2013

Effectiveness of Multi-View Face Images and Anthropometric Data In Real-Time Networked Biometrics

Sriram Sankar
West Virginia University

Effectiveness of Multi-View Face Images and Anthropometric Data In Real-Time Networked Biometrics

by

Sriram Sankar

Thesis submitted to the
College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Master of Science
in
Electrical Engineering

Vinod Kulathumani, Ph.D., Chair
Xin Li, Ph.D.
Arun Ross, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2012

Keywords: Biometrics, camera networks, network biometrics,

Abstract

Effectiveness of Multi-View Face Images and Anthropometric Data In Real-Time Networked Biometrics

by

Sriram Sankar
Master of Science in Electrical Engineering

West Virginia University

Vinod Kulathumani, Ph.D., Chair

Over the years, biometric systems have evolved into a reliable mechanism for establishing identity of individuals in the context of applications such as access control, personnel screening and criminal identification. However, recent terror attacks, security threats and intrusion attempts have necessitated a transition to modern biometric systems that can identify humans under unconstrained environments, in real-time. Specifically, the following are three critical transitions that are needed and which form the focus of this thesis: (1) In contrast to operation in an offline mode using previously acquired photographs and videos obtained under controlled environments, it is required that identification be performed in a real-time dynamic mode using images that are continuously streaming in, each from a potentially different view (front, profile, partial profile) and with different quality (pose and resolution). (2) While different multi-modal fusion techniques have been developed to improve system accuracy, these techniques have mainly focused on combining the face biometrics with modalities such as iris and fingerprints that are more reliable but require user cooperation for acquisition. In contrast, the challenge in a real-time networked biometric system is that of combining opportunistically captured multi-view facial images along with soft biometric traits such as height, gait, attire and color that do not require user cooperation. (3) Typical operation is expected to be in an open-set mode where the number of subjects that enrolled in the system is much smaller than the number of probe subjects; yet the system is required to generate high accuracy.

To address these challenges and to make a successful transition to real-time human identification systems, this thesis makes the following contributions: (1) A score-based multi-modal, multi-sample fusion technique is designed to combine face images acquired by a multi-camera network and the effectiveness of opportunistically acquired multi-view face images using a camera network in improving the identification performance is characterized; (2) The multi-view face acquisition system is complemented by a network of Microsoft Kinects for extracting human anthropometric features (specifically height, shoulder width and arm length). The score-fusion technique is augmented to utilize human anthropometric data and the effectiveness of this data is characterized. (3) The performance of the system is

demonstrated using a database of 51 subjects collected using the networked biometric data acquisition system.

Our results show improved recognition accuracy when face information from multiple views is utilized for recognition and also indicate that a given level of accuracy can be attained with fewer probe images (lesser time) when compared with a uni-modal biometric system.

Acknowledgments

I would first like to thank my committee chair and advisor, Dr.Vinod Kulathumani, for giving me the opportunity to work with him and his students. This thesis would not be possible without his constant guidance and support. I would also like to thank Dr. Xin Li and Dr.Arun Ross for being on my committee. I have been fortunate to have had the opportunity to take courses with all of my committee members, and their teachings have been essential to my understanding of the subject. Next, I would also like to thank my colleagues with whom I've had the pleasure of working alongside.

Finally, I would like to express my gratitude to my mother, father and sister. Their support seemingly has no limit, and has been much appreciated throughout my life.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Over the years biometric systems have evolved into one of the most reliable means of human identification. But, due to the recent terror attacks and various security threats, there is an increasing need for a transition to modern biometric systems that establish identity of subjects in real-time. An example application is one where biometric data corresponding to a handful of individuals have been provided to the system, and it is required to generate an alert if one of these individuals has been located within the monitored region (with high accuracy and preferably before the individual has moved too far away from the region). Examples of such scenarios can be found in disaster relief, monitoring secured choke points, port of entries, humanitarian aid centres etc,. Human identification in real-time pose challenges that are illustrated below:

- Camera based modern surveillance have been setup that stream opportunistically extracted face and other data to a server. The data from these cameras is expected to be of different quality. Fusion algorithms have to be designed that collaboratively identify subjects with multiple installations that provide different views of the same subject in real-time.

- In the context of real-time human identification, the probe that is matched cannot be expected to have been enrolled apriory (largely Open Set). Even in such scenarios, the biometric systems are expected to maintain high accuracy.

- Investigation has to be done to check whether the presence of multiple samples of different views of the subject can be exploited to maintain the accuracy of biometric identification system.

- Multimodal biometric system have been installed that compensates the non-universality and scalability issues of uni-modal systems. Most of these multi-modal system use face with other traits such as iris and finger-print. These systems are highly accurate but need user-cooperation. Instead, with the use of camera surveillance system, soft-features[1], such as attire, gait, anthropometric measurements can be utilized to aid the biometric system in identification.

Keeping in view the above mentioned challenges, this thesis makes the following contributions.

1.2 Thesis Contribution

Following are the main contribution of this thesis

- To exploit the advantages of multiple samples for face, profile-face and partial-profile-face streamed to the server, multi sample score fusion scheme has been adopted. This Multi-Sample-Score-Fused face recognition system has been used in conjunction with body measurements extracted.

- The Multi-view collaborative face detection system in [2] has been augmented with Microsoft Kinects [3] to extract anthropometric features. Novel heuristics have been used to extract the anthropometric features of the subject in the scene and these

measurements have used in collaboration with multi-sample score fused face recognition system resulting in enhanced accuracy.

- Performance of the system has been analyzed in *Closed set* and *Open-set* scenarios. Particularly, in Open-set mode of operation when expected size of the probes is very high when compared to the gallery. The contribution of the extracted body measurements on the accuracy is clearly seen.

- Data is collected for 51 subjects with this modified multi-view face and anthropometry acquisition system. Results indicate that verification and identification performance of multi sample score fused face recognition system is better than multi-sample unimodal face(frontal and non-frontal) recognition systems. Appreciable reduction in False Positive Identification Rate(FPIR) can be observed when face based biometric system is combined with anthropometric measurements.

1.3 Thesis Organization

The rest of the thesis is divided into 6 chapters. Chapter 2 covers related work and background information of various techniques that has been used for face recognition and human identification using anthropometric measurements. Chapter 3 describes our collaborative face acquisition system along with the heuristics that have been used for extraction of anthropometric measurements. Chapter 4 deals with various fusion strategies that have been used and various results obtained. Chapter 5 describes the impact of using anthropometric measurements on face recognition and the framework in which the two systems are integrated. Finally conclusion and future work is presented in chapter 6.

Chapter 2

Background and Related Work

2.1 Background

Biometric systems are human identification system that maps the humans personal, behavioral or biological characteristics to the persons identity. Such characteristics are called traits. The traditional systems that use smart cards, magnetic strips, passwords, PIN etc.., such tokens can be lost, stolen or easily forgotten. Further, these systems do not map the person and his identity, but, map the identity to the token. Some of the most commonly used traits are face, iris, fingerprint, gait, key-stroke, DNA etc..

For a real-time human identification system it is important that the system is covert and non-intrusive in acquiring features that can uniquely identify subjects. Face is one such biometric that has been used in covert surveillance systems that have been deployed at public places, state borders, secured checkpoints etc. A generic face recognition system consists of 2 parts Face detection and Face recognition.

- **Face detection**: Face detection system specifies the location of face region in the image and extracts the face from the image. Changes in pose, expression, illumination of the face, and occlusion of facial features due to accessories and other subjects in the scene affect the accuracy of the face detector.

- **Face Recognition**: An ideal face recognition system identifies the subject if the subject is enrolled and rejects the subject if the subject is not enrolled into the system.

  The quality of the face images extracted by the face detector and the robustness of the feature extractor are two main factors that affect the accuracy of the face recognition system.

2.2 Related Work

2.2.1 Face detection

The techniques that are used in face detection are mainly divided into two types Feature based and Image based face detection.

- **Feature Based Face Detection** These techniques deal with segmentation of face using visual features such as pixel properties, grey-scale and color information. The use of low-level features results in high false detections. Some examples are face detection using Edges as described in [4, 5, 6], face detection using gray information are seen in [7, 8] where intensity changes on the face near the eyes are used to detect the face, Skin color analysis[9, 10] has also been used for face detection. Face detection using Active contours is another subset of feature based face detection. Contours are formed based on the interaction between the image edges and the shape model. These segments on the face are utilized in face localization based on further heuristics. There are three types of shape models Snakes[11], Deformable templates [12] and Point distribution models [13]. Each one of them has its own significance.

- **Image Based Face Detection** The algorithms in this category does not search any feature but takes the images as input to localize the face in the image. Techniques

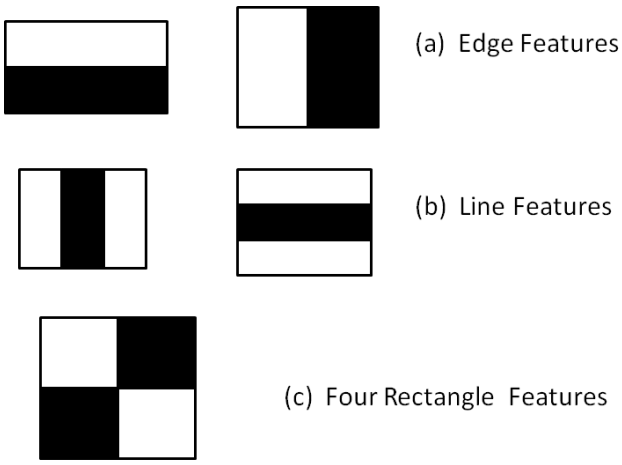


Figure 2.1: Haar-like Features

like PCA have been applied to face detection in [14]and considerable results have been obtained. Machine learning techniques such as Neural Networks [15], SVMs, Hidden Markov Models [16, 17] have also been used on the sub-image to classify the image as face and non-face.

2.2.2 Viola-Jones Face Detector

The recent development in face detection and most widely use face detector is Viola Jones Face detector[18]. This detector uses weak haar-like features and use Boosting, machine learning technique to make the resultant classifier powerful. Viola-Jones face detector divides the input image into small size integral images. The integral image is be used to compute simple Haar-like rectangular features. The features are defined as the (weighted) intensity difference between black and white rectangles as shown in Fig2.1. AdaBoosting[19] is used to combine such weak features and get a classifier with the high resultant accuracy. The overall process of classifying an integral image forms a decision tree, which was called a *cascade*. As shown in Fig2.2 the input sub-windows pass a series of nodes during detection. Each node makes a binary decision whether the window will be kept for the next stage or reject immediately. We use OpenCV Library[20] to detect faces and side faces in the scene.

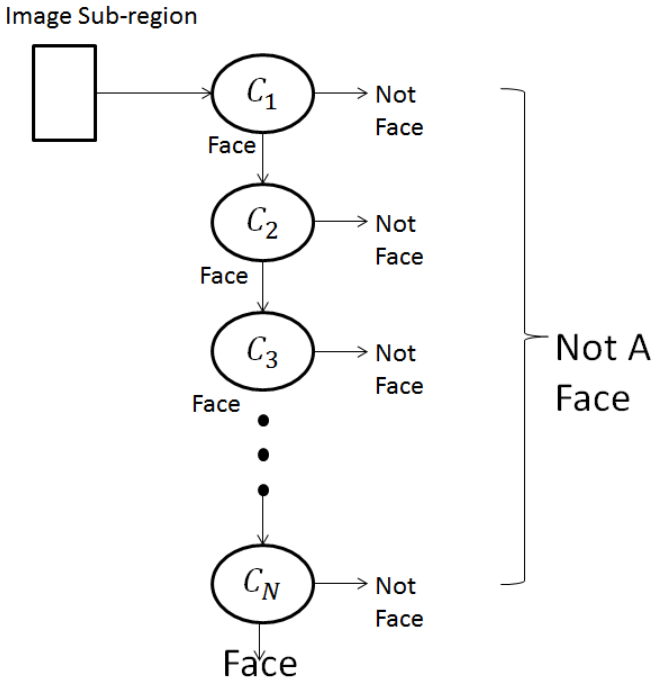


Figure 2.2: Viola Jones cascade for Face detection

2.2.3 Face recognition

Face recognition is done on 2D frontal face images. 2D face recognition methods suffer from pose and illumination changes. 2D-Image based face recognition techniques can be further classified into Appearance based and Model based methods.

- **Appearance Based Methods**

  Also called Holistic methods use the complete face image as the feature vector. Sub-space analysis methods find the most representative basis vectors for the face. Some of the most popular algorithms are PCA [21], LDA[22] and ICA[23]. Linear Binary Pattern LBP[24] is one of the most powerful feature representation algorithms that uses the local histogram as the feature to compare two faces, hence invariant to illumination. The combination of any two of the above mentioned techniques such as LBP-PCA[25], PCA-LDA[22] are also used.

