Dense Semantic and Topological Correspondence of 3D Faces without Landmarks

Zhenfeng Fan[1,2][0000−0003−2299−7596], Xiyuan Hu[1,2⋆][0000−0002−7095−6986], Chen Chen[1,2][0000−0001−8297−6549], and Silong Peng[1,2,3][0000−0002−3594−5843]

[1] Insititute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[2] University of Chinese Academy of Sciences, China
[3] Beijing Visytem Co. Ltd.
{fanzhenfeng2016,xiyuan.hu,chen.chen,silong.peng}@ia.ac.cn

Abstract. Many previous literatures use landmarks to guide the correspondence of 3D faces. However, these landmarks, either manually or automatically annotated, are hard to define consistently across different faces in many circumstances. We propose a general framework for dense correspondence of 3D faces without landmarks in this paper. The dense correspondence goal is revisited in two perspectives: semantic and topological correspondence. Starting from a template facial mesh, we sequentially perform global alignment, primary correspondence by template warping, and contextual mesh refinement, to reach the final correspondence result. The semantic correspondence is achieved by a local iterative closest point (ICP) algorithm of kernelized version, allowing accurate matching of local features. Then, robust deformation from the template to the target face is formulated as a minimization problem. Furthermore, this problem leads to a well-posed sparse linear system such that the solution is unique and efficient. Finally, a contextual mesh refining algorithm is applied to ensure topological correspondence. In the experiment, the proposed method is evaluated both qualitatively and quantitatively on two datasets including a publicly available FRGC v2.0 dataset, demonstrating reasonable and reliable correspondence results.

Keywords: 3D face · dense correspondence · point set registration.

1 Introduction

Current progresses in economical sensors and advanced algorithms have pushed many 3D facial analysis technologies into real-life applications [1, 2]. Compared to its 2D counterpart, 3D face contains extra geometric information that can be utilized to resolve problems caused by illumination, pose, and expression variations. One of the most fundamental and challenging problems in the 3D facial analysis context is dense correspondence, which can contribute to many applications of 3D faces [3–5]. The benefit of dense correspondence is generally two-fold: 1) the one-to-one correspondence of points among different faces allows them to be organized in the same vector space, enabling further data analysis as a whole;

2) compared to sparse representations such as landmarks only, dense representations can capture subtle structures of faces, providing more discriminative information.

Inherited from the 2D literature, the methods using landmarks [6–10] have also been very popular for 3D face correspondence. Landmarks are some fiducial points on a face, such as the nose tip and the eye corner, which are very distinctive around their local contexts and can be detected either manually or automatically. The number of landmarks varies from a few tens to a hundred in different applications. It is generally considered that landmarks should dominate over all other points and be used as a guidance to establish full dense point-to-point correspondence. To date, many techniques have been proposed for accurate and robust detection of landmarks in 2D situations as well as in 3D [11–13].

Unfortunately, there are some major bottlenecks in using landmarks for 3D face correspondence. Firstly, manual annotation of landmarks is a laborious task, which is impractical for large database. Secondly, automatic landmark labeling is challenging especially for data with non-neutral expressions and variations in age and ethnicity. While it is easy to locate a landmark on one subject, it may be difficult to locate it on another subject with significant difference to ensure the same semantic meaning. Finally, it is even unable to define consistent number of landmarks across all subjects with data defects and noises, let alone the precision of these landmarks. All these considerations are demanding for practical applications.

This work aims at establishing dense point-to-point correspondence of 3D facial shapes without landmarks. We do not use texture here because texture is not always consistent with the underlying shape morphology, e.g. for the eyebrows, although the texture may provide with useful information for detection of some landmarks. Furthermore, shape is generally more reliable and stable feature under pose and illumination variations. Our goal is to seek for *reasonable* and *reliable* shape correspondence rather than *strict* and *standard* result. Strict correspondence may intuitively and physically exist between different data of the same subject, however, this problem hardly leads to a standard answer across different subjects. We restrict the correspondence problem to facial shapes and clarify the problem in two perspectives. One is *semantic correspondence* which guarantees that the corresponded points share the same semantic meaning. The other is *topological correspondence* which guarantees that the corresponded points lie in the same local context. An example for two corresponded faces A and A' is illustrated in Fig 1. It shows the corresponded points p and p' both denote the nose tip, and are both located at the center of their corresponded vicinity points.

We propose a fully automatic framework to achieve the semantic and topological correspondence goals for 3D faces. A template warping strategy is adopted here. First a face is chosen as a template and used to initially register the target faces, which is referred to as *global alignment* (Sec 3.1). Then preliminary correspondence (Sec 3.2) and robust deformation (Sec 3.3) from the template to the target are alternated and iterated for a few times to obtain a primary correspondence result. At this stage, *semantic correspondence* is guaranteed.
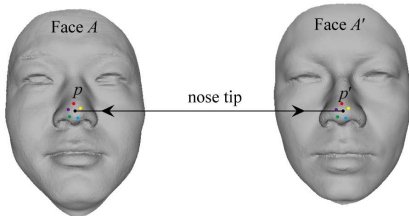
Fig. 1. Example of two corresponded faces in which points with the same color are in correspondences.

Finally, a contextual mesh refining algorithm is applied to ensure *topological correspondence* (Sec 3.4).

