3D face recognition based on pose and expression invariant alignment ☆

Naeem Iqbal Ratyal [a], Imtiaz Ahmad Taj [a], Usama Ijaz Bajwa [b,*], Muhammad Sajid [a]

[a] Vision and Pattern Recognition Systems Research Group, Mohammad Ali Jinnah University, Islamabad, Pakistan
[b] Department of Computer Science, COMSATS Institute of Information Technology, Lahore, Pakistan

ARTICLE INFO

ABSTRACT

In this paper we present a novel pose and expression invariant approach for 3D face registration based on intrinsic coordinate system characterized by nose tip, horizontal nose plane and vertical symmetry plane of the face. It is observed that distance of nose tip from 3D scanner is reduced after pose correction which is presented as a quantifying heuristic for proposed registration scheme. In addition, motivated by the fact that a single classifier cannot be generally efficient against all face regions, a two tier ensemble classifier based 3D face recognition approach is presented which employs Principal Component Analysis (PCA) for feature extraction and Mahalanobis Cosine (MahCos) matching score for classification of facial regions with weighted Borda Count (WBC) based combination and a re-ranking stage. The performance of proposed approach is corroborated by extensive experiments performed on two databases: GavabDB and FRGC v2.0, confirming effectiveness of fusion strategies to improve performance.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Biometric data tends to classify human beings through their distinctive physiological and behavioral characteristics. Among well-known biometric modalities like iris, voice, fingerprint, palm print and gait, human face is considered an effective biometric for a wide range of vital applications in numerous and diverse domains. Biometric data acquisition process has always been thought as an affront to privacy of a subject and requires subject's cordial cooperation whereas facial biometric data is noninvasive and socially well accepted. Although face recognition using 2D intensity images has been studied intensively in the last decades and the majority of implemented face recognition systems are based on 2D images [1], it is still a very challenging task to recognize people using 2D images under diverse circumstances of pose, expression and illumination variations. With the emergence of reliable and inexpensive 3D scanners and the use of 3D facial structure information, the researchers generally conclude that some of the nuisance factors related to 2D images can be overcome by using 3D scans due to the reason that 3D data can more easily be pose corrected and it is not affected by illumination changes [2]. However, to handle facial expression variations, we still require evolution of algorithms, irrespective of using 2D or 3D data, because expressive faces complicate the face recognition by creating higher intra-class variance than inter-class variance. The main algorithms which have been evolved for 2D face recognition are based on holistic and local features of a facial image. In holistic techniques, face recognition is performed using the entire face. Three leading algorithms in this category include PCA,

---

Linear Discriminant Analysis (LDA) and Independent Component Analysis (ICA) [3]. The major local feature based approaches of 2D face recognition are Local Binary Pattern (LBP) [4], Scale Invariant Feature Transform (SIFT) [3] and Gabor Wavelets [5]. The appearance based paradigm has also been explored to find strength of 3D face recognition and the most popular, appearance based 3D methods are PCA, LDA and ICA [1]. Local feature based 3D matching exploits properties of local descriptive points, curves and regions. The study [1,6] reported point based methods that exploited meshSIFT algorithm to detect facial key points. Another study [7] explored the curvelet transform for detecting salient points on the face scans to build multi-scale local surfaces. The curve based method [8] proposed a Riemannian framework for analyzing facial shapes from radial curves emanating from the nose tip whereas the study [9] is a representative work of region based methods suitable for occlusion handling. The most crucial phase of any 3D face recognition algorithm is registration of facial surfaces and the final accuracy of the results greatly depends on the quality of the alignment module. In this work, we aim to present (i) an efficient face registration method based on intrinsic coordinate system and (ii) a face recognition method based on ensemble classifier that incorporates rank based fusion.

The first contribution of our study is a novel 3D registration algorithm. The proposed algorithm does not register two face scans to each other using the conventional registration process, rather transforms each 3D scan to an intrinsic coordinate system. This reference coordinate system is based on the nose tip, vertical symmetry plane and horizontal nose plane of the face. A novel method is presented for finding vertical symmetry plane and horizontal nose plane. Angles are measured between intrinsic and world coordinate systems which are then used to align the face surface. A novel method with mathematical validation is presented for quantitative analysis of our proposed registration method which is the second contribution of this study. The third contribution is a region based two tier, ensemble classification method where face is divided into regions like other approaches [10,11]. PCA based classifier using MahCos distance metric is implemented for classification of each region where PCA is used to extract compact feature vectors. PCA is a subspace method which provides dimensionality reduction while relying on a set of basis vectors which correspond to maximum variance direction of the image data [12]. It is observed that combining results of multiple overlapping regions using an ensemble classifier produces excellent recognition results. For performance evaluation of proposed face registration and recognition method, GavabDB 3D database [13] of 61 subjects and FRGC v2.0 3D database [14] comprising of 4007 images from 466 subjects with pose and expression variations are used which are the most commonly used databases for 3D face recognition.

This study is organized as follows. Section 2 presents an overview of related work. Proposed methodology for 3D face registration and recognition is presented in Section 3. Section 4 is dedicated for experiments and results and a description of the used 3D face databases. Conclusion and future work is presented in Section 5.

2. Related work

The related work comprises of two parts. In the first part, work related to 3D face registration is addressed. The second part focuses on 3D face recognition, i.e. the classification of 3D facial data.

2.1. 3D face registration

Facial registration transforms the facial features in such a way that they can be reliably compared. A few algorithms exist in literature [11,15,16] for registering 3D face scans which can be broadly divided into (i) One-to-all registration (ii) Registration to average face model (AFM) and (iii) Registration to intrinsic coordinate system.

In one-to-all registration, 3D point cloud of a probe face is registered to every face in the gallery. Iterative Closest Point (ICP) registration algorithm [15] falls into this category. In the ICP algorithm, the point clouds are registered by minimizing distance between them iteratively. The distance between the point clouds is minimized by rotating and translating one of the point clouds relative to the other. The closest point between the surfaces is determined by finding distance from each point of the first point cloud to the second point cloud and averaging all distances. Because in one-to-all registration a face is registered to every other face in the gallery, ICP is not suitable in face identification (one-to-many matches). On the other hand it is suitable in a verification scenario (one-to-one match). A limitation of ICP is that it needs initial course registration for convergence. Another disadvantage is the fact that it is relatively slow. ICP based approaches have been used in many leading research papers [17,18]. Simulated Annealing (SA) is another registration approach in one-to-all category which was implemented in [19]. SA [19] is based on stochastic algorithm that is used for local search. Initially centers of the mass of two faces are aligned. Then a course alignment is accomplished with an SA based searching method. For course alignment an estimation measure is reduced which uses MSAC robust estimator along with Mean Square Error (MSE) of the corresponding points of two faces to be compared. After that, a precise alignment is obtained based on a searching algorithm which uses Surface Interpenetration Measure (SIM) as the evaluation criterion. The drawback of SA based registration is its computational time which is comparable to ICP. For registration to an AFM [11], AFM is constructed by locating landmarks on the face. Average of landmark coordinates is calculated followed by the Procrustes analysis [11]. Then transformed landmark points are re-averaged for getting an AFM. This method is suitable for both of face identification and verification scenarios. A considerable disadvantage of the AFM based registration is its low accuracy rates because some of the spatial information is lost while creating AFM and a probe may be less accurately registered to an AFM compared to registering it directly to a true positive. The third method, registration to an intrinsic coordinate system requires landmark localization on 3D scans.

