Boosting Local Shape Matching for Dense 3D Face Correspondence

Zhenfeng Fan[1,2], Xiyuan Hu[1,2 *], Chen Chen[1,2], and Silong Peng[1,2,3]

[1]Institute of Automation, Chinese Academy of Sciences
[2]University of Chinese Academy of Sciences
[3]Beijing Visytem Co. Ltd.
{fanzhenfeng2016,xiyuan.hu,chen.chen,silong.peng}@ia.ac.cn

Abstract

Dense 3D face correspondence is a fundamental and challenging issue in the literature of 3D face analysis. Correspondence between two 3D faces can be viewed as a non-rigid registration problem that one deforms into the other, which is commonly guided by a few facial landmarks in many existing works. However, the current works seldom consider the problem of incoherent deformation caused by landmarks. In this paper, we explicitly formulate the deformation as locally rigid motions guided by some seed points, and the formulated deformation satisfies coherent local motions everywhere on a face. The seed points are initialized by a few landmarks, and are then augmented to boost shape matching between the template and the target face step by step, to finally achieve dense correspondence. In each step, we employ a hierarchical scheme for local shape registration, together with a Gaussian reweighting strategy for accurate matching of local features around the seed points. In our experiments, we evaluate the proposed method extensively on several datasets, including two publicly available ones: FRGC v2.0 and BU-3DFE. The experimental results demonstrate that our method can achieve accurate feature correspondence, coherent local shape motion, and compact data representation. These merits actually settle some important issues for practical applications, such as expressions, noise, and partial data.

1. Introduction

Dense correspondence seeks canonical representations of data such that both global and local structures of them can be compared. It can be viewed as a *non-rigid registration* problem [55], which is fundamental in the field of 3D face analysis. While *landmark correspondence* [29, 32] only matches sparse and anatomically meaningful key-points for different faces, dense correspondence matches all points that are sufficient for detailed descriptions of the whole facial regions. Accurate point-to-point correspondence contributes to many applications of 3D faces [18, 41, 5, 42, 66, 56, 39, 60, 28, 24, 13, 14, 33].

However, dense 3D face correspondence remains a challenging task, particularly for data with large expressions. It can be argued that this problem is implicit both mathematically and physically. In the mathematical view, unlike the rigid case, the non-rigid registration problem belongs to a much larger class that has no explicit formulation, and large expressions add another layer of complexity for the registration considering large deformations. In the physical view, while locating landmarks on 3D faces can be guided by the common knowledge of the anatomical structures, correspondence of points on smooth regions (such as the check and forehead) has no solid definition. This also raises difficulties in assessing the correspondence results.

The state-of-the-art literature [2, 35, 46, 28, 37, 63, 62, 65] commonly uses triangle mesh data and template-warping strategies to establish dense correspondence between different faces. The mesh connection relationship (topology) is intrinsically fixed during the correspondence process, which is beneficial for accurate point-to-point correspondence. We adopt the template-warping strategy in this work. First an arbitrary but noiseless face is chosen as the template. After the correspondences of all faces are completed, we use the average face of all neutral faces as the template.

A single-template strategy cannot guarantee good results for large expression variations. To address this problem, some existing works [37, 36, 51] treat the chin as articulate shape and use blendshapes to model open-mouth motions,

*All correspondences should be forwarded to Xiyuan Hu. This work is supported by the National Key R&D Program of China (2017YFC0803505), the Natural Science Foundation of China (61571438), and the Open Project of National Engineering Laboratory for Forensic Science (2017NELKFKT02).

while others [27, 45, 58, 16, 7, 26] use statistical face models to regularize the correspondence process. These methods incorporate some prior knowledge of face and show impressive results capable of modeling expressions. However, this is a chicken-and-egg problem, since the prior knowledge is usually learned from some accurately corresponded face samples.

This study aims at building dense point-to-point correspondence of 3D faces without a prior model. We propose an explicit deformation model to minimize the weighted least-square error for the rigid alignment of some seed points. The seed points are initialized by a few facial landmarks, and are then augmented gradually to boost local shape matching to build overall correspondence. Extensive experiments on several datasets demonstrate that the proposed method can achieve accurate and reliable results.

The main contributions of this work are: 1) we propose an explicit formulation of the deformation process, and it guarantees not only exact seed point matching but also coherent motions of neighboring points; 2) we construct a global-to-local hierarchical registration strategy for accurate localization of seed points.

2. Related work

There are numerous rigid/non-rigid registration methods for 3D shapes in the literature. For a comprehensive overview one can refer to [55, 57, 40]. In this section, we only cover the most related works and mainly focus on 3D faces.

2.1. 2D vs. 3D methods

Most previous methods [42, 28, 2, 27, 37, 51, 45, 7] work directly on the 3D space of facial shapes, since full information of the raw data is kept on this original domain. As alternatives, some other methods [30, 59, 11, 31, 53, 38] find reasonable mappings from the 3D space to canonical 2D domains. In the seminal work of 3D Morphable Model (3DMM) [5, 6], Blanz and Vetter propose to map the 3D face to the 2D cylindrical coordinate, encode both shape and texture features, and use a regularized optic-flow based algorithm for dense correspondence. Despite the efficiency of the 2D methods by reducing the dimensionality, one potential pitfall is that some original 3D shape features are lost. Therefore, we deal with the dense correspondence problem on 3D in this work.

2.2. Modeling shape motions

Correspondence of two faces can be viewed as a deformation process that one moves to the other either locally or globally. Patel and Smith [46] use a thin-plate spline (TPS) warp to build correspondence with the help of some manual annotations. [43, 42, 45] simply regularize the offset of each individual point by enforcing local smoothness constraint. [44, 64] incorporate functional maps into the correspondence process, and guarantee smooth local motion by the low frequency basis of the eigenfunctions of Laplace-Beltrami operators. To sum up, it is universally acknowledged that the registration process should satisfy *coherent local motion*. Amberg *et al*. [2] propose an *optimal step non-rigid iterative closest points* (NICP) algorithm and model the shape motion as locally affine transformation. The NICP is commonly treated as a benchmark method and has many effective variants [45, 16, 42, 17, 35]. Most recently, it has been used to build some well-known statistical facial models, such as the Basel face model (BFM) [47, 25] and the large scale facial model (LSFM) [8, 9].