We evaluate the proposed method both qualitatively and quantitatively on two databases including a publicly available FRGC v2.0 [14] database and our newly collected HQ3DP database in the experiment. It shows that our method can achieve reasonable and reliable correspondence results. Except for a fully automatic algorithm for 3D face dense correspondence, the main contribution of this work is two-fold: 1) we construct a template warping strategy that takes both global and local features into consideration for accurate semantic correspondence, where the meaning of "semantic" we designated here is more than just landmark correspondence; 2) we consider topological correspondence and propose a mesh refining algorithm to achieve it, and this notion may give new insights towards the final solid definition for dense correspondence of 3D faces.

2 Related work

For a comprehensive overview of 3D shape correspondence methods which are not limited to faces one can refer to [15, 5]. Here we only review and summarize some state-of-the-art representative algorithms for dense point correspondence of 3D faces.

In the seminal work of 3D Morphable Model (3DMM) [3, 16], Blanz and Vetter propose a novel regularized optic flow based algorithm for dense point correspondence. The 3D coordinates of the facial surface are projected to a cylindrical "UV" map, with both the third dimension "radius" (i.e. distance to the cylindrical axis) and the texture encoded as "pixel intensities" with proper weights. Coarse-to-fine and local smoothing strategies are applied in the computational process. Despite its efficiency by reducing the dimensionality of this problem from 3D to 2D, the potential pitfalls are 1) the mapping from 3D coordinates to 2D "UV" space may not be one-to-one, 2) mapping from 3D to 2D and then back to 3D requires resampling and may lose precision of original data.

As an alternative, a thin plate splines (TPS) [17] warp is employed by Patel and Smith for the correspondence task with some manually labeled annotations. Gilani et al.[13] also use the TPS for estimating and minimizing the local bending energy for registering of some seed points which can be much more than

landmarks, and further incorporating them into the full shape correspondence. These approaches are considered more robust to handle large variations in facial identity.

Amberg et al.[18] propose an optimal step nonrigid iterative closest point (NICP) algorithm for the registration of 3D shapes. It is used as a benchmark algorithm for face correspondence in some prior arts [19–22]. More recently in the work of Booth et al.[23, 24], landmark detections are automated using the 3D textures in different viewing directions. And then, these landmarks are used to guide the local affine deformations of NICP with a proper parameter setting. They also compare the performance of several methods for 3D face correspondence and reach a conclusion that NICP is superior to the TPS and optic flow based algorithms.

In the rich literature of 3D face correspondence, many algorithms [6, 25–27] give impressive results by their capability to model identity and expression variations or both. For example, Bolkart and Wuhrer [27] propose a multilinear groupwise model for 3D face correspondence to decouple both identity and expression variations. However, a well-established correspondence model is a prerequisite in these methods and the correspondence problem is considered in the restrictive space expressed by the model. Although insightful and practically useful, a chicken-and-egg problem still remains to be solved. A vital question is, if the corresponded model has some errors, they may accumulate and propagate in their subsequent usages.

Existing methods are mostly focusing on semantic correspondence. For example, there is functional face proposed by Zhang et al. [28] who incorporate multiple features into the correspondence strategy by Laplace-Beltrami operators. Some geodesic preserving methods [29, 30] have also shown merits over Euclidean distance based ones on modeling expression variations. We are not going to categorize them by the algorithm details, for example, by whether they employ one-to-one or one-to-many strategies, or they use textures or not. Semantic correspondence may be notably defined in some regions of the face such as the eyes corners. However in other regions such as the cheek and forehead, it is even not uniquely identifiable by a human. In this work, we are seeking for a more definite solution for this problem. The similarity of local data structure, referred as topological correspondence in this work, is an extra constraint we take into consideration.

3   Method

3.1   Global alignment

An arbitrary but well enough (noiseless) face is chosen as the template and used to initially register the target faces on the first stage, referred as global alignment here. Global alignment aims at aligning two facial shapes such that the true point-to-point correspondence is close to the nearest one. We use a modified iterative closest point (ICP) method to achieve global alignment. ICP [31] is a
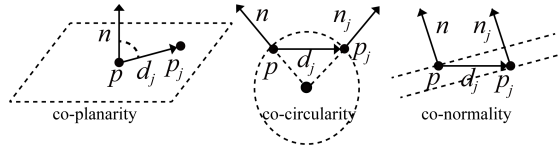
Fig. 2. Meanings of co-planarity, co-circularity, and co-normality. The three sub-figures lead to zero values for each of the three terms, respectively.

simple yet efficient point set registration algorithm that iterates between closest-point correspondence and rigid-motion estimation until convergence. Given two sets of corresponded points $P = \{p_1, p_2, ..., p_n\}$ and $Q = \{q_1, q_2, ..., q_n\}$ in $\mathbb{R}^3$, the rigid-motion estimation is a constraint minimization problem with respect to the optimal rotation $\mathbf{R}$ and translation $\mathbf{t}$,

$$\{\mathbf{R}, \mathbf{t}\} = \arg\min_{\mathbf{R}, \mathbf{t}} \sum_{i=1}^{n} \|\mathbf{R}p_i + \mathbf{t} - q_i\|_2^2. \tag{1}$$

And this problem can be solved efficiently by singular value decomposition.

