Asia Pacific Academy of Science Pte. Ltd.

Metaverse 2024, 5(1), 2476.
https://doi.org/10.54517/m.v5i1.2476

Article

Automatic labeling of 3D facial acupoint landmarks

Junjie Yin[1], Meie Fang[1,*], Weiyin Ma[2]

[1] Metaverse Research Institute, School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou 510006, China
[2] Department of Mechanical Engineering, City University of Hong Kong, Hong Kong, China
*** Corresponding author:** Meie Fang, fme@gzhu.edu.cn

CITATION

Yin J, Fang M, Ma W. Automatic labeling of 3D facial acupoint landmarks. Metaverse. 2024; 5(1): 2476.
https://doi.org/10.54517/m.v5i1.2476

ARTICLE INFO

Received: 8 January 2024
Accepted: 6 March 2024
Available online: 19 March 2024

Abstract: As special marks on a human face, facial landmarks reflect the facial features of various parts of the face, which is crucial in biomedicine and medical imaging. In addition, facial landmarks are also important features in computer vision such as face detection, face recognition, facial pose estimation, and facial animation. In this paper, we construct a 3D facial acupoint annotated dataset by labeling 37 facial acupoints on 846 neutral face triangle mesh on the FaceScape dataset. Based on these annotated data, we use a feature template matching method to realize the automatic annotation of 37 acupoints on triangle meshes. We used 40 meshes as the training set to extract the geometric patterns of 3D acupoints and then measured the performance of the automatic labeling algorithm on 20 meshes and 806 meshes as the test sets. In the training process, we extract the tangent plane for each landmark, project the neighbor vertices of the landmark to the tangent plane, and construct the feature image with $R \times R$ resolution through the bounding box of the projected points. In the testing process, we use the pattern images extracted during training to find the average features and use them as a guide to optimize the predicted landmarks. The experimental results show that our automatic acupoint labeling method has achieved good results.

Keywords: landmark detection; local geometric feature; acupoints; dataset

1. Introduction

The human face is as innate as other biological characteristics of the human body (such as fingerprints, iris, etc.), and its uniqueness provides the necessary prerequisite for the identity of human beings. The characteristics of the human face are usually abstracted into facial landmarks. Therefore, fields such as biomedicine [1], medical imaging [2], and computer vision will widely use facial landmarks to deconstruct and analyze facial features. In medicine, such as orthodontics, X-ray cephalometry is a simple and effective way to observe craniofacial structure and internal connections. Through the marking points described on the X-ray cephalometric film, the line distances, angles, and proportions are measured to understand the structure and relationship of soft and hard tissues such as craniofacial and teeth can be found to find cost-effective indicators to quickly and accurately guide clinical diagnosis and treatment. In computer vision tasks such as face recognition, registration, craniofacial analysis [3], 3D face reconstruction [4], facial expression analysis [5], head pose estimation [6], and many other tasks, we first automatically located some facial landmarks on the input face data, such as eye corners, nose tip, mouth corners, face contour, etc., and then complete complex visual tasks through the processing and analysis of these landmarks.

Facial landmarks can reflect the characteristics of various parts of the face. With the development of technology and the increase in accuracy requirements, the

number of facial key points has grown from 5 points in the early stage to more than 200 points. In visual tasks, the basic function of the key points is to describe the structure of the face. Therefore, some regions that appear to have special structures are usually selected subjectively as landmarks, and one or several points are used to describe a specific part of the face, such as one landmark to represent the nose tip, two eye corners to identify the eyes, and two mouth corners are used to represent the mouth, which provides convenience for the structured description of the face. In medical auxiliary tasks, the selection of facial landmarks needs to consider the deeper facial skull structure, not just the facial surface, such as the pronasale (the most prominent point of the nose, that is, the tip of the nose), the orbitale (the lowest point of the inferior orbital margin), and the supramental (the most concave point of the bone between the inferior alveolar margin point and the premental point). These points can also describe the tip of the nose, eyes, mouth, and other parts, but it is more difficult to obtain skull information by X-ray than only considering the face's surface. Although the selection of these landmarks is different, it is essentially a structured description of the face. Visual tasks consider more facial surface information because it is easier to obtain, while in medical tasks, facial skull information is considered to describe the facial structure more accurately. If there is a set of landmarks that are easy to obtain the face surface information and have the accuracy of the craniofacial structure, it is beneficial to both visual and medical tasks.