These landmarks are compared with the corresponding 3D points of the intrinsic coordinate system. The resulting transformation is then applied to whole point cloud. This method greatly depends on landmarks which may be less accurately located in case of non-frontal and expressive faces. This issue has been discussed in detail in [16]. A study [10] presents a registration method based on intrinsic coordinate system by finding vertical symmetry plane of the face, slope of the nose bridge and nose tip. These features can be called the landmark structures instead of landmarks, which mark position only. The choice of landmark structures is beneficial because they remain stable even under pose variations and facial expressions. These landmark structures define an intrinsic coordinate system. The registration algorithm presented in this study is based on intrinsic coordinate system using the nose tip, horizontal nose plane and vertical symmetry plane of the face. It is different from [10] in many respects including (i) The referred study finds origin of coordinate system by fitting 3D nose template on the symmetry plane and using Normalized Cross Correlation (NCC) as a matching criterion whereas in our study, nose tip is a point which is at a minimum distance from the 3D scanner. (ii) In [10], the projection plane is tilted around x-axis for searching best position of the nose around the symmetry plane in y-direction to calculate an estimate of angle around x-axis whereas in our study, we find an angle around x-axis geometrically by detecting the angle between vertical symmetry plane and y-axis of world coordinate system. (iii) The study [10] generates range images to find symmetry plane and angles of rotation around y and z-axis such that projection plane becomes perpendicular to the symmetry plane. Our proposed algorithm finds horizontal nose plane using which, the rotation angle around y-axis is determined. The proposed methodology finds the measure of rotation around z-axis by localizing landmarks on inner eye corners of range images in xy-plane.

2.2. 3D face recognition

A thorough survey of 3D face recognition research is given in [2]. Review of the 3D face recognition approaches from the perspective of their role in a fusion scheme is presented in the following section.

The study [18] proposed the use of two feature vectors based on extraction of six, four-level magnitude sub regions using dual-tree complex wavelet transform and six magnitude sub images. In both cases features were combined into a large matrix prior to LDA based subspace learning. In the testing phase Cosine distance was used for establishing two similarity matrices. Then similarity matrices were fused using sum rule and nearest neighbor classifier was used for recognition. The authors tested their algorithm using FRGC v2.0 database and obtained identification rate of 98.6% for neutral faces. In the study [20] a 2D–3D feature level fusion algorithm is proposed based on PCA using canonical correlation analysis (CCA) to match a 2D or 3D probe face image with a fused gallery face image which contains features from both of 2D and 3D modalities. The authors employed a database of 115 subjects having neutral and expressive faces using Euclidean distance classifier for the experiments and achieved 55% and 85% recognition accuracy using CCA alone and CCA–PCA based fusion algorithm respectively. Another prominent example of fusion of 2D images and 3D depth maps is presented in [21] where sum rule was used and significantly better results were obtained after fusion of individual modalities. The authors evaluated the performance of the system based on Elastic Bunch Graph Matching (EBGM) and subspace methods such as PCA, LDA and Nonnegative Matrix Factorization (NMF) and reported 89% recognition accuracy using a database of 261 subjects. The study [22] presented a face recognition approach based on fusion of three parallel face recognizers using sets of facial contour features extracted around nose tip and inner eye corners. The authors fused the three face recognizers at the feature and match score level to boost the performance of resulting unified face recognition system. They achieved 99.09% rank-1 recognition rate using an already roughly registered database of 116 subjects. A study based on fusion of many overlapping face regions is presented in [10] equipped with an approach in which 60 different face regions were selected and fusion was realized using majority voting. The region classifiers employed PCA–LDA method for feature extraction and likelihood ratio as matching score respectively. The algorithm has been evaluated on the FRGC v2.0 database and rank-1 recognition rate of 99% has been reported in this study. A recent region based study [9] presented a technique to deal with occlusions covering the facial surface. The authors considered two problems which are occlusions and missing data handling. In the classification stage, they employed a regional scheme to fuse the results using the product rule. On the neutral subsets from two databases with realistic facial occlusions, the authors obtained 100% identification accuracies whereas occlusion, expression and pose subsets were reported to have relatively lower identification rates in the same study. Our study is focused on 3D face recognition in the presence of varying poses between probe and gallery images. The studies [9,10,18] achieved good performance in 3D face recognition by using a region based fusion approach considering 40, 60 and 6 face regions respectively, whereas the study [22] presented fusion of 3 parallel face recognizers. Our proposed ensemble classifier based approach takes into account fifteen overlapping regions composed of three sets of five regions each and it targets to explore contribution of overlapping regions toward providing complementary information using fusion. The recognition results of our proposed method are comparable to the best reported results of state of the art studies [1,17,18,23] whereas it is computationally less expensive.

3. Proposed methodology

The framework of our proposed methodology is shown in Fig. 1 which presents whole 3D face recognition system. The face registration, preprocessing, feature extraction and classification stages are depicted in the same figure.
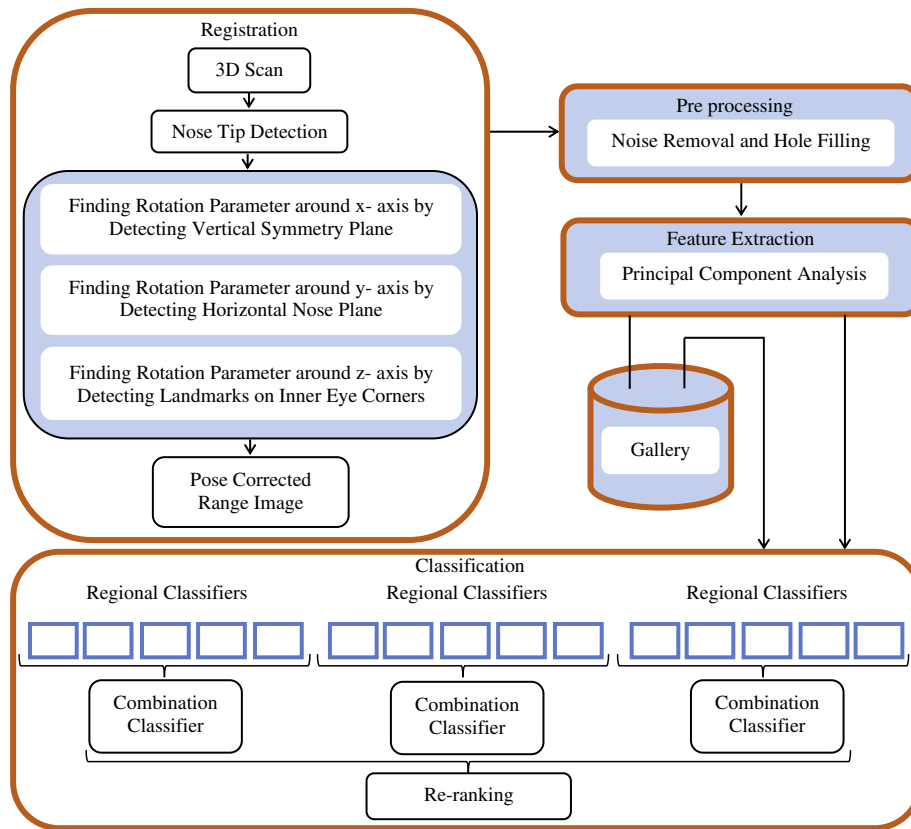
Fig. 1. Block diagram of proposed 3D face registration and recognition algorithm.

