

# Pose and Expression Invariant Alignment based Multi-View 3D Face Recognition

**Naeem Ratyal<sup>1</sup>, Imtiaz Taj<sup>1</sup>, Usama Bajwa<sup>2</sup>, Muhammad Sajid<sup>1</sup>**

<sup>1</sup> Vision and Pattern Recognition Systems Research Group, Capital University of Science and Technology (CUST)  
Islamabad, Pakistan

[e-mail: naemratyal@hotmail.com; imtiaztaj@cust.edu.pk; sajid\_faiz@hotmail.com]

<sup>2</sup> Department of Computer Science, COMSATS Institute of Information Technology, Lahore, Pakistan  
[e-mail: usama@usamajaz.com]

\*Corresponding author: Muhammad Sajid

*Received January 28, 2018; revised April 4, 2018; accepted April 28, 2018;  
published October 31, 2018*

---

## **Abstract**

In this study, a fully automatic pose and expression invariant 3D face alignment algorithm is proposed to handle frontal and profile face images which is based on a two pass coarse to fine alignment strategy. The first pass of the algorithm coarsely aligns the face images to an intrinsic coordinate system (ICS) through a single 3D rotation and the second pass aligns them at fine level using a minimum nose tip-scanner distance (MNSD) approach. For facial recognition, multi-view faces are synthesized to exploit real 3D information and test the efficacy of the proposed system. Due to optimal separating hyper plane (OSH), Support Vector Machine (SVM) is employed in multi-view face verification (FV) task. In addition, a multi stage unified classifier based face identification (FI) algorithm is employed which combines results from seven base classifiers, two parallel face recognition algorithms and an exponential rank combiner, all in a hierarchical manner. The performance figures of the proposed methodology are corroborated by extensive experiments performed on four benchmark datasets: GavabDB, Bosphorus, UMB-DB and FRGC v2.0. Results show marked improvement in alignment accuracy and recognition rates. Moreover, a computational complexity analysis has been carried out for the proposed algorithm which reveals its superiority in terms of computational efficiency as well.

---

**Keywords:** 3D FR, 3D alignment, profile face, SVM, unified classifier

## 1. Introduction

**B**iometrics refers to measurable physiological and behavioral attributes that can be used for automatic recognition of human beings. Over the last few decades; face, iris, gait, finger and palm print, based recognition approaches have been extensively employed in biometric systems. Face recognition is beneficial because it is non-invasive, socially acceptable and applicable to non-cooperative scenarios. The terms FV and FI mean that a subject's face is matched to claimed identity face only, or to the whole gallery respectively. The challenges of 2D FR include pose, illumination, and expression (PIE) variations. Pose variations are caused by changes in head rotations. Illumination variations occur due to changes in shadows and highlights, and reversal of contrast gradients, whereas, expression variations change geometric positions of the facial features. PIE variations result in distinct changes in facial appearances leading to degradation of system performance dramatically. In the recent past, 3D face acquisition and recognition techniques have largely evolved, and 3D shape based facial representations are considered to be much more robust to varying poses, and independent of illumination variations. Therefore, FR using 3D models is proving to be promising especially under unconstrained acquisition scenarios of real world. However, evolution of FR algorithms is required to differentiate between expression deformations and interpersonal disparities. Encouraged by these evidences, many 3D FR approaches have been proposed and experimented in the last few years as given in the work of Bowyer et al. [1] and literature reviews [2-5]. The existing 3D FR approaches can be categorized into holistic [6-8] and local feature based [3-4,9-11] domains.

The most crucial stage in any 3D FR approach is face alignment and the resulting recognition accuracy primarily depends on robustness of the alignment module. In alignment phase, facial features are transformed such that they can be reliably matched. A few 3D face alignment approaches [12-15] are: (i) Iterative Closest Point (ICP) [12], (ii) Simulated Annealing (SA) [13], (iii) Average Face Model (AFM) [14] and (iv) ICS [15].

ICP based approaches align two 3D faces by minimizing the distance between them iteratively, whereas, SA [13] algorithm uses a stochastic approach employing a local search method. Drawbacks of ICP include: (i) initial course alignment and, (ii) slow convergence; while SA's limitation is its excessive time consumption. The drawbacks of ICP and SA methods limit the use of these approaches to FV where, only one-to-one alignment to the claimed identity image is needed. They become problematic in FI where, one-to-many alignment events to the entire gallery are required. In AFM [14] based alignment, the AFM is constructed by locating and averaging the facial landmarks; whereas, the probe image is aligned to AFM only once. This aspect empowers this method to be used as an alignment technique in both of FV and FI scenarios. A significant downside of the AFM based algorithm is the probe image's less accurate alignment, as spatial information is lost during the averaging process. The fourth method, namely ICS; mainly involves localization of landmarks on 3D facial images, comparison of landmarks to the corresponding landmarks on ICS and a transformation stage to finalize the alignment. The weakness of ICS is the low accuracy of landmark detection procedure, particularly for the face images with pose and expression variations. A single alignment event required for a probe to align it to ICS makes this technique appropriate for FV as well as FI scenarios. Another alignment method is proposed in the work of Wang et al. [16] to place each face model to a standard orientation. It does not align a probe face to every face contained in the gallery; therefore, it can be

employed for both of FV and FI efficiently. This alignment method uses the facial symmetry plane which is determined using Principal Component Analysis (PCA) and ICP. Based on the normal of the symmetry plane, the nose tip, and nose bridge direction; six degrees of freedom are fixed in a 3D face to obtain a standard alignment posture.

Solutions enabling recognition of subjects with missing facial parts are now attracting an increasing interest. Profile image based FR is an example of such a case where profile face images are used in the recognition process. Profile face images are defined as the face images where a subject presents his/her head rotated approximately  $+90^\circ$  (right profile face image) or  $-90^\circ$  (left profile face image) around the vertical axis in  $xz$  plane [17]. In the proposed study, our aim is to present: (i) an efficient two-pass face alignment algorithm to deal with neutral and expressive 3D faces captured at frontal and profile poses. Frontal face images are captured with subjects facing towards camera at different angles, whereas, profile face images are captured from the face side. (ii) FV and FI algorithms employing SVM and unified classifiers respectively, for classification of 3D face images. The contributions and novelty of the proposed algorithm are as under.

1. The first contribution of this study is a novel two-pass 3D alignment approach which does not align two faces to each other employing the conventional procedure of alignment. In contrast, it uses a coarse to fine approach to accurately align the 3D face images to an ICS in the first pass at a very low computational cost of a single 3D rotation. The face images are aligned at fine level in the second alignment pass where they are rotated in the range of  $+2^\circ$  to  $-2^\circ$  in  $xz$  and  $yz$  planes separately, to minimize the nose tip-scanner distance in both planes. The proposed approach is (i) computationally less expensive than ICP, SA and the approach proposed in [16], and (ii) more accurate than AFM and ICS.

2. The second contribution of this study is a multi-view full face region based classification method where two sets of 3D facial images are synthesized at  $0^\circ$ ,  $\pm 10^\circ$ ,  $\pm 20^\circ$  and  $\pm 30^\circ$  and matched using SVM. For SVM based training and testing, Nearest neighbor (NN) classifier using Mahalanobis Cosine (MahCos) matching score is implemented to generate pairwise scores from two sets of synthesized images after PCA [6] based subspace learning. The proposed SVM classifier based multi-view FV approach produces excellent results than the contemporary approaches (Section 4.5).

3. The third contribution is a unified classifier based FI method which fuses results from seven base classifiers, two parallel FR algorithms, and an exponential rank combiner in a hierarchical manner to enhance the recognition accuracy. The proposed approach exhibits better accuracy than the existing methods (Section 4.5).

The performance of the proposed face alignment and recognition method is evaluated using four benchmark databases; namely, GavabDB [17], Bosphorus [18], UMB-DB [19] and FRGC v2.0 [20]. These databases carry pose and expression variations and are commonly utilized for developing 3D FR algorithms. For example, the algorithms presented in state-of-the-art studies [21-23,2] employed FRGC v2.0 database while those presented in Refs. [24,3,25,9,26-27] are based on both of FRGC v2.0 and GavabDB databases. Similarly the studies [9,26,28] employed GavabDB, Bosphorus and FRGC v2.0 databases while the study [10] used UMB-DB to evaluate the algorithm.

Rest of the study is organized as follows: The related work is discussed in Section 2. Section 3 deals with the details of proposed 3D alignment and FR methodology. Experiments and results are given in Section 4, whereas, the findings are concluded in Section 5.

## 2. Related Work

For a thorough survey of research in the field of 3D and multimodal 2D+3D FR, the reader is referred to the studies [1,5,29]. Review of the 3D, multi-view and fusion based FR algorithms is presented as following.

The study [30] proposed to synthesize several facial variations by utilizing a morphable model using 2D images. The advantage of this study is estimation of novel face views using 2D shape and texture information. On the other hand, the estimated face parameters limit the final recognition accuracies. The study [31] proposed a PCA based 3D face synthesis approach to generate new faces based on a reference face model. The advantage of this study is that it preserves face size information which helps in the recognition process. On its down side, the method requires correct segmentation of face area, establishment of accurate correspondences to the reference face model, and the selection of appropriate PCA training faces. The study [16] proposed a new 3D FR approach named as Collective Shape Difference Classifier (CSDC) by computing Signed Shape Difference Map (SSDM) between two aligned 3D faces. Due to computational efficiency and high recognition rates, this method is promising for deployment in practical biometric systems based on 3D FR. A drawback of the method is its failure when too many nose points are missing. Another shortcoming of the method is that it is validated using a single 3D face database and testing on other databases with varying conditions is missing. The study [32] proposed a fully automatic 3D FR system using multi-view 2.5D facial images. The approach employs directional maximum, to estimate the nose tip and the pose angle simultaneously. Face images are recognized using ICP based approach corresponding to best located nose tip. This method is beneficial because of classifying face images with large pose variations, whereas, ICP based computational complexity is its limitation.

A prominent algorithm based on fusion of 2D and 3D features is proposed in the study [33] employing PCA and canonical correlation analysis (CCA). The drawback of this algorithm is that it has been tested only on a single already aligned face database and, therefore, needs testing using unaligned face images from other 3D face databases. Another fusion based study is given in Ref. [34] equipped with an approach where match scores of each subject were combined for both of 2D albedo, and depth images. The use of 2D/3D images obtained using photometric stereo (a 3D reconstruction method) is the strength of this method. However, the method lacks experimentation/comparison on other benchmark 2D/3D databases. A recent facial region based study [10] proposed a method with the strength to handle missing data and occlusions. The disadvantages of the method are that, it can tackle small pose variations and is region based. A major limitation of the region based studies is computational complexity because each face region is to be matched to the whole gallery. Another limitation is that the face regions do not take into account holistic characteristics of the face. Please note that the following studies [35-36,2,28] are also region based on their downside.

A study based on fusion of results acquired from several overlapping facial regions, is proposed in Ref. [35] employing PCA-Linear Discriminant Analysis (LDA) based features and decision level fusion (majority voting). The strength of the method is its suitability to be used in biometric template protection techniques. The study [36] proposed a facial recognition system (FRS) which employed fusion of three face classifiers using feature and match score level fusion methods. The advantages of this method are; the use of pose invariant contour features, and saving in face template storage resources. The region based study [2] proposed the use of facial representation based on dual-tree complex wavelet transform (DT-CWT) and six facial sub regions. This study employed sparse feature

representation and is, therefore, computationally efficient which is its strength. A recent circular region based study [28] proposed an effective keypoint detection and matching methodology using histograms of principle curvature measures. The advantage of this study is its competitive performance on pose subsets.

Our 3D FR based study is focused on using multi-view face images. Inspired by the studies [30-32], the proposed FRS targets to enhance classification accuracies using, complementary information obtained from synthesized multi-view faces, and fusion strategies (motivated by [33-36,10,2,28]). The results obtained from our proposed methodology are better than the results reported by other state-of-the-art studies [2,3,10,21-24] in terms of all the evaluation criteria.

