




## Research Article

# Deeply Learned Pose Invariant Image Analysis with Applications in 3D Face Recognition

**Naeem Ratyal** <sup>1,2</sup> **Imtiaz Ahmad Taj** <sup>2</sup>  
**Muhammad Sajid** <sup>1,2</sup> **Anzar Mahmood** <sup>1</sup>  
**Sohail Razzaq**<sup>3</sup> **Saadat Hanif Dar**<sup>4</sup> **Nouman Ali**<sup>4</sup> **Muhammad Usman**<sup>1</sup>  
**Mirza Jabbar Aziz Baig**<sup>1</sup> and **Usman Mussadiq** <sup>1</sup>

<sup>1</sup>Department of Electrical Engineering, Mirpur University of Science and Technology (MUST), Mirpur-10250, AJK, Pakistan

<sup>2</sup>Vision and Pattern Recognition Systems Research Group, Capital University of Science and Technology, Islamabad-45750, Pakistan

<sup>3</sup>Department of Electrical Engineering, COMSATS University Islamabad, Abbottabad Campus, Abbottabad-22060, Pakistan

<sup>4</sup>Department of Software Engineering, Mirpur University of Science and Technology (MUST), Mirpur-10250, AJK, Pakistan

Correspondence should be addressed to Muhammad Sajid; [sajid\\_faiz@hotmail.com](mailto:sajid_faiz@hotmail.com)

Received 19 February 2019; Accepted 27 May 2019; Published 17 June 2019

Academic Editor: Bogdan Smolka

Copyright © 2019 Naeem Ratyal et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Face recognition aims to establish the identity of a person based on facial characteristics and is a challenging problem due to complex nature of the facial manifold. A wide range of face recognition applications are based on classification techniques and a class label is assigned to the test image that belongs to the unknown class. In this paper, a pose invariant deeply learned multiview 3D face recognition approach is proposed and aims to address two problems: face alignment and face recognition through identification and verification setups. The proposed alignment algorithm is capable of handling frontal as well as profile face images. It employs a nose tip heuristic based pose learning approach to estimate acquisition pose of the face followed by coarse to fine nose tip alignment using  $L_2$  norm minimization. The whole face is then aligned through transformation using knowledge learned from nose tip alignment. Inspired by the intrinsic facial symmetry of the Left Half Face (LHF) and Right Half Face (RHF), Deeply learned (d) Multi-View Average Half Face (d-MVAHF) features are employed for face identification using deep convolutional neural network (dCNN). For face verification d-MVAHF-Support Vector Machine (d-MVAHF-SVM) approach is employed. The performance of the proposed methodology is demonstrated through extensive experiments performed on four databases: GavabDB, Bosphorus, UMB-DB, and FRGC v2.0. The results show that the proposed approach yields superior performance as compared to existing state-of-the-art methods.

## 1. Introduction

Face recognition is the ability of a biometric system to scan, store, and identify [1] human faces. It has become a fertile field for researchers due to various applications in computer vision and pattern recognition [2]. It has found numerous real-time applications in access control, security, surveillance, criminal identification, fraud detection, and even in human computer interaction [3, 4] based on machine learning algorithms [5]. The advantage of face recognition over other biometric modalities is noninvasive acquisition, social acceptance, and appropriateness for noncooperative scenarios [6]. The term

face identification is coined for matching a subject's face template with every face template in the gallery [7]. On the other hand, face verification aims to identify a face template against the claimed identity [8]. The challenging problems of face recognition include Pose, Expression, and Illumination (PIE) variations [9]. Despite the great success made over the past, robust face recognition using 2D intensity images is still a significant challenging problem in the presence of PIE variations [10]. Recently, with the rapid growth of 3D acquisition, imaging, visualization, and reconstruction [11, 12] techniques, 3D shape analysis, and processing such as surface registration and shape retrieval have been extensively studied. 3D shapes

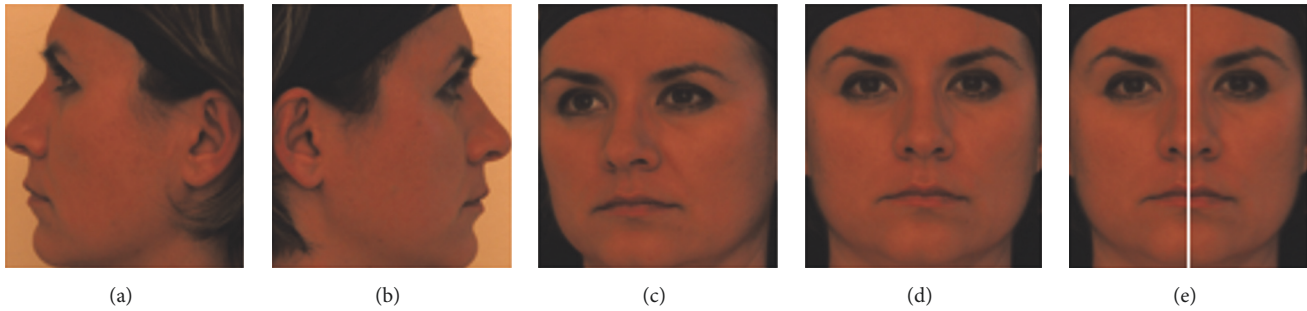


FIGURE 1: Subject Bosphorus: bs000, (a) LPF; (b) RPF; (c) FF; (d) face showing facial symmetry in (e) LHF and RHF.

make a wide array of new kinds of application possible, e.g., 3D biometrics, 3D medical imaging, 3D remote sensing, virtual reality, augmented reality, and 3D human-machine interaction [13]. As a special application of 3D shape analysis and processing, the key issue of 3D face recognition has also been widely addressed and identified to be much more robust to varying poses and illumination changes [14]. Therefore, pose invariant face recognition using 3D models is proving to be promising especially in case of in-depth pose variations along  $x$ -,  $y$ -, and  $z$ -axis under unconstrained acquisition scenarios of real world [15]. Encouraged by these lines of evidence, many 3D face recognition approaches have been evolved and experimented in the last few years as given in the work of Bowyer et al. [16] and the literature reviews [17–20].

The existing 3D face recognition approaches can be grouped into holistic, local feature-based, and hybrid domains [20]. Under the holistic paradigm, Principal Component Analysis (PCA) [21, 22], Linear Discriminant Analysis (LDA) [21, 23], and Independent Component Analysis (ICA) [24] are based on subspace learning. Local feature-based methods employ features from local descriptive points, curves, and regions. In point based methods, a curvelet transform based study is proposed in the research work [18] to find salient points on the face images for building multiscale local surfaces. Similarly, facial key points are detected exploiting meshSIFT algorithm in the point based methods [19, 25]. A prominent curve based method employing Riemannian framework is presented in the paper [26], whereas, the study [27] is a representative region based approach for occlusions and missing data handling problem. Hybrid approaches employ both of holistic and local feature-based methods [28] or combination of 2D and 3D images in the face recognition process [14].

The most crucial stage in any 3D face recognition algorithm is face alignment and the resulting accuracy primarily depends on the robustness of the alignment module [29]. In alignment phase, facial features are transformed such that they can be reliably matched. A few 3D alignment techniques [29–34] existing in literature are based on Iterative Closest Point (ICP) [30], Intrinsic Coordinate System (ICS) [31], Simulated Annealing (SA) [32], and Average Face Model (AFM) [29].

Solutions enabling recognition of subjects captured under arbitrary poses are now attracting an increasing interest. Profile image based face recognition [35] is an example of

such a case where left or right profile face images are used in the recognition process. Left profile face (LPF) images are defined as the face images where a subject presents his/her face rotated  $-90^\circ$  (please see Figure 1(a)), whereas, in right profile face (RPF) images the subject's face is rotated  $+90^\circ$  around the vertical axis in  $xz$  plane [36] (Figure 1(b)). Please note that frontal face (FF) images are captured with subjects facing towards camera at different angles as shown in Figure 1(c). The face alignment and recognition of LPF, RPF, and FF is a challenging problem in 2D intensity images. In the proposed study, our aim is to present a novel 3D face alignment and recognition algorithm to deal with FF, LPF, and RPF images. For face identification, Multi-View Whole Face (MVWF) images are synthesized to integrate real 3D facial feature information that boosts the face recognition accuracy. Motivated by the intrinsic symmetry of a face [37] (Figure 1(d)) exhibited by the LHF and RHF images (Figure 1(e)), Multi-View LHF (MVLHF) images and Multi-View RHF (MVRHF) images are combined into Multi-View Average Half Face (MVAHF) images. Subsequently, d-MVAHF features are extracted and employed in the face recognition process using dCNN. For a comparative evaluation, experimental results are also reported for d-MVWF, d-MVLHF, and d-MVRHF features on all four databases.

For performance evaluation of the proposed approach, four benchmark databases, namely, GavabDB [36], Bosphorus [38], UMB-DB [39], and FRGC v2.0 [40], have been used in this study. These databases carry pose and expression variations and are commonly utilized for developing 3D face recognition algorithms. For example, the algorithms presented in state-of-the-art studies [17, 41–43] employed FRGC v2.0 database while those presented in Refs. [19, 26, 28, 44–46] are based on both of FRGC v2.0 and GavabDB databases. Similarly, the studies [26, 28, 47] employed GavabDB, Bosphorus, and FRGC v2.0 databases while the study [27] used UMB-DB to evaluate the algorithm. The main contributions and novelty of the proposed algorithm are as follows:

- (1) The first contribution of this study is a novel 3D alignment algorithm that can deal with neutral and expressive FF, LPF, and RPF images in face recognition applications. The proposed algorithm differs from the conventional alignment approaches in two aspects: (i) It does not align two face images to each

other; rather, it is capable of aligning a standalone probe face image (PFI); and (ii) it employs a nose tip heuristic based pose learning approach. The pose learning approach first estimates acquisition pose of the PFI. Subsequently, an  $L_2$  norm minimization based coarse to fine alignment approach is employed that initially aligns the nose tip of the PFI. This is followed by a transformation step to align the whole facial surface in a single 3D rotation. The proposed algorithm is referred to as pose learning-based coarse to fine (PCF) alignment algorithm in the rest of the study.

- (2) The second contribution is a novel deeply learned approach for image analysis with applications in 3D face recognition. The proposed d-MVAHF-based face identification and d-MVAHF-SVM-based face verification approach employs face images oriented at  $0^\circ$ ,  $10^\circ$ ,  $20^\circ$ , and  $30^\circ$ . For a comparative evaluation, the proposed approach is tested using d-MVWF images, d-MVLHF images, and d-MVRHF images oriented at  $0^\circ$ ,  $\pm 10^\circ$ ,  $\pm 20^\circ$ , and  $\pm 30^\circ$ ;  $0^\circ$ ,  $-10^\circ$ ,  $-20^\circ$ , and  $-30^\circ$ ; and  $0^\circ$ ,  $10^\circ$ ,  $20^\circ$ , and  $30^\circ$ , respectively. The proposed algorithm is also validated using deeply learned multiview LPF (d-MVLPF) and deeply learned multiview RPF (d-MVRPF) images oriented at  $0^\circ$ ,  $-10^\circ$ ,  $-20^\circ$ , and  $-30^\circ$ ; and  $0^\circ$ ,  $10^\circ$ ,  $20^\circ$ , and  $30^\circ$ , respectively.
- (3) Third contribution is (i) the study of the role of pose learning and nose tip alignment in reducing the computational complexity of the PCF alignment algorithm for face recognition applications and (ii) the computational complexity analysis of the d-MVAHF-based face recognition compared to d-MVWF based face recognition for biometric applications.

Rest of the study is organized as follows: Related work is presented in Section 2. Section 3 deals with the details of proposed 3D alignment and face recognition algorithm. Experiments and results are given in Section 4, whereas the discussion and conclusions are presented in Sections 5 and 6, respectively.

## 2. Related Work

For a thorough survey of research in the area of 3D face recognition and applications, the reader is referred to the studies [16, 48]. The related work in the context of both components of the current study, i.e., 3D face alignment and 3D face recognition, is discussed separately.

*2.1. 3D Face Alignment Algorithms.* Review of the existing 3D face alignment algorithms, namely, ICP [30], ICS [31], SA [32], and AFM [29], is given as follows:

ICP [30, 34] based algorithm aligns two 3D faces by minimizing the distance between them iteratively. Limitations of the ICP include initial course alignment and slow convergence. The drawbacks of ICP technique limit its applicability to only verification setup where the PFI is to be aligned to the claimed identity image only [15]. They

become an issue in identification setup where a probe is to be aligned to the whole gallery. The second method, alignment to an ICS [31], mainly involves localization of landmarks on 3D facial images, comparison of landmarks to corresponding points on ICS, and a transformation phase to finish alignment. The downside of this method includes low accuracy in localization of landmarks, especially 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 identification as well as verification scenarios [15]. SA [32] algorithm employs a stochastic technique using a local search based approach and its drawback is excessive time consumption [33]. Similar to ICP, this method is suitable for verification setup only. In AFM [29] based alignment, the AFM is constructed by demarcating and averaging landmarks on the facial images; and the probe image is aligned to AFM only once. This aspect empowers this method to be used as an alignment technique in both of face identification and verification setups [15]. A significant disadvantage of the AFM based method is probe image's less accurate alignment to an AFM due to loss of spatial information involved in averaging process [15]. Another fast and effective face alignment method is proposed in the work of Wang et al. [34] to place each face model to a standard position and orientation. It does not align a probe image to every image in the gallery; therefore, it can be employed for both of face verification and identification efficiently. This alignment method is based on the facial symmetry plane which is determined using 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.

*2.2. 3D Face Recognition Algorithms.* Review of the 3D face recognition methods from the perspective of their role in developing multiview and fusion based face recognition algorithms is presented as follows.

The study [49] proposed to synthesize the various facial variations by utilizing a morphable model that augments the existing training set comprising a single frontal 2D image of each subject. The morphable model face is a vector space representation based parametric model of the faces. In the vector space, any convex combination of shape and texture vectors describes a human face. For a single face image, 3D shape, pose, texture, and illumination, etc., are automatically estimated by the algorithm. The recognition task is realized by measuring the Mahalanobis score between the fitting model and the shape and texture parameters of the models contained in the gallery. The authors performed identification experiments on two publicly available databases, namely, FERET and CMU-PIE, and achieved a recognition rate of 95.9% and 95%, respectively.