A novel Go-ICP [32, 33] algorithm, which is an improved version of ICP for 3D shape registration, guarantees an optimum solution even without a good initialization by a branch and bound searching strategy in the parameter space. By virtue of it, detection of some fiducial points (e.g. the nose tip and some other sparse landmarks) for pose estimation of 3D face which is done in many existing works is circumvented here, making the proposed method one step further to be automatic.

### 3.2 Dichotomic preliminary correspondence

To achieve semantic correspondence, we treat the points on a face differently. The points on a face are heuristically categorized into two groups. Points that belongs to the complex regions where the semantic meaning is clear form one group, and the rests that belongs to smooth regions where the semantic meaning is not clearly distinctive form the other. Since the face is an ellipsoid-like surface, the smooth regions can be modeled to be planar or circular. Let $p_j = p + d_j$ $(j = 1, ..., k)$ be the neighbors of point p and let n_j $(j = 1, ..., k)$ and n be the corresponding normals. Inspired by the work of [34, 35], we define entropy of a point p on a facial surface as

$$M(p) = M_p(p) + M_c(p) + M_n(p), \tag{2}$$

which is the sum of three terms (also illustrated in Fig 2).
1. *Co-planarity* measures the extent that neighboring vertices lie in each other's tangent planes.

$$M_p(p) = \sum_{j=1}^{k} \left(n^T d_j\right)^2 / k \tag{3}$$
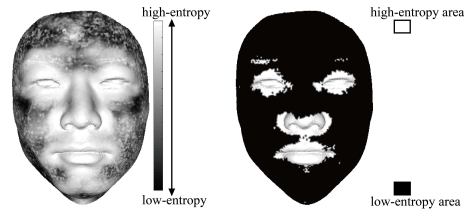
Fig. 3. Mapping of entropy on a 3D face. The left and right sub-figures are an entropy-intensity map with colorbar and a binary map for two classes of points, respectively.

2. *Co-circularity* measures the extent that neighboring vertices share similar curvatures.

$$M_c(p) = \sum_{j=1}^{k} \left[(n + n_j)^T d_j \right]^2 / k \tag{4}$$

3. *Co-normality* measures the extent that neighboring vertices hold the same normal.

$$M_n(p) = \sum_{j=1}^{k} \| n - n_j \|^2 / k \tag{5}$$

The neighbors of p can be defined as points within a local patch around it. Note that all the vectors are normalized to unit length in the computation process. We then classify the points on the template face into two groups by sorting their entropies. Entropy gives a measure to the local distinctiveness for semantic meaning. In this paper we define the top 10% of all points as high-entropy points and they are a well cover for the eyes, nose, and mouth regions. Fig 3 shows an entropy-intensity map for all points and a binary map for the two groups of points on a 3D face.

We simply apply the closest-point rule to the low-entropy points for preliminary correspondence for computational efficiency. However as to the high-entropy points, we apply a different local ICP of kernelized version for semantic correspondence. A Gaussian kernel is adopted here to emphasis local features. The optimal rotation **R** and translation **t** are estimated by

$$\{\mathbf{R}, \mathbf{t}\} = \arg\min_{\mathbf{R}, \mathbf{t}} \sum_{i=1}^{k} K(p_i) \|\mathbf{R} p_i + \mathbf{t} - q_i\|_2^2. \tag{6}$$

Compared to Eq 1, the objective function is reweighted by a kernel function $K(p_i)$. The effect of Gaussian kernel is illustrated in Fig 4. The local shape differences between the template and the target face are commonly manifested as stretching or shrinking of curves to accompany with expression and identity variations. It shows that compared to uniform weighting, Gaussian weighting as a function of distance to the corresponding point, is capable of matching finer local details.
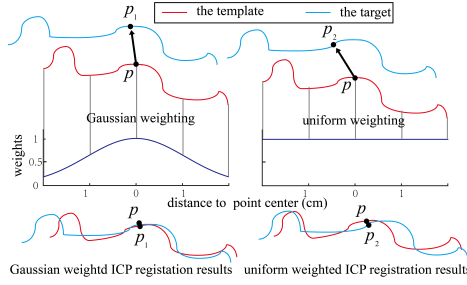
Fig. 4. Effect of Gaussian weighting compared to standard ICP (referred as uniform weighting). The point p is corresponded to p_1 and p_2 by the two different ICP weighting strategies, respectively.

3.3 Nonrigid robust deformation

The template face is deformed into the target face to obtain the correspondence result after preliminary correspondence is established. Before formulation of the deformation model, we first give the meaning of some symbols.

- V and P denotes the point sets of the template face and the target face, respectively.
- v_i is a point on V and its preliminary correspondence on P is p_{v_i}.
- o_i denotes the deformation offset to be solved.
- w_i denotes the weight assigned to each corresponded point pair.