3.1. 3D face registration method

For 3D face registration, world coordinate system is represented by *x*, *y* and *z*-axis as in Fig. 2 with origin defined at nose tip *O*. The figure shows a model face [13] and intrinsic coordinate system defined by *a*, *b* and *c*-axis at the same origin. Three angles and an origin are necessary to define an intrinsic coordinate system. Origin is the nose tip and three angles defining the intrinsic coordinate system are *α*, *β* and *γ*. Angle *α* is inscribed between nasal bridge slope represented by *b*-axis of intrinsic coordinate system and *y*-axis of world coordinate system in *yz* plane. *β* is the angle between *c*-axis and *z*-axis of intrinsic and world coordinate systems respectively describing face rotation in *xz* plane. Similarly *γ* is the angle between *a*-axis and
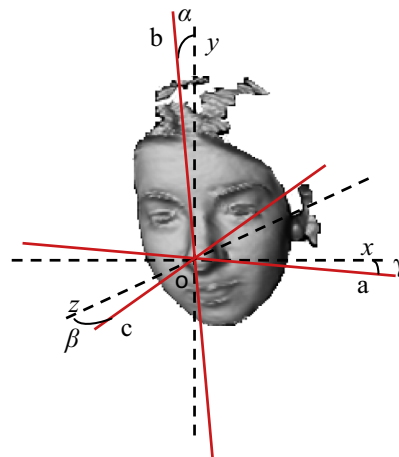

Fig. 2. World and intrinsic coordinate systems with angles *α*, *β* and *γ* in *yz*, *xz* and *xy* planes respectively.

x-axis of intrinsic and world coordinate systems in *xy* plane. Angle *α* gives estimation of face rotation around *x*-axis whereas angles *β* and *γ* contribute toward rotation estimation around *y* and *z*-axis respectively. After finding angles *α*, *β* and *γ*, each point cloud is rotated using these parameters and transformed so that its intrinsic coordinate system axes become concurrent with the world coordinate system axes. This process brings 3D scans in frontal position. The proposed algorithm finds nose tip, vertical symmetry plane and horizontal nose plane to determine angles *α*, *β* and *γ* and it is robust for neutral as well as expressive faces due to stability of landmark structures.

3.1.1. Face localization and nose tip detection

Face of an individual is localized and cropped by using nose tip heuristic. In our methodology, first of all nose tip of 3D scans is found. Nose tip is the point which is nearest to the 3D scanner when scans are captured. While finding nearest point there exist a number of problems due to noise scenarios. In many subjects hairs on forehead come nearer to the scanner instead of nose. Similarly, in case of female subjects, hairs which are spread around neck or ears become nearer to scanner instead of nose. In order to cope with such noise problems due to hair, the algorithm searches nose tip in an approximate Region of Interest (ROI). Similarly subjects having expression known as "puffy" in literature (Fig. 3(a)), present another problem. In such subjects the nearest point may lie on lips instead of nose as shown in Fig. 3(b). To deal with this issue, first of all each 3D point cloud is rotated around *x*-axis using rotation matrix given in Eq. (15) with *α* = 15°, *β* = 0° and *γ* = 0° for all scenarios. This process initially brings the nose tip nearest to the scanner (see Fig. 3(c)). After finding the nose tip, original point cloud is utilized for further processing. Our proposed nose detection method does not require any manual intervention and is fully automatic.

3.1.2. Horizontal nose plane detection

The rotation parameter *β* is determined using horizontal nose plane of face. For finding this plane, *x*, *y* and *z* coordinates of nose tip at origin, *O(x, y, z)*, are determined. All values of *x* are found in *xy* plane at the nose tip by fixing *y*-coordinate of the 3D point cloud. This process extracts all feature points lying parallel to *x*-axis passing through the nose tip. These feature points constitute horizontal nose plane which passes through face region at the nose tip horizontally as shown in Fig. 4(a). On the horizontal nose plane, where it touches the face, three points are found at equal distances on both sides of nose. Averages of these points at both sides are taken separately to compensate for outliers and are labeled as *B* and *Q* in Fig. 4(a).

To find angle of rotation *β* around *y*-axis <*AOB* and <*POQ* are determined by the algorithm. To describe significance of finding <*AOB* and <*POQ* in determining *β*, following mathematical reasoning is presented.

In Fig. 4(b), $\overline{AP}$ defines *x*-axis while $\overline{oO}$ defines *z*-axis. $\overline{OB}$ and $\overline{OQ}$ represent the lines joining origin and points *B* and *Q* which are determined by the proposed algorithm. Arc $\widehat{BQ}$ is the path along which nose tip can move and point *o* represents center of the arc. $\overline{BQ}$ is the line joining points *B* and *Q* through arc center *o*. The arc $\widehat{BQ}$, radius $\overline{oO}$ and angle *β* are related by the following equation.

$$s = r\beta \tag{1}$$

where *s* is arc length, *r* is radius of the circle of which the arc is a part and *β* is the angle subtended by end points of the arc.

Finding *β* gives

$$\beta = \frac{s}{r} \tag{2}$$

To align the nose tip with the point where *x* = 0 in *xz* plane, following objective function is minimized.

$$\text{Objective Function} = \frac{s}{r} \tag{3}$$

Referring to Fig. 4(b), it can be observed that

$$\overline{AP} \| \overline{BQ} \tag{4}$$

and

$$\overline{OB} \| \overline{ST} \tag{5}$$

Therefore

$$< AOB = < BoT = < SoQ = \beta_1 \tag{6}$$

Referring to Eq. (4)

$$\overline{OQ} \| \overline{UV} \tag{7}$$

Therefore

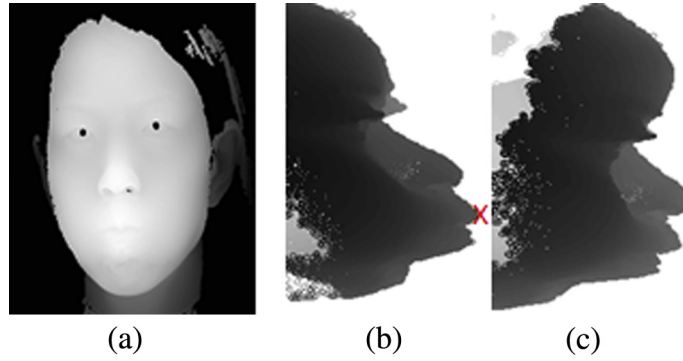$$< POQ = < QoV = < UoB = \beta_2 \tag{8}$$

Fig. 3. Puffy face of subject 04202d564.abs. (a) *xy* view, (b) *yz* view, (c) *yz* view at angle 15°.

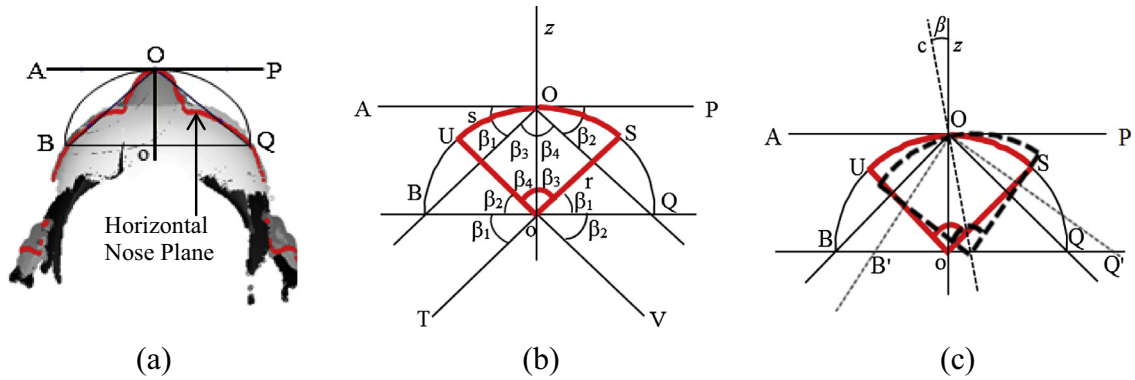**Fig. 4.** Subject 02463d550.abs (a) range image, (b) frontal and (c) frontal and rotated view face geometry in *xz* plane.