The study [50] proposed a 3D face recognition algorithm where PCA based 3D face synthesis approach is employed to generate new faces based on a reference face model. The approach preserves important 3D size information present in the input face and achieves better alignment of facial points using 3D scaling of the generic reference face. The algorithm

uses “one minus the cosine of the angles” among the PCA model parameters as the matching score. The experiments were performed using FRGC face database and 92-96% verification rates at 0.001% FAR and rank-1 identification accuracies between 94 and 95% were obtained.

The study [51] proposed a fully automatic face recognition system using multiview 2.5D facial images. The approach employs a feature extractor using the directional maximum, to find the nose tip and pose angle simultaneously. Face images are recognized using ICP based approach corresponding to best located nose tip. The experiments were performed on the MSU and the UND databases obtaining 96.2% and 97% identifications rates, respectively.

The study [34] proposed Collective Shape Difference Classifier (CSDC) based approach using a summed confidence as a similarity measure. The authors computed Signed Shape Difference Map (SSDM) between two aligned 3D faces as a mediate depiction for comparison of facial shapes. They used three types of features, to encode the characteristics and local similarity between them. They constructed three strong classifiers using most discriminative local facial features by boosting and training as weak classifiers. The experiments were carried out on FRGC v2.0 yielding verification rates better than 97.9% at 0.001% FAR and rank-1 recognition rates above 98%.

A study based on fusion of results acquired from several overlapping facial regions is proposed in the paper [15] employing decision level fusion (majority voting). PCA-LDA based method was used for extraction of features, whereas likelihood ratio was used as matching criterion to classify the individual regions. The author conducted experiments using FRGC v2.0 3D database to evaluate efficacy of the algorithm and reported 99% rank-1 recognition rate and 94.6% verification rate at 0.1% FAR, respectively.

Another fusion based study is given in the paper [52] equipped with an approach where match scores of each subject were combined for both of 2D albedo and depth images. Experimental results are reported by employing PCA, LDA, and Nonnegative Matrix Factorization (NMF) based subspace methods and Elastic Bunch Graph Matching (EBGM). Among the experiments, the best results were reported for sum rule based score level fusion. The authors achieved 89% recognition accuracy on a database of 261 subjects.

A recent region based study [27] proposed a method to handle occlusions covering the facial surface, employing two databases containing facial images with realistic occlusions. The authors addressed two problems, namely, missing data handling and occlusions, and improved the classification accuracy at score level using product rule. In the experiments, 100% classification results were obtained for neutral subsets, whereas, in the same study, pose, expression, and occlusion subsets achieved relatively low classification accuracies.

The study [53] proposed a facial recognition system (FRS) which employed fusion of three face classifiers using feature and match score level fusion methods. The features used by classifiers were extracted at facial contours around inner eye corners and the nose tip. The classification task was performed in LDA subspace by using Euclidean distance based

INN classifier. Experiments were performed on a coregistered 2D-3D image database acquired from 116 subjects, and rank-1 recognition rate of 99.09% was obtained by the authors.

A prominent algorithm based on fusion of 2D and 3D features is proposed in the study [54] which uses PCA, employing canonical correlation analysis (CCA) to learn mapping between a 2D image and its respective 3D scan. The algorithm is capable of classifying a probe image (whether it is 2D or 3D), by matching it to a gallery image, modeled by fusion of 2D and 3D modalities containing features from both sides. The authors performed experiments using a database of 115 subjects which contains neutral and expressive pairs of 2D images and 3D scans. They employed Euclidean distance classifier for the classification and obtained 55% classification accuracy using CCA algorithm alone. Their results were improved to 85% by using CCA-PCA algorithm.

The study [17] is a representative work of region based face recognition methods. The study proposed the use of facial representation based on dual-tree complex wavelet transform (DT-CWT) and six subregions. In this study NN classifier was employed in the classification stage and the authors achieved identification rate of 98.6% for neutral faces on the FRGC v2.0 database. Similarly, verification rate of 99.53% at 0.1% FAR was obtained for neutral faces on the same database.

A recent circular region based study [47] proposed an effective 3D face keypoint detection and matching framework using three principle curvature measures. The local shape of the face around each 3D keypoint was comprehensively described by histograms of the principal curvature measures. Similarity comparison between facial surfaces was established by matching local shape descriptors by sparse representation based reconstruction method and score level fusion. The evaluation of the algorithm was performed on GavabDB, FRGC v2.0, and Bosphorus databases obtaining 100% (neutral subset), 99.6% (neutral subset), and 98.6% (pose subset) recognition rates, respectively.

The proposed study is focused on aligning the PFI employing the PCF alignment algorithm. It targets to enhance classification accuracies using complementary information obtained from d-MVAHF-based features acquired from synthesized MVAHF images. The results obtained from our proposed methodology are better than the state-of-the-art studies [17, 19, 27, 41–44] in terms of all the evaluation criteria employed by these studies.

### 3. Materials and Methods

The proposed system consists of face alignment, identification, and verification components implemented through PCF alignment algorithm, d-MVAHF, and d-MVAHF-SVM-based methodologies, respectively. The following sections explain the proposed algorithm in detail.

*3.1. The Proposed PCF Alignment Algorithm.* An illustration of the PCF alignment algorithm is presented in Figure 2(a). It employs nose tip heuristic in the pose learning step and aligns the PFI in  $xz$ ,  $yz$ , and  $xy$  planes separately. The procedure to determine the nose tip is described in the following paragraphs.

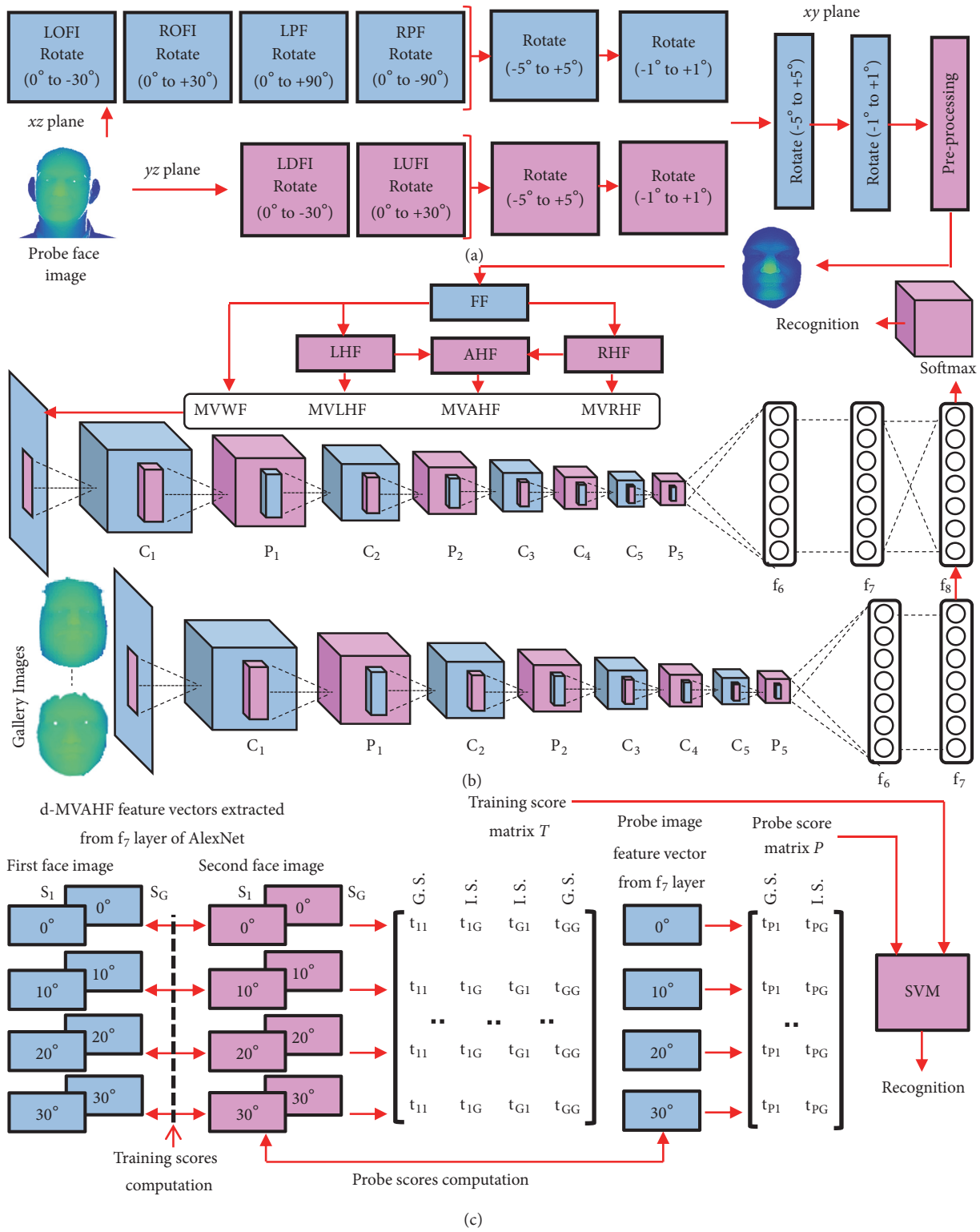


FIGURE 2: The proposed framework: (a) PCF alignment algorithm; (b) d-MVAHF-based face identification algorithm; (c) d-MVAHF-SVM-based face verification algorithm.

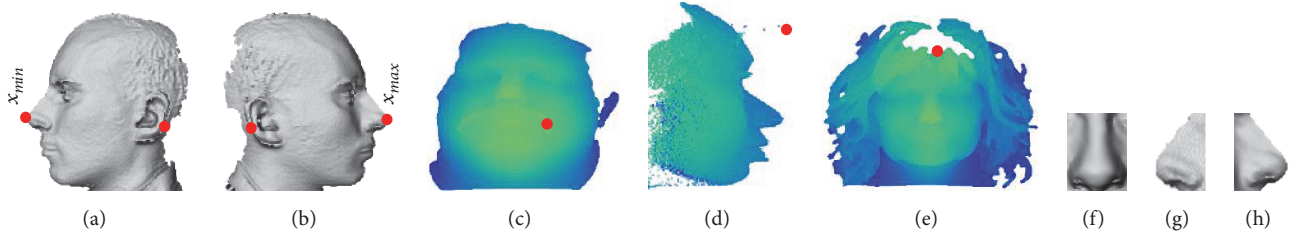


FIGURE 3: Examples of incorrectly detected nose tips on (a, b) ears; (c) lips area; (d) z-axis noise; (e) forehead hairs. Nose templates: (f) frontal; (g) left; (h) right.

**3.1.1. Nose Tip Detection Technique.** Nose tip detection is a specific facial feature detection problem in depth images. The study [55] proposed a nose tip detection technique for FF images based on histogram initialization and triangle fitting and obtained a detection rate of 99.43% on FRGC v2.0 database. In contrast to the study [55], the proposed study marks the nose tip as the nearest captured point from 3D scanner to the face and is used to localize, align, and crop the PFI. Several problems were faced in detecting the nose tip as follows:

One of the problems was incorrect nose tip detection in LPF or RPF images where it was detected on ears or some other facial parts as shown on ear of RPF of subject GavabDB: cara26\_derecha and ear of LPF of subject GavabDB: cara26\_izquierda in Figures 3(a) and 3(b), respectively. In order to handle this problem, the PFI was first classified as FF, LPF, or RPF using a convolutional neural network (CNN), and then nose tip was detected employing three different strategies for each of FF, LPF, or RPF. The CNN was trained for a three-class problem for FF, LPF, or RPF classification task. The PFI was used as input to the CNN which produced an  $N$  dimensional vector as the output where  $N$  is the number of classes. The CNN architecture was comprised of two convolutional layers followed by batch normalization and max pooling stages. The CNN also included two fully connected layers at the end. The first one contained 1024 units while the second fully connected layer with three units performed as the output layer with the softmax function. The architecture of the CNN for a PFI is shown in Figure 4. The CNN classifies the PFI of size  $h \times w$  as FF, LPF, or RPF using the final feature vector  $S = \{S_1^p, S_2^p, \dots, S_{h^p, w^p}^p\}$  computed for the layer  $p$ . Based on the classification of the PFI, the nose tip is determined as follows:

- (1) For FF images, the facial point at the minimum distance from the 3D scanner along  $z$ -axis is marked as the nose tip.
- (2) For LPF, the facial point having the minimum coordinate value along  $x$ -axis ( $x_{min}$ ) is defined as the nose tip.
- (3) For RPF, the facial point having the maximum coordinate value along  $x$ -axis ( $x_{max}$ ) is marked as the nose tip.

Another problem of the nose tip detection process was incorrect detection of the nose tip in those subjects which were captured with leaning forward or backward faces. In the leaning forward faces, the nose tip was detected on forehead, whereas, in leaning backward faces, it was detected on chin or lips area (See Figure 3(c) for subject FRGC v2.0: 04233d510). Similarly, noise scenarios played an adverse role in detecting the nose tip. For example, in some of the face images the  $z$ -axis noise occurring in the face acquisition process was marked as the nose tip as shown in Figure 3(d) for the subject FRGC v2.0: 04217d461. Another such scenario was regarding female subjects where hairs on forehead or spread around their neck or ears were marked as the nose tip as shown in Figure 3(e) for the subject FRGC v2.0: 04470d297.