Then, the deformation problem is formulated as

$$\{o_i | i \in V\} = \underset{\{o_i | i \in V\}}{\arg\min} \{\sum_{i \in V} w_i \left\| p_{v_i} - (v_i + o_i) \right\|_2^2 + \lambda \cdot \sum_{i \in V} \sum_{j \in r_i^{(1)}} \left\| o_i - o_j \right\|_2^2 \}, \quad (7)$$

where $r_i^{(1)}$ denotes the 1-ring neighbors of the point v_i and it contains $N_{r_i^{(1)}}$ points. The second term in Eq 7 is a local smoothing constraint, which renders the template face to deform into the target face gradually in each small step. For each step, we set $v_i \leftarrow v_i + o_i \, (i \in V)$ and preliminary correspondence is rebuilt by the method described in Sec 3.2. The iteration is terminated when the residual error $\sum_{i \in V} \left\| p_{v_i} - v_i \right\|_2^2$ is smaller than a predefined threshold ε.
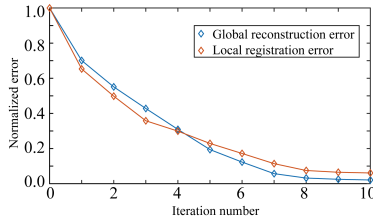
Solution to the deformation problem. The overall solution to Eq 7 leads to a linear system:

$$[\mathbf{A}_{ij}]_{N \times N} \cdot [\mathbf{O}_{ij}]_{N \times 3} = [\mathbf{B}_{ij}]_{N \times 3}, \quad (8)$$

where

$$\mathbf{A}_{ij} = \begin{cases} w_i + \lambda \cdot N_{r_i^{(1)}} & if \, i = j \\ -\lambda & if \, i \neq j \, and \, j \in r_i^{(1)} \\ 0 & otherwise \end{cases}, \quad (9)$$

$$[\mathbf{O}_{ij}]_{N \times 3} = [o_1, ..., o_N]^T, \quad (10)$$

Fig. 5. Global reconstruction error of the target face and local registration error of high-entropy points for an example in Fig 6. The errors are normalized to $[0, 1]$.

and

$$[\mathbf{B}_{ij}]_{N \times 3} = [w_1 \cdot (p_{v_1} - v_1), ..., w_N \cdot (p_{v_N} - v_N)]^T. \tag{11}$$

N is the total number of points. It is obvious that $\mathbf{A}$ is a *sparse*, *square*, and *strict diagonal-dominant* matrix such that the system is well-posed, computational efficient, and has a unique solution.

Weighting differently for each point pair. To ensure more reliable preliminary correspondence customized for nonrigid deformation, the weight for each point pair is set by some intuitive rules as follows.

1. Weights for high-entropy points should be larger than some other points such that the semantic correspondence of high-entropy points serves as a guide for the deformation process.

2. If the distance between the corresponded points is larger than a threshold or the inner product between the normal vectors of the corresponded points is smaller than a threshold, we set the weight to be zero to reject unreasonable correspondences.

3. If the corresponded point of the target face lies on the boundary of the mesh, the weight is set to zero.

These rules guarantee reasonable deformation to reconstruct the target face. Of course more rules can be devised to meet some specified needs practically. It is different to some landmark guided nonrigid deformation methods [6, 18] in two perspectives. One is that the detection and preliminary correspondence of the high-entropy points are automated by a local ICP of kernelized version. The other is, since high-entropy points are much more than landmarks, the interactions of them with their nearby ones lead to robust solutions which are more tolerant to individual correspondence errors. The preliminary correspondences of high-entropy points also renew themselves for smaller local registration errors (see Fig 5) as the deformation towards the target proceeds. As shown in Fig 6, it provides a possible way to model both identity and mild expression variations.

3.4   Contextual mesh refinement

We assume two corresponded faces should have similar local mesh structure to ensure topological correspondence. The following steps constitute the algorithm
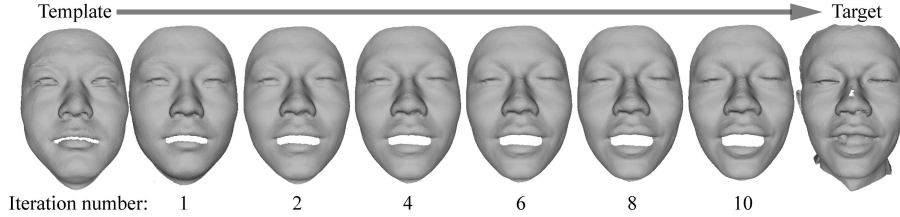
Fig. 6. The template face deforms into the target face in each small step.

to achieve topological correspondence. It starts from a corresponded mesh by the nonrigid deformation described above.

1. Project the 1-ring neighbors of each point of the template and the target face into their local tangent planes, respectively.

2. For each point v_i of the template, register its projected 1-ring neighbors to its corresponded projected 1-ring neighbors on the target such that the point of the template is now located at p'_{v_i}.

3. Renew p_{v_i} to p'_{v_i} but with a local smoothing strategy determined by

$$\{o_i|i \in P\} = \underset{\{o_i|i\in P\}}{\arg\min}\{\sum_{i\in P} \left\|p'_{v_i} - (p_{v_i} + o_i)\right\|_2^2 + \eta \cdot \sum_{i\in P}\sum_{j\in r_i^{(1)}} \left\|o_i - o_j\right\|_2^2\}, \quad (12)$$

such that $p_{v_i} \leftarrow p_{v_i} + o_i \ (i \in P)$.

4. Loop from step 1 to step 3 until $\sum_{i\in P} \left\|p'_{v_i} - p_{v_i}\right\|_2^2$ is smaller than a predefined threshold ζ.

Measuring the mesh structure on the local tangent plane greatly simplifies the problem from 3D to 2D. Furthermore, renewing each point of the target face on its local tangent plane enables the mesh to avoid shrinking. Fig 7(a) depicts the steps visually. Fig 7(b) shows an example of how the local mesh structure is changed after topological correspondence by the proposed algorithm.