- **Model Based Methods**

  Faces are represented by models. These models are used to fit to the input test faces and the nearest match is determined. EBGM[26], AAM[27],3D Morphable Models[28] are some of the examples of model based face recognition. In EBGM each face is represented as graphs with nodes at the landmark points(corners of eyes, lips, tip of nose etc.) and the edges labelled with the 2D distance vectors. Gabor wavelets, both magnitude and direction, are used to label the nodes. The goal of the matching is to extract a graph from the image the maximizes the graph similarity function. Active Appearance models is a statistical model that combines the shape variation and grey-level appearance information in the feature vector. Matching of AAM is to find the model parameters for the input face and use a distance metric to classify the face.

2.2.4 Anthropometric Measurements

Extracting anthropometric measurements from an image can be said equivalent to finding the distance between two points on the body. This has two issues to be solved. First one being, accurate identification points on the body and second one, accurate conversion of 2D image co-ordinate to 3D world-co-ordinate to get the actual distance between the two points. CAESER[29] database documents 2D as well as 3D scans of 5000 subjects with 73 anthropometric landmarks on the body, these landmarks points are taken as ground truth in defining and developing our algorithm. In this thesis we intend to extract points from the silhouette that is obtained after the foreground segmentation of the body. Foreground segmentation of a image is a computationally expensive process. Simple background subtraction also segments the foreground, but, in clustered environments the accuracy of this background segmentation decreases. To convert the 2D image co-ordinates to 3D world co-ordinate, the camera has to be calibrated. [30] discusses various available methods for camera calibration. Depth using a single camera can also be obtained using vanishing points and vanishing lines as shown in [31], but requires knowledge of the scene beforehand. Using stereo camera for

depth depends on accuracy of locating corresponding points on both the cameras. Stereo camera such as Pointgery BunBlebee[32] can be used but these are very expensive. Stereo cameras made using off the shelf cameras are not as accurate as bumblebee and are difficult to calibrate and maintain, since, small disturbance in their position after calibration, results in high error in the depth. Other depth cameras such as the Time of Flight depth camera are also very expensive and not suitable for applications as ours.

We overcome these problems by using Microsoft Kinect[3].Kinect has the advantage of being inexpensive compared to other depth camera and yet be accurate. Kinect has 3D depth sensor that provides the depth at each location in the image. OpenNi[33] API provides us accesses to the depth and RGB streams of Kinect. This depth segmentation of Kinect is utilized in foreground segmentation of human body and Pin hole[34] camera model is used along with depth from the kinect to convert the 2D image locations to 3D co-ordinates. The distance between these points gives the anthropometric measurements. Various heuristics that are used to find points on the human body is given in detail in chapter 3. The Kinect is calibrated beforehand using OpenCV camera calibration toolbox.

Chapter 3

System Setup and Data Extraction

In this chapter we will discuss various aspects of the system setup, *Collaborative face data acquisition system* and aspects related to extraction of anthropometric data.

3.1 System Setup

Our system for collaborative human identication consists of a network of N_c cameras with overlapping views that are all focused on a critical regions such as entrances to public places and narrow corridors or walkways. For our specific experimental setting and evaluations, we use a network of 3 cameras located along an arc of radius 10 feet. We use the Microsoft Kinect cameras which have a visual spectrum sensor as well as an IR camera that can provide depth information. We use the visual spectrum camera for face image acquisition and the 3d depth sensor only for anthropometric data acquisition. Each camera is locally connected to a a laptop for computation related to face detection and anthropometry. The cameras are deployed on tripods at a height of 7 feet from the ground with their principal axes making an angle of 20° with the horizontal plane and with a separation of 6 feet between the cameras along the arc as shown in Fig.3.1. The angles made by the principal axes of cameras C_2 and C_3 with that of camera C_1 are 40° and 80° respectively. The relative orientations between the cameras (the angles between principal axes of each pair of cameras) are assumed to be

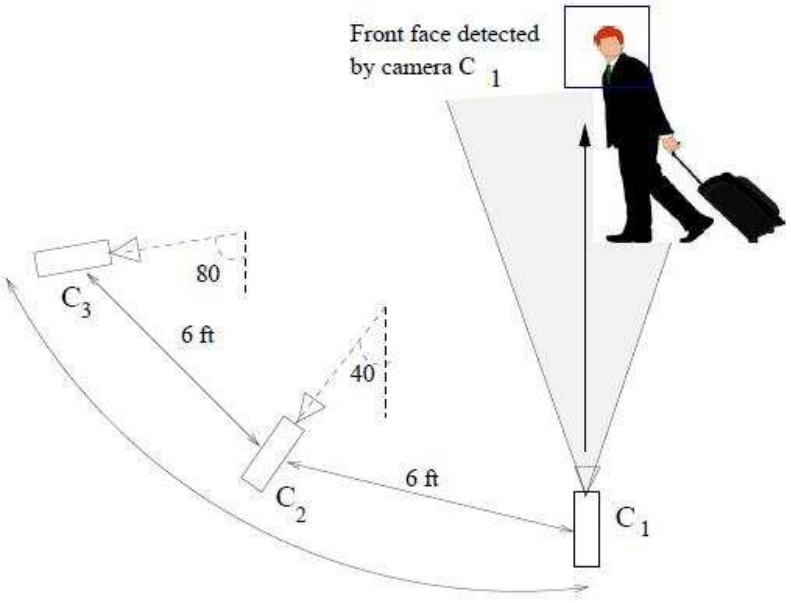


Figure 3.1: Collaborative Camera System

known. We assume that a clock synchronization algorithm is running on the nodes but we note that the clocks of any two nodes may not be in perfect synchrony. Let t_s denote the maximum clock synchronization error between any pair of cameras. This implies that the local clocks of any two cameras can be at most t_s apart.

The cameras are connected wirelessly to a fusion center where the transmitted face images and anthropometric data are collected and may be used for identication. In terms of face acquisition, our goal is to acquire face images corresponding to the following poses: frontal, left (or right) prole, partial left (or partial right) prole. We use the yaw angle (that measures the rotation of a face image along the vertical axis) to dene front, prole and partial prole faces (illustrated in Fig. **??**). We define a face image of a subject acquired by a camera to be frontal if the yaw angle made by the subjects pose ranges from $-22°$ to $22°$, the yaw angle range that OpenCV face detector detects as frontal-face. We define a face image of a subject acquired by a camera to be partial left (partial right) prole if the yaw angle made by the the subjects pose ranges from $-22°$ to $-70°$ ($22°$ to $70°$). We define a face image of a

Figure 3.2: Classification of Frontal and Non-frontal faces

subject acquired by a camera to be left (right) prole if the yaw angle made by the subjects pose ranges from $-70°$ to $-120°(70°$ to $120°)$. We have used the term side face to denote any non-frontal pose of the face. In terms of anthropometric data, our goal is to measure the height and shoulder width of a subject moving through the scene.

3.2 Data Extraction

3.2.1 Collaborative face acquisition

The performance analyses of collaborative face acquisition has been discussed in detail in [2]. We use Microsoft Kincet[3] instead of Logitech Pro 9000 cameras that were used in [2]. The advantages of replacing the Logitech cameras with Microsoft Kinect [3] are:

- Microsoft Kinect has RGB and Depth cameras, internally synchronized and acquiring frames with a resolution of 640 X 480 at 30 frames per second.
- Using OpenNI [33] with Kinects gives the access to Skeleton Points on the subjects with which the search space for the Face detector is reduced, hence, reducing the false

positive detections for the front face.

- OpenNI also segments the subjects from the background. This foreground segmentation is done on the depth image. The silhouette that is obtained after the foreground segmentation as shown in Fig3.3(a) is used to extract the anthropometric information on the subjects present in the scene. It has to be noted here that the camera that captures maximum shoulder width is used. The fusion of measurements from other cameras will be handled in the future.

3.2.2 Anthropometric Data Extraction

A detailed description of how Kinect uses its IR camera for obtaining depth data and hence provide a 3d image of the scene can be found in [35]. Our data acquisition system builds upon the OpenNI [33] framework provided for obtaining skeleton points of human subjects within the eld of view of the camera . The OpenNI software extracts the depth image from a scene and uses that to obtain a background subtracted image of the human silhouette (shown in Fig3.3(a)) and 3d locations of several skeletal joints along the human body. In Fig. 3.3(b), the skeletal points corresponding to left shoulder, right shoulder, head, feet and the center of gravity as extracted by OpenNI are shown. However, the skeletal points provided by the OpenNI framework have primarily been used for obtaining a subjects pose information and classifying gestures [36]. We observe that the skeletal points returned by OpenNI are not precise enough to be used as soft biometric measurements. Therefore, in our framework, we utilize these points only as a basis to more precisely estimate the height and shoulder width.

Degree of convexity: We first obtain a measure of convexity for all points located along the contour of the human silhouette. The contour is assumed to be a closed curve. Let P denote the set of N points along the contour and let p_0, p_1, .., p_{N-1} denote these points sorted along the clockwise direction. Note that p_0 and p_{N-1} are adjacent points. For any point p_i, let p_{i+K} and p_{i-K} denote the K^{th} point and its neighbor in the clock-wise

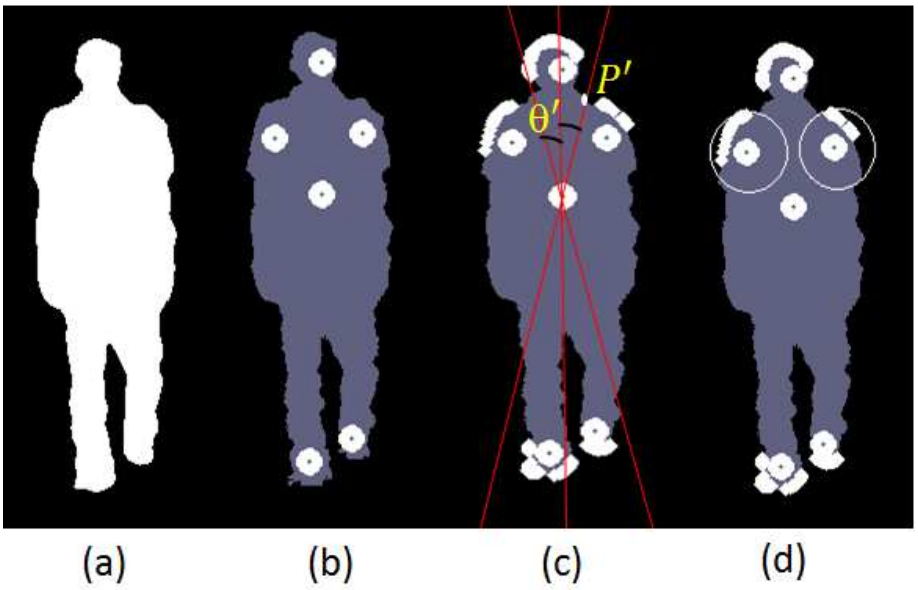


Figure 3.3: Real-time measurements of height and shoulder width: (a)2-d silhouette extracted using the depth image obtained from Kinect (b) Head, Shoulder(Right,Left), Leg(Right Leg) Obtained from OpenNI (c) Determination of head and feet points (d) Determination of shoulder points

and anti-clockwise direction(K points far) respectively. The degree of convexity at a point p_i (denoted as θ_{convp_i}) is given by the angle formed by segment $\mathbf{p_{K-i}p_i}$ with the segment $\mathbf{p_ip_{i+K}}$, measured in a direction that includes all the points inside the silhouette. This is illustrated in Fig. 3.4. K is a constant that provides some separation between the points being considered to measure the degree of convexity. We have chosen $K = 10$.

Shoulder width: Let x_{RS} and x_{LS} be the right and left shoulder points as returned by OpenNI on the human silhouette and let d_s denote the distance between these points. With these points as center, we draw a circle of radius $d_s/2$ and identify all points along the contour of the human body that lie within the two circles. This is shown in Fig. 3.3(d). The points p_L and p_R with the lowest values for θ_{convp_i} in the curves formed by x_{RS} and x_{LS} are chosen as the left and right shoulder points respectively. The distance between p_L and p_R in the 3d coordinate system is used as the shoulder width corresponding to this subject.

Height: The subjects vertical orientation is determined as the direction that is perpen-

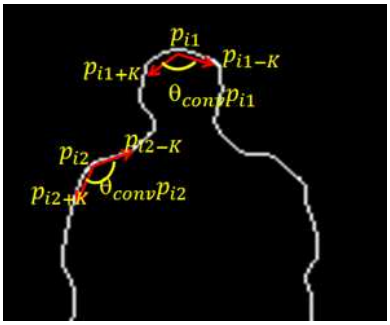


Figure 3.4: Determining degree of convexity for points along contour

dicular the two shoulder points p_L and p_R. Let x_{cg} denote the center of gravity returned by OpenNI. The line parallel to this orientation and passing through x_{cg} is used as the axis of gravity. To determine the height, we obtain a set of points long the contour which form an angle of less than 15 deg with the axis of gravity.This is shown in Fig. 3.3(c). Two sets of points are obtained, near the head and feet respectively. Then, we determine the most convex point along these two sets and use these as the head and feet points respectively. The points are transformed to the 3d coordinates and the distance between them is used as the height of a subject. Note that measurements are made by multiple cameras at multiple instances during the time that a subject remains within the field of view, but only the view with highest shoulder-width is considered as the correct measurement. In the other views the complexity of the contour is very less and hence could result in high error. We plot a histogram of all these measurements with a bin size of $10mm$. The average of the most frequent histogram bin is used as the corresponding measurement for subject height and shoulder width.