2.3. Effect of landmarks

Dense correspondence of 3D faces should prioritize landmark correspondence, since landmarks are the most prominent feature points. Landmarks control the holistic facial shape and help to deal with expressions. In the correspondence process, many existing methods [5, 2, 45, 16, 26, 46, 1] use some sparsely corresponded landmarks to guide the overall deformation process to achieve dense correspondence. These landmarks are detected either manually or automatically. Recently, Fan *et al*. [22] define high-entropy points to replace the landmarks, and these points can be considered as denser landmarks. A problem induced by landmarks is that coherent local motion cannot be strictly guaranteed, especially when there is a large gap between the landmark distributions of two faces. A typical case is in registrations of data with extreme expressions considering large deformations, where landmark and closest-point correspondence may contradict with each other. This leads to incoherent shape motions around the landmarks during the deformation process. We propose an elegant formulation for shape motions to solve this problem.

3. Rigid-motion estimation between corresponded point sets using weighted least-square method

Registration of two point sets generally involves a rigid-motion estimation process. Given two corresponded point sets $P = \{p_1, p_2, ..., p_n\}$ and $Q = \{q_1, q_2, ..., q_n\}$ in $\mathbb{R}^3$, the rigid-motion estimation is represented as a constrained minimization problem with respect to optimal rotation $\mathbf{R}$ and translation $\mathbf{T}$:

$$\{\mathbf{R}, \mathbf{T}\} = \underset{\mathbf{R}\in SO(3), \mathbf{T}\in\mathbb{R}^3}{\arg\min} \sum_{i=1}^{n} w_i \left\|(\mathbf{R}p_i + \mathbf{T}) - q_i\right\|_2^2, \quad (1)$$

where $SO(3)$ denotes the space of Givens matrices in $\mathbb{R}^{3\times 3}$ and $w_i (i = 1, 2, ..., n)$ is the weight for each point pair.

The solution procedures are outlined as follows.

1. Compute the weighted centroids of both point sets:

$$\bar{p} = \frac{\sum_{i=1}^{n} w_i p_i}{\sum_{i=1}^{n} w_i},\ \bar{q} = \frac{\sum_{i=1}^{n} w_i q_i}{\sum_{i=1}^{n} w_i}. \quad (2)$$

2. Compute the centered vectors:

$$x_i = p_i - \bar{p},\ y_i = q_i - \bar{q}\,(i = 1, 2, ..., n). \quad (3)$$

3. Compute the $\mathbb{R}^{3\times3}$ covariance matrix $\mathbf{S} = \mathbf{X}\mathbf{W}\mathbf{Y}^T$, where $\mathbf{X} = (x_1, x_2, ..., x_n)$, $\mathbf{Y} = (y_1, y_2, ..., y_n)$, and $\mathbf{W} = \mathbf{diag}(w_1, w_2, ..., w_n)$.
4. Compute the singular value decomposition $\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.
5. Compute the optimal rotation $\mathbf{R}$ and translation $\mathbf{T}$:

$$\mathbf{R} = \mathbf{V}\begin{pmatrix} 1 & & \\ & 1 & \\ & & \det(\mathbf{V}\mathbf{U}^T) \end{pmatrix}\mathbf{U}^T,\ \mathbf{T} = \bar{q} - \mathbf{R}\bar{p}. \quad (4)$$

The rigid-motion estimation process of the well-known *iterative closest points* (ICP) [4, 15] algorithm for rigid object registration can be considered as a special case in which $w_i = 1(i = 1, 2, ..., n)$. Contrary to that of rigid registration, a non-uniform weighting strategy provides better flexibility for non-rigid registration of 3D faces.

4. Proposed method

Dense point-to-point correspondence can be viewed as a non-rigid registration problem and has to be domain specific. When it applies to 3D human face, the common knowledge of one's anatomical structures should be considered. The widely acknowledged facial landmarks, of which the number varies from a few tens to a hundred, are fiducial points for the anatomically meaningful structures. Therefore, the landmarks provide clues for correspondence of the most prominent facial features, particularly for the sense organs. However, we cannot accurately locate the vessels, muscles, and bones on most areas of the skin surface data that are not adjacent to the eyes, mouth, and nose regions. Furthermore, issues on expressions, noise, and partial data should be settled for practical applications. In this section, we introduce an automatic, adaptive, and accurate method for dense correspondence of 3D faces.

4.1. Template deformation using seed points

In a template deformation perspective [5, 2, 45, 16, 26, 46, 1, 22], each specific landmark should dominate over other points in a local region around it. Correspondences of all the landmarks control the holistic deformation and local dominance of each specific landmark controls the regional deformation. The landmark-guided deformation brings up such a fundamental question: *given sparse correspondences of some seed points, how, at least one step further, to extend them to achieve dense correspondences of all points?*

The following conditions define the basic requisites of a desirable deformation for dense correspondence.
i. The registration error should be in a decreasing trend.
ii. The corresponded seed points should be matched exactly.
iii. The neighboring points should have coherent motions.

Condition **i** guarantees that the deformation behaves in a convergent manner, and condition **ii** and **iii** guarantee stable feature correspondence while preserving local structures of the 3D face. In fact, many state-of-the-art methods [1, 43, 2, 34] for non-rigid registration are designed delicately towards these goals. For example, in the NICP [2], these conditions are modeled as three energy terms for distance, landmarks, and stiffness, respectively. These terms are then fed into a common optimization objective function. However, the general frameworks using control (seed) points fail to satisfy condition **ii** and **iii** simultaneously, but rather perform as a trade-off between them. The reason lies in the inconsistent preliminary correspondences of control points and others. To be specific, different laws for the correspondences of landmarks and their surroundings hardly lead to coherent local motions even with large efforts for regularization. Instead, we propose a unified law to meet these conditions.

Suppose P and Q are two dense point sets representing the template and the target face, respectively. We model the deformation to be composed of individual elements for *locally rigid motions*, as

$$p_i \leftarrow \mathbf{R}_i p_i + \mathbf{T}_i\ (i \in P), \quad (5)$$

where $\{\mathbf{R}_i, \mathbf{T}_i\}(i \in P)$ are the optimal rotation and translation for each point. They are estimated by a weighted rigid alignment of the corresponded seed point sets $\tilde{P} \subset P$ and $\tilde{Q} \subset Q$, as

$$\{\mathbf{R}_i, \mathbf{T}_i\} = \underset{\mathbf{R}_i \in SO(3), \mathbf{T}_i \in \mathbb{R}^3}{\arg\min} \sum_{p_j \in \tilde{P}, q_j \in \tilde{Q}} \frac{1}{d_{ij}^2} \|\mathbf{R}_i p_j + \mathbf{T}_i - q_j\|_2^2, \quad (6)$$

where $d_{ij} = \|p_i - p_j\|(i \in P, j \in \tilde{P})$ is the distance between each specified point and each seed point.