4   Experiments

Datasets. We use two datasets for the evaluation of the results. One is our newly collected database (**HQ3DP**) which includes 330 high-resolution 3D face prototypes by a structured light device. This system merges scans from 4 different directions in a well-controlled environment. The database is originally for building a statistical model and the subjects are mostly in neutral expressions. It is used for some experiments because of the high quality on the mouth, nose, and eyes regions. The other is a benchmark 3D face recognition grand challenge database (**FRGC v2.0** [14]) from University of Notre Dame. It is a publicly available database with mild facial expressions used by many existing works, enabling comparison of the correspondence results quantitatively. It is
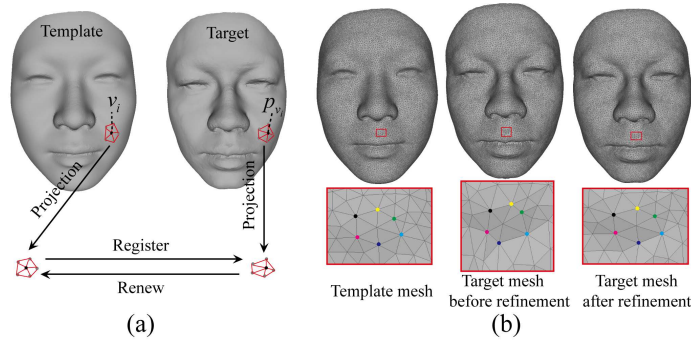
Fig. 7. (a)The steps for mesh refinement;(b)Effects of topological correspondence for an example. Note that the local structure of the output target mesh by the refining algorithm resembles the template mesh.

Table 1. Average runtime for the reconstruction of a target face.

Global alignment	Template warping	Mesh refinement	Total time
108.6(s)	30.2(s)	25.8(s)	164.6(s)

worth mentioning that some famous and publicly available databases (e.g. BU-3DFE [36], Bosphorus [37]) have made a great contribution to the progress for 3D facial analysis. We choose the FRGC v2.0 because its relative high resolution and the subjects are mostly free of occlusions.

Parameter setting. The template face is resampled to include $20,969$ points in the experiments. The parameter λ controls the influence of the template shape and should be decreasing. We set λ to decrease from 1 to 0.1 by a step size 0.1. Note that it should also be affected by the resolution of the template face. When the resolution becomes higher, λ should also be larger to maintain the same effective range. The setting of parameter η is trivial because it shows little influence to the final results. We set η to be a constant 0.5.

Computational time. The computational time increases linearly with the number of points for the representation of a face. Tab 1 summarize the runtime (VC++ implementation, Core i7-6700k single thread) for the reconstruction of a target face from a template face with $20,969$ points. Since there are many per-vertex operations that are independent to each other, parallel implementation can be developed to accelerate the process.

4.1   Visualizing the correspondence result

The correspondence result is evaluated qualitatively first. We adopted three intuitive criteria for it.

Criterion 1: Sharpness of the average face. It can be argued that if the correspondence results for all the faces are semantically accurate, the average
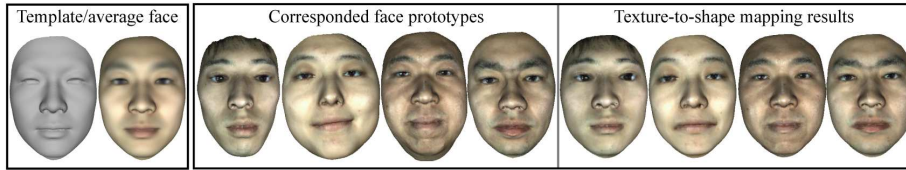
Fig. 8. Average face and texture-to-shape mappings.
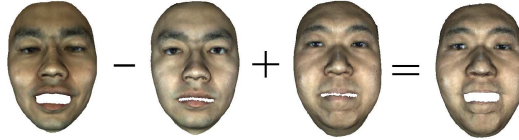

Fig. 9. An example of expression transfer result.

face should be clear on the edges (e.g. edges for the eyes, nose, and mouth region). The average face can be further used as the template face. It shows in Fig 8 that the average face of 330 faces in HQ3DP maintains the key structure of human face, which indicates good correspondence.

Criterion 2: Texture-to-shape mapping. Some existing works [4, 28] transfer the texture of one face to the shape of another face to assess the correspondence result qualitatively. Following it we extract several corresponded face prototypes in HQ3DP, and paste their textures to the average shape, respectively. Fig 8 shows a well match of semantic meaning between shape and textures.

Criterion 3: Expression transfer. A robust correspondence method should also be able to deal with mild facial expressions. We calculate the error between a corresponded neutral face and face with expression of the same identity, and then add it to another neutral face of different identity. It shows a realistic synthetic result in Fig 9 for expression transfer.