### 3. Proposed Methodology

The proposed FRS consists of three modules: (i) alignment, (ii) verification and (iii) identification as illustrated in Fig. 1 along with multi-view model parameters.

#### 3.1 Two-Pass 3D Face Alignment Method

Accuracy of final results of FRS significantly relies on the accuracy of the alignment component. In this study, a two-pass coarse to fine alignment algorithm is proposed which is based on ICS and MNSD. The proposed alignment approach is based on the intuition that when a face is aligned in a frontal orientation, its nose tip-scanner distance (the model parameter) is the minimal. Moreover, the computational complexity involved in finding MNSD using a brute force approach (which seeks all possible solutions) would be reduced by the course to fine approach. Firstly, face images are coarsely aligned in  $xz$  plane using ICS based alignment. Then they are aligned in  $xz$  plane using MNSD based fine approach (section 3.1.3). Similarly the face images are aligned in  $yz$  plane using ICS and MNSD based coarse to fine alignment approach (section 3.1.4). In the end face images are aligned in  $xy$  plane using the ICS based alignment approach (section 3.1.5). It is notable that alignment in  $xy$  plane is based on single pass, as in  $xy$  plane the nose-tip scanner distance remains unchanged.

##### 3.1.1 Intrinsic and World Coordinate Systems

An ICS is defined by three angles and an origin represented by angles  $\rho$ ,  $\theta$  and  $\phi$  and nose tip 'O' in Fig. 2 and is represented by  $u$ ,  $v$  and  $w$ -axis in the same figure. Similarly the world coordinate system is shown by  $x$ ,  $y$  and  $z$ -axis defined at the same nose tip 'O' as the origin in Fig. 2. A 3D model is also shown in Fig. 2, where  $l$  shows the MNSD.

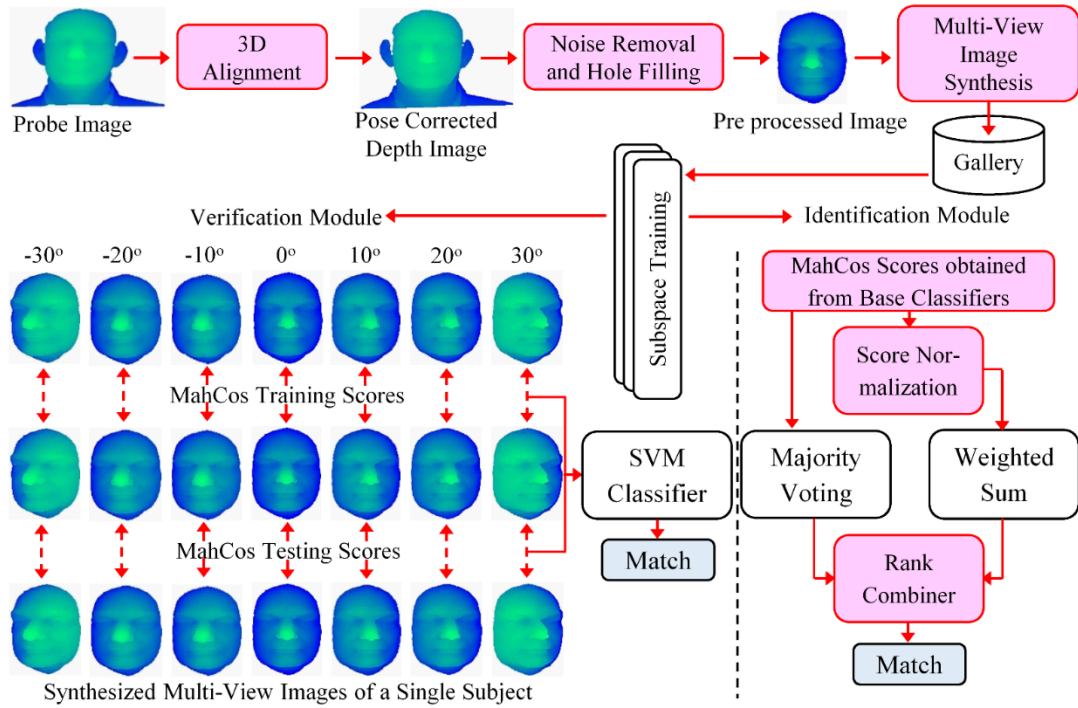


Fig. 1. Illustrative diagram of the proposed 3D FRS along with multi-view model parameters

Angle  $\theta$  is the angle inscribed between  $z$ -axis of world and  $w$ -axis of ICS in  $xz$  plane. Angle  $\rho$  is subtended between  $y$  and  $v$ -axis, whereas, angle  $\phi$  is inscribed between  $x$  and  $u$ -axis of world and ICS in  $yz$  and  $xy$  planes respectively. Angle  $\rho$  is used to estimate rotation parameter around  $x$ -axis, whereas, angle  $\theta$  and  $\phi$  are used for pose estimation around  $y$  and  $z$ -axis respectively. The procedure to determine nose tip and rotation angles  $\rho$ ,  $\theta$  and  $\phi$  is described in the following paragraphs.

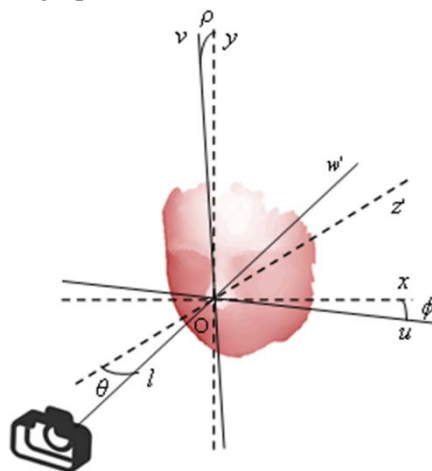
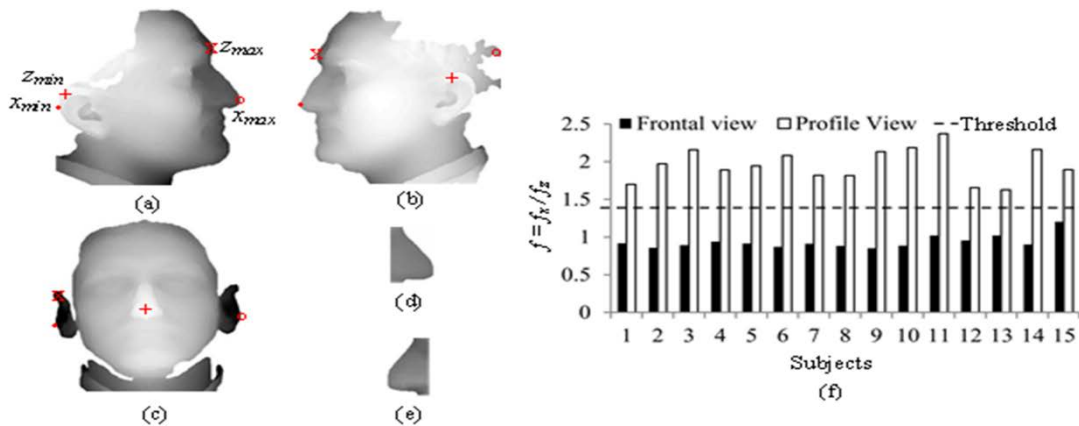


Fig. 2. Intrinsic and world coordinate systems along with subject ‘cara17\_frontal1’ from GavabDB

### 3.1.2 Nose Tip Detection Algorithm:

Nose tip detection is a specific and important feature detection issue in range images. The study [37] proposed a training free method to detect nose tip for frontal face images based on generating profile face images. The authors employed a histogram initialization and triangle fitting based approach and obtained a detection rate of 99.43% using FRGC v2.0 database. In contrast to the study [37], our study localizes nearest detected point from the 3D scanner, as the nose tip and is used for cropping the subject's face. In this study, nose tip detection process in profile face images is also addressed using a novel approach because, the nearest point in these faces is detected on ears or some other face region as presented in **Fig. 3 (a)** and **(b)**. Referring to **Fig. 3**, the proposed approach computes the differences  $f_x = x_{max} - x_{min}$  and  $f_z = z_{max} - z_{min}$  defining the length (L) and width (W) of a profile face image along  $x$ -axis (**Fig. 3 (a)** and **(b)**) while W and L of the frontal face along  $z$ -axis (**Fig. 3 (c)**). Then a function  $f = f_x / f_z$  is computed which returns a higher value than a threshold (empirically determined) for profile face images and lower for frontal face images. The threshold is calculated from a training set of 15 subjects using 1 frontal, 1 right profile and 1 left profile image for each subject. Higher value of function  $f$  in the former case is attributed to the lower value of half face width in the denominator. The experimental results for threshold calculation are shown in **Fig. 3 (f)**.



**Fig. 3.** Minimum and maximum distances of the face from 3D scanner (a) right profile (b) left profile (c) frontal view of subject 'cara6\_frontal1' (d) right oriented nose template of subject 'cara5\_derecha' (e) left oriented nose template of subject 'cara5\_izquierda' from GavabDB (f) experimental results

Using the threshold defined by the function  $f$ , the profile face images are separated from frontal face images. In order to classify between right or left profile, right and left nose templates (please see **Fig. 3 (d)** and **(e)**) are correlated with the profile face image. The right and left oriented nose templates of a single subject are randomly selected from the nose templates of six randomly chosen (three male and three female) subjects from the GavabDB database on satisfactory experimental performance. The profile face image is classified as right or left using greater value of the maximum correlation coefficient employing Normalized Cross Correlation (NCC). The right and left profile face images are rotated in the range of  $0^\circ$  to  $-90^\circ$  and  $0^\circ$  to  $+90^\circ$  respectively, using a step size of  $10^\circ$ , and then from  $-10^\circ$  to  $+10^\circ$  using a step size of  $1^\circ$  to mark the nose tip. The nose tip is marked at a point of minimum distance from the 3D scanner. The whole procedure requires 4 landmarks to

classify the profile face images from frontal face images:  $x_{min}$ ,  $x_{max}$ ,  $z_{min}$  and  $z_{max}$  (please see [Fig. 3\(a\)](#) and [\(b\)](#)).

In order to estimate the robustness of nose tip detection method, we use two features: (i) depth map variance, and (ii) highest peak of the depth image histogram. For a frontal, and correctly detected nose tip; former is the minimum while latter is the maximum. The mentioned features are determined using nine facial orientations along  $x$ -axis at  $0^\circ$ ,  $60^\circ$  and  $-60^\circ$  each with  $y$ -axis orientation at  $0^\circ$ ,  $60^\circ$  and  $-60^\circ$ . Finally, the mentioned features from all nine orientations are selected and the corresponding nose tip is declared as the correctly detected nose tip. We have manually observed that the proposed algorithm correctly detected the nose tips of face images from GavabDB, Bosphorus, UMB-DB and FRGC v2.0 databases. The proposed algorithm also correctly detected the nose tip in those images where the nearest point from the scanner to the face image was detected (without using the proposed algorithm) at forehead, lips, chin, right profile face and left profile face, as detailed in [Table 1](#). For frontal face images, the proposed alignment approach uses nose tip along with 2 landmarks in  $xz$  plane, and 1 in  $yz$  plane, whereas, nose tip and 1 landmark in  $yz$  plane is employed in case of profile face images.