Obtaining landmarks from face data usually requires manual labeling by participants, which is extremely laborious and tedious work, and the repeated annotation work of the same personnel for a long time can easily bring errors by attention or muscle memory. Therefore, the automatic labeling algorithm of landmarks has also become a popular research field. Research on facial landmark localization has a long history, from traditional methods such as active shape models (ASM) [7], activate appearance models (AAM) [8], etc., to machine learning methods such as regression models. With the rapid development of deep learning, a variety of neural network structures are also used for the automatic localization of landmarks.

In this paper, we first select a part of face acupoints as facial landmarks, and manually annotate acupoints on 846 triangle meshes in the FaceScape dataset [9], forming a 3D facial acupoint labeling dataset. Based on this data, we propose an automatic landmark annotation algorithm based on a small amount of data, which achieves good results on the test data.

In summary, this article has the following contributions:

- Construction of an annotated dataset containing 846 labeled data of 37 acupoint landmarks on the FaceScape dataset.
- Development of an automatic labeling algorithm for 37 3D facial acupoint landmarks based on local geometric pattern image matching method, and the algorithm is decoupled from the number of landmarks, making it adaptable to any type and number of landmarks.

2. Related works

2.1. Landmark selection strategy

The selection strategy of facial landmarks revolves around the parts of the face that have the most prominent features and then selects the most representative point from these parts as a feature point.

The Farkas craniofacial anthropometry system [10] is a landmark scheme widely used in medicine, clinical genetics, plastic surgery, and oral maxillofacial surgery. Dr. Farkas created a measurement system based on anthropometric landmarks. These landmarks are located on the midline or are symmetrically distributed about the midline of the face, covering important parts such as the eyes, nose, mouth, and ears (as shown in **Figure 1** and **Table 1**). Based on Farkas's system, Qiao et al. [11] define 26 soft tissue landmarks and use these landmarks to extract facial phenotype measurements to analyze the genetic and geographical associations. Liang et al. [12] used the landmarks ex and en to extract the eye phenotype. In addition, Farkas landmarks are also widely used in the study of human face morphology [13,14], and there are also many variants and derivatives [15].

Table 1. Landmarks of the Farkas system.

Midline landmarks		**Symmetrical landmarks along the midline**			
Label	**Name**	**Label**	**Name**	**Label**	**Name**
g	Glabella	ft	Frontotemporale	sba	Subaurale
n	Nasion	sa	Superaurale	al	Alare
prn	Pronasale	zy	Zygion	go	Gonion
sn	Subnasale	ex	Exocanthion	ch	Cheilion
ls	Labiale superius	en	Endocanthion		
li	Labiale inferius	ps	Palpebrale superius		
pg	Pogonion	pi	Palpebrale inferius		
gn	Gnathion	mf	Maxillofrontale		

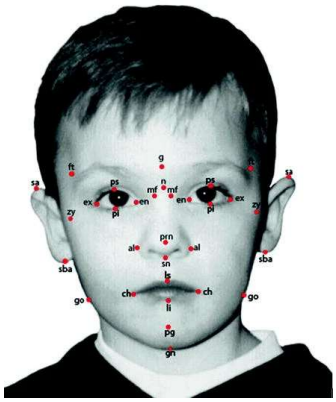


Figure 1. Landmarks defined by the Farkas system [16].

Compared with Farkas' craniofacial landmarks, which can be easily obtained from face images, cephalometric landmarks require X-rays to obtain skull images. The landmarks in cephalometric measurement can be divided into two categories:

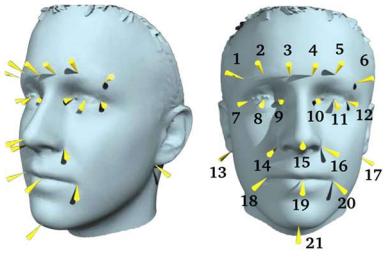
one is anatomical, which truly represents some anatomical structures of the skull. The other type is extended. This type of landmark is obtained by extending the anatomical landmark points on the cephalogram.

