Abstract—We present an accurate and robust framework for detecting and segmenting faces, localizing landmarks and achieving fine registration of face meshes based on the fitting of a facial model. This model is based on a 3D Point Distribution Model (PDM) that is fitted without relying on texture, pose or orientation information. Fitting is initialized using candidate locations on the mesh, which are extracted from low-level curvature-based feature maps. Face detection is performed by classifying the transformations between model points and candidate vertices based on the upper-bound of the deviation of the parameters from the mean model. Landmark localization is performed by finding the transformation that minimizes the deviation of the model from the mean shape. Face registration is obtained using prior anthropometric knowledge and the localized landmarks. The performance of face detection is evaluated on a database of faces and non-face objects where we achieve an accuracy of 99.6%. We also demonstrate face detection and segmentation on objects with different scale and pose. The robustness of landmark localization is evaluated with noisy data and by varying the number of shapes and model points used in the model learning phase. Finally, face registration is compared with the traditional Iterative Closest Point (ICP) method and evaluated through a face retrieval and recognition framework on the GavabDB dataset, where we achieve a recognition rate of 87.4% and a retrieval rate of 83.9%.

I. INTRODUCTION

The analysis of 3D face meshes is important in many applications, especially in the biometric and medical fields. Such applications aim to accurately relate information from different meshes in order to compare them. For example, the automated comparison of face meshes is desirable in the evaluation of facial surgery interventions and for quantifying and localizing changes in two meshes. A common approach to compare meshes is by a rigid registration, where two or more meshes are fitted in exact alignment with one another. The Iterative Closest Point (ICP) algorithm [1] is a commonly used approach for achieving rigid registration, that is based on the closest point associations from one mesh to the other. However, in the presence of large deformations and outliers, the performance of ICP registration degrades since the global minimum of the cost function often does not correspond to the optimal registration. The ICP algorithm also suffers from the limitation of requiring roughly aligned meshes in order to converge in terms of the mean square error (MSE).

The localization of specific anthropometric locations (landmarks) and regions on face meshes often plays an important part in these applications. Landmarks can aid the ICP algorithm in achieving rough alignment of meshes, and by themselves provide valuable semantic information. In biometric applications, landmarks are often instrumental in the generation of signatures for faces [2]. The accuracy of localizing landmarks is however limited by the prior knowledge of orientation and pose of the faces, and also by the availability of a texture map. The detection and segmentation of faces from meshes is also often an important step prior to landmark localization and registration. Face segmentation is required when a mesh contains more than one person or includes other body parts.

In this paper, we propose a robust framework to accurately detect and segment 3D face meshes, localize landmarks, and perform effective rigid registration in the presence of deformations. The proposed algorithms are based on the fitting of a facial model (PDM) that eliminates the need for prior knowledge of orientation and pose of the meshes and relaxes the constraints on feature map thresholding [3]. The PDM represents the shape of the region of interest on the face. Model fitting is achieved through the transformations between the model points and candidate vertices on the mesh. Candidate vertices include candidate inner eye and nose tip vertices, which are isolated using suitable feature maps that highlight the curvature properties. For face detection, we classify model transformations as face-fit based on their parameters. Landmark localization on the detected face is achieved through refining the model-fit by minimising the shape deviation from the mean shape. The 3D face registration algorithm incorporates prior anthropometric knowledge to achieve fine registration through the detection of suitable landmarks and regions that are then used in the ICP registration. The pipeline of the proposed approach is shown in Fig. 1.

The paper is organized as follows. Section II discusses related works on 3D facial analysis and registration. The databases used in this work are presented in Section III. Section IV describes the creation of the facial model. The proposed approach for face detection and landmark localization is described in Section V, whereas the face registration algorithm is described in Section VI. Section VII discusses the experimental results and the validation of the algorithms. Finally, in Section VIII we draw conclusions.

II. RELATED WORK

In this section we discuss and classify existing approaches on 3D face detection, landmark localization, face registration and statistical models for face analysis.

Colombo et al. [6] perform 3D face detection by first identifying candidate eyes and noses, and by then using the
eyes and nose regions in a classifier. However, the authors highlight that their method is highly sensitive to the presence of outliers and holes around the eyes and nose regions. Most existing approaches ([14], [15]) target face localization, rather than detection, where the presence and number of faces is known. In [14], face localization is performed by finding the nose tip and segmentation is done through a cropping sphere centered at the nose tip. This approach is highly restrictive to the database used, as each input mesh is assumed to contain only one frontal faces. Moreover, the cropping sphere has a fixed radius over the entire database and hence the segmentation is not robust to scale variance. In [15], 3D point clustering using texture information is performed to localize the face. This method relies on the availability of a texture map and the authors state reduction in stability with head rotations greater than ±45° from the frontal pose.