**Table 1.** Nearest points from the scanner to the face without using the proposed nose-tip detection algorithm

Database	Forehead	Lips	Chin	Right profile	Left Profile
GavabDB	17	1	29	61	61
Bosphorus	126	9	79	210	210
UMB-DB	29	1	1		
FRGC v 2.0	10	22	2		

### 3.1.3 Alignment in $xz$ Plane

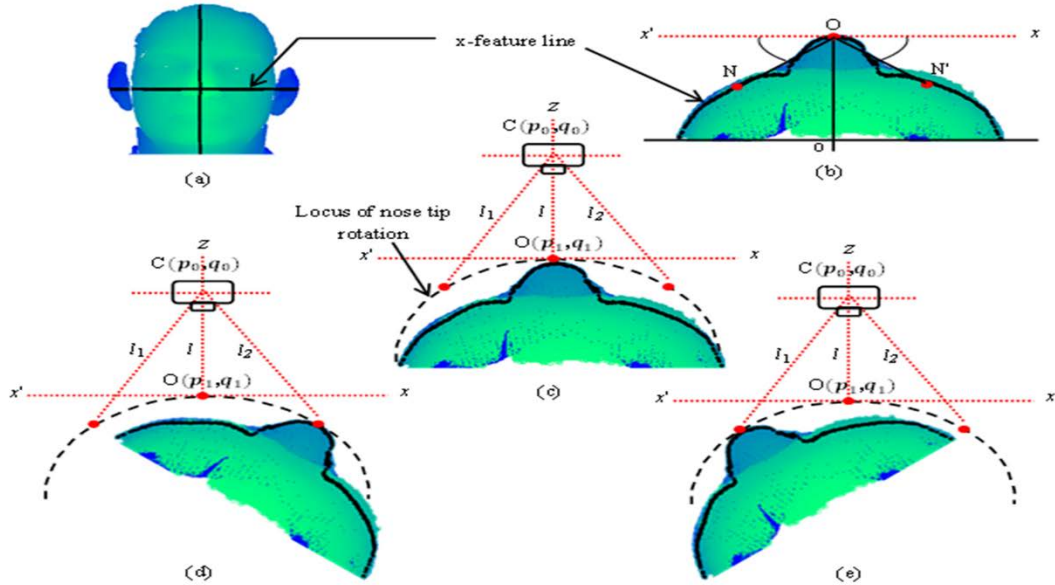
**First Pass:** (i) In the first pass, the frontal face images are aligned in  $xz$  plane by computing the correction angle  $\theta$  employing the procedure outlined below.

1. Find coordinates of the nose tip  $O(x, y, z)$ .
2. At nose tip level, select feature points in the direction of  $x$ -axis along contour of the face. These feature points constitute  $x$ -feature line and are shown in [Fig. 4\(a\)](#) and [\(b\)](#) in  $xy$  and  $xz$  planes respectively.
3. On the  $x$ -feature line, determine one landmark on each side of the nose tip at a distance of  $P = 45$  pixels, which are labeled as  $N$  and  $N'$  in [Fig. 4\(b\)](#). Measure angles  $x'ON$  and  $xON'$  (please see [Fig. 4\(b\)](#)), and calculate half of their difference angle which gives the required correction angle  $\theta$  in  $xz$  plane. e.g. if angle  $x'ON$  is  $34^\circ$  and angle  $xON'$  is  $48^\circ$  then half of difference of angles becomes  $7^\circ$ . Thus, an increase of  $7^\circ$  in angle  $x'ON$  and decrease of  $7^\circ$  in angle  $xON'$  sets the value of both equal to  $41^\circ$ .
4. Rotate the 3D face image using 3D rotation matrix [38] (equation 1) by substituting the computed value of angle  $\theta$ , and  $\rho = \phi = 0^\circ$ . This procedure aligns the 3D face images in  $xz$  plane.

$$A = \begin{bmatrix} \cos \theta \cos \phi & -\cos \rho \sin \phi + \sin \rho \sin \theta \cos \phi & \sin \rho \sin \phi + \cos \rho \sin \theta \cos \phi \\ \cos \theta \sin \phi & \cos \rho \cos \phi + \sin \rho \sin \theta \sin \phi & -\sin \rho \cos \phi + \cos \rho \sin \theta \sin \phi \\ -\sin \theta & \sin \rho \cos \theta & \cos \rho \cos \theta \end{bmatrix} \quad (1)$$



5. In order to handle errors in selection of landmarks, measure angles  $x'ON$  and  $xON'$ . Both angles must be equal, pointing to the fact that the image has been aligned correctly. If angles  $x'ON$  and  $xON'$  are not equal, then again select a landmark at a distance of  $P = P + 1$  pixels at both sides of the nose tip and repeat the above algorithm.



**Fig. 4.** Alignment algorithm using depth image of subject '02463d550.abs' from FRGC v2.0  
First pass: (a) frontal view in  $xy$  plane (b) top view in  $xz$  plane, Second pass: (c) aligned (d-e) unaligned

(ii) It was explained in nose tip detection process (section 3.1.2) that profile face images are rotated in the range of  $0^\circ$  to  $-90^\circ$  and  $0^\circ$  to  $+90^\circ$  respectively, to determine the nose tip. Therefore, they become in frontal view and are coarsely aligned in  $xz$  plane in this step.

**Second Pass:** The second pass of the alignment algorithm finely aligns both of the frontal and profile face images in  $xz$  plane using the approach depicted in Fig. 4 (c), (d) and (e). The algorithm aligns the face images at a MNSD by rotating them in the range of  $-2^\circ$  to  $+2^\circ$  using a step size of  $0.5^\circ$ . The nose-tip scanner distance is further reduced by rotating the face images in the range of  $-0.5^\circ$  to  $+0.5^\circ$  using a step size of  $0.1^\circ$ . The rationale behind the second pass strategy is discussed as follows. Referring to the Fig. 4 (c), (d) and (e), the nose-tip scanner distance  $l_1$  or  $l_2$  of an unaligned face has a larger value and is reduced for a nose tip aligned with the  $z$ -axis line. The face alignment causes an essential reduction in the nose-tip scanner distance meaning that the nose-tip scanner distance  $l$  must have a smaller value than  $l_1$  or  $l_2$  in Fig. 4 (c), (d) and (e). This fact can be verified from Fig. 4 (c), (d) and (e) by observing that among all distances such as  $l_1$  and  $l_2$ , the distance  $l$  between the nose tip  $O(p_1, q_1)$  and the 3D scanner point  $C(p_0, q_0)$  is a perpendicular to the line  $x'x$ . Since the perpendicular distance from any point on a line segment is always the minimum, therefore,  $l$  is smaller than  $l_1$  or  $l_2$ . For an aligned face, the distance between the points  $C(p_0, q_0)$  and  $O(p_1, q_1)$  is calculated using the distance formula as given in equation 2.

$$l = \sqrt{(p_0 - p_1)^2 + (q_0 - q_1)^2} \quad (2)$$

### 3.1.4 Alignment in yz Plane

**First Pass:** In the first pass, the proposed alignment approach aligns both of frontal and profile face images in yz plane by determining the correction angle  $\rho$  using the procedure outlined below.

1. Find the nose tip  $O(x, y, z)$  coordinates.
2. Through the nose tip, select feature points in the direction of y-axis along contour of the face. These feature points constitute y-feature line and are shown in **Fig. 5(a)** and **(b)** in  $xy$  and  $yz$  planes respectively.
3. On the y-feature line, determine one landmark on the nose bridge at a distance of  $Q = 30$  pixels from the nose tip which is shown by point  $t$  in **Fig. 5(b)**. Measure the correction angle  $\rho = \angle yOt$  (please see **Fig. 5(b)**) to adjust the head tilt  $\rho$  at  $30^\circ$ . This step aligns the face image in an upright frontal view [35].
4. Rotate the 3D face image by substituting the computed value of angle  $\rho$ , and  $\theta = \phi = 0^\circ$  in equation 1. This process aligns the faces in yz plane.
5. In order to handle errors in selection of the landmark, measure angle  $\rho = \angle yOt$ . This angle must be equal to  $30^\circ$  pointing to the fact that the image has been correctly aligned. If angle  $\rho$  is not equal to  $30^\circ$ , then again select a landmark at a distance of  $Q = Q + 1$  pixels from the nose tip and repeat the above algorithm.

**Second Pass:** The second alignment pass finely aligns both of the frontal and profile face images by rotating them in yz plane. The algorithm first aligns the face images at a MNSTD by rotating them in the range of  $-2^\circ$  to  $+2^\circ$  using a step size of  $0.5^\circ$ . The mentioned distance is further reduced by rotating the face images in the range of  $-0.5^\circ$  to  $+0.5^\circ$  with a step size of  $0.1^\circ$ .

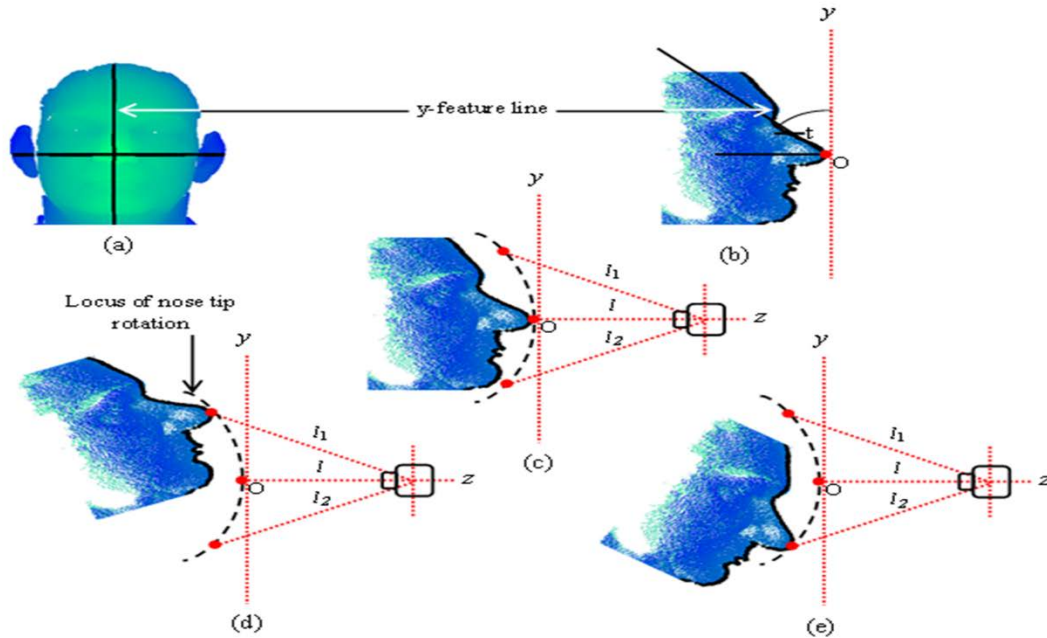
### 3.1.5 Alignment in xy Plane

1. Rotate both of the right and left halves of the face image around z-axis in the range of  $-5^\circ$  to  $+5^\circ$  using a step size of  $1^\circ$ .
2. For each of the above mentioned rotations, mirror any of the right or left half face in  $xy$  plane and slide on the other half such that they are entirely overlapped.
3. For the pixels at the same grid position  $(i, j)$  of overlapping half faces, find the sum of differences,  $D_{z(i,j)}$ , using the values less than the threshold  $T$  (to rule out any possible outliers in z-direction, because of face acquisition process) such that

$$S = \sum_{i,j} \begin{cases} 0 & D_{z(i,j)} > T \\ D_{z(i,j)} & otherwise \end{cases} \quad (3)$$

4. Select the rotation angle corresponding to the minimum value of  $S$  (equation 3) indicating a good match.
5. Around the angle determined in step 4, repeat steps 1 to 4 by rotating the face in the range of  $-1^\circ$  to  $+1^\circ$  using a step size of  $0.1^\circ$ . e.g. if the value determined in step 4 is  $-2^\circ$ , then rotate in the range of  $-3^\circ$  to  $-1^\circ$ . The outcome of the whole process is the required correction angle  $\phi$ .
6. Rotate the face using equation 1 by substituting the computed value of angle  $\phi$ , and  $\rho = \theta = 0^\circ$ . This procedure aligns face images in  $xy$  plane.

**(ii)** In the  $xy$  plane, the profile face images aligned in  $xz$  and  $yz$  planes come up as half faces. The other half of the probe profile face is synthesized by AFM based approach and the procedure outlined above for frontal face images is adopted for finding correction angle  $\phi$ .