As < *BoO* and < *QoO* are right angles, therefore

$$< OoU = 90° - \beta_2 = \beta_4 \tag{9}$$

and

$$< OoS = 90° - \beta_1 = \beta_3 \tag{10}$$

Subtracting Eq. (10) from Eq. (9)

$$< OoU - < OoS = \beta_1 - \beta_2 = \beta_4 - \beta_3 \tag{11}$$

After measuring angles, <*AOB* and <*POQ*, their difference (hence difference of angles β_3 and β_4) is calculated and half of the value of this difference angle is our required angle β. The reason of taking half value lies in the fact that decreasing 5° from <*AOB* increases 5° in <*POQ* and vice a versa, e.g. if <*AOB* = 45° and <*POQ* = 35° then half of difference of angles turns out to be 5°. Thus decreasing 5° from <*AOB* sets <*AOB* = 40° and an increase of 5° in <*POQ* adjusts its value equal to 40°. To give the reader a better understanding of angle β, it is drawn in Fig. 4(c) between *c*-axis and *z*-axis of intrinsic and world coordinate systems respectively, where *c*-axis represents a rotated view and *z*-axis shows a frontal view. It has been discussed in Section 3.1 and Eq. (3) that minimizing value of angle β to zero, folds *c*-axis of intrinsic coordinate system to become concurrent with the *z*-axis of world coordinate system thus correcting the face pose in *xz* plane.

3.1.3. Vertical symmetry plane detection

Vertical symmetry plane is significant because it is used to find angle α. Angle α brings faces which are leaning forward or backward in a position so that they may be treated as frontal. The pivot for finding vertical symmetry plane is nose tip. At the nose tip, all points along *y*-axis are found by fixing *x*-coordinate of the 3D point cloud. This way all the points are found which have a fixed value of *x* but values of *y* are varying. By finding corresponding *z*-coordinates against all selected *y*-coordinates, vertical symmetry plane is constructed as shown in Fig. 5(a). For finding α, 3D point cloud is plotted in *yz* plane as shown in Fig. 5(a), where $\overline{yy}'$ represents *y*-axis and $\overline{zz}'$ represents *z*-axis. From the nose tip a point *N* is determined on Nose Bridge at the vertical symmetry plane. Thus a central angle α, measured by the algorithm, is subtended by end points of arc *l* at the nose tip as shown in Fig. 5(a). Radius ρ, arc *l* and angle α are related as.
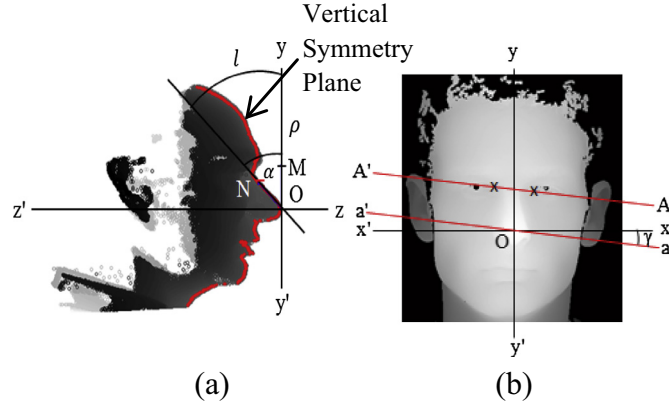
Fig. 5. Range images (a) subject 02463d556.abs in *yz* plane, (b) subject 04201d374.abs in *xy* plane.

$$l = \rho\alpha \tag{12}$$

where l is arc length, ρ is radius of the arc and α is the angle inscribed by end points of the arc. Solving for α

$$\alpha = \frac{l}{\rho} \tag{13}$$

The objective function given in Eq. (14) is maximized to adjust the head tilt at 30° because this places the face in an upright position, resulting in a frontal view [10].

$$\text{Objective Function} = \frac{\rho}{l} \tag{14}$$

The third angle γ rotates the face in *xy* plane. This angle is measured by localizing landmarks on inner eye corners of the face as shown in Fig. 5(b). A line *A′A* is drawn to join landmarks and the line *a′a* is drawn parallel to *A′A* which passes through the determined nose tip. This process defines angle $\gamma = xOa$ with *x*-axis of world coordinate system to be used as correction parameter in *xy* plane.

The 3D pose of an individual is corrected by substituting values of estimated rotation parameters α, β and γ in rotation matrix [24] given in Eq. (15). The process is repeated until convergence is reached. i.e. angles α becomes 30° while angles β and γ achieve a value equal to zero.

$$A = \begin{bmatrix} \cos\beta\cos\gamma & -\cos\alpha\sin\gamma + \sin\alpha\sin\beta\cos\gamma & \sin\alpha\sin\gamma + \cos\alpha + \cos\alpha\sin\beta\cos\gamma \\ \cos\beta\sin\gamma & \cos\alpha\cos\gamma + \sin\alpha\sin\beta\sin\gamma & -\sin\alpha\cos\gamma + \cos\alpha\sin\beta\sin\gamma \\ -\sin\beta & \sin\alpha\cos\beta & \cos\alpha\cos\beta \end{bmatrix} \tag{15}$$

3.2. 3D face recognition method

The range images have some artifacts like noise and gaps, so the range images are preprocessed before their classification. First of all median filtering is applied for removal of sharp spikes which occur during the face scanning. Afterwards, interpolation is employed to fill the holes on the face followed by low pass filtration at the end to smooth the facial irregularities.

3.2.1. Regional classifier

The face regions cropped as shown in Fig. 6 are classified by PCA based classifier using the MahCos distance metric. MahCos distance is the Cosine distance computed in the Mahalanobis space. MahCos distance between two vectors u and v of image space is explained by Eqs. (16) and (17) as described in [12].

$$S_{MahCos}(u, v) = -\frac{m \cdot n}{|m||n|} = -\frac{\sum_{i=1}^{N}(m_i n_i)}{\sqrt{\sum_{i=1}^{N}(m_i)^2}\sqrt{\sum_{i=1}^{N}(n_i)^2}} = -\frac{\sum_{i=1}^{N}\left(\frac{u_i}{\sigma_i}\frac{v_i}{\sigma_i}\right)}{\sqrt{\sum_{i=1}^{N}\left(\frac{u_i}{\sigma_i}\right)^2}\sqrt{\sum_{i=1}^{N}\left(\frac{v_i}{\sigma_i}\right)^2}} \tag{16}$$

where $m_i = \frac{u_i}{\sigma_i}$, $n_i = \frac{v_i}{\sigma_i}$ and σ_i is standard deviation of ith dimension.

In this case higher similarity means higher score. Thus, the actual distance is calculated as given below.

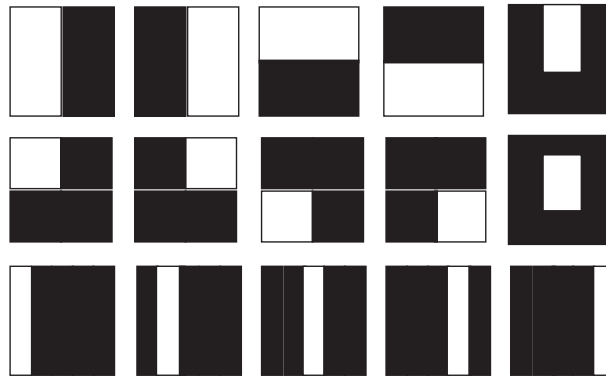$$D_{MahCos}(u, v) = 1 - S_{MahCos}(u, v) \tag{17}$$

Fig. 6. Region creation for fusing results.

3.2.2. Fusion of multiple regions

By the selection of multiple small regions on the face, any error caused by a single region can be compensated by fusing the matching scores from multiple regions, thus making the recognition more robust to artifacts, wrinkles, facial hair or expression variations [25]. Similar methodology has been employed by [10,11]. It has been deduced in this study that small sized regions produce low recognition rates and large sized overlapping regions yield high recognition rates. Large and overlapping regions have also been used in [10]; therefore we explored the fusion of many relatively large overlapping regions. The regions were chosen in such a way that they would allow stable features to compare different types of local variations. Such regions are cropped by leaving out left or right side of the face, which are less visible under large rotations around the *y*-axis. Other cropped regions leave out upper or lower part of the face because of variations in caps, hair, etc. or in expressions of the mouth. The face regions used in this study are shown in Fig. 6 where the five regions in first row are defined as set S_1 while five regions in second and third row are defined as set S_2 and S_3 respectively. White areas show the facial regions that are included in the experiments. The region assemblage densely covers the whole face.