Although modeling non-rigid registration as locally rigid motions is not a new idea, we explicitly formulate the correspondence process as deformation guided by sparsely corresponded seed points. We also model the influences of the seed points to be inversely proportional to the squared distances to the specified point, which ensures local deformations guided by the correspondence of each seed point. The merits of this model are elaborated as follows.

- It can be perfectly customized to fit the three conditions above. Firstly, we will boost this deformation model for incremental shape matching (in Sec. 4.3) to meet condition **i**. Secondly, the weight $\frac{1}{d_{ij}^2}$ approaches infinity when

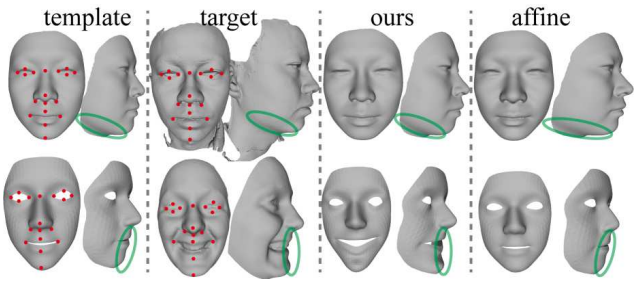

Fig. 1. Comparative results between the proposed deformation and the affine one for the alignment of 3D faces using a few corresponded landmarks (red dots). The bottom and the top row are two different examples with and without expressions, respectively. Pay attention to the areas with green circles.

p_i is a specific seed point ($i = j$). According to the solution procedures in Sec. 3 for Eq. 6, the resulted deformation $\{\mathbf{R}_i, \mathbf{T}_i\}$ will first align the seed point p_i to q_i, and then estimate a compatible rotation $\mathbf{R}_i$ centred on $p_i(q_i)$. Thus condition **ii** holds exactly. Finally, two neighboring points p_i and p_k will result in similar deformations $\{\mathbf{R}_i, \mathbf{T}_i\}$ and $\{\mathbf{R}_k, \mathbf{T}_k\}$ according to Eq. 6, which satisfies condition **iii**.

- It can be considered as an interpolation algorithm based on the correspondence of all the seed points. Compared to some existing methods for dense correspondence, the extrapolation is more robust by imposing rigidness of face structures, which renders its applicability for data with missing parts. Moreover, since the face is a piecewise smooth surface, adaptive augmentation of seed points with large registration errors can recover the overall correspondence in a few iterations. This avoids the need for traversing through all the points, leading to not only robustness to noise but also efficient optimization.
- It is consistent with an intuitive, plausible, but indefinite idea for 3D face dense correspondence: the correspondence of landmarks should be solid and exact, and be prioritized, while the relative locations to the landmarks should be considered for points on the smooth regions.

We depict in Fig. 1 some results of preliminary template deformations via Eq. 5 and 6 using a few initialized seed points (landmarks), compared with those by a common affine alignment [23, 2, 51]. It shows that our model is both more *stable for extrapolation* and more *flexible for large deformation*.

4.2. Augmentation and correspondence of seed points

Seed points are defined as the corresponded control points for the deformation in Sec. 4.1. In this section, we propose two algorithms for adaptive augmentation and accurate correspondence of the seed points, respectively.

Seed point augmentation. We initialize the seed points by some landmarks on the template as shown in Fig. 1. The seed points are then augmented by some newly selected ones step by step, in order to match the corresponded facial surfaces as much as possible. Given a template mesh $\mathcal{S} = (\mathcal{V}_0, \mathcal{E}_0)$ and a target mesh $\mathcal{T} = (\mathcal{V}_1, \mathcal{E}_1)$, where $\mathcal{V}_i$ and $\mathcal{E}_i (i = 0, 1)$ denote the vertices and edges of the meshes, respectively, **Algorithm** 1 gives the details of adaptive seed point selection in each step. It can be summarized briefly as: select points with the largest registration errors but away from each other with a certain distance threshold ρ. The purpose is to enable fast convergence while avoiding redundant local registrations.

Algorithm 1 Seed points selection.

Input:
The template mesh $\mathcal{S} = (\mathcal{V}_0, \mathcal{E}_0)$;
The target mesh $\mathcal{T} = (\mathcal{V}_1, \mathcal{E}_1)$;
Output:
Selected point set $\mathcal{V}_c$;
Compute the distance $d_i (i \in \mathcal{V}_0)$ between each point $v_i (i \in \mathcal{V}_0)$ and $\mathcal{T}$;
Initialize $\mathcal{Q} = (d_i, v_i)(i \in \mathcal{V}_0)$ and $\mathcal{V}_c = \emptyset$;
repeat
  Find $d_{i_0} = \max d_i (i \in \mathcal{Q})$, and include the corresponding v_{i_0} in $\mathcal{V}_c$;
  Remove $\{(d_i, v_i) | \|v_i - v_{i_0}\| \leq \rho\}$ from $\mathcal{Q}$;
until $\mathcal{Q} = \emptyset$
return $\mathcal{V}_c$.

Seed point correspondence. Inspired by some of the ideas on local registrations in [12, 22], we employ reweighted ICP [50] for accurate correspondence of seed points. The points on the template surface are weighted differently as Gaussians of their distances to the corresponding seed point v_0:

$$w_i = e^{\frac{-\|v_i - v_0\|^2}{2\sigma^2}} (i \in P), \tag{7}$$

where w_i is the weight for each point v_i on the template P, and σ is the standard deviation of the Gaussian function. We use five Gaussians with decreasing deviations to register the template to the target, as shown in Fig. 2 (a). Large σ flattens the weights and fits globally, while small σ accounts more for localized shape features. The registration process is proceeded with a cascade manner towards smaller σ. Fig. 2 (b) shows the evolution of closest-point[1] correspondence for a seed point by registrations with different $\sigma (\sigma = \{+\infty, 40, 30, 20, 10\} mm)$, where accurate correspondence of this seed point is finally achieved.