**Fig. 5.** Alignment algorithm using depth image of subject '02463d550.abs' from FRGC v2.0  
 First pass: (a) frontal view in  $xy$  plane (b) side view in  $yz$  plane, Second pass (c) aligned (d-e) unaligned

### 3.2 3D Face Recognition

Depth face images carry artifacts such as noise and gaps which are handled through the preprocessing step. The related noise removal process suggested in this work consists of three major steps: (i) the noise in 3D data generally occurs along the viewing direction ( $z$ -axis) of the sensor, hence, the sharp spikes introduced in the depth images during the image acquisition process are removed using median filtering, (ii) the facial image holes are filled using interpolation, and (iii) facial irregularities are smoothed through low pass filtering at the end of preprocessing phase. Please note that we do not employ multi-view images which are captured at various angles in the acquisition phase. Rather, the novel views at  $0^\circ$ ,  $\pm 10^\circ$ ,  $\pm 20^\circ$  and  $\pm 30^\circ$  are synthesized by rotating the face image oriented at  $0^\circ$  around  $y$ -axis, and are arranged in two sets of seven depth face images. Intuitively, by using the additional information of synthesized multi-view face images of the same subject, the FR accuracy would be improved. The features provided by multi-view face images are stable to study several local face variations. The occluded and less visible regions of frontal face images also become visible in multi-view images. An example of multi-view face images used in our study is depicted in [Fig. 1](#) showing the complementary 3D information which helps to enhance the recognition accuracies.

#### 3.2.1 Classifier Fusion

The results obtained from several classifiers can be fused to improve the recognition accuracies using numerous methods [39] such as sensor, feature, rank, decision and score level fusion [40-41]. Score level fusion methods are categorized as: (i) transformation-based e.g. weighted sum-rule (ii) density-based e.g. likelihood ratio test with Gaussian Mixture

Model and (iii) classifier based e.g. SVM based fusion approach. The verification set up in our study is realized using SVM which is explained in the following section.

### 3.2.2 SVM Classifier

Intuitively, an SVM defines a decision surface in the training phase to maximize the separation between the closest points termed as support vectors in a higher dimensional space. For a binary classification problem, an OSH,  $w_x + b = 0$ , with maximum margins separates the training vectors obtained from two classes  $(x_1, y_1), \dots, (x_i, y_i)$  where  $x_i \in R_n$  and  $y_i \in \{1, -1\}$ . The following objective function is minimized to define OSH having constraints  $y_i[(w \cdot x_i) + b] \geq 1 - \xi_i, \xi_i \geq 0$  for  $i = 1, \dots, k$ .

$$\phi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^k \xi_i \quad (4)$$

where  $\xi_i$  are slack variables used to penalize the errors when data are not separable linearly, and  $C$  is regularization constant. Now sign of the following OSH function can be used for classification of a test point.

$$f(x) = \sum_{i=1}^k y_i a_i K(x, x_i) + b \quad (5)$$

where  $a_i \geq 0$  are the Lagrangian multipliers and  $b$  is computed by the optimization problem mentioned above. In equation 5  $K$  is the kernel trick for class separation in a higher dimensional space,  $x_i$  is the  $i$ th training class and  $x$  is the test class. It is experimentally observed in this study that kernel SVM yields better recognition rates than the linear SVM. Therefore, radial basis function (RBF) kernel is used in our study as given below where  $\sigma^2$  is spread of RBF.

$$K(x, x_i) = \exp \left[ \frac{-\|x - x_i\|^2}{2\sigma^2} \right] \quad (6)$$

### 3.2.3 Face Verification Algorithm

FV algorithm employs SVM based classification using MahCos scores computed between depth face images which are aligned and synthesized by the proposed algorithm at 7 orientations. For each of the face image, the original feature space is  $500 \times 500$  which is reduced using PCA. After PCA based subspace learning, the feature space is reduced to  $G \times 1$  where  $G$  is the size of the gallery. In order to train the SVM, the algorithm uses two neutral face images of each subject to compute MahCos training scores. MahCos score is the Cosine score between two vectors  $s$  and  $t$  of image space and is calculated in the Mahalanobis space as described by equations 7 & 8 [42].

$$d_{MahCos}(s, t) = \frac{m \cdot n}{|m| |n|} = \frac{\sum_{i=1}^N (m_i n_i)}{\sqrt{\sum_{i=1}^N (m_i)^2} \sqrt{\sum_{i=1}^N (n_i)^2}} = \frac{\sum_{i=1}^N \left( \frac{s_i}{\sigma_i} \frac{t_i}{\sigma_i} \right)}{\sqrt{\sum_{i=1}^N \left( \frac{s_i}{\sigma_i} \right)^2} \sqrt{\sum_{i=1}^N \left( \frac{t_i}{\sigma_i} \right)^2}} \quad (7)$$

where  $m_i = \frac{s_i}{\sigma_i}$ ,  $n_i = \frac{t_i}{\sigma_i}$  and  $\sigma_i$  is standard deviation of the  $i$ th dimension. The higher

$$D_{MahCos}(s, t) = 1 - d_{MahCos}(s, t) \quad (8)$$

similarity reflects higher score in this case. Thus, the final distance is computed as under. Referring to **Fig. 6**, MahCos scores are computed between the first neutral image of each subject and second neutral image of the whole gallery  $G$ . The scores are computed by arranging images in (training, gallery) pairs oriented at  $(0^\circ, 0^\circ)$ ,  $(10^\circ, 10^\circ)$ ,  $(-10^\circ, -10^\circ)$ ,  $(20^\circ, 20^\circ)$ ,  $(-20^\circ, -20^\circ)$ ,  $(30^\circ, 30^\circ)$ ,  $(-30^\circ, -30^\circ)$  and stored in row 1 to 7 of a training score matrix respectively. Each element  $S_{ij}$  represents the score computed between image  $i$  to image  $j$  where  $i, j \in \{1, 2, \dots, G\}$ . It is clear that the element  $S_{ij}$  (for  $i=j$ ) corresponds to the MahCos score computed between an image and itself and, therefore, represents genuine score. All other scores  $S_{ij}$  computed between image  $i$  and  $j$  represent imposter scores. Focusing on the score matrix, it is revealed that all genuine scores e.g  $S_{11}$  corresponding to all 7 orientations constitute  $7 \times 1$  dimensional column vectors of genuine scores. Similarly, all imposter scores e.g  $S_{1G}$  corresponding to all 7 orientations constitute  $7 \times 1$  dimensional column vectors of imposter scores. The entire procedure results into creation of  $7 \times 1$  dimensional genuine and imposter column vectors which are referred to as training vectors. For an example gallery of 10 subjects, there will be  $G \times G$  total,  $G$  genuine and  $G^2 - G$  imposter training vectors totaling into 100, 10 and 90 respectively. In the classification phase, MahCos probe scores are computed between the aligned and synthesized probe image and second neutral image of the whole gallery (please see **Fig. 6**). The computed scores between (probe, gallery) pairs oriented at  $(0^\circ, 0^\circ)$ ,  $(10^\circ, 10^\circ)$ ,  $(-10^\circ, -10^\circ)$ ,  $(20^\circ, 20^\circ)$ ,  $(-20^\circ, -20^\circ)$ ,  $(30^\circ, 30^\circ)$  and  $(-30^\circ, -30^\circ)$  are arranged in row 1 to row 7 of the probe score matrix. Consequently, probe score matrix is populated with  $7 \times 1$  dimensional, 1 genuine and  $G-1$  probe vectors as shown in **Fig. 6**. Based on the genuine and imposter (training and probe vectors), the SVM classifies the probe image against the whole gallery. To search for the best possible model parameters, hyper plane parameter tuning is used. With RBF kernel, we used  $C=0.01$ , and  $\gamma=0.001$  to achieve the best verification accuracy.

### 3.2.4 Face Identification Algorithm

FI algorithm is based on a three stage unified classifier that employs  $500 \times 500$  dimensional depth face images oriented at  $0^\circ, 10^\circ, -10^\circ, 20^\circ, -20^\circ, 30^\circ, -30^\circ$ . PCA is employed to reduce the dimensionality of the feature space to  $G \times 1$  where  $G$  is the size of the gallery. Unlike the FV algorithm which employs SVM based classification using two neutral images per subject to compute MahCos training scores, the FI algorithm employs MahCos score based NN classification approach as base classifiers using one neutral image per subject.

Referring to **Fig. 7**, at first stage of the unified classifier, MahCos scores between probe and gallery images are computed for (probe, gallery) pairs oriented at  $(0^\circ, 0^\circ)$ ,  $(10^\circ, 10^\circ)$ ,  $(-10^\circ, -10^\circ)$ ,  $(20^\circ, 20^\circ)$ ,  $(-20^\circ, -20^\circ)$ ,  $(30^\circ, 30^\circ)$ ,  $(-30^\circ, -30^\circ)$ . Thus, 7 MahCos score based base classifiers produce 7,  $(1 \times G)$  dimensional) matching (row) score vectors. The results obtained from first stage are fused using (i) two parallel face recognition algorithms, namely: weighted sum rule and majority voting, and (ii) a rank combiner module at second and third stage, respectively.

**Score Level Fusion:** In the weighted sum method of score level fusion, the computed scores by a base classifier are multiplied by an empirically computed weight for the same base classifier. The obtained multiplication results are then added up and arranged in the form of a ranked list as the output. Mathematically

$$d_x = \sum_{i=1}^m W_i \cdot d_{ix} \quad (9)$$

where  $W_i$  is the weight and  $d_{ix}$  is the MahCos score vector of the  $i$ th classifier, while  $d_x$  is the total weight of all classifiers for probe image  $x$ .  $W_i$  is determined as under.

$$W_i = \frac{\text{Recognition accuracy obtained from each base classifier}}{\sum \text{Recognition accuracies obtained from all base classifiers}} \quad (10)$$

For score normalization, min-max normalization approach is employed which maps scaled original distribution of matching scores to the interval [0, 1]. If  $\max(X)$  and  $\min(X)$  respectively represent maximum and minimum raw matching scores, then the normalized score is calculated as

$$x' = \frac{x - \min(X)}{\max(X) - \min(X)} \quad (11)$$

where  $x \in X$  and  $X$  is the set of computed scores by a base classifier, and  $x'$  represents the normalized value of  $x$ .

**Decision Level Fusion:** Majority voting based decision level fusion is naïve in the nature where the base classifiers contribute one vote each and the classification decision is based on a total of more than half votes. In the context of our application, the classification is based on four or more votes. The output of majority voting is a ranked list of obtained class labels.