Such problems were handled by searching the nose tip in an approximate Region of Interest (ROI). The ROI on the already classified FF, LPF, or RPF images was determined by measuring two features: (i) maximum value of depth map histogram and (ii) maximum value of the correlation coefficient of Normalized Cross Correlation (NCC). The former feature was measured using  $z$ ,  $-x$ , and  $x$  depth map histograms for each of the FF, LPF, or RPF in the respective order, whereas the latter was measured by correlating the corresponding frontal, left or right oriented nose templates (please see Figures 3(f), 3(g), and 3(h) for subject GavabDB: cara26\_frontal2, izquierda, and derecha, respectively) with FF, LPF, or RPF images. The nose templates were randomly selected from ten randomly chosen five male and female subjects each from the GavabDB database on satisfactory experimental results. For measuring the depth map histograms and correlation coefficient values, the PFI was rotated between  $40^\circ$  and  $-40^\circ$  with a step size of  $-40^\circ$  around  $x$ -axis adjusting the  $y$ -axis orientation at  $40^\circ$  to  $-40^\circ$  with the same step size resulting into nine facial orientations. The intuition behind this strategy is to search an upright position of the face because for such a position maximum number of depth values accumulate into a single bin of the depth map histogram and the correlation coefficient of the NCC returns maximum value among all nine facial positions. Consequently, the nose tip was correctly detected as the nearest captured point from 3D scanner to the face using an approximate ROI.

The proposed algorithm correctly detected the nose tips of face images from GavabDB, Bosphorus, UMB-DB, and FRGC v2.0 databases including all those cases where the nose

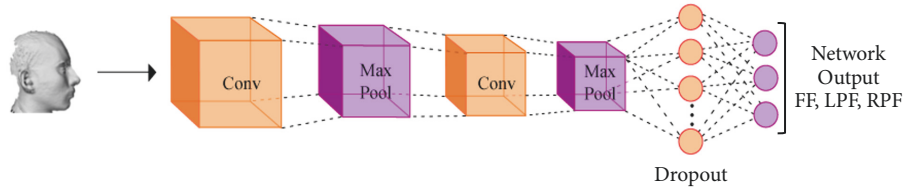


FIGURE 4: Illustration of CNN for FF, LPE, and RPF classification task.

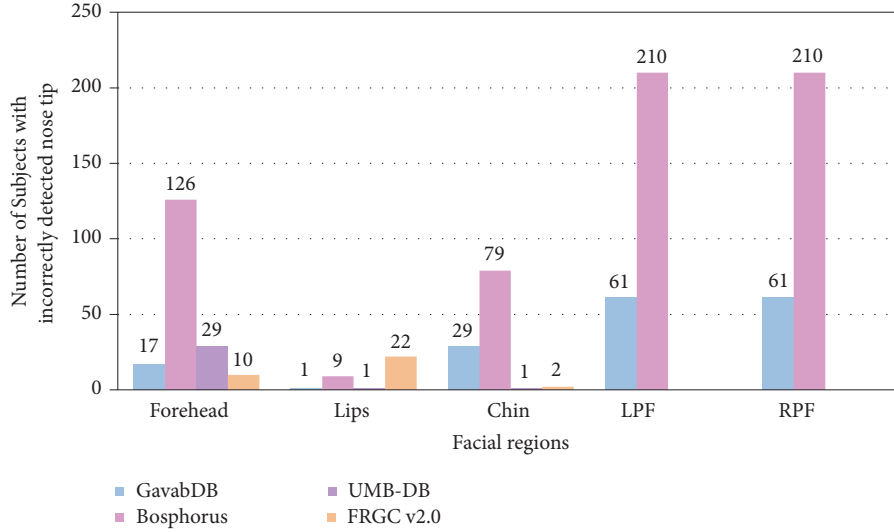


FIGURE 5: Incorrectly detected nose tips without employing the proposed nose tip detection technique.

tip was incorrectly detected at forehead, lips, chin, LPE, or RPF as detailed in Figure 5.

**3.1.2. Face Alignment Algorithm.** It was mentioned in the start of this section that the PCF alignment algorithm aligns the PFI in  $xz$ ,  $yz$ , and  $xy$  planes separately. The alignment in  $xz$  and  $yz$  planes employs  $L_2$  norm minimization calculated between the nose tip and the 3D scanner. The alignment in  $xy$  plane employs a different strategy based on  $L_2$  norm minimization calculated between the LHF image and flipped RHF image.

In order to explain the PCF alignment algorithm in  $xz$  and  $yz$  planes, the PFI is shown in Figure 6 with three nose tip positions 1, 2, and 3 in both planes separately. Intuitively, it can be observed in Figure 6 that the face image is aligned when the nose tip is set in line with the optic axis of the 3D scanner at position 1. Conversely, when it is not in line with the optic axis of the 3D scanner at position 2 or 3, the face image is not aligned. It can be observed in Figure 6 that  $L_2$  norm at nose tip position 1 is a perpendicular from the nose tip to the 3D scanner, which is not the case at nose tip positions 2 and 3. The perpendicular distance from a point on a line is always the shortest which leads to the conclusion that when PFI is aligned at position 1, the  $L_2$  norm is computed as the minimum and shorter than the corresponding values of  $L_2$  norms at positions 2 and 3. Therefore, alignment of the PFI causes an essential reduction in the  $L_2$  norm computed between the nose tip and the 3D scanner. The  $L_2$  norm

between nose tip position 1 ( $N(m_1, n_1)$ ) and the 3D scanner point  $S(m_0, n_0)$  is calculated as given in equation (1).

$$\|d\|_2 = \sqrt{(m_1 - m_0)^2 + (n_1 - n_0)^2} \quad (1)$$

### 3.1.3. Alignment in $xz$ Plane

**(1) Pose Learning.** First of all, the capture pose of the probe face image is learned to determine whether to rotate it clockwise or anticlockwise to align it at minimum  $L_2$  norm. For this purpose, only the nose tip of the probe face image is rotated clockwise at  $-1^\circ$  and corresponding  $L_2$  norm is measured between nose tip and 3D scanner. For example, a nose tip oriented at  $-1^\circ$  or  $30^\circ$  is rotated clockwise at  $-2^\circ$  or  $29^\circ$ , respectively, to measure the  $L_2$  norm. It is notable that a negative angle of rotation (e.g.,  $-2^\circ$ ) turns a probe face image (Figure 7(a)) clockwise in  $xz$  and  $yz$  planes and anticlockwise in  $xy$  plane as shown in Figures 7(b)–7(d).

As a result of clockwise rotation, if  $L_2$  norm is decreased (Figure 8(a)), the probe face image is classified as left oriented face image (LOFI) (Figure 8(c)). Similarly, if  $L_2$  norm is increased (Figure 8(b)), the probe face image is classified as right oriented face image (ROFI) as shown in Figure 8(d). Please note that, rotating the nose tip at  $1^\circ$  instead of  $-1^\circ$ , a decrease in  $L_2$  norm classifies the probe face image as ROFI, whereas an increase in  $L_2$  norm classifies it as LOFI. In this study we adjust this parameter at  $-1^\circ$ .

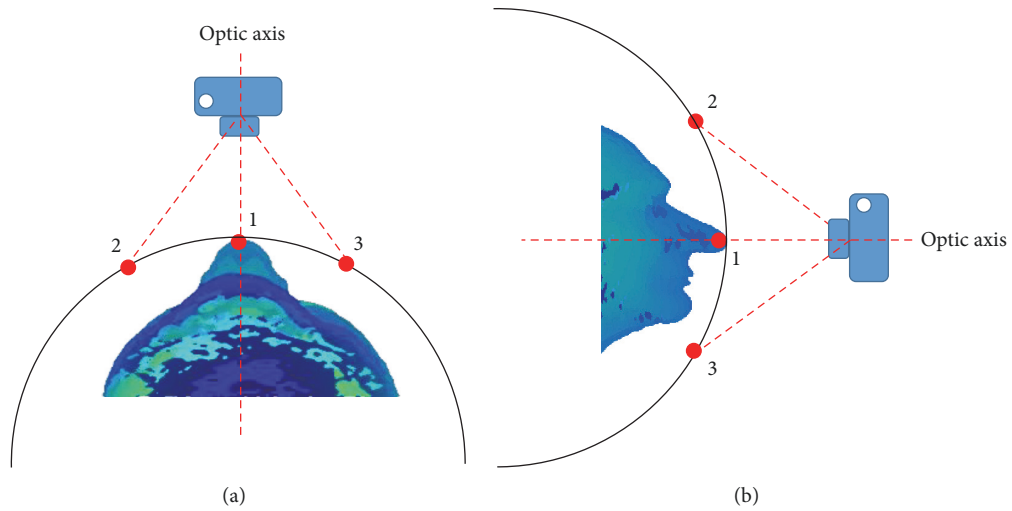


FIGURE 6: PCF alignment algorithm showing an aligned image at minimum  $L_2$  norm in (a)  $xz$ ; (b)  $yz$  plane.

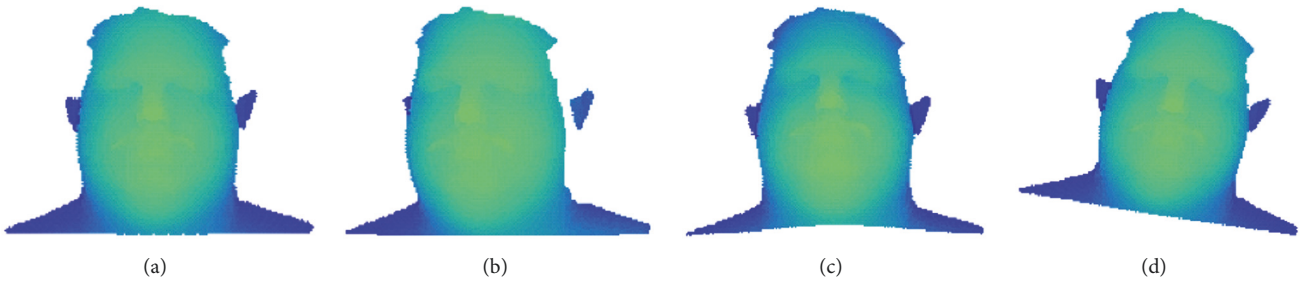


FIGURE 7: (a) 3D scan of subject FRGC v2.0:04233d396 rotated in (b)  $xz$ ; (c)  $yz$ ; (d)  $xy$  plane at  $-2^\circ$ .

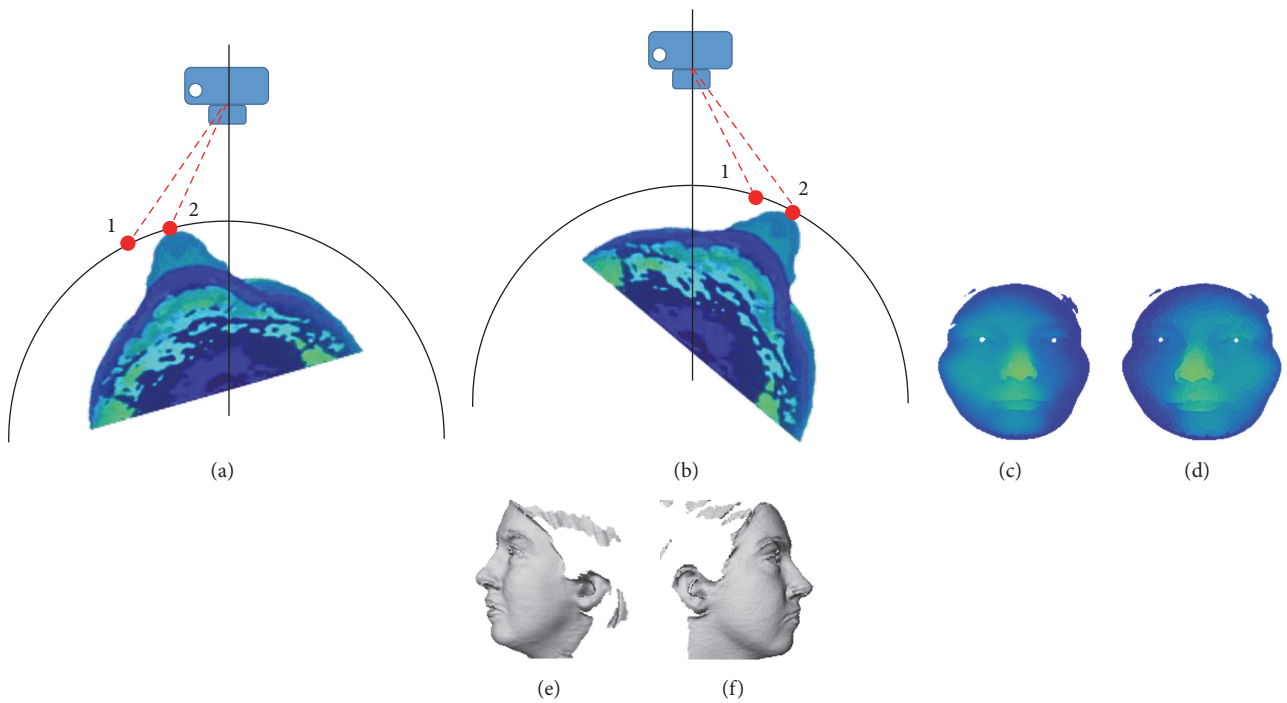


FIGURE 8: (a, b) Pose learning in  $xz$  plane; (c) LOFI; (d) ROFI; (e) LPF; (f) RPF. (a, b, c, d) Subject FRGC v2.0: 04221d553; (e, f) subject GavabDB: caral\_izquierda, derecha.



