

For example, if the photo measures 4,000 by 3,000 pixels, simply multiplying the two numbers gives 12 million, or a 12-megapixel (MP) photo.

In the previous BOSS setup, the CVIP lab gathered images in a stereo setup (using two Canon EOS 7D cameras to capture an image), which provides a total of 18 MP per camera. These images consisted of 5184 pixels wide by 3456 pixels high (5184*3456 = 17,915,904 ≈ 18MP). The images gathered from these cameras are of high-resolution; they provide crisp picture quality. The BOSS system has been analyzed and evaluated using these high resolution images; hence, a logical question is: how would the system perform with low resolution images? This question was one of the motivations behind this thesis. This issue would motivate investigating the use of smart phones and portable devices, in general, for facial biometrics. This thesis, along with its test parameters and data collection, have revolved around using the iPhone 4 which boasts a low resolution camera which generates a 5 MP (2592*1936) image; this should not to be confused with the iPhone 4s which boasts an 8 MP camera.

Apart from resolution, there are other parameters which are challenging for a facial recognition system. These challenges include varying pose (pitch, roll, and yaw), illumination, expression, and occlusion, in addition to the distance of the subject from the camera.

Pitch refers to the head angle rotating up and down (subject looking up and looking down). Roll refers to the body staying straight while the head is rotated to the left and right past the median line of the body. Yaw refers to the head being turned to the left and right, causing partial occlusion of each side of the face. Figure 38 represents the different poses (see Figure 7)

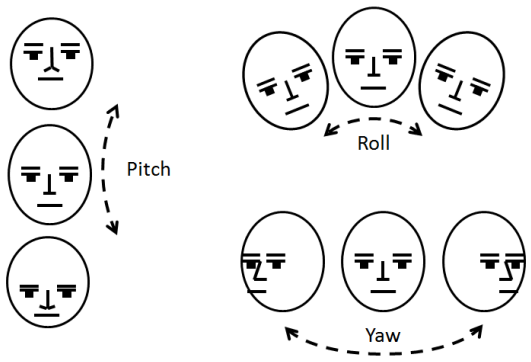


FIGURE 38 - Different Facial Pose

Illumination refers to varying light in a scene. Depending on where the light is in regard to a subject's face. Illumination affects the performance of face recognition systems. For example, a light source above and behind a person would produce a shadowing of the face, causing facial features to be hidden and information lost; however, if light is in front of a subject (i.e. a spot light shining on the face) all facial features would be shown.

Facial expression is another challenge prevalent in face recognition. On average, a human has 43 muscles in their face. These muscles are capable of expressing emotions ranging from happiness and sadness to fear and disgust. Each one of these expressions may cause a person's facial features to change dramatically. Since facial recognition is used by comparing facial features from a probe with features of subjects enrolled in a database (gallery), severe distortion of one's face would cause failure in identification. It should be mentioned that facial expression has not been modeled in the BOSS.

Occlusion refers to the full or partial covering of a face. This can be as simple as a female's hair strand, to as severe as a masked robber. Occlusion causes facial features to be hidden; which introduces uncertainties in the facial feature extraction phase of a face recognition system.

B. BOSS Implementation (single channel imported image)

In Chapter 2, the BOSS was described in use with a dual channel stereo setup, where the input image was taken online, from two cameras connected to the system. In this section, the input images were taken offline by an iPhone 4 camera, and then imported to a BOSS equipped computer via the use of an applet called "Quickshot with Dropbox" which uploads an image from the smart phone into the hard disk.

As stated before, the BOSS is capable of importing images onto the system without the need of taking the picture from the GUI itself. Once an image is captured by the iPhone 4 and imported to the computer via the app "Quickshot," it is then imported into the BOSS as the input image; this is the face acquisition step. Apart from the acquisition of the images, the pipeline between the stereo image and single image are very similar.

The face detection phase is almost the same as stated in chapter 2, where the Viola-Jones algorithm is used in order to detect a face and used to detect facial features as well; however, code for the BOSS had to be modified, specifically the scaling factor ratios described previously, in order to properly detect faces from this new sensor, the iPhone 4 (Appendix B). For the current setup, the ratios were adjusted as follows: ½ initially, then by 1/3, and lastly by ¼ the original image size. Before adjusting these ratios, the BOSS was incapable of detecting faces from an iPhone 4 input image because the face size was too small.

Once a possible face has been detected, the scoring algorithm for false positive reduction, discussed in the previous section, is used.

Apart from the modified face detection code, the main difference between the dual channel and single channel configuration is the lack of a second false positive reduction step in

the single channel configuration. Since there are not two images, the system does not use a left and right image Gabor signature to determine which face corresponds to one another between the two images. While this may increase the chance of False Positive faces when using the BOSS for multiple face detection, this thesis tested the BOSS in the single face detection mode (i.e. there was only one face present in a given image).

Similarly to the previous chapter, once a face is detected, the Adaboost classifier is used in order to detect 9 facial feature points, and then ASM is used to find the output 68 facial feature points. Again, a feature vector is made for each of the three signatures discussed previously (Gabor, LBP, 3D sparse reconstruction). Each feature vector is then compared to the feature vectors of previously enrolled subjects and the system outputs a decision, Figure 39.

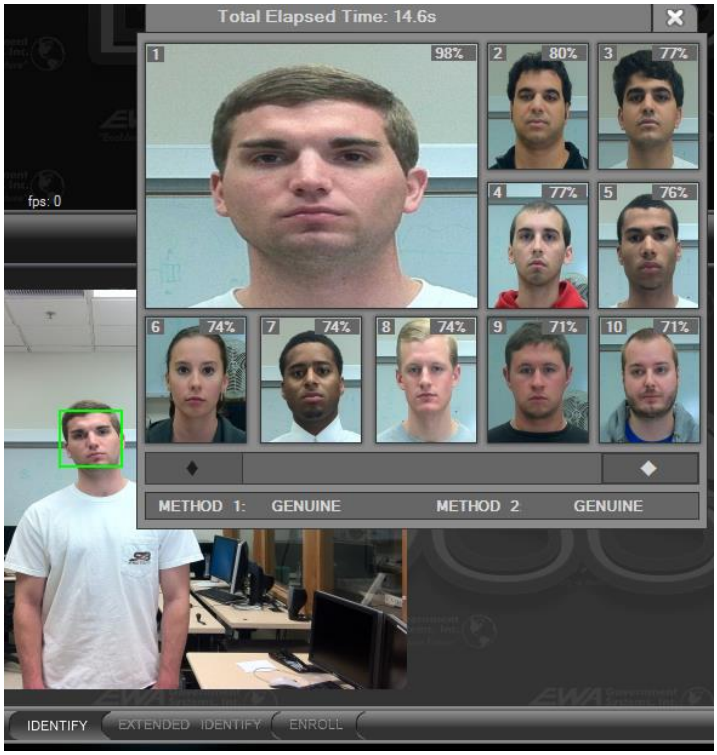


FIGURE 39 – True Positive Identification; Confidence Level = 98%
Distance = 5 feet, Illumination ON, Roll = +15°

Notice that apart from outputting a Confidence Level (CL) between the probe and identified face, the system uses two methods in order to state whether the identified face is an imposter (not in the database), or genuine (in the database). Method 1 is calculated by first seeing if Rank 1 has a CL $\geq$ 30%, and also if the CL between Rank 1 and Rank 2 are above a certain point. Method 2 is calculated by first seeing if Rank 1 has a CL $\geq$ 30%, and if the slope of Rank 1, Rank2, and Rank3 is above a certain threshold. It was found that Method 2 was more accurate due to Rank 1 and Rank 2 being close to each other.

C. Testing

1. Test Set Up

This thesis investigated the question: can the BOSS be used on low resolution cameras? In order to answer this question, a new database needed to be acquired, using low resolution cameras. As stated before, the low resolution camera in questions is the 5 MP camera housed on the iPhone 4. In order to challenge the BOSS, test parameters had to be made in order to create a database of varying pose, illumination, expression, and distance similar to what would be used in the wild. The test setup is described in detail below.

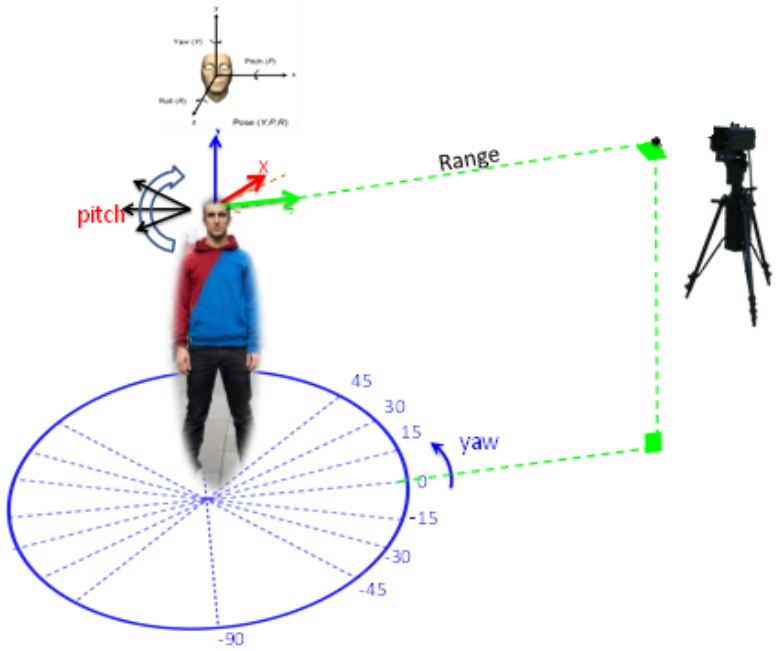


FIGURE 40 - A single channel (individual) data collection setup

Using a tripod with adjustable leg lengths, the height of the camera from the ground remained at a constant five feet and two inches. The tripod was perfectly level, using a leveling device built onto its structure. Figure 40 is a representation of the test set up. Testing was done on a total of 21 subjects. These subjects varied in sex, height, weight, and ethnicity. Of course, half male and half female were ideal for testing the system; however a lack of female participation/interest created the need to gather more males for testing. There were a total of 8 female subjects along with 13 male subjects imaged.

2. Test Parameters

This section describes the test parameters used to challenge the BOSS. First, the subjects were enrolled into the BOSS database from an image taken from a distance of 5 feet away with 0° of pose as well as Illumination ON. Once the subject was enrolled into the database, they were

asked to change 3 parameters of pose (pitch, roll, and yaw), which was described earlier in this chapter, as well as varying other parameters.

The first parameter was varying the yaw (0°, ± 15°). The second parameter was varying the roll (0°, ± 15°). The third parameter was varying pitch (0°, ± 15°). It should be mentioned that, aside from yaw, it is very difficult to standardize these angles in variation. The subjects were photographed giving two different expressions for each of the aforementioned poses, the expressions consisted of normal (no expression) and smiling. This parameter was introduced in order to evaluate the BOSS performance on expression, which the system has not been modeled on. For each of these different pose/expression variations, the subject was introduced to the fourth parameter; varying illumination. This varying illumination was produced by placing a spotlight pointed toward the subject's face. The fifth parameter was varying distance from the camera (5, 10, and 15 feet). Figure 41 is a panorama view of the testing station. In order to measure the varying yaw, a protractor was printed from the internet. Once the angles were verified to be accurate, a protractor was taped at each of the three distances. The tape lines represent the +15° and -15° of yaw marks. This particular image is a representation of the 10 feet distance, +15° of yaw, normal expression, and illumination ON image.

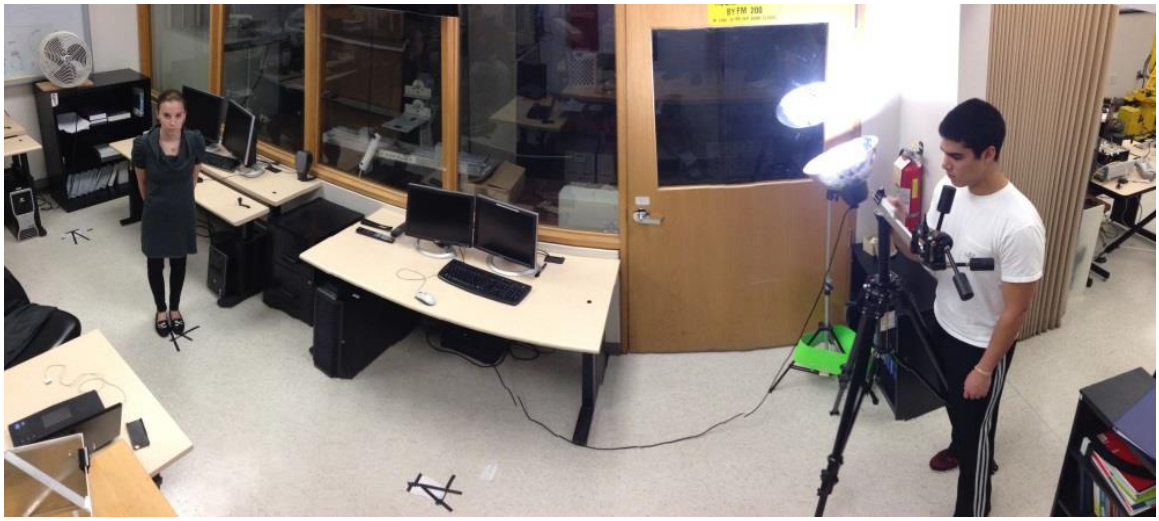


FIGURE 41 - Testing Station

With seven different pictures regarding pose, two different expressions for each differing pose, i.e. 14 pictures for each varying illumination, or 28 pictures per varying distance, resulting in 84 pictures for each subject.