4.3. Boosting local shape matching

We propose to boost the local shape matching between the template and the target face by incremental deformation (in Sec. 4.1) based on the corresponded seed points

[1] Note that the seed point correspondence is not necessarily on the target surface but depends on the registration result.

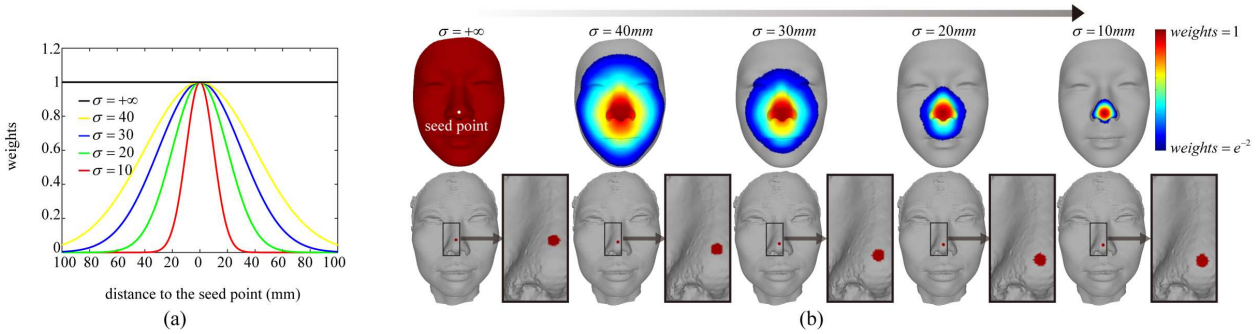

Fig. 2. Different Gaussian functions (a) and the global-to-local registration process (b) for the correspondence of a seed point (white/red dot at the nose tip). The top and the bottom row in (b) show the weighted template and the results on the target, respectively.

(in Sec. 4.2), to finally achieve dense correspondence. The main algorithm is summarized in **Algorithm** 2, where there is a hyper-parameter $nIter$ denoting the iteration times. Ideal stopping criterion should be set according to the registration error, but we fix $nIter = 20$ considering noisy data and computational efficiency in this work. Fig. 3 shows a typical example where the template face is deformed into the target face gradually. We also show together the registration errors, selected seed points, and profile views for the deformation process. This process actually bears a hierarchy structure: it first fits the global shape and then proceeds to boost local shape matching. Note that it is an analogy to the process of wearing a sheet mask on one's face: the unmatched parts of the surface are smoothed gradually.

Algorithm 2 Boosting local shape matching.

Input:
The template mesh $\mathcal{S} = (\mathcal{V}_0, \mathcal{E}_0)$ and some manually selected landmarks $\mathcal{V}_s \subset \mathcal{V}_0$;
The target mesh $\mathcal{T} = (\mathcal{V}_1, \mathcal{E}_1)$;
Output:
Deformed template mesh $\mathcal{S}$;
1: Establish correspondences $\tilde{\mathcal{V}}_s$ on $\mathcal{T}$ for all landmarks $\mathcal{V}_s$ on $\mathcal{S}$ using the reweighed ICP in Sec. 4.2;
2: Deform $\mathcal{S}$ by Eq. 5 and 6 based on the correspondences $\tilde{\mathcal{V}}_s \Leftrightarrow \mathcal{V}_s$;
3: **for** $i \in \{2, 3, ..., nIter\}$ **do**
4: Select seed points $\mathcal{V}_c$ by **Algorithm** 1 and increase the seed points by $\mathcal{V}_s \leftarrow (\mathcal{V}_s \cup \mathcal{V}_c)$;
5: Establish correspondences $\tilde{\mathcal{V}}_c$ on $\mathcal{T}$ for the augmented seed points $\mathcal{V}_c$ on $\mathcal{S}$ using the reweighed ICP in Sec. 4.2, and let $\tilde{\mathcal{V}}_s \leftarrow (\tilde{\mathcal{V}}_s \cup \tilde{\mathcal{V}}_c)$;
6: Deform $\mathcal{S}$ by Eq. 5 and 6 based on the correspondences $\tilde{\mathcal{V}}_s \Leftrightarrow \mathcal{V}_s$;
7: **end for**
8: **return** $\mathcal{S}$.

Implementation details.

- Considering the face scale varies from individual to individual, we normalize the template face by a scaling factor based on a few corresponded landmarks (seed points) between two faces before it is fed into the correspondence process in Step 3 of **Algorithm** 2 . The optimization process is given by

$$\{s, \mathbf{R}, \mathbf{T}\} = \underset{s\in\mathbb{R}^1, \mathbf{R}\in SO(3), \mathbf{T}\in\mathbb{R}^3}{\arg\min} \sum_{p_i\in\mathcal{V}_s, q_i\in\tilde{\mathcal{V}}_s} \|(s\mathbf{R}p_i + \mathbf{T}) - q_i\|_2^2, \tag{8}$$

  which can be solved efficiently by alternative estimations of the rigid motion $\{\mathbf{R}, \mathbf{T}\}$ and the scaling factor s.
- Considering missing parts of the raw target data, we enforce an additional principle into the seed point selection **Algorithm** 1: first exclude the candidate points on the template face whose closest points lie on the boundary of the target mesh.
- Fig. 4 shows two different templates used for database with no expressions and with expressions, respectively. Except for the template differences, we also use geodesic distance instead of Euclidean distance for d_{ij} in Eq. 6 and $\|.\|$ in Eq. 7, to handle eyes and mouth separations. The geodesic distances for all point pairs are computed using a fast heat-flow based method [19].
- The seed point correspondence process is accelerated by some pre-organized data structures [3, 10], as well as trimming and downsampling of data. The template mesh is first downsampled by a factor positively correlated to the Gaussian deviation σ. And then we eliminate the parts on the template mesh whose weights are smaller than e^{-2} as shown in Fig. 2 (b) (the top row). Moreover, we skip fitting by Gaussian weights with large σ in the latter iterations of the boosting **Algorithm** 2, since the global structures of faces are already well aligned.