Once faces are detected and segmented, landmark localization is often used for face analysis. Many existing approaches rely on accurately locating corresponding landmarks or regions to perform a rough alignment of meshes. In [13] a curvature based "surface signature" image is used to locate salient points for the estimation of the rigid transformation. Shape-based matching with geodesic interpolation is used in [16] to localize corresponding points on surface pairs. In both these approaches, the localized points are not anthropometrically specific and the correspondence may not be accurate if deformations are present in one of the surfaces. Accurate landmark localization is also important for multi-modal face recognition. Kakadiaris et al. [17] perform face recognition with an annotated model that is non-rigidly registered to face meshes through a combination of ICP, simulated annealing [18] and elastically adapted deformable model fitting [19]. Rigid alignment is achieved with the use of spin images [20] but constraints are imposed on the initial orientation of the face. An attempt to overcome the limitation of facial expressions in recognition is presented in [5], through the matching fusion of multiple face regions. The localization of the regions here is once again imposed with constraints on orientation and rigid thresholds for curvature features. The dependence on prior knowledge of feature map thresholds, orientation and pose is evident in most existing methods for landmark localization on meshes ([12], [21], [22]). Some recent approaches ([23], [24]), rely on multimodal data (3D+2D) for localizing landmarks, however texture information is not always available. A multimodal feature extraction approach is presented in [25] where the nose tip is extracted using 3D data, and then other feature points are localized with 3D+2D information. The localization accuracy here is highly dependent on the initial localization of the nose tip and the availability of a texture map.

The ICP algorithm is widely used for face registration and there has been many attempts to overcome its limitations mentioned in the previous section. Ayyagari et al. [4] present a method for automatic registration of 3D face point sets through a Gaussian field criterion. While this method overcomes the limitation of the undifferentiable ICP cost function and the requirement for prior rough alignment, the parameter of the Gaussian criterion needs to be tuned to reach an appropriate trade-off between residual error and convergence. Other global registration methods exist ([7], [8], [11]), some of which use the ICP, but are inappropriate when large local deformations (due to facial expressions or medical conditions) and outliers (due to the acquisition process) are present. Many variants to the original ICP algorithm have been proposed to improve speed and convergence ([26]–[28]), but without removing the limitation of requiring roughly aligned surfaces. In Table I we summarize and compare existing methods for landmark detection and registration.

In this work we use a statistical shape model to overcome the above mentioned limitations. Statistical models such as Active Shape Models (ASM), Active Appearance Models (AAM) and 3D Morphable Models (3DMM) are extensively used for face analysis ([29], [30]). The shape model used in these approaches, called Point Distribution Model (PDM), aim to perform image interpretation using prior statistical knowledge of the shape to be found. In AAMs, texture information is also modeled and associated with the corresponding shape model. 3DMMs are closely related to AAMs where a 3D model is used to estimate the 3D parameters in a 2D image and to recognize and segment objects ([30], [31]). The PDM has...
only recently been adapted for 3D volumetric data ([32], [33]) and reconstruction of 3D meshes [34]. However, their use with polygonal meshes is largely limited due to the fact that PDMs are mainly used with the associated texture model, which is not always available. Hutton et al. [35] use a hybrid of the ICP and ASM fitting to achieve non-rigid registration of a dense surface model on 3D faces. While this method does not require texture, it still imposes constraints on the initial orientation of the face and is not scale invariant.

### III. DATABASES

In this work we use 3,339 3D scans from 5 databases, namely the BU-3DFE database [36], the GavabDB [37], the NTU 3D Model Database ver.1 [38], the AIM@SHAPE Shape Repository\(^1\) and an in-house dataset obtained by optical surface scanning [39].

In particular, the overall dataset includes 2,934 polygonal face meshes from the BU-3DFE, GavabDB and in-house databases, and 405 non-face objects (to be used for the evaluation of face detection) from the NTU-database and the AIM@SHAPE repository. The face meshes have varying expressions and intensities of expressions, and varying degrees of noise (outliers and holes), facial deformations and presence of other body parts.

The BU-3DFE database contains face meshes from 100 individuals (56 females and 44 males) with 25 face meshes per person. This is a face expressional database where the 25 meshes per person corresponds to five expressions (anger, disgust, fear, happiness, sadness and surprise) with four degrees (intensities) of expressions and one neutral expression. This database is provided with ground-truth of annotated landmarks for each face, which include anthropometric landmarks with points interpolated between them, which we used for the training and evaluation of our facial model. Figure 2 shows a sample subject with the range of expressions and intensities, together with sample landmarks.

The GavabDB database consists of 427 face meshes, corresponding to 7 scans of 61 subjects with different poses and facial expressions. The dataset provides a challenging collection of faces with the samples of each subject having varying degrees of noise, holes and presence of other body parts and clothing. Figure 3 shows sample scans from the GavabDB database.

The in-house database mainly consists of scans of patients with facial deformations, whereas the NTU-database contains 3D models of a wide range of objects. The AIM@SHAPE Shape Repository is another growing database of object scans that aims at aiding the modeling, processing, and interpretation of digital shapes. We use 400 scans from the NTU-database and 5 scans from the AIM@SHAPE repository. Figure 4 shows sample object scans from these two databases.

### IV. 3D FACIAL MODEL

We use a 3D facial model based on a PDM to represent the shape of the region of interest that includes the required landmarks, along with statistical information of the shape variation across the training set. With the PDM, we build a parameterized model, \( \Omega = \Upsilon(b) \), where \( \Omega = \{\omega_1, \omega_2, \ldots, \omega_N\} \), with \( \omega_i = (x_i, y_i, z_i) \) representing each landmark. The vector \( b \) holds the parameters which can be used to vary the shape and \( \Upsilon \) defines the function over the parameters. We use \( N \) manually annotated landmarks representing the points of interest, from a training set of \( L \) face meshes, thus resulting in \( L \) training shapes where each \( k^{th} \) shape is a \( 3 \times N \) element vector, \( \Omega^k = \{\omega_{1}^{k}, \omega_{2}^{k}, \ldots, \omega_{N}^{k}\} \).