D. Data Collection

This section describes the data collection process. Once the test parameters and test set-up was finalized, volunteers were sought out to be imaged using the CVIP Lab biometric IRB consent form. Each subject was then photographed using an iPhone 4 at various distances and varying parameters. After each subject had been photographed, the images were imported to a BOSS equipped computer through the applet "Quickshot with Dropbox." Upon enrolling each subject into the BOSS (using the Illumination ON, No Expression, 0° pose, distance of 5 feet photograph), each of the remaining 83 photos were used as an input to the system. Figure 42 is a snap shot of 6 subjects enrolled into the database using the iPhone 4 camera.

FIGURE 42 - Example of Enrolled Subjects in BOSS database

With each image output, an Excel spread sheet was populated (Appendix C). The output data collected included subject Rank as well as Confidence Level (CL). The data collected also stated such things as if Method 1 and Method 2 had the subject as "imposter" or "genuine." This data can be translated into whether the system decision was in fact a true positive, false positive or false negative. This thesis was only concerned with whether a subject was properly identified (i.e. Rank1); therefore, data from the spread sheet was converted to binary (1 or 0). Binary 1 refers to a true positive of Rank 1 and Method 2 properly identifying the decision as "genuine" (recall, Method 2 was found to be more accurate), Binary 0 was distributed to any failures in the system (i.e. incorrect face under Rank 1).

E. Results

In this section, the output data collected from the BOSS on the 21 enrolled subjects is described. 84 images were captured from each of the 21 subjects photographed, coming out to a total of 1,764 input images to the BOSS. From these 1,764 images, a total of 1,176 (66.67%) were properly identified as true positive. Of the 588 images that the BOSS did not identify, 20 were caused by failure in the face detection phase (i.e. face was not detected).

Upon completion of the data collection process stated in the previous chapter, MATLAB was used in order to generate plots that analyzed the outcomes in many different ways. An example of this source code is in Appendix D.

Figure 43 below depicts the BOSS recognition rate in regard to varying distance. While keeping the other parameters (pose, illumination, and expression) fixed, curves were produced to show how distance affected the system. As expected, recognition rate of a subject was generally best at the shortest distance of 5 feet (red line). When processing the data, however, it was unusual to see that the 10 feet distance (blue line) tended to have a worse recognition rate than the 15 feet distance (green line). Upon further research, this problem has been attributed to the lighting in the test room. In the test room, fluorescent lights were above and in front of the subjects at the 5 feet and 15 feet distance; however, there was no light above the subject at the 10 feet distance. It is believed that the fluorescent light in front of the 15 feet mark caused another parameter that had not initially been accounted for, on the distance of 10 feet. Since there was a large amount of light behind the subjects, shadowing occurred on their faces, creating an occlusion parameter that had not been accounted for. Even though this was discovered afterwards, it was in itself a discovery of the effect illumination has on image quality, and did not warrant repeating the data acquisition at that particular distance.

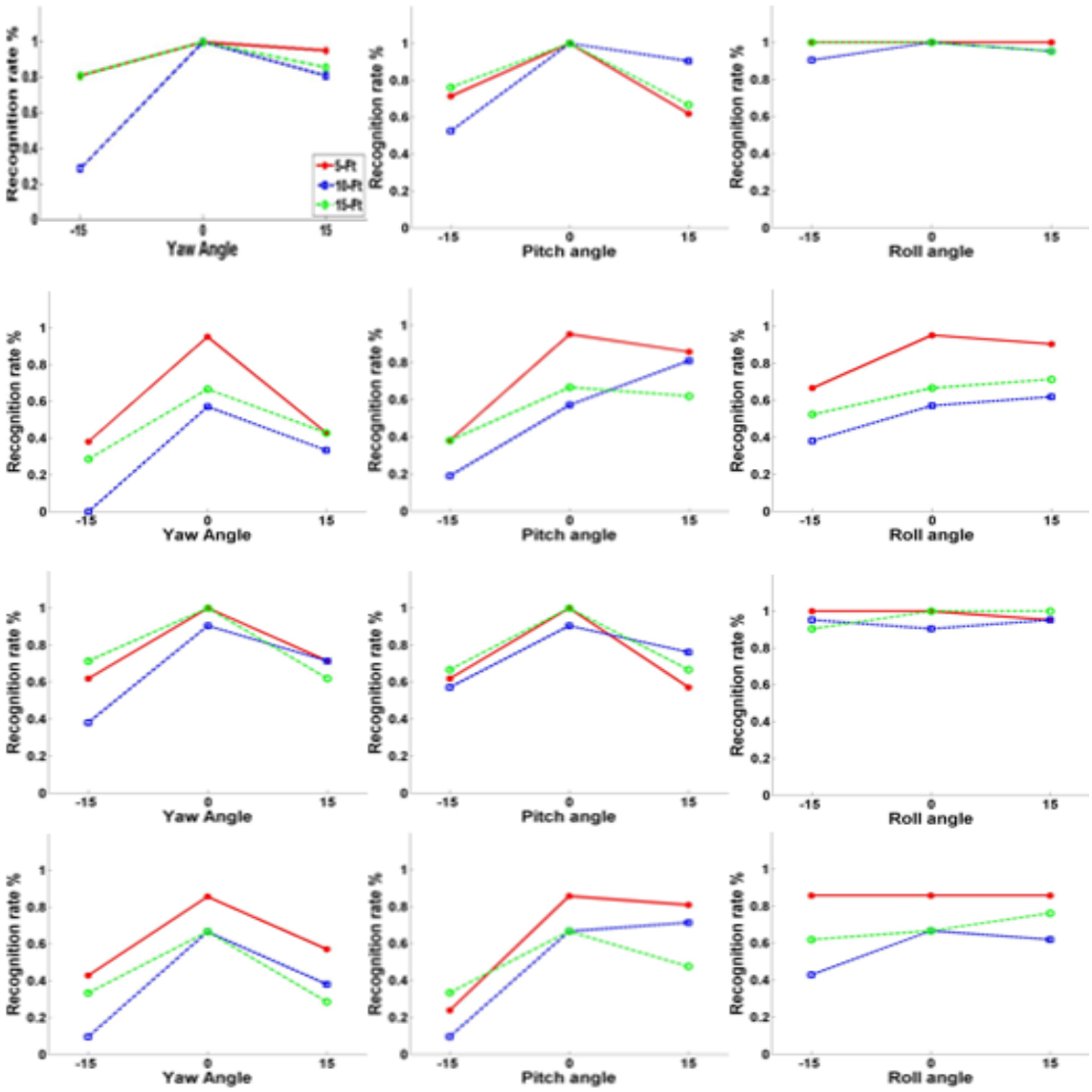


FIGURE 43 –Effect of Distance on the BOSS. Red represents 5 feet, Blue Represents 10 feet, Green represents 15 feet. First Row: Illumination ON/No Expression; Second Row: Illumination ON/Smiling; Third Row: Illumination OFF/No Expression; Fourth Row: Illumination OFF/Smiling

During the data analysis, another trend soon became apparent. Apart from the illumination issue at the distance of 10 feet, the recognition rate for a subject at a yaw of -15° was consistently less than any other parameter. Again, an unforeseen parameter was to blame for this low issue. Referring back to Figure 41 it is apparent that the back room illumination, or lack thereof, caused a complication in face acquisition. While the two issues mentioned were not

accounted for, they added to the “Face Recognition in the Wild” definition; image acquisition in “the wild” is not ideal and therefore these unaccounted parameters served as a challenge for the uncertainties in the imaging process.

Along with evaluating the BOSS in regard to varying distance, it was also necessary to evaluate the system on the varying illumination parameters; stated in the test parameters section of this chapter. While keeping the parameters of pose, distance, and expression fixed, curves were produced to show how illumination affected the system. From Figure 44, it is apparent that recognition rate was very similar during the illumination ON and Illumination OFF tests. Each two rows of this figure represent a specified distance as well as no expression and smiling, respectively. While recognition rate was not consistently above 80% for these tests, the recognition rate did not vary significantly when varying the light. It should be noted that the yaw of -15° again affected the system recognition rate negatively.

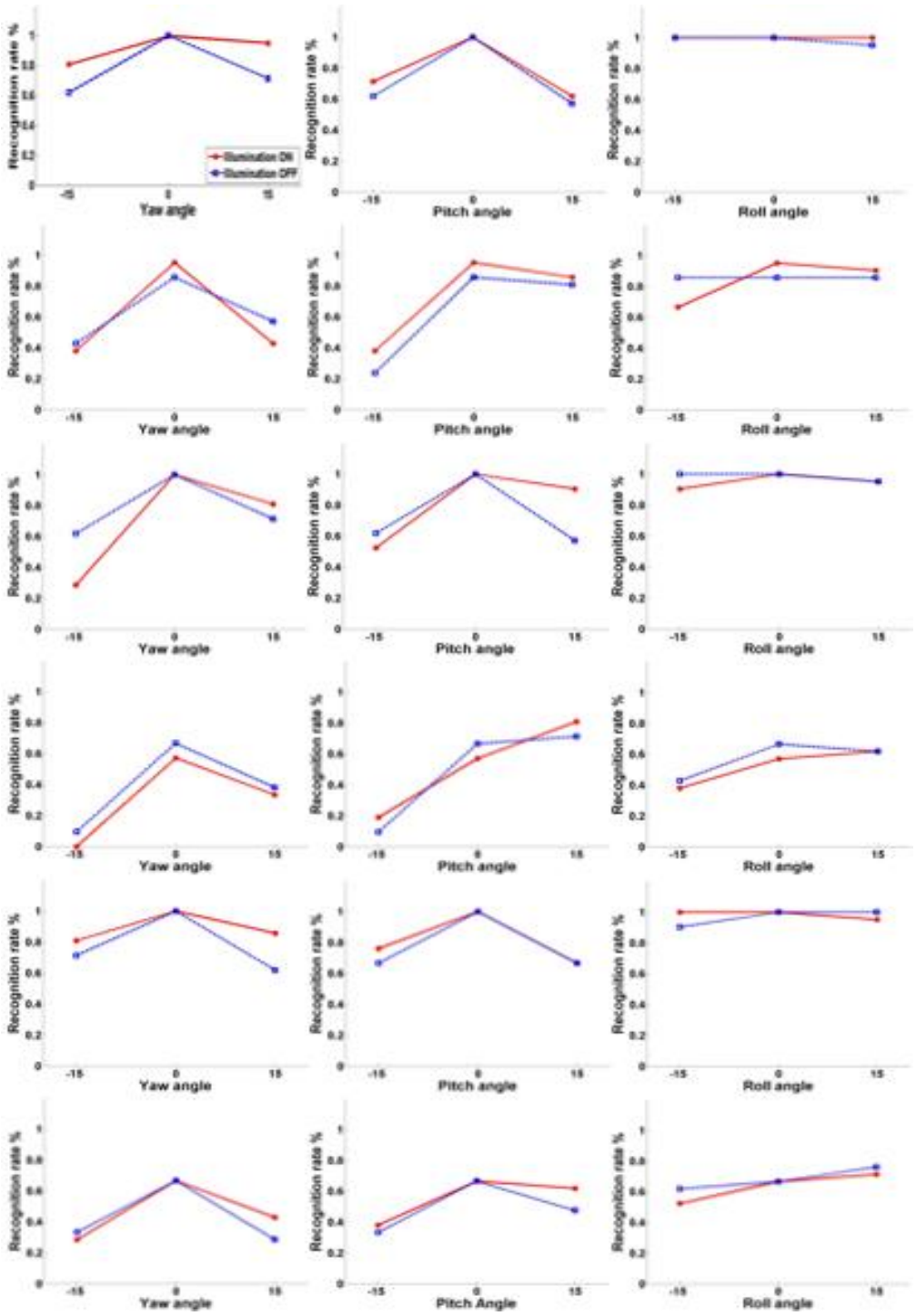


FIGURE 44 – Effect of Illumination on the BOSS. Red represents Illumination ON, Blue Represents Illumination OFF. First Row: 5 ft/NO Expression; Second Row: 5 ft /Smiling; Third Row: 10 ft/NO Expression; Fourth Row: 10 ft /Smiling; Fifth Row: 15 ft/NO Expression; Sixth Row: 15 ft /Smiling

From the test parameters, the BOSS was also evaluated in regard to varying expression; these expressions were no smile and smile. As stated before, the challenge of expression has not been modeled on the BOSS; while this is not ideal, the statement "Face Recognition in the Wild" requires the system to be evaluated on varying parameters that would be present in an uncontrolled environment; this includes varying expression.

Figure 45 depicts the recognition rate of the BOSS during varying expression. In order to evaluate the system on varying expressions, pose, illumination, and distance was held constant per graph. Each two rows of this figure represent a specific distance (starting with 5 feet) and whether illumination was ON or OFF, respectively.