## (2) Coarse Alignment

- (i) *LOFI*: based on the outcome of the above step, the nose tip of a LOFI is rotated in the range of  $0^\circ$  to  $-30^\circ$  (clockwise) with a step size of  $-10^\circ$  and corresponding  $L_2$  norms are recorded. For example, if a LOFI is captured at an orientation of  $30^\circ$ , the nose tip is rotated between  $(30^\circ + 0^\circ = 30^\circ)$  and  $(30^\circ + (-30^\circ) = 0^\circ)$ . Similarly, the nose tip of a LOFI captured at an orientation of  $1^\circ$  is rotated between  $(1^\circ + 0^\circ = 1^\circ)$  and  $(1^\circ + (-30^\circ) = -29^\circ)$ . In both cases the nose tip is aligned at  $0^\circ$  corresponding to minimum  $L_2$  norm. However, the nose tips of the LOFI captured at  $29^\circ, 28^\circ, 27^\circ, 26^\circ, 25^\circ, 24^\circ, 23^\circ, 22^\circ,$  and  $21^\circ$  do not pass through the  $0^\circ$  position; therefore, they are aligned at  $-1^\circ, -2^\circ, -3^\circ, -4^\circ, -5^\circ, +5^\circ, +4^\circ, +3^\circ, +2^\circ,$  and  $+1^\circ$ , respectively (please see Table 1), and are aligned in step 3 at fine level.
- (ii) *ROFI*: the nose tip of a ROFI is rotated in the range of  $0^\circ$  to  $+30^\circ$  (anticlockwise) with a step size of  $10^\circ$  and corresponding  $L_2$  norms are recorded. For a ROFI captured at an orientation of  $-30^\circ$  or  $-1^\circ$ , the nose tip is rotated between  $(-30^\circ + 0^\circ = -30^\circ)$  to  $(-30^\circ + 30^\circ = 0^\circ)$  and  $(-1^\circ + 0^\circ = -1^\circ)$  to  $(-1^\circ + 30^\circ = 29^\circ)$ , respectively. The nose tip is aligned at  $0^\circ$  corresponding to minimum  $L_2$  norm in both of the cases. However, the nose tips of the ROFI captured at  $-29^\circ, -28^\circ, -27^\circ, -26^\circ, -25^\circ, -24^\circ, -23^\circ, -22^\circ,$  and  $-21^\circ$  are aligned at  $1^\circ, 2^\circ, 3^\circ, 4^\circ, 5^\circ, -5^\circ, -4^\circ, -3^\circ, -2^\circ,$  and  $-1^\circ$ , respectively (please see Table 1), and are aligned in step 3 at fine level.
- (iii) *LPF*: the nose tip of an LPF (Figure 8(e)) is rotated in the range of  $0^\circ$  to  $+90^\circ$  (anticlockwise) with a step size of  $10^\circ$  and corresponding  $L_2$  norms are recorded. For an LPF captured at an orientation of  $-90^\circ$ , the nose tip is rotated between  $(-90^\circ + 0^\circ = -90^\circ)$  and  $(-90^\circ + 90^\circ = 0^\circ)$  and is aligned at  $0^\circ$  corresponding to minimum  $L_2$  norm. However, the nose tips of the LPF captured at  $-89^\circ, -88^\circ, -87^\circ, -86^\circ, -85^\circ, -84^\circ, -83^\circ, -82^\circ,$  and  $-81^\circ$  are aligned at  $1^\circ, 2^\circ, 3^\circ, 4^\circ, 5^\circ, -5^\circ, -4^\circ, -3^\circ, -2^\circ,$  and  $-1^\circ$ , respectively (please see Table 1), and are aligned in step 3 at fine level.
- (iv) *RPF*: the nose tip of a RPF (Figure 8(f)) is rotated in the range of  $0^\circ$  to  $-90^\circ$  (clockwise) with a step size of  $-10^\circ$  and corresponding  $L_2$  norms are recorded. If a RPF is captured at an orientation of  $90^\circ$ , the nose tip is rotated between  $(90^\circ + 0^\circ = 90^\circ)$  and  $(90^\circ + (-90^\circ) = 0^\circ)$  and is aligned at  $0^\circ$  corresponding to minimum  $L_2$  norm. However, the nose tips of the RPF captured at  $89^\circ, 88^\circ, 87^\circ, 86^\circ, 85^\circ, 84^\circ, 83^\circ, 82^\circ,$  and  $81^\circ$  are aligned at  $-1^\circ, -2^\circ, -3^\circ, -4^\circ, -5^\circ, +5^\circ, +4^\circ, +3^\circ, +2^\circ,$  and  $+1^\circ$ , respectively (please see Table 1), and are aligned in step 3 at fine level.

Please note that, for a ROFI captured at  $-25^\circ$ , a LOFI captured at  $25^\circ$ , an LPF captured at  $-85^\circ$ , or a RPF captured at  $85^\circ$ , the nose tip can get aligned at  $5^\circ$  or  $-5^\circ$  because minimum  $L_2$  norm is equal at both orientations. However, we have aligned the nose tip at  $5^\circ$  in this study. The face images captured at  $\pm 75^\circ, \pm 65^\circ, \dots, \pm 5^\circ$  are aligned using the same alignment procedure.

(3) *Fine Alignment*. The nose tip of the LOFI, ROFI, LPF, and RPF is rotated in the range of  $-5^\circ$  to  $5^\circ$  with a step size of  $1^\circ$ . This means that nose tip aligned at  $-5^\circ$  is rotated between  $((-5^\circ) + (-5^\circ) = -10^\circ)$  and  $((-5^\circ) + (5^\circ) = 0^\circ)$  to catch the  $0^\circ$  position. On the other hand, the nose tip aligned at  $5^\circ$  is rotated between  $((5^\circ) + (-5^\circ) = 0^\circ)$  and  $((5^\circ) + (5^\circ) = 10^\circ)$  to catch the  $0^\circ$  position. After aligning the nose tip at  $0^\circ$ , it is rotated in the range of  $-1^\circ$  to  $1^\circ$  with a step size of  $0.1^\circ$  to achieve an accurate final alignment at a minimum  $L_2$  norm. Finally, the whole probe face image is rotated and aligned at an angle corresponding to the alignment of the nose tip. i.e., if the nose tip is aligned at  $13^\circ$ , then the whole face image is rotated at  $13^\circ$  and is finally aligned in  $xz$  plane.

3.1.4. Alignment in  $yz$  Plane

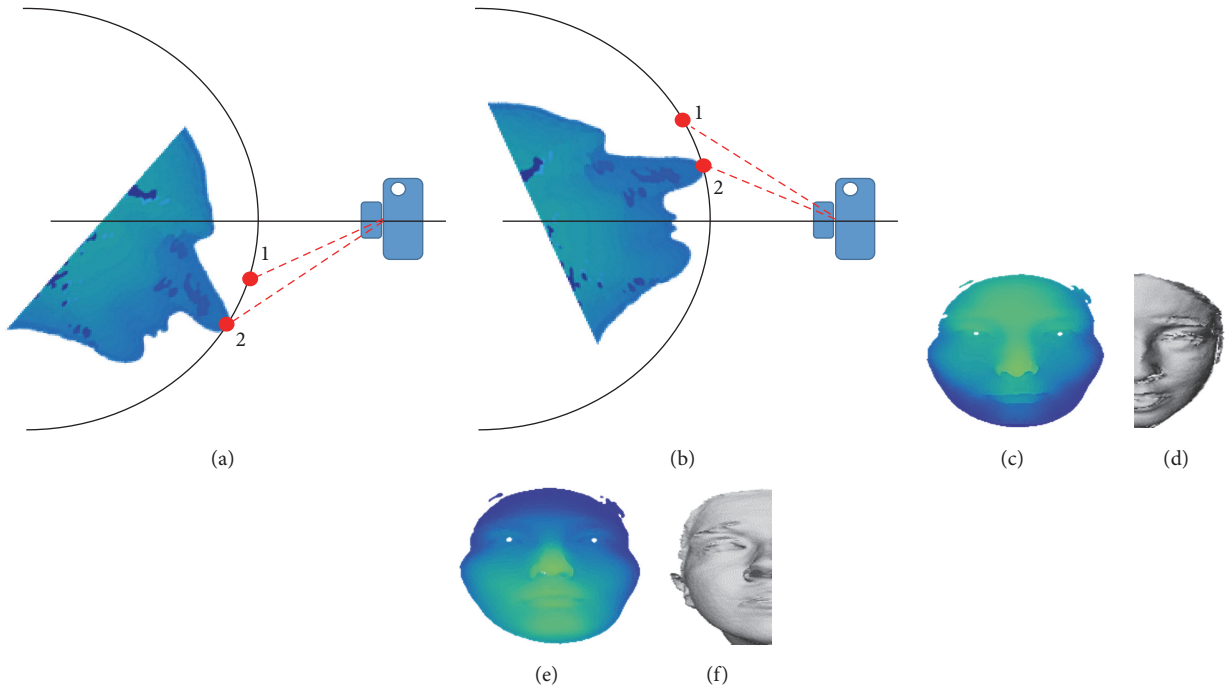
(1) *Pose Learning*. In  $yz$  plane, the capture pose of the probe face image aligned in  $xz$  plane is learned at first to align it at a minimum  $L_2$  norm. For this purpose, only nose tip of the probe face image is rotated upwards (clockwise) at  $-1^\circ$  and corresponding  $L_2$  norm is measured. If  $L_2$  norm is decreased (Figure 9(a)), the probe face image is classified as looking down face image (LDFI) (Figures 9(c) and 9(d)). On the other hand, if  $L_2$  norm is increased (Figure 9(b)), it is classified as looking up face image (LUFU) as shown in Figures 9(e) and 9(f). Please note that, rotating the nose tip at  $1^\circ$  instead of  $-1^\circ$ , a decrease in  $L_2$  norm classifies a probe face image as LUFU, whereas an increase in  $L_2$  norm classifies it as LDFI. In this study we adjust this parameter at  $-1^\circ$ .

## (2) Coarse Alignment

- (i) *LUFU*: in coarse alignment phase, the nose tip of a LUFU is rotated in the range of  $0^\circ$  to  $+30^\circ$  downwards (anticlockwise) with a step size of  $10^\circ$  and corresponding  $L_2$  norms are recorded. If a LUFU is captured at an orientation of  $-30^\circ$ , the nose tip is rotated between  $-30^\circ$  and  $0^\circ$ . If a LUFU is captured at an orientation of  $-1^\circ$ , the nose tip is rotated between  $-1^\circ$  and  $29^\circ$ . In both cases, the nose tip is aligned at  $0^\circ$  corresponding to minimum  $L_2$  norm. However, the nose tips of the LUFU captured at  $-29^\circ, -28^\circ, -27^\circ, -26^\circ, -25^\circ, -24^\circ, -23^\circ, -22^\circ,$  and  $-21^\circ$  do not pass through  $0^\circ$  position. They are aligned at  $1^\circ, 2^\circ, 3^\circ, 4^\circ, 5^\circ, -5^\circ, -4^\circ, -3^\circ, -2^\circ,$  and  $-1^\circ$  respectively (please see Table 1), and are aligned in step 3 at fine level.
- (ii) *LDFI*: the nose tip of a LDFI is rotated in the range of  $0^\circ$  to  $-30^\circ$  upwards (clockwise) with a step size of  $-10^\circ$  and corresponding  $L_2$  norms are recorded. For a LDFI captured at an orientation of  $30^\circ$  or  $1^\circ$ , the nose tip is rotated between  $30^\circ$  to  $0^\circ$  and  $1^\circ$  to  $-29^\circ$ , respectively. The nose tip is aligned at  $0^\circ$  corresponding to minimum  $L_2$  norm in both of the cases. However, the nose tips of the LDFI captured at  $29^\circ, 28^\circ, 27^\circ, 26^\circ, 25^\circ, 24^\circ, 23^\circ, 22^\circ,$  and  $21^\circ$  are aligned at  $-1^\circ, -2^\circ, -3^\circ, -4^\circ, -5^\circ, +5^\circ, +4^\circ, +3^\circ, +2^\circ,$  and  $+1^\circ$ , respectively (please see Table 1), and are aligned in step 3 at fine level. It is worth mentioning that the face images captured at  $\pm 25^\circ, \pm 15^\circ, \dots, \pm 5^\circ$  are handled

TABLE 1: Acquisition pose of the face and respective alignment positions given in bold case (all values in degrees).

LPF										RPF									
LOFI/LDFI										ROFI/LUFI									
90	80	70	60	50	40	30	20	10	<b>0</b>	-90	-80	-70	-60	-50	-40	-30	-20	-10	<b>0</b>
89	79	69	59	49	39	29	19	9	<b>-1</b>	-89	-79	-69	-59	-49	-39	-29	-19	-9	<b>1</b>
88	78	68	58	48	38	28	18	8	<b>-2</b>	-88	-78	-68	-58	-48	-38	-28	-18	-8	<b>2</b>
87	77	67	57	47	37	27	17	7	<b>-3</b>	-87	-77	-67	-57	-47	-37	-27	-17	-7	<b>3</b>
86	76	66	56	46	36	26	16	6	<b>-4</b>	-86	-76	-66	-56	-46	-36	-26	-16	-6	<b>4</b>
85	75	65	55	45	35	25	15	<b>5</b>	<b>-5</b>	-85	-75	-65	-55	-45	-35	-25	-15	-5	<b>5</b>
84	74	64	54	44	34	24	14	<b>4</b>	<b>-6</b>	-84	-74	-64	-54	-44	-34	-24	-14	-4	<b>6</b>
83	73	63	53	43	33	23	13	<b>3</b>	<b>-7</b>	-83	-73	-63	-53	-43	-33	-23	-13	-3	<b>7</b>
82	72	62	52	42	32	22	12	<b>2</b>	<b>-8</b>	-82	-72	-62	-52	-42	-32	-22	-12	-2	<b>8</b>
81	71	61	51	41	31	21	11	<b>1</b>	<b>-9</b>	-81	-71	-61	-51	-41	-31	-21	-11	-1	<b>9</b>

FIGURE 9: (a, b) Pose learning in  $yz$  plane; (c, d) LDFI; (e, f) LUFI. (a, b, c, e) Subject FRGC v2.0: 04221d553; (d, f) subject GavabDB: cara1\_izquierda, derecha.

using the alignment procedure mentioned in coarse alignment phase of  $xz$  plane.

(3) *Fine Alignment.* The nose tip of LUFI or LDFI is rotated in the range of  $-5^\circ$  to  $5^\circ$  with a step size of  $1^\circ$  to catch the  $0^\circ$  position as discussed in fine alignment phase of  $xz$  plane. Similarly, in order to align the nose tip at fine level, it is rotated in the range of  $-1^\circ$  to  $1^\circ$  with a step size of  $0.1^\circ$  to achieve an accurate final alignment at a minimum  $L_2$  norm. In the end, whole probe face image is rotated at an angle corresponding

to the alignment of the nose tip and is finally aligned in  $yz$  plane.