4.2   Compactness of bases derived by PCA

Fig 7(b) has already shown more reasonable local mesh structure for the representation of a corresponded face visually. To demonstrate the effectiveness of this mesh refining algorithm quantitatively, we apply PCA to 330 corresponded faces in HQ3DP database after Procrustes alignment. Fig 10 shows the percentage of energy preserved by varying the number of principal components before and after the mesh refining algorithm in Sec 3.4. We also implement the NICP algorithm [18] for comparison. It shows that there is no significant differences between sparse landmark guided correspondence (NICP) and denser high-entropy guided one. However, the proposed mesh refining algorithm has made the bases more compact to a considerable extent. It is worth mentioning that this algorithm can start from the correspondence results by any of the existing algorithms.
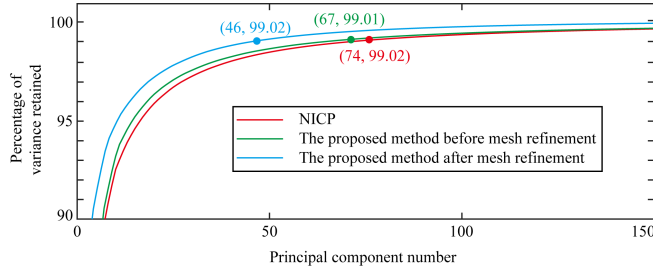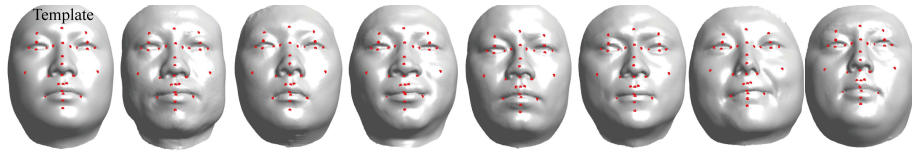
Fig. 10. Compactness of the fitted model.

Fig. 11. Landmark correspondences. The left sub-figure is the manually annotated template and the rests are the landmark detection results.

Compactness of PCA components implies less variability of the local mesh structure for all the corresponded faces. The proposed algorithm, derived from the assumption of similar local mesh structure for topological correspondence, inherently meet this goal. One practical significance is for the 3DMM fitting process, more compact bases may indicate less chance to be trapped in local minima, enabling better 3D reconstructed results.

4.3   Landmark correspondence

The proposed algorithm is different from some existing landmark guided ones. Nevertheless, it can also be used for 3D face landmark detection. A merit of the proposed algorithm is that it can detect any number of user defined landmarks. The method is simple: manually label some landmarks on the template face and then find the points with the same indexes on other target faces as their respective landmarks. Fig 11 shows some landmark detection results qualitatively on HQ3DP database, which are accurate and competitive with manual annotation. Note that we do not give specific anthropometric meaning to these landmarks here.

To quantitatively assess the proposed method, we also compare the landmark detection results with some previous literatures on the FRGC v2.0 database. The manual annotations from [11, 38] are used as the ground truth for comparison. Fig 12(a) shows the template we used and the anthropometric meanings of the annotations. Tab 2 shows the comparison for the correspondence results. It demonstrates that our method achieves state-of-the-art performance and fulfills the requirement for semantic correspondence. A little degradation for the
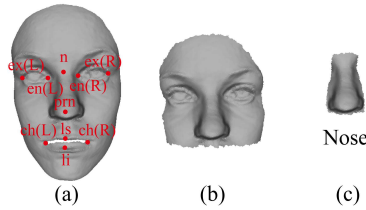
Fig. 12. Templates used for (a) correspondence, (b) face recognition for the full shape, and (c) face recognition for the nose region.

Table 2. Comparative results of the mean and standard deviation (SD) of landmark localization error (mm) for 4007 images on FRGC v2.0 dataset.

Landmark	Segundo et al. [39]	Creusot et al. [11]	Gilani et al. [13]	ours
	Mean/SD	Mean/SD	Mean/SD	Mean/SD
ex(L)	-	5.87/3.11	4.50/2.97	**2.62/1.54**
en(L)	3.69/2.26	4.31/2.44	3.12/2.09	**2.53/1.66**
n	-	4.20/2.07	3.63/2.02	**2.43/1.36**
ex(R)	-	6.00/3.03	3.74/2.79	**2.60/1.71**
en(R)	3.35/2.33	4.29/2.03	2.73/2.14	**2.49/1.65**
prn	2.73/1.39	3.35/2.00	2.68/1.48	**2.11/1.17**
ch(L)	-	5.47/3.45	5.31/2.05	**2.93/2.14**
ch(R)	-	5.64/3.58	4.38/2.08	**2.84/2.17**
ls	-	4.23/3.21	3.31/2.65	**2.35/2.86**
li	-	5.46/3.29	**4.02/3.80**	4.35/3.93

lower lip midpoint (**li**) may result from large noises in the mouth region for this database. Note that we do not use textures for locating it, although textures should be helpful and are used for accurate manual annotations.

4.4 3D Face Recognition

Since more accurate correspondence results tend to give better 3D face recognition performances, we extend the proposed method into a challenging face recognition task on the database FRGC v2.0. The ROC III mask defined in [14] is used for the 3D shape based face recognition experiment. Under this mask, the gallery and probe sets include a total number of 4,007 3D scans of 466 subjects collected during Fall 2003 and Spring 2004, respectively. It is also an extremely challenging experiment conducted by many previous works because of the cross-seasonal variations imposed. We follow a PCA+LDA face recognition pipeline [40] after correspondence of all faces are established. Since the database meets some common problems such as being occluded by hair or hat and masked by tremendous noises especially around the mouth region, the faces are cropped after correspondences are established. Fig 12(b) and Fig 12(c) shows the shapes of the cropped faces for face recognition based on the full shape and the nose region, respectively. To compare the performance with the previous literatures,

Table 3. Comparisons of face recognition performances for the ROC III experiments on FRGC v2.0 dataset.

