

METHODOLOGY ARTICLE

Open Access

Automatic landmark annotation and dense correspondence registration for 3D human facial images

Jianya Guo, Xi Mei and Kun Tang*

Abstract

Background: Traditional anthropometric studies of human face rely on manual measurements of simple features, which are labor intensive and lack of full comprehensive inference. Dense surface registration of three-dimensional (3D) human facial images holds great potential for high throughput quantitative analyses of complex facial traits. However there is a lack of automatic high density registration method for 3D faical images. Furthermore, current approaches of landmark recognition require further improvement in accuracy to support anthropometric applications.

Result: Here we describe a novel non-rigid registration method for fully automatic 3D facial image mapping. This method comprises two steps: first, seventeen facial landmarks are automatically annotated, mainly via PCA-based feature recognition following 3D-to-2D data transformation. Second, an efficient thin-plate spline (TPS) protocol is used to establish the dense anatomical correspondence between facial images, under the guidance of the predefined landmarks. We demonstrate that this method is highly accurate in landmark recognition, with an average RMS error of ~ 1.7 mm. The registration process is highly robust, even for different ethnicities.

Conclusion: This method supports fully automatic registration of dense 3D facial images, with 17 landmarks annotated at greatly improved accuracy. A stand-alone software has been implemented to assist high-throughput high-content anthropometric analysis.

Keywords: 3D face, Facial morphology, Registration, Landmark localization, Dense correspondence

Background

Large-scale, high-throughput phenotyping is becoming increasingly important in the post-genomics era. Advanced image processing technologies are used more and more for collecting deep and comprehensive morphological data from different organisms, such as yeast [1], plants [2], worm [3] as well as mice [4]; and for different body parts such as brain [5,6], lung [7] and face [8-13]. Especially for the brain 3D image registration, a novel elastic registration (HAMMER) of magnetic resonance images of the brain has greatly facilitated the medical research of brain [6,14]. A recent work that combined florescent labeling and non-rigid registration achieved registration accuracy up to 2 μ m in drosophila

brain, which potentially allows functional analyses at individual neuron resolution [5]. The soft tissue of the human face is a complex geometric surface composed of many important organs, including eyes, nose, ears, mouth, etc. Given its essential biological functions, the human face has been a key research subject in a wide range of fields including anthropology [15], medical genetics [8,9,16,17], forensics [18,19], psychology [20,21], aging [22,23] and the upcoming quantitative genomics [24,25], etc.. Nonetheless, for a long period of time period the rich quantitative traits of face have not been made full use of. Previous anthropometric studies have been largely based on tedious manual measurements of dozens of distances between a set of landmarks, which were subjectively determined by the observers' eyes and were thus error prone and sensitive to individual differences [26-28]. In the past few years, efforts have been paid to discover the genetic determinants of normal facial variations either by examining

* Correspondence: tangkun@picb.ac.cn
CAS-MPG Partner Institute and Key Laboratory for Computational Biology,
Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences,
Shanghai, China

candidate genes or via genome-wide association studies [24,25,29-33]. Although high resolution 3D images were taken in some of these studies, the landmarks were still manually annotated [25]; and simple landmark coordinates or landmark-distances were used as the major phenotype data [24,25,33]. Such practices unavoidably led to loss of resolution and statistic power. In short, the lack of quantitative methods to capture the facial morphology in high definition and full automation has hindered the biological researches in human face. In fact, in the field of computer vision, many technologies have been developed for landmark recognition and dense point registration. Nonetheless, few have been successfully applied in the biological studies of human face. This is largely due to the different requirements between computer vision and biological studies. Quantitative biological analyses of face require the face registration to follow the principle of anatomical correspondence and landmarks have to be localized at high accuracies. However such rules are often not satisfied in the computer vision methods. For image registration, many methods rely on rigid transformation, such as the Iterative Closest Point (ICP) [34]. ICP uses affine transformations, including rotation and translation to find the closest corresponding points between two surfaces. The registration based on ICP does not fully capture the anatomical variability, especially when faces to be compared differ significantly in shape or expression. For landmark localization, there exist many automatic methods [18,35-39]. Landmark localization methods based on ICP suffer from the intrinsic incompatibility with anatomical correspondence [40,41]. At some point, many landmark localization approaches use local, curvature-based facial features due to their invariance to surface translation and rotation. The two most frequently adopted features are the HK curvature and the shape index [35-37,42]. However, curvature-based descriptors often suffer from surface irregularities, especially near eye and mouth corners [43]. Other studies have used appearance-based methods where the facial features are modeled by basis vectors calculated from transformations such as Principal Component Analysis (PCA) [44,45], Gabor wavelets [46,47], or the Discrete Cosine Transform (DCT) [39]. However, the lowest mean localization errors (root mean square error, RMS) these approaches can achieve were around 3 ~ 5 mm [18,35-39], not accurate enough for high resolution morphometric analyses.

For biological inference, anatomical correspondence has to be established. This can be achieved by non-rigid transformations. A common method for deforming 3D surfaces is the thin-plate spline (TPS) algorithm [48]. The process of using TPS warping involves minimizing a bending energy function for a transformation over a set of fiducial points (landmarks), thereby bringing the corresponding fiducial points on each surface into alignment with each

other. A dense registration method has been developed based on TPS, and was successfully used to detect many face dysmorphology caused by rare genetic defects such as Noonan, 22q11 deletion, Bardet-Biedl and Smith-Magenis syndromes [8-13]. This approach therefore demonstrated great importance in biological and medical research. However, the TPS based registration has a key limitation that restrained its wide use in large-scale 3D facial datasets: A set of landmarks have to be localized as the anchoring points before TPS can be carried out. Methods have been developed to combine ICP-based landmark annotation and TPS warping to fully automate the registration [40,41]. However, the landmark correspondences found by ICP are not exactly anatomically homologous, as previously discussed.

In this study, we develop an automatic registration method which combines a novel solution of landmark localization and an efficient protocol of TPS-based surface registration. The landmark localization mainly employs PCA to extract landmarks on surfaces by use of both shape and texture information. For the surface registration, a new TPS warping protocol that avoids the complication of inverse TPS warping (a compulsory procedure in the conventional registration method) is used to resample the meshes according to the reference mesh. We show that this method is highly accurate and robust across different ethnicities. We also propose a new spherical resampling algorithm for re-meshing surfaces which efficiently removes the caveats and improves the mesh structure. Furthermore, the associated texture is also included in the registered data for visualization and various analyses.

Methods

Ethics statement

Sample collection in this study was carried out in accordance with the ethical standards of the ethics committee of the Shanghai Institutes for Biological Sciences (SIBS) and the Declaration of Helsinki, and has been specifically surveyed and approved by SIBS. A written statement of informed consent was obtained from every participant, with his/her authorizing signature. The participants, whose transformed facial images are used in this study as necessary illustrations of our methodology, have been shown the manuscript and corresponding figures. Aside from the informed consent for data sampling, a consent of publication was shown and explained to each participant and their authorizing signature was obtained as well.

The 3D face data set

Three-dimensional facial images were acquired from individuals of age 18 to 28 years old, among which 316 (114 males and 202 females) were Uyghurs from Urumqi, China

and 684 (363 males and 321 females) were Han Chinese from Taizhou, Jiangsu Province, China. Another training set which did not overlap with the first 1000 sample faces, consisted of 80 Han Chinese, 40 males and 40 females from Taizhou, Jiangsu Province, China. The participants were asked to pose an approximately neutral facial expression, and the 3D pictures were taken by the 3dMDface® system (www.3dmd.com). Each facial surface was represented by a triangulated, dense mesh consisting of ~30000 vertices, with associated texture (Figure 1).