This research work presents a two-tier fusion methodology. The regional classifiers as described in Section 3.2.1 are used for classification based on individual regions. The main motivation of the proposed architecture is to exploit the complementary information given by the regional classifiers based on different regional information input. At the first step, the output from 5 regional classifiers in a row (please see Fig. 6) is fused with a combination classifier that uses WBC method. Then results of combination classifiers for all three region sets are fused at second stage with a re-ranking approach into a single decision. WBC method and re-ranking approach are explained in the following section.

3.2.3. Fusion techniques

There are many ways to fuse the results of a pool of classifiers. Several approaches are given in the studies [10,26] e.g. Dempster-Shafer Theory of Evidence, weighted majority voting and selection of the best combination of classifiers, etc. In the study [10], five levels of fusion are explained as following:

(1) Sensor level fusion: Sensor level fusion fuses raw data from different sensors prior to feature extraction.
(2) Feature level fusion: Extracted features of a subject are combined in this technique.
(3) Rank level fusion: It is fusion of results at rank level. e. g. Borda Count (BC) method.
(4) Decision level fusion: It is the fusion of decisions of the different classifiers producing class labels. e.g. majority voting.
(5) Score level fusion: In this method scores produced by individual classifiers are fused. e.g. sum and product rule.

The BC method [26] is based on the summation of ranks assigned to each of the regions by respective regional classifiers. The image with the lowest rank sum is declared a match. A limitation of the BC method is that, it can be strongly affected by regions which give poor quality contribution toward final result. For example, if each of four regions reports a rank-one match for a person and 5th region reports a 300th rank due to noise or occlusions, then total rank sums to 304 which will probably result in a mismatch.

To overcome this drawback of BC method, WBC fusion method is implemented as the first tier of ensemble classifier which considers only the top ranked matches. For this purpose a generalization of the rank-sum method is presented to transform a set of ranks $\Re$ into the set $f(\Re)$ using a function f which may be any nonlinear monotonically increasing function such that.

$$\Re = \{r, r+1, r+2, r+3, \ldots, R\} \tag{18}$$

where r and R represents rank 1 and rank N results respectively, and

$$f(\Re) = \{f(r), f(r+1), f(r+2), f(r+3)\ldots, f(R)\} \tag{19}$$

In this study, the function $f(r)$ is proposed as

$$f(r) = (s - r)^w \tag{20}$$

where $s = 6$ and $w = 2$. These values of s and w have been chosen empirically because maximum performance was achieved with this selection. The motivation to use such a function f is to penalize the classes at the bottom of a ranked list. Also, higher weight assignments to top ranked classes returned experimentally higher results than the traditional BC method. When a new probe image is to be classified, the regional classifiers contribute a rank toward the ranked list against each of gallery images. The combination classifiers compute the corresponding ranks using WBC method for regions shown in Fig. 6. Thus, sums of ranked lists against each of gallery images are calculated and rank-1 is assigned to maximum value among all of the sums, rank-2 to next lower value of sum and so on. Re-ranking approach is employed at second ensemble stage, to fuse results produced by combination classifiers for all three region sets. In the re-ranking approach, outputs of combination classifiers are fused by assigning an empirically calculated weight to the combination classifiers and a weight to the output rank of the combination classifier. Finally these ranks are re-ordered to produce a ranked list where each ranked label represents a unique fused result. The mathematical description for such re-ranking [26] is expressed below.

$$W_{T_x} = \sum_{i=1}^{c} W_{K_i} \cdot W_{R_{x_i}} \tag{21}$$

where c is the number of combination classifiers, W_{K_i} is empirically calculated weight for a combination classifier, x is the probe image and W_{T_x} is total weight calculated against the probe image x. $W_{R_{x_i}}$ is the weight assigned to a rank from the output of a combination classifier for the probe image x given as:

$$W_{Rxi} = \frac{1}{e^{Rx}} \tag{22}$$

where Rx is the rank of the corresponding matched label. Hence, the farther the matched label is in rank, lesser is its rank weight toward final weight calculation. Final re-ordered ranks are determined with the total weights contributed by each of combination classifiers computed for each of the gallery images against a probe image.
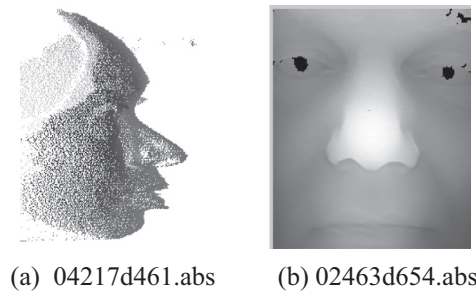
4. Experiments and results

In this section two sets of experiments are reported for each of GavabDB [13] and FRGC v2.0 [14] 3D databases where each set includes one registration experiment explaining implementation of our registration method and two face recognition experiments conducted to evaluate the results of proposed face recognition method based on fusion of classifiers. These databases are briefly described as under.

4.1. FRGC v2.0 database

FRGC v2.0 [14] is a publically available license based database which consists of 50,000 recordings divided into training and validation partitions. It supports 6 experiments, however, we focus on experiment 3 in which gallery and probe sets consist of both the shape and texture images for each subject. Experiment 3 allows for an assessment of the contribution of the shape and textures to the performance of 3D facial imagery [14]. 3D scans are provided in form of four matrices of size 480×640 each, out of which three represent x, y and z coordinates of scans whereas z is the distance from the scanner. The fourth matrix denotes binary representation of valid 3D points in x, y and z matrices. Both male and female subjects with age 18 years and above are included in the database. Facial images of individuals have been acquired with frontal view and minimal pose variations. Almost sixty percent of the subjects have neutral expressions whereas others have expressions of happiness, disgust, surprise, sadness and inflated cheeks or puffy mouth [2]. Images have been captured at varying lengths from the scanner; therefore resolution of the scans is variable. Some of the subjects have occlusions such as hair on face but none of them is wearing glasses [3]. 2D images are also provided in one-to-one correspondence to the 3D scans and the scans have artifacts such as spikes and holes as shown in Fig. 7(a) and (b) respectively. The FRGC v2.0 database contains 3D face scans which are separated in three sets, namely, Spring2003 set (comprised of 943 scans of 277 subjects), Fall 2003 and Spring 2004 sets (consisting of 4007 scans of 466 subjects in total).

4.2. GavabDB database

The GavabDB [13] database is comprised of 427 [17] facial scans from 61 Caucasian subjects among which 45 are male and remaining are female. For each subject there are nine facial scans even with large pose variations and accentuated expressions. The database includes two frontal scans for each subject with neutral expression while another two scans are captured with neutral facial expression where subjects are looking up or down at nearly $+35°$ and $-35°$ respectively. Among the remaining five, three scans present smile, laugh or a random expression and two are right and left profile scans that are acquired at $+90°$ and $-90°$ rotation respectively.

(a) 04217d461.abs   (b) 02463d654.abs

Fig. 7. 3D scan showing (a) spikes, (b) holes.

4.3. Registration results

Experiments are performed to estimate registration parameters using the proposed algorithm. In first step, existing angles describing face rotation in *xz*, *yz* and *xy* planes are found for all subjects. Then required pose correction parameters (angles) are computed and applied to correct the 3D pose of the subjects. A novel method for verification of registration results is proposed in the following section.