\(^1\)http://shapes.aim-at-shape.net/index.php
Training shapes are aligned and scaled to the same coordinate frame to eliminate global transformations. Then we perform statistical analysis on the shape variations that are present in the training shapes only. We use Procrustes analysis [40] to align the training shapes to their mutual mean in a least-squares sense via similarity transformations. Procrustes alignment minimizes $D$, the sum of distances of each shape $\Omega^k$ to the mean $\bar{\Omega} = \frac{1}{L} \sum_{k=1}^{L} \Omega^k$, i.e., $D = \sum_{k=1}^{N} |\omega^k - \bar{\omega}|^2$. At each iteration, $\bar{\Omega}$ is scaled such that $|\bar{\Omega}| = 1$.

We then use Principal Component Analysis (PCA) to estimate the variations of the shape cloud, formed by the training points in the $(L \times 3 \times N)$ - dimensional space, along the principal axes of the point cloud. PCA provides an efficient parameterization of the shape model through dimensionality reduction. The principal axes and corresponding variations are represented by the eigenvectors and eigenvalues of the covariance $Z$ of the data:

$$Z = \frac{1}{L-1} \sum_{k=1}^{L} (\Omega^k - \bar{\Omega})(\Omega^k - \bar{\Omega})^T. \quad (1)$$

If $\phi$ contains the $t$ eigenvectors corresponding to the largest eigenvalues, then any shape similar to those in the training set can be approximated using

$$\Omega \approx \bar{\Omega} + \phi b, \quad (2)$$

where $\phi = (\phi_1 | \phi_2 | \ldots | \phi_t)$ and $b$ is a $t$ dimensional vector given by $b = \phi^T(\Omega - \bar{\Omega})$. The value of $t$ is chosen such that the model represents certain proportion (e.g. 98%) of the shape variance, ignoring the rest as noise. The variance of the $i^{th}$ parameter, $b_i$, across the training set is given by $\lambda_i$. The mean shape is obtained when all parameters are set to zero.

We use the BU-3DFE database and the corresponding ground-truth for the training of our facial model. We use 48 ground-truth landmarks from the eyebrows, eyes and nose regions (Fig. 5) provided with the dataset and include an additional landmark at the nose-tip. Our training set used is composed of landmarks from 150 faces, corresponding to all the expressions of 6 individuals, out of the total of 2500 faces of 100 individuals. We retain 98% of the training variance, which corresponds to 45 eigen-modes ($t = 45$), and ignore 2% variation as noise [29]. By varying the first 3 parameters ($b_1$, $b_2$ and $b_3$) separately, we can generate shape examples as shown in Fig. 6. It can be seen that large variations in the shape and scale are regulated by varying the first parameter alone (Fig. 6 (top)). The second parameter mainly affects the shape of the nose (Fig. 6 (middle)), while the third parameter mainly affects the orientation of the eyebrows (Fig. 6 (bottom)).

\[ \text{V. FACE DETECTION AND LANDMARK LOCALIZATION} \]

The statistical information we obtained with the PDM is used to test candidate positions on a new mesh to detect faces, landmarks and facial regions. To fit the model $\Omega$, we isolate candidate vertices on a face mesh using curvature-based feature maps, avoiding the need of a texture map. The inner eye and nose tip areas on a face are normally unique based on local curvature and can be robustly isolated from other vertices. A block diagram of the model fitting algorithm is shown in Fig. 7.

\[ \text{A. Isolation of candidate vertices} \]

In order to characterize the curvature property of each vertex on the face mesh we compute two feature maps, namely the \textit{shape index} and the \textit{curvedness index} [41]. These maps are derived based on the principal curvature values, $\kappa_1(\cdot)$ and $\kappa_2(\cdot)$, at all the vertices of the mesh using differential geometry. The shape index, $\rho$, at a vertex $v_i$, is defined as

$$\rho(v_i) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \left( \frac{\kappa_1(v_i) + \kappa_2(v_i)}{\kappa_1(v_i) - \kappa_2(v_i)} \right), \quad (3)$$

where $\kappa_1(v_i) \geq \kappa_2(v_i)$; $\rho(\cdot) \in [0, 1]$. The feature map generated by $\rho(\cdot)$ can describe subtle shape variations from concave to convex thus providing a continuous scale between salient shapes. However, $\rho(\cdot)$ does not give an indication of the
scale of curvature present at each vertex. For this reason, an additional feature is introduced, the curvedness of a surface. The curvedness of a surface, \( \gamma(v_i) \), at a vertex \( v_i \), is defined as

\[
\gamma(v_i) = \sqrt{\kappa_1^2(v_i) + \kappa_2^2(v_i)}.
\]

The low-level feature maps are computed after Laplacian smoothing that reduce outliers arising from the scanning process. A comparison between feature maps generated with a smoothed and non-smoothed surface scan is shown in Fig. 8 (left-middle).