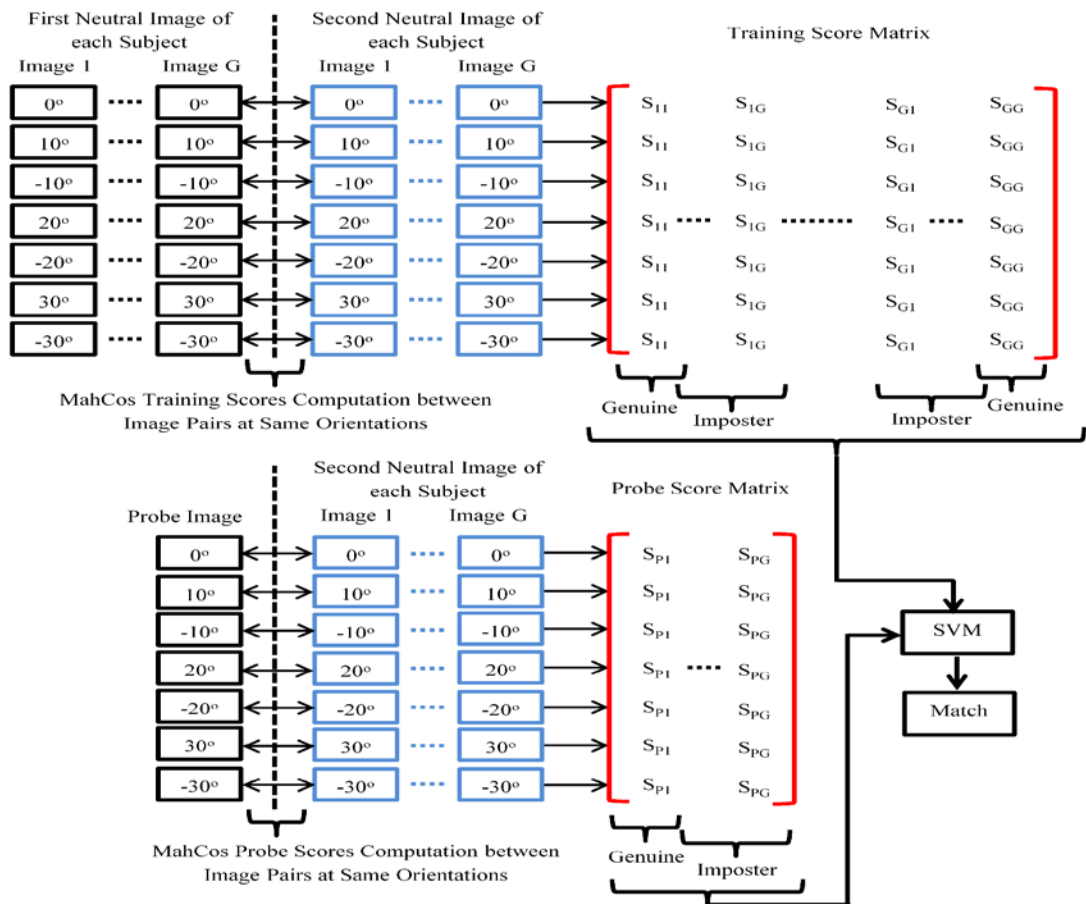


Fig. 6. Face verification algorithm

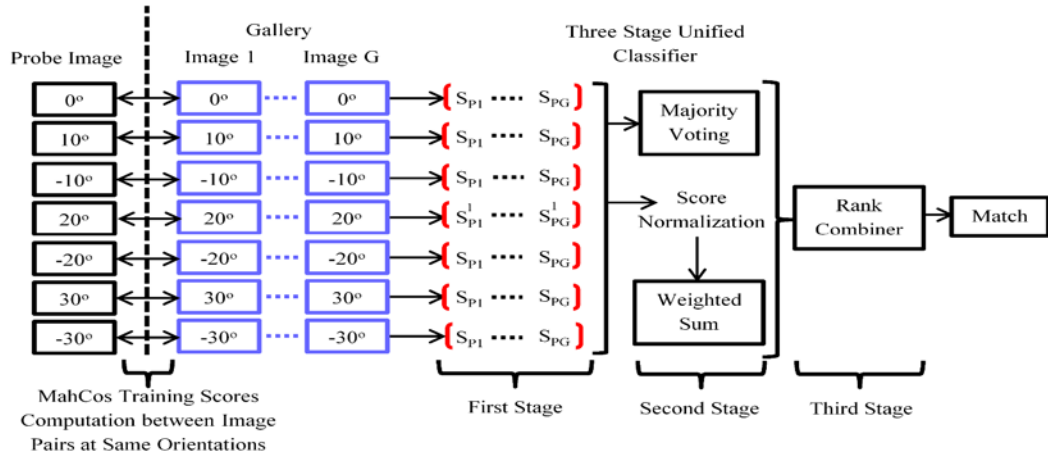


Fig. 7. Face identification algorithm

**Rank Combiner:** The rank combiner module is employed to fuse results yielded by the weighted sum rule and majority voting based classifiers; whereas, each of these classifiers and the ranks produced by them are assigned empirically computed weights. The mathematical description for the rank combiner component [3] [9] of the unified classifier is given below.

$$W_t = \sum_{j=1}^m W_{c_j} \cdot W_r \quad (12)$$

where  $W_t$  is the total weight computed for the probe image  $x$ .  $W_{c_j}$  is the weight assigned to classifier  $j$  determined using equation 10. Thus, recognition accuracy obtained from classifier  $j$  is divided by the total recognition accuracies obtained from both classifiers.  $W_r$  is the weight assigned to a ranked list produced by the classifier  $j$  implementing a parallel FR algorithm for probe image  $x$  against the gallery, as under:

$$W_r = e^{-Rx} \quad (13)$$

where  $Rx$  represents the corresponding matched class rank. Therefore, the classes obtaining decreasing ranks in an order, are assigned exponentially decreasing weights respectively. Final ranks are computed by the rank combiner using the total weights; contributed by weighted sum rule and majority voting based classifiers, for a probe image against the gallery.

#### 4. Experiments and Results

Our proposed algorithm targets to design an efficient approach by performing experiments on four databases, namely, GavabDB, Bosphorus, UMB-DB and FRGC v2.0. One experiment is conducted to implement the proposed alignment algorithm. Similarly FV and FI experiments are performed to implement proposed FR algorithm using min-max normalized scores. In the verification experiments the performance is reported as verification

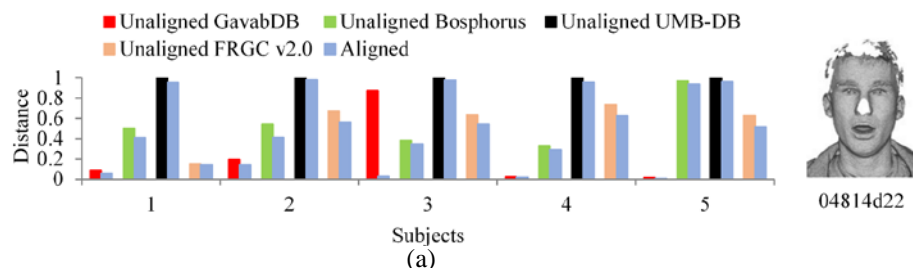
rate (VR) at 0.1% false accept rate (FAR), whereas, in the identification experiments the performance is reported as rank-1 identification rate (IR). We used 5 distinct classifiers including SVM, Random Forest (RF), Decision Tree (DT), Adaboost and Convolutional Neural Network (CNN) [43]. The SVM classifier defines a decision surface in the training phase to maximize the distance to the support vectors. In contrast, the RF based classifier is an ensemble growing tree-based method that operates by constructing a multitude of decision trees in training, and outputting the matched class. For details of RF, DT, Adaboost and CNN, the reader is referred to the study [43]. The considered 3D databases, GavabDB [17], Bosphorus [18], UMB-DB [19] and FRGC v2.0 [20] are reviewed in the following section along with description of experiments and results.

#### 4.1 3D Face Databases

The GavabDB [17] database contains 549 3D facial images acquired from 45 male and 16 female subjects under pose and expression variations (including profile face images). The Bosphorus database [18] contains a total of 4666 scans with pose and expression variations collected from 61 male and 44 female subjects. The UMB-DB database [19] is composed of 1473 3D depth images of 142 [10] subjects. Similarly, FRGC v2.0 3D database [20] [44] contains 4006 images with pose and expression variations.

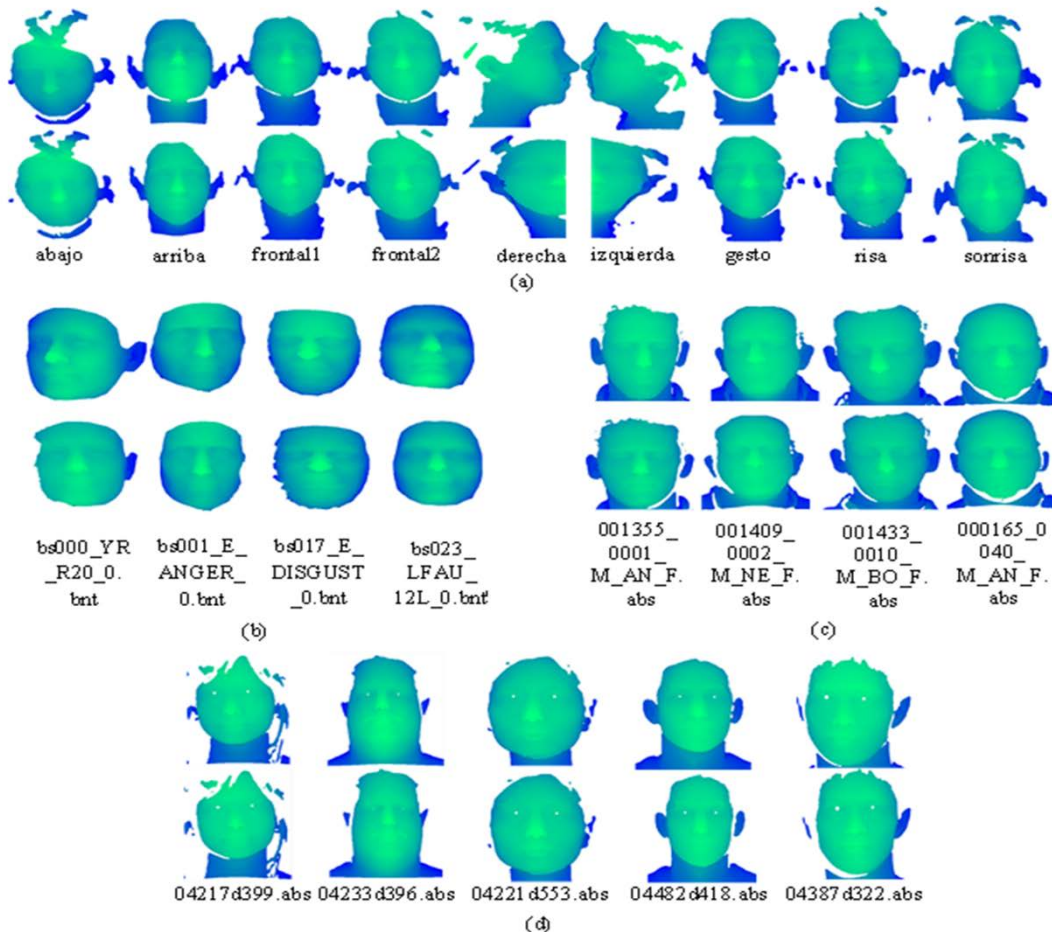
#### 4.2 Face Alignment Experiments

Using the proposed two-pass algorithm, alignment experiments are performed on GavabDB, Bosphorus, UMB-DB and FRGC v2.0 databases to align the faces at the MNSD. For evaluation of the alignment accuracy, there is no existing evaluation criterion. One method that can be employed is human judgment but human judgment method is not automatic. Therefore, the proposed alignment method is evaluated using the ‘nose tip-scanner distance evaluation method’. It is observed in the experiments that the results of the ‘nose tip-scanner distance evaluation method’, and manual judgment are quite similar; and that ‘nose tip-scanner distance evaluation method’ is a promising automatic criterion to check alignment accuracy. The minimized and normalized distances for 5 unaligned images from each of GavabDB, Bosphorus, UMB-DB, and FRGC v2.0 databases are shown in Fig. 8. Fig. 9 (a), (b), (c) and (d) depict original and aligned depth images of subject ‘cara1’ from GavabDB database and some of the subjects from Bosphorus, UMB-DB and FRGC v2.0 databases respectively. The proposed algorithm accurately minimized nose tip-scanner distances of 99.82%, 100% (non-occluded), 100% and 99.95% subjects from above mentioned four databases, respectively. The nose tip was not detectable for one subject in GavabDB database and two subjects in FRGC v2.0 database due to missing nose tip region; else the accuracy of the proposed algorithm would have been 100% for each of these databases. Please note that the face images with missing nose tip region were manually aligned. An example face image with missing nose tip region is shown in Fig. 8(b)



**Fig. 8.** (a) Nose tip-scanner distances for 5 subjects from (i) GavabDB: ‘cara1\_abajo’ to ‘cara1\_forntal2’ (ii) FRGC v2.0: ‘04202d566.abs’ to ‘04202d574’ (iii) UMB-DB: ‘000001\_0190\_F\_NE\_F.abs’ to ‘000005\_0190\_F\_NE\_F’ (iv) Bosphorus: ‘bs000\_CAU\_A22A25\_0.bnt’ to ‘bs000\_E\_ANGER\_0.bnt’ (b) Subject with missing nose tip from FRGC v2.0 database





**Fig. 9.** Example 3D faces: original: (row 1, 3, 5), Aligned: (row 2, 4, 6) from (a) GavabDB (b) Bosphorus (c) UMB-DB (d) FRGC v2.0

### 4.3 Face Recognition Experiments

#### 4.3.1 Experiments on GavabDB Database

(i) The performance of the FV algorithm is evaluated using the protocol used in Ref. [24] where ‘frontal1’ image of each subject is included in the gallery and image ‘frontal2’ is used as probe. Referring to section 3.2.3, the neutral face image ‘abajo’ is also enrolled in the gallery with ‘frontal1’ to calculate pairwise training scores, whereas, the images ‘frontal2’ and ‘frontal1’ are used for pairwise probe score computation for the N vs. N experiment. We also repeated this experiment by using RF, DT, Adaboost and CNN.