In computer vision, facial landmarks have a wide range of applications, such as high-precision face pose alignment [17,9], accurate analysis of facial features and expressions based on landmarks, processing, and synthesis of specific parts to achieve entertainment functions such as beautification and animation [18,19].

The number of landmarks used in computer vision has developed from the initial few to hundreds of landmarks. The selection of landmarks is mainly to locate the key parts of the face [6]. Therefore, compared with medical landmarks, there will be some points that are not on the surface of the face, such as the center of the eye, the center of the open mouth, etc., such as the landmarks of the eye center in the AFLW dataset [20] are not points on the surface of the face (as shown in **Figure 2**). In addition, the number of landmarks between different datasets is usually different, as shown in **Table 2**, so it is challenging to use across different datasets. These landmarks selection schemes are also widely used in 3D tasks.

Table 2. Some sample datasets with different landmarks.

Dataset	Faces	Landmarks	Dimension
ALFW [20]	25,993	21	2D
300-W [21]	3837	68	2D
AFW [22]	205	6	2D
Helen [23]	2330	194	2D
	10,993	68	2D
Menpo [24]	3852	39	2D
	11,971 + 280 k	84	3D
AFLW2000-3D [25]	2000	68	3D
FaceScape [9]	847	68	3D



(a) Side View (b) Frontal Face

Figure 2. Landmarks labeled in the AFLW dataset [20].

These landmarks' configuration is diverse and closely related to the facial surface. Still, it has a weak relationship with the craniofacial surface, so it is difficult for direct utilization among different medical applications.

2.2. Landmarks detection

Manual annotation of landmarks is tedious and time-consuming repetitive work. To solve this problem, a large number of automatic landmark detection/annotation

algorithms have been proposed. These landmark detection algorithms can be divided into three categories [26]: holistic methods, Constrained Local Model (CLM) methods, and regression-based methods.

Holistic methods use facial appearance information to construct a global facial shape model for landmark detection. The statistical model AAM (Active Appearance Model [8]) established by Principal Component Analysis (PCA) is used to locate the landmarks by fitting the learned appearance and shape models in the test images. Most holistic methods focus on improving the fitting algorithm [27,28]. In addition, AAM can also be optimized by adjusting the representation of features [29,30]. Constrained Local Methods use global facial shape features and unique local features around each landmark to infer the location of the landmark. The CLM [31] finds the landmarks by minimizing misalignment errors. Regression-based methods learn the mapping of landmarks directly from images, usually without explicitly constructing a feature model [32].

Due to the convenience of image data acquisition and use, automatic landmarks detection algorithm is more widely studied in 2D. Zhang et al. [33] use heatmap-offset regression to perform landmark detection on face images under unconstrained conditions. Jeong et al. [34] propose a facial landmark detection algorithm for real-world driving situations by integrating a locally weighted random forest regressor with random sampling consistency and an explicit global shape model. With the rapid development of deep learning, a large number of networks have also been proposed for 2D facial landmark detection [18,35,36]. Compared with 2D, the research on 3D facial landmark detection is relatively new. Salazar et al. [37] predicts landmarks through a network that learns the surface statistical information around each landmark and the connection structure of landmarks on 3D face scan data. Papazov et al. [38] predicts landmarks on RGB-D images through 3D local shape descriptors. Berends et al. [39] use two consecutive DiffusionNet models to achieve automatic labeling of cephalometric landmarks. Perakis et al. [40] uses 3D local shape descriptors to extract candidate landmarks, and identify landmarks by matching the candidate landmarks with the FLM (Facial Landmark Model) of facial anatomical landmarks.

3. Landmarks and method

3.1. Acupoint landmarks

Acupoints are important locations in traditional Chinese medicine. It represents the position where the energy from the internal organs and meridians of the human body is infused on the surface. Acupoints are not only the reaction points of diseases but also the stimulation points of acupuncture, massage, and other clinical medicine.