Due to the BOSS not possessing any expression modeling, it was expected that the "smiling" expression would have a recognition rate less than that of the no expression parameter. It can be said that expression does in fact affect the BOSS performance due to the large difference between the parameter's recognition rates; a larger difference in recognition rate was noticed, when comparing the previous illumination parameter. Again, the very low recognition rate at yaw of -15° should be noted.

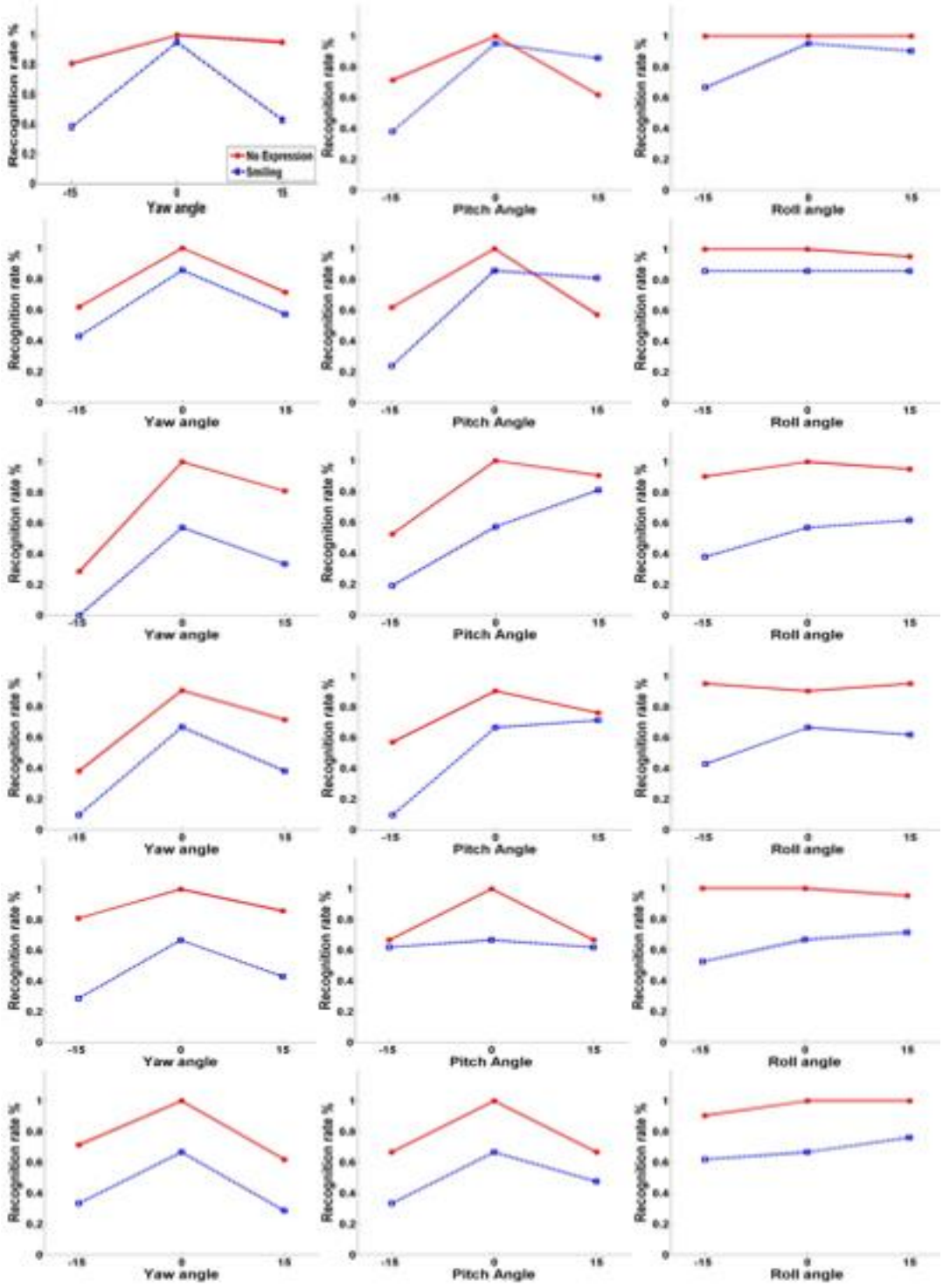


FIGURE 45 - Effect of Expression on the BOSS. Red represents NO Expression, Blue Represents Smiling. First Row: 5 ft/Illumination ON; Second Row: 5 ft /Illumination OFF; Third Row: 10 ft/Illumination ON; Fourth Row: 10 ft/Illumination OFF; Fifth Row: 15 ft/Illumination ON; Sixth Row: 15 ft/Illumination OFF

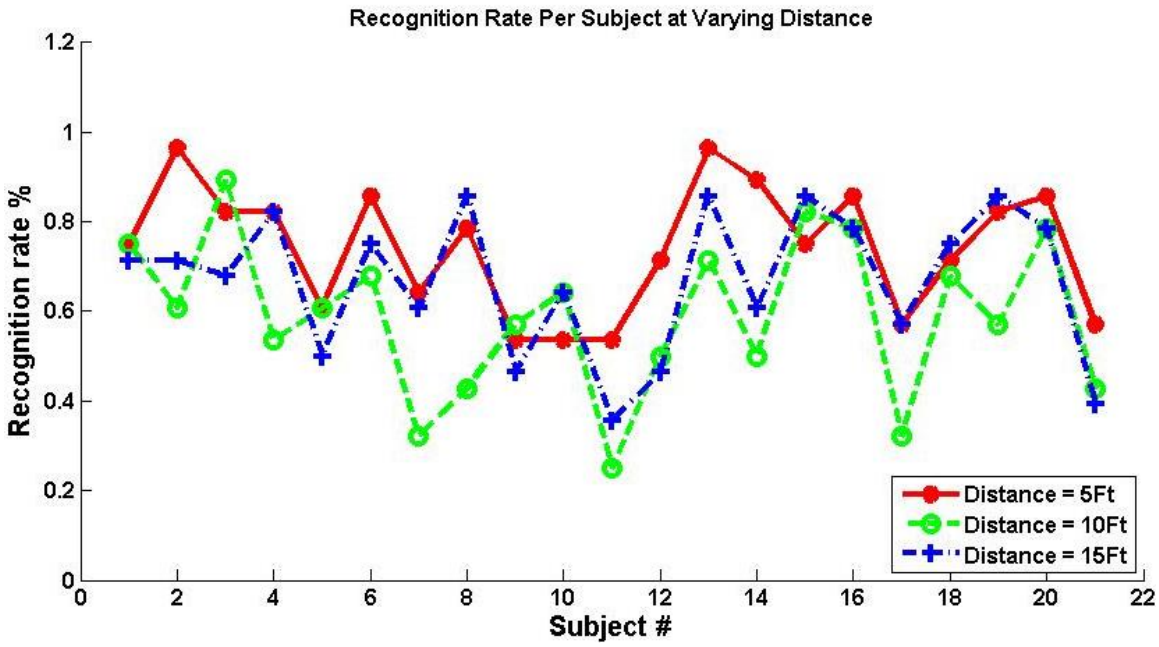


FIGURE 46 - Recognition of Each Subject at Each Distance

Figure 46 represents each subject's recognition rate in regard to varying distance. Every parameter previously mentioned was used in order to calculate the overall system performance per subject. It is clear that the closest distance of 5 feet had the best recognition rate for each varying parameter, as expected. While 7 subjects had a better recognition rate at other distances (2 at 10 feet and 5 at 15 feet) these are considered outliers. Again, note the low recognition rate of the 10 feet distance. As stated above, this low performance can be attributed to the test room lighting.

Comparing Figure 46 with the data collected, it is apparent that subject recognition rate can be low due to the subject themselves. While this thesis was intended to evaluate the BOSS on unconstrained scenarios, certain test parameters were made in order to test the BOSS up to a certain degree. While yaw was capable of being quantified easily (taping protractor to the floor), there was no apparent way to keep pitch and roll to ±15°. Using the naked eye, subjects were

told to tilt their head more or less in a certain way to roughly estimate these two pose parameters. It became apparent, especially with non-experienced imaging subjects, that many were unable to keep their faces at 0° for one pose parameter while testing the system on another pose (i.e. keeping face pitch at 0° while rotating +15° of yaw. This introduced a variation of both parameters to the system, and can be reasons for these subjects' low recognition rate.

In order to evaluate the BOSS more effectively on specific parameters in the future, the test area would need to be moved where there is uniform lighting, in order to test the system on varying illumination more precisely, as well as avoiding the problems occurred at a distance of 10 feet and yaw of -15°. In addition, it is believed that being able to precisely measure pose angles would greatly increase the BOSS performance. While 21 subjects may have been sufficient for a small sample evaluation, future evaluations of the BOSS should include many more subjects. The subjects ranged from varying weight, height, gender, skin color, age, and ethnicity.

F. Summary

In this chapter the BOSS pipeline was evaluated using low resolution images captured from an iPhone camera. The challenges of pose, illumination, expression, resolution, and occlusion were defined and described in detail, and their effect on image acquisition. These challenges were induced by photographing 21 subjects of varying sex, height, weight, and ethnicity, while changing parameters such as pose (yaw, pitch, and roll), illumination, expression, and distance. The test setup was described in great detail, as well as the data collection process. These subjects provided 1700+ images which were used to design, test, and evaluate the entire BOSS pipeline. Face detection rates were obtained, and the feature descriptors

for image-based and 3D reconstructions, from images, were used for facial representation. Stereo (dual channel) was not implemented. The system was evaluated over three distances indoors.

Intensive testing and data analysis illustrated the challenges of fully automated face recognition in the wild; yet, it also motivated use of widespread devices in modern day life, such as smart phones, to perform useful facial biometric tasks. The following chapter will address this matter further.

V. FACIAL BIOMETRICS ON PORTABLE DEVICES & SMART PHONES

A. Introduction

In this chapter we discuss feasible facial biometrics on the cell phone and how cloud computing may be used for distributed facial biometrics in various practical applications, including surveillance, security, disaster relief and healthcare. We should state at the outset that facial biometrics in the wild is expensive in computing and one has to be modest when asking smart phones with limited storage and CPU power to perform like a BOSS system. Yet, the technology of smart phones is improving and the algorithms of facial biometrics are developing.

References on facial biometrics on smart phones and cloud computing is sketchy at best; there is neither standard nor details of systems performance and scenarios of use. This thesis builds on the experience gained from BOSS and discovers smart phones and cloud computing in two respects: 1) image-based facial biometric algorithms that would be able to implement a "single-channel" version of BOSS on cell phones; 2) study of potential use of cloud computing to build a distributed biometric network. The first issue will highlight an implementation of the CVIP Lab approach (e.g., Rara et al., 2010 [48]) from generating 3D images from 2D images

and a database. The second issue will discuss network topologies that can incorporate cell phones as nodes of "smart biometrics" units.

B. Building a "Single-Channel" BOSS for Cell Phones

1. Image-Based Computing

In his MS thesis, Rara, 2006 [51] investigated data reduction techniques for face recognition and suggested that principal component analysis (PCA), independent component analysis (ICA) and linear discriminant analysis (LDA) may be used individually or together in order to perform face recognition, using the Eigen faces [5.5] or the Fisher faces approach [5.6].

Following the same steps in Nes, 2003 [50] and Rara, 2006 [51], the Eigen faces or the fisher faces approach can be executed in two steps: a) pre-processing and b) construction of the Eigen/Fisher faces, using a database. The recognition, based on PCA, ICA or LDA, can be then conducted using any distance similarity measure (e.g. Mahalanobis Distance, the cosine distance or the least square error distance). The quantities have been programmed in various forms (e.g., Matlab, C++, C#); a Java-based approach would be more suitable for cell phones. Below we describe the general approach.

The preprocessing step is crucial in any face recognition system; it removes superficial image noise that may result in degradation of classification accuracy. Each face image in the databases undergoes the following normalization procedure: (a) integer to float conversion (b) geometric normalization (c) masking (d) histogram equalization and (e) pixel normalization. As in Nes, 2003 [50], the geometric normalization consists of lining up the eye coordinates of the face because it is inherent for humans to tilt the face sideways when posing for a camera. Figure 47 shows a prototype of an unaligned and aligned face, in terms of eye coordinates. The angle α

in the original image can be defined as *arctan* (h_{diff}/ w_{diff}), where h_{diff} is the vertical eye coordinates difference and w_{diff} is the horizontal eye coordinates difference. A positive angle would require a counter-clockwise rotation while a negative angle will result into a clockwise rotation of the image. The next step will be to scale the image by setting the distance between the eyes on a user-defined constant. The scaling factor will be *eyedistance*/w_{diff}, where *eyedistance* is a user-defined constant and w_{diff} is the measured new horizontal eye coordinates.