(ii) For the FI experiment, experimental setting of [27] is considered to conduct three N vs. N experiments. According to the mentioned protocol, the image ‘frontal1’ belonging to each of 61 subjects is enrolled in the gallery, whereas, the images ‘frontal2’, rotated looking down and rotated looking up are used as probe sets.

(iii) For FI of profile face images, the proposed algorithm considers four base classifiers for right ( $0^\circ$ ,  $10^\circ$ ,  $20^\circ$ ,  $30^\circ$ ) and left ( $0^\circ$ ,  $-10^\circ$ ,  $-20^\circ$ ,  $-30^\circ$ ) profile views. The scores of four base

classifiers are fused by weighted sum rule based classifier, whereas, majority voting based classifier decides on three votes to classify a true positive. The FV and FI results produced by the proposed approach for N vs. N experiments are given in [Table 2](#).

**Table 2.** Rank-1 recognition rates (%) using GavabDB database, U and W represent un-weighted and weighted recognition rates respectively

Proposed methodology	VR <sup>(1)</sup>	Rotated Looking down		Rotated Looking up		Frontal Face		Right Profile		Left Profile	
		U	W	U	W	U	W	U	W	U	W
SVM classifier	100					-	-	-	-	-	-
Score based fusion	-	91.8	95.08	93.44	96.72	93.44	96.72	77.05	80.33	88.52	91.80
Decision based fusion	-		93.44		95.08	-	95.08	-	78.69	-	90.16
Unified classifier	-	95.08	98.4	96.72	100	96.72	100	80.33	83.61	91.8	95.08

<sup>(1)</sup> CNN, Adaboost, RF and DT produced VR of 98.4%, 96.72%, 93.44% and 90.16%, respectively.

### 4.3.2 Experiments on Bosphorus Database

The proposed identification algorithm is evaluated by performing N. vs. N experiment using experimental protocol of Ref. [10] where the gallery set consists of first neutral images of each subject (105 scans) and the probe set contains rest of 194 neutral images, and the challenging pose subsets in separate experiments. The performance of the proposed multi-view synthesis based identification approach is given in [Table 3](#).

**Table 3.** Rank-1 recognition rates (%) using Bosphorus and UMB-DB databases, U and W represent un-weighted and weighted recognition rates respectively

Proposed methodology	Bosphorus								UMB-DB	
	Frontal Face		YR <90° 525 images		YR =90° 210 images		Overall 1365 images		Frontal Face	
	U	W	U	W	U	W	U	W	U	W
Score based fusion	94.29	96.19	89.3	92.4	80.5	83.3	89.5	92.4	95.07	96.48
Decision based fusion		94.29		90.3		81.4		90.5		95.07
Unified classifier	98.1	100	91.2	95.1	83.8	86.7	92.7	95.9	97.89	99.3

### 4.3.3 Experiments on UMB-DB Database

For evaluation of the proposed multi-view synthesis based FI algorithm, we employ the experimental protocol of Ref. [10] to create the N. vs. N. experiment where the gallery is comprised of one neutral image per subject (142 scans) and the probe set contains all remaining neutral scans (299 scans). The performance figures of our approach are given in [Table 3](#).

### 4.3.4 Experiments on FRGC v2.0 Database

(i) We investigated the performance of proposed FV algorithm by creating N vs. N experiment using 2469 neutral scans from FRGC v2.0 database [21]. This database consists of 370 subjects having at least two neutral face images [25]. Therefore, 740 face images (two per subject) are enrolled in the gallery, whereas, first two stored neutral images are enrolled

for the subjects having more than two neutral face images. Rest of the neutral images are included in the probe set. We also performed this experiment by using RF, DT, Adaboost and CNN.

(ii) The performance of the FI algorithm is evaluated using experimental protocol of Ref. [21] for N vs. N experiment. In this experiment, probe set is created using 2003 neutral images, whereas, 466 images (first neutral image for each of 466 subjects) are enrolled in the gallery. The results of the proposed FV and FI algorithms for N vs. N experiments are shown by receiver operating characteristic (ROC) curves in Fig. 10 (a) and cumulative match characteristic (CMC) curves in Fig. 10 (b). Please note that recognition rates obtained from score based fusion and unified classifier were increased by 2.36% and 2.58% respectively using weight assignment strategy. Examples of successful and failure cases from GavabDB, Bosphorus, UMB-DB and FRGC v2.0 are shown in Fig. 11. Referring to Fig. 11 (f), (g), (h), (l), (m), and (o), the failure is because of the hair occluding the face images; whereas, it is because of motion artifacts in case of Fig. 11(q).

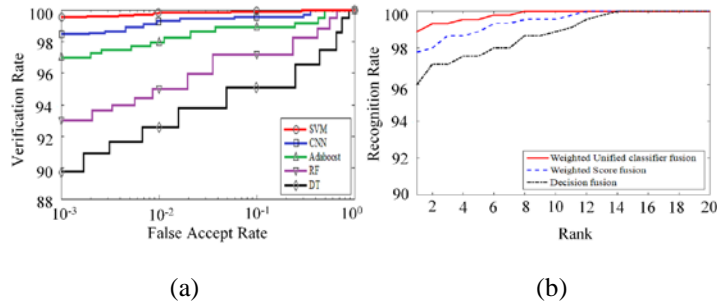


Fig. 10. (a) ROC (b) CMC curves of the proposed approach using FRGC v2.0 for N vs. N experiment

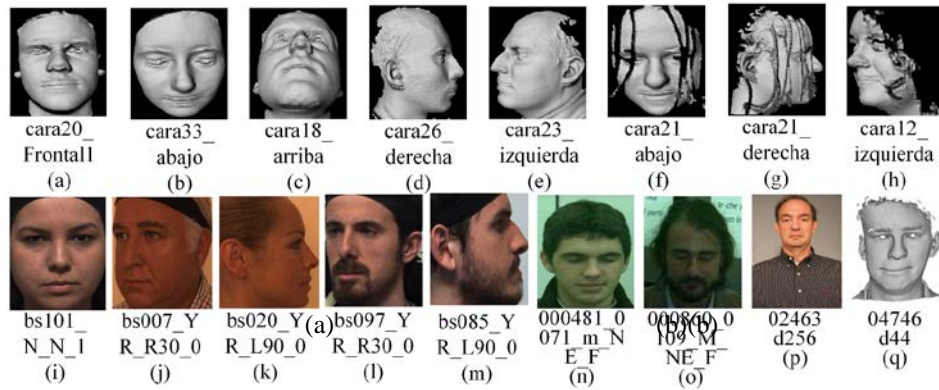


Fig. 11. Successful cases from GavabDB: (a) frontal face (b) rotated looking down (c) rotated looking up (d) right profile (e) left profile, Bosphorus: (i) frontal face (j) YR < 90° (k) YR = 90°, UMB-DB: (n) frontal face, FRGC v2.0: (p) frontal face; Failure cases from GavabDB: (f) rotated looking down (g) right profile (h) left profile, Bosphorus: (l) YR < 90° (m) YR = 90°, UMB-DB: (o) frontal face, FRGC v2.0: (q) frontal face

#### 4.4 Computational Complexity Analysis

Computational complexity of the proposed FR approach is discussed in Table 4. Please note that, 9 multiplication along with 6 addition operations are involved to rotate a single 3D point. Subtraction of 1 in the terms  $V_F - 1$  and  $V_P - 1$  is due to the face image aligned at 0° in the alignment phase. Referring to equation 6; for FV algorithm, 7 more multiplications, 13 additions and 1 load operation are involved for each of the 163 support vectors.

The complexity of the proposed approach is compared with that of ICP algorithm (neglecting complexity of the initial course alignment) which employs Mean Squared Error (MSE) distance computation [25] involving 3 multiplication and 6 addition operations. It is assumed that for best ICP based alignment, 11 rotations in each of x, y and z directions at  $0^\circ$ ,  $\pm 2^\circ$ ,  $\pm 4^\circ$ ,  $\pm 6^\circ$ ,  $\pm 8^\circ$  and  $\pm 10^\circ$  (total 1331 rotations) are sufficient. Table 4 reveals that the proposed methodology is computationally more efficient than ICP.

It is obvious that the proposed multi-view methodology is computationally extensive compared to the case if single view images at  $0^\circ$  are used for recognition. It is notable that the training and projection is usually performed offline, whereas, the distance calculation and recognition is accomplished online [42]. Therefore, ignoring the PCA based projection step in Table 4, the complexity of the multi-view algorithm turns out to be 8.87% greater than the single view case. On the other hand, this increased complexity improves the recognition accuracy to 12.17% for the experiment given in section 4.3.4 (ii). This trend remains persistent in the other experiments performed in this study.

**Table 4.** Computational complexity analysis for the proposed methodology using  $M=300000$ ,  $m \times n=500 \times 500$ ,  $m_1 \times n_1=500 \times 250$ ,  $G$ =Gallery of 466 subjects,  $g$ =reduced feature space= $G$ ,  $V_F=7$  synthesized frontal faces,  $V_P=4$  synthesized profile faces

		Multiplications		Additions		
		Frontal Faces	Profile Faces	Frontal Faces	Profile Faces	
Proposed Methodology	Nose Tip Detection	-	$M \times 9 \times 40^{(1)}$	$M \times 1^{(2)}$	$M \times 6 \times 40 + M \times 40 \times 1$	
	Alignment	First Pass	$M \times 9$	-	$M \times 6$	-
		Second Pass	$M \times 9 \times 72^{(3)}$	$M \times 9 \times -52^{(4)}$	$M \times 6 \times 72$	$M \times 6 \times 52$
	Multi-View Synthesis	$M \times 9 \times (V_F - 1)$	$M \times 9 \times (V_P - 1)$	$M \times 6 \times (V_F - 1)$	$M \times 6 \times (V_P - 1)$	
	PCA	$(m \times n) \times G \times V_F$	$(m_1 \times n_1) \times G \times V_P$	$(m \times n) \times G \times V_F$	$(m_1 \times n_1) \times G \times V_P$	
	MahCos Distance	$g \times G \times V_F$	$g \times G \times V_P$	$g \times G \times V_F$	$g \times G \times V_P$	
	Total	1.03E+09	4.9E+08	9.6E+08	4.17E+08	
ICP		$M \times 1331 \times 3 \times 466 = 5.58E+11$		$M \times 1331 \times 6 \times 466 = 1.12E+12$		

<sup>(1)</sup> rotations in  $xz$  plane, <sup>(2)</sup> accounts for additions of  $M$  points based on comparisons of  $z$  values for MNSD, <sup>(3)</sup> 20, 20 and 32 rotations in  $xz$ ,  $yz$  and  $xy$  planes respectively, <sup>(4)</sup> 20 and 32 rotations in  $yz$  and  $xy$  planes. In this case, there is no  $xz$  plane rotation as profile face images are aligned in  $xz$  plane in nose tip detection process, Time complexity (Seconds): alignment=23.36, multi-view synthesis=1.92, MahCos distance=0.06, Unified classifier=0.13 Total=25.47 computed on a standard P4 (1.8GHz) computer with 8GB RAM.