Referring to Fig. 8 when a subject's nose tip is not aligned with the line joining *z*-axis of 3D scanner and nose of 3D face scan, it has a larger distance measured from the scanner to nose tip. On the other hand when the subject's nose tip is aligned with the *z*-axis line, its distance is reduced from the scanner. This means when a subject has a non-frontal pose, the distance of nose tip from the scanner has a larger numerical value. But after pose correction, the nose tip distance must be reduced from the scanner as shown in Fig. 8. Using distance formula between two points, the distance of point $O(x_0, z_0)$ to nose tip $O(x_4, z_4)$ in Fig. 8 is given as.

$$d_1 + d_4 = \sqrt{(x_0 - x_1)^2 + (z_0 - z)^2} + \sqrt{(x_1 - x_4)^2 + (z - z_4)^2} \tag{23}$$

The distance of point $O(x_0, z_0)$ to point $O(x_2, z)$ is given as

$$d_2 = \sqrt{(x_0 - x_2)^2 + (z_0 - z)^2} \tag{24}$$

and the distance of point $O(x_0, z_0)$ to point $O(x_5, z_5)$ is given as

$$d_3 + d_5 = \sqrt{(x_0 - x_3)^2 + (z_0 - z)^2} + \sqrt{(x_3 - x_5)^2 + (z - z_5)^2} \tag{25}$$

The points $O(x_1, z)$, $O(x_2, z)$ and $O(x_3, z)$ lie at the same straight line *x′x*. It is clear from Fig. 8 that distance d_2 between nose tip and the point $O(x_0, z_0)$ is a perpendicular to line *x′x* through the point $O(x_2, z)$. The perpendicular distance from a point on a line is always the shortest distance. So all other nose tip distances such as d_1 and d_3 are greater than d_2. Moreover, distances d_4 and d_5 in Fig. 8 further increment the distances d_1 and d_3 respectively. So it is evident that when a 3D scan is rotated from a non-frontal to frontal position such that the nose tip lies perpendicularly on line *x′x*, the distance of nose tip from the
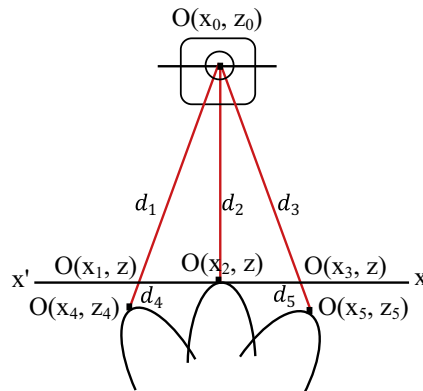
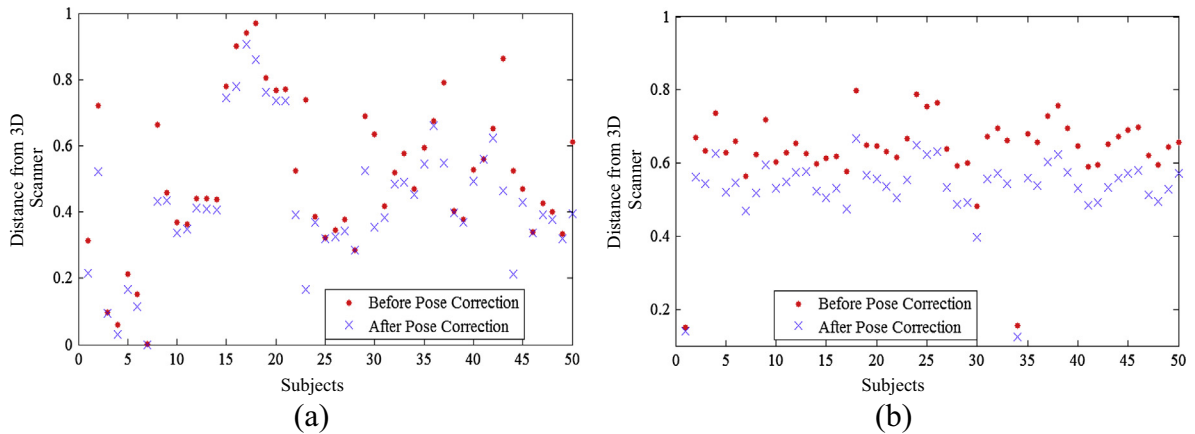**Fig. 8.** Distance between nose tip and 3D scanner.

Fig. 9. Distance of nose tip from scanner for first 50 scans (a) GavabDB: cara1_abajo to cara8_abajo (b) FRGC v2.0:02463d546.abs to 04202d564.abs.
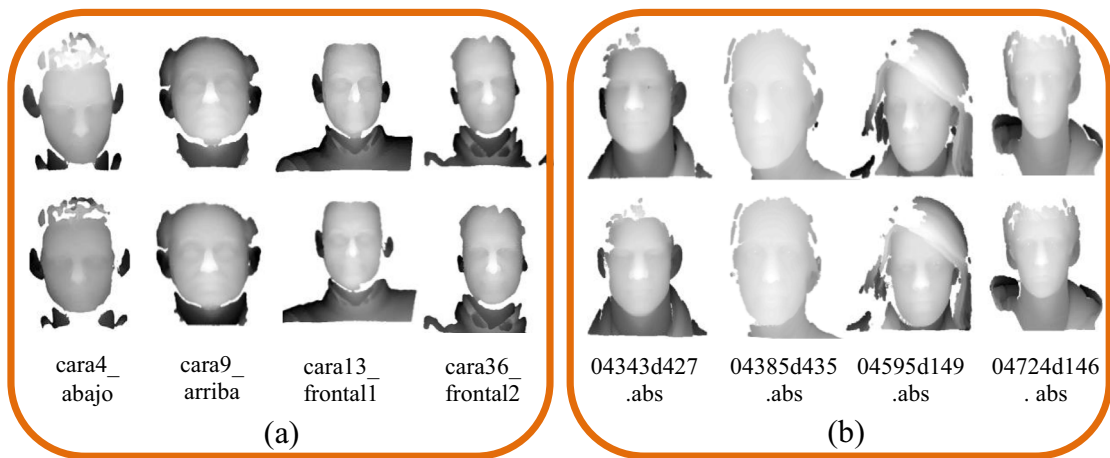

Fig. 10. Example 3D Scans: original (top row) and pose corrected (bottom row) (a) GavabDB (b) FRGC v2.0.

Table 1
Rank-1 recognition rates using GavabDB database. N/A entries present that corresponding studies do not employ region sets.

Author, Year	Gallery size	Probe size	Region set S_1	Region set S_2	Region set S_3	Rank-1 recognition rate (%)
Mahoor and Abdel-Mottaleb, 2009 [17]	61	61	N/A	N/A	N/A	95
Zhang et al., 2014 [1]	183	61	N/A	N/A	N/A	100
Proposed methodology	61	61	90.16%	86.89%	81.97%	95.08
Proposed methodology	183	61	96.72%	91.80%	85.25%	100

scanner is decreased. In other words, when pose of a face is corrected from non-frontal to frontal position, its nose tip distance from the scanner is essentially reduced. Our proposed algorithm achieved 99.77% and 99.95% registration accuracy based on distance reduction using GavabDB and FRGC v2.0 databases respectively.

The nose tip was not detectable for one subject in GavabDB database and two subjects in the FRGC v2.0 database else our registration accuracy would have been 100%. The subjects in GavabDB and FRGC v2.0 databases carry almost all types of facial variations like non frontal pose, several types of facial expressions, varying distance of subjects from the scanner and occlusions. A plot of normalized distances before and after applying our registration algorithm for first 50 scans from GavabDB and FRGC v2.0 databases is shown in Fig. 9(a) and (b) respectively whereas Fig. 10 shows original as well as pose corrected range images of some of the subjects.