### 3.1.5. Alignment in $xy$ Plane

(1) *Coarse Alignment.* The PFI is rotated in the range of  $-5^\circ$  to  $+5^\circ$  with a step size of  $1^\circ$  around  $z$ -axis. For each rotation, it is cropped into LHF and RHF images using the nose tip heuristic. The flipped RHF image is shifted along LHF image in  $xy$  plane and corresponding  $L_2$  norm is computed

for each rotation at pixel values of the same grid position  $P_{i,j}$ . In order to rule out the outliers due to  $z$ -axis noise, pixel values less than a threshold  $T$  are considered in the  $L_2$  norm computation as given in equation (2). The face image is coarsely aligned at an angle corresponding to the minimum value of  $L_2$  norm which represents a good match.

$$P_{i,j} = \begin{cases} 0 & P_{i,j} > T \\ P_{i,j} & \text{otherwise} \end{cases} \quad (2)$$

(2) *Fine Alignment.* The face image is aligned at fine level by rotating it in the range of  $-1^\circ$  to  $+1^\circ$  with a step size of  $0.1^\circ$  using the procedure described above. The LPF and RPF which come up as LHF and RHF images after alignment in  $xz$  and  $yz$  planes (see Figures 9(d) and 9(f)) are aligned in  $xy$  plane in a similar fashion.

3.2. *d-MVAHF-Based 3D Face Recognition.* For face recognition, the depth images were preprocessed to deal with noise and gap based artifacts. The sharp spikes presented in depth face images due to the face capture process were removed using median filtering. Finally, the facial holes were filled employing interpolation and facial irregularities were smoothed through low pass filtering at the end. The aligned whole face images were then rotated at  $0^\circ$ ,  $\pm 10^\circ$ ,  $\pm 20^\circ$ , and  $\pm 30^\circ$  to synthesize MVWF images. Similarly, LHF and RHF images were rotated at  $0^\circ$ ,  $-10^\circ$ ,  $-20^\circ$ , and  $-30^\circ$  and  $0^\circ$ ,  $10^\circ$ ,  $20^\circ$ , and  $30^\circ$  around  $y$ -axis to synthesize MVLHF and MVRHF images, respectively. MVLHF images were flipped and shifted along respective MVRHF images such that they were completely overlapped (flipped MVRHF images can also be shifted along MVLHF images equally). Subsequently, facial depth values on the same grid positions were averaged and complementary facial feature information provided by the nonoverlapping facial regions was retained to obtain more complete global information for each view separately. The outcome of the whole process was a set of four MVAHF images oriented at  $0^\circ$ ,  $10^\circ$ ,  $20^\circ$ , and  $30^\circ$ . The motivation behind using MVAHF images instead of MVWF images is as follows: (i) Facial feature information carried by a half face image is similar to that of the flipped other half face image due to intrinsic facial symmetry of the LHF and RHF. (ii) RHF region is gradually occluded by rotating a whole face image at  $-10^\circ$ ,  $-20^\circ$ , and  $-30^\circ$ . Similarly, LHF region is occluded by rotating the whole face image at  $10^\circ$ ,  $20^\circ$ , and  $30^\circ$ . The occluded face regions poorly contribute in the face recognition process. On the other hand, computational complexity of the system is two-fold. (iii) The multiview 3D information corresponding to MVWF images remains available by combining the facial information obtained from MVLHF and MVRHF images into MVAHF images. (iv) The synthesized MVAHF images provide stable features to evaluate the local variations and also include feature information from occluded facial regions less visible in frontal view images. Figure 10 readily shows the complementary face information through example synthesized MVAHF images employed for improving the face recognition accuracy.

3.2.1. *d-MVAHF-Based Face Identification Algorithm.* An overview of the proposed d-MVAHF-based 3D face recognition algorithm is given in Figure 2(b). To extract d-MVAHF features using dCNN, an MVAHF image of the size  $h \times w$  is processed through a deep network architecture known as AlexNet [56]. A pretrained AlexNet based deep network architecture was selected because of its better performance. AlexNet consists of five convolutional layers represented as C1, C2, C3, C4, C5 followed by three pooling layers denoted by P1, P2, P3, and three fully connected layers indicated by f6, f7, f8. Fully connected layers employ dropouts for regularization. Each convolutional layer is followed by a rectified linear unit (ReLU). The AlexNet architecture is graphically represented in Figure 2(b). The MVAHF-based facial features are extracted using the second to last fully connected layers followed by the normalization process. The output of layer  $k$  is a set  $A^k = \{a_1^k, a_2^k, a_3^k, \dots, a_n^k\}$  of MVAHF-based facial features.

The procedure for implementing the proposed approach is outlined as follows:

- (1) For each MVAHF image, a 2048-dimensional d-MVAHF feature vector was extracted from the f7 layer of AlexNet.
- (2) Matching scores between probe and gallery MVAHF images were calculated by comparing the respective  $L_2$  normalized d-MVAHF feature vectors. The matching scores were arranged as a matching-score matrix  $S$  of size  $m \times n$ , where  $m$  and  $n$  denote the size of probe and gallery sets in the respective order. The matrix  $S$  has a negative polarity reflecting that lower values of matching scores represent higher level of similarity between the probe and gallery images, and vice versa. This step produced four matching-score matrices  $S_j$ , for each of the normalized d-MVAHF feature vectors corresponding to AHF images oriented at  $0^\circ$ ,  $10^\circ$ ,  $20^\circ$ , and  $30^\circ$ .
- (3) Each of the matching-score matrices  $S_j$  was normalized before fusion in f8 layer of the AlexNet. For score normalization, min-max normalization rule was utilized to normalize each row for mapping original scores distribution to the interval  $[0, 1]$ . If maximum and minimum row specific values of raw matching scores are  $\max(S_{jrow})$  and  $\min(S_{jrow})$ , respectively, then normalized scores are computed as given in equation (3).

$$S_{jrow} = \frac{S_{jrow} - \min(S_{jrow})}{\max(S_{jrow}) - \min(S_{jrow})} \quad (3)$$

- (4) The four normalized matching-score matrices corresponding to the four MVAHF images were then fused using score based fusion to produce a combined matching-score matrix  $S_{row}$  as given in equation (4).

$$S_{row} = \sum_{j=1}^4 w_j S_{jrow} \quad (4)$$

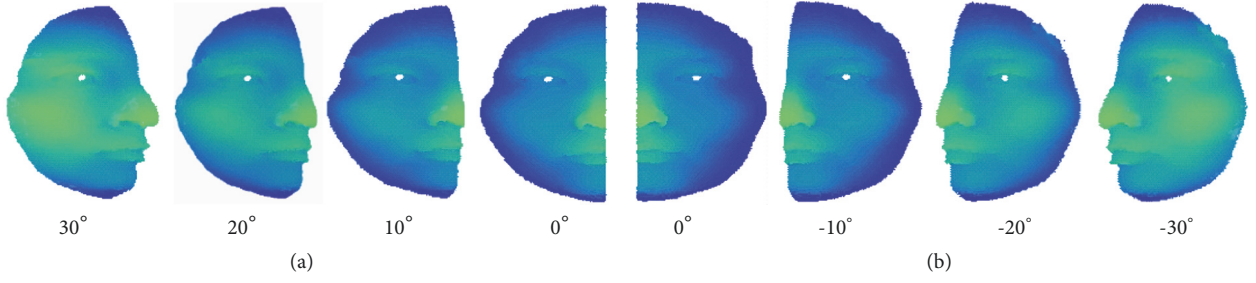


FIGURE 10: 3D scan of subject FRGC v2.0: 04221d553. (a) RHF images; (b) LHF images.

where  $w_j$  represents the weight assigned to the  $j$ th MVAHF image using the recognition accuracies obtained from MVAHF images as given in equation (5):

$$w_j = \frac{r_j}{\sum_{j=1}^4 r_j} \quad (5)$$

where  $r_j$  represents the recognition accuracies of the  $j$ th MVAHF image against the gallery. We can use the recognition accuracies in test phase as a given PFI is first converted into MVAHF images oriented at  $0^\circ$ ,  $10^\circ$ ,  $20^\circ$ , and  $30^\circ$ . Then, each of the mentioned MVAHF images is classified against the gallery and leads to four recognition accuracies which are subsequently used to compute the weights in equation (5). This procedure is similar as employed for each of the training images in the training phase. For example, if the recognition accuracies obtained from MVAHF images oriented at  $0^\circ$  are maximum then the corresponding matching score matrix is assigned the maximum weight. The matching score matrix  $S_{row}$  was again normalized as  $S'_{row}$  using the min-max rule as given in equation (3).

- (5) The normalized matching scores obtained from  $S'_{row}$  were utilized in the Softmax layer of the AlexNet to compute the final recognition accuracies.
- (6) The whole process was repeated to classify MVWF, MVLHF, and MVRHF images.

**3.2.2.  $d$ -MVAHF-SVM-Based Face Verification Algorithm.** For a binary classification problem such as face verification, SVM aims to employ a hyperplane  $w_x + b = 0$  having maximum margins, termed as optimal separating hyper plane (OSH) that separates training vectors of two classes  $(x_1, y_1), \dots, (x_i, y_i)$  where  $x_i \in R_n$  and  $y_i \in \{1, -1\}$  in a higher dimensional space. The objective function of the form given in equation (6) is minimized to obtain the OSH with constraints  $y_i[(w \cdot x_i) + b] \geq 1 - \xi_i$ ,  $\xi_i \geq 0$  for  $i = 1, \dots, k$ .

$$\theta(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^k \xi_i \quad (6)$$

where  $\xi_i$  are slack variables used to penalize errors if the data are not linearly separable, and  $C$  is the regularization

constant. Now sign of the following OSH surface function can be used to classify a test point.

$$f(x) = \sum_{i=1}^k y_i a_i K(x, x_i) + b \quad (7)$$

where  $a_i \geq 0$  are corresponding support vectors Lagrangian multipliers and  $b$  is determined by above-mentioned optimization problem. In equation (7),  $K$  is the kernel trick used to transform nonseparable data onto a higher dimensional space where it becomes linearly separable by a hyperplane,  $x_i$  is the  $i$ th training sample, and  $x$  is the test sample. It is experimentally observed in this study that radial basis function (RBF) kernel based SVM produces better recognition accuracies than the linear SVM and is of the form given in equation (8) where  $\sigma^2$  is spread of RBF.

$$K(x, x_i) = \exp \left[ -\frac{\|x - x_i\|^2}{2\sigma^2} \right] \quad (8)$$

The proposed face verification algorithm employs  $d$ -MVAHF-SVM-based classification approach using two neutral face images of each subject. In order to train the SVM, MahCos scores were computed between four  $d$ -MVAHF feature vectors of each image extracted using AlexNet as shown in Figure 2(b). MahCos score between two vectors  $s$  and  $t$  of image space is defined as the Cosine score calculated in the Mahalanobis space, as given in equations (9) and (10) [57].

$$\begin{aligned} d_{MahCos}(s, t) &= -\frac{m \cdot n}{|m| |n|} = -\frac{\sum_{i=1}^N (m_i n_i)}{\sqrt{\sum_{i=1}^N (m_i)^2} \sqrt{\sum_{i=1}^N (n_i)^2}} \\ &= -\frac{\sum_{i=1}^N ((s_i/\sigma_i)(t_i/\sigma_i))}{\sqrt{\sum_{i=1}^N (s_i/\sigma_i)^2} \sqrt{\sum_{i=1}^N (t_i/\sigma_i)^2}} \end{aligned} \quad (9)$$

where  $m_i = s_i/\sigma_i$ ,  $n_i = t_i/\sigma_i$  and  $\sigma_i$  is standard deviation of  $i$ th dimension. In this case, higher similarity yields higher score. Thus, the actual MahCos score is computed as given in equation (10).

$$D_{MahCos}(s, t) = 1 - d_{MahCos}(s, t) \quad (10)$$

Referring to Figure 2(c), MahCos scores were computed between the first neutral image of each subject and second

neutral image of the whole gallery  $G$ . The scores were computed by using (training, gallery) pairs of d-MVAHF feature vectors for images oriented at  $(0^\circ, 0^\circ)$ ,  $(10^\circ, 10^\circ)$ ,  $(20^\circ, 20^\circ)$ , and  $(30^\circ, 30^\circ)$  to populate rows 1 to 4 of a training score matrix  $T$ . Each element  $t_{ij}$  represents the score computed between d-MVAHF feature vectors of image  $i$  to image  $j$  where  $i, j \in \{1, 2, \dots, G\}$ . The element  $t_{ij}$  (for  $i = j$ ) represents genuine MahCos score computed between an image and itself, whereas the scores  $t_{ij}$  (for  $i \neq j$ ) represent imposter scores. The genuine scores (e.g.,  $t_{11}$ ) and the imposter scores (e.g.,  $t_{1G}$ ) corresponding to all four orientations constitute  $4 \times 1$  dimensional column vectors of genuine and imposter scores and are referred to as training vectors. For an example gallery of 20 subjects, there will be  $G \times G$  (400) total,  $G$  (20) genuine, and  $G^2 - G$  (380) imposter training score vectors.

In the classification phase, MahCos probe scores were computed between the d-MVAHF feature vector of PFI and second neutral image of the whole gallery as shown in Figure 2(c). The computed (probe, gallery) scores between d-MVAHF feature vector pairs of images oriented at  $(0^\circ, 0^\circ)$ ,  $(10^\circ, 10^\circ)$ ,  $(20^\circ, 20^\circ)$ , and  $(30^\circ, 30^\circ)$  were used to populate rows 1 to 4 of the probe score matrix  $P$  with  $4 \times 1$  dimensional, one genuine, and  $G-1$  probe score vectors (see Figure 2(c)). Based on the training of genuine and imposter d-MVAHF feature vectors, the SVM classifies the PFI against the gallery. Similar procedure was adopted to classify MVWF, MVLHF, and MVRHF images.