## 4.5 Comparison to Existing Approaches

The FV and FI results of the proposed approach are compared to state-of-the-art earlier studies in Fig. 12. Referring to Fig. 12 for GavabDB database, the study [45] presented a sparse representation framework for FR using low level geometric features. The study [24] proposed a FR technique based on 3D keypoint extraction, and sparse comparison based similarity evaluation. The study [3] employed meshSIFT for detection of facial keypoints, and sparse representation for classification. The algorithm [27] encoded different types of features and modalities into a compact representation using covariance based descriptors and employed geodesic distance in the matching process. Finally, the study [28] presented a 3D face keypoint detection and matching approach based on principle curvatures where matching was performed using local shape descriptors, sparse representation based reconstruction method and score level fusion.

Referring to Fig. 12 for Bosphorus and UMB-DB databases, the study [10] reported results using facial depth information and ICP based algorithms. The methodology presented in the study [46] is based on local descriptor extraction, and matching according to differential surface measurements. The study [47] employed three keypoint descriptors corresponding to surface differential quantities, and multitask sparse representation based fine-grained matching algorithm. The technique presented in the study [28] is discussed in above paragraph, whereas, the study [48] proposed to implement a 3D deformable model to be fitted to unseen face images for FR.

Referring to Fig. 12 for FRGC v2.0 database, the methodology given in Ref. [22] proposed to integrate global and local geometric cues for FR using Euclidean distance based classification. The study [21] is focused on a graph based approach using iso-geodesic stripes and 3D weighted walkthrough (3DWW) descriptors. The study [23] proposed a resolution invariant approach using local features where scale space extrema based on shape index images are detected and classified using SVM classifier. Finally; a DT-CWT, LDA and NN based FR approach is implemented in the study [2], whereas, the 3D FR methods proposed in the studies [28,47-48] are discussed in above paragraphs of this section. The proposed, two-pass 3D alignment and recognition approach has yielded better recognition accuracies than the studies compared in Fig. 12. The ROC curves obtained by the comparative methods have also been compared with the proposed method in Fig. 13. The area under curve (AUC) computed for the comparative methods is 96.10%, 98.03%, 98.8%, 99.64% and 99.93% for Ref. [22], Ref. [21], Ref. [23], Ref. [2] and this study, respectively which also reflects superiority of our method.

### 4.6 Results Related Discussion

The proposed study addresses the problem of 3D alignment and multi-view 3D FR. The important findings and impact of model parameters are discussed as under:

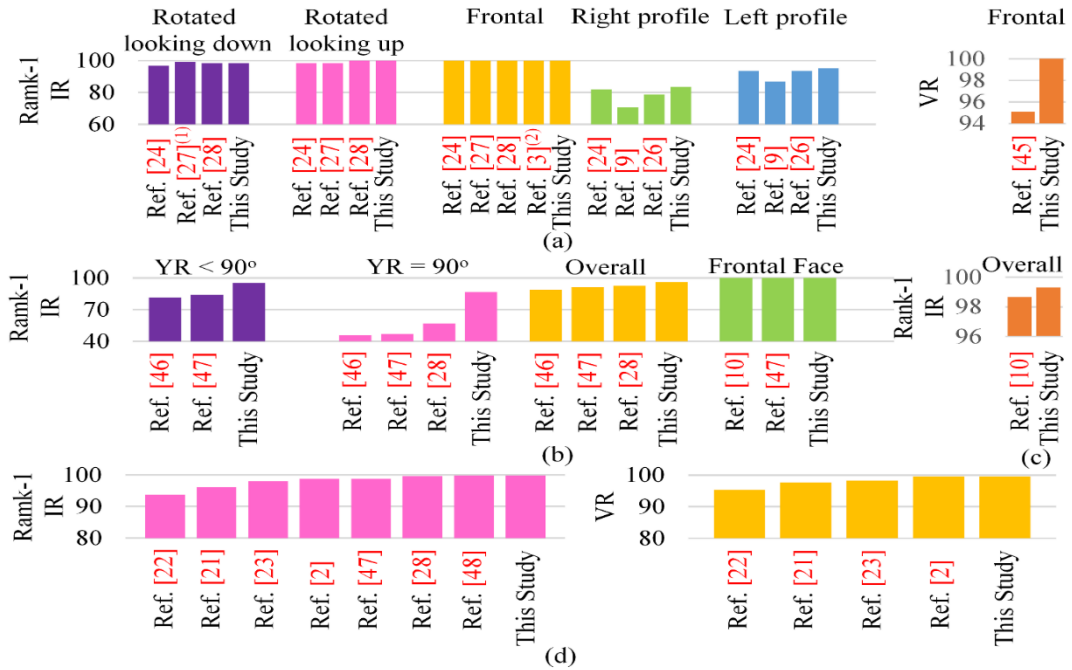
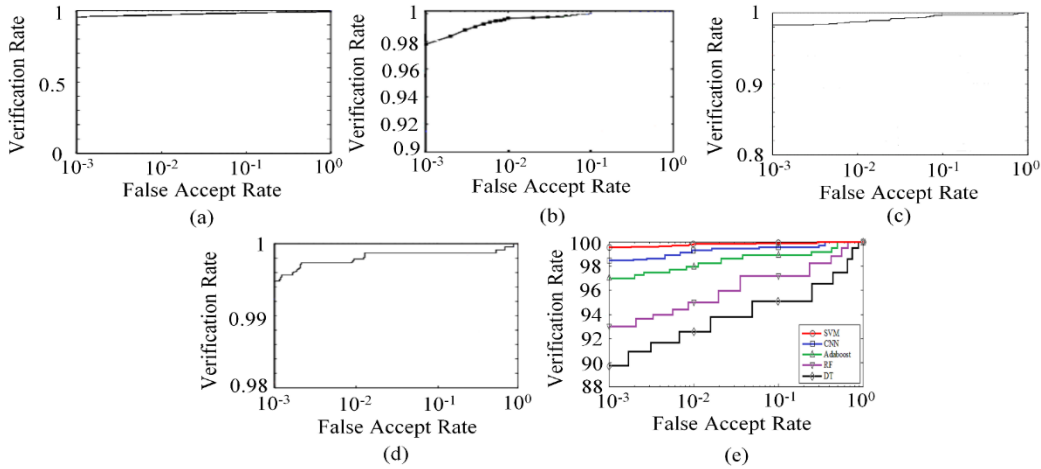


Fig. 12. Comparison of recognition rates using (a) GavabDB, (b) Bosphorus, (c) UMB-DB, (d) FRGC v2.0  
 (1) Typo error, IR can be 98.4% instead of 99.18% (2) In [3], gallery comprising of 183 subjects is employed.



**Fig. 13.** Comparison of ROC curves using FRGC v2.0 (a) Ref. [22], 95.37% (b) Ref. [21], 97.7% (c) Ref. [23], 98.3% (d) Ref. [2], 99.53% (e) This study 99.57%

(i) In the nose tip detection algorithm, it is observed that by increasing the value of threshold the profile face images are misclassified as frontal while by decreasing it frontal images are misclassified as profile face images. Therefore, face alignment and recognition accuracy is decreased in either case.

(ii) As discussed in Section 3.1.3 (ii), the nose-tip scanner distance parameter is reduced after face alignment. The impact of this parameter can be clearly observed in Fig. 4 (c), (d), (e) and Fig. 5 (c), (d), (e). The mentioned impact is validated by the quantitative results as shown in Fig. 8. The two-pass alignment algorithm successfully aligned the 3D face images from GavabDB and FRGC v2.0 databases achieving 99.82% and 99.95% alignment accuracies respectively while 100% accuracy was achieved for each of non-occluded subsets of Bosphorus and UMB-DB databases.

(iii) In order to analyze the impact of multi-view model parameters, it can be visualized in Fig. 1 that a face image oriented at  $0^\circ$  contains maximum feature information. As the face is gradually rotated from  $0^\circ$  to  $\pm 10^\circ$ ,  $\pm 20^\circ$  and  $\pm 30^\circ$  (with a step size of  $\pm 10^\circ$ ), the facial features start to occlude and FR accuracy would decrease intuitively. This intuition has been validated using the experimental protocol given in section 4.3.4 (ii) and the results are shown in Table 5. The results reveal that face images oriented at  $0^\circ$  produced best single FR accuracies which gradually decreased for face images rotated from  $0^\circ$  to  $\pm 30^\circ$ . This result trend remains same in all other similar experiments performed in this study. In order to determine the optimal step size of multi-view model parameters, an experiment was performed using experimental protocol given in Section 4.3.4 (ii), adjusting the step size to  $\pm 1^\circ$ . It was observed that the recognition accuracy remained same for a step size of  $\pm 1^\circ$  from existing orientation. This is because, a meager rotation of  $\pm 1^\circ$  does not change the facial discriminative feature information. It is further observed that a step size of at least  $\pm 10^\circ$  results into a change leading to an optimal final recognition accuracy. To validate our findings, we have shown the corresponding accuracies for a step size of  $\pm 5^\circ$  in Table 5.

**Table 5.** Rank-1 recognition rates using FRGC v2.0 (a) step size (b) recognition rates (%)

(a)	$0^\circ$	$+5^\circ$	$-5^\circ$	$+10^\circ$	$-10^\circ$	$+15^\circ$	$-15^\circ$	$+20^\circ$	$-20^\circ$	$+25^\circ$	$-25^\circ$	$+30^\circ$	$-30^\circ$
(b)	87.1	87.1	87.1	86.1	86.5	86.1	86.5	85.4	85.2	85.4	85.2	82.6	83.5

(iv) It is deduced that the SVM based FV algorithm outperformed the best single recognition accuracy producing a VR of 100% on GavabDB database, and 99.57% at 0.1% FAR using FRGC v2.0 database for the N vs. N experiments. SVM with min-max normalization performed better than RF, DT, Adaboost and CNN. This is because min-max normalization improves SVM performance by assigning higher weights to the features with greater separation between genuine and imposter scores. Secondly, SVM with min-max normalization is more flexible in terms of the size of training and test sets compared to other methods. (v) It is evident from experiments that the proposed FI algorithm clearly obtained improved recognition accuracies than the Single base classifier or parallel FR algorithms, whereas, weighted sum rule based classifier performed better than the majority voting based classifier. Our approach yielded rank-1 IR of 100% using the GavabDB database while 99.8% using FRGC v2.0 database for the N vs. N experiments, whereas, IR of 95.08%, 83.61%, 98.4% and 100% were obtained for left profile, right profile, rotated looking down and rotated looking up subsets of GavabDB database respectively. Similarly, IR of 100% and 99.3% were obtained using Bosphorus and UMB-DB databases respectively for N vs. N experiment. The rank-1 identification results for pose subsets of Bosphorus database are 95.1% (for  $YR < 90^\circ$ ), 86.7% (for  $YR = 90^\circ$ ) and 95.9% (overall). (vi) Experiments show that face alignment and recognition accuracies yielded by the proposed approach are comparable in all databases and that it obtained better performance for both of FV and FI experiments than those compared in [Fig. 12](#).

Despite of its excellent recognition accuracies, a limitation of the proposed two-pass alignment and FR approach is its incapability to align face images with missing nose region. This limitation may be overcome by synthesizing the missing nose region.

## 5. Conclusion

In this study, a fully automatic two-pass 3D face alignment and recognition algorithm is presented and its performance is demonstrated on GavabDB, Bosphorus, UMB-DB and FRGC v2.0 databases. Several novelties were contributed in this study including: (i) two-pass coarse to fine 3D face alignment algorithm, that accurately aligned 3D face images to an ICS in the first pass at a MNSD which is further minimized in the second alignment pass. (ii) SVM classifier based multi-view FV technique using scores computed between pairs of synthesized multi-view faces. (iii) three stage unified classifier based FI method that hierarchically combines results of seven base classifiers, two score and decision based parallel FR algorithms and an exponential rank combiner. Experimental results show the effectiveness of the proposed methodology to align and recognize frontal and profile face images, compared to the state-of-the-art methods. In the future, we intend to improve FRS performance by (i) synthesizing multi-view face images in yz plane (ii) employing local features alone, and a combination of holistic and local features and (iii) investigating impact of diverse ensemble classifiers.