Method	Modality	Verification rate	Identification rate	Average rate
Osaimi et al. [41]	full shape	94.05	96.52	95.29
Alyüz et al. [42]	full shape	86.1	-	-
Wang et al. [43]	full shape	98.04	98.3	98.17
Berretti et al. [44]	full shape	-	94.15	-
Queirolo et al. [45]	full shape	96.6	**99.6**	98.1
Spreeuwers [46]	full shape	94.6	99.0	96.8
Drira et al. [47]	full shape	97.1	97.0	97.05
Mohammadzade et al. [4]	full shape	**99.2**	-	-
Ours	full shape	97.7	99.5	**98.6**
Emambakhsh et al. [48]	nose region	93.5	97.9	95.7
Spreeuwers [46]	nose region	83.7	94.5	89.1
Ours	nose region	**95.8**	**98.5**	**97.2**

both face identification and verification are conducted. Face identification results are obtained by nearest-neighbor search to get **Rank-1** performance. And face verification results are obtained by interpolating the ROC curves, in which verification rates under 0.1% false alarm rate are shown. The comparative results are included in Tab 3. It shows that our method achieves both the best face verification (95.8%) and identification (98.5%) performances for the nose region and the best average performance (98.6%) for the full shape, respectively, which demonstrates reliable correspondences.

5 Discussion and Conclusion

The proposed algorithm for correspondence can handle mild facial expression variations. A transitive strategy instead of the one-template strategy is preferred for faces with large expressions, such as the BU-3DFE [36] database. "Transitive" means correspondences are transferred from low-level to high-level expressions. With the established correspondences of a number of sampling faces, it would be easy to learn a prior model of 3D faces and then generalize to circumstances with occlusions and partial data. Although the inter-individual correspondence problem may not have strict and standard answer, the experiments demonstrate reasonable and reliable results by the proposed method. We hope this work including the notion of semantic and topological correspondence, will be helpful for a final solid definition for dense correspondence of 3D faces.

6 Acknowledgement

References

1. Corneanu, C.A., Simon, M.O., Cohn, J.F., Guerrero, S.E.: 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) (2016) 1548–1568
2. Matthews, I., Xiao, J., Baker, S.: 2d vs. 3d deformable face models: Representational power, construction, and real-time fitting. International journal of computer vision **75**(1) (2007) 93–113
3. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3d faces. In: Proceedings of the 26th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co. (1999) 187–194
4. Mohammadzade, H., Hatzinakos, D.: Iterative closest normal point for 3d face recognition. IEEE transactions on pattern analysis and machine intelligence **35**(2) (2013) 381–397
5. Tam, G.K., Cheng, Z.Q., Lai, Y.K., Langbein, F.C., Liu, Y., Marshall, D., Martin, R.R., Sun, X.F., Rosin, P.L.: Registration of 3d point clouds and meshes: a survey from rigid to nonrigid. IEEE transactions on visualization and computer graphics **19**(7) (2013) 1199–1217
6. Pan, G., Zhang, X., Wang, Y., Hu, Z., Zheng, X., Wu, Z.: Establishing point correspondence of 3d faces via sparse facial deformable model. IEEE Transactions on Image Processing **22**(11) (2013) 4170–4181
7. Gilani, S.Z., Mian, A., Eastwood, P.: Deep, dense and accurate 3d face correspondence for generating population specific deformable models. Pattern Recognition **69** (2017) 238–250
8. Romdhani, S.: Face image analysis using a multiple features fitting strategy. PhD thesis, University_of_Basel (2005)
9. Guo, J., Mei, X., Tang, K.: Automatic landmark annotation and dense correspondence registration for 3d human facial images. BMC bioinformatics **14**(1) (2013) 232
10. Grewe, C.M., Zachow, S.: Fully automated and highly accurate dense correspondence for facial surfaces. In: European Conference on Computer Vision, Springer (2016) 552–568
11. Creusot, C., Pears, N., Austin, J.: A machine-learning approach to keypoint detection and landmarking on 3d meshes. International journal of computer vision **102**(1-3) (2013) 146–179
12. Song, M., Tao, D., Sun, S., Chen, C., Maybank, S.J.: Robust 3d face landmark localization based on local coordinate coding. IEEE Transactions on Image Processing **23**(12) (2014) 5108–5122
13. Zulqarnain Gilani, S., Shafait, F., Mian, A.: Shape-based automatic detection of a large number of 3d facial landmarks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2015) 4639–4648
14. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on. Volume 1., IEEE (2005) 947–954
15. Castellani, U., Bartoli, A.: 3d shape registration. In: 3D Imaging, Analysis and Applications. Springer (2012) 221–264
16. Blanz, V., Vetter, T.: Face recognition based on fitting a 3d morphable model. IEEE Transactions on pattern analysis and machine intelligence **25**(9) (2003) 1063–1074