Chapter 4

Multi-View Face Recognition

In this chapter we will discuss various aspects of biometric fusion. The various fusion techniques that have been used to utilize the presence of multiple evidences are described in this chapter.

4.1 Biometric Fusion

The system described in the previous chapter sends frontal and non-frontal faces to the server with their respective timestamps. At the server the front face is matched with the corresponding side faces(profile and partial profile faces) using the nearest timestamps. Since, we have multiple evidences(multiple samples and multiple modalities) of the same subject it would be appropriate to take decision collaboratively, with face and side face identification systems. The collaboration in this kind of system can be done in following ways.

- **Multi-Sample system**: In this type of system, the decision is take collaboratively using multiple samples of one biometric acquired during a short period of time.
- **Multi-Modal system**: In this type of system, more than one modality of biometric traits are used to arrive at a decision on the input subjects identity.

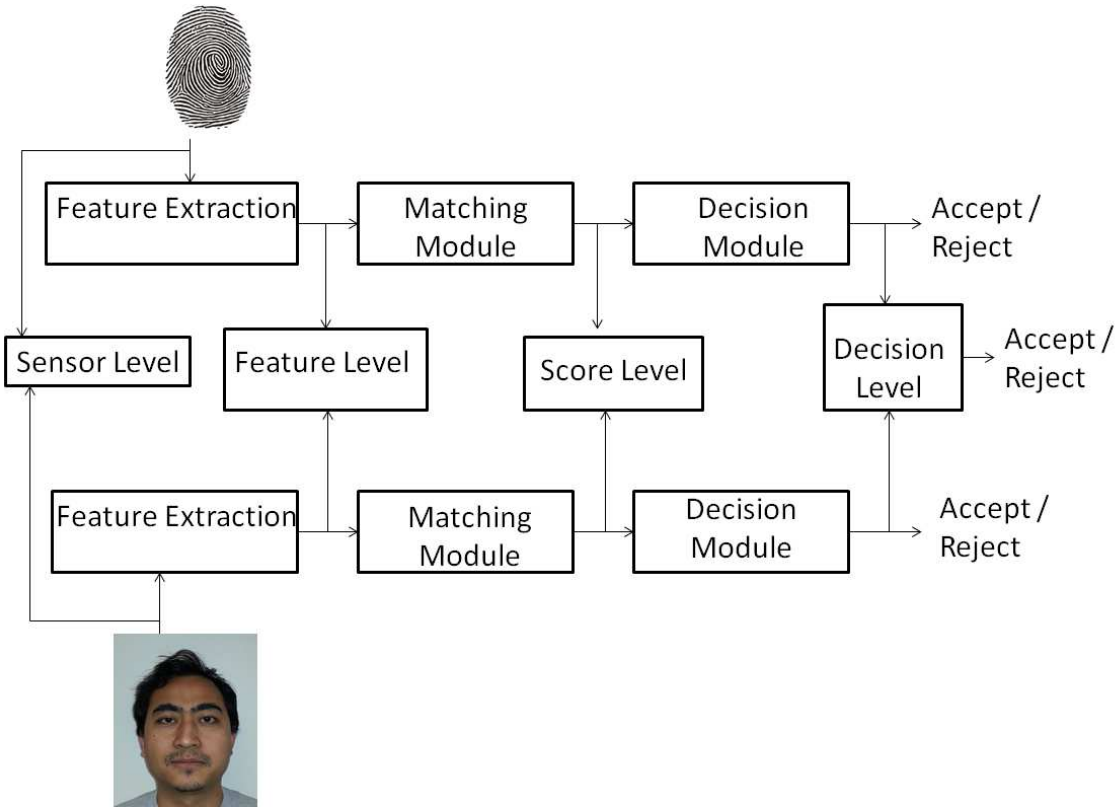


Figure 4.1: Collaborative Side Face Detection With Epipolar Line

Based on the module at which the fusion is done the biometric fusion can be of 5 types.The different levels of fusion are shown in the Fig4.1. Following are the various levels of fusion along with examples of multi-sample fusion at each of these levels:

- Sensor Level Fusion: Fusion where complete face image is constructed using frontal and non-frontal images with techniques such as homography, as in [37]
- Feature Level Fusion:Fusion where the gallery samples of the identified subject are updated as in [38][39]
- Score Level Fusion: Taking best match score when the probe is matched with each gallery image can be said as a example of score based multi-sample fusion as in [40]
- Decision and Rank Level Fusion: With multiple samples a majority voting scheme can

be said to be an example of decision or rank level fusion for example [41]

In this thesis we present the analysis of a *Hybrid fusion* scheme in both verification and identification. In our *Hybrid fusion* scheme we use the score level multi-sample fusion by considering the minimum match score for each gallery subject and use this score for score-level multi modal fusion to take advantage frontal and non-frontal faces obtained form the collaborative face acquisition system as described earlier.

4.2 Fusion Methodology

As mentioned above our system detects front faces, partial profile face and profile faces. For frontal face recognition, we use the PittPatt Face Recognition Software Development Kit from Pittsburgh Pattern Recognition[42]. This SDK provides recognition tools that extract templates from faces and compare templates to compute similarity scores. When using this software, there is no need to explicitly align the frontal images before feeding to the recognition algorithm. The underlying algorithm is also robust to slight variations in the pose and illumination changes. The PittPatt face recognition algorithm extracts the template of the test image and compares it with the templates of the face images in the database.

For non-frontal face recognition, we use the *Local Binary Patterns*(LBP) based classier which considers both shape and texture information to represent the face images [24], [43]. The LBP operator forms labels for the image pixels by thresholding a p pixel neighborhood around each pixel in comparison with the center pixel of the neighborhood, and considering the result as a binary number. This results in a p bit label for each neighborhood, with 2^p possible values. A histogram of these 2^p labels is then used as the image descriptor. Since its introduction, the LBP operator has been extended to use different neighborhood sizes and shapes. In our system, we have specically considered a circular (8, 3) neighborhood, i.e., 8 sample points uniformly separated along a circle of radius 3 around each pixel. Furthermore, we use an extension of LBP, namely uniform LBP, in which only a subset of the 2^p labels are

used in forming the histogram feature. Specically, only patterns in which there are at most 2 bitwise transitions from 0 to 1 or vice-versa are considered as uniform. A separate label is used for each of these uniform patterns, and one label is used for all the other patterns. In an (8, 3) neighborhood this results in 59 uniform patterns. In forming the LBP feature vector, each face image is rst divided into 5 X 5 equisized smaller sub-blocks (or cells). Division of a face image image into smaller cells allows us to retain spatial information in the face image. Local 8 bit binary patterns are extracted and a separate histogram is obtained for each cell. Let R_j denote the j^{th} cell where $1 \leq$ j ≤ 25. Let $f(x, y)$ denote the label of pixel (x, y), where $f(x, y)$ ranges from 1 to 59. Let H_{ij} denote the histogram frequency of the label i (1 $\leq$ i ≤ 59)in region R_j. Let I(A) = 1 if predicate A is true and 0 otherwise. Thus we have:

$$H_{ij} = \sum_{xy} I[f(x, y) == i] * I[(x, y) \in R_j]$$

The histogram frequencies H_{ij} for cell j are normalized as:

$$\dot{H}_{ij} = \frac{H_{ij}}{\sum_{i=1..,59} H_{ij}}$$

The histogram frequencies across all cells are then concatenated into a single histogram that efciently represents the face image. During matching, scores are calculated using the nearest neighbor classication technique with Chi-square (χ^2) statistical estimate for dissimilarity measure. Specic ally, if O_i denotes the observed frequency of label i, M_i denotes the expected frequency of label i, the number of labels are denoted by L, then the χ^2 dissimilarity measure between the observed sequence O and the expected sequence M is given by:

$$\chi^2(S, M) = \sum_i \frac{(O_i - M_i)^2}{O_i + M_i} \ , \ i = 0, 1, 2, 3, 4, 5...L$$

The LBP scores are then normalized so that they are homogeneous and their range lies within $[0, 1]$ using the min-max normalization technique [44].

$$\dot{s} = \frac{s - min(s_k)_{k=1,2..n}}{max(s_k)_{k=1,2..n} - min(s_k)_{k=1,2..n}}$$

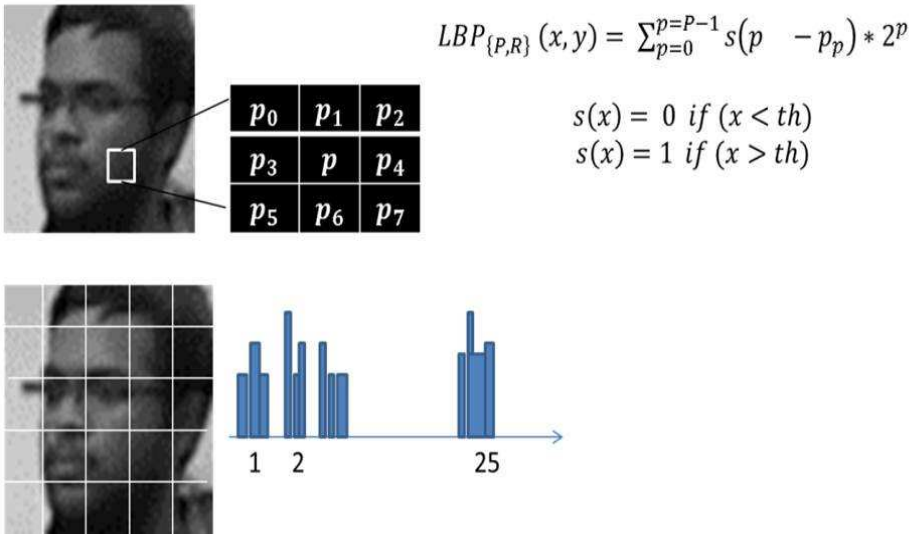


Figure 4.2: Local Binary Pattern

Note however that the PittPatt front face matcher assigns a similarity score. In order to be consistent with the front face scores, the similarity scores $(d_k)_{k=1,..n}$ are converted to dissimilarity scores as

$$(d_k = 1 - s_k)_{k=1,..n}$$

In has to be noted here that frontal, partial-profile and profile face images are compared with their respective frontal,partial-profile and profile gallery set. Inter-modal matching has not been explored in this thesis and will be considered in future.

4.2.1 Multi Sample Score Fusion

Let G denote the gallery set of subjects against which probe images are to be matched. In each trial, a given probe subject moves through the region covered by the camera network and multiple probe images are obtained corresponding to each face pose. We first describe how the multiple probes of a given pose are fused to yield a single score (multi-sample fusion) and then describe how the scores from multiple views are fused. In a given trial, let P_f denote

the set of front face probe images acquired. Let $g \in G$ denote a gallery subject. Let N_g denote the set of gallery images for subject g and let $g_i \in N_g$ denote the i^{th} gallery image for subject g. Let $s_f^{(gi,j)}$ denote the match score for the j^{th} probe image with g_i,where $j \in P_f$. The fused score corresponding among all probe images with respect to gallery image g_i of subject g is obtained by taking the minimum across all probe images. Thus,

$$s_f^{g_i} = \underset{j \in P_f}{MIN}(s_f^{(g_i,j)})$$

Similar multi-sample fused scores are obtained with respect to partial prole face images and prole face images (denoted as s_{pp} and s_{pf} respectively). We then use a weighted score fusion strategy to get an overall score so for each subject in a given trial. Let w_f , w_{pp} and w_{pf} denote the weights for front face score, partial prole score and prole score respectively. Thus, we have:

$$s_o{}^{g_i} = w_f.s_f^{g_i} + w_{pp}.s_{pp}^{g_i} + w_{pf}.s_{pf}^{g_i}$$

In the experimentation phase, the weights are determined by using a set of subjects separately for training. For the trials involving the training subjects, we obtain a matrix of matching scores corresponding to each probe image against every gallery subject. Using these scores, we determine the Equal Error Rate (EER), i.e., the recognition accuracy when the false acceptance rate is equal to the false rejection rate. For different combinations of w_f , w_{pp} and w_{pf}, the EER is determined for the fusion based classier and the combination of weights that gives the least EER is selected. This combination of weights is used on the set of test subjects as evaluated in latter sections.