## References

- [1] K. Bowyer, K. Chang and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1-15, 2006. [Article \(CrossRef Link\)](#)
- [2] X. Wang, Q. Ruan, Y. Jin and G. An, "Three-dimensional face recognition under expression variation," *EURASIP Journal on Image and Video Processing*, 2014:51, 2014.

[Article \(CrossRef Link\)](#)

- [3] L. Zhang, Z. Ding, H. Li, Y. Shen and J. Lu, “3D face recognition based on multiple keypoint descriptors and sparse representation,” *PLOS ONE*, vol. 9, no. 6: e100120, 2014.  
[Article \(CrossRef Link\)](#)
- [4] S. Elaiwat, M. Bennamoun, F. Boussaid and A. El-Sallam, “3-D face recognition using curvelet local features,” *IEEE Signal Processing Letters*, vol. 21, pp. 172–175, 2014.  
[Article \(CrossRef Link\)](#)
- [5] S. Soltanpour, B. Boufama and Q.J.Wu, “A survey of local feature methods for 3D face recognition,” *Pattern Recognition*, 2017. [Article \(CrossRef Link\)](#)
- [6] K. I. Chang, K. W. Bowyer and P. J. Flynn, “An evaluation of multimodal 2D+3D face biometrics,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 619-624, April 2005. [Article \(CrossRef Link\)](#)
- [7] C. BenAbdelkader and P. A. Griffin, “Comparing and combining depth and texture cues for face recognition,” *Image and Vision Computing*, vol. 23, no. 3, pp. 339–352, 2005.  
[Article \(CrossRef Link\)](#)
- [8] C. Heshner, A. Srivastava and G. Erlebacher, “A novel technique for face recognition using range imaging,” in *Proc. of International Symposium on Signal Processing and Its Applications*, vol. 2, pp. 201-204, Jul. 2003. [Article \(CrossRef Link\)](#)
- [9] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi and R. Slama, “3D face recognition under expressions, occlusions and pose variations,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2270–2283, 2013. [Article \(CrossRef Link\)](#)
- [10] N. Alyüz, B. Gökberk and L. Akarun, “3-D face recognition under occlusion using masked projection,” *IEEE Transactions on Information Forensics and Security*, vol. 8, pp. 789–802, 2013. [Article \(CrossRef Link\)](#)
- [11] D. Smeets, J. Keustermans, D. Vandermeulen and P. Suetens, “MeshSIFT: local surface features for 3D face recognition under expression variations and partial data,” *Computer Vision and Image Understanding*, vol. 117, no. 2, pp. 158–169, 2013. [Article \(CrossRef Link\)](#)
- [12] P. J. Besl and N. D. McKay, “A method for registration of 3-D shapes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.  
[Article \(CrossRef Link\)](#)
- [13] C. C. Queirolo, L. Silva, O. R. Bellon and M. P. Segundo, “3D face recognition using simulated annealing and the surface interpenetration measure,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 206–219, 2010. [Article \(CrossRef Link\)](#)
- [14] N. Alyüz, B. Gökberk and L. Akarun, “Regional registration for expression resistant 3-D face recognition,” *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 425-440, Sep. 2010. [Article \(CrossRef Link\)](#)
- [15] T. Papatheodorou and D. Rueckert, “3D face recognition,” *I-Tech Education and Publishing*, Vienna 2007. [Article \(CrossRef Link\)](#)
- [16] Y. Wang, J. Liu and X. Tang, “Robust 3D face recognition by local shape difference boosting,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 10, pp. 1858-1870, 2010. [Article \(CrossRef Link\)](#)
- [17] A. B. Moreno and A. Sanchez, “GavabDB: a 3D face database,” in *Proc. of Second COST Workshop on Biometrics on the Internet: Fundamentals, Advances and Applications*, pp. 77-82, 2004.
- [18] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiiktutan, B. Gokberk, B. Sankur and L. Akarun, “Bosphorus database for 3D face analysis,” *Biometrics and Identity Management*, pp. 47-56, 2008. [Article \(CrossRef Link\)](#)
- [19] A. Colombo, C. Cusano and R. Schettini, “UMB-DB: A database of partially occluded 3D faces,” in *Proc. of International Conference on Computer Vision*, pp. 2113–2119, 2011.



[Article \(CrossRef Link\)](#)

- [20] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek, "Overview of the face recognition grand challenge," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 947-954, 2005. [Article \(CrossRef Link\)](#)
- [21] S. Berretti, A. Del Bimbo and P. Pala, "3D face recognition using isogeodesic stripes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2162-2177, 2010. [Article \(CrossRef Link\)](#)
- [22] F. R. Al-Osaimi, M. Bennamoun and A. Mian, "Integration of local and global geometrical cues for 3D face recognition," *Pattern Recognition*, vol. 41, no. 3, pp. 1030-1040, 2007. [Article \(CrossRef Link\)](#)
- [23] G. Chang and Y. Wang, "Robust 3D face recognition based on resolution invariant features," *Pattern Recognition Letters*, vol. 32, no. 7, pp. 1009-1019, 2011. [Article \(CrossRef Link\)](#)
- [24] S. Berretti, A. Del Bimbo and P. Pala, "Sparse matching of salient facial curves for recognition of 3D faces with missing parts," *IEEE Transactions on Information Forensics and Security*, vol. 8, pp. 374-389, 2012. [Article \(CrossRef Link\)](#)
- [25] M. H. Mahoor and M. Abdel-Mottaleb, "Face recognition based on 3D ridge images obtained from range data," *Pattern Recognition*, vol. 42, no. 3, pp. 445-451, 2009. [Article \(CrossRef Link\)](#)
- [26] D. Huang, M. Ardabilian, Y. Wang and L. Chen, "3-D face recognition using eLBP-based facial description and local feature hybrid matching," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1551-1565, Oct. 2012. [Article \(CrossRef Link\)](#)
- [27] W. Hariri, H. Tabia, N. Farah, A. Benouareth and D. Declercq, "3D face recognition using covariance based descriptors," *Pattern Recognition Letters*, vol. 78, pp. 1-7, 2016. [Article \(CrossRef Link\)](#)
- [28] Y. Tang, H. Li, X. Sun, J. M. Morvan and L. Chen, "Principal curvature measures estimation and application to 3D face recognition," *Journal of Mathematical Imaging and Vision*, pp. 1-23, 2017. [Article \(CrossRef Link\)](#)
- [29] A. F. Abate, M. Nappi, D. Riccio and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1885-1906, 2007. [Article \(CrossRef Link\)](#)
- [30] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063-1074, 2003. [Article \(CrossRef Link\)](#)
- [31] T. Russ, C. Boehnen and T. Peters, "3D face recognition using 3D alignment for PCA," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1391-1398, 2006. [Article \(CrossRef Link\)](#)
- [32] X. Lu and A. K. Jain, "Automatic feature extraction for multiview 3D face recognition," in *Proc. of International Conference on Automatic Face and Gesture Recognition*, pp. 585-590, April, 2006. [Article \(CrossRef Link\)](#)
- [33] P. Kamencay, R. Hudec, M. Benco and M. Zachariasova, "2D-3D face recognition method based on a modified CCA-PCA algorithm," *International Journal of Advanced Robotic Systems*, vol. 11, no. 36, pp. 1-8, 2014. [Article \(CrossRef Link\)](#)
- [34] S. Zafeiriou, G. A. Atkinson, M. F. Hansen, W. A. P. Smith, V. Argyriou, M. Petrou, M. L. Smith and L. N. Smith, "Face recognition and verification using photometric stereo: The photoface database and a comprehensive evaluation," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 121-135, 2013. [Article \(CrossRef Link\)](#)
- [35] L. Spreeuwers, "Fast and accurate 3D face recognition using registration to an intrinsic coordinate system and fusion of multiple region classifiers," *International Journal of Computer Vision*, vol. 93, pp. 389-414, 2011. [Article \(CrossRef Link\)](#)

- [36] S. Jahanbin, R. Jahanbin and A. C. Bovik, "Passive three dimensional face recognition using iso-geodesic contours and procrustes analysis," *International Journal of Computer Vision*, vol. 105, pp. 87–108, 2013. [Article \(CrossRef Link\)](#)
- [37] X. Peng, M. Bennamoun and A. S. Mian, "A training-free nose tip detection method from face range images," *Pattern Recognition*, vol. 44, pp. 544–558, 2011. [Article \(CrossRef Link\)](#)
- [38] G. G. Slabaugh, "Computing Euler angles from a rotation matrix," Technical Report, 1999. Available at: [Article \(CrossRef Link\)](#) (Date of access: 07-April-2016).
- [39] U. I. Bajwa, I. A. Taj and M. W. Anwar, "A unified classifier for robust face recognition based on combining multiple subspace algorithms," *Optics Communications*, vol. 285, no. 21-22, pp. 4324-4332, Oct. 2012. [Article \(CrossRef Link\)](#)
- [40] A. A. Ross, A. K. Jain and K. Nandakumar, "Handbook of Multibiometrics (Levels of fusion in biometrics)," *Springer*, US, 2006. [Article \(CrossRef Link\)](#)
- [41] A. A. Ross, A. K. Jain and K. Nandakumar, "Handbook of Multibiometrics (Score level fusion)," *Springer*, US, 2006. [Article \(CrossRef Link\)](#)
- [42] U. I. Bajwa, I. A. Taj, M. W. Anwar and X. Wang, "A multifaceted independent performance analysis of facial subspace recognition algorithms," *PLOS ONE*, vol. 8, no. 2:e56510, 2013. [Article \(CrossRef Link\)](#)
- [43] O. Sagi, L. Rokach, "Ensemble learning: A survey," *Wires Data Mining and Knowledge Discovery*, e1249, 2018. [Article \(CrossRef Link\)](#)
- [44] A. S. Mian, M. Bennamoun and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1927-1943, 2007. [Article \(CrossRef Link\)](#)
- [45] X. Li, T. Jia and H. Zhang, "Expression-insensitive 3D face recognition using sparse representation," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2575–2582, 2009. [Article \(CrossRef Link\)](#)
- [46] S. Berretti, N. Werghi, A. Del Bimbo and P. Pala, "Matching 3D face scans using interest points and local histogram descriptors," *Computers and Graphics*, vol. 37, no. 5, pp. 509–525, 2013. [Article \(CrossRef Link\)](#)
- [47] H. Li, D. Huang, J.M. Morvan, Y. Wang and L. Chen, "Towards 3D face recognition in the real: a registration-free approach using fine-grained matching of 3D keypoint descriptors," *International Journal of Computer Vision*, vol. 113, no. 2, pp. 128–142, 2015. [Article \(CrossRef Link\)](#)
- [48] S.Z. Gilani, A. Mian and P. Eastwood, "Deep, dense and accurate 3D face correspondence for generating population specific deformable models," *Pattern Recognition*, vol. 69, pp. 238-250, 2017. [Article \(CrossRef Link\)](#)



**Dr. Naeem Ratyal** received Ph.D degree in Electrical Engineering from CUST, Pakistan in 2016. His research interests are 3D face recognition, image processing and computer vision.



**Dr. Imtiaz Taj** received Ph.D. degree from Hokkaido University, Japan in 2001. He has supervised four Ph.D. dissertations in the fields of computer vision and pattern recognition.



**Dr. Usama Bajwa** obtained Ph.D. degree in Computer Engineering from Center for Advanced Studies in Engineering, Pakistan in 2013. His research field is, image processing and computer vision.



**Dr. Muhammad Sajid** received his PhD degree in 2016 in Electrical Engineering from CUST, Pakistan. His main research interests include face recognition, image processing, and biometrics.