5. Experiments

Datasets. We use three datasets, including two publicly available ones for the evaluation of the proposed method. 1) One is the **FRGC v2.0** [48] database from University of Notre Dame. It includes a total number of $4,007$ 3D scans of 466 subjects collected during Fall 2003 and Spring

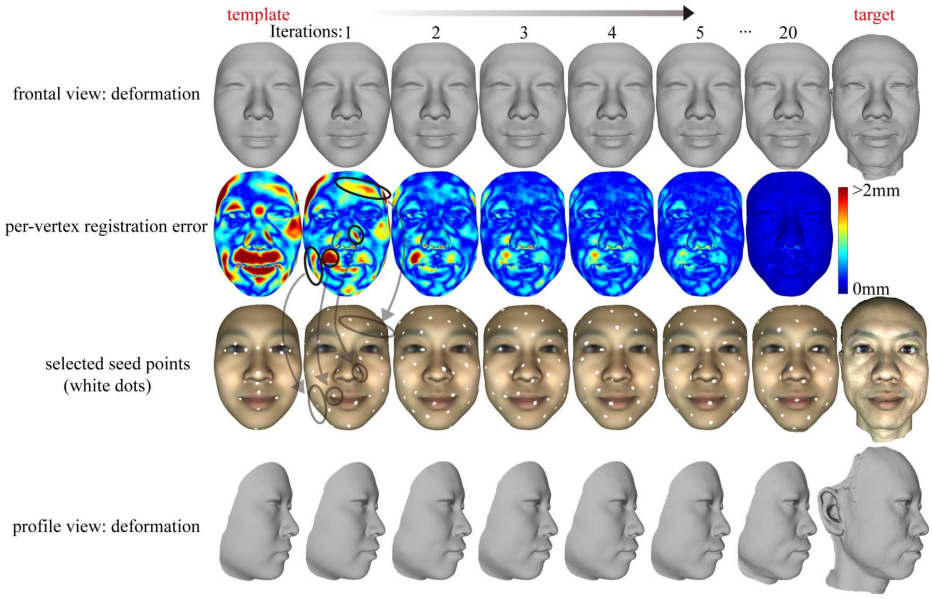

Fig. 3. An example for the dense correspondence process by the proposed method: the template face is deformed into the target face in 20 iterations. The arrows denote that the selected seed points are associated with large regional registration errors.

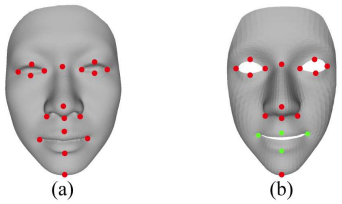

Fig. 4. Two different templates for database with expressions (b) and without expressions (a). The initialized landmarks are marked as dots, where the green dots need manual correspondences for the BU-3DFE dataset.

2004. The point clouds in it have relatively high resolutions for the frontal view compared with the profile views, and most of the scans are in neutral expressions. Many existing works include this database, enabling comparison across the literature quantitatively. 2) Another database is the **BU-3DFE** [61], which is made publicly available in 2006. It includes a total number of $2,500$ 3D scans of 100 subjects with various expressions (neutral, happiness, surprise, fear, sadness, disgust, and angry) on 4 different levels. It is also a benchmark database for 3D facial expression research. The resolutions of the raw data are around $10,000$ vertices per face. 3) In addition, we collect a new database including hundreds of high-resolution 3D faces with a modern structured light device. The collection system merges scans from 4 different directions, and the resolutions are around $300,000$ vertices per face together with pixel-wise textures. The subjects in this database are mostly in neutral expressions under a well-controlled environment.

Parameter setting. We have tested different parameters and finally set $nIter = 20$, $\sigma = \{+\infty, 40, 30, 20, 10\}mm$, and $\rho = 25mm$ in all of our experiments. The setting of σ is trivial if we follow the global-to-local scheme. And larger ρ results in fewer selected seed points but more iterations $nIter$, according to which we have made a trade-off.

Computational time. On a machine with Core-i5-6600k CPU (3.5GHZ, single thread), a MATLAB implementation of the proposed algorithm takes about 100 seconds for a template mesh with $19,334$ vertices. Note that almost all the operations in this algorithm are per-vertex optimizations that are independent to each other. Therefore, a more efficient version of parallel implementation can be developed.

Assessing the result of 3D face dense correspondence is not an easy task since there is no universal rules for the ground-truth in the literature. In this section, we evaluate the proposed method in three perspectives: feature correspondence, structural correspondence, and correspondence on some practical issues.

5.1. Feature correspondence

We define "feature" here as the most significant feature on the face that human can distinguish, particularly for the eyes, nose, and mouth regions. Two experiments are conducted for the evaluation of it.

1) Texture transfer. Some current works [42, 64] paste the textures of some corresponded faces on a template face to assess the results qualitatively. Following them we transfer the textures between different faces. Fig. 5 shows a well matching of features for three corresponded samples in our

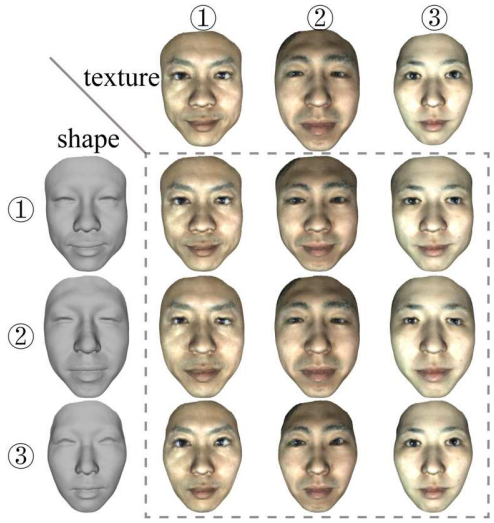

Fig. 5. Texture transfer results by combinations of different facial shapes and textures.

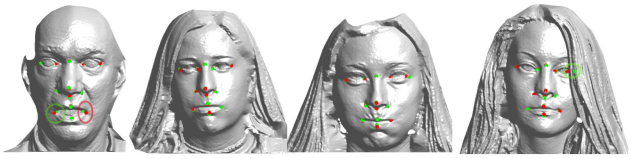

Fig. 6. Some examples for landmark detection on FRGC v2.0. Our results and the manually labeled ones are marked as red and green dots, respectively. And the red circle shows a better result than the manual annotation, while the green circles indicate inaccurate detections.

database.