4.3 Data Collection and Experimental Analysis

We used the camera network setup as described in chapter 2 and performed an experiment using 51 subjects. In each trial, a subject was asked to walk facing one of the cameras starting from a distance of 15 feet from the camera at about 2 feet per second. Multi-view face images

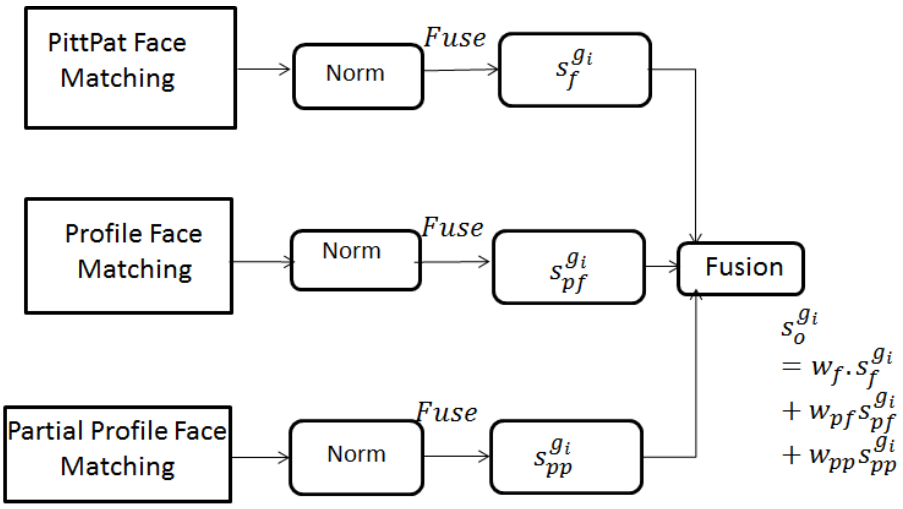


Figure 4.3: Multi-Modal Fusion

were extracted by the cameras using the technique described in chapter 2 and transmitted to a fusion center. At the same time measurements of shoulder width and height were also transmitted to the fusion center. 3 trials were carried out for each of the 51 subjects. Only the biometric data that is extracted by the cameras when each subject is between a distance of 13 feet and 9 feet from the cameras is used for evaluation of identication accuracy. Note that these are typically low resolution images and the size of the images ranged from 20-by-20 pixels to 45-by-45 pixels and a subject is in the scene for about 2 - 3 seconds when this data is acquired. Using the network setup of Microsoft Kinect cameras and 2.4Ghz laptops for local computation, frames were captured at 25 fps. With our multi-view face acquisition software, front face images were detected at approximately 25fps and side face images were acquired at 22 fps. Thus, in each trial an average of 70 front face images and 55 prole and partial prole face images are acquired. A random set of 10 images are acquired from a separate trial for each subject from all 3 views to be used as gallery images for the subjects.

The performance analysis is performed in 2 modes *Closed set* mode and *Open set* mode.

4.3.1 Closed Set Analysis

As described above, the matching scores ($s_f^g, s_{pp}^g, s_{pf}^g$) and the multi-view fused matching score (s_o^g) are obtained for each probe subject against all gallery subjects ($g \in G$) in all the trials. We characterize the closed set identication performance for the system by plotting an ROC curve that shows the Genuine Acceptance Rate (GAR) against the False Acceptance Rate (FAR). Fig. 4.4[a] shows the closed set ROC curve for the system separately for front face images, partial prole face images, prole face images without Multi-sample fusion and ROC with Multi-sample fusion is shown in Fig. 4.4[b].

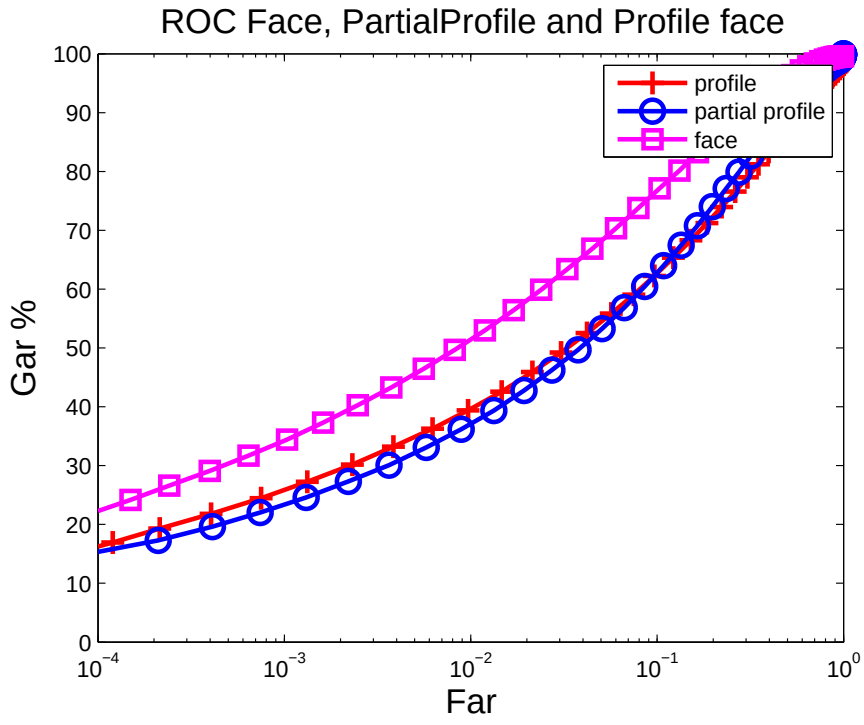
It clearly shows that the performance improves considerably when *Multi Sample Fusion* and weighted sum multi-modal fusion is applied.

Fig 4.5 show the performance of the multi-sample fusion with weighted sum multi-view fusion on low resolution and high resolution images that are recorded as the subjects walks towards the camera in the scene.

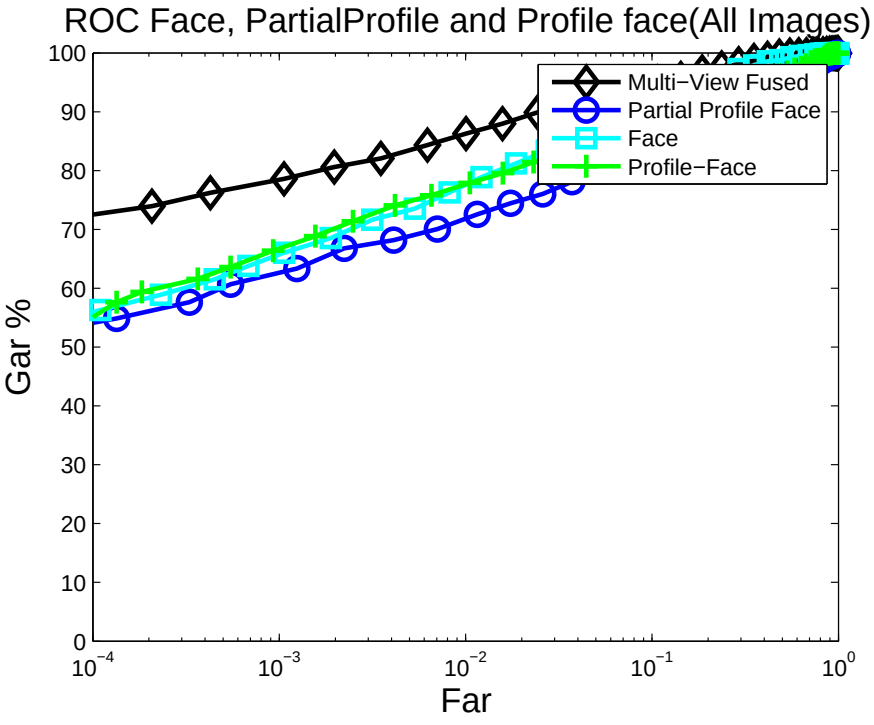
4.3.2 Open Set Analysis

In this section, we evaluate the system performance by using an open-set analysis. We iteratively select one subject to be left out from the list of 41 subjects at a time and compare the matching scores generated for that probe subject with the matching scores generated for all other probe subjects. For different thresholds, we compute the number of times where the subject that is left out is falsely identied and the number of times the genuine subject is identied. We repeat these computations iteratively for each of the 41 subjects being considered as the subject being left out and the average over all these measurements are used to generate a plot of the True Positive Identication Rate (TPIR) against the False Positive Identication rate (FPIR). This graph is shown in Fig. 4.6 separately for front face images, partial prole face images, prole face images and when the scores are fused using the multi-sample fusion and weighted score fusion strategy described in earlier.

Fig 4.7 show the performance of the multi-sample fusion with weighted sum multi-view

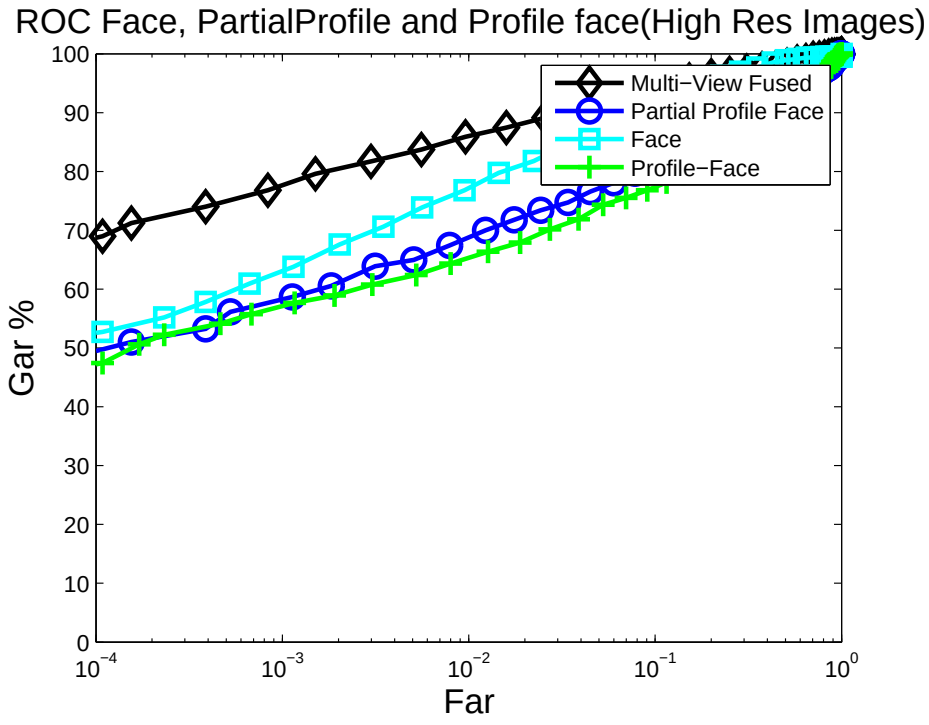


(a)

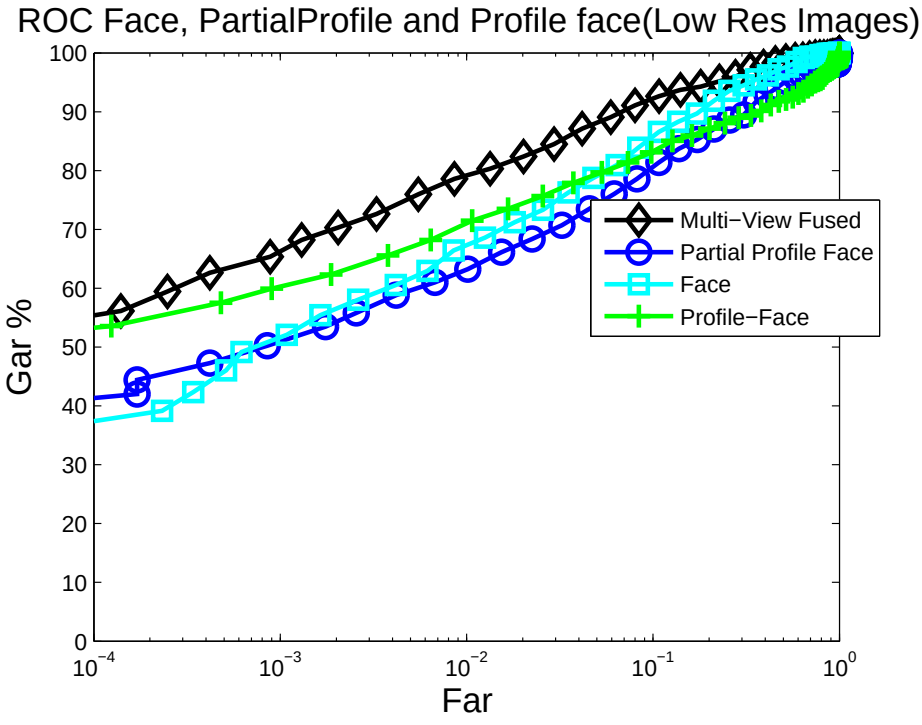


(b)

Figure 4.4: Closed Set ROC (a)without Multi-Sample Fusion (b)With Multi-Sample Fusion and Multi-Modal(view) Fusion



(a)



(b)

Figure 4.5: Closed Set ROC with Multi-Sample Weighted Score Fusion with (a)High Resolution Images (b)Low Resolution Images