## 4. Results

The objective of this component of the study is to investigate the performance of proposed face alignment and recognition algorithm. Four databases, namely, GavabDB, Bosphorus, UMB-DB, and FRGC v2.0, are employed in the experiments. On each of these databases, face alignment, identification, and verification experiments are conducted to implement the proposed methodology. In the face identification and verification experiments the performance is reported as rank-1 identification rate and verification rate at 0.1% false accept rate (FAR) in the respective order. The considered 3D face databases, GavabDB [36], Bosphorus [38], UMB-DB [39], and FRGC v2.0 [40], are reviewed in the following section along with description of the experiments and results.

### 4.1. 3D Face Databases

**GavabDB Database.** The GavabDB [36] database contains 549 3D facial images acquired using Minolta VI-700 laser sensor, from 45 male and 16 female Caucasian subjects. Each subject is acquired 9 times under various facial expressions and large pose variations. The database contains six neutral images for each subject among which two, named “carai\_frontal1” and “carai\_frontal2,” are captured under frontal view. Another two are taken where a subject is looking up or down at angles  $+35^\circ$  or  $-35^\circ$  named “carai\_arriba” and “carai\_abajo,” respectively. Remaining two neutral images are scanned from right or left side at angles  $+90^\circ$  or  $-90^\circ$ , respectively, which are named “carai\_derecha” and “carai\_izquierda,” respectively. The three nonneutral

images “carai\_gesto,” “carai\_risa,” and “carai\_sonrisa” present a random gesture chosen by the subjects, accentuated laugh, and a smile, respectively. The GavabDB database carries several types of facial variations including variations in pose, expressions, occlusions, and resolution.

**The Bosphorus Database.** The Bosphorus database [38] is a multipose 3D face database constructed to enable testing of realistic and extreme pose variations, expression variations, and typical occlusions that may occur in real life. Each subject is captured with approximately 13 poses, 34 expressions (such as happiness, sadness, and surprise), and 4 occlusions. The database contains a total of 4666 scans collected from 61 male and 44 female subjects, including 29 professional actors/actresses. The 3D scans were acquired using Inspeck Mega Capturor II 3D and processed to remove holes and spikes and to crop the facial area.

**UMB-DB Database.** The UMB-DB database [39] is composed of 1473 3D depth images of 142 [27] subjects including 98 male and 45 female subjects, mostly in the age range of 19 to 50 years. Almost all of the acquired subjects are Caucasian with a few exceptions. Each subject is included with a minimum of three neutral, nonneutral (angry, smiling, and bored), and occluded acquisitions, with a size of  $480 \times 640$ . The Minolta Vivid 900 laser scanner is used to capture 2D and 3D images simultaneously. Face images have been captured in several indoor locations with uncontrolled lighting conditions. The database is released without any processing such as noise reduction or hole filling.

**FRGC v2.0 Database.** FRGC v2.0 3D database [40] is a publically available license based database. It supports 6 experiments among which our study is focused on Experiment 3, designed for 3D shape and texture analysis. The face scans are acquired at varying lengths from the scanner with variable resolution, frontal view, and minimal pose variations by a Minolta Vivid 900/910 series sensor. The scans are available in the form of four matrices of the size  $480 \times 640$ . The matrices represent  $x$ ,  $y$ ,  $z$  coordinates of faces and a binary representation showing valid points of the  $x$ ,  $y$ ,  $z$  matrices (whereas  $z$  is the facial distance from the scanner). The database contains male and female subjects aged 18 years and above. About sixty percent of the subjects carry neutral expressions, and others carry expressions of happiness, sadness, surprise, disgust, and inflated cheeks. Some of the subjects carry occlusions (such as hair, spikes, and holes on face), but none of them is wearing glasses [58].

**4.2. Face Alignment Experiments.** Using the proposed PCF algorithm, alignment experiments are performed on GavabDB, Bosphorus, UMB-DB, and FRGC v2.0 databases to align the faces at the minimum  $L_2$  norm between nose tip and 3D scanner. In order to evaluate the alignment accuracy of face images, there is no existing evaluation criterion. One method that can be employed is human judgment but human judgment method is not automatic. Therefore,  $L_2$  norm minimization evaluation method is employed in this

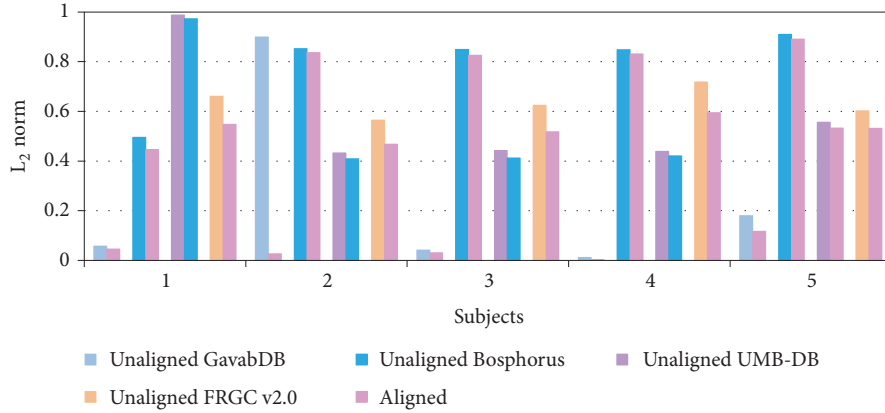


FIGURE 11: PCF algorithm based minimized  $L_2$  norms shown for five subjects after alignment.

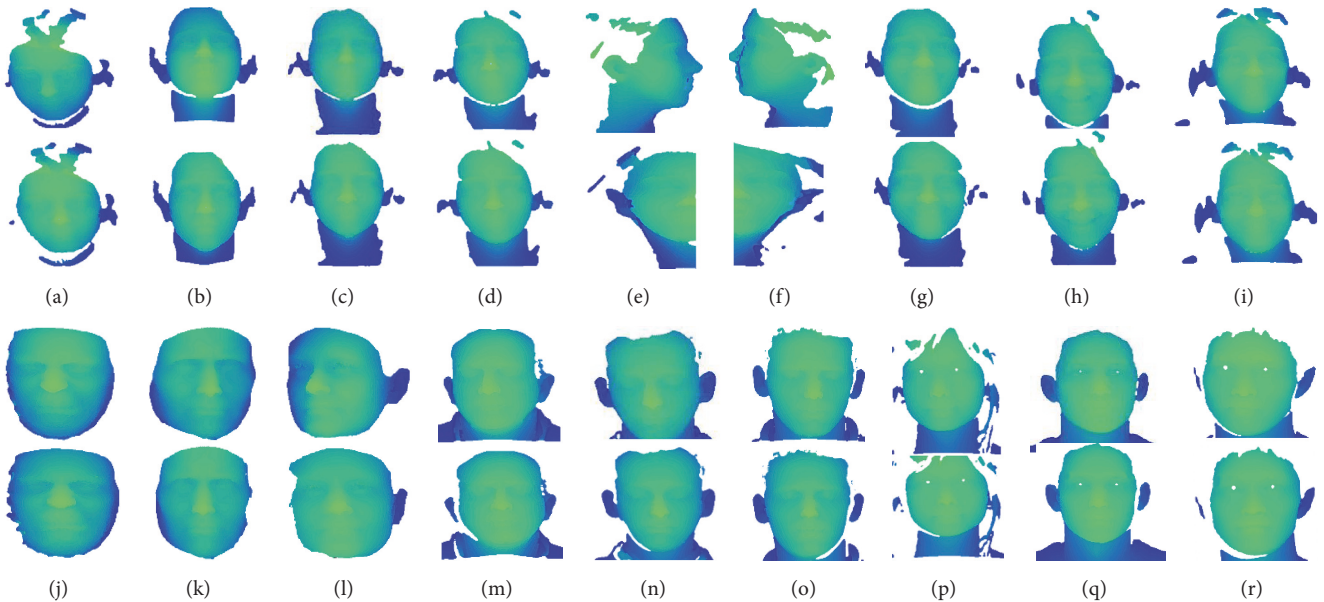


FIGURE 12: Example 3D face images: original: (rows 1, 3); aligned: (rows 2, 4).

study. It is observed in the experiments that the results of the  $L_2$  norm minimization evaluation method and manual judgment are quite similar and that the mentioned method is a promising automatic criterion to check alignment accuracy.

The minimized and normalized  $L_2$  norms for five unaligned images of subjects GavabDB: cara1\_gesto to cara2\_abajo, Bosphorus: bs000\_E\_DISGUST\_0 to bs000\_E\_SURPRISE\_0, UMB-DB: 000006\_0190\_F\_BO\_F to 000012\_0024\_M\_AN\_F and FRGC v2.0: 04203d436 to 04203d444 are shown in Figure 11. Figure 12 depicts example, original, as well as aligned face images from GavabDB: cara1\_(a) abajo (b) arriba (c) frontal1 (d) frontal2 (e) derecha (f) izquierda (g) gesto (h) risa (i) sonrisa, Bosphorus: (j) bs017\_E\_DISGUST\_0 (k) bs001\_E\_ANGER\_0 (l) bs000\_YR\_R20\_0, UMB-DB: (m) 001409\_0002\_M\_NE\_F (n) 001433\_0010\_M\_BO\_F (o) 001355\_0001\_M\_AN\_F, and FRGC v2.0: (p) 04217d399 (q) 04482d418 (r) 04387d322, respectively. The proposed PCF alignment algorithm accurately aligned and minimized  $L_2$

norms of 99.82%, 100% (nonoccluded), 100%, and 99.95% subjects from GavabDB, Bosphorus, UMB-DB, and FRGC v2.0 databases, respectively.

4.3. Face Recognition Experiments. The protocols and results of face recognition experiments are given using four databases as follows:

4.3.1. Experiments on GavabDB Database

- (1) For the identification setup, experimental protocol of [46] is considered to perform N vs. N experiments using d-MVWF, d-MVLHF, d-MVRHF, and d-MVAHF images. 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.

TABLE 2: Unweighted (U) and weighted (W) recognition rates (%) using GavabDB database.

Proposed methodology	Rank-1 Identification rates										Verification rates
	FF		Rotated looking up		Rotated looking down		LPF		RPF		
	U	W	U	W	U	W	U	W	U	W	
d-MVWF	96.7	100	96.7	100	95.1	98.4	-	-	-	-	100
d-MVLHF	95.1	98.4	93.4	96.7	93.4	96.7	91.8	95.1	-	-	96.7
d-MVRHF	93.4	96.7	95.1	98.4	91.8	95.1	-	-	80.3	83.6	98.4
d-MVAHF	96.7	100	96.7	100	95.1	98.4	-	-	-	-	100

TABLE 3: Unweighted (U) and weighted (W) rank-1 identification rates (%) using Bosphorus and UMB-DB databases.

Proposed methodology	Bosphorus										UMB-DB	
	FF		YR <sup>1</sup> < 90°				YR= 90°		Overall		Frontal Face	
			525 images				210 images		1365 images			
	U	W	U	W	U	W	U	W	U	W	U	W
d-MVWF	97.1	100	92.2	95.4	-	-	93.1	96	96.5	99.3		
d-MVLHF	95.2	98.1	91.4	94.5	84.3	87.1	91.8	94.9	93.7	97.2		
d-MVRHF	96.2	99	91	94.1			91.3	94.4	94.4	97.9		
d-MVAHF	97.1	100	92.2	95.4	-	-	93.1	96	96.5	99.3		

<sup>1</sup>YR is yaw rotation (along  $y$ -axis in  $xz$  plane).

- (2) For identification of profile face images, this study employs d-MVLPF and d-MVRPF images for each of the 61 subjects.
- (3) For evaluation of face verification algorithm, the protocol used in the study [44] is followed where “frontal1” image of each subject is enrolled in the gallery to follow the experimental protocol mentioned for this database and the image “frontal2” is used as probe. Referring to Section 3.2.2, two neutral images per subject are used to calculate d-MVWF, d-MVLHF, d-MVRHF, and d-MVAHF-based training scores for SVM classifier in the training phase. Therefore, the neutral image “abajo” is included as second image along with “frontal1” of the gallery for computing pairwise training scores, whereas “frontal2” and “frontal1” are used for pairwise probe score calculation for N vs. N verification experiments. The face identification and verification performance of the proposed methodology for N vs. N experiments is given in Table 2.

**4.3.2. Experiments on Bosphorus Database.** Using Bosphorus database, the proposed d-MVAHF identification algorithm is evaluated by performing N vs. N experiments on d-MVWF, d-MVLHF, d-MVRHF, and d-MVAHF images using experimental protocol of the study [27]. In the mentioned protocol the gallery set consists of first neutral scan of each subject (105 scans), whereas the probe set is created using the remaining 194 neutral scans and the challenging pose scans in separate experiments. The performance of the proposed identification approach is given in Table 3.

**4.3.3. Experiments on UMB-DB Database.** For evaluation of the proposed d-MVAHF identification algorithm, we employ the experimental protocol of the study [27] to create the N. vs. N. experiments using d-MVWF, d-MVLHF, d-MVRHF, and d-MVAHF images where the gallery set is comprised of one neutral scan per subject (142 scans) and the probe set contains all remaining neutral scans (299 scans). The performance of our proposed methodology is given in Table 3.