4.4. Face recognition results

Recognition is a task where the Facial Recognition System (FRS) is presented a probe image to attempt to match it with a reference image in the gallery. Closed set face recognition experiments have been performed on the GavabDB and FRGC v2.0
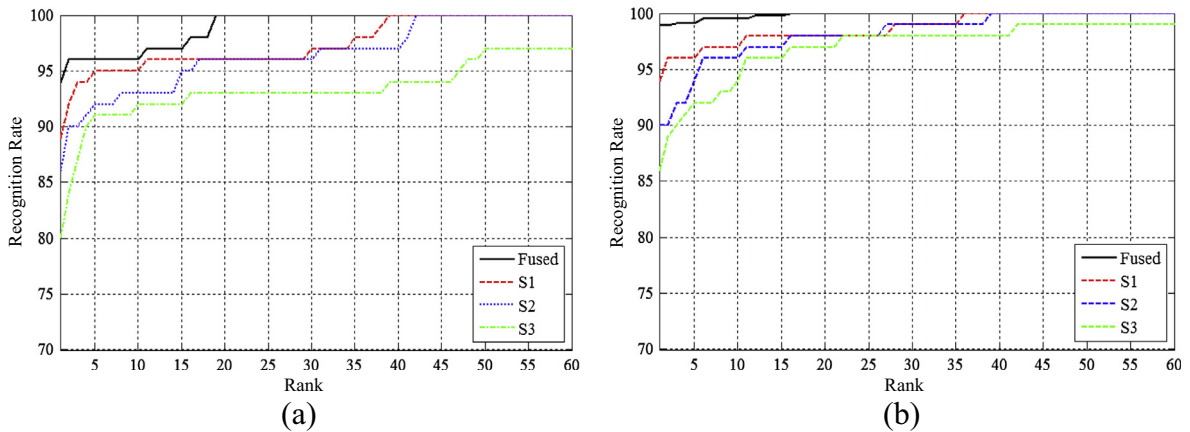
Fig. 11. Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database (a) 1st experiment, (b) 2nd experiment.

Table 2
Rank-1 recognition rates using FRGC v2.0 database. N/A entries present that gallery and probe sizes of corresponding study are not explicitly available.

Author, Year	Gallery size	Probe size	Rank-1 recognition rate (%)
Mahoor and Abdel-Mottaleb, 2009 [17]	370	370	93.7
Osaimi et al., 2007 [23]	466	1944	93.78
Wang et al., 2014 [18]	N/A	N/A	98.71
Proposed methodology (1st Experiment)	466	1944	94.15
Proposed methodology (2nd Experiment)	466	1944	98.93

databases. In closed set recognition, it is guaranteed that for every probe image there is a guaranteed match present in the gallery. The performance of proposed methodology is investigated using the neutral 3D face images. The poses of all subjects are corrected by using the algorithm presented in this study. Two face recognition experiments are conducted using GavabDB database. Experimental protocol of [17] is considered for the 1st experiment where 2nd neutral scan of each subject is enrolled in the gallery and 1st is used as probe. For the 2nd experiment, experimental protocol of [1] is employed where three neutral scans of each subject are included in the gallery and remaining neutral scan is used as probe. It has been explained that our fusion method employs WBC based combination classifiers at tier 1 of classification process. Referring to Section 3.2.2, the facial regions in set S_1 are of large size whereas regions cropped for S_2 and S_3 are of relatively small size. The impact of region size can be seen by relating rank-1 results of combination classifiers for sets S_1, S_2 and S_3 in Table 1. It is concluded that regions of large size lead to better face recognition rates. Re-ranking scheme is used at the second tier of classification process to fuse the results of combination classifiers resulting in 95.08% and 100% rank-1 score for 1st and 2nd experiment respectively.

The FRGC v2.0 contains 466 subjects and almost 2410 point clouds with neutral expression. For some of the subjects, more than two neutral images are captured with a time lapse of one week between them. In the 1st experiment using FRGC v2.0 database, 466 images (one image per subject) are included in the gallery and 1944 are considered as probe. This strategy for splitting gallery and probe sets has been previously employed in the study [23]. The combination classifier for region set S_1 achieved 89.02% rank-1 recognition rate while S_2 and S_3 obtained 86.1% and 80.24% rank-1 recognition rates respectively whereas re-ranking resulted in 94.15% rank-1 score.

In the 2nd experiment, experimental protocol of [18] is considered where 943 scans from FRGC v1 are used in the training phase. The results of combination classifiers for region sets S_1, S_2, S_3 and re-ranking are improved to 94.21%, 90.99%, 85.84% and 98.93% respectively. The improved classification accuracy (in case of GavabDB database as well) comes from the enrollment of multiple training images of known class labels in the gallery as the performance substantially degrades if one gallery image per subject is used in training phase [27]. The cumulative match characteristic (CMC) plots are given in Fig. 11(a) and (b) to depict the behavior of considered classifiers.

4.5. Comparison with other algorithms

Table 1 and 2 compare the recognition results of the proposed methodology to state of the art previous studies which used the GavabDB and FRGC v2.0 databases respectively. Referring to Table 1 the algorithm presented in [1] is based on meshSIFT for facial key points detection with sparse representation based classification and the study [17] presented a 3D face recognition approach based on 3D binary ridge images, principal maximum curvature and ICP matching technique.

Table 3
Computational complexity comparison for image of size $M \times N = 500 \times 500$.

Mahoor and Abdel-Mottaleb [17]				Proposed study
ICP	ICP (fast)	ICP (14% points)	ICP (fast 14% points)	
Multiplications				
$(M \times N) \times (1331 \times 3) \times 370$ $= 369.35 \times 10^9$	$M \log (N) \times (1331 \times 3) \times 370 = 6.63 \times 10^9$	$(M \times N) \times (0.14)^2 \times (1331 \times 3) \times 370$ $= 7.24 \times 10^9$	$0.14 M \times \log (0.14 \times N) \times (1331 \times 3)$ $\times 370 = 0.63 \times 10^9$	$(M \times N) \times 18$ $+ (M \times N) \times 370 \times 5 = 0.467 \times 10^9$
Addition				
$(M \times N) \times (1331 \times 6) \times 370$ $= 738.71 \times 10^9$	$M \log (N) \times (1331 \times 6) \times 370 = 13.25 \times 10^9$	$(M \times N) \times (0.14)^2 \times (1331 \times 6) \times 370$ $= 14.48 \times 10^9$	$0.14 M \times \log (0.14 \times N)$ $\times (1331 \times 6) \times 370 = 1.27 \times 10^9$	$(M \times N) \times 13$ $+ (M \times N) \times 370 \times 5 = 0.466 \times 10^9$

Referring to Table 2, besides the study [17], authors [23] integrated local and global geometric cues for 3D face recognition utilizing PCA and 2D histograms of tensors with Euclidean distance matching while the study [18] employed an approach based on DT-CWT and LDA with nearest neighbor classifier. Our proposed intrinsic coordinate system based registration and ensemble classifier based face recognition algorithm has realized a better performance with rank-1 recognition result starting at 98.93%.

The complexity of the proposed algorithm considering computationally intensive operations involves 1 addition operation for the nose tip detection for each image point whereas for each of puffy face correction and registration processes, 9 multiplication and 6 addition operations are required for each image point. PCA based face recognition algorithm requires $(M \times N) \times 370 \times 5$ multiplication and addition operations each for a gallery of 370 images whereas the multiplication factor of 5 compensates for the overlapping regions (4 half face regions in region set S_1 and 11 quarter face regions in region sets S_1, S_2 and S_3 approximately equal to 5 whole face regions) as given in Fig. 6. Among the algorithms [1,17,18,23] which are compared in Tables 1 and 2, the only available complexity of the study [17] (neglecting initial course registration complexity) is primarily based on MSE distance calculation which utilizes 3 multiplication and 6 addition operations for each of 1331 ($11 \times 11 \times 11$) rotations at $0°$, $\pm2°$, $\pm4°$, $\pm6°$, $\pm8°$ and $\pm10°$ in x, y and z directions assuming that 11 iterations in each of x, y and z directions successfully result in the best geometric alignment.

Referring to Section 2.1, ICP based approaches register each probe image to every gallery image whereas the proposed algorithm registers a probe image to the intrinsic coordinate system only once. The complexity calculations for the proposed algorithm and the study [17] are given in Table 3 which clearly reveals that proposed face registration and recognition algorithm is computationally more efficient than the study [17].