Workflow

The workflow is briefly described as follows (Figure 2). Starting with a set of raw 3D face scans, the nose tip is first automatically localized on each face using a sphere fitting approach and pose normalization is performed to align all sample faces to a uniform frontal view. For the landmark annotation, the six most salient landmarks were first manually labeled on a set of training samples; Principal Component Analysis (PCA) was then employed to localize these 6 landmarks on the sample surfaces and 11 additional landmarks were heuristically annotated afterwards. A reference face was then chosen, re-meshed using spherical sampling, and TPS-warped to each sample face using the 17 landmarks as fiducial points. A dense, biological correspondence was thus built by re-meshing the sample face according to the reference face. The correspondence is further improved by using the average face as the reference and repeating the registration process.

Preliminary nose tip localization and pose normalization

In 3D facial image processing, pose normalization and landmark localization are highly dependent on each other since pose normalization is typically guided by landmarks. The features commonly used for pose correction are the nose tip and inner eye corners as they are easier to detect [35], less sensitive to pose variation, and invariant to facial expressions [42,44,49,50]. On the other hand, most existing landmark localization approaches rely on the assumption of frontal or approximately frontal poses and are therefore sensitive to roll

and yaw rotation [38,42,51]. In order to fully automate the pose normalization and landmark annotation, we first identify the most robust and prominent landmark, the nose tip.

Since the area around the nose tip can be approximated as a semi-sphere with a diameter specific to nose, we try to identify the nose tip by fitting a sphere around every vertex using its surrounding vertices. A vertex is likely the nose tip if its neighboring points fit a sphere very well and the sphere diameter approaches the specific value of the nose tip. As this method is insensitive to the pose of the face, the spherical area on nose tip can be seen as a rotation invariant descriptor (RID). The algorithm is described as follows. Let us denote a facial mesh composed of N points by $\mathbf{F} = \{\mathbf{p}_i\}$ for $i = 1, \dots, N$. Suppose \mathbf{S} is the set of M points that are within distance R around the point \mathbf{p}_j ($1 \leq j \leq N$). The best fit sphere \mathbf{T} around \mathbf{p}_j is therefore determined by two parameters, namely the center $\mathbf{O} = (a, b, c)$ and radius r . Another parameter e is the average residual distance of the M points to the best fit sphere. e describes how well the set of M points fit onto a sphere. A detailed description of sphere fitting and the calculation of e can be found in Additional file 1. The smaller e is, the better \mathbf{S} fits a sphere. The two parameters, r and e , are calculated for every point. In order to form a proper sphere of radius r around each vertex, the included distance to adjacent points (R) must be slightly larger than the radius of the sphere (r) as it is assumed that not every point will lie on the sphere. On the other hand, r should be chosen with good consistency across genders and ethnic backgrounds, thereby establishing a uniform criterion for all faces. To determine the optimal R and r , we randomly chose 50 face images from each of the four groups: Han male, Han female, Uyghur male and Uyghur female. All the 200 images were manually annotated for the nose tip. For decreasing R radius values starting at 18 mm (a value more than big enough to cover the nose tip region), the sphere fitting was carried out and the best fit r values were obtained for each image. For every R value, the average r was then calculated for each group as well as for all four groups (Additional file 2: Figure S1). The global



Figure 1 The surface used in our research. (a) The coordinate system used in our research (red, green and blue axes stand for x, y and z axes respectively). (b) An example scan with 17 landmarks marked by the colored spots. The red spots are the 6 most salient landmarks, namely the inner and outer corners of the eyes and both corners of the mouth, the blue spots indicate the other 11 landmarks used in this study. (c) Raw mesh details around the nose tip.

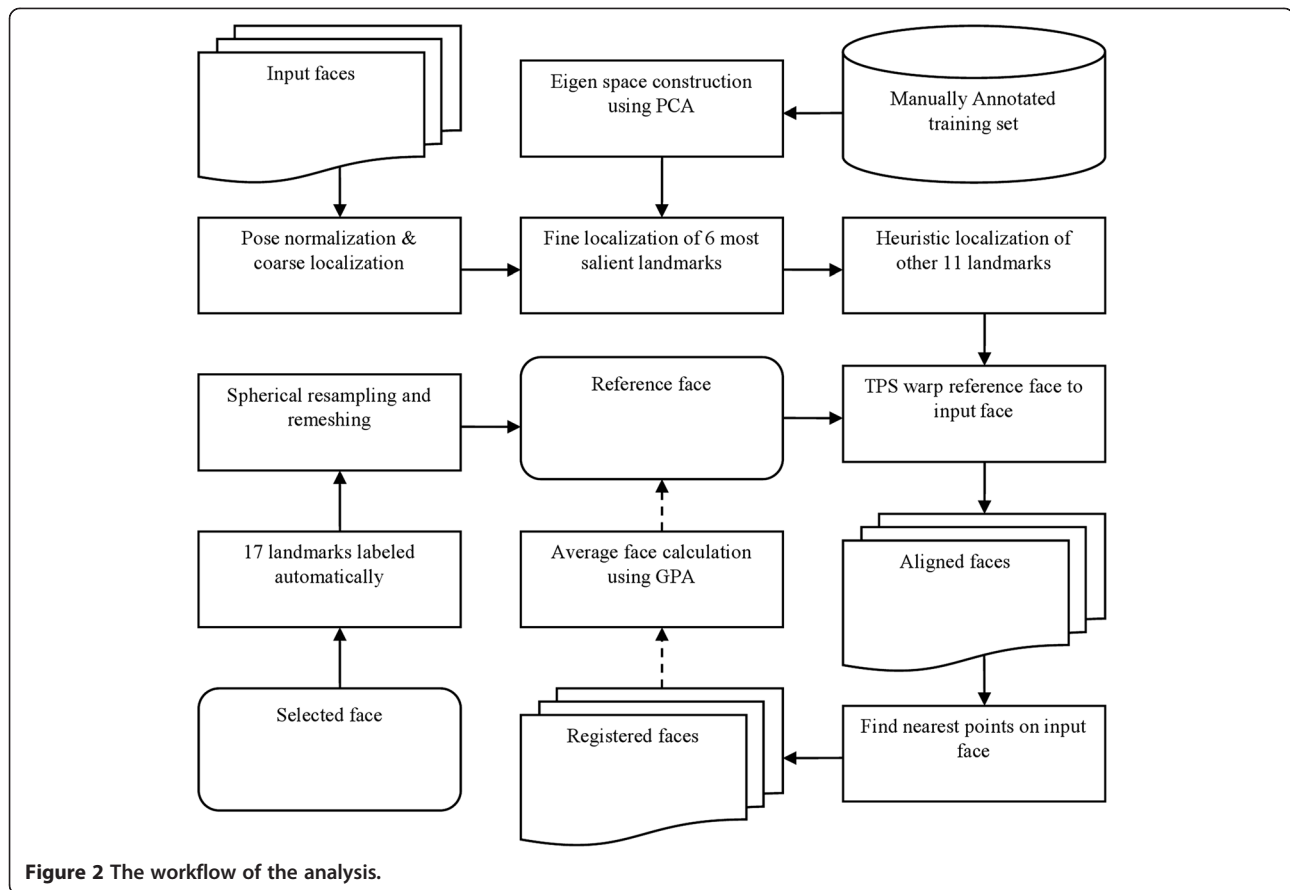


Figure 2 The workflow of the analysis.