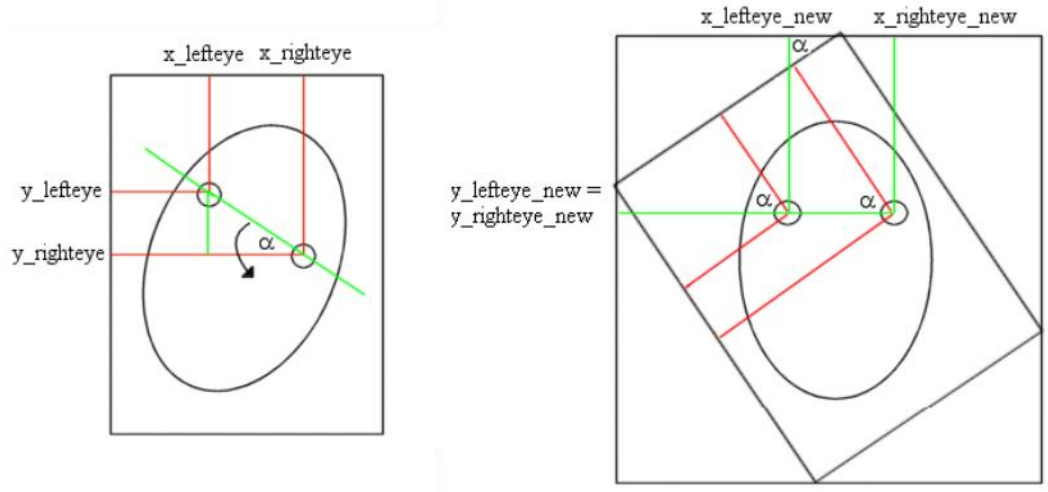


FIGURE 47 - Original (left) Image and Rotated Image (right) (adapted from [50] [51])

The cropping of the image follows Figure 48, with the desired size defined by *norm_height* and *norm_width*. The parameter *eyerow* defines how many pixels in vertical direction above the eyes should the cropping start. The equations for the new eye coordinates, following a counter-clockwise positive rotation α, consists of the following:

$$\begin{aligned} x_{new} &= x*\cos(\alpha)+y*\sin(\alpha) \\ y_{new} &= (imagewidth-x)*\sin(\alpha)+y*\cos(\alpha) \end{aligned} \tag{0.1}$$

The equations for the new eye coordinates following a clockwise negative rotation α are:

$$x_{new} = x * \cos(-\alpha) + (imageheight - y) * \sin(-\alpha)$$
$$y_{new} = x * \sin(-\alpha) + y * \cos(-\alpha) \quad (0.2)$$

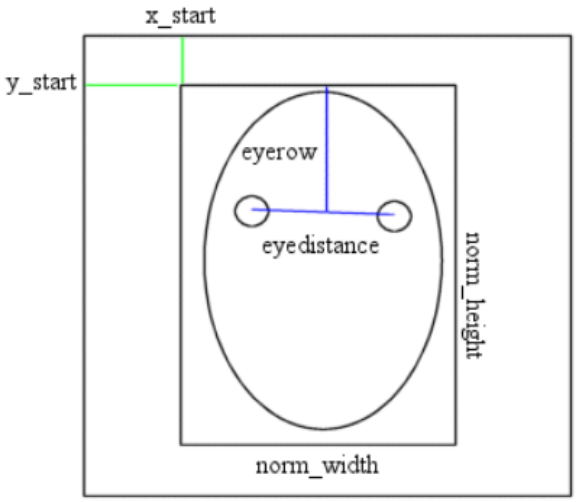


FIGURE 48 - Cropping of the Image (e.g., [50][51])

The starting coordinates for cropping are:

$$x_{start} = x_{lefteye\_new} * scale - (normwidth - \text{eyedistance})/2$$
$$y_{start} = y_{lefteye\_new} * scale - eyerow \quad (0.3)$$

The resulting images are shown, after masking and histogram equalization, in Figure 49.

FIGURE 49 -FERET Images of a Subject after Normalization Steps (Rara, 2006 [51])

Of course, fancier version of cropping based on Active Appearance Modeling (AAM) may be implemented as well (Elhabian and Farag, 2009 [52]). This approach constructs an AAM

model around the landmarks, and generates a mesh, under wish the facial information is cropped (e.g., Figure 50).

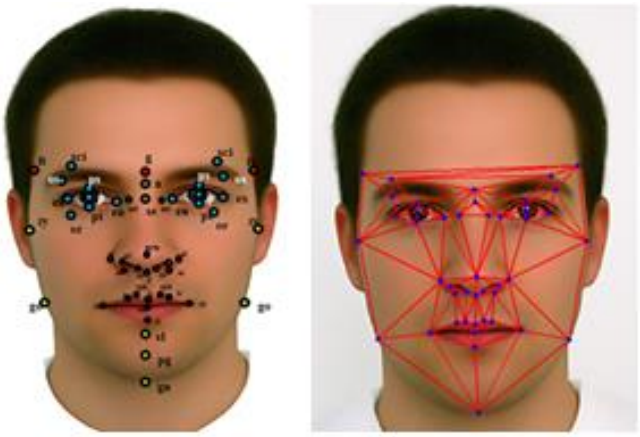


FIGURE 50 - Face cropping based on Active Appearance Modeling (AAM) (e.g., [53])

The advantage of AAM modeling is that it is beneficial for face synthesis using approaches like the Lukas-Kanade algorithm [53] (see also Mathews and Baker, 2004 [54]).

We could densify the mesh (increase the number of vertices) used in cropping (see Figure 51) [55]. The PCA, ICA and LDA approaches used in Eigen Faces/Fisher Faces would necessitate image registration; hence, proper cropping using specified landmarks would simplify this process. Ideally, feature-based approaches for face recognition would be such that the features from the cropped regions would be normalized in a manner that is less sensitive to registration errors.

In so far as implementing the image-based approaches (e.g., Eigen/Fisher Faces), which is the focus of this chapter, we may use any of the well-established implementations in the literature and port to cell phones. The implementation in Xi, 2012 [49] is based on Java.

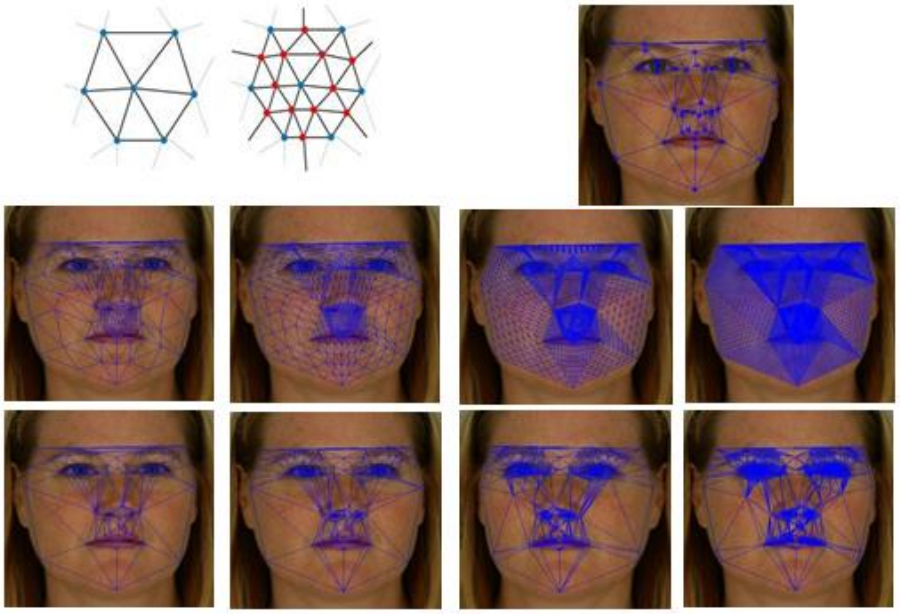


FIGURE 51 - Densified meshes starting from level 1 to level 4. Top row shows the output of loop subdivision, while the bottom row shows the meshes after filtration using a cornerness criterion (CVIP Lab 2011 Report, pp. 12[55])

2. 3D Reconstruction

The approach developed by Rara, et al., 2009 [48] at the CVIP Lab is most adequate for generating 3D versions of cropped facial regions. We briefly describe this approach below.

Using the concept of spherical harmonics, we can efficiently represent the set of images of objects under varying illumination as a linear combination of harmonic images [56]. Then the image I_i of p_i is:

$$I_i = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} l_{lm}\, b_{lm}(p_i) \qquad (5.4)$$

where p_i denotes the ith point on the surface of the object, and $b_{lm}(p_i)$ are the harmonic images. The equations for the 1st nine harmonic images are (e.g., Basri and Jacobs, 2003 [56]):

$$
\begin{aligned}
b_0 &= c_0\lambda & b_1 &= c_1\lambda{\cdot}n_x & b_2 &= c_2\lambda{\cdot}n_y \\
b_3 &= c_3\lambda{\cdot}n_z & b_4 &= c_4\lambda{\cdot}\left(3n_{z^2}-1\right) & b_5 &= c_5\lambda{\cdot}n_{xy} \\
b_6 &= c_6\lambda{\cdot}n_{xz} & b_7 &= c_7\lambda{\cdot}n_{yz} & b_8 &= c_8\lambda{\cdot}\left(n_{x^2}-n_{y^2}\right)
\end{aligned}
\tag{5.5}
$$

where λ denotes albedo and $\boldsymbol{n} = (n_x, n_y, n_z)$ is the surface normal, $(\cdot)$ is a component-wise operator, and $n_{xy} = n_x \cdot n_y$.

As an example Figure 52 shows the first nine Spherical Harmonics (SPH) constructed on USF database ([48][57]).

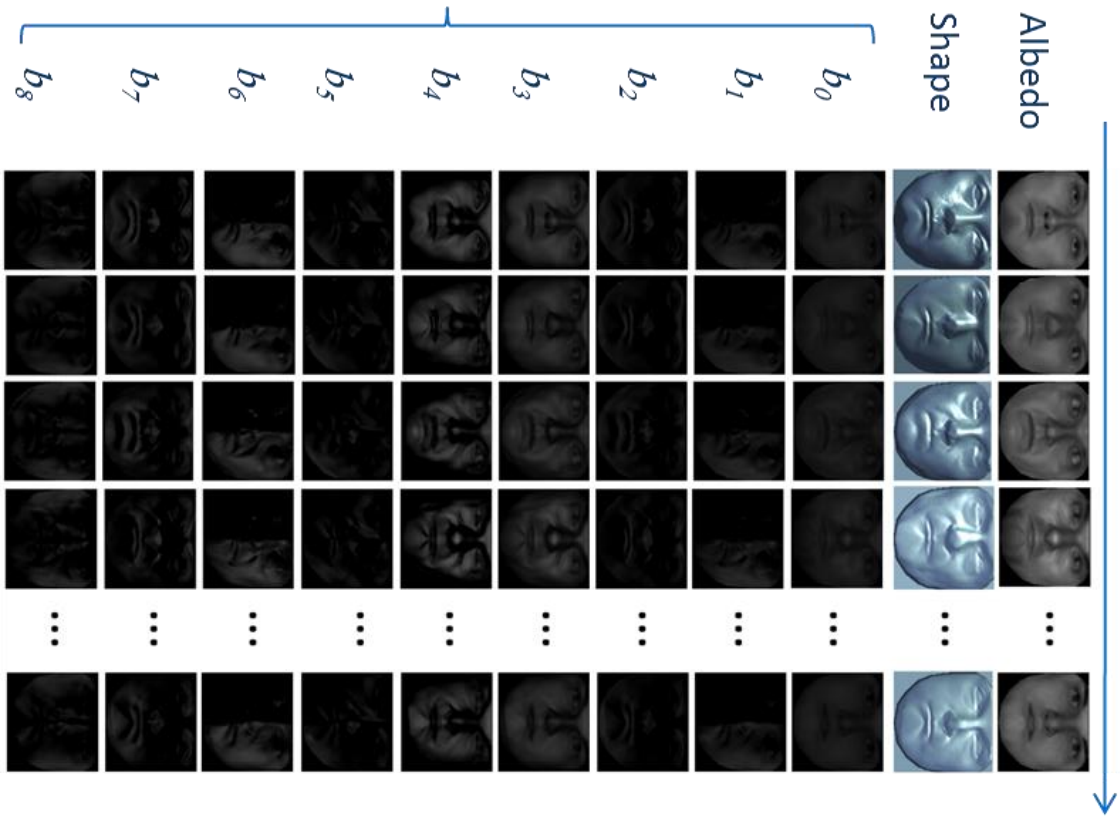


FIGURE 52 - Nine Spherical Harmonics generated from database of albedo and shape

The goal is to reconstruct 3D facial shape from a single 2D input image of general and unknown lighting. Common shape-from-shading algorithms require/ estimate the light source direction under the assumption of single light source. However real-life applications have multiple and unknown light sources. Rara et al., 2010 [48] developed a new statistical shape-from-shading framework for images of unknown illumination, we make use of recent results that

general lighting can be expressed using low-order spherical harmonics for convex Lambertian surfaces. Using Partial Least Squares, 3D face reconstruction is accomplished in a computationally effective manner.

Figure 53 shows the spherical harmonic projection images for different subjects with known 3D shape and albedo given an input image, where the projection images share the same illumination as the input image.