Acupoints are usually located in the depressions of the skull or the pores where muscles and bones meet and are widely distributed on the surface of the body. Hence, we believe that using facial acupoints as facial landmarks can not only describe the shape of the human face but also have the accuracy of the skull structure of the human face. So that visual tasks and medical tasks can have a unified landmarks framework.

Acupoints are usually located in the depressions of the skull or the pores where

muscles and bones meet and are widely distributed on the surface of the body. Hence, we believe that using facial acupoints as facial landmarks can not only describe the shape of the human face but also have the accuracy of the skull structure of the human face. So that visual tasks and medical tasks can have a unified landmarks framework.

We selected 37 facial acupoints as landmarks and labeled them on the 846 neutral face triangle meshes of the FaceScape dataset [9] (see **Table 3** basic information about acupoint landmarks, and **Figure 3**). The selected facial acupoints are evenly distributed on the face and are distributed in both flat and complex parts. These landmarks are symmetrically distributed with the midline of the face as the axis (5 landmarks are distributed along the midline).

Table 3. Basic information about acupoint landmarks.

Index	Label	Name
0 (M)	GV24^{+}	Yintang
1 (M)	GV25	Suliao
2 (M)	GV26	Shuigou
3 (M)	GV27	Duiduan
4 (M)	CV24	Chengjiang
5/6 (L/R)	GB14	Yangbai
7/8 (L/R)	TE23	Sizhukong
9/10 (L/R)	BL2	Zanzhu
11/12 (L/R)	EX-HN5	Temple
13/14 (L/R)	GB1	Tongziliao
15/16 (L/R)	BL1	Jingming
17/18 (L/R)	ST1	Chengqi
19/20 (L/R)	ST2	Sibai
21/22 (L/R)	ST7	Xiaguan
23/24 (L/R)	LI20	Yingxiang
25/26 (L/R)	SI18	Quanliao
27/28 (L/R)	ST3	Juliao
29/30 (L/R)	LI19	Kouheliao
31/32 (L/R)	ST6	Jiache
33/34 (L/R)	ST4	Dicang
35/36 (L/R)	ST5	Daying

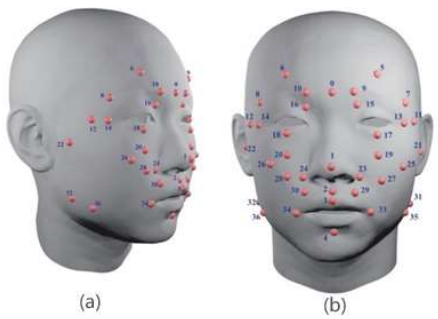


Figure 3. The 37 acupoint landmarks. **(a)** Left front view; **(b)** Front view.

3.2. Sample points processing

We use $M = (V, E, F)$ to describe a triangle mesh with m vertices $\{v_0, v_1, \cdots, v_{m-1}\}$. To describe the local feature of a given landmark v_l, we improve the framework proposed by Wang et al. [41] by finding all the neighbors of the landmarks within the radius r as sample points S. The whole method is shown in **Figure 4**. If the number of vertices within r is less than a threshold, then the 2-ring vertices of the landmarks are picked as sample points.

$$S = \left\{v \in V: \left|\left|v - v_l\right|\right|_2 < r\right\} \text{or} \{v \in 2 - \text{ring}(v_l)\} \tag{1}$$

After obtaining the sample points S, we retrieve the unit normal n_l of the landmark v_l, and project these points onto the tangent plane of the landmarks. Considering a point $p \in S$, the signed distance from p to the tangent plane of v_l is calculated according to the equation:

$$dis = (p - v_l)n_i \tag{2}$$

and then the projection p' on the tangent plane can be obtained by:

$$p' = p - dis \cdot n_i \tag{3}$$

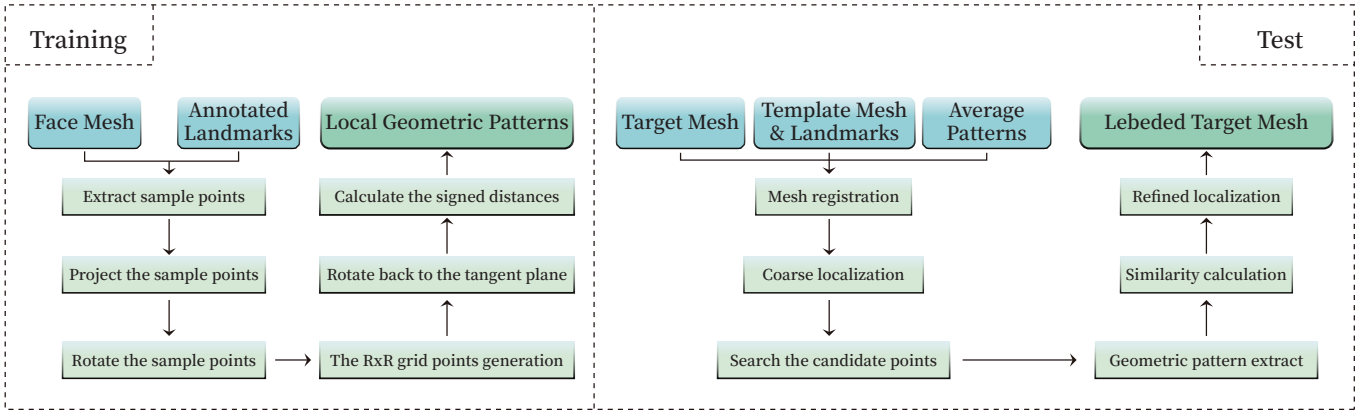


Figure 4. Flowchart of training and landmarks prediction.

All points in S are projected onto the tangent plane of the landmarks through the above processing to obtain S_T.

To use these sampling points to produce a regular image grid, we rotate S_T so that the tangent plane of v_l after rotation is perpendicular to the z-axis. We first calculate the rotation matrix R from the landmark's unit normal n_l to the z-axis denoted as $u_z = (0,0,1)$. Considering arbitrary unit normal vector $u = (a, b, c)$ as shown in **Figure 5** starting from the origin, we first construct the rotation matrix R_x of u bout the x-axis. Let $u' = (0, b, c)$ be the projection of u on the yoz plane, then:

$$R_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{c}{d} & -\frac{b}{d} \\ 0 & \frac{b}{d} & \frac{c}{d} \end{bmatrix} \tag{4}$$

where α denotes the rotation angle corresponding to u rotating around x-axis to the xoz plane, and $d = |u'| = \sqrt{b^2 + c^2}$. Next, let $u'' = (a, 0, d)$ be the vector after u rotates around x-axis, then the rotation matrix R_y of u'' around y-axis is:

$$R_y(\beta) = \begin{bmatrix} d & 0 & -a \\ 0 & 1 & 0 \\ a & 0 & d \end{bmatrix} \tag{5}$$

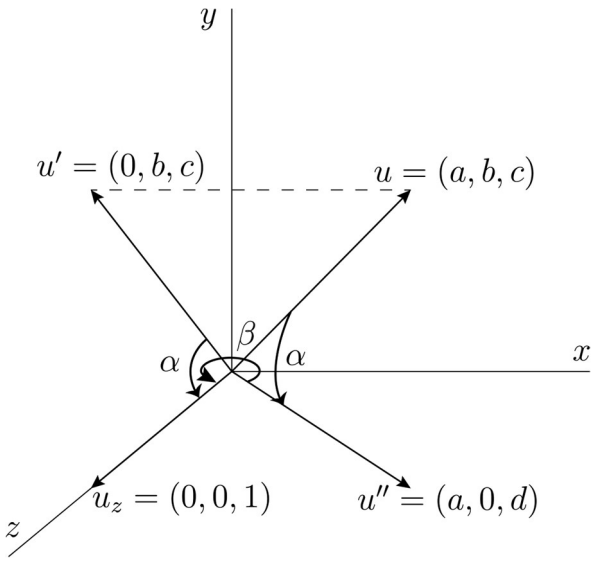


Figure 5. Rotation of a unit vector aligned to the *z*-axis.

With R_x and R_y, we can rotate any unit normal vector onto the u_z. Apply $R = R_y(\beta)R_x(\alpha)$ transformation on S_T, we can get final sample points S_P located on the plane perpendicular to the *z*-axis as shown in **Figure 6c**.