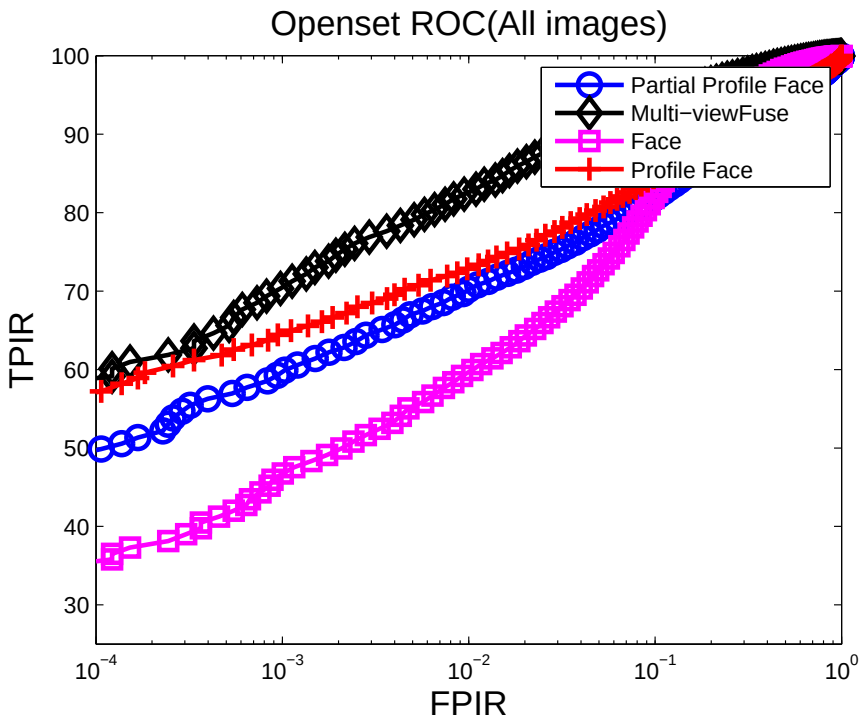
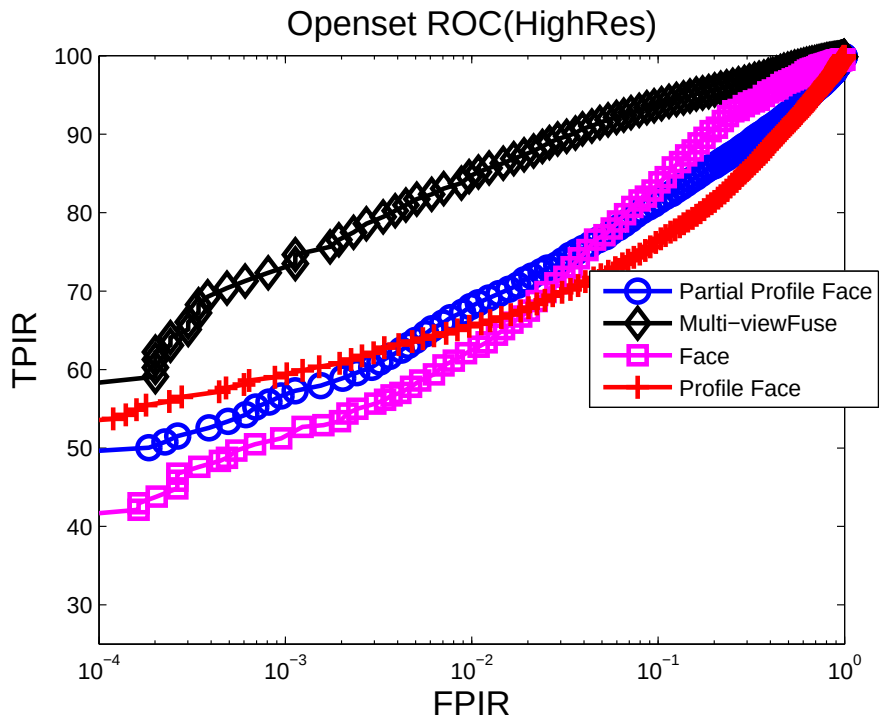
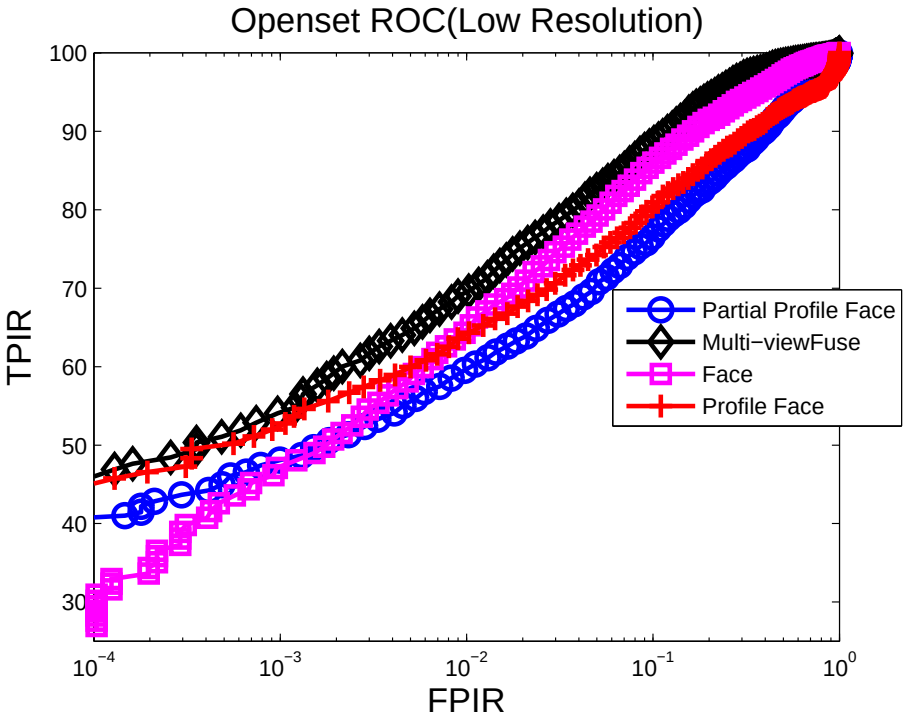


Figure 4.6: Open Set ROC(Muiti-Sample and Weighted Sum Fused)

fusion on low resolution and high resolution images that are recorded as the subjects walks towards the camera in the scene.



(a)



(b)

Figure 4.7: Open Set ROC(Muiti-Sample and Weighted Sum Fused) (a)High Resolution Images (b)Low Resolution Images

Chapter 5

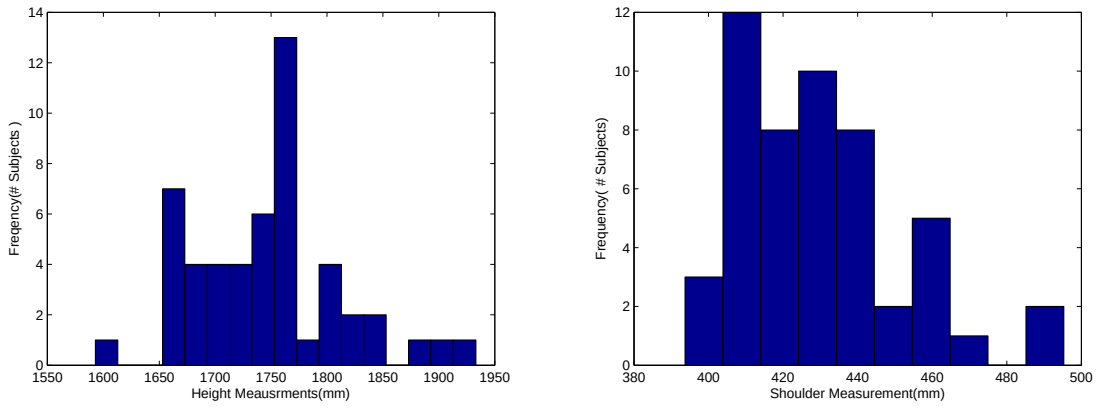
Impact of Anthropometric Measurements on Face Recognition

In this chapter we discuss how the anthropometric information extracted as discussed in chapter 3 are used as a secondary biometric recognition system to enhance the performance of the face based biometric system described in the previous chapter.

Fig 5.1 show the histogram distribution of the true measurements of the 51 subjects that is collected beforehand and stored in the data base. we see that the height and shoulder width of the subjects are concentrated in the range of 1650 1725 mm and 410 430 mm respectively. Also the mean error in the measurement of height and shoulder width for the 51 subjects is observed to be 19.5 mm and 8 mm respectively.

The Fig5.2 shows the CMC curve when Height and Shoulder width are used in identification. In can be seen that the performance when the 2 measurements are fused improves to 80.9% accuracy at rank 10. With a bigger data set and a data set with wider range of measurements, we think the accuracy will be higher than 80.9% hence improving the overall system accuracy.

The Fig 5.3 is a schematic of the system framework that incorporates the anthropometric biometric data for face recognition. The depth data extracted is used to estimate the height and the shoulder width of the subject in the scene as described in chapter 3. Let H_i and



(a) (b)

Figure 5.1: Histogram of Measurements of 51 subjects (a)Height(b)Shoulder Width

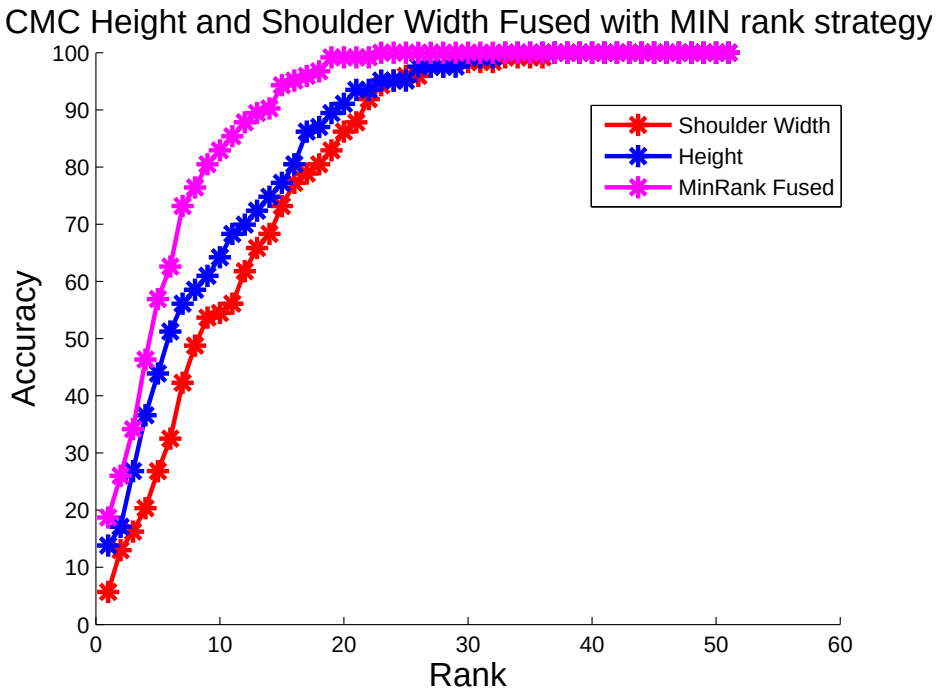


Figure 5.2: Framework for Integration of Anthropometry and Face Recognition System

SH_i be the height and shoulder width of the subject i in the scene. These measurements are compared to the true measurements of all the N subjects that are enrolled into the system. The set $I = \{i_1, i_2, i_3, .., i_k\}$ gives the set of k subjects that are short-listed based on *Min-Error* that is generated when extracted measurements are compared with the measurements in the data-base. I_H and I_{SH} is the set of subjects short-listed with Height and Shoulder. I_{anthro} is the set of probable subjects is given as

$$I_{anthro} = I_H \bigcap I_{SH}$$

The size of the subset I_{anthro} is denoted and K. Hence instead of scoring the input face against all the 41 subjects as done in the previous chapter, we compare the input test subject with $K(\ll 41)$ to identify the subject in the scene. Hence, we expect that the performance of the face recognition system will improve or atleast be equal as the data base size K decreases. Apart form accuracy, when the data base size increases the time taken to establish the identity will be much less when soft-biometric features such as body measurements are used.

This analysis is preformed in both *Close Set* and *Open Set* scenarios.

5.0.3 Close Set Analysis With Anthropometry

From the following plots, it can be seen that with subset size $K = 11$ the performance of the Anthropometry + Face Recognition system(Multi-Sample and Multi-View Fused) is almost as good as the biometric system that uses the Face Recognition system alone(all 41 subjects in database). The following plots int Fig 5.4 show the performance with $K = 11$, $K = 14$ and $K = 16$. Fig 5.5 show the performance with high and low resolution images.

As described above the performance of the face recognition system depends on the size subset I_{anthro}, K. The following plots Fig 5.6 ad Fig 5.7show the variation of the GAR(%) with change in K and two thresholds FAR = 0.01 and FAR = 0.1.