To reduce the computational overhead through the reduction of outlier candidate vertices, we first decimate the original mesh and then we average the feature maps across vertex neighbors according to

\[
\tilde{\rho}(v_i) = \frac{1}{P} \sum_{p \in \mathcal{P}(v_i)} \rho(v_p),
\]

and

\[
\tilde{\gamma}(v_i) = \frac{1}{P} \sum_{p \in \mathcal{P}(v_i)} \gamma(v_p),
\]

where \( \mathcal{P}(v_i) \) is the set of \( P \) neighboring vertices of \( v_i \).

If \( \tilde{\gamma}(v_i) > \gamma_s \), then \( v_i \) is in a salient high-curvature region. The condition \( \tilde{\rho}(v_i) < \rho_e \) identifies concave regions; while \( \tilde{\rho}(v_i) > \rho_n \) identifies convex regions. We can therefore relax thresholds to segregate candidate inner eye vertices from the nose tip ones. The thresholds \( \gamma_s = 0.1 \), \( \rho_e = 0.3 \) and \( \rho_n = 0.7 \) were found to be adequate for all the 2,934 faces from the BU-3DFE, GavabDB and in-house databases.

We used second-order neighborhoods for feature averaging and a decimation of 80\%. Note that decimation needs to be done after the extraction of the feature maps, otherwise the resulting features would not characterize the original surface (Fig. 8 (right)). Likewise, we perform the neighborhood averaging of the feature maps after decimation, as if it was done before, the consistency of features in a neighbourhood would remain and outlier candidate vertices would not be eliminated. Note that the smoothed and decimated mesh is only used for the isolation of the candidate vertices, whereas the original mesh is used for the PDM fitting. Figure 9 shows examples of scans with isolated candidate vertices: regions in green are candidate nose-tip vertices and regions in red are candidate eye vertices.

B. Face detection

We use the isolated candidate vertices to estimate candidate positions for the fitting of the PDM \( \Omega \). Face detection is performed by classifying model fits as face-fit or non-face-fit, based on the deviation of \( b \). We fit \( \Omega \) onto a new mesh \( \Psi_i \) by performing similarity transformations of the model, estimated
using three control points of the mean shape $\Omega$. The control points are the inner eye points ($\omega_r$ and $\omega_l$) and the nose tip point ($\omega_n$), with $\{\omega_r, \omega_l, \omega_n\} \in \Omega$. A further reduction in outlier candidate combinations is performed at this stage by checking the triangle formed by each combination of 2 inner eye ($\omega_r, \omega_l$) and 1 nose tip ($\omega_n$). A plausible inner eye-nose triangle should be acute angled with

\[
\begin{align*}
    d_{rl}^2 + d_{rn}^2 &> d_{ln}^2, \\
    d_{rl}^2 + d_{ln}^2 &> d_{rn}^2, \\
    d_{rn}^2 + d_{ln}^2 &> d_{rl}^2,
\end{align*}
\]

where $d_{rl}, d_{rn}$ and $d_{ln}$ are the lengths of the sides of the triangle. Figure 10 shows examples of plausible acute angled candidate combinations in white and outlier combinations in yellow.

Plausible combinations of the candidate inner eye vertices and candidate nose tip vertices on $\Psi_i$ are used as target points to transform the model. Next the remaining points of $\Omega$ are moved to the closest vertices on $\Psi_i$. $\Omega$ is then projected back into the model space and the parameters of the model, $b$, are updated. This selective search is performed until a face-fit is found. A face-fit is defined as one that results in all parameters $b_i$ satisfying the condition

\[ |b_i| \leq +d_{m}\sqrt{\lambda_i}, \quad \forall \ i = 1, \ldots, t. \tag{7} \]

where $d_{m}$ is a suitable limit on the allowed standard deviation.

Figure 11 shows an analysis of the deviation of the model parameters with face-fits and non-face-fits. For face-fits, ground-truth from 2350 faces of the BU-3DFE database (that were not used in training the model) was projected into the model space and the parameters $b$ calculated. Non-face fits are based on the model parameters from the model fitting on outlier face candidate vertices. The Equal Error Rate (EER) is at 11.80 standard deviations. To make our algorithm robust to non-faces, we restrict the range in the face search to 3 standard deviations ($d_{m} = 3$). While the ground-truth parameters are seen to deviate above 3 standard deviations, model fitting is still obtained on these faces with a marginal decrease in accuracy. Figure 12 shows two example faces with ground-truth in red and face-fit in green where the fitting was restricted with $d_{m} = 3$. In these two examples, 60% (27 out of 45) ground-truth parameters were found to be greater than 3 standard deviations.

For detecting multiple faces, all candidate vertices within the bounds of a fit model are removed and model fitting is repeated with the remaining candidate vertices. For face segmentation, we place a sphere of radius $r$ at the midpoint of the line joining the nasal bridge and nose tip, and the intersection boundary of the face and the sphere is used to segment the face [14]. To account for different face sizes, $r$ is set in proportion to the size of the fit model. A histogram analysis of the ratios of top and bottom face heights to nose length was performed using 2,927 faces from the BU-3DFE and GavabDB databases. Figure 13 shows the histograms of the ground-truth ratios of top face height $l_t$ to nose length $l_n$ (blue), and bottom face height $l_b$ to nose length $l_n$ (red).
The length of the nose $l_n$ is defined as the Euclidean distance between the nose tip and nasal bridge landmarks. A larger variation in the bottom face height ratio is noticed due to expressions causing jaw movements. To make our algorithm robust to all face sizes, we set the value of $\rho_l = 2.6 l_n$.