3.3. Local feature descriptor

After obtaining the sampling points S_P, we can generate a set of points in a squared grid with resolution R and the maximum area centered at the landmark based on these points. For a given landmark $v_l = (x, y, z)$ and the corresponding S_{P_l}, we first calculate the half-side length a of the grid:

$$a = \min\big(\min(x_{\max} - x, x - x_{\min}), \min(y_{\max} - y, y - y_{\min})\big) \tag{6}$$

where $x_{\min}, x_{\max}, y_{\min}, y_{\max}$ represent the boundaries of the axis-aligned bounding box of S_{P_l}. To generate a grid with resolution R, the unit length h of the grid is:

$$h = \frac{2a}{(R-1)} \tag{7}$$

Then the point set G_{p_l} on the square grid of given landmarks v_l is:

$$G_{p_l} = \{(x_{\min} + i \cdot h, y_{\min} + j \cdot h | i, j = 0,1, \cdots, R-1\} \tag{8}$$

Next, we rotate G_{p_l} back to the tangent plane of v_l through R^{-1} and get G_{T_l} as shown in **Figure 6f**. The whole process of sample point processing and square grid generation is show in the **Figure 6.** We cast two rays along the normal in both directions of v_l originating at the points in G_{T_l} and find the signed distance from the grid point to the surface of the mesh as the local feature. This feature represents the signed distances from points near the landmark to its tangent plane, forming a pattern image.

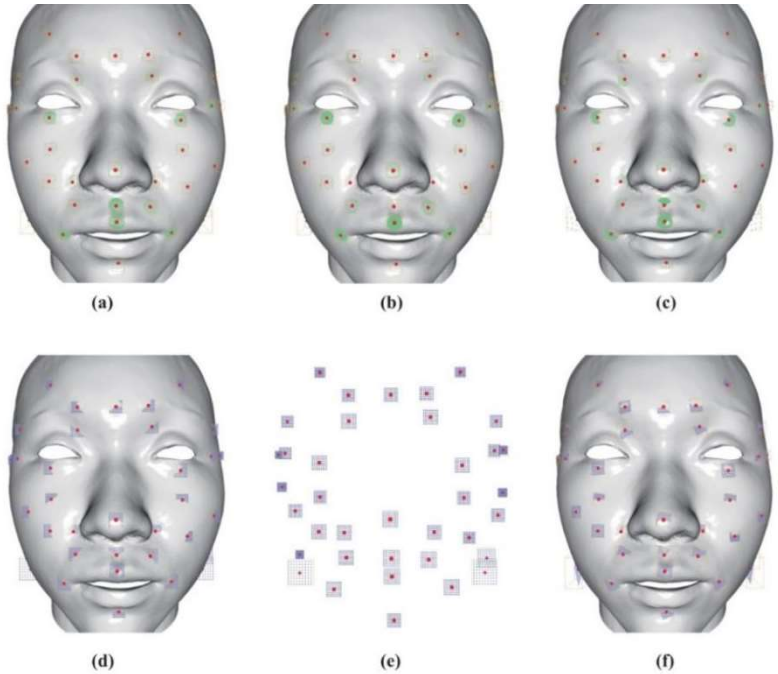


Figure 6. The processing of sample points ultimately generates a square grid used to calculate the local features of landmarks; **(a)** sampling points S are composed of vertices within the radius r of each landmark or the 2-ring neighbors; **(b)** sampling points S projected onto the tangent plane of each landmark, denote as S_T; **(c)** sampling points S_T are rotated to the plane perpendicular to the z-axis with the landmark as the origin, denote as S_P; **(d)** generate $R \times R$ grid points ($R = 9$) in the bounding box of S_P; **(e)** grid points without occlusion; **(f)** the grid points that rotate back to the landmark's tangent plane are used to calculate the final features.
The red dots represent landmarks, the green dots represent the vertices near the landmarks and processed, the blue dots represent the points in the generated square grid, each point is ultimately represented as an image pixel, and the orange box represents the Axis-Aligned Bounding Box.