#### 4.3.4. Experiments on FRGC v2.0 Database

- (1) For evaluation of face identification algorithm, experimental protocol of the study [41] is employed for N vs. N experiments using d-MVWF, d-MVLHF, d-MVRHF, and d-MVAHF images from FRGC v2.0 database which contains 2469 neutral images [41]. In these experiments, probe set is created using 2003 neutral images, whereas first neutral image of each of the 466 subjects is enrolled in the gallery.
- (2) Face verification algorithm was investigated by creating N vs. N experiments using the d-MVWF, d-MVLHF, d-MVRHF, and d-MVAHF images. The FRGC v2.0 database comprises 370 such subjects that have at least two neutral images [45]. Therefore, two images per subject (740 images) are included in the gallery to calculate SVM training scores. In case of the subjects that have more than two neutral images, the first two of the stored neutral images are contained in the gallery. All the remaining neutral face images are used as probe set. The performance of the proposed identification and verification algorithms is given by cumulative match characteristic (CMC) curves

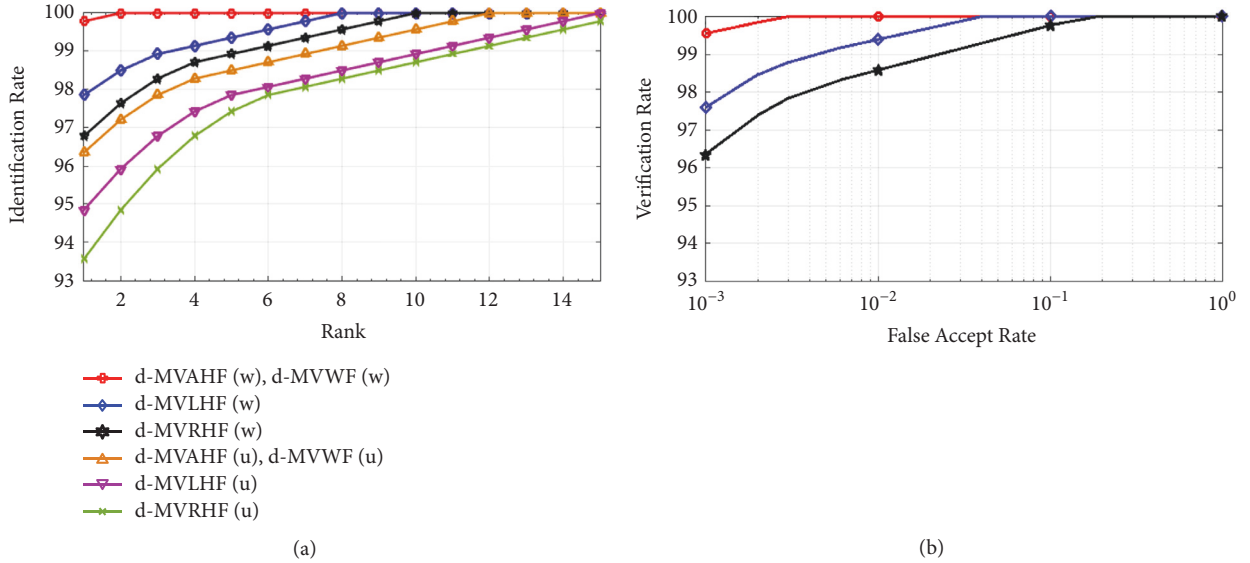


FIGURE 13: (a) CMC; (b) ROC curves showing weighted (w) and unweighted (u) face identification and verification rates for FRGC v2.0 database, respectively.

in Figure 13(a) and receiver operating characteristic (ROC) curves in Figure 13(b).

**4.4. Computational Complexity Analysis.** Computational complexity analysis of the proposed algorithm is given in terms of Big- $O$  notation as follows.

- (1) The computational complexity of the proposed PCF alignment algorithm is of the order of  $O(m)$ , where  $m$  represents total number of facial depth points in the point cloud.
- (2) For d-MVAHF-based face identification, the total time complexity of AlexNet is calculated in terms of all of its convolutional layers as  $O(\sum_{j=1}^n y_{j-1} x_j^2 y_j z_j^2)$ . Here,  $n$  represents the number of convolutional layers,  $y_{j-1}$  is the number of input channels of the  $j$ th layer,  $y_j$  is the number of filters of the  $j$ th layer,  $x_j$  is the spatial size of the filters, and  $z_j$  denotes the size of the output feature map.
- (3) For the d-MVAHF-SVM-based face verification setup, the computational complexity involves complexity of the AlexNet mentioned above along with complexity of the SVM classifier which is of the order of  $O(\log(n))$ . The computational complexity analysis shows that the feature extraction stage using AlexNet is computationally the most demanding and expensive stage of the proposed face identification and verification algorithms.
- (4) The experiments were performed on a P4 computer with an Intel core i7 1.8Ghz CPU and 8GB of RAM. The computational complexity in terms of computation time is shown in Table 4. The time computed after feature extraction by the Alexnet with its own classifier in face identification is higher compared to using SVM classifier in classification phase for

face verification. This is because Alexnet classifier generates the complex decision boundaries in the feature space for classification. On the other hand, SVM only takes into account the global matching scores resulting into lower computation time.

**4.5. Comparison with Existing Algorithms.** The performance of the proposed approach is compared with the existing state-of-the-art earlier studies in the following.

*GavabDB.* Referring to Table 5, the study [26] proposed a Riemannian framework based face recognition approach to analyze facial shapes using radial curves emanating from the nose tip. The study [28] reported face recognition results employing multiscale extended Local Binary Pattern descriptors and a hybrid matching method using local features. The study [44] proposed a face recognition approach using 3D keypoint extraction and sparse comparison based similarity evaluation. The algorithm proposed in the study [46] encoded different types of facial features and modalities into a compact representation using covariance based descriptors where face recognition was performed using a geodesic distance based approach. The study [47] presented a 3D face keypoint detection and matching approach based on principle curvatures. In this study matching was performed using local shape descriptors, sparse representation based reconstruction method, and score level fusion. The approach proposed in Ref. [59] employed 3D binary ridge images along with principal maximum curvature and ICP based matching. The study [60] proposed a sparse representation based framework for face recognition using low level geometric features.

*Bosphorus.* The approach presented in the study [27] reported face recognition accuracies employing facial



TABLE 4: Time complexity of the proposed approach in seconds.

Preprocessing	MVAHF synthesis	Feature extraction	Classification		Total	
			Face recognition	Face verification	Face recognition	Face verification
0.451	0.089	1.024	0.029	0.021	1.593	1.585

TABLE 5: Recognition accuracies comparison for the proposed and existing approaches using GavabDB, Bosphorus, and UMB-DB databases.

Algorithms	GavabDB					Verification rates	Bosphorus				UMB-DB	
	FF	Rank-1 Identification rates		LPF	RPF		FF	Rank-1 Identification rates			Overall	FF
		Rotated looking up	Rotated looking down					YR <sup>1</sup> < 90°	YR = 90°			
Existing	100	98.4	96.7	93.4	81.9	82.3	100	81.6	45.7	88.6	98.7	
	[44]	[44]	[44]	[44]	[44]	[59]	[27]	[61]	[61]	[61]	[27]	
	100	98.4	99.2	86.9	70.5	95.1	100	84.1	47.1	91.1	98	
	[46]	[46]	[46]	[26]	[26]	[60]	[62]	[62]	[62]	[62]	[39]	
	100	100	98.4	93.4	78.7	-	-	94.8	57.1	92.8	-	
	[47]	[47]	[47]	[28]	[28]			[63]	[47]	[47]		
Proposed	d-MVLHF	98.4	96.7	96.7	95.1 <sup>2</sup>	-	96.7	98.1	94.5	87.1 <sup>2</sup>	94.9	97.2
	d-MVRHF	96.7	98.4	95.1	-	83.6 <sup>2</sup>	98.4	99	94.1		94.4	97.9
	d-MVWF/ d-MVAHF	100	100	98.4	95.1	83.6	100	100	95.4	-	96	99.3

<sup>1</sup> YR is yaw rotation (along  $y$ -axis in  $xz$  plane).

<sup>2</sup> LPF, RPF, and face images at YR = 90° turn into LHF and RHF, respectively, after alignment.

depth information and ICP algorithm and the study [47] is mentioned in above paragraph. The face recognition methodology given in the paper [61] extracted local descriptors to perform matching according to differential surface measurements. The study [62] employed surface differential measurement based keypoint descriptors to perform face recognition using multitask sparse representation based fine-grained matching algorithm. The study [63] proposed to fit 3D deformable model to unseen PFIs for face recognition.

*UMB-DB*. The study [27] is discussed in above paragraph whereas the recognition accuracies reported in the paper [39] are based on an approach employing PCA.

*FRGC v2.0*. Referring to Table 6, the study [17] is focused on using DT-CWT and LDA based face recognition approach. The study [41] proposed to employ isogeodesic stripes and 3D weighted walkthrough (3DWW) descriptors in the face recognition process. The methodology proposed in the study [42] integrated global and local geometric cues for face recognition employing Euclidean distance based classifier. Finally, the study [43] proposed a local features based resolution invariant approach to classify scale space extrema using SVM classifier, whereas the studies [47, 62, 63] are discussed with approaches presented in Table 5. The proposed d-MVAHF-based 3D face recognition approach has yielded better results than the existing state-of-the-art studies given in Tables 5 and 6.

## 5. Discussion

The proposed study covers the problem of 3D face alignment and face recognition with applications in identification and verification scenarios. The former employs PCF approach while the latter is based on d-MVAHF images. The performance of these two algorithms is discussed separately.

### 5.1. PCF Alignment Algorithm

- (1) The proposed PCF alignment algorithm achieved 99.82% and 99.95% alignment accuracy on GavabDB and FRGC v2.0 databases, respectively. Similarly, an accuracy rate of 100% was obtained on nonoccluded subsets of Bosphorus and UMB-DB databases each. The nose tip was not detectable for one subject in GavabDB database and two subjects in FRGC v2.0 database; else the accuracy of the proposed alignment algorithm would have been 100% for each of these databases. The excellent level of accuracies is attributed to the fine alignment performed at a step size of 0.1°.
- (2) The proposed alignment algorithm is very effective for face recognition applications because it rotates the nose tip in correct direction to save computational cost. This rotation in correct direction is because of pose learning aspect of the proposed approach. e.g., pose learning of a LOFI or LUFi correctly dictates the

TABLE 6: Recognition accuracies comparison for the proposed and existing approaches using FRGC v2.0 database.

	Existing algorithms							Proposed algorithm		
	[17]	[41]	[42]	[43]	[47]	[62]	[63]	d-MVLHF	d-MVRHF	d-MVWF/d-MVAHF
Face identification	98.7	96.1	93.8	98	99.6	98.7	99.8	97.9	96.8	99.8
Face verification	99.5	97.7	95.4	98.3		-		97.6	96.4	99.6

algorithm to rotate the nose tip to the right side or downwards for alignment.

- (3) The proposed PCF alignment algorithm is computationally very efficient. Referring to Section 3.1.3, it first aligns the nose tip only employing 35 (3+11+21) rotations in each of  $xz$  and  $yz$  planes. Then whole face image is aligned in a single 3D rotation in each plane (instead of 35 rotations) using the knowledge learned from the nose tip alignment. Please note that aligning the whole face instead of nose tip only at the cost of 35 rotations is computationally very expensive. For example, a 3D face image composed of 0.3 million depth points requires  $0.3 \times 35 = 10.5$  million rotations. The computational efficiency is attributed to alignment of nose tip prior to the whole face image.

### 5.2. d-MVAHF-Based 3D Face Recognition

- (1) The proposed d-MVAHF-based 3D face recognition approach obtained rank-1 identification rates of 100%, 100%, 98.4%, 95.1%, and 83.6% for FF, rotated looking up, rotated looking down, LPE, and RPF subsets of GavabDB database, respectively. Using Bosphorus database, rank-1 identification rates of 100%, 95.4%, 87.1%, and 96% were obtained for FF,  $YR < 90^\circ$ ,  $YR = 90^\circ$ , and overall experiments. Similarly, rank-1 identification rate of 99.3% was obtained for FF experiment on UMB-DB database, whereas rank-1 identification rate of 99.8% was achieved using FRGC v2.0 database.

The proposed d-MVAHF-SVM-based face verification approach achieved a verification rate of 100% and 99.57% at 0.1% FAR for FF experiments using GavabDB and FRGC v2.0 databases, respectively. The improved identification and verification rates of the proposed study compared to the studies [17, 26–28, 39, 41–44, 46, 47, 61–63] and [17, 41–43, 59, 60], respectively, are attributed to d-MVAHF-based approach, whereas the mentioned studies neither used deep learning nor employed multiview approach.

- (2) Using d-MVAHF images, recognition accuracies equivalent to that of d-MVWF images were achieved at a reduced computational cost of 71%. This is because d-MVWF-based approach employed seven synthesized whole face images of a subject oriented at  $0^\circ$ ,  $\pm 10^\circ$ ,  $\pm 20^\circ$ , and  $\pm 30^\circ$ . On the other hand, d-MVAHF-based approach integrated 3D facial information of seven MVWF images into four MVAHF

images oriented at  $0^\circ$ ,  $10^\circ$ ,  $20^\circ$ , and  $30^\circ$  which is equivalent to using two whole face images.

- (3) Comparative evaluation was also performed employing d-MVLHF and d-MVRHF based face identification and verification approaches. For d-MVLHF based approach, the identification accuracies of FF, rotated looking up, and rotated looking down experiments and verification accuracies were decreased by 1.63%, 3.41%, 1.76%, and 3.41%, respectively, using GavabDB database. For d-MVRHF based approach the mentioned accuracies were decreased by 3.41%, 1.63%, 3.47%, and 1.63%, respectively. For FF,  $YR < 90^\circ$ , and overall experiments of Bosphorus database, the d-MVLHF, and d-MVRHF based identification accuracies were decreased by 1.94%, 0.95%, and 1.16% and 1.01%, 1.38%, and 1.69%, respectively. Similarly, the d-MVLHF and d-MVRHF based identification accuracies on UMB-DB database were decreased by 2.16% and 1.43%, respectively, for FF experiment. For the same experiment on FRGC v2.0 database, the d-MVLHF and d-MVRHF based identification rates were reduced by 1.94% and 3.1%, whereas, the verification rates were reduced by 2.05% and 3.32%, respectively. The reduction in recognition accuracies is because of noise, or motion artifacts introduced at the time of face image acquisition.
- (4) Weight assignment strategy enhanced unweighted rank-1 identification rates by 3.56%, 3.24%, 3.45%, and 3.41% in the experiments performed on GavabDB, Bosphorus, UMB-DB, and FRGC v2.0 databases, respectively. This enhancement is because of assigning more weights to better performing MVAHF images (please see equation (5)).
- (5) Experimental results suggest that integration of the knowledge learned from MVWF images into d-MVAHF images boosts the face recognition accuracies. This is attributed to the fact that multiview face images provide more facial feature information for classification than the case of single view facial features.
- (6) Experimental results of the PCF alignment and d-MVAHF-based 3D face recognition algorithms are comparable in all four employed databases. These databases contain several types of variations such as gender, pose, age, noise, and resolution variations (Section 4.1). This indicates that the proposed methodology is capable of aligning and classifying subjects captured with several mentioned variations.