average r value of 11.7 mm were chosen for the lowest variance across different sexes and ethnicities, denoted as r_0 ; and the corresponding R of 15 mm was also chosen and denoted as R_0 . The two spherical parameters can then be combined with the optimal r_0 radius into one statistic (f) which describes how well a given point fits the criteria for a nose tip:

$$f = e(r_0 + |r - r_0|) \quad (1)$$

The f value should be very small around the nose tip region. Indeed, we found that small f values congregated around the nose tip area (Figure 3). More interestingly, the global minima of the f values consistently appeared close to the manually annotated nose tip across hundreds of observations. We therefore use the point with the minimum f value to approximate the nose tip (Figure 3). The left column, a Han Chinese male in the middle column and an Uyghur female in the right column. Top row, the f values are shown as color gradients. Warm colors indicate convex sphere fitting, while the cold colors indicate concave to the reader. The f values deviating more from 0 are marked with greater color intensity. Central row, the minimum convex f values plotted for different individuals,

which can be seen to coincide with the manually annotated nose tips shown in the bottom row.

The pose correction becomes easy once the nose tip has been located. Correcting the pose basically consists of resetting the viewing coordinate system where an origin point and two axes must be defined. In some studies, the ICP matches are applied [52,53]. Other studies try to find landmarks (i.e. inner eye corners) other than the nose tip to determine the pose [38,48]. However, in this study we followed a rather practical solution in which all vertices within 50mm of the nose tip are used to correct the pose via the Hotelling transformation [52,53].

Localization of the six most salient landmarks using PCA

Here we propose a novel landmark localization method. The basic idea is to transform the 3D shape and texture data into a 2D space. A 2D PCA algorithm is then used to identify the six most salient landmarks, namely the inner and outer corners of the eyes and both corners of the mouth (Figure 1b). First, the image texture is converted to the YCbCr color space, in which the Y component defining the gray scale intensity is calculated as $y = 0.30r + 0.59 + 0.11b$. Only the gray scale values are used as color information for this step. For any 3D face image, the plane defined by the x and y axes is defined as the target 2D

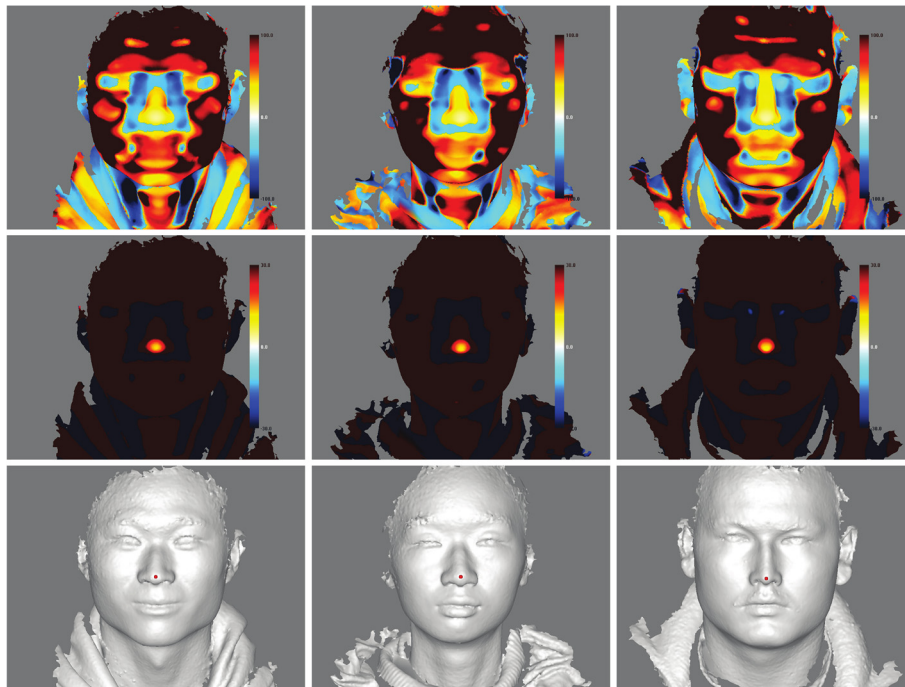


Figure 3 Nose tip localization using RID illustrated for three individuals. a Han Chinese female in the left column, a Han Chinese male in the middle column and an Uyghur female in the right column. Top row, the f values are shown as color gradients. Warm colors indicate convex sphere fitting, while the cold colors indicate concave to the reader. The f values deviating more from 0 are marked with greater color intensity. Central row, the minimum convex f values plotted for different individuals, which can be seen to coincide with the manually annotated nose tips shown in the bottom row.

space. The 3D surface and its corresponding 2D texture are then resampled on a uniform square grid at a 1mm resolution to obtain the corresponding z coordinate values and gray scale values. These values are directly mapped to the target 2D space (Figure 4). In order to minimize the data noise, the z coordinate and gray scale values are denoised using a 3×3 decision-based median filter [37]. Only the values of the outer most layer are transformed to 2D following the z buffering algorithm, particularly for the areas where the 3D surface folds into a multilayer along

the z-axis [54]. Holes that may occur inside the surface are closed by bicubic interpolation as previously described [55]. The interpolation process was done separately on texture and the 2.5D image data. The resulting 2D image combines both shape and texture information, which serves as the basis for the PCA-based landmark localization. The PCA analysis is a commonly used approach for accurate pattern recognition in 2D data [54,56,57]. It involves retrieving the feature signatures by dissecting the training data with PCA, followed by



Figure 4 The signature patch for the left lip corner illustrated in the 2D space. (a) The z coordinate values mapped into the 2D space. (b) The gray scale values mapped into the 2D space.

projecting the sample data into the PCA eigenspace to determine the similarity. In this study, the landmark signature is obtained by defining a patch of a given size, say $s\text{mm} \times s\text{mm}$, centered around the manually annotated landmark in the training set (Figure 4). Each patch therefore contains s^2 z coordinate values, which are then concatenated into a vector and normalized to have zero mean and unit length. We define it as $Z = (z_1, z_2, \dots, z_{s \times s})$. The same number of gray scale values are also concatenated into a vector and normalized to have unit length. We define it as $Y = (y_1, y_2, \dots, y_{s \times s})$. Z and Y can be combined together to specify the shape and texture properties around the landmark:

$$P = (z_1, y_1, z_2, y_2, \dots, z_{s \times s}, y_{s \times s})^T \quad (2)$$

P is then calculated for the signature eigenspace \mathbf{U} using PCA (see Additional file 1 for details). To find the landmarks in a sample face, a patch of $s\text{mm} \times s\text{mm}$ is similarly defined for every point in the corresponding 2D grid, and a sample patch vector P_s is derived following equation (2). P_s is subsequently projected to the space \mathbf{U} to evaluate its closeness to the origin point of \mathbf{U} . In this study, two measurements of closeness are used, the reconstruction error e and the Mahalanobis distance d (see Additional file 1 for details). Sample points with smaller values for e and d are more likely to be a valid landmark. Therefore, the sample point corresponding to the minimum product value of e and d is defined as the landmark in our work. The patch size parameter s inevitably affects the final localization accuracy. We formally evaluated the dependence of the accuracy on the patch size, as illustrated in Additional file 3: Figure S2. Briefly, we checked the distances of the automatically annotated landmarks to the manually annotated ones, which we define as the localization error, for a random set of 100 individuals of different ethnicities and genders (25 individuals from each of the four groups: Han male, Han female, Uyghur male and Uyghur female). We found that in general, the localization error decreases with patch size (Additional file 3: Figure S2). However the error reaches a minimum when s is around 21mm and further increasing in the patch size does not reduce the error. Therefore we use the s value of 21mm throughout this study. To further optimize the computational efficiency, we narrow down the search for each landmark to a corresponding "landmark zone" on each sample face. Briefly, an arithmetic mean is calculated for each landmark across the training set, and projected onto the 2D space. Rectangular areas around the projection points are then defined as the landmark zones, with their sizes set experimentally (i.e. by training through a large number of faces) to ensure all real landmarks are encompassed. Therefore, the

search for a particular landmark is done only within this landmark zone.

Heuristic localization of ten additional landmarks

Given the annotation of the six most salient landmarks, the pose of the surface can be fine tuned again. The reference plane is set to be the best fit plane to the six landmarks by least squares. The normal to the reference plane is set to be the z axis, and the y axis is given by the projection of the line going through the centers of lip corners and the eye corners onto the reference plane. The x axis is uniquely determined afterwards.

After the pose correction, 10 additional landmarks are identified heuristically by using geometric relations and texture constraints and the nose tip position is also updated. These landmarks include soft tissue nasion, alares, subnasale, labiale superius (upper lip point), stomion (the middle point between the upper and lower lip), labiale inferius (lower lip point), pogonion (chin point), and earlobe tips. The nose tip can be fine tuned according to the more uniformly defined coordinate system across all sample surfaces. Briefly, a semi-sphere is refitted around the previous nose tip and the point that minimizes the z coordinate error is chosen as the new nose tip. The subnasale point can be located by finding the inflection point with the minimum angle right below the nose tip. The alare points are the inflection points with the minimum local angles going horizontally away from the nose tip. Similar angle heuristics are applied to the detection of labiale superius, inferius, and stomion, with additional texture information in the YCbCr color space. For example, the labiale superius should locate the position on the border line where the Cr values below the line are greater (more red) than those above. Noticing that the region around the nasion point is approximately saddle-shaped and that of the chin point is ellipsoidal or sphere-shaped, both characterized by the two-way symmetry, we therefore locate the two points by finding the maximum local symmetry scores. The earlobe points are easily found by locating the tips with sheer slopes along the z-axis.

Spherical resampling and surface remeshing

During the 3D image acquisition, the surface meshes often suffer from mesh structure irregularities and/or defects such as mesh holes (see Figure 1c for example). Surface remeshing is often used to solve such problems [58]. In this work, we apply spherical sampling to the reference surface to obtain a well-structured mesh. Spherical sampling is preferred as human faces are approximately ellipsoidal. We first perform a spherical parameterization to the surface using the geographic coordinates. Given a vertex (x_ρ, y_ρ, z_ρ) on the original surface mesh, the spherical parameterization $(\rho_\rho, \theta_\rho, \phi_\rho)$ can be obtained as follows:

$$\begin{aligned}\rho_i &= \sqrt{2x_i^2 + y_i^2 + z_i^2} \\ \theta_i &= \arcsin(y_i/\rho_i) \\ \phi_i &= \arctan(\sqrt{2}x_i/z_i)\end{aligned}\quad (3)$$

The x-coordinate is multiplied by a factor $\sqrt{2}$ before the coordinate conversion, to compensate for the face aspect ratio (height to width) [53]. When plotted against θ and ϕ , the parameterized surface unfold into a nearly flat plane. This surface is then trimmed with an oval path to remove the irregular edges and re-sampled from a uniform square grid with an interval of 0.005 for both θ and ϕ . The re-sampled data points are then converted back to the Cartesian coordinate system to define a new surface mesh.

Surface registration for dense correspondence

In order to preserve the anatomical correspondence across the facial surfaces, we adopted the idea of the TPS-based registration method proposed previously [59]. In that study, all surfaces were first manually annotated for a set of landmarks. The sample surfaces and the reference were all TPS warped to the cross-sample average landmarks. Each sample surface was then re-meshed by the closest points to the reference vertices, and further inverse TPS warped back to the original shape. Mathematically, TPS warping is not invertible. Although an approximation exists, it is computationally intensive and error prone [60]. In our study, we designed an alternative scheme. First, a well-structured surface with few defects is chosen as the reference face, and spherically remeshed as described above. Then only the reference surface is TPS warped to each sample surface, taking the 17 landmarks as the fiducial points. The TPS warping is done as previously described [13]. Thereafter the vertices on the reference surface find their closest projections on the sample surface, which define the new mesh vertices of the sample surface [13,61]. The dense correspondence is established after all the sample surfaces are remeshed using the same reference. This approach eliminates the need for inverse TPS warping, and enhances the computational efficiency as well.

Results

Accuracy of the landmark localization

In this section we demonstrate the accuracy of the proposed algorithm for automatic landmark localization. The accuracy is measured by the deviation of the automatically annotated landmarks from those manually annotated.

A subset of the sample surfaces were picked randomly and manually annotated with the 17 landmarks by the experimenter who did the same to the training set. Automatic landmark annotation was also performed independently. The surfaces missing some features such as the earlobes

were removed from further analysis. This left 115 Han Chinese (56 males, 59 females) and 124 (48 males, 76 females) Uyghur for the evaluation. The mean and standard deviation (SD) of the annotation errors measured in Euclidean distance, as well as the root mean square (RMS) errors were calculated (Table 1). As can be seen from Table 1, most landmarks have mean errors between 1mm and 1.5mm, indicating rather high accuracy. Most of the SD values are below 1mm, suggesting good consistency of the annotation across different samples. The RMS error is within the range of 1.1 ~ 2 mm for most measurements. Greater errors are found for the Pogonion (~1.8 mm mean error for both the Han Chinese and Uyghur) and the two earlobe tips (mean error 2 ~ 3 mm, SD error 1.6 ~ 2.2 mm and RMS error 2.6 ~ 3.6 mm). Pogonion and earlobes are both strongly affected by facial/head hair, which may account for the relatively larger errors and standard deviations. It is worth noticing that all the error values are similar between the Han Chinese and Uyghur samples despite the use of the Han Chinese training set. Given the substantial genetic and morphological differences between these two ethnic populations, this indicates good robustness of our method when applied to different ethnicities.

Robustness of the registration method

One way to evaluate the robustness of the registration method is to determine how the use of different references would affect the correspondence mapping. We performed such an evaluation, as shown in Figure 5. First, we obtained the average Han Chinese male and female faces by registering all the Han Chinese samples to the same reference surface, followed by obtaining the average meshes across either gender group (average face calculation is explained in more detail in the next section). These average faces are point-to-point corresponded. We can see the two average faces differ substantially in their shape (Figures 5a and c). To test the robustness of the registration method, a test face (Figure 5b) is chosen randomly from the data set, and registered separately using either average face as the reference. Euclidian distances are calculated for the pairing points between the two newly registered meshes. One expects to see small differences between the two registration results if the method is robust. Figure 5d shows that most parts have point-wise errors much less than 0.9 mm, which indicates the robustness of our registration method with varying references. Certain regions like eyebrows exhibit greater errors, most likely due to the mesh irregularities caused by facial hair.