3.4. Coarse-to-fine localization

We define the landmark localization problem as searching for vertices with the most similar local feature descriptors mentioned above. First, we extract the local feature pattern image of each landmark for all meshes on the training set, and calculate the average feature patterns as a guide for prediction refinement.

A template mesh M_a is selected from the training set. For each target mesh M_t in the test set, we first use the ICP algorithm to align the M_t mesh with M_a. For each landmark of M_a, Knn is used to search the nearest vertex in M_t as the coarse location of landmarks in M_t. For each coarsely located landmark, we search the vertices within radius r of the landmark or the 2-ring neighbors, the same as the generation of the sampling points S_T.

For each candidate point v_{l_p} of landmark v_l, we calculate the local features of v_{l_p} and generate a pattern image. We then calculate the MS-SSIM [42] similarity between the pattern and the average feature pattern and use the candidate point with the maximum similarity as the final optimized landmark.

4. Experiments

4.1. Data and parameters

Unlike deep learning methodologies that heavily rely on randomized partitioning techniques to ensure model generalization, our method operates on a different paradigm. Our method, rooted in local geometric pattern image matching, does not necessitate an identical training process akin to deep learning models. Instead, it focuses on capturing sufficient information from the training dataset to effectively represent local geometric features. Consequently, random partitioning techniques become unnecessary, as the method's efficacy stems from its ability to discern and leverage local geometric patterns without extensive training data. Therefore, we used a simple division of training set and test set in the experiments.

In our experiment, we used the neutral face mesh No.1–40 in the FaceScape dataset as the training set to extract the local feature of each acupoint landmark. In the process of extracting features, we construct the sample points S_T with the search radius $r = 4$, the minimum neighbor threshold is 5, and the feature pattern resolution $R = 9$. The odd resolution can ensure that the center of each pattern represents the distance of 0. The average patterns of the final extracted feature points are shown in **Figure 7**.

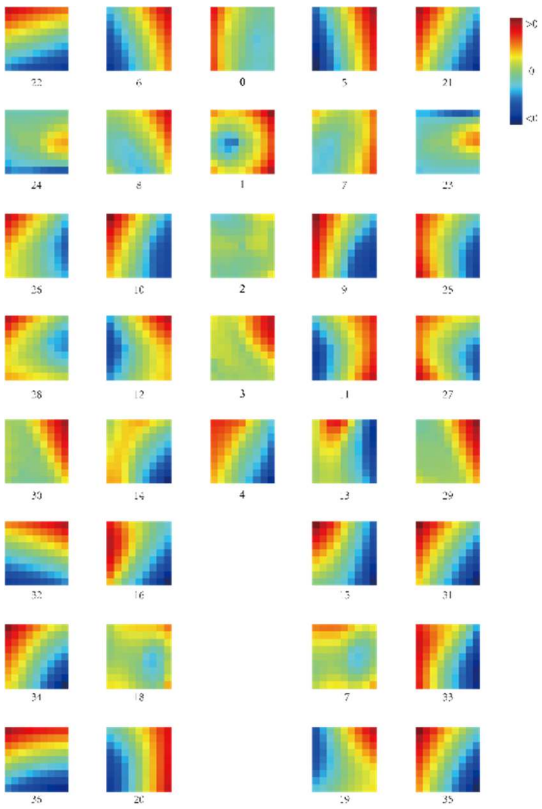


Figure 7. The 9 × 9 resolution average feature patterns of 37 acupoint landmarks in the training set.

4.2. Implement detail

We use the Python version of Open3D [43] to build a landmark labeling

program, which is used to label acupoint landmarks on a 3D face mesh. To obtain higher efficiency, our automatic landmark labeling algorithm is implemented in C++.

To facilitate the representation and visualization of feature patterns, during the feature extraction process, we perform special normalization processing on the distance represented by the features. We map the positive and negative distance values to $[0, 0.5]$ and $[-0.5, 0]$ respectively, thus ensuring that the distance value represented by the middle of the pattern is 0 (odd resolution).

Each pixel represents the signed distance from a certain position near the landmark to the tangent plane, red represents the positive distance, and the corresponding grid point on the tangent plane is located outside the face, while the blue distance is negative and the corresponding grid point is located inside the face. The middle column represents the patterns of the acupoints located on the midline of the face, and the left and right sides represent the patterns of the right and left faces, respectively.