Figure 14 illustrates the segmentation procedure with the cropping sphere centered on the nose tip. The colored portion of the face shows the segmented region.

### C. Landmark localization

We perform landmark localization through refining the model fit in the region of a detected face-fit. First, the 2nd-order neighborhood of the control points $\omega_r, \omega_l$, and $\omega_t$ are isolated as shown in Fig. 15. These vertices now form the new candidate vertices that are used for an exhaustive model fitting search. The final model fit is the transformation with the minimum deviation from the mean shape, while respecting the constraint of Eq. (7).

The face detection and landmark localization algorithm is summarized in Algorithm 1.

---

**Algorithm 1 Face detection and landmark localization**

- $E$: Set of candidate eye vertices; $F$: Set of candidate nose vertices
- $x$: number of candidate eye vertices; $y$: number of candidate nose vertices

1. for $i \leftarrow 1, x$ do
2. $\omega_r = E(i)$
3. for $j \leftarrow 1, y$ do
4. $\omega_l = E(j)$
5. for $k \leftarrow 1, y$ do
6. $\omega_n = F(k)$
7. Estimate $\Omega: \min \theta \leftarrow D = |\alpha_r - \omega_r| + |\alpha_l - \omega_l| + |\alpha_n - \omega_n|$
8. $\Omega = \Omega\theta\tau\chi(\Omega)$
9. for $p \leftarrow 1, N$ do
10. $\omega(p) = C_{\omega}(\omega(p))$
11. end for
12. $\Omega = \Omega^{-1}\theta\tau\chi(\Omega)$
13. $b = \phi^0(\Omega - \Omega')$
14. end for
15. end for
16. end for

Transformation satisfying, $-3\sqrt{\lambda_1} \leq b_i \leq +3\sqrt{\lambda_1}, \forall i = 1, ..., t$, chosen as face-fit.

Transformation with minimum $\nu$, where $\nu = \sum b_i$ is chosen as best fit.

---

**VI. FACE REGISTRATION**

We register face meshes based on rigid registration with prior anthropometric knowledge, that utilizes an adaptation of the ICP algorithm. We use the detected landmarks from the PDM fitting to first perform coarse registration of the meshes using three key landmarks $\alpha_r$, $\alpha_l$ and $\omega_n$ (Fig. 17 (left)), by computing the best fit mapping in a least squares sense. The landmarks are then used to segment specific stable regions on the face, that are robust to expressions and facial deformations. Stable regions include the region around the inner eye points $\alpha_r$ and $\alpha_l$, the nasal bridge between these points and around the eyebrow region (Fig.17 (right)), as they are composed mainly of hard tissue. Vertices from these regions are isolated using the localized landmarks of the fitted model in a similar way as illustrated in Fig. 15. These regions are then used to perform fine registration. Figure 16 shows a block diagram of the registration approach.

Once the reference scan and the test scan are registered, we measure the distance between the two meshes to evaluate the accuracy of the registration. To this end, the symmetric Hausdorff distance [43] is used. The Hausdorff distance has been previously used to good effect with 3D meshes as a similarity measure for face recognition [44], [45]. Let $S$ and $S'$ be the two facial meshes and $\partial(v_i, S')$ be the distance between a vertex $v_i \in S$ and $S'$. If we define

$$\partial(v_i, S') = \min_{v_j \in S'} \|v_i - v_j\|^2,$$

then the Hausdorff distance, $\partial(S, S')$, is given by

$$\partial(S, S') = \max_{v_i \in S} \partial(v_i, S'),$$

and the symmetric Hausdorff distance, $\partial_s$, is then given by

$$\partial_s(S, S') = \max[\partial(S, S'), \partial(S', S)].$$

The symmetric Hausdorff distance provides a more accurate estimate of the distance than the traditional one-sided distance computation. Figure 18 shows an illustration of the registration approach.
VI. EXPERIMENTAL RESULTS

In this section we demonstrate the performance of the proposed algorithms for face detection, landmark localization and face registration. Face detection is evaluated with different configurations of models and training on a database of object and face scans, while face segmentation is demonstrated on object scans that include 3D scenes with multiple faces. The landmark localization is evaluated at the variation of the number of shapes and number of model points used in learning the model, and with the addition of noise. The effectiveness of the face registration algorithm is shown through a comparison with the traditional ICP method and an alternative approach to find stable regions for fine registration. The accuracy of the registration is also evaluated through a face retrieval framework.

A. Face detection

The accuracy of the proposed face detector is evaluated on a dataset of 827 meshes of faces and non-faces. The dataset was composed of 427 faces from the GavabDB database and 400 object scans from the NTU-dataset. We varied the allowed parameter limit ($d_m$) from 0 to 5 with different combinations of training sets, $L$, and configurations of model points, $N$. Figure 19 shows the different model configurations used. It was seen that better accuracy was obtained on using a larger training set (Fig. 20 (left)). Lower number of training samples results in a decrease in true acceptance rates, due to the limited shape variability captured by the model. On the other hand, model configurations with lower number of points also result in lower true acceptance rates along with higher false acceptance rates (Fig. 20 (right)). This is because models with lower number of points have a limited face-shape description leading to more false positives. We restrict our model to 150 training samples and 49 points to limit the complexity in training and model fitting. Figure 21 shows the precision, recall and accuracy on varying $d_m$ with the best model configuration ($L=150$, $N=49$). The optimum results were obtained with $d_m = 3$ and $d_m = 3.5$ where all the faces were classified correctly, with only 2 non-faces being detected as faces. Figure 22 shows examples of face detections from the GavabDB database, with the corresponding segmented faces. Note that we are able to detect faces even in the presence of large holes near the eye regions, thus overcoming the limitation presented in [6]. The average runtime for the PDM fitting over the GavabDB database was 121 seconds on a 3.20 GHz Intel Pentium 4 CPU.