(7) The performance of face recognition degrades significantly when the input images are of low resolution such as images captured by surveillance cameras or from a large distance [64]. This is because of unavailability of the discriminating information present in the high resolution face images. On the other hand, face recognition accuracies improve with the increasing resolution of PFIs [65]. There are two standard approaches to handle this problem: (1) downsampling approach where the resolution of gallery images is downsampled to the resolution of PFIs and (2) super resolution approach where the low resolution of PFIs is improved into higher resolution images [64]. The proposed d-MVAHF-based approach can be employed to recognize low resolution depth images. Referring to Tables 5 and 6, as the proposed approach outperforms existing approaches using high resolution PFIs, it would perform better than the existing approaches to handle low resolution PFIs. This is because the initial layers of dCNNs can effectively learn low level features encountered in low resolution images (for example, lines, dots, etc.). In contrast, the later layers tend to learn high level features like shapes and objects based on low level features.

## 6. Conclusions

In this paper, a novel approach based on deeply learned pose invariant image analysis with applications in 3D face recognition is presented. The PCF alignment algorithm employed the following: (i) pose learning approach using nose tip heuristic to estimate acquisition pose of the face; (ii)  $L_2$  norm minimization based coarse to fine approach for nose tip alignment; and (iii) a transformation step to align the whole face image incorporating the knowledge learned from nose tip alignment. Face recognition algorithm was implemented in both of identification and verification setups. The dCNN based face identification algorithm was implemented using d-MVAHF images, whereas the verification algorithm was employed using d-MVAHF-SVM-based methodology. The experimental performance was evaluated using four benchmark 3D face databases, namely, GavabDB, Bosphorus, UMB-DB, and FRGC v2.0.

In conclusion, it was observed that (i) the proposed PCF alignment algorithm is capable of correctly aligning the frontal and profile face images, (ii) its pose learning aspect is very effective to find correct direction of rotation for facial alignment, (iii) it is computationally very efficient due to alignment of the nose tip first, (iv) LHF and RHF based intrinsic facial symmetry is a promising measure to evaluate d-MVAHF-based face recognition, (v) d-MVAHF images and d-MVWF images produced similar recognition accuracies, (vi) MVLHF images and MVRHF images yielded relatively decreased recognition rates compared to MVAHF images, (vii) weight assignment strategy significantly enhanced the recognition rates, (viii) deeply learned facial features possess more discriminative power compared to handcrafted features, (ix) experimental results show that

the real 3D facial feature information integrated in the d-MVAHF images significantly enhanced the face recognition accuracies, (x) the proposed PCF alignment and d-MVAHF-based face recognition is computationally efficient compared to d-MVWF image based face recognition, and (xi) the frontal and profile face recognition accuracies produced by the proposed methodology are better than existing state-of-the-art methods and are comparable in all databases for both of identification and verification experiments.

As a future direction, we plan to (i) develop 3D face alignment algorithm using deep learning based approach and (ii) reduce the number of synthesized multiview face images such that the computational complexity of the system is further reduced and overall system performance can be enhanced.

## Data Availability

Previously reported face image datasets including the GavabDB, Bosphorus, UMB-DB, and FRGC v2.0 have been used to support this study. The datasets are available upon request from the sponsors. The related datasets are publicly available at the following links: (1) GavabDB: <http://archive.is/2K19W>, (2) Bosphorus: <http://bosphorus.ee.boun.edu.tr/Home.aspx>, (3) UMB-DB: <http://www.ivl.disco.unimib.it/minisites/umbdb/request.html>, and (4) FRGC v2.0: <https://cvrl.nd.edu/projects/data/#face-recognition-grand-challenge-frgc-v20-data-collection>.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Authors' Contributions

Naeem Ratyal, Muhammad Sajid, Anzar Mahmood, and Sohail Razzaq conceived the idea and contributed in the experimentation process and writing of manuscript including tables and figures. Imtiaz Ahmad Taj, Saadat Hanif Dar, Nouman Ali, Muhammad Usman, Mirza Jabbar Aziz Baig, and Usman Mussadiq took part in organizing the manuscript and conducting experiments to compute time complexity. All authors contributed to the final preparation of the manuscript.

## Acknowledgments

The authors are thankful to the organizers of GavabDB, Bosphorus, UMB-DB, and FRGC, for provision of the databases for research purposes.

## References

- [1] M. Sajid, N. Iqbal Ratyal, N. Ali et al., "The impact of asymmetric left and asymmetric right face images on accurate age estimation," *Mathematical Problems in Engineering*, vol. 2019, Article ID 8041413, 10 pages, 2019.
- [2] M. Bessaoudi, M. Belahcene, A. Ouamane, A. Chouchane, and S. Bourennane, "Multilinear Enhanced Fisher Discriminant

- Analysis for robust multimodal 2D and 3D face verification,” *Applied Intelligence*, vol. 49, no. 4, pp. 1339–1354, 2019.
- [3] E. Basaran, M. Gökmen, and M. E. Kamasak, “An efficient multiscale scheme using local Zernike moments for face recognition,” *Applied Sciences (Switzerland)*, vol. 8, no. 5, article no. 827, 2018.
  - [4] S. Z. Gilani and A. Mian, “Learning from millions of 3D scans for large-scale 3D face recognition,” in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1896–1905, Salt Lake City, UT, USA, June 2018.
  - [5] A. Irtaza, S. M. Adnan, K. T. Ahmed et al., “An ensemble based evolutionary approach to the class imbalance problem with applications in CBIR,” *Applied Sciences (Switzerland)*, vol. 8, no. 4, article no. 495, 2018.
  - [6] N. Dagnes, E. Vezzetti, F. Marcolin, and S. Tornincasa, “Occlusion detection and restoration techniques for 3D face recognition: a literature review,” *Machine Vision and Applications*, vol. 29, no. 5, pp. 789–813, 2018.
  - [7] S. Ramalingam, “Fuzzy interval-valued multi criteria based decision making for ranking features in multi-modal 3D face recognition,” *Fuzzy Sets and Systems*, vol. 337, pp. 25–51, 2018.
  - [8] M. Sajid, N. Ali, S. H. Dar et al., “Data augmentation-assisted makeup-invariant face recognition,” *Mathematical Problems in Engineering*, vol. 2018, Article ID 2850632, 10 pages, 2018.
  - [9] J. Kittler, P. Koppen, P. Kopp, P. Huber, and M. Ratsch, “Conformal mapping of a 3d face representation onto a 2D image for CNN based face recognition,” in *Proceedings of the 11th IAPR International Conference on Biometrics, ICB 2018*, pp. 124–131, Australia, February 2018.
  - [10] M. Bessaoudi, M. Belahcene, A. Ouamane, and S. Bourennane, “A novel approach based on high order tensor and multi-scale locals features for 3D face recognition,” in *Proceedings of the 4th International Conference on Advanced Technologies for Signal and Image Processing, ATSSIP 2018*, pp. 1–5, Tunisia, March 2018.
  - [11] F. Liu, R. Zhu, D. Zeng, Q. Zhao, and X. Liu, “Disentangling Features in 3D Face Shapes for Joint Face Reconstruction and Recognition,” in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5216–5225, Salt Lake City, UT, USA, June 2018.
  - [12] A. T. Tran, T. Hassner, I. Masi, E. Paz, Y. Nirkin, and G. Medioni, “Extreme 3D face reconstruction: seeing through occlusions,” in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3935–3944, Salt Lake City, UT, USA, June 2018.
  - [13] N. Pears, Y. Liu, and P. Bunting, *3D Imaging, Analysis and Applications*, vol. 3, Springer, Berlin, Germany, 2012.
  - [14] N. Werghi, C. Tortorici, S. Berretti, and A. Del Bimbo, “Boosting 3D LBP-Based face recognition by fusing shape and texture descriptors on the mesh,” *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 5, pp. 964–979, 2016.
  - [15] L. Spreeuwens, “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, no. 3, pp. 389–414, 2011.
  - [16] K. W. 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.
  - [17] X. Wang, Q. Ruan, Y. Jin, and G. An, “Three-dimensional face recognition under expression variation,” *Eurasip Journal on Image and Video Processing*, vol. 2014, no. 51, 2014.
  - [18] S. Elaiwat, M. Bennamoun, F. Boussaid, and A. El-Sallam, “3-D face recognition using curvelet local features,” *IEEE Signal Processing Letters*, vol. 21, no. 2, pp. 172–175, 2014.
  - [19] 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, Article ID e100120, pp. 1–9, 2014.
  - [20] S. Soltanpour, B. Boufama, and Q. M. J. Wu, “A survey of local feature methods for 3D face recognition,” *Pattern Recognition*, vol. 72, pp. 391–406, 2017.
  - [21] A. Ouamane, A. Chouchane, E. Boutellaa, M. Belahcene, S. Bourennane, and A. Hadid, “Efficient tensor-based 2D+3D face verification,” *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 11, pp. 2751–2762, 2017.
  - [22] 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, 2005.
  - [23] 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.
  - [24] C. Heshier, A. Srivastava, and G. Erlebacher, “A novel technique for face recognition using range imaging,” in *Proceedings of the 7th International Symposium on Signal Processing and Its Applications, ISSPA 2003*, vol. 2, pp. 201–204, France, July 2003.
  - [25] 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.
  - [26] 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.
  - [27] N. Alyuz, B. Gokberk, and L. Akarun, “3-D face recognition under occlusion using masked projection,” *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 5, pp. 789–802, 2013.
  - [28] 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, 2012.
  - [29] 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, 2010.
  - [30] 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.
  - [31] T. Papatheodorou and D. Rueckert, *3D Face Recognition*, I-Tech Education and Publishing, Vienna, Austria, 2007.
  - [32] C. C. Queirolo, L. Silva, O. R. P. Bellon, and M. Pamplona 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.
  - [33] C. C. Queirolo, L. Silva, O. R. P. Bellon, and M. P. Segundo, “3D face recognition using the surface interpenetration measure: a comparative evaluation on the FRGC database,” in *Proceedings of the 2008 19th International Conference on Pattern Recognition, ICPR 2008*, USA, December 2008.

- [34] 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.
- [35] K. Cao, Y. Rong, C. Li, X. Tang, and C. C. Loy, "Pose-robust face recognition via deep residual equivariant mapping," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5187–5196, Salt Lake City, UT, USA, June 2018.
- [36] A. B. Moreno and A. Sanchez, "GavabDB: a 3D face database," in *Proceedings of the Second COST Workshop on Biometrics on the Internet: Fundamentals, Advances and Applications*, pp. 77–82, 2004.
- [37] M. Lewis, "Factors affecting the perception of 3D facial symmetry from 2D projections," *Symmetry*, vol. 9, no. 10, p. 243, 2017.
- [38] A. Savran, N. Alyüz, H. Dibeklioglu et al., "Bosphorus database for 3D face analysis," in *Biometrics and Identity Management*, vol. 5372 of *Lecture Notes in Computer Science*, pp. 47–56, Springer Berlin Heidelberg, Berlin, Germany, 2008.
- [39] A. Colombo, C. Cusano, and R. Schettini, "UMB-DB: a database of partially occluded 3D faces," in *Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops, ICCV Workshops 2011*, pp. 2113–2119, Spain, November 2011.
- [40] P. J. Phillips, P. J. Flynn, T. Scruggs et al., "Overview of the face recognition grand challenge," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, pp. 947–954, USA, June 2005.
- [41] 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.
- [42] 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, 2008.
- [43] G. Zhang and Y. Wang, "Robust 3D face recognition based on resolution invariant features," *Pattern Recognition Letters*, vol. 32, no. 7, pp. 1009–1019, 2011.
- [44] S. Berretti, A. Del Bimbo, and P. Pala, "Sparse matching of salient facial curves for recognition of 3-D faces with missing parts," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 2, pp. 374–389, 2013.
- [45] 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.
- [46] 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.
- [47] 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*, vol. 59, no. 2, pp. 211–233, 2017.
- [48] 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.
- [49] 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.
- [50] T. Russ, C. Boehnen, and T. Peters, "3D face recognition using 3D alignment for PCA," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2006*, pp. 1391–1398, USA, June 2006.
- [51] X. Lu and A. K. Jain, "Automatic feature extraction for multiview 3D face recognition," in *Proceedings of the FGR 2006: 7th International Conference on Automatic Face and Gesture Recognition*, pp. 585–590, UK, April 2006.
- [52] S. Zafeiriou, G. A. Atkinson, M. F. Hansen et al., "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.
- [53] 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, no. 1, pp. 87–108, 2013.
- [54] P. Kamencay, R. Hudec, M. Bencko, 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.
- [55] X. Peng, M. Bennamoun, and A. S. Mian, "A training-free nose tip detection method from face range images," *Pattern Recognition*, vol. 44, no. 3, pp. 544–558, 2011.
- [56] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS '12)*, pp. 1097–1105, Lake Tahoe, Nev, USA, December 2012.
- [57] 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, Article ID e56510, 2013.
- [58] A. S. Mian, M. Bennamoun, and R. Owens, "An efficient multi-modal 2D-3D hybrid approach to automatic face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1927–1943, 2007.
- [59] M. H. Mahoor, *A multi-modal approach for face modeling and recognition [Ph.D. dissertation]*, 2008, Ph.D. dissertation.
- [60] X. Li, T. Jia, and H. Zhang, "Expression-insensitive 3D face recognition using sparse representation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2575–2582, 2009.
- [61] 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.
- [62] 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.
- [63] 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.
- [64] S. Biswas, K. W. Bowyer, and P. J. Flynn, "Multidimensional scaling for matching low-resolution face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 2019–2030, 2012.
- [65] M. Jian and K.-M. Lam, "Simultaneous hallucination and recognition of low-resolution faces based on singular value decomposition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 11, pp. 1761–1772, 2015.