2) Landmark correspondence. We compare our method with some existing works for the detection of landmarks on **FRGC v2.0**. The ground-truth results are the manual annotations provided by Creusot *et al*. [20]. First we manually label some landmarks on a template, and then we apply the correspondence algorithm in Sec. 4.2. Considering inaccurate manual labeling on the template, we further apply the algorithm to the neighboring points around a specified landmark and replace it by the one with the minimum error. Fig. 6 shows some examples and Table 1 presents the comparative results. Checking the visual results in Fig. 6 carefully, we observe that our method is competitive with the manual annotation owing to the weighted rigid registration of local features. Our method also achieves state-of-the-art performance compared to the most recent works.

5.2. Structural correspondence

The meaning of structural correspondence is two-fold: 1) in a narrow sense we think that the local mesh structures are altered with similar patterns across different faces,

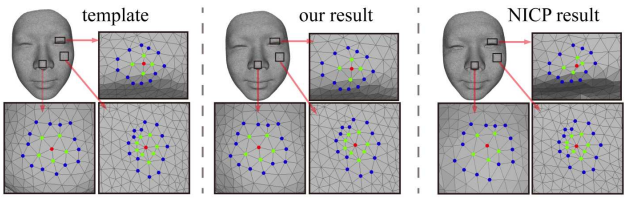

Fig. 7. Comparisons for the detailed mesh structures between the corresponded results by our method and that by the NICP. The 2-ring neighbors of a corresponded point are marked with colored dots for better viewing.

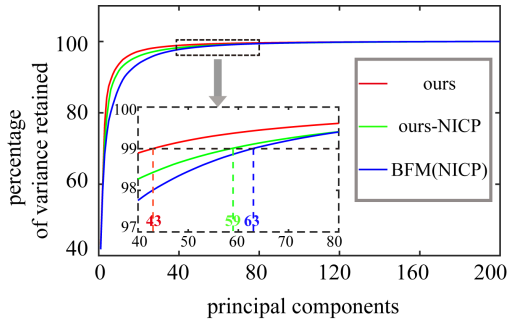

Fig. 8. Compactness of the fitted PCA model by our method compared with that by the NICP and the BFM (also by the NICP).

which is consistent with a universal idea of "coherent local motion"; 2) in a broad sense we think that all facial data as a group should gain certain benefits from the reconstructed canonical representations, and we also view it by the minimum-description length (MDL) principle [49] as used by Davies *et al*. [21] for statistical shape modeling.

In Sec. 4.1, we have discussed a fundamental problem for shape deformation and our formulation provides an elegant answer to it. To view the effect of our formulation, we show a corresponded example in Fig. 7 together with the one reconstructed by NICP [2]. Checking the mesh structures in details, we observe that our result shows similar patterns between the template and the target mesh almost everywhere. In contrast, the NICP cannot guarantee such a coherent result.

We further apply PCA to 200 corresponded samples in our database after Procrustes alignment, following the same routine of 3DMM [5]. Compactness of data is evaluated using the MDL principle: *the less, the better*. We also implement the NICP for these facial samples for comparison. Fig. 8 depicts the percentage of energy preserved by varying the number of principal components, together with that from a publicly available BFM model [47]. It manifests that our method leads to much more compact PCA bases, as the number of principal components to explain 99% energy decreases from $\mathbf{59}(\mathbf{63})$ to $\mathbf{43}$. Note that compactness is a crucial property of data for dimensional reduction.

Table 1. Comparative results of the mean and standard deviation (mean/SD) of landmark localization error (mm) for 4007 images on the FRGC v2.0 dataset. The best and the second results are shown in red and blue, respectively. Symbols for landmarks: Ex/En-outer/inner eye corner, N-nose bridge saddle, Prn-nose tip, Sn-nasal base, Ac-nose corner, Ch-mouth corner, Ls/Li-upper/lower lip midpoint.

Landmarks	Ex	En	N	Prn	Sn	Ac	Ch	Ls	Li
Segundo *et al*. [52]	-	3.5/2.3	-	2.7/1.4	-	5.3/1.9	-	-	-
Creusot *et al*. [20]	5.9/3.1	4.3/2.2	4.2/2.1	3.4/2.0	3.7/3.1	4.8/3.6	5.6/3.5	4.2/3.2	5.5/3.3
Sukno *et al*. [54]	4.6/2.7	3.5/1.7	2.5/1.6	2.3/1.7	2.7/1.1	2.6/1.4	3.9/2.8	3.3/1.8	4.6/3.4
Fan *et al*. [22]	2.6/1.6	2.5/1.7	2.4/1.4	2.1/1.2	-	-	2.9/2.2	2.4/2.9	4.4/3.9
Gilani *et al*. [27]	2.5/1.9	2.4/1.2	2.5/1.5	2.2/1.8	3.4/1.1	3.0/2.4	2.5/1.8	2.4/3.1	3.5/3.7
Ours	2.1/1.9	1.9/1.0	2.4/1.2	1.8/1.2	1.8/0.9	1.9/0.9	2.8/2.5	2.0/2.2	4.3/3.1
Improvement	16%/-19%	21%/17%	0%/14%	14%/0%	33%/18%	27%/36%	-12%/-39%	17%/-22%	-23%/6%

Fig. 9. Examples for problematic data. The top row shows the raw scans with noise (*e.g*. stripe, spike, and Gaussian) and missing parts, and the bottom row shows the fitted results.

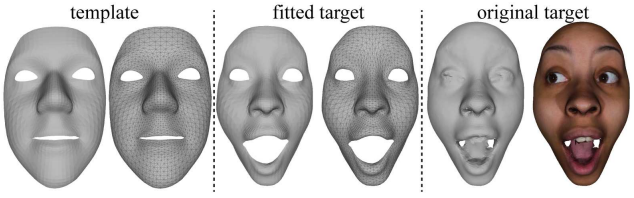

Fig. 10. An example for large expression.

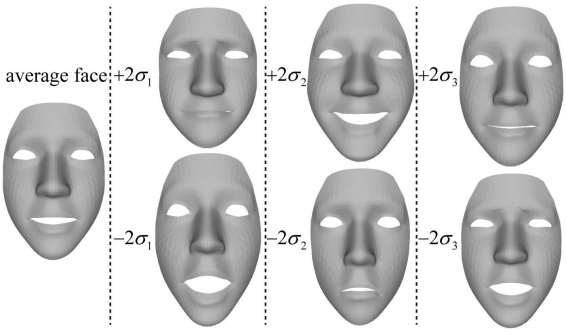

Fig. 11. Variations ($\pm$2SD) of the first three components of the expression-PCA model for the BU-3DFE database.