The average faces calculation with the 3D face registration

We applied the proposed 3D face registration method to the whole 3D face sample set. In total 363 male and 321 female Han Chinese and 114 male and 202 female Uyghur

Table 1 Mean error and standard deviation of 17 automatically annotated landmarks

Landmarks	Han			Uyghur		
	Mean error (mm)	SD error (mm)	RMS error	Mean error (mm)	SD error (mm)	RMS error
Right eye outer corner	1.339	0.947	1.641	1.511	1.091	1.864
Right eye inner corner	1.192	1.155	1.660	1.280	0.914	1.573
Left eye inner corner	1.162	0.809	1.416	1.489	0.896	1.738
Left eye outer corner	1.148	0.738	1.365	1.507	1.008	1.814
Right lip corner	0.995	0.581	1.153	1.424	0.817	1.642
Left lip corner	1.012	0.569	1.161	1.147	0.693	1.341
Nose tip	0.761	0.747	1.067	1.113	0.549	1.242
Nasion	1.487	0.654	1.625	1.604	0.833	1.808
Right alare	1.310	0.752	1.511	1.034	0.742	1.273
Left alare	1.480	0.798	1.682	1.128	0.577	1.268
Lip center	1.189	0.715	1.388	1.145	0.812	1.404
Upper lip	1.270	0.769	1.485	1.727	1.138	2.069
Lower lip	1.380	0.855	1.624	1.501	0.941	1.772
Subnasale	1.299	1.263	1.812	0.999	0.574	1.153
Pogonion	1.809	1.878	2.608	1.785	0.882	1.992
Right earlobe tip	2.074	1.658	2.656	2.866	2.18	3.601
Left earlobe tip	2.678	1.640	3.141	2.835	2.054	3.501

were included in this analysis. All surfaces were automatically annotated. One Han Chinese face with few caveats and fully exposed skin was chosen as the reference, to which all the sample faces were registered. The Generalized Procrustes Analysis (GPA) was then used to align all the registered surfaces to a common coordinate [62]. The

average faces were then calculated as the average meshes colored by the corresponding average texture pixels across all the samples in each group. Figure 6 shows the average faces of the four groups. As can be seen, the average faces well retain the morphological and textural features of each group. Specifically, the Uyghur are characterized by a

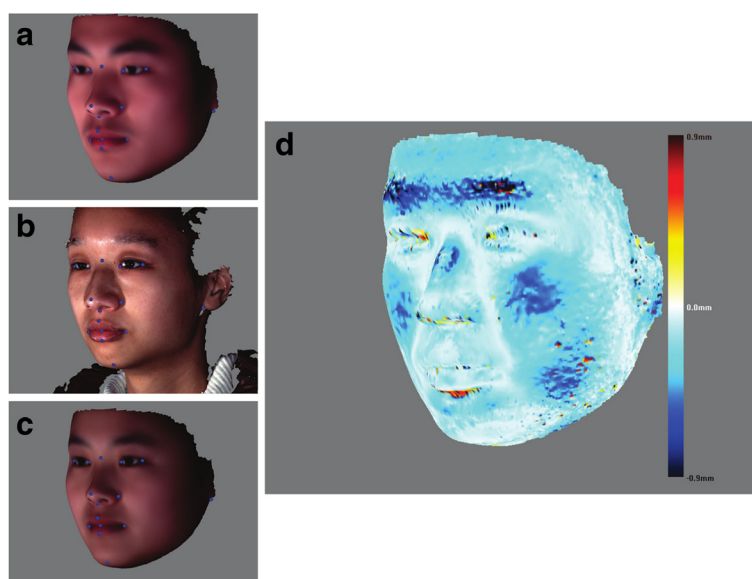


Figure 5 Evaluation of the robustness of the registration method. The average face of either gender is used as the reference to register the sample surface, and the registration results are compared. **(a)** The average face of male Han Chinese, **(b)** The sample face to be registered, **(c)** The average face of female Han Chinese, **(d)** The comparison of the two registration results. The differences are represented in color gradients, with the darker colors denoting greater pointwise differences.

more protruding nose and eyebrow ridges while Han Chinese have wider cheeks. The skin pigmentation also seems lighter for the Uyghur compared to the Han Chinese. This difference could not be confirmed as the environmental light was not well controlled when the 3D images were taken.

Discussion

In this work we propose a fully automatic registration method for high resolution 3D facial images. This method combines automatic landmark annotation and TPS-based registration. Previous TPS based automatic registration methods suffer from low accuracy in landmark localization [40,41], and are not suitable for anthropometric studies. For the problem of landmark localization, most time-honored solutions deal with only 2.5D data, leaving out the texture information. In particular, Perakis et al. described a method that made use of a comprehensive list of local shape descriptors, and achieved a precision of around 4 mm [63]. Szeptycki et al. combined curvature analysis with a generic face model in a coarse-to-fine workflow, which enabled rotation invariant 3D landmark annotation at a precision of around 10 mm [37]. On the other hand, D'Hose et al. made use of the Gabor wavelets to extract curvature information for coarse landmark localization, followed by an ICP-based fine mapping [47]. This study achieved an overall precision level of a bit over 3 mm [47]. Hutton et al. developed a method called the dense surface model (DSM), which hybridized the ICP optimization and active shape model (ASM) fitting to enable the automatic

registration of 3D facial surfaces [12]. They demonstrated that for the ten studied landmarks, the estimated positions using the DSM method have relatively small RMS errors (~ 3 mm) from the manual annotations. In this study, we constructed a novel PCA based landmark localization method, which made use of both the 3D geometric and 2D texture information, and achieved much lower landmark RMS errors, 1.7 ~ 1.8 mm on average, for a bigger number (17) of landmarks (Table 1). If the less salient landmarks, such as the earlobe tips, are excluded from the analysis, the errors will decrease further (Table 1). The novel use of both shape and texture information played a key role in improving the landmark localization accuracy. We found that the positions of some salient landmarks such as the eye corners are ambiguous even manually when the texture is removed. Texture gives rich information about the facial anatomical layout, such as the boundaries of different skin/tissue types. In fact, texture is almost the only information source for pattern recognition in 2D images and has been shown to give good performance. We projected both the shape and texture data into the 2D space, where the well-established PCA algorithm was used to detect the key landmarks. We also made use of the texture information for detecting certain other landmarks. Furthermore, due to the use of simple and optimized algorithms, the landmark annotation is also very efficient and does not require large amounts of memory. Hundreds of surfaces can be annotated within several minutes on a standard Windows PC. It is known that PCA can give wrong results on multi-modal feature distributions. This is