17. Patel, A., Smith, W.A.: 3d morphable face models revisited. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE (2009) 1327–1334
18. Amberg, B., Romdhani, S., Vetter, T.: Optimal step nonrigid icp algorithms for surface registration. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, IEEE (2007) 1–8
19. Cheng, S., Marras, I., Zafeiriou, S., Pantic, M.: Statistical non-rigid icp algorithm and its application to 3d face alignment. Image and Vision Computing **58** (2017) 3–12
20. Li, H., Sumner, R.W., Pauly, M.: Global correspondence optimization for non-rigid registration of depth scans. In: Computer graphics forum. Volume 27., Wiley Online Library (2008) 1421–1430
21. Ferrari, C., Lisanti, G., Berretti, S., Del Bimbo, A.: A dictionary learning based 3d morphable shape model. IEEE Transactions on Multimedia (2017)
22. Ferrari, C., Lisanti, G., Berretti, S., Del Bimbo, A.: Dictionary learning based 3d morphable model construction for face recognition with varying expression and pose. In: 3D Vision (3DV), 2015 International Conference on, IEEE (2015) 509–517
23. Booth, J., Roussos, A., Zafeiriou, S., Ponniah, A., Dunaway, D.: A 3d morphable model learnt from 10,000 faces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 5543–5552
24. Booth, J., Roussos, A., Ponniah, A., Dunaway, D., Zafeiriou, S.: Large scale 3d morphable models. International Journal of Computer Vision (2017) 1–22
25. Vlasic, D., Brand, M., Pfister, H., Popović, J.: Face transfer with multilinear models. In: ACM transactions on graphics (TOG). Volume 24., ACM (2005) 426–433
26. Cheng, S., Marras, I., Zafeiriou, S., Pantic, M.: Active nonrigid icp algorithm. In: Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on. Volume 1., IEEE (2015) 1–8
27. Bolkart, T., Wuhrer, S.: A groupwise multilinear correspondence optimization for 3d faces. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 3604–3612
28. Zhang, C., Smith, W.A., Dessein, A., Pears, N., Dai, H.: Functional faces: Groupwise dense correspondence using functional maps. In: CVPR. (2016) 5033–5041
29. Sidorov, K.A., Richmond, S., Marshall, D.: Efficient groupwise non-rigid registration of textured surfaces. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE (2011) 2401–2408
30. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Expression-invariant 3d face recognition. In: international conference on Audio-and video-based biometric person authentication, Springer (2003) 62–70
31. Besl, P.J., McKay, N.D., et al.: A method for registration of 3-d shapes. IEEE Transactions on pattern analysis and machine intelligence **14**(2) (1992) 239–256
32. Yang, J., Li, H., Jia, Y.: Go-icp: Solving 3d registration efficiently and globally optimally. In: Proceedings of the IEEE International Conference on Computer Vision. (2013) 1457–1464
33. Yang, J., Li, H., Campbell, D., Jia, Y.: Go-icp: a globally optimal solution to 3d icp point-set registration. IEEE transactions on pattern analysis and machine intelligence **38**(11) (2016) 2241–2254
34. Szeliski, R., Tonnesen, D.: Surface modeling with oriented particle systems. Volume 26. ACM (1992)
35. Linsen, L.: Point cloud representation. Univ., Fak. für Informatik, Bibliothek (2001)

36. Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.J.: A 3d facial expression database for facial behavior research. In: Automatic face and gesture recognition, 2006. FGR 2006. 7th international conference on, IEEE (2006) 211–216

37. Savran, A., Alyüz, N., Dibeklioğlu, H., Çeliktutan, O., Gökberk, B., Sankur, B., Akarun, L.: Bosphorus database for 3d face analysis. In: European Workshop on Biometrics and Identity Management, Springer (2008) 47–56

38. Szeptycki, P., Ardabilian, M., Chen, L.: A coarse-to-fine curvature analysis-based rotation invariant 3d face landmarking. In: Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on, IEEE (2009) 1–6

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

40. Gupta, S., Markey, M.K., Bovik, A.C.: Anthropometric 3d face recognition. International journal of computer vision **90**(3) (2010) 331–349

41. Al-Osaimi, F., Bennamoun, M., Mian, A.: An expression deformation approach to non-rigid 3d face recognition. International Journal of Computer Vision **81**(3) (2009) 302–316

42. Alyuz, N., Gokberk, B., Akarun, L.: Regional registration for expression resistant 3-d face recognition. IEEE Transactions on Information Forensics and Security **5**(3) (2010) 425–440

43. Wang, Y., Liu, J., Tang, X.: Robust 3d face recognition by local shape difference boosting. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(10) (2010) 1858–1870

44. Berretti, S., Del Bimbo, A., Pala, P.: 3d face recognition using isogeodesic stripes. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(12) (2010) 2162–2177

45. Queirolo, C.C., Silva, L., Bellon, O.R., Segundo, M.P.: 3d face recognition using simulated annealing and the surface interpenetration measure. IEEE transactions on pattern analysis and machine intelligence **32**(2) (2010) 206–219

46. Spreeuwers, L.: Fast and accurate 3d face recognition. International Journal of Computer Vision **93**(3) (2011) 389–414

47. Drira, H., Amor, B.B., Srivastava, A., Daoudi, M., Slama, R.: 3d face recognition under expressions, occlusions, and pose variations. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(9) (2013) 2270–2283

48. Emambakhsh, M., Evans, A.: Nasal patches and curves for expression-robust 3d face recognition. IEEE transactions on pattern analysis and machine intelligence **39**(5) (2017) 995–1007