5.3. Practical issues

The proposed method has been successfully applied to problematic data with large noise and missing parts, as well as data with large expressions. One of the key reasons is that we devise a shape-motion formulation which is both stable for extrapolation and flexible for large deformation. The other reason lies in our strategy for correspondence of seed points: it matches local patches hierarchically and avoids searching for correspondence of each individual point.

Fitting to noisy and partial data. Fig. 9 shows some corresponded results for noisy and partial data. We can see that this method can handle problematic data with different types of noise and large missing regions. Furthermore, it is not difficult to generalize this method to data with occlusions, since occlusions and missing parts are similar problems.

Fitting to expressions. Fig. 10 shows the fitted result (with detailed mesh structure) of a face with an extreme expression in the **BU-3DFE** database. We also apply our algorithm to the whole database and build an expression-PCA model as shown in Fig. 11. Our method is not fully automatic in this case. It involves the manual annotations (guided by textures) of four landmarks around the mouth region as shown in Fig. 4 (b), but only for the first iteration. We suggest that this process can be automated by incorporating the state-of-the-art 2D landmark detection methods based on the texture information. Note that it is even not possible for humans to identify the correct locations of these landmarks based on shapes only for many samples in this database.

6. Conclusion

We present a robust algorithm for dense correspondence of 3D face in this paper. This algorithm models dense point-to-point correspondence as deformation guided by a number of seed points, and the seed points are augmented to boost shape matching step by step, to finally achieve dense correspondence. Extensive experiments on several datasets, and in different perspectives, demonstrate the effectiveness of the proposed algorithm. Since 3D face is representative in the field of 3D shape analysis, it is possible to generalize this algorithm to the common non-rigid registration problem.

References

[1] B. Allen, B. Curless, and Z. Popović. The space of human body shapes: reconstruction and parameterization from range scans. *ACM Transactions on Graphics*, 22(3):587–594, 2003.

[2] B. Amberg, S. Romdhani, and T. Vetter. Optimal step nonrigid icp algorithms for surface registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

[3] J. L. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.

[4] P. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[5] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, pages 187–194. ACM, 1999.

[6] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003.

[7] T. Bolkart and S. Wuhrer. A groupwise multilinear correspondence optimization for 3d faces. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3604–3612. IEEE, 2015.

[8] J. Booth, A. Roussos, A. Ponniah, D. Dunaway, and S. Zafeiriou. Large scale 3d morphable models. *International Journal of Computer Vision*, 126(2-4):233–254, 2018.

[9] J. Booth, A. Roussos, S. Zafeiriou, A. Ponniah, and D. Dunaway. A 3d morphable model learnt from 10,000 faces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5543–5552. IEEE, 2016.

[10] G. Borgefors. Hierarchical chamfer matching: A parametric edge matching algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(6):849–865, 1988.

[11] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *International Journal of Computer Vision*, 64(1):5–30, 2005.

[12] B. J. Brown and S. Rusinkiewicz. Global non-rigid alignment of 3-d scans. *ACM Transactions on Graphics*, 26(3):21, 2007.

[13] C. Cao, Y. Weng, S. Lin, and K. Zhou. 3d shape regression for real-time facial animation. *ACM Transactions on Graphics*, 32(4):41, 2013.

[14] X. Cao, Z. Chen, A. Chen, X. Chen, S. Li, and J. Yu. Sparse photometric 3d face reconstruction guided by morphable models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2018.

[15] Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155, 1992.

[16] S. Cheng, I. Marras, S. Zafeiriou, and M. Pantic. Active non-rigid icp algorithm. In *Proceedings of the IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, volume 1, pages 1–8. IEEE, 2015.

[17] S. Cheng, I. Marras, S. Zafeiriou, and M. Pantic. Statistical non-rigid icp algorithm and its application to 3d face alignment. *Image and Vision Computing*, 58:3–12, 2017.

[18] C. A. Corneanu, M. O. Simón, J. F. Cohn, and S. E. Guerrero. Survey on rgb, 3d, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(8):1548–1568, 2016.

[19] K. Crane, C. Weischedel, and M. Wardetzky. Geodesics in heat: A new approach to computing distance based on heat flow. *ACM Transactions on Graphics*, 32(5):152, 2013.

[20] C. Creusot, N. Pears, and J. Austin. A machine-learning approach to keypoint detection and landmarking on 3d meshes. *International Journal of Computer Vision*, 102(1-3):146–179, 2013.

[21] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. A minimum description length approach to statistical shape modeling. *IEEE Transactions on Medical Imaging*, 21(5):525–537, 2002.

[22] Z. Fan, X. Hu, C. Chen, and S. Peng. Dense semantic and topological correspondence of 3d faces without landmarks. In *European Conference on Computer Vision*, pages 541–558. Springer, 2018.

[23] J. Feldmar and N. Ayache. Rigid, affine and locally affine registration of free-form surfaces. *International Journal of Computer Vision*, 18(2):99–119, 1996.

[24] C. Ferrari, G. Lisanti, S. Berretti, and A. Del Bimbo. A dictionary learning based 3d morphable shape model. *IEEE Transactions on Multimedia*, 19(12):2666–2679, 2017.

[25] T. Gerig, A. Morel-Forster, C. Blumer, B. Egger, M. Luthi, S. Schönborn, and T. Vetter. Morphable face models-an open framework. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pages 75–82. IEEE, 2018.

[26] S. Z. Gilani, A. Mian, and P. Eastwood. Deep, dense and accurate 3d face correspondence for generating population specific deformable models. *Pattern Recognition*, 69:238–250, 2017.

[27] S. Z. Gilani, A. Mian, F. Shafait, and I. Reid. Dense 3d face correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(7):1584–1598, 2018.

[28] S. Z. Gilani, F. Shafait, and A. Mian. Shape-based automatic detection of a large number of 3d facial landmarks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4639–4648. IEEE, 2015.

[29] G. G. Gordon. Face recognition based on depth and curvature features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 808–810. IEEE, 1992.

[30] C. M. Grewe and S. Zachow. Fully automated and highly accurate dense correspondence for facial surfaces. In *European Conference on Computer Vision*, pages 552–568. Springer, 2016.

[31] X. Gu, S. Wang, J. Kim, Y. Zeng, Y. Wang, H. Qin, and D. Samaras. Ricci flow for 3d shape analysis. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1–8. IEEE, 2007.