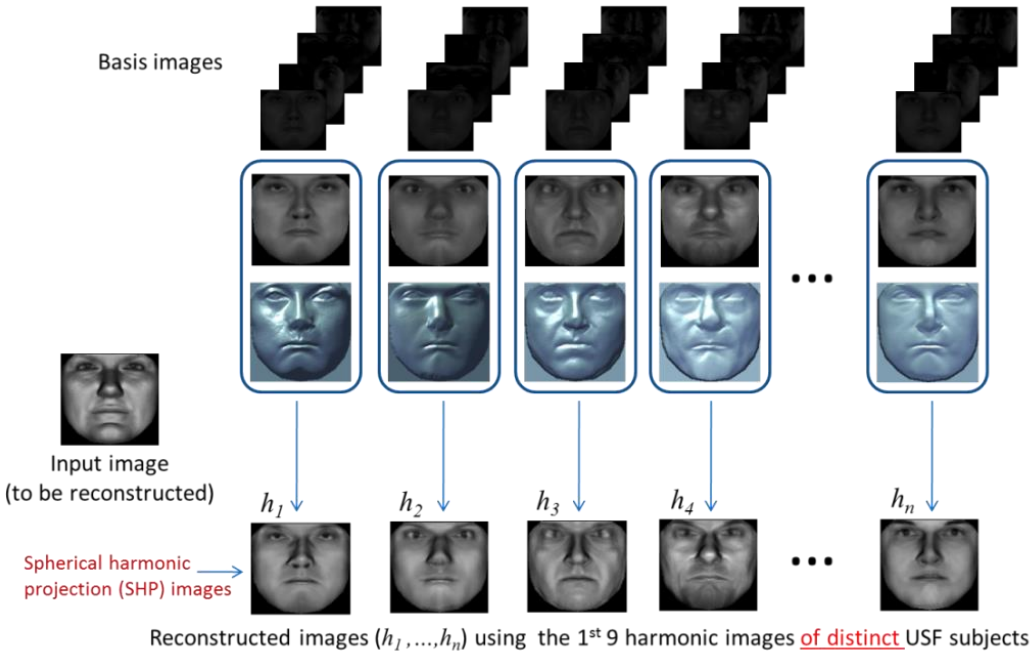


FIGURE 53 - Spherical harmonics for objects under varying illumination

Figure 54 illustrates the framework ([48]) for model-based shape recovery for general and uknown lighting. Figure 55 shows the performance of the method on the USF and Yale database. Visual inspection on the Yale database reconstructions reveal realistic recovered shape and albedo.

This is the approach used in the BOSS project and is programmed in C++ and C#. It would be the most logical approach to deploy over cell phone using Java.

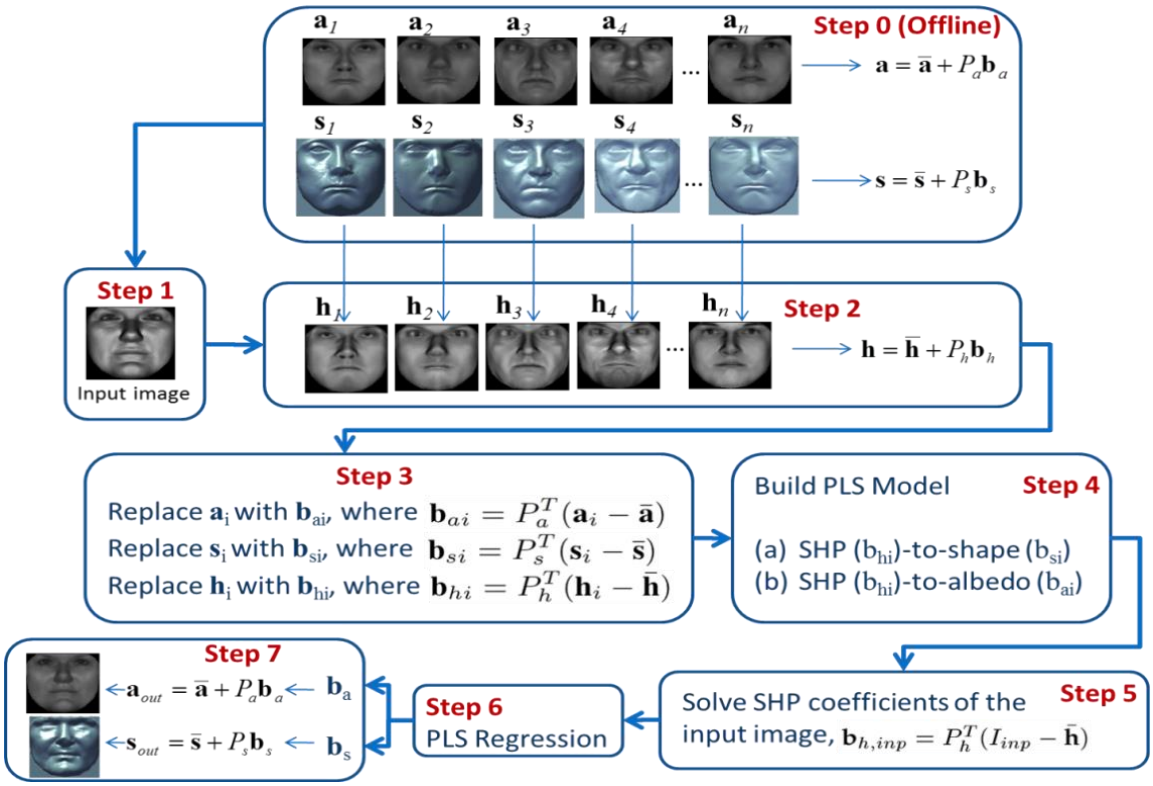


FIGURE 54 - Block diagam of our statistical-shape-from-shading

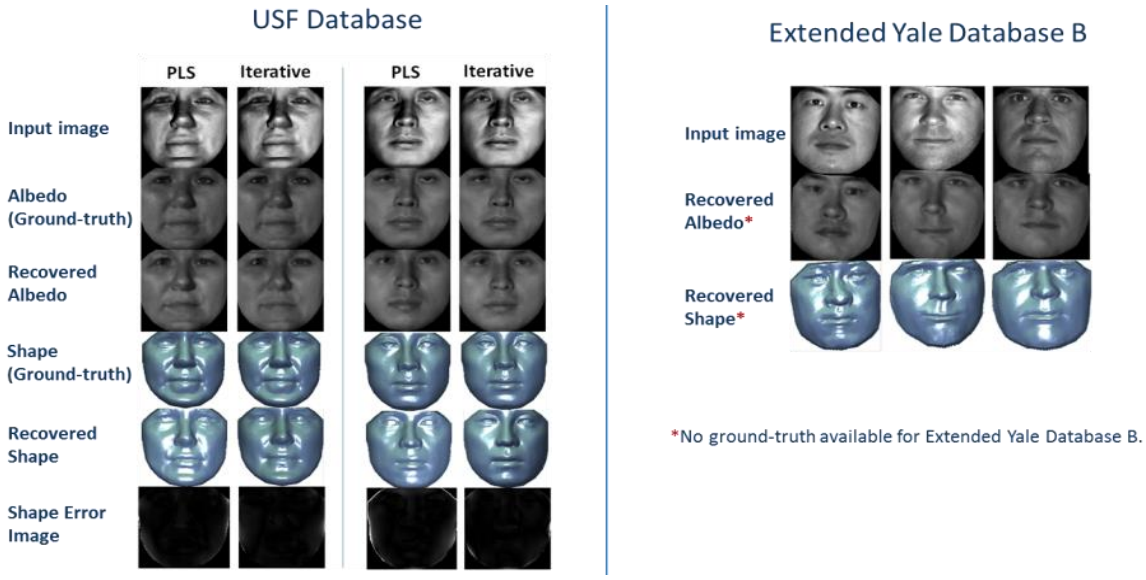


FIGURE 55 - Experimental results, (left) using groundtruth shape and albedo of the USF database and (right) using the extended Yale database.

3. Fusion of Approaches

The image based and 3D reconstruction approaches can also be fused to in order to get features from each approach and a decision fusion method can combine the two approaches. The two approaches do not require a huge database for design. Sparse 3D reconstruction may be optimal, as the information holding the largest discrimination are within a few landmarks; hence a dense mesh (see Figure 51) may not be necessary.

A test bed to develop a cell-phone implementation would use much more data than what was used in Chapter 4 to evaluate BOSS using the iPhone 4 lenses. There is no standard database on social media (e.g., Facebook, google, etc.) that possesses the variations in illumination, pose and expressions required for training the 3D reconstruction approach above; yet, one expects that such a database would be available soon in lieu of the challenge studies held in conjunction with computer vision and biometric meetings.

C. Summary

This chapter considered image-based approaches for facial biometrics on the cell phone. Classical approaches using PCA, ICA and LDA used in Eigen/Fisher Faces would be possible to deploy on cell phones (e.g., Xi [49]); a 3D approach would be challenging for dense reconstruction. A sparse reconstruction approach will be most adequate. One of the issues with facial biometrics on cell phone is the standards and lack of availability of test databases. Follow-up research would need to address these issues. Deploying biometrics on the cloud is feasible, yet would depend on the circumstance of the networks and the intended application.

VI. CONCLUSIONS AND FUTURE DIRECTIONS

A. Conclusions

Face recognition in the wild connotes recognition of individuals unabated by age, pose, illumination, expression (A-PIE), and uncertainties from the imaging scenario (e.g., distance, crowd, action) or mechanism of imaging (still or video cameras, or partial information from non-traditional sources, such as a newspaper photo, face-book image, etc.). In that sense of generality, the information content in an image of an individual is challenged to identify the individual, by the computer, under uncertainty. It is a daunting task and very interesting domain of research. In this thesis, the term "Face Recognition in the Wild" has been defined as unconstrained face recognition under A-PIE+; the (+) will connote any alterations to the design scenario of the face recognition system.

The thesis used the BOSS project at the CVIP Lab as a kernel to study and evaluate face recognition in the wild. This chapter is a summary of BOSS, the research plan, and conclusions of testing it using a low resolution iPhone 4 camera. The chapter also summarizes ideas on using smart phones for face recognition. Finally, the chapter contains recommendations to further work in the domain of face recognition in the wild.

The thesis took a view of evaluating BOSS in a scenario different than the testing scenario used in summer 2012. The same BOSS project was used in this evaluation; however, scaling factors for input images were modified for use with smaller, low resolution images. In particular, the following approach for sensors and testing were used:

- a low resolution camera (iPhone 4) is used rather than the Canon EOS 7D high resolution camera;
- portable data gathering and computing (images are transported to the hard disk using "Quickshot for Dropbox" applet connection) and a quad core Alienware laptop is used for computation, instead of the 8 CPU units on which the system was tested;
- indoor scenario for data collection, and distance of 5, 10, and 15 feet;
- data collected for angles (0, 15° and - 15°) for pose, two illuminations and two expressions; and
- Twenty-one subjects (mix of gender, ethnicity, and skin color).

The same system thresholds were used in evaluating the performance.

The main findings of the test were:

i. low-resolution cameras and a laptop may be used to implement a portable version of BOSS;
ii. performance degrades with distance;
iii. moderate pose did not degrade the performance significantly;
iv. moderate expression led to some degradation in the performance but not significantly; and

v. large gallery reduces the speed of execution; hence, the need for optimal search methods.

Given the findings of this thesis, a smart phone option may be feasible, given the constraints of the distance and the imaging conditions. Chapter 5 studied basic ingredients of face detection, cropping and recognition using smart phones.

B. Future Directions

The research under detection, facial feature representation and matching is ongoing elsewhere, and the literature is quite rich; in fact, there are a number of journals and annual meetings on biometrics. The research trend in facial biometrics exploits all advances in related fields such as image analysis, computer vision and machine learning. The applications of facial biometrics dictate focus on particular frameworks suitable for the circumstances of data, desired accuracy levels, and speed of execution.

From the biometrics technology prospective, sensors will always improve in accuracy and miniaturization; hence, portable facial biometrics will evolve and will improve in accuracy and speed of performance.

With the use of Cloud computing, the perceived applications of facial biometrics on cell phones is their deployment in scenarios such as disaster relief, crowd control, law enforcement on highways, and in surveillance of certain individuals. Healthcare applications and telemedicine may also include biometrics for verification of individuals, and even prescription of medicine. This would require proper use of network topologies and standards. Smart phone biometrics could also be capable of performing useful home-based medical services, which would benefit the elderly, people with special needs, and in fighting drug abuses.

REFERENCES

1. Paul Ekman and Erika Rosenberg, Ed., What the Face Reveals, Oxford University Press, Oxford, UK, 2005.
2. T. Kanade, "Computer Recognition of Human Faces," Birkhauser, 1973.
3. D. Terzopoulos, Y. Lee and M. Vasilescu, "Model-based and image-based methods for facial image synthesis, analysis and recognition," Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004, pp., 3-8.
4. R. Brunelli and T. Poggio, Face Recognition: Features versus Templates, IEEE Transactions on PAMI, 1993, Vol.15 (10):1042-1052.
5. P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," Proc. Computer Vision and Pattern Recognition, CVPR 2001, pp. 1-8.
6. Erik Hjelmås and B. Low, "Face Detection: A Survey", Computer Vision and Image Understanding, CVIU, 2001, Vol. 83, pp. 236-274.
7. Ming-Hsuan Yang, D. Kriegman and N. Ahuja, "Detecting faces in images: a survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI, 2002, vol.24, no.1, pp.34-58.
8. J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. Journal of the Optical Society of America A, 2(7):1160–1169, July 1985.
9. T. Ojala, M. Pietikainen, and T. Maenpaa, (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence (24): 971-987.
10. D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," Int. J. Comput. Vision, 2004, vol. 60, no. 2, pp. 91-110.
11. Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), 2008, Vol. 110, No. 3, pp. 346–359.
12. M. Turk ,and A. Pentland, "Face recognition using eigenfaces", in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1991, pp. 586-591.
13. H. Ling, S. Soatto, N. Ramanathan, W. Jacobs, "A Study of Face Recognition as People," Int. Conference on Computer Vision, ICCV, 2007, pp. 1-8.