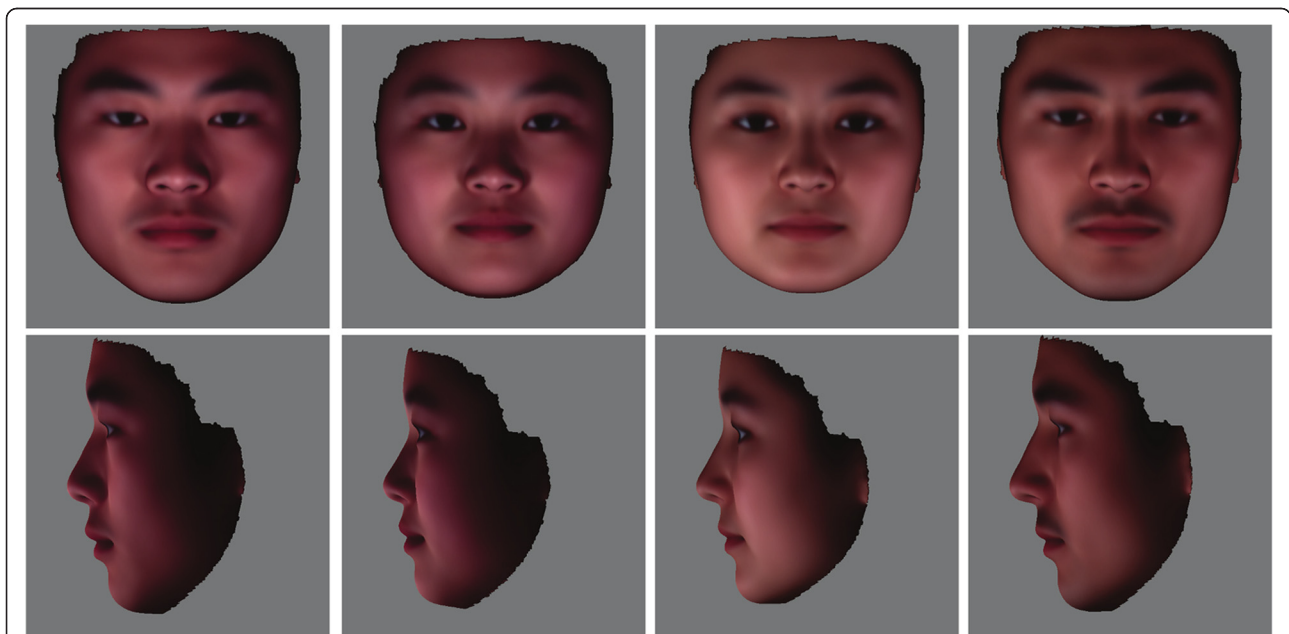


Figure 6 Front and profile views of the average faces. From left to right: Han Chinese male, Han Chinese Female, Uyghur female, and Uyghur male.

particular the case when the surface data is taken under very different poses or light conditions, or image data contains non-Gaussian noises like spikes, holes and nevus. Under such conditions, PCA gives no guarantee that the modes of maximum variance (the principal component axis) are accounted for by the features of interest rather than noises. However, such problems are minimized in our study. First of all, for the purpose of anthropometric analyses, all surface images are supposed to be taken in laboratory environment where pose and light condition are well controlled to suppress unwanted variations. Moreover, several pre-processing and classification method were adopted to increase the specificity: 1. Median filter is used to reduce non-Gaussian noises; 2, both local patch vectors are normalized to have zero mean and unit length to eliminate bias to either modality; 3, the reconstruction error e was used along with the Mahalanobis distance for pattern classification. After all, the landmark recognition errors presented in Table 1 were assessed across all the individuals in the test panels. Unless obvious imaging errors were observed and the corresponding images removed, the errant landmark localizations were not specifically filtered from the test panels. The low average recognition errors therefore support the robustness of our method. The landmark localization may be further improved in the performance. For example, in the PCA based localization, one may apply bigger patch sizes but use a subset of points within each patch to construct the signature vector P . This may reduce the redundant information and result in better accuracy or efficiency. On the other hand, the gain in accuracy may also be partially attributed to the higher image resolution of our data (~30,000 vertices per surface on average) compared to the previous work (~10,000 vertices per surface). Furthermore, we also proposed a new protocol for the TPS-based registration, whereby the TPS warping was only applied to the reference face while the sample faces remained undeformed and thus avoided the step of inverse TPS warping, thereby further increasing the efficiency of our method. It is interesting to note that both the automatic landmark annotation and the TPS based registration steps work equally well for two different ethnicities, namely Han Chinese and Uyghur, in spite of the fact that they are substantially different in both genetic background and facial appearance. Han Chinese are representative of East Asian populations while Uyghur is an ancient admixture population whose ancestries came from both East Asians and Caucasians (European people) [64]. As a result, Uyghur participants exhibited many Caucasian facial features such as sunken eyes and high nose ridge, etc. (Figure 6). This method was also tested on other ethnic groups and showed consistent robustness (data not shown). Such ethnicity independency is very important when this method is used to study the cross population facial morphological variations in humans.

It should be noted that the aim of this study is not to propose a general scheme of 3D surface registration. Rather, our method combines various pragmatic solutions to construct an optimized pipeline for high-throughput registration of dense 3D facial images. To the authors' knowledge, this is the only fully automatic non-rigid registration method that aligns dense 3D face images at a landmark accuracy of ~2 mm. In fact, high resolution non-rigid registration methods that maximize the anatomical correspondence can greatly promote the biological and medical researches of the corresponding organs/features. A formal TPS based 3D face registration method, DSM, has revealed great potential in series of medical genetic studies of face dysmorphology [8-13]. In the field of brain research, efficient non-rigid methods, such as HAMMER [6], TPS-HAMMER [65] and Brainaligner [5] were developed for specific image data, and successfully applied in detection of aging/disease induced brain morphological changes [66] and delineation of ultrastructures [67] and neuronal circuits [68] of brain. Given the high definition and accuracy, our method may have many potential applications, such as quantitative characterization of human face diversity and detection of genetic/environmental factors that can induce facial shape changes. Furthermore, as this method is robust to diverse ethnicities, it is particularly suitable for studying evolution and divergence of human face among different populations. A standalone software has been implemented for this method, and is freely available for academic use upon request.

In the future, the anatomic correspondence can be further improved by including additional features such as the eyebrows, eyelid lines, and lip lines as landmarks. These features may provide discrimination power towards different facial expressions.

Conclusions

In summary, this study proposes a new scheme to build accurate and robust anatomical correspondence across dense surfaces of 3D facial images; and it was implemented into a fully automatic and efficient registration package. This method enables high-throughput capture and analysis of the wide ranging and yet fine detailed variations within human facial morphology. Such comprehensive and high resolution phenotypic data should be valuable in anthropological, disease diagnosis, and forensic studies of human facial morphology.

Additional files

Additional file 1: Appendix.

Additional file 2: Figure S1. Mean r value of different human groups with respect to different R values. For each group, 50 faces were analyzed. For each face, the nose tip was first annotated and all neighboring points within distance R were used to calculate the r value.

Additional file 3: Figure S2. Localization error (mean error) of the 6 most salient landmarks with respect to the varying patch sizes. 100 faces were analyzed, of which 25 were Han male, 25 were Han female, 25 were Uygur male, and 25 were Uygur female.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

GJ designed the new algorithm of landmark localization, implemented and tested all the methods and drafted the manuscript; MX contributed to the design of the landmark localization algorithm and manuscript writing; TK initiated and supervised this project, contributed to the image collection and manuscript writing. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the Cutting Edge Research Project (Grant No. 2011KIP201) from the CAS-SIBS Outstanding Junior Researchers, the Key Research Direction Grant (No. KSCX2-EW-Q-1-12) from the CAS Knowledge Innovation Project and the Max-Planck-Gesellschaft Partner Group Grant. The funders had no role in study design, data collection or analysis, decision to publish, or preparation of the manuscript. Prof. Dr. Mark Stoneking from the Max Planck Institute for Evolutionary Anthropology has made valuable contributions to this study and provided proof-reading of the manuscript.