Figure 23 (a) shows visual face detection and segmentation

VII. EXPERIMENTAL RESULTS

In this section we demonstrate the performance of the proposed algorithms for face detection, landmark localization and face registration. Face detection is evaluated with different configurations of models and training on a database of object and face scans, while face segmentation is demonstrated on object scans that include 3D scenes with multiple faces. The landmark localization is evaluated at the variation of the number of shapes and number of model points used in learning the model, and with the addition of noise. The effectiveness of the face registration algorithm is shown through a comparison with the traditional ICP method and an alternative approach to find stable regions for fine registration. The accuracy of the registration is also evaluated through a face retrieval framework.

A. Face detection

The accuracy of the proposed face detector is evaluated on a dataset of 827 meshes of faces and non-faces. The dataset was composed of 427 faces from the GavabDB database and 400 object scans from the NTU-dataset. We varied the allowed parameter limit ($d_m$) from 0 to 5 with different combinations of training sets, $L$, and configurations of model points, $N$. Figure 19 shows the different model configurations used. It was seen that better accuracy was obtained on using a larger training set (Fig. 20 (left)). Lower number of training samples results in a decrease in true acceptance rates, due to the limited shape variability captured by the model. On the other hand, model configurations with lower number of points also result in lower true acceptance rates along with higher false acceptance rates (Fig. 20 (right)). This is because models with lower number of points have a limited face-shape description leading to more false positives. We restrict our model to 150 training samples and 49 points to limit the complexity in training and model fitting. Figure 21 shows the precision, recall and accuracy on varying $d_m$ with the best model configuration ($L=150$, $N=49$). The optimum results were obtained with $d_m = 3$ and $d_m = 3.5$ where all the faces were classified correctly, with only 2 non-faces being detected as faces. Figure 22 shows examples of face detections from the GavabDB database, with the corresponding segmented faces. Note that we are able to detect faces even in the presence of large holes near the eye regions, thus overcoming the limitation presented in [6]. The average runtime for the PDM fitting over the GavabDB database was 121 seconds on a 3.20 GHz Intel Pentium 4 CPU.

Figure 23 (a) shows visual face detection and segmentation

VII. EXPERIMENTAL RESULTS

In this section we demonstrate the performance of the proposed algorithms for face detection, landmark localization and face registration. Face detection is evaluated with different configurations of models and training on a database of object and face scans, while face segmentation is demonstrated on object scans that include 3D scenes with multiple faces. The landmark localization is evaluated at the variation of the number of shapes and number of model points used in learning the model, and with the addition of noise. The effectiveness of the face registration algorithm is shown through a comparison with the traditional ICP method and an alternative approach to find stable regions for fine registration. The accuracy of the registration is also evaluated through a face retrieval framework.

A. Face detection

The accuracy of the proposed face detector is evaluated on a dataset of 827 meshes of faces and non-faces. The dataset was composed of 427 faces from the GavabDB database and 400 object scans from the NTU-dataset. We varied the allowed parameter limit ($d_m$) from 0 to 5 with different combinations of training sets, $L$, and configurations of model points, $N$. Figure 19 shows the different model configurations used. It was seen that better accuracy was obtained on using a larger training set (Fig. 20 (left)). Lower number of training samples results in a decrease in true acceptance rates, due to the limited shape variability captured by the model. On the other hand, model configurations with lower number of points also result in lower true acceptance rates along with higher false acceptance rates (Fig. 20 (right)). This is because models with lower number of points have a limited face-shape description leading to more false positives. We restrict our model to 150 training samples and 49 points to limit the complexity in training and model fitting. Figure 21 shows the precision, recall and accuracy on varying $d_m$ with the best model configuration ($L=150$, $N=49$). The optimum results were obtained with $d_m = 3$ and $d_m = 3.5$ where all the faces were classified correctly, with only 2 non-faces being detected as faces. Figure 22 shows examples of face detections from the GavabDB database, with the corresponding segmented faces. Note that we are able to detect faces even in the presence of large holes near the eye regions, thus overcoming the limitation presented in [6]. The average runtime for the PDM fitting over the GavabDB database was 121 seconds on a 3.20 GHz Intel Pentium 4 CPU.

Figure 23 (a) shows visual face detection and segmentation
results on a scene containing the head and neck of three people, with varied orientation and hair coverage. The automatically segmented faces are displayed below. A composite scan with three faces at different scales and the corresponding segmented faces is shown in Fig. 23 (b). Notice the scale invariance of the facial model fitting and subsequent face segmentation. Finally, Fig. 23 (c) shows the face detection results on non-face objects. It is possible to notice that despite changes in scale and facial expressions, the faces are detected and correctly segmented from the rest of the body. A composite scan with three faces at different scales and the corresponding segmented faces is shown in Fig. 23 (b). Notice the scale invariance of the facial model fitting and subsequent face segmentation. Finally, Fig. 23 (c) shows the face detection results on non-face objects. It is possible to notice that despite changes in scale and facial expressions, the faces are detected and correctly segmented from the rest of the body.