14. J. Wang, Y. Shang, G. Su, and X. Lin, "Age Simulation for Face Recognition," Proc. Int'l Conf. Pattern Recognition, pp. 913-916, 2006.
15. J. Suo, F. Min, S. Zhu, S. Shan, and X. Chen, "A Multi-Resolution Dynamic Model for Face Aging Simulation," Computer Vision and Pattern Recognition, pp. 1-8, 2007.
16. U. Park, Y. Tong and A. Jain, "Age-Invariant Face Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI, 2010, vol.32, no.5, pp. 947-954.
17. Eslam Mostafa, and Aly Farag, "Dynamic Weighting of Facial Features for Automatic Pose-Invariant Face Recognition," in Proceedings of Ninth Conference on Computer and Robot Vision, 2012, pp. 411-416.
18. Shirren Elhabian, Eslam Mostafa, Ham Rara, and Aly Farag, "Non-Lambertian Model-based Facial Shape Recovery from Single Image Under Unknown General Illumination," in Proceedings of Ninth Conference on Computer and Robot Vision, 2012, pp. 252-259.
19. Aly Farag, Mike Miller, Ham Rara, Shireen Elhabian, Mostafa Abdelrahman, Eslam Mostafa, Ahmed Elbarkouky and Moumen Elmelegy, Biometric Optical Surveillance System (BOSS), TR-12-2012, CVIP Lab, University of Louisville, KY.
20. Eslam Mostafa, Asem Ali, Naif Alajlan, and Aly Farag, "Pose invariant approach for face recognition at distance," in Proceedings of European Conference on Computer Vision - Volume Part VI, 2012, pp. 15-28.
21. Eslam Mostafa, Moumen Elmelegy, and Aly Farag, "Passive Single Image-based Approach for Camera Steering in Face Recognition at-a-Distance Applications," in Proceedings of IEEE Conference on Biometrics: Theory, Applications and Systems, 2012, pp. 371-376.
22. P. N. Belhumeur, D. J. Kriegman, What is the set of images of an object under all possible lighting conditions?, Int'l Journal of Computer Vision, vol. 28, no. 3, pp. 245-260, 1998.
23. R. Ramamoorthi, Analytic PCA construction for theoretical analysis of lighting variability in images of a Lambertian object, IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI), 2002, Vol. 24, No. 10, pp. 1322-1333.
24. R. Basri, D. W. Jacobs, Lambertian reflectance and linear subspaces, Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.25, no.2, pp. 218- 233, Feb 2003.

25. S. Elhabian and A. Farag, “Analytic Bilinear Appearance Subspace Construction for Modeling Image Irradiance under Natural Illumination and Non-Lambertian Reflectance,” in Proceedings of Computer Vision and Pattern Recognition, CVPR, 2013.
26. P. Ekman and W. Friesen. Pictures of Facial Affect. Palo Alto, CA: Consulting Psychologist, 1976.
27. C. Izard, L. Dougherty, and E. Hembree. A system for identifying affect expressions by holistic judgments. In Unpublished Manuscript, University of Delaware, 1983.
28. P. Ekman and W. Friesen. The Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, San Francisco, 1978.
29. Y. Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. IEEE Trans. on Pattern Analysis and Machine Intell., 23(2):1–19, 2001.
30. V. Blanz, and T. Vetter, "Face recognition based on fitting a 3D morphable model," IEEE Trans. on Pattern Anal. and Mach. Intell., vol. 25, no. 9, pp. 1063-1074, Sep. 2003.
31. Barry-John Theobald, Iain Matthews, Jeffrey F. Cohn, and Steven M. Boker, “Real-time expression cloning using appearance models,” International Conference on Multimodal Interfaces. ACM, November 2007.
32. Leslie Farkas, Anthropometry of the Head and Face, 2nd Edition, Lippincott-Raven Publishers 1994.
33. American National Standard for Information Systems (ANSI/NIST) – Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information. NIST – Special Publication 500-271, May 2007.
34. Yoav Freund and Robert E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” Journal of Computer and System Sciences, Vol. 55(1) pp.119-139, 1997.
35. T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI, 2006, vol. 28, no. 12, pp. 2037-2041.
36. Bicego, M.; Lagorio, A.; Grosso, E.; Tistarelli, M., "On the Use of SIFT Features for Face Authentication," Computer Vision and Pattern Recognition Workshop, 2006, pp. 35-35.

37. P. Dreuw, P. Steingrube, H. Hanselmann, H. Ney, and G. Aachen, G, “SURF-Face: Face Recognition Under Viewpoint Consistency Constraints,” British Machine Vision Conference, BMVC’09, 2009, pp. 1-11.
38. Eslam Mostafa, and Aly Farag, "Dynamic Weighting of Facial Features for Automatic Pose-Invariant Face Recognition," in Proceedings of Ninth Conference on Computer and Robot Vision, 2012, pp. 411-416.
39. Shirren Elhabian, Eslam Mostafa, Ham Rara, and Aly Farag, "Non-Lambertian Model-based Facial Shape Recovery from Single Image Under Unknown General Illumination," in Proceedings of Ninth Conference on Computer and Robot Vision, 2012, pp. 252-259.
40. A. Ali, M. Miller, T. Starr, and A. Farag. Passive stereo-based 3d human face reconstruction at a distance. Technical report, CVIP Lab, Univ. of Louisville, Jan. 2010.
41. P. N. Belhumeur, J. Hespanha, and D. J. Kriegman, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell., 19(7):711–720, 1997.
42. T. Cootes and C. Taylor. Statistical models of appearance for computer vision. Technical report, Univ. of Manchester, UK, Mar. 2004.
43. G. Medioni, J. Choi, C.-H.Kuo, and D. Fidaleo. Identifying noncooperative subjects at a distance using face images and inferred three-dimensional face models. IEEE Trans. Sys. Man Cyber. Part A, 39(1):12–24, 2009.
44. H. Rara, S. Elhabian, A. Ali, T. Gault, M. Miller, T. Starr, and A. Farag.A framework for long distance face recognition using dense - and sparse-stereo reconstruction. In ISVC ’09: Proceedings of the 5th International Symposium on Advances in Visual Computing, pages 774–783, Berlin, Heidelberg, 2009. Springer-Verlag.
45. H. Rara, S. Elhabian, A. Ali, M. Miller, T. Starr, and A. Farag. Face recognition at-a-distance based on sparse stereo reconstruction. Computer Vision and Pattern Recognition Workshop, 0:27–32, 2009.
46. Jing Xiao, Simon Baker, Iain Matthews, Takeo Kanade, "Real-Time Combined 2D+3D Active Appearance Models," Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, pp. 535-542, 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04) - Volume 2, 2004.

47. Amberg, B.; Blake, A.; Vetter, T.; , "On compositional Image Alignment, with an application to Active Appearance Models," Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on , vol., no., pp.1714-1721, 20-25 June 2009.
48. Ham Rara, Shireen Elhabian, Thomas Starr, and Aly Farag, "3D Face Recovery from Intensities of General and Unknown Lightning Using Partial Least Squares," *Proc. of 2010 IEEE International Conference on Image Processing (ICIP)*, pp. 4041-4044, 2010.
49. Kai Xi, Biometric Security System Design: From Mobile to Cloud Computing Environment, PhD Dissertation, University of New South Wales, Australia, 2012.
50. A. Nes, "Hybrid Systems for Face Recognition," MS Graduate Thesis, Norwegian University of Science and Technology, June 2003.
51. Ham Rara, "Dimensionality Reduction Techniques in Face Recognition," MS Thesis, CVIP Lab, University of Louisville, May 2006.
52. Shireen Elhabian and Aly Farag, "Anthropometric-based Face Recognition At-a-Distance Based on Sparse Stereo Reconstruction," TR-12-2009, CVIP Lab, University of Louisville, December 17, 2009.
53. Bruce D. Lucas and Takeo Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," International Joint Conference on Artificial Intelligence, pages 674–679, 1981.
54. Iain Matthews and Simon Baker, "Active Appearance Models Revisited," International Journal of Computer Vision, Volume 60 Issue 2, November 2004, pp. 135 – 164.
55. Computer Vision and Image Processing Report, (CVIP Lab TR-11-6), University of Louisville, June 11, 2011 pp. 12.
56. R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," IEEE Trans. PAMI., vol. 25, no. 2, pp. 218–233, 2003.
57. S. Sarkar, USF HumanID 3D Face Database, University of South Florida, Tampa, FL, USA.

APPENDIX A: BOSS Hardware List

Hardware Equipment List:

1. Canon EOS 7D digital SLR camera, with live view
2. All the software described in this document is written in C#, C++, or MATLAB with compiled code as well as source code delivered to customer. Some code were developed in Matlab and later translated to C++ or C#.
3. PC processor: an Intel i7-3960X Extreme Edition Sandy Bridge-E 3.3 GHz Processor, with 32GB is memory installed.
4. Pan Tilt Unit: PTU D100E manufactured by FLIR Systems, Inc. With a maximum payload of 25lbs and an angular resolution of 0.0075˚ the D100E provides an ideal platform for accurate and smooth positioning of the optics.
5. A V-Infinity model supplying 24VDC @ 8.3A was chosen to power the PTU's.

Camera: The imaging hardware chosen for the BOSS system is a Canon EOS 7D digital SLR camera. The 7D has a maximum image resolution of 18 megapixels, auto white balance, auto focus, and live view functionality among other features. The SDK allows for the development of custom software to control and access theses pertinent features making the 7D an ideal choice for software controlled, semi-automatic long-range facial recognition application system. The wide variety of prime and zoom lenses that are compatible with the 7D allows BOSS to be configured for a wide range of standoff distances.

PTU: A high resolution PTU (Pan-Tilt Unit) (PTU) is required to accurately position the cameras of the BOSS system. An emphasis was placed on the angular resolution capability of the PTU given the large standoff distances required for the BOSS system deployment. The PTU would also have to be rated for the load of the 7D digital camera and largest lens anticipated for maximum standoff distance.

The unit found to be most appropriate was the PTU D100E manufactured by FLIR Systems, Inc. With a maximum payload of 25lbs and an angular resolution of 0.0075˚ the D100E provides an ideal platform for accurate and smooth positioning of the optics. It is a serial device communicating via the RS232 protocol through a protocol converting USB adapter. A built-in command set offers both ASCII and binary formats. This command set supports real-time control at up to 60 commands/second with very low and predictable latencies that are ideal for subject tracking.

Power Supplies: The BCU (Biometric Collection Unit) requires a power supply for the camera, pan-tilt unit, and data transmission module. Power supplies can be large and heavy making it critical to choose one that is not over -designed for the application. The power supply needed to convert 110-120VAC to DC. The pan-tilt unit is the highest voltage component of the BCU, therefore it was necessary to choose a power supply that would support this requirement. Voltage requirement for other devices would be met by designing a simple circuit board to step down and regulate the power supply's output to the appropriate voltages. Although this introduces inherent inefficiencies in the system regarding power loss, it is preferred over incorporating multiple power supplies. In addition, since the power is ultimately supplied from an outlet, the source is

effectively not limited. The power distribution circuit was designed to be as efficient as possible. A V-Infinity model supplying 24VDC @ 8.3A was chosen to power the BCU. This model has a built-in fan drive to power DC fans for convective cooling of the housing that contains the data transmission module, circuit board components and the power supply itself. There is an over temperature shut down feature to protect the unit from catastrophic damage.

CPU: The processing unit will dictate how fast instructions are processed and ultimately, how quickly a result is presented to the user. Since the application deals with many large images and ultimately very large amounts of data, memory for the system is also very important. Because of this, the fastest single package processor was selected along with an ample amount of RAM. The processing power for the system comes from an Intel Core i7-3960X Extreme 3.3GHz Six-Core CPU. This rests inside of an ASUS P9X79 over-clockable motherboard. Memory for the system is provided by 32GB of G. Skill Ripjaws 240-Pin DDR3 SDRAM and a Corsair Force Series 3 480GB Solid State Hard Drive. The system has been overclocked with the BIOS tool to run stable at ~4.2 GHz. Graphics are rendered and displayed through the EVGA GeForce GTX 580 1536MB 384-bit GDDR5 Graphics Card which supports CUDA development. Power to each component is supplied by the Thermaltake Toughpower 1475W Power Supply. All components are housed inside a Mid Tower Silverstone Kublai KL04W computer case to help minimize the size and reduce shipping costs.