Received: 17 January 2013 Accepted: 15 July 2013

Published: 22 July 2013

References

- Ohya Y, Sese J, Yukawa M, Sano F, Nakatani Y, Saito TL, Saka A, Fukuda T, Ishihara S, Oka S: **High-dimensional and large-scale phenotyping of yeast mutants.** *Proc Natl Acad Sci USA* 2005, **102**(52):19015.
- Tobias C, Roberto H, Falk S: **HTPheno: an image analysis pipeline for high-throughput plant phenotyping.** *BMC Bioinforma* 2011, **12**(1):148.
- Long F, Peng H, Liu X, Kim SK, Myers E: **A 3D digital atlas of *C. elegans* and its application to single-cell analyses.** *Nat Methods* 2009, **6**(9):667–672.
- Kristensen E, Parsons TE, Hallgrímsson B, Boyd SK: **A novel 3-D image-based morphological method for phenotypic analysis.** *Biomedical Engineering, IEEE Transactions* 2008, **55**(12):2826–2831.
- Peng H, Chung P, Long F, Qu L, Jenett A, Seeds AM, Myers EW, Simpson JH: **BrainAligner: 3D registration atlases of *Drosophila* brains.** *Nat Methods* 2011, **8**(6):493–500.
- Shen D, Davatzikos C: **HAMMER: hierarchical attribute matching mechanism for elastic registration.** *IEEE Trans Med Imaging* 2002, **21**(11):1421–1439.
- Li B, Christensen GE, Hoffman EA, McLennan G, Reinhardt JM: **Establishing a normative atlas of the human lung: intersubject warping and registration of volumetric CT images.** *Acad Radiol* 2003, **10**(3):255–265.
- Hammond P: **The use of 3D face shape modelling in dysmorphology.** *Arch Dis Child* 2007, **92**(12):1120.
- Hammond P, Hutton T, Allanson J, Buxton B, Karmiloff-Smith A, Patton M, Pober B, Smith A, Tassabehji M: **3D dense surface models identify the most discriminating facial features in dysmorphic syndromes.** Toronto, Canada: 54th Annual Meeting of the American Society for Human Genetics; 2004.
- Hammond P, Hutton TJ, Allanson JE, Buxton B, Campbell LE, Clayton-Smith J, Donnai D, Karmiloff-Smith A, Metcalfe K, Murphy KC, et al: **Discriminating power of localized three-dimensional facial morphology.** *Am J Hum Genet* 2005, **77**(6):999–1010.
- Hammond P, Hutton TJ, Allanson JE, Campbell LE, Hennekam RC, Holden S, Patton MA, Shaw A, Temple IK, Trotter M, et al: **3D analysis of facial morphology.** *Am J Med Genet A* 2004, **126A**(4):339–348.
- Hutton TJ, Buxton BF, Hammond P: **Automated registration of 3D faces using dense surface models.** Citeseer: Proceedings of British Machine Vision Conference; 2003:439–448.
- Hutton TJ, Buxton BF, Hammond P, Potts HWW: **Estimating average growth trajectories in shape-space using kernel smoothing.** *Medical Imaging, IEEE Transactions on* 2003 2003, **22**(6):747–753.
- Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, Christensen GE, Collins DL, Gee J, Hellier P, et al: **Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration.** *Neuro Image* 2009, **46**(3):786–802.
- Farkas LG, Katic MJ, Forrest CR: **International anthropometric study of facial morphology in various ethnic groups/races.** *J Craniofac Surg* 2005, **16**(4):615.
- Weinberg SM, Neiswanger K, Richtsmeier JT, Maher BS, Mooney MP, Siegel MI, Marazita ML: **Three-dimensional morphometric analysis of craniofacial shape in the unaffected relatives of individuals with nonsyndromic orofacial clefts: a possible marker for genetic susceptibility.** *Am J Med Genet A* 2008, **146**(4):409–420.
- Hammond P, Forster-Gibson C, Chudley AE, Allanson JE, Hutton TJ, Farrell SA, McKenzie J, Holden JJA, Lewis MES: **Face and brain asymmetry in autism spectrum disorders.** *Mol Psychiatry* 2008, **13**(6):614–623.
- Albert AM, Ricanek K Jr, Patterson E: **A review of the literature on the aging adult skull and face: implications for forensic science research and applications.** *Forensic Sci Int* 2007, **172**(1):1–9.
- Kayser M, De Knijff P: **Improving human forensics through advances in genetics, genomics and molecular biology.** *Nat Rev Genet* 2011, **12**(3):179–192.
- Meyer-Marcotty P, Alpers GW, Gerdes A, Stelzig-Eisenhauer A: **Impact of facial asymmetry in visual perception: a 3-dimensional data analysis.** *Am J Orthod Dentofacial Orthop* 2010, **137**(2):168–e161.
- Little AC, Jones BC, DeBruine LM: **Facial attractiveness: evolutionary based research.** *Philosophical Transactions of the Royal Society B: Biological Sciences* 2011, **366**(1571):1638–1659.
- Ramanathan N, Chellappa R, Biswas S: **Computational methods for modeling facial aging: a survey.** *J Vis Lang Comput* 2009, **20**(3):131–144.
- Fu Y, Guo G, Huang TS: **Age synthesis and estimation via faces: a survey.** *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 2010 2010, **32**(11):1955–1976.
- Liu F, van der Lijn F, Schurmann C, Zhu G, Chakravarty MM, Hysi PG, Wollstein A, Lao O, de Bruijne M, Ikram MA, et al: **A genome-wide association study identifies five loci influencing facial morphology in Europeans.** *PLoS Genet* 2012, **8**(9):e1002932.
- Paternoster L, Zhurov AI, Toma AM, Kemp JP, St Pourcain B, Timpson NJ, McMahon G, McArdle W, Ring SM, Smith GD, et al: **Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position.** *Am J Hum Genet* 2012, **90**(3):478–485.
- Richardson ER: **Racial differences in dimensional traits of the human face.** *Angle Orthod* 1980, **50**(4):301–311.
- Allanson JE, O'Hara P, Farkas LG, Nair RC: **Anthropometric craniofacial pattern profiles in down syndrome.** *Am J Med Genet* 1993, **47**(5):748–752.
- Dryden I, Mardia K: *Statistical analysis of shape.* Chichester: Wiley; 1998.
- Yamaguchi T, Maki K, Shibasaki Y: **Growth hormone receptor gene variant and mandibular height in the normal Japanese population.** *Am J Orthod Dentofacial Orthop* 2001, **119**(6):650–653.
- Tomoyasu Y, Yamaguchi T, Tajima A, Nakajima T, Inoue I, Maki K: **Further evidence for an association between mandibular height and the growth hormone receptor gene in a Japanese population.** *Am J Orthod Dentofacial Orthop* 2009, **136**(4):536–541.
- Weinberg S, Naidoo S, Bardi K, Brandon C, Neiswanger K, Resick J, Martin R, Marazita M: **Face shape of unaffected parents with cleft affected offspring: combining three-dimensional surface imaging and geometric morphometrics.** *Orthod Craniofac Res* 2009, **12**(4):271–281.
- Ermakov S, Rosenbaum MG, Malkin I, Livshits G: **Family-based study of association between ENPP1 genetic variants and craniofacial morphology.** *Ann Hum Biol* 2010, **37**(6):754–766.
- Boehringer S, Van Der Lijn F, Liu F, Günther M, Sinigerova S, Nowak S, Ludwig KU, Herberz R, Klein S, Hofman A: **Genetic determination of human facial morphology: links between cleft-lips and normal variation.** *Eur J Hum Genet* 2011, **19**(11):1192–1197.
- Besl PJ, McKay ND: **A method for registration of 3-D shapes.** *IEEE Trans Pattern Anal Mach Intell* 1992, **14**(2):239–256.
- Creusot C, Pears N, Austin J: **3D face landmark labelling.** Proceedings of the ACM workshop on 3D object retrieval. New York, NY, USA: ACM; 2010:27–32.
- Dibeklioglu H, Salah AA, Akarun L: **3D facial landmarking under expression, pose, and occlusion variations.** 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems; 2008. Arlington, VA, USA: IEEE; 2008:1–6.