4.3. Automatic labeling results

We tested on two subsets of the FaceScape dataset, with a small subset containing 20 models (No.41–60), and a large subset containing 806 models except for the training data and No.832 (the mesh was missing in the dataset version used), and the MSE of each acupoint is shown in **Table 4**.

Table 4. Acupoint landmark prediction results.

Index	Label	MSE_1 (mm)	MSE_2 (mm)	Index	Label	MSE_1 (mm)	MSE_2 (mm)	Index	Label	MSE_1 (mm)	MSE_2 (mm)
0	$GV24^+$	3.33761	4.16259	13	GB1L	2.86373	6.36939	25	SI18L	14.9271	9.27154
1	GV25	2.71117	2.86771	14	GB1R	5.5479	9.55427	26	SI18R	12.8831	13.383
2	GV26	4.39703	3.28801	15	BL1L	4.20445	9.51578	27	ST3L	6.23471	4.71092
3	GV27	3.33836	2.28356	16	BL1R	3.59924	9.82582	28	ST3R	4.82106	5.37582
4	CV24	3.89653	6.46715	17	ST1L	6.76117	3.93621	29	LI19L	5.19442	6.2019
5	GB14L	8.73421	6.70675	18	ST1R	5.1002	3.57489	30	LI19R	5.46766	9.24654
6	GB14R	8.26157	8.27526	19	ST2L	9.91773	4.49015	31	ST6L	7.95975	11.9356
7	TE23L	4.73751	5.11096	20	ST2R	7.6565	4.50048	32	ST6R	9.92264	9.77193
8	TE23R	5.12839	5.85634	21	ST7L	10.7996	15.0106	33	ST4L	3.2399	4.94159
9	BL2L	3.48487	4.50023	22	ST7R	11.5619	17.3381	34	ST4R	2.95033	3.86951
10	BL2R	3.70636	7.64659	23	LI20L	2.06342	3.13582	35	ST5L	9.37778	18.148
11	EX-HN5L	11.1605	5.02405	24	LI20R	2.47917	2.90459	36	ST5R	8.47732	18.4812
12	EX-HN5R	10.8742	5.81509								

MSE represents the mean square error between the ground-truth (manually labeled points) and the automatically labeled points. MSE1 and MSE2 are the results of the test set containing 20 and 806 models, respectively.

The experimental results show that the error of acupoint landmarks distributed on and near the middle line of the human face is smaller than acupoints far away from the middle line. This is because during the face alignment process, the registration effect is better in the middle part of the face, and the width of different faces causes the closer to the outside of the face, the lower the degree of registration, which causes the coarse localization of the landmark far away from the real point,

resulting in a larger MSE of the optimized landmarks. This can be improved by coarse positioning of the facial landmarks through non-rigid registration.

4.4. Algorithm comparison

Due to the particularity of the landmarks we use, it is difficult to compare with the accuracy of other methods, so we compare with other methods from an algorithmic point of view as shown in **Table 5**.

Table 5. Algorithm comparison.

Methods	Template	Landmarks	Candidate landmarks	Final prediction	Extensibility
Perakis P et al. [40]	No	8/5	Using shape index and spin image to determine candidates.	Fitting into FLM8/5 model.	Poor
Ingale AK et al. [44]	No	16	Segment mesh, find sharp edge vertices as candidates.	The point within the specific mesh cluster that satisfies certain constraints.	Poor
Ours	Yes	37	Rigid face registration, the nearest neighbor's neighborhood of template facial landmarks as candidates.	The candidates with the highest matching degree to the template landmark patterns.	Good

Perakis constructed FLM models for the sets of 8 and 5 landmarks respectively, selecting candidate landmarks through shape index and spin image, and searching the combination of these candidate points to fit into corresponding FLM. Because a different number of landmarks needs to build different FLM models, and with the increase of the number of landmarks, the search space size of candidate combinations will increase exponentially, so the extensibility of landmarks is poor.

Ingale et al. [44] first segment the face by the clustering algorithm, and find the