B. Landmark localization

To evaluate the accuracy of the landmark localization, we compare the landmarks localized on 2350 faces of the BU-3DFE database with the provided ground-truth. Note that these faces are independent of the 150 faces used in training the model. We measure the model fitting accuracy based on the localization errors of 5 key landmarks, i.e., outer eye points $(\alpha_{or}, \alpha_{ol})$, inner eye points $(\alpha_r, \alpha_l)$ and nose tip point $(\alpha_n)$. To account for different head sizes, the error is normalized by the distance between $\alpha_r$ and $\alpha_l$ in each scan. The mean normalized error for each landmark, across the test set, is estimated along with the standard deviation and used in the final comparison. We introduce a detection failure criterion, wherein if the distance between a landmark and the ground-truth is larger than a certain threshold ($\tau_P = 0.5$), it is deemed to be a failure.

Figure 24 (top) highlights the influence of the size, $L$, of the training set on the overall performance. It can be noticed that the accuracy improves, with a decrease in both the mean and standard deviation of the error, with the use of a larger training set $L$. This is because more shape variability is captured in the model without incurring in over-training. Figure 24 (middle) shows the influence of varying the number of model points $N$. Once again an improvement is seen on increasing the number of model points $N$, as a better description of the shape of interest is captured. The corresponding percentage of failed detections is shown in Table II. The evaluation of the robustness of the proposed landmark detection method can be
seen in Fig. 24 (bottom). The figure shows the influence of additive white noise with variance $\sigma$. It can be seen that the algorithm achieves stable detections even up to $\sigma = 1$, with only a marginal increase in the mean and standard deviation of the error. The 3 control points $\alpha_I$, $\alpha_l$ and $\alpha_n$ are found to be most stable as they are used to estimate the model fitting, while the outer eye points ($\alpha_{or}$, $\alpha_{ol}$) have a higher mean error with a larger standard deviation.

We also analyzed the effect of expressions on the accuracy of landmark localization. Figure 25 shows the mean normalized error and standard deviation with the 7 expressions (Anger, Disgust, Fear, Happiness, Neutral, Sadness and Surprised). The results show that least error was obtained with neutral and surprise expressions, while the highest error was obtained with anger, distress and happy expressions.

We compared our approach with a state-of-the-art method replicating [21], [22] and found an overall improvement of 62% in localization accuracy. The main reasons for the improvement is due to the invariance of the model fitting to pose variations.

C. Face registration

We compare the proposed face registration approach with a registration method based on the traditional ICP, and perform both qualitative and quantitative evaluations. The proposed registration approach will be referred to as S-ICP, while the ICP registration will be referred to as R-ICP.

Sample results are shown in Fig. 26. The R-ICP algorithm samples randomly 200 vertices from across the scan, while S-ICP uses around 100 vertices from the specific local regions for the ICP step. In both methods the ICP step is executed for 50 iterations. The first row of Fig. 26 shows the registration results with two identical meshes, one of which is slightly rotated in the YZ direction (where Y is the vertical axis and Z is the depth axis). In this case R-ICP outperforms S-ICP as there are no local deformations and the global minimum of the distance function corresponds with the best match.

Rows 2 and 3 of Fig. 26 show the matching of two meshes of the same person, taken at different time instances. In the second row, S-ICP clearly distinguish the regions of change around the cheeks and the eyes. This is consistent with the morphological knowledge of the person under study as the cheeks region consists of soft tissue which has changed over time, while the upper regions of the face consist of hard tissue which is known to be invariant. A change is seen in the eye region as the eyes are closed in one scan, and open

![Fig. 24. Normalized distance (error) between automatically detected landmarks and ground-truth landmarks: (Left) mean ; (Right) standard deviation; (Top) comparison with varying L (N=49); (middle) comparison with varying N (L=150); (bottom) comparison with additive white noise with variance $\sigma$ (L=150, N=49)]

![Fig. 25. Effect of expressions on landmark localization accuracy. Normalized distance (error) between automatically detected landmarks and ground-truth landmarks: (Left) mean ; (Right) standard deviation. (Key - AN: Anger; DI: Disgust; FE: Fear; HA: Happiness; NE: Neutral; SA: Sadness and SU: Surprise)]

![Table II

<table>
<thead>
<tr>
<th>N</th>
<th>$\alpha_r$</th>
<th>$\alpha_l$</th>
<th>$\alpha_{or}$</th>
<th>$\alpha_{ol}$</th>
<th>$\alpha_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>12.0%</td>
<td>10.2%</td>
<td>13.6%</td>
<td>14.4%</td>
<td>2.4%</td>
</tr>
<tr>
<td>27</td>
<td>11.6%</td>
<td>9.1%</td>
<td>12.5%</td>
<td>12.5%</td>
<td>1.4%</td>
</tr>
<tr>
<td>37</td>
<td>11.3%</td>
<td>7.9%</td>
<td>12.5%</td>
<td>12.5%</td>
<td>1.4%</td>
</tr>
<tr>
<td>49</td>
<td>9.3%</td>
<td>7.6%</td>
<td>12.5%</td>
<td>12.5%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>
in the other. R-ICP fails due to the presence of these local deformations and outliers. The third row shows the accuracy of using anthropometric information in the registration. Here the person changes pose with an intentional deformation of the left cheek so that the entire mouth region is displaced. The best registration is achieved by S-ICP, accurately highlighting the regions where deformations occurred.