APPENDIX B: Modified BOSS Code

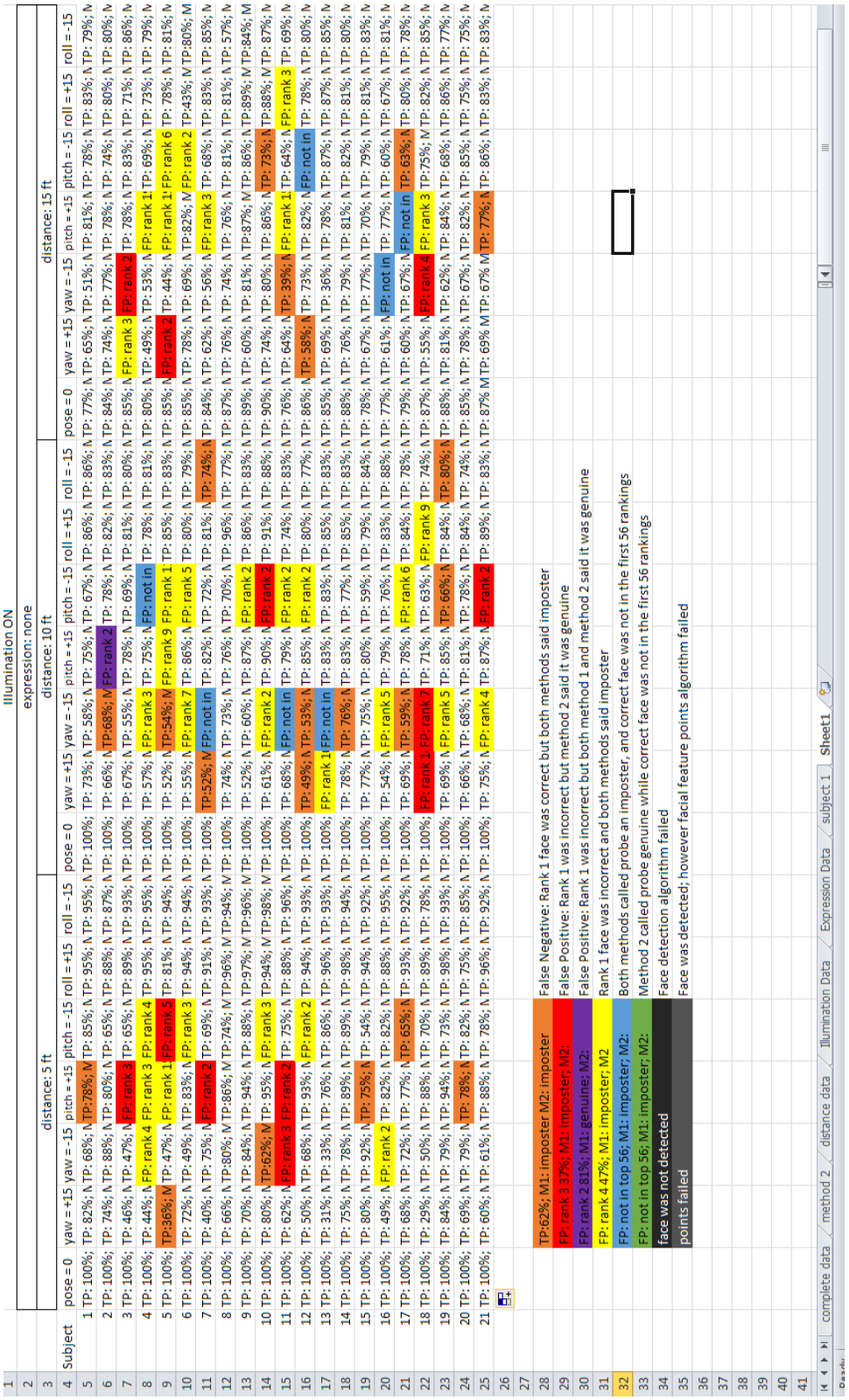
```
    public void FaceDetectGray_Illum_Skin(Image<Bgr, byte> imC, out List<FaceData> Faces, out List<FaceData>
MaxFace, string xml_fnameF,
        string xml_fnameL, string xml_fnameR, string xml_fnameM, Matrix<float> smodel)
    {
        Image<Gray, byte> im = imC.Convert<Gray, byte>();
        Faces = new List<FaceData>(); // Init
        MaxFace = new List<FaceData>();

        // Multiresolution params
        Matrix<Single> Ratio = new Matrix<Single>(3, 1);
        Matrix<Single> params2 = new Matrix<Single>(6, 3);
        if (imC.Width < 1000)
        {
            Ratio[0, 0] = 1; Ratio[1, 0] = 2; Ratio[2, 0] = 3;
            params2[5, 0] = 1.0f; params2[5, 1] = 2.0f; params2[5, 2] = 3.0f;
        }
        else
        {
            Ratio[0, 0] = 6; Ratio[1, 0] = 7; Ratio[2, 0] = 8;
            params2[5, 0] = 6.0f; params2[5, 1] = 7.0f; params2[5, 2] = 8.0f;
        }
        if (imC.Width < 2600)
        {
            Ratio[0, 0] = 2; Ratio[1, 0] = 3; Ratio[2, 0] = 4;
            params2[5, 0] = 2.0f; params2[5, 1] = 3.0f; params2[5, 2] = 4.0f;
        }

        params2[0, 0] = 1.1f; params2[0, 1] = 1.1f; params2[0, 2] = 1.1f;
        params2[1, 0] = 2.0f; params2[1, 1] = 2.0f; params2[1, 2] = 2.0f;
        params2[2, 0] = 0.0f; params2[2, 1] = 0.0f; params2[2, 2] = 0.0f;
        params2[3, 0] = 35.0f; params2[3, 1] = 30.0f; params2[3, 2] = 25.0f;
        params2[4, 0] = 35.0f; params2[4, 1] = 30.0f; params2[4, 2] = 25.0f;
```

APPENDIX C: Sample of Data Sheet

Sample of Data Analysis Spread Sheet (Illumination ON, No Expression, Varying Pose, Varying Distance):



	Illumination ON																							
	expression: none																							
	distance: 5 ft							distance: 10 ft							distance: 15 ft									
Subject	pose = 0	yaw = +15	yaw = -15	pitch = +15	pitch = -15	roll = +15	roll = -15	pose = 0	yaw = +15	yaw = -15	pitch = +15	pitch = -15	roll = +15	roll = -15	pose = 0	yaw = +15	yaw = -15	pitch = +15	pitch = -15	roll = +15	roll = -15
1	TP: 100%;	TP: 82%;	TP: 68%;	N TP: 78%;	M TP: 85%;	N TP: 95%;	N TP: 95%;	TP: 100%;	TP: 73%;	N TP: 58%;	N TP: 75%;	N TP: 67%;	N TP: 86%;	N TP: 86%;	TP: 77%;	N TP: 65%;	N TP: 51%;	N TP: 81%;	N TP: 78%;	N TP: 83%;	N TP: 79%; N
2	TP: 100%;	TP: 74%;	N TP: 88%;	N TP: 80%;	N TP: 65%;	N TP: 88%;	N TP: 87%;	TP: 100%;	TP: 66%;	N TP:68%;	M FP: rank 2	TP: 78%;	N TP: 82%;	N TP: 83%;	N TP: 84%;	N TP: 74%;	N TP: 77%;	N TP: 78%;	N TP: 74%;	N TP: 80%;	N TP: 80%; N
3	TP: 100%;	TP: 46%;	N TP: 47%;	N FP: rank 3	TP: 65%;	N TP: 89%;	N TP: 93%;	TP: 100%;	TP: 67%;	N TP: 55%;	N TP: 78%;	N TP: 69%;	N TP: 81%;	N TP: 80%;	N TP: 85%;	N FP: rank 3	FP: rank 2	TP: 78%;	N TP: 83%;	N TP: 71%;	N TP: 86%; N
4	TP: 100%;	TP: 44%;	N FP: rank 4	FP: rank 3	FP: rank 4	TP: 95%;	N TP: 95%;	TP: 100%;	TP: 57%;	N FP: rank 3	TP: 75%;	N FP: not in	TP: 78%;	N TP: 81%;	N TP: 80%;	N TP: 49%;	N TP: 53%;	N FP: rank 1!	TP: 69%;	N TP: 73%;	N TP: 79%; N
5	TP: 100%;	TP:36%;	M TP: 47%;	N FP: rank 1	FP: rank 5	TP: 81%;	N TP: 94%;	TP: 100%;	TP: 52%;	N TP:54%;	M FP: rank 9	FP: rank 1	TP: 85%;	N TP: 83%;	N TP: 85%;	N FP: rank 2	TP: 44%;	N FP: rank 1!	FP: rank 6	TP: 78%;	N TP: 81%; N
6	TP: 100%;	TP: 72%;	N TP: 49%;	N TP: 83%;	N FP: rank 3	TP: 94%;	N TP: 94%;	TP: 100%;	TP: 55%;	N FP: rank 7	TP: 86%;	N FP: rank 5	TP: 80%;	N TP: 79%;	N TP: 85%;	N TP: 78%;	N TP: 69%;	N TP:82%;	M FP: rank 2	TP:43%;	M TP:80%; M
7	TP: 100%;	TP: 40%;	N TP: 75%;	N FP: rank 2	FP: rank 2	TP: 69%;	N TP: 91%;	TP: 100%;	TP:52%;	M FP: not in	TP: 82%;	N TP: 72%;	N TP: 81%;	N TP: 74%;	N TP: 84%;	N TP: 62%;	N TP: 56%;	N FP: rank 3	TP: 68%;	N TP: 83%;	N TP: 85%; N
8	TP: 100%;	TP: 66%;	N TP:80%;	M TP:86%;	M TP:74%;	M TP:96%;	M TP:94%;	M TP: 100%;	TP: 74%;	N TP: 73%;	N TP: 76%;	N TP: 70%;	N TP: 96%;	N TP: 77%;	N TP: 87%;	N TP: 76%;	N TP: 76%;	N TP: 74%;	N TP: 76%;	N TP: 81%;	N TP: 57%; N
9	TP: 100%;	TP: 70%;	N TP: 84%;	N TP: 94%;	N TP: 88%;	N TP:97%;	M TP:96%;	M TP: 100%;	TP: 52%;	N TP: 60%;	N TP: 87%;	N FP: rank 2	TP: 86%;	N TP: 83%;	N TP: 89%;	N TP: 60%;	N TP: 81%;	N TP:87%;	M TP: 86%;	N TP:89%;	M TP:84%; M
10	TP: 100%;	TP: 80%;	N TP:62%;	M TP: 95%;	N FP: rank 3	TP:94%;	M TP:98%;	M TP: 100%;	TP: 61%;	N FP: rank 2	TP: 90%;	N FP: rank 2	TP: 91%;	N TP: 88%;	N TP: 90%;	N TP: 74%;	N TP: 80%;	N TP: 86%;	N TP: 73%;	N TP:88%;	M TP: 87%; N
11	TP: 100%;	TP: 62%;	N FP: rank 3	FP: rank 2	TP: 75%;	N TP: 88%;	N TP: 96%;	N TP: 100%;	TP: 68%;	N FP: not in	TP: 79%;	N FP: rank 2	TP: 74%;	N TP: 83%;	N TP: 76%;	N TP: 64%;	N TP: 39%;	N FP: rank 1!	TP: 64%;	N FP: rank 3	TP: 69%; N
12	TP: 100%;	TP: 50%;	N TP: 68%;	N TP: 93%;	N FP: rank 2	TP: 94%;	N TP: 93%;	N TP: 100%;	TP: 49%;	N TP: 53%;	N TP: 85%;	N FP: rank 2	TP: 80%;	N TP: 77%;	N TP: 86%;	N TP: 58%;	N TP: 73%;	N TP: 82%;	N FP: not in	TP: 78%;	N TP: 80%; N
13	TP: 100%;	TP: 31%;	N TP: 33%;	N TP: 76%;	N TP: 86%;	N TP: 96%;	N TP: 93%;	N TP: 100%;	FP: rank 1	FP: not in	TP: 83%;	N TP: 83%;	N TP: 85%;	N TP: 83%;	N TP: 85%;	N TP: 69%;	N TP: 36%;	N TP: 78%;	N TP: 87%;	N TP: 87%;	N TP: 85%; N
14	TP: 100%;	TP: 75%;	N TP: 78%;	N TP: 89%;	N TP: 89%;	N TP: 98%;	N TP: 94%;	N TP: 100%;	TP: 78%;	N TP: 76%;	N TP: 83%;	N TP: 77%;	N TP: 85%;	N TP: 83%;	N TP: 88%;	N TP: 76%;	N TP: 79%;	N TP: 81%;	N TP: 82%;	N TP: 81%;	N TP: 80%; N
15	TP: 100%;	TP: 80%;	N TP: 92%;	N TP: 75%;	N TP: 54%;	N TP: 94%;	N TP: 92%;	N TP: 100%;	TP: 77%;	N TP: 75%;	N TP: 80%;	N TP: 59%;	N TP: 79%;	N TP: 84%;	N TP: 78%;	N TP: 67%;	N TP: 70%;	N TP: 79%;	N TP: 81%;	N TP: 81%;	N TP: 83%; N
16	TP: 100%;	TP: 49%;	N FP: rank 2	N TP: 82%;	N TP: 82%;	N TP: 88%;	N TP: 95%;	N TP: 100%;	TP: 54%;	N FP: rank 5	TP: 79%;	N TP: 76%;	N TP: 83%;	N TP: 88%;	N TP: 77%;	N TP: 61%;	N FP: not in	TP: 77%;	N TP: 60%;	N TP: 67%;	N TP: 81%; N
17	TP: 100%;	TP: 68%;	N TP: 72%;	N TP: 77%;	N TP: 65%;	N TP: 93%;	N TP: 92%;	N TP: 100%;	TP: 69%;	N TP: 59%;	N TP: 78%;	N FP: rank 6	TP: 84%;	N TP: 78%;	N TP: 79%;	N TP: 60%;	N TP: 67%;	N FP: not in	TP: 63%;	N TP: 80%;	N TP: 78%; N
18	TP: 100%;	TP: 29%;	N TP: 50%;	N TP: 88%;	N TP: 70%;	N TP: 89%;	N TP: 78%;	N TP: 100%;	FP: rank 1!	FP: rank 7	TP: 71%;	N TP: 63%;	N FP: rank 9	TP: 74%;	N TP: 87%;	N TP: 55%;	N FP: rank 4	FP: rank 3	TP:75%;	M TP: 82%;	N TP: 85%; N
19	TP: 100%;	TP: 84%;	N TP: 79%;	N TP: 94%;	N TP: 73%;	N TP: 98%;	N TP: 93%;	N TP: 100%;	TP: 69%;	N FP: rank 5	TP: 85%;	N TP: 66%;	N TP: 84%;	N TP: 80%;	N TP: 88%;	N TP: 81%;	N TP: 62%;	N TP: 84%;	N TP: 68%;	N TP: 86%;	N TP: 77%; N
20	TP: 100%;	TP: 69%;	N TP: 79%;	N TP: 78%;	N TP: 82%;	N TP: 75%;	N TP: 85%;	N TP: 100%;	TP: 66%;	N TP: 68%;	N TP: 81%;	N TP: 78%;	N TP: 84%;	N TP: 74%;	N TP: 85%;	N TP: 78%;	N TP: 82%;	N TP: 85%;	N TP: 75%;	N TP: 75%;	N TP: 75%; N
21	TP: 100%;	TP: 60%;	N TP: 61%;	N TP: 88%;	N TP: 78%;	N TP: 96%;	N TP: 92%;	N TP: 100%;	TP: 75%;	N FP: rank 4	TP: 87%;	N FP: rank 2	TP: 89%;	N TP: 83%;	N TP: 87%	M TP: 69%	M TP: 67%	M TP: 77%;	N TP: 86%;	N TP: 83%;	N TP: 83%; N