[32] S. Gupta, M. K. Markey, and A. C. Bovik. Anthropometric 3d face recognition. *International Journal of Computer Vision*, 90(3):331–349, 2010.

[33] X. Huang, S. Zhang, Y. Wang, D. Metaxas, and D. Samaras. A hierarchical framework for high resolution facial expression tracking. In *the IEEE Conference on Computer Vision and Pattern Recognition Workshop*, pages 22–22. IEEE, 2004.

[34] B. Jian and B. C. Vemuri. Robust point set registration using gaussian mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1633–1645, 2011.

[35] H. Li, R. W. Sumner, and M. Pauly. Global correspondence optimization for non-rigid registration of depth scans. *Computer Graphics Forum*, 27(5):1421–1430, 2008.

[36] H. Li, T. Weise, and M. Pauly. Example-based facial rigging. *ACM Transactions on Graphics*, 29(4):32, 2010.

[37] T. Li, T. Bolkart, M. J. Black, H. Li, and J. Romero. Learning a model of facial shape and expression from 4d scans. *ACM Transactions on Graphics*, 36(6):194, 2017.

[38] Y. Lipman and T. Funkhouser. Möbius voting for surface correspondence. *ACM Transactions on Graphics*, 28(3):72, 2009.

[39] F. Liu, D. Zeng, Q. Zhao, and X. Liu. Joint face alignment and 3d face reconstruction. In *European Conference on Computer Vision*, pages 545–560. Springer, 2016.

[40] B. Maiseli, Y. Gu, and H. Gao. Recent developments and trends in point set registration methods. *Journal of Visual Communication and Image Representation*, 46:95–106, 2017.

[41] I. Matthews, J. Xiao, and S. Baker. 2d vs. 3d deformable face models: Representational power, construction, and real-time fitting. *International Journal of Computer Vision*, 75(1):93–113, 2007.

[42] H. Mohammadzade and D. Hatzinakos. Iterative closest normal point for 3d face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):381–397, 2013.

[43] A. Myronenko and X. Song. Point set registration: Coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2262–2275, 2010.

[44] M. Ovsjanikov, M. Ben-Chen, J. Solomon, A. Butscher, and L. Guibas. Functional maps: a flexible representation of maps between shapes. *ACM Transactions on Graphics*, 31(4):30, 2012.

[45] G. Pan, X. Zhang, Y. Wang, Z. Hu, X. Zheng, and Z. Wu. Establishing point correspondence of 3d faces via sparse facial deformable model. *IEEE Transactions on Image Processing*, 22(11):4170–4181, 2013.

[46] A. Patel and W. A. Smith. 3d morphable face models revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1327–1334. IEEE, 2009.

[47] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter. A 3d face model for pose and illumination invariant face recognition. In *Proceedings of the IEEE International Conference on Advanced video and Signal Based Surveillance*, pages 296–301. IEEE, 2009.

[48] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 947–954. IEEE, 2005.

[49] J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.

[50] S. Rusinkiewicz and M. Levoy. Efficient variants of the icp algorithm. In *Proceedings of the IEEE International Conference on 3-D Digital Imaging and Modeling*, pages 145–152. IEEE, 2001.

[51] A. Salazar, S. Wuhrer, C. Shu, and F. Prieto. Fully automatic expression-invariant face correspondence. *Machine Vision and Applications*, 25(4):859–879, 2014.

[52] M. P. Segundo, L. Silva, O. R. P. Bellon, and C. C. Queirolo. Automatic face segmentation and facial landmark detection in range images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 40(5):1319–1330, 2010.

[53] K. A. Sidorov, S. Richmond, and D. Marshall. Efficient groupwise non-rigid registration of textured surfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2401–2408. IEEE, 2011.

[54] F. M. Sukno, J. L. Waddington, and P. F. Whelan. 3-d facial landmark localization with asymmetry patterns and shape regression from incomplete local features. *IEEE Transactions on Cybernetics*, 45(9):1717–1730, 2015.

[55] G. K. Tam, Z.-Q. Cheng, Y.-K. Lai, F. C. Langbein, Y. Liu, D. Marshall, R. R. Martin, X.-F. Sun, and P. L. Rosin. Registration of 3d point clouds and meshes: a survey from rigid to nonrigid. *IEEE Transactions on Visualization and Computer Graphics*, 19(7):1199–1217, 2013.

[56] L. Tran and X. Liu. Nonlinear 3d face morphable model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2018.

[57] O. Van Kaick, H. Zhang, G. Hamarneh, and D. Cohen-Or. A survey on shape correspondence. *Computer Graphics Forum*, 30(6):1681–1707, 2011.

[58] D. Vlasic, M. Brand, H. Pfister, and J. Popović. Face transfer with multilinear models. *ACM Transactions on Graphics*, 24(3):426–433, 2005.

[59] Y. Wang, M. Gupta, S. Zhang, S. Wang, X. Gu, D. Samaras, and P. Huang. High resolution tracking of non-rigid motion of densely sampled 3d data using harmonic maps. *International Journal of Computer Vision*, 76(3):283–300, 2008.

[60] T. Weise, S. Bouaziz, H. Li, and M. Pauly. Realtime performance-based facial animation. *ACM Transactions on Graphics*, 30(4):77, 2011.

[61] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato. A 3d facial expression database for facial behavior research. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pages 211–216. IEEE, 2006.

[62] Y. Zeng, C. Wang, X. Gu, D. Samaras, and N. Paragios. Higher-order graph principles towards non-rigid surface registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(12):2416–2429, 2016.

[63] Y. Zeng, C. Wang, Y. Wang, X. Gu, D. Samaras, and N. Paragios. Dense non-rigid surface registration using high-order graph matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2010.

[64] C. Zhang, W. A. Smith, A. Dessein, N. Pears, and H. Dai. Functional faces: Groupwise dense correspondence using functional maps. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5033–5041. IEEE, 2016.

[65] H. Zhang, A. Sheffer, D. Cohen-Or, Q. Zhou, O. Van Kaick, and A. Tagliasacchi. Deformation-driven shape correspondence. In *Computer Graphics Forum*, volume 27, pages 1431–1439. Wiley, 2008.

[66] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li. Face alignment across large poses: A 3d solution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 146–155. IEEE, 2016.