For quantitative analysis, a scan was modified with synthetic deformations and then registered to its original self using both methods. Fig. 27 shows the results with a scan in which cylindrical deformations were made, to partially overlap the semantic regions, with a total MSE = 1.00. S-ICP provides the best match with a measured MSE = 1.10, while R-ICP fails with measured error of 1.31.

We also quantitatively evaluate the accuracy of S-ICP in a face recognition and retrieval framework. To this end, the symmetric Hausdorff distance is used as a similarity measure \( \Delta \). The Hausdorff distance estimation is based on different regions of the face, towards which, the face was divided into 6 regions (\( R_1 \) - \( R_6 \)), separating the forehead, eyes, nose, cheek, mouth and chin areas, as shown in Fig. 28. For a given selection of regions \( S_R \in \Psi \) and \( S'_R \in \Psi' \), the similarity \( \Delta \) is defined as,

\[
\Delta = \max \left[ \sum_{v_i \in S_R} \max [\partial(v_i, \Psi')], \sum_{v'_i \in S'_R} \max [\partial(\Psi', v'_i)] \right].
\]

The performance of S-ICP, with the use of different regions in the similarity estimation, is evaluated in terms of recognition and retrieval accuracies and compared with the ICP registration. Recognition here refers to the accuracy of the retrieved rank-1 identity, while retrieval refers to accuracy of retrieving faces of the same person with most similarity. The retrieval accuracy is measured using the average dynamic precision (ADP) [46]. The ADP is defined as

\[
ADP = \frac{1}{S} \sum_{i=1}^{S} T_i,
\]

where \( T_i \) is the number of true positives, with \( T_i \leq i \), and \( S \) is the scope size which refers to the total number of expected true positives. \( S \) is set to 7 in our experiments since we have 7 samples per person. For example, if for a given query the retrieved results correspond to [1, 1, 0, 1, 1, 0, 1] until rank-7 (where 1 is a true positive and 0 is a true negative), the \( ADP = 1 + 1 + 0.67 + 0.75 + 0.8 + 0.67 + 0.71 = 5.56/7 = 0.794 \).

The GavabDB database is used for this evaluation, as the faces in this database present challenging situations due to the large variations in presence of noise and outliers in the scans of the same person. We use the forehead, eyes and nose regions (\( R_1 \), \( R_2 \) and \( R_3 \)) and their combinations to evaluate the recognition and retrieval performance and avoid the remaining regions as they are most affected by expressions.

Figure 29 shows the rank-1 recognition and retrieval accuracy in terms of the ADP of S-ICP and compares it with R-ICP. We can see that while R-ICP outperforms S-ICP in some region combinations for recognition accuracy, S-ICP performs better in terms of retrieval accuracy. The overall best results were obtained by S-ICP on using regions \( R_1 + R_2 + R_3 \) with recognition accuracy of 87.4% and retrieval accuracy of 83.9%. Figure 30 (a-b) shows the retrieval results with example queries exhibiting mild and exaggerated expressions. In both cases the proposed approach retrieves more true positive within the top 6 ranks, while ICP retrieves only 2 correct samples in

Fig. 26. Change detection with natural deformations: (a) test and reference meshes, (b) ICP based registration, (c) proposed registration

Fig. 27. Change detection with synthetic deformations: (a) test and reference scans. Change detection results based on: (b) ICP registration, (c) proposed registration

Fig. 28. Regions used for similarity estimation in the face recognition and retrieval framework

Fig. 29.
the presence of mild expressions and completely fails in the presence of an exaggerated expression. Figure 30 (c) presents an example of a failure mode, where the proposed approach fails due to the presence of a hole in the region used for registration.

VIII. CONCLUSION

We presented a novel approach to detect and segment 3D faces, localize landmarks and achieve fine registration of face meshes, based on the fitting of a Point Distribution Model. Face detection is achieved by classifying model fits as face fit and non-face fit based on the model parameters. Landmark localization is performed by finding the model fit that minimizes the deviation of the model from the mean shape. The face registration algorithm performs accurate registration of 3D faces with incorporation of prior anthropometric knowledge. The face detection is efficient on evaluation with a database of faces and non-faces, and is also demonstrated on scenes with multiple faces and large scale variations. The performance of the landmark localization approach was evaluated with different parameters, models and number of training samples. The algorithm is effective in the fitting of the model and shows significant improvement over a state-of-the-art approach. The proposed face registration method shows accurate registration of face meshes in the presence of deformations, and outperforms methods based on the traditional ICP. Our future work includes the use of a hierarchical model to overcome the problem of occlusions. We also aim at optimizing the candidate points’ detection, using different geometric shape descriptors, for the isolation of facial features other than the inner eye and nose tip points. This will enable better rejection of outlier candidate positions for the fitting of the model and improve the speed of the algorithm.

ACKNOWLEDGMENT

We would like to acknowledge Dr. Lijun Yin, Department of Computer Science, The State University of New York at Binghamton, and The Research Foundation of State University of New York, for providing the BU-3DFE database.

REFERENCES