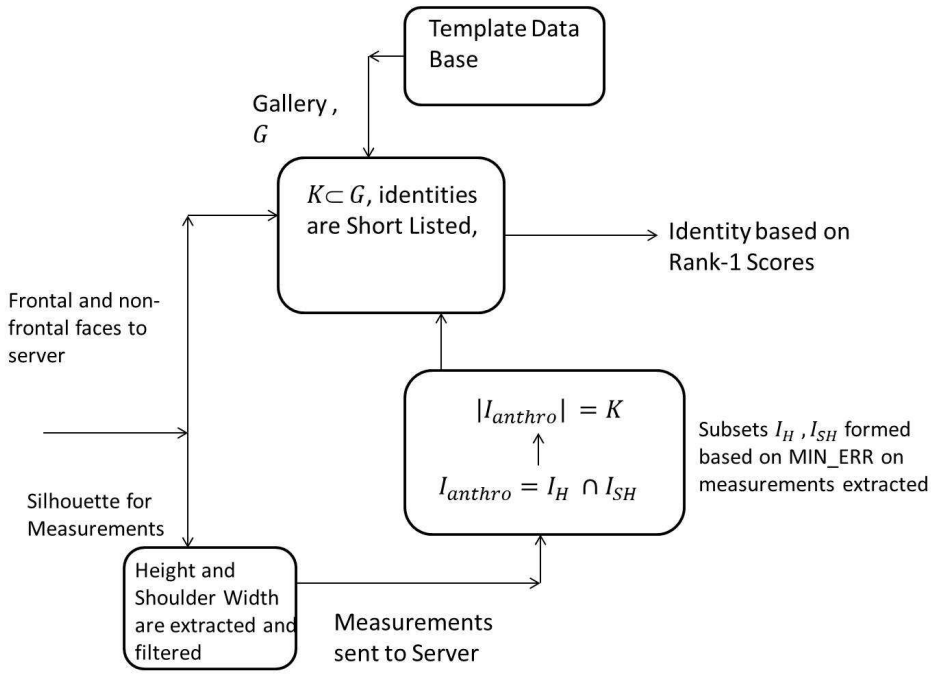


Figure 5.3: Framework for Integration of Anthropometry and Face Recognition System

5.0.4 Open Set Analysis With Anthropometry

The same can observed with the Open set recognition also. The performance of the face recognition system is sustained even after reducing the search space by K subjects. This results in large reduction in the time taken in identifying the subject in the scene when size of the database increases. The following plots Fig 5.8 show the Openset ROC with reduced data set size at $K = 11$,$K = 14$ and $K = 16$. Fig 5.9, shows the performance after the data base size reduction with high and low resolutions.

As described above the performance of the face recognition system depends on the size subset I_{anthro}, K. The following plots Fig 5.10 ad Fig 5.11show the variation of the GAR(%) with change in K and two thresholds FAR = 0.01 and FAR = 0.1.

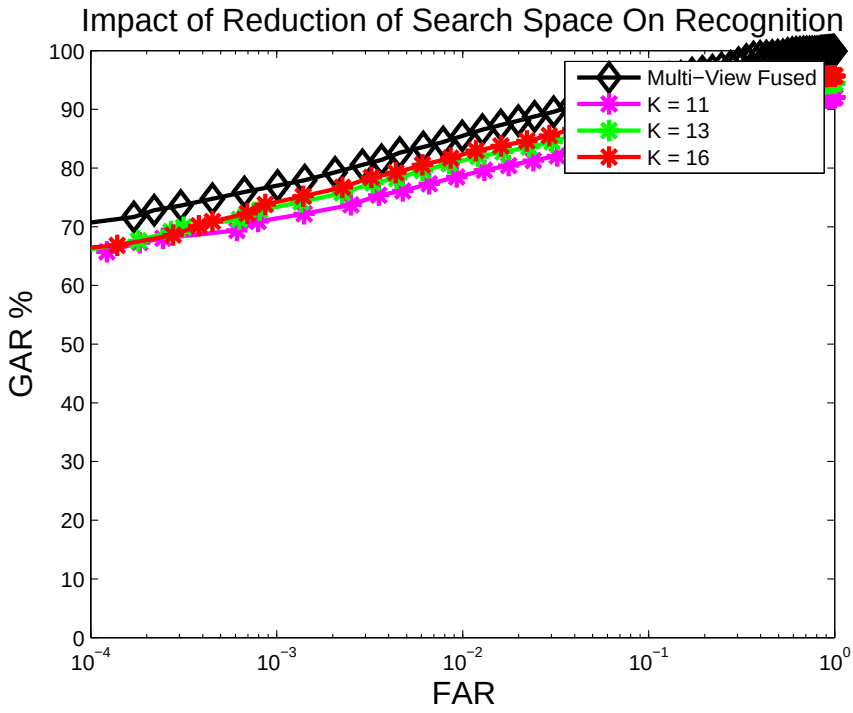
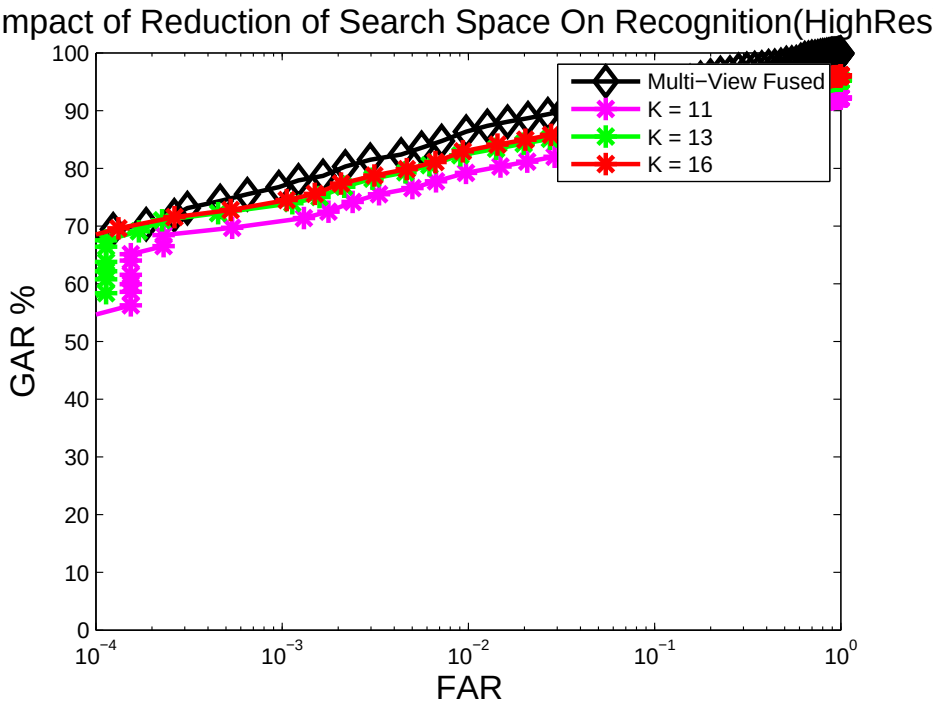
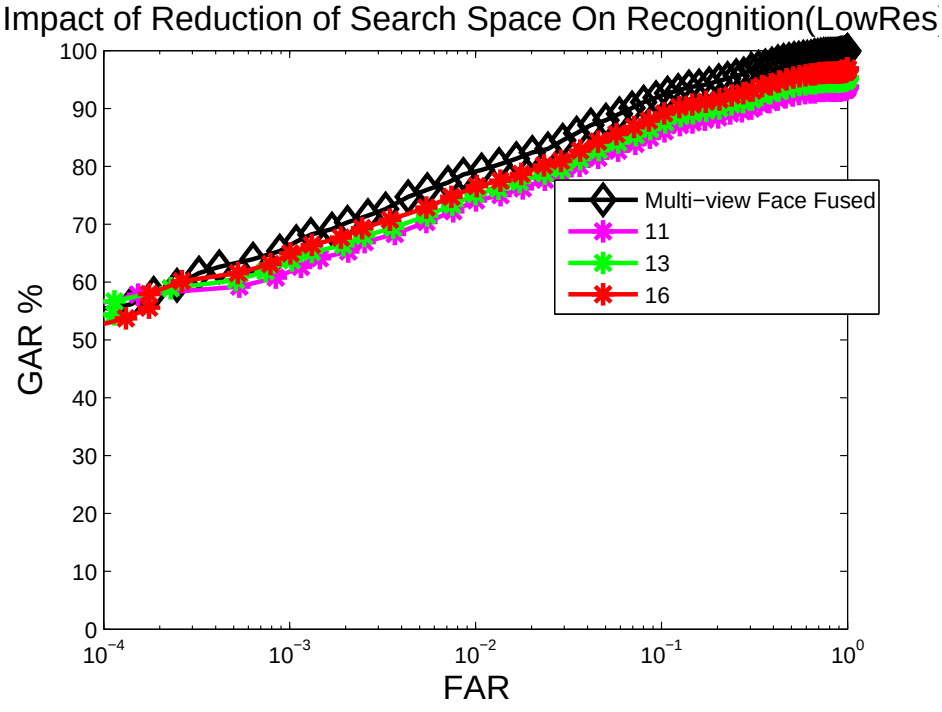


Figure 5.4: Closed Set ROC(Muiti-Sample and Weighted Sum Fused)With Anthropometry and with all images



(a)



(b)

Figure 5.5: Closed Set ROC(Muiti-Sample and Weighted Sum Fused)with Anthropometric Measurements (a)High Resolution Images (b)Low Resolution Images

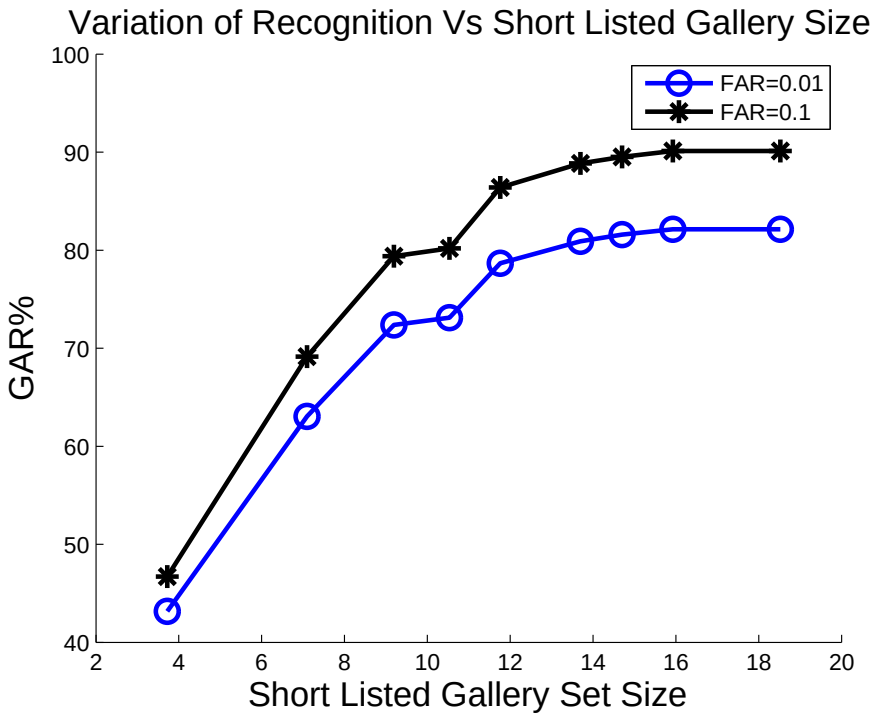
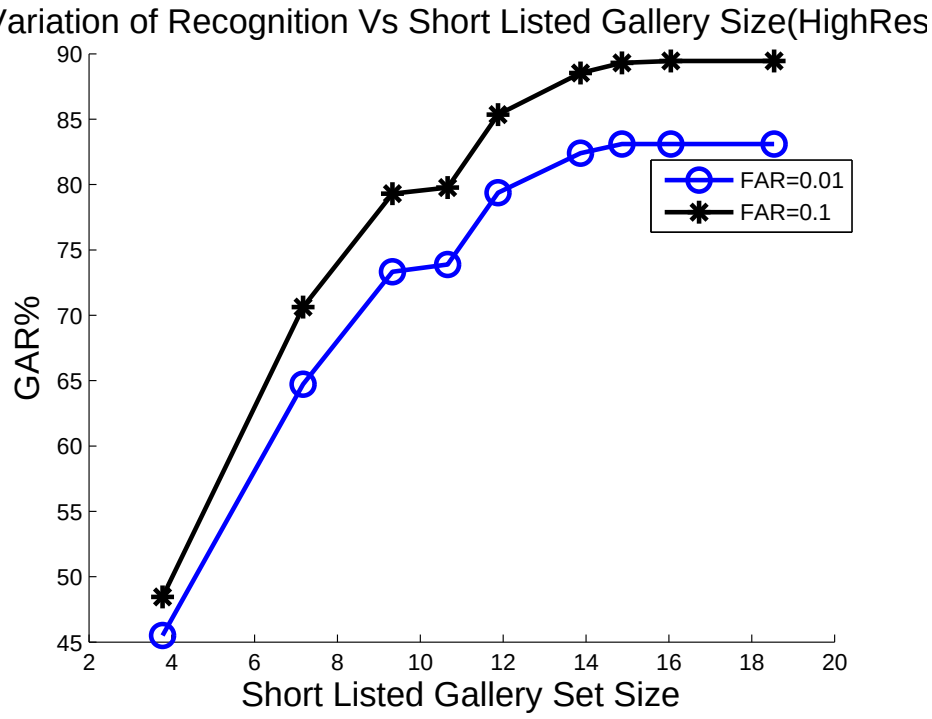
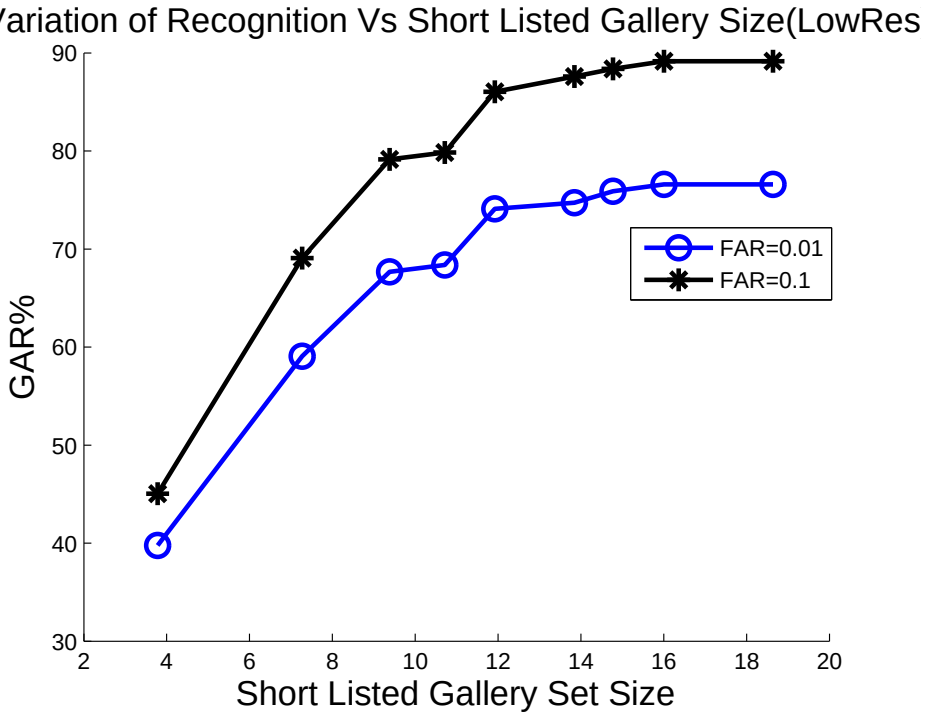