5. Conclusion and future work

We presented 3D face registration and recognition algorithm where the robust, accurate and fast face registration method registers the 3D point cloud to the intrinsic coordinate system. Our algorithm is capable of correcting the 3D poses of neutral as well as expressive faces. Several novelties were contributed in this work explicitly: (i) Method for finding vertical symmetry plane and horizontal nose plane. (ii) Idea of representing reduced nose tip distance from 3D scanner as a measure of pose correction. (iii) Two tier ensemble classifier based face recognition.

The recognition was based on 15 regional classifiers, used as 3 sets of 5 classifiers. By employing two different fusion techniques, namely, WBC and re-ranking, the results of regional classifiers were fused into 3 combination classifiers which were then combined into a final score. In conclusion, firstly we observe that ensemble classifier has clearly outperformed the constituent regional classifiers. Secondly, it is found that fusion methods enhance the performance of regional classifiers using rank based approaches. Thirdly, it is deduced that regions of large size result into better recognition performance than the small cropped regions. Finally, it is evident from results that increasing the number of regional classifiers might not improve the result accuracy of the ensemble classifier. Rather the results of ensemble classifier may get worse by increasing the number of constituent regional classifiers. Our registration method corrected the poses of 99.77% and 99.95% subjects and the ensemble classifier achieved 100% and 98.93% rank-1 recognition rate using GavabDB and FRGC v2.0 databases respectively. It is observed that techniques presented in this study are comparable in both databases.

Future investigation is needed to study the impact of diversity of ensemble classifiers, better performing regions and feature based region level classification techniques for fusing their results into an ensemble classifier.

Acknowledgments

References

[1] Zhang L, Ding Z, Li H, Shen Y, Lu J. 3D face recognition based on multiple keypoint descriptors and sparse representation. PLoS ONE 2014;9(6):e100120.
[2] Bowyer K, Chang K, Flynn P. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. Comput Vis Image Underst 2006;101(1):1–15.
[3] Mian AS, Bennamoun M, Owens R. An efficient multimodal 2D–3D hybrid approach to automatic face recognition. IEEE Trans Pattern Anal Mach Intell 2007:1584–601.
[4] Ojala T, Pietikainen M, Harwood D. A comparative study of texture measures with classification based on feature distributions. Pattern Recogn 1996;29:51–9.
[5] Lades M, Vorbruggen JC, Buhmann J, Lange J, Malsburg Cvd, Wurtz RP, et al. Distortion invariant object recognition in the dynamic link architecture. IEEE Trans Comput 1993;42(3):300–11.
[6] Smeets D, Keustermans J, Vandermeulen D, Suetens P. MeshSIFT: local surface features for 3D face recognition. Comput Vis Image Underst 2013;117:158–69.
[7] Elaiwat S, Bennamoun M, Boussaid F, El-Sallam A. 3-D face recognition using curvelet local features. IEEE Signal Process Lett 2014;21:172–5.
[8] Drira H, Amor BB, Srivastava A, Daoudi M, Slama R. 3D face recognition under expressions, occlusions, and pose variations. IEEE Trans Pattern Anal Mach Intell 2013;35:2270–83.
[9] Alyüz N, Gökberk B, Akarun L. 3-D face recognition under occlusion using masked projection. IEEE Trans Inf Forensics Secur 2013;8:789–802.
[10] Spreeuwers L. Fast and accurate 3D face recognition using registration to an intrinsic coordinate system and fusion of multiple region classifiers. Int J Comput Vision 2011;93:389–414.
[11] Alyüz N, Gökberk B, Akarun L. Regional registration for expression resistant 3-D face recognition. IEEE Trans Inf Forensics Secur 2010;5(3):425–40.

[12] Bajwa UI, Taj IA, Anwar MW, Wang X. A multifaceted independent performance analysis of facial subspace recognition algorithms. PLoS ONE 2013;8(2):e56510.
[13] Moreno AB, Sanchez A. Gavab: a 3d face database. In: Proceedings of the 2nd COST workshop on biometrics on the internet: fundamentals, advances and applications; 2004.
[14] Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW, Chang J, Hoffman K, et al. Overview of the face recognition grand challenge. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR); 2005.
[15] Besl PJ, McKay ND. A method for registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intell 1992;14(2):239–56.
[16] Papatheodorou T, Rueckert D. 3D face recognition. Vienna: I-Tech Education and Publishing; 2007. p. 417–46.
[17] Mahoor MH, Abdel-Mottaleb M. Face recognition based on 3D ridge images obtained from range data. Pattern Recogn 2009;42(3):445–51.
[18] Wang X, Ruan Q, Jin Y, An G. Three-dimensional face recognition under expression variation. EURASIP J Image Video Process 2014;2014:51.
[19] Queirolo CC, Silva L, Bellon ORP, Segundo MP. 3D face recognition using the surface interpenetration measure: a comparative evaluation on the FRGC database. In: Proceedings of the 19th international conference on pattern recognition (ICPR); 2008.
[20] Kamencay P, Hudec R, Benco M, Zachariasova M. 2D–3D face recognition method based on a modified CCA-PCA algorithm. Int J Adv Rob Syst 2014;11(36):1–8.
[21] Zafeiriou S, Atkinson GA, Hansen MF, Smith WAP, Argyriou V, Petrou M, et al. Face recognition and verification using photometric stereo: the photoface database and a comprehensive evaluation. IEEE Trans Inf Forensics Secur 2013;8(1):121–35.
[22] Jahanbin S, Jahanbin R, Bovik AC. Passive three dimensional face recognition using iso-geodesic contours and procrustes analysis. Int J Comput Vision 2013;105:87–108.
[23] Al-Osaimi FR, Bennamoun M, Mian A. Integration of local and global geometrical cues for 3D face recognition. Pattern Recogn 2007;41(3):1030–40.
[24] Slabaugh GG. Computing Euler angles from a rotation matrix. Technical report; 1999. <http://www.gregslabaugh.name/publications/euler.pdf> [accessed 03.09.14].
[25] Faltemier T, Bowyer K, Flynn P. A region ensemble for 3-D face recognition. IEEE Trans Inf Forensics Secur 2008;3(1):62–73.
[26] Bajwa UI, Taj IA, Anwar MW. A unified classifier for robust face recognition based on combining multiple subspace algorithms. Optics Commun 2012;285(21–22):4324–32.
[27] Mohammadzade H, Hatzinakos D. An expression transformation for improving the recognition of expression-variant faces from one sample image per person. In: Proceedings of the IEEE 4th international conference on biometrics: theory applications and systems (BTAS); 2010.

Naeem Iqbal Ratyal received B.Sc. degree in electrical engineering from UAJ&K, Pakistan, in 1998, and M.Sc. degree in electrical engineering from UET Taxila, Pakistan, in 2008. Currently, he is a Ph.D. candidate at Mohammad Ali Jinnah University. His research interests are 3D face recognition, image processing and computer vision.

Imtiaz Ahmad Taj received Ph.D. degree from Hokkaido University, Japan, in 2001. He is presently serving as Professor at Mohammad Ali Jinnah University and supervising research group of Vision and Pattern Recognition Systems. He has authored and co-authored 50+ research publications of international repute. He has supervised four Ph.D. dissertations in the fields of computer vision, pattern recognition and biometrics.

Usama Ijaz Bajwa is currently serving as an Assistant Professor at Department of Computer Science, COMSATS Institute of Information Technology. He did his M.S. & Ph.D. from CASE Islamabad, Pakistan. After completing his M.S., he served as a visiting researcher at Medical Imaging Lab, University of South Wales, United Kingdom. He has authored/co-authored 24 international conference and journal research papers.

Muhammad Sajid received his B.Sc. and M.Sc. degrees in electrical engineering from UAJ&K, Pakistan, in 2002 and 2008, respectively. He is currently pursuing Ph.D. at Mohammad Ali Jinnah University. His main research interests lie in the fields of face recognition, image processing and biometrics.