37. Szeptycki P, Ardabilian M, Chen L: *A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking*, IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems: 2009. Washington, DC, USA: IEEE; 2009:1–6.
38. Chang KJ, Bowyer W, Flynn PJ: *Multiple nose region matching for 3D face recognition under varying facial expression*. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 2006 2006, **28**(10):1695–1700.
39. Salah AA, Inar H, Akarun L, Sankur B: *Robust facial landmarking for registration*. *Ann Telecommun* 2007, **62**(1–2):1608–1633.
40. Ifranoglu MO, Gokberk B, Akarun L: *3D shape-based face recognition using automatically registered facial surfaces*, Proceedings of the 17th International Conference on Pattern Recognition: 2004. Washington, DC, USA: IEEE; 2004:183–186.
41. Schneider DC, Eisert P, Herder J, Magnor M, Grau O: *Algorithms for automatic and robust registration of 3d head scans*. *Journal of Virtual Reality and Broadcasting* 2010, **7**:7.
42. Colombo A, Cusano C, Schettini R: *3D face detection using curvature analysis*. *Pattern recognition* 2006, **39**(3):444–455.
43. Kakadiaris I, Passalis G, Toderici G, Murtuza N, Theoharis T: *3D face recognition*. Edinburgh, UK: Proceedings of the British Machine Vision Conference: 2006; 2006:869–868.
44. Nair P, Cavallaro A: *Matching 3D faces with partial data*. Leeds, UK: Proc British Machine Vision Conference: 2008; 2008:1–4.
45. Abate AF, Nappi M, Riccio D, Sabatino G: *2D And 3D face recognition: a survey*. *Pattern Recognit Lett* 2007, **28**(14):1885–1906.
46. Wang Y, Chua CS, Ho YK: *Facial feature detection and face recognition from 2D and 3D images*. *Pattern Recognit Lett* 2002, **23**(10):1191–1202.
47. D'Hose J, Colineau J, Bichon C, Dorizzi B: *Precise localization of landmarks on 3d faces using gabor wavelets*, First IEEE International Conference on Biometrics: Theory, Applications, and Systems. Crystal City, VA, USA: IEEE; 2007:1–6.
48. Bookstein FL: *Principal warps: thin-plate splines and the decomposition of deformations*. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 1989 1989, **11**(6):567–585. Washington, DC, USA.
49. Sun Y, Yin L: *Automatic pose estimation of 3D facial models*, 19th International Conference on Pattern Recognition: 2008. Tampa, FL, USA: IEEE; 2008:1–4.
50. Lu X, Jain AK, Colbry D: *Matching 2.5 D face scans to 3D models*. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 2006 2006, **28**(1):31–43. Washington, DC, USA.
51. Salah AA, Alyz N, Akarun L: *Registration of three-dimensional face scans with average face models*. *Journal of Electronic Imaging* 2008, **17**:011006.
52. ter Haar FB, Veltkamp RC: *A 3D face matching framework for facial curves*. *Graph Model* 2009, **71**(2):77–91.
53. Mian AS, Bennamoun M, Owens R: *An efficient multimodal 2D-3D hybrid approach to automatic face recognition*. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 2007, **29**(11):1927–1943.
54. Tsalkanidou F, Tzovaras D, Srinivas MG: *Use of depth and colour eigenfaces for face recognition*. *Pattern Recognit Lett* 2003, **24**(9–10):1427–1435.
55. Mian AS, Bennamoun M, Owens R: *Keypoint detection and local feature matching for textured 3D face recognition*. *Int J Comput Vis* 2008, **79**(1):1–12.
56. Chang K, Bowyer K, Flynn P: *Face recognition using 2D and 3D facial data*. Santa Barbara, California, USA: ACM Workshop on Multimodal User Authentication: 2003; 2003:25–32.
57. Turk MA, Pentland AP: *Face recognition using eigenfaces*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition: 1991. Maui, HI, USA: IEEE; 1991:586–591.
58. Alliez P, Ucelli G, Gotsman C, Attene M: *Recent advances in remeshing of surfaces*. *Shape analysis and structuring* 2008:53–82.
59. Hutton TJ, Buxton BR, Hammond P: *Dense surface point distribution models of the human face*, IEEE Workshop on Mathematical Methods in Biomedical Image Analysis: 2001. Kauai, HI, USA: IEEE; 2001:153–160.
60. Johnson H, Christensen G: *Landmark and intensity-based, consistent thin-plate spline image registration*, Information Processing in Medical Imaging: 2001. London, UK: Springer; 2001:329–343.
61. Schneider PJ, Eberly DH: *Geometric tools for computer graphics*. Morgan Kaufmann Pub; 2003.
62. Gower JC: *Generalized procrustes analysis*. *Psychometrika* 1975, **40**(1):33–51.
63. Perakis P, Passalis G, Theoharis T, Kakadiaris IA: *3D Facial landmark detection & face registration*. Tech Rep: University of Athens; 2011.
64. Xu S, Huang W, Qian J, Jin L: *Analysis of genomic admixture in Uyghur and its implication in mapping strategy*. *Am J Hum Genet* 2008, **82**(4):883–894.
65. Wu G, Yap PT, Kim M, Shen D: *TPS-HAMMER: improving HAMMER registration algorithm by soft correspondence matching and thin-plate splines based deformation interpolation*. *NeuroImage* 2010, **49**(3):2225–2233.
66. Lao Z, Shen D, Xue Z, Karacali B, Resnick SM, Davatzikos C: *Morphological classification of brains via high-dimensional shape transformations and machine learning methods*. *NeuroImage* 2004, **21**(1):46–57.
67. Yang JS, Awasaki T, Yu HH, He Y, Ding P, Kao JC, Lee T: *Diverse neuronal lineages make stereotyped contributions to the Drosophila locomotor control center, the central complex*. *J Comp Neurol* 2013, **521**(12):Sp1.
68. Helmstaedter M, Mitra PP: *Computational methods and challenges for large-scale circuit mapping*. *Curr Opin Neurobiol* 2012, **22**(1):162–169.

doi:10.1186/1471-2105-14-232

Cite this article as: Guo et al.: Automatic landmark annotation and dense correspondence registration for 3D human facial images. *BMC Bioinformatics* 2013 **14**:232.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