Figure 5.6: Impact of Variation of Reduced Subset Size on Closed Recognition



(a)



(b)

Figure 5.7: Impact of Variation of Reduced Subset Size on Closed Recognition with (a)High Resolution Images (b)Low Resolution Images

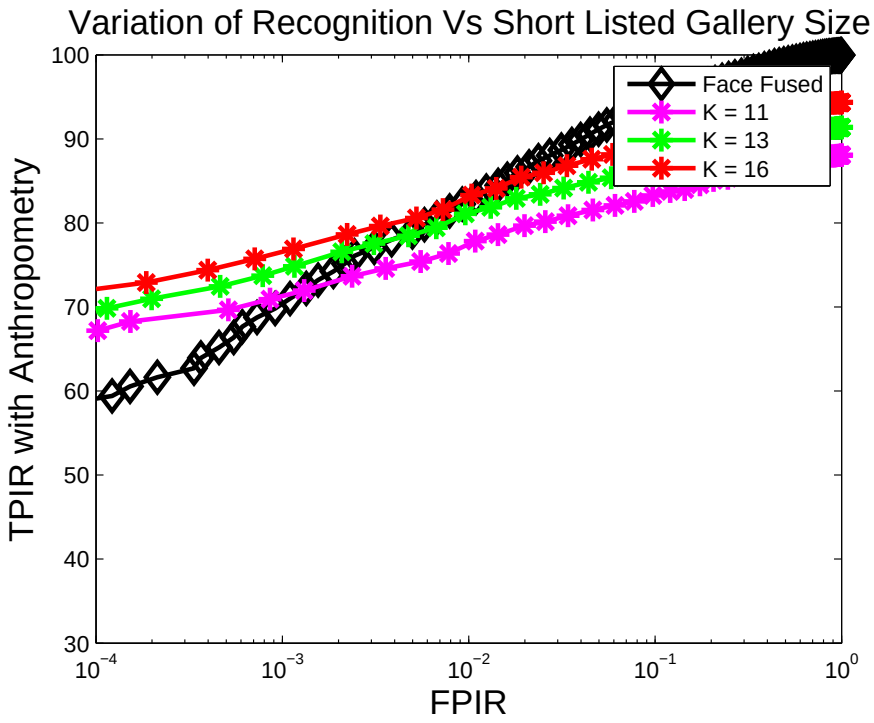
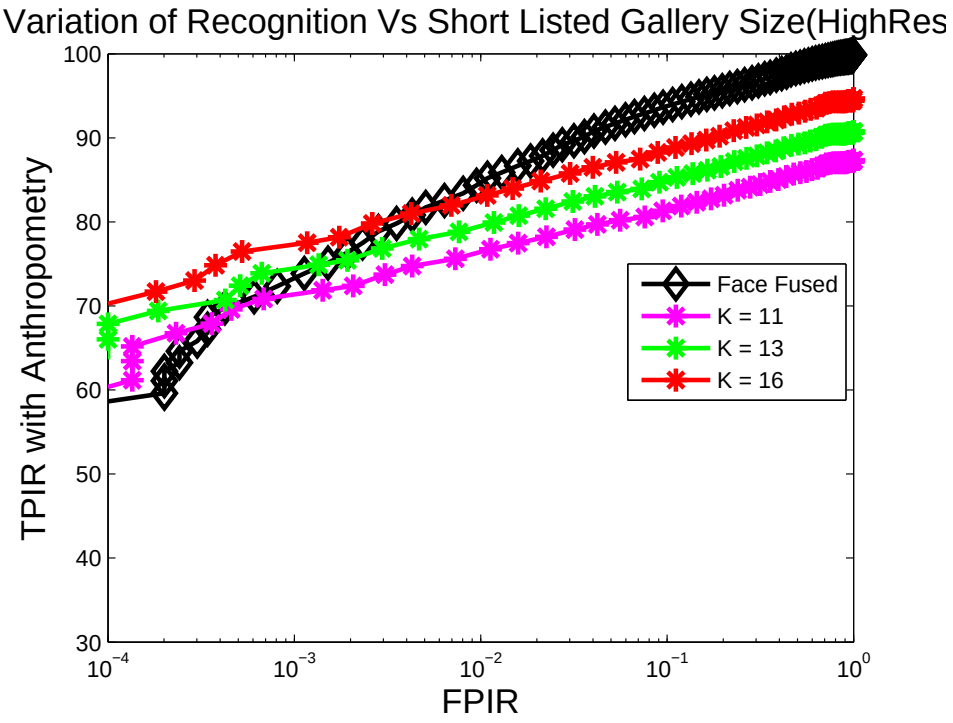
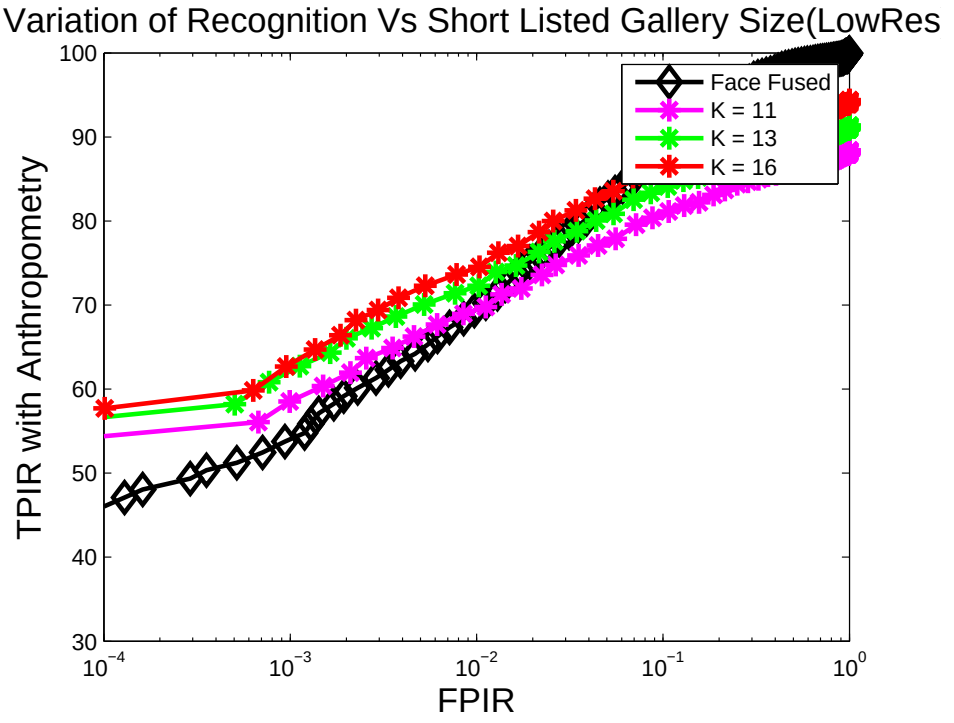


Figure 5.8: Open Set ROC(Muiti-Sample and Weighted Sum Fused)With Anthropometry with all images



(a)



(b)

Figure 5.9: Open Set ROC(Muiti-Sample and Weighted Sum Fused) with Anthropometry (a)High Resolution Images (b)Low Resolution Images

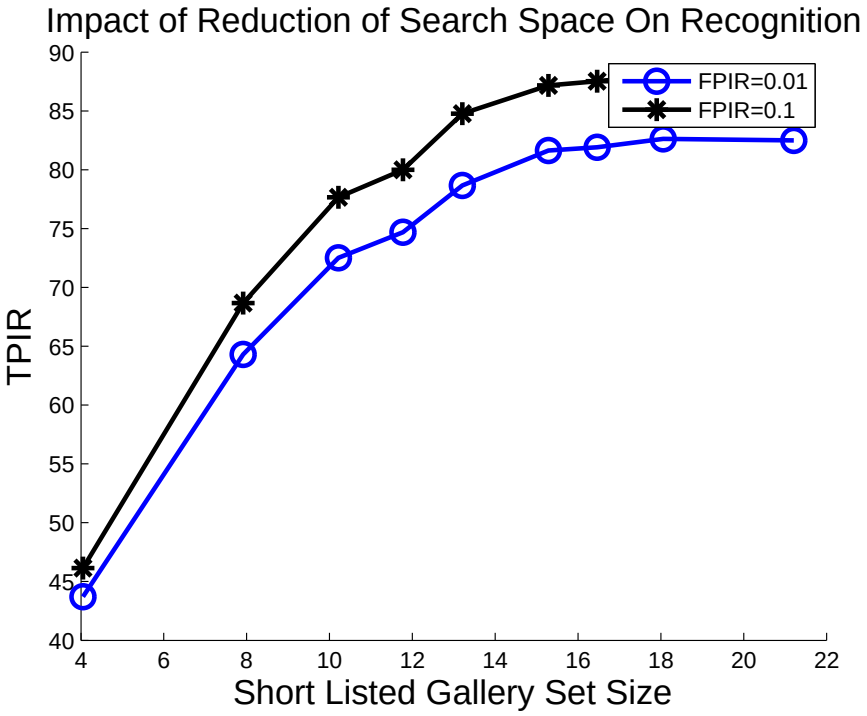
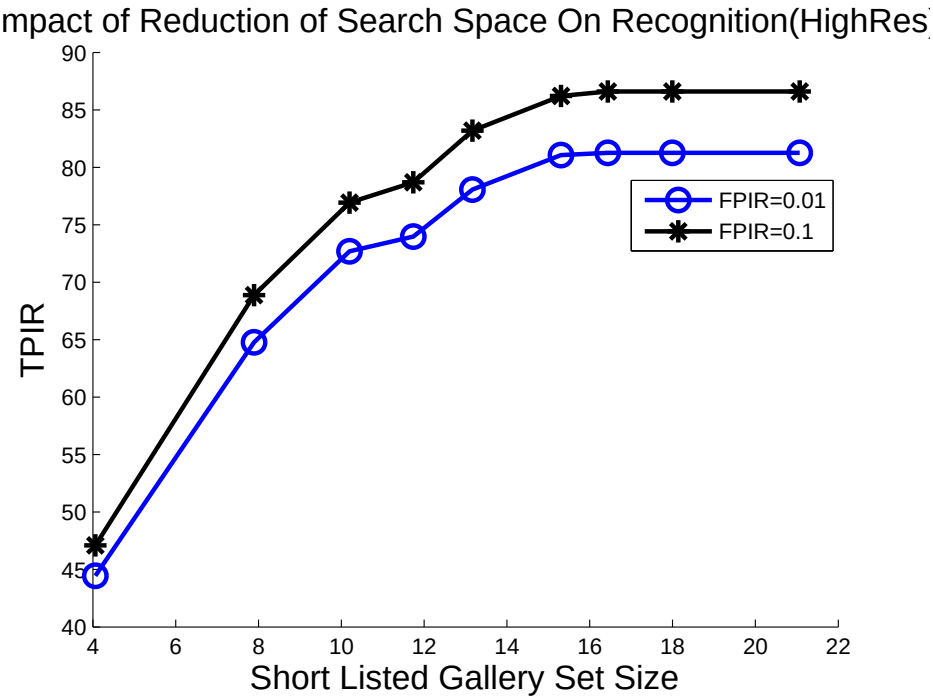
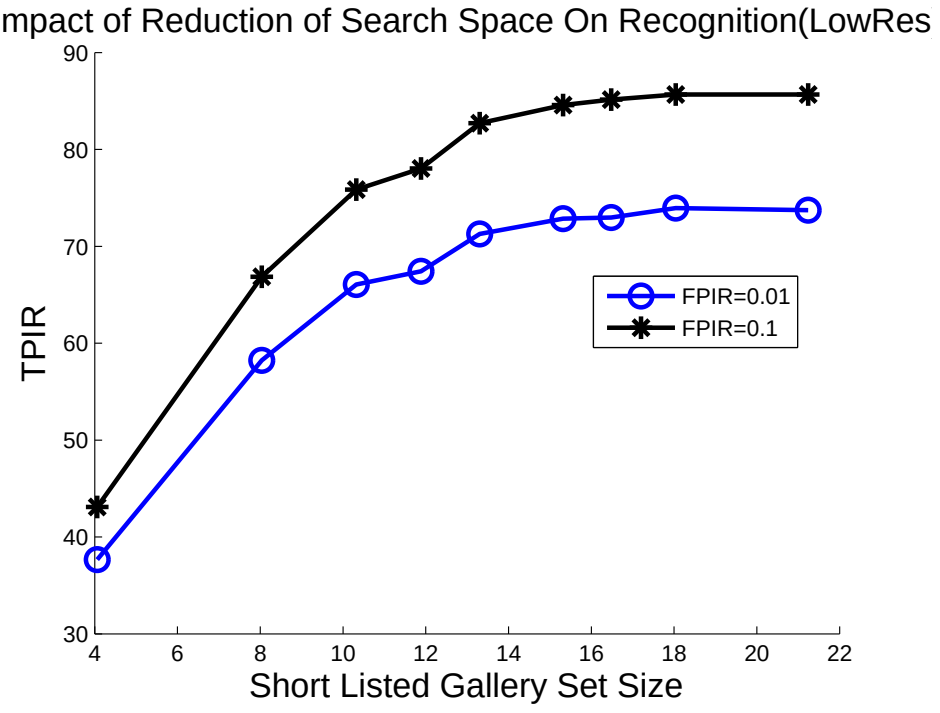