TP:62%; M1: imposter M2: imposter	False Negative: Rank 1 face was correct but both methods said imposter
FP: rank 3 37%; M1: imposter; M2:	False Positive: Rank 1 was incorrect but method 2 said it was genuine
FP: rank 2 81%; M1: genuine; M2:	False Positive: Rank 1 was incorrect but both method 1 and method 2 said it was genuine
FP: rank 4 47%; M1: imposter; M2	Rank 1 face was incorrect and both methods said imposter
FP: not in top 56; M1: imposter; M2:	Both methods called probe an imposter, and correct face was not in the first 56 rankings
FP: not in top 56; M1: imposter; M2:	Method 2 called probe genuine while correct face was not in the first 56 rankings
face was not detected	Face detection algorithm failed
points failed	Face was detected; however facial feature points algorithm failed

APPENDIX D: Sample code for recognition rate curves

Sample MATLAB code produced to plot recognition rate curves (15 feet, illumination off, changing expression):

```
close all;
clc;
clear all;

%15 feet, Illumination OFF, Changing Expression

data_1exp = [0.714285714000000,1,0.619047619000000;
   0.333333333000000,0.666666667000000,0.285714286000000;
   0.666666667000000,1,0.666666667000000;
   0.333333333000000,0.666666667000000,0.476190476000000;
   0.904761905000000,1,1;
   0.619047619000000,0.666666667000000,0.761904762000000;];

angles = [-15 0 15];
% for i = 1:2:36
figure(1)
i = 1;
   plot(angles, data_1exp(i,:),'-r*','LineWidth',3,'MarkerSize', 10); hold on;
   plot(angles, data_1exp(i+1,:),'--bs','LineWidth',3,'MarkerSize', 10); hold on;
   axis ([-20 20 0 1.2]);

   Ha = gca
   set(Ha,'XTickMode','manual');
   set(Ha,'XTick',[-15 0 15]);
   set(Ha, 'fontweight','bold', 'Fontsize', 20);
   box off;

   xlabel('Yaw angle', 'fontsize', 25);
   ylabel('Recognition rate %','fontsize', 25);
   saveas(gcf,'Yaw_15ft_illumOFF.jpg','jpg');

% end
figure(2)
i = 3;
   plot(angles, data_1exp(i,:),'-r*','LineWidth',3,'MarkerSize', 10); hold on;
   plot(angles, data_1exp(i+1,:),'--bs','LineWidth',3,'MarkerSize', 10); hold on;
   axis ([-20 20 0 1.2]);

   Ha = gca
   set(Ha,'XTickMode','manual');
   set(Ha,'XTick',[-15 0 15]);
   set(Ha, 'fontweight','bold', 'Fontsize', 20);
   box off;

   xlabel('Pitch Angle', 'fontsize', 25);
   ylabel('Recognition rate %','fontsize', 25);
   saveas(gcf,'Pitch_15ft_illumOFF.jpg','jpg');

figure(3)
i = 5;
```

```
    plot(angles, data_1exp(i,:),'-r*','LineWidth',3,'MarkerSize', 10); hold on;
    plot(angles, data_1exp(i+1,:),'--bs','LineWidth',3,'MarkerSize', 10); hold on;
    axis ([-20 20 0 1.2]);

    Ha = gca
    set(Ha,'XTickMode','manual');
    set(Ha,'XTick',[-15 0 15]);
    set(Ha, 'fontweight','bold', 'Fontsize', 20);
    box off;

    xlabel('Roll angle', 'fontsize', 25);
    ylabel('Recognition rate %','fontsize', 25);
    saveas(gcf,'Roll_15ft_illumOFF.jpg','jpg');
aa=1;
```

VITA

Mostafa Farag - Received the Bachelors of Science (BS) in Electrical and Computer Engineering, *Magna Cum Laude*, from the University of Louisville in December 2012. He is currently studying for the Masters of Engineering degree (MENG), in Electrical and Computer Engineering at University of Louisville, and is expected to complete his research in Fall 2013. Throughout his study at the University of Louisville he has maintained above 3.8/4.0 GPA, has been on the Dean's List and Governor Scholar. As part of his education, he spent three-semester Co-Op at GE and Louisville Gas and Electric (LG&E), where he worked on trouble shooting of GE newest home refrigerator, and studied the maintenance schedule of the high voltage distribution units for LG&E.

In spring 2012, Mostafa Farag joined the Computer Vision and Image Processing Laboratory (CVIP Lab), as a research assistant on the Biometric Optical Surveillance System (BOSS) project, where he conducted data collection for testing the BOSS system. In spring 2013, he joined Dr. James Graham's Laboratory as a research assistant on a Tele-Health project. Since summer 2013, he joined the CVIP Lab as a research student to complete the research of his MENG thesis on Face Recognition in the Wild, where he has been developing criterion to evaluate facial biometrics systems, such as BOSS, using portable computing and low resolution cameras in unconstrained environments. At the CVIP, he also worked on the Lung Project, and tested the Lab's computer aided design (CAD) system on a number of low contrast CT scans of the chest. This work was presented at the annual meetings Research! Louisville in September 24-27, and at the Brown Cancer Center (BCC) Annual Meeting on October 25, 2013. His poster "Computerized Reading of Chest CT Scans" won the 3rd place award for graduate research students at the BCC meeting. He has published a conference paper: "Reliability and Cyber-Security Assessment of Telehealth Systems" Computer Applications in Industry and Engineering (CAINE-2013) conference, with Dr. James Graham, and will send his findings on Facial Biometrics in the Wild to a technical conference end of this fall 2013.

Extracurricular activities of Mostafa Farag includes: memberships of Tau Beta Pi, IEEE and Kappa Sigma Fraternity. He held secretary of IEEE student chapter in his senior year, and Social Chair and Treasurer of the Kappa Sigma Fraternity.

Mostafa Farag plans to pursue employment with the aviation and aerospace industry. He also plans to pursue further graduate studies in the domain of electrical and computer engineering and computer science.

UNIVERSITY of LOUISVILLE

Mostafa Farag
1006 S Brook St. Louisville, KY 40203
(502) 526-7747 |**mostafa.farag01@gmail.com**

EDUCATION
Dec 2013 — *J.B. Speed School of Engineering, UofL, Louisville, KY*
M. Eng, Electrical Engineering, *GPA*: 3.82/4.0

Dec 2012 — **B. S, Electrical Engineering,** *GPA***:** 3.85/4.0
May 2008 — **Diploma, Ballard High School, Louisville, KY**, *GPA:* 3.7/4.0

EXPERIENCE

Aug '10 – Aug '11 — **LG&E and KU Services Co., Simpsonville, KY**
Co-op III: Transmission Operations
Shadowed principal engineers and was responsible for developing the training documents for the Electric System Coordinators (ESC). These documents are still used by LG&E to train their ESC.

Co-op II: Transmission Operations
Analyzed the reliability of the electric grid to environmental-related outages as well as planned maintenance outages. Verified line ratings between LG&E and KU and TVA along with creating Slider Diagrams in PSSe comparable to the OpenNet electrical grid.

Jan '10 – May '10 — **General Electric, Louisville, KY**
Co-op I: Power Electronics
Designed and tested a buck converter in order to determine the optimal voltage necessary for the Mako II clothes washer inverter. Tested and documented inverter reliability in accordance to six sigma (6 σ) regulations.

OTHER EXPERIENCE

May '13 – Present — **Graduate Thesis – University of Louisville**
- Designing a metric to evaluate the Biometric Optical Surveillance System (BOSS) for facial Recognition, using low resolution and unconstrained imaging, and portable computing.
- Study computer-assisted diagnosis (CAD) systems for analysis of lung nodules in chest CT.

Dec '12 – May '13 — **Graduate Research Assistant – University of Louisville**
Researched different network schemes and privacy issues facing telehealth vendors. Co-authored a conference paper diagnosing possible security risks to vendor's systems, with ways to mitigate security failure; and co-authored two presentations at Research!Louisville and the Brown Cancer Center.

June ’12 – Dec ’12	**Research Assistant – University of Louisville** Worked for Computer Vision and Image Processing (CVIP) Lab assisting Ph.D students in autonomous robotics and facial biometrics research. Developed code for object detection and recognition, was used in CAD system for facial detection of crowds in an uncontrolled environment.
Jan ‘12 – May ‘12	**Senior Design Project – University of Louisville** Project leader; worked with 3 other students to design and build a functioning Solar Powered Charging Station in order to convert solar energy to usable power for charging electric vehicles.
Jan ‘12 – May ‘12	**Computer Design – University of Louisville** Designed custom Digital Integrated Circuits using VHDL. Programmed and demoed Spartan 3-E Automotive FPGA.
Aug ‘11 – Dec ‘11	**Autonomous Robotics Design – University of Louisville** Designed autonomous robot using LEGO® MINDSTORMS®; group placed third of seven in final competition against other classmates.
Jan ‘11 – May ‘11	**Embedded System Design – University of Louisville** Designed and tested various systems using the Motorola MC68HC11microcontroller and Assembly language.
SKILLS	SPICE, MATLAB, Simulink, PLECS, LabVIEW, PSSe, AutoCAD, Solid Edge, Programming (VHDL, C, Java, XML, HTML), Microsoft Office, and familiar with mobile computing
ACTIVITIES & HONORS	• J.B. Speed School Outstanding Student Award • Dean’s Scholar/List throughout college career • Treasurer of Mu-Eta Chapter of The Kappa Sigma Fraternity • Secretary of IEEE Student Branch at U of L • Tau Beta Pi, National Engineering Honor Society • Volunteer at CIVIC Organizations • Governor’s Scholarship (University of Louisville)

PUBLICATION Karla Welch, J. Chris Foreman, James H. Graham, Mostafa Farag, Melinda Whitfield Thomas, Phil Womble, "Reliability and Cyber-Security Assessment of Telehealth Systems," Proc. 26th Intl. Conference on Computer Applications in Industry and Engineering (CAINE-2013), Los Angeles, Sept. 2012, pp. 65-70.

PRESENTATIONS Research!Louisville

- Presented on CAD systems for analysis of lung nodules in chest CT

James Graham Brown Cancer Center Retreat

- Presented on CAD systems for analysis of lung nodules in chest CT
- Won Third Place for Graduate Student Research

REFERENCES

1. Professor James Graham – Department of Electrical and Computer Engineering, J. B. Speed School of Engineering, University of Louisville, E-mail: james.graham@louisville.edu
2. Professor Larry Tyler – Department of Engineering Fundamentals, J. B. Speed School of Engineering, University of Louisville, E-mail:ldtyle01@louisville.edu
3. Mr. Mike Miller – Research Engineer, Computer Vision and Image Processing Laboratory; CVIP Lab, University of Louisville, E-mail: mike.miller@louisville.edu
4. Jane Tanner – Academic Advisor, Department of Electrical and Computer Engineering, J. B. Speed School of Engineering, University of Louisville, E-mail: jltann01@louisville.edu
5. Professor Aly A. Farag – Department of Electrical and Computer Engineering, J. B. Speed School of Engineering, University of Louisville, E-mail: aly.farag@louisville.edu
6. William H. Bicknell – Engineering Manager, GE Appliances, Appliance Park, E-mail: William.Bicknell@ge.com