Figure 5.10: Impact of Variation of Reduced Subset Size on Open Recognition



(a)



(b)

Figure 5.11: Impact of Variation of Reduced Subset Size on Open Recognition with (a)High Resolution Images (b)Low Resolution Images

Chapter 6

Conclusions

6.1 Conclusions

This thesis presents a *Multi-Sample-Multi-Modal* face recognition system constructed using modified collaborative face acquisition system[2] using Microsoft Kinects[3]. This system collaboratively acquires frontal and non-frontal faces automatically labels them and sends the faces to the server. The availability of multiple evidences at the server has been fused in two ways. The performance of the two methods of fusion namely, Multi-modal fusion and Multi-sample fusion have been analyzed. Weighted sum score fusion technique has been adopted for Multi-modal fusion, fusing frontal face match scores obtained form PitPatt[42] and side face match scores obtained form Local Binary Pattern (LBP) matcher. Multi-sample fusion has been performed by taking minimum match score generated by probe images in a particular trial. When both of these fusion schemes are combined, we get, Multi-Sample-Weighted Sum Score fusion, the performance of this fusion methodology is evaluated on both closed and open-set scenarios. The identity of a probe subject is established as the subject walks into the system. The identity is established at various instances during each trail and FPIR is used to evaluate the identification performance in openset scenario. This face acquisition system is augmented by anthropometric measurements that are extracted from the silhouette. The effectiveness of using these measurements in conjunction with face recognition

system is also analyzed in closed and openset cases.

The Results indicated that the Multi-sample fused uni-modal system performs better than a unimodal system that takes into consideration all the faces(frontal and non-frontal) acquired . Multi-Sample-Weighted sum score fusion performs better than unimodal Multi-sample fused face-identification system. Identification performance in openset has been evaluated with one subject left out strategy. The anthropometric measurements have been integrated with face recognition system ROC curves indicate that the same recognition performance can be obtained with the search space reduced to 16 subjects rather that 41. The impact of reduction of search space with performance at different FAR(0.10 and 0.010) have been analysed. All the above mentioned analysis has been done with low and high resolution images collected in real time.

6.2 Future Work

In this thesis only one strategy for Multi-sample fusion has been explored, in future various other fusion strategies such as variance and standard deviation in the match scores obtained during a trial have to be explored to fuse multiple samples. Quality metrics such as motion blur, which is one of the most important cause for lower accuracy in real-time identification has to be used in fusion to make the system more robust to real-time identification errors. Anthropometric measurements have shown considerable improvement in human identification when used with face identification system.

Carefully revisiting the heuristics and formulating new and more effective heuristics can improve verification and identification performance to a great extant.

Apart from above mentioned extensions to the current work, the fusion of measurements obtained at different views has not been considered in this thesis. The advantages of having multiple measurements form multiple views could be used to collaboratively to extract more anthropometric measurements rather than just the height and shoulder-width.

Multi-modal matching of face images for example, matching frontal face with profile-face

images has not been explored in this thesis. In can be understood that, this needs detail study of the feature vectors that will be used to describe frontal and non-frontal face images such that they could be matched. Hence, considered over the scope of our real-time face recognition system. It can also be seen that this kind of system(collaborative face acquisition) is independent of the algorithm used for face(frontal, non-frontal) matching. Hence, can be used with any face(frontal, non-frontal) matching algorithms can be used at the server.

References

[1] Anil K. Jain, Sarat C. Dass, Karthik Nandakumar, and Karthik N, "Soft biometric traits for personal recognition systems," in *Proceedings of International Conference on Biometric Authentication, Hong Kong*, 2004, pp. 731–738.

[2] S. Parupati, R. Bakkannagari, S. Sankar, and V. Kulathumani, "Collaborative acquisition of multi-view face images in real-time using a wireless camera network," in *Distributed Smart Cameras (ICDSC), 2011 Fifth ACM/IEEE International Conference on*, aug. 2011, pp. 1 –6.

[3] Microsoft, *Microsoft Kinect*, November 2010.

[4] T. Sakai, M. Nagao, and Takeo Kanade, "Computer analysis and classification of photographs of human faces," in *Proc. First USA-JAPAN Computer Conference*, 1972, pp. 55–62.

[5] I. Craw, H. Ellis, and J.R. Lishman, "Automatic extraction of face-features," *Pattern Recognition Letters*, vol. 5, no. 2, pp. 183 – 187, 1987.

[6] J. Choi, S. Kim, and P. Rhee, "Facial components segmentation for extracting facial features," in *AVBPA99*, 1999.

[7] R. Hoogenboom and M. Lew, "Face detection using local maxima," in *Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on*, oct 1996, pp. 334 –339.

[8] C. Wong, D. Kortenkamp, and M. Speich, "A mobile robot that recognizes people," in *Tools with Artificial Intelligence, 1995. Proceedings., Seventh International Conference on*, nov 1995, pp. 346 –353.

[9] S. Mckenna, S. Gong, and J. J. Collins, "Face tracking and pose representation," in *In BMVC*, 1996, pp. 755–764.

[10] M. Hunke and A. Waibel, "Face locating and tracking for human-computer interaction," in *Signals, Systems and Computers, 1994. 1994 Conference Record of the Twenty-Eighth Asilomar Conference on*, oct-2 nov 1994, vol. 2, pp. 1277 –1281 vol.2.

[11] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *INTERNATIONAL JOURNAL OF COMPUTER VISION*, vol. 1, no. 4, pp. 321–331, 1988.

[12] Alan L. Yuille, "Deformable templates for face recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 59–70, Jan. 1991.

[13] P. Nair and A. Cavallaro, "3-d face detection, landmark localization, and registration using a point distribution model," *Multimedia, IEEE Transactions on*, vol. 11, no. 4, pp. 611–623, 2009.

[14] H.M. El-Bakry and M. Hamada, "Fast principal component analysis for face detection using cross-correlation and image decomposition," in *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, june 2009, pp. 751 –756.

[15] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705 –741, may 1995.

[16] S.M.Marchand and B.Merialdo, "Pseudo two-dimensional hidden markov models for face detection in colour images," in *IN PROCEEDINGS OF THE AUDIO- AND VIDEO-BASED BIOMETRIC PERSON AUTHENTICATION (AVBPA'99*, 1999, pp. 13–18.

[17] Lingmin Meng, T.Q. Nguyen, and D.A. Castanon, "An image-based bayesian framework for face detection," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, 2000, vol. 1, pp. 302 –307 vol.1.

[18] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, vol. 1, pp. I–511 – I–518 vol.1.

[19] Yoav Freund and Robert E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Proceedings of the Second European Conference on Computational Learning Theory*, London, UK, UK, 1995, EuroCOLT '95, pp. 23–37, Springer-Verlag.

[20] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[21] R.Psarrou, S.Gong, and M.Walter, "Eigenfaces for recognition," in *In J of Cognitive Neuroscience v3 nr1*, 1991, pp. 71–86.

[22] W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant analysis of principal components for face recognition," in *Proceedings of the 3rd. International Conference on Face & Gesture Recognition*, Washington, DC, USA, 1998, FG '98, pp. 336–, IEEE Computer Society.

[23] Chengjun Liu and Harry Wechsler, "Comparative assessment of independent component analysis (ica) for face recognition," in *International Conference on Audio and Video Based Biometric Person Authentication*, 1999, pp. 22–24.

[24] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037 –2041, dec. 2006.

[25] M.J. Nordin and A.A.K.A. Hamid, "Combining local binary pattern and principal component analysis on t-zone face area for face recognition," in *Pattern Analysis and Intelligent Robotics (ICPAIR), 2011 International Conference on*, june 2011, vol. 1, pp. 25 –30.

[26] Kin-Wang Cheung, Jiansheng Chen, and Yiu-Sang Moon, "Pose-tolerant non-frontal face recognition using ebgm," in *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*, 29 2008-oct. 1 2008, pp. 1 –6.

[27] T. F. Cootes and C. J. Taylor, "Cj.taylor, "active shape models - "smart snakes," in *in Proceedings of the British Machine Vision Conference*, 1992.

[28] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 9, pp. 1063 – 1074, sept. 2003.

[29] *Civilian American and European Surface Anthropometry Resource ProjectCAESAR.*

[30] E.E. Hemayed, "A survey of camera self-calibration," in *Advanced Video and Signal Based Surveillance, 2003. Proceedings. IEEE Conference on*, july 2003, pp. 351 – 357.

[31] P. Parodi and G. Piccioli, "3d shape reconstruction by using vanishing points," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 2, pp. 211 –217, feb 1996.

[32] "Bumblebee," .

[33] OpenNI organization, *OpenNI User Guide*, November 2010, Last viewed 19-01-2011 11:32.

[34] Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, New York, NY, USA, 2 edition, 2003.

[35] K. Khoshelham and S.O. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications.," *Sensors (Basel)*, vol. 12, no. 2, pp. 1437–54, 2012.

[36] L. Gallo, A.P. Placitelli, and M. Ciampi, "Controller-free exploration of medical image data: Experiencing the kinect," in *Computer-Based Medical Systems (CBMS), 2011 24th International Symposium on*, 2011, pp. 1–6.

[37] O. Boumbarov, S. Panev, I. Paliy, P. Petrov, and L. Dimitrov, "Homography-based face orientation determination from a fixed monocular camera," in *Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), 2011 IEEE 6th International Conference on*, 2011, vol. 1, pp. 399–403.

[38] Li Qiang, Zhengding Qiu, Dongmei Sun, and Yanqiang Zhang, "Subspace framework for feature-level fusion with its application to handmetric verification," in *Signal Processing, 2006 8th International Conference on*, 2006, vol. 4, pp. –.

[39] Arun A. Ross and Rohin Govindarajan, "Feature level fusion of hand and face biometrics," pp. 196–204, 2005.

[40] K. Nandakumar, Yi Chen, S.C. Dass, and A.K. Jain, "Likelihood ratio-based biometric score fusion," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 2, pp. 342–347, 2008.

[41] B. Gokberk and L. Akarun, "Comparative analysis of decision-level fusion algorithms for 3d face recognition," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, 2006, vol. 3, pp. 1018–1021.

[42] *Pittsburgh Pattern Recognition*.

[43] G. Heusch, Y. Rodriguez, and S. Marcel, "Local binary patterns as an image preprocessing for face authentication," in *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, 2006, pp. 6 pp.–14.

[44] Anil Jain, Karthik Nandakumar, and Arun Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270 – 2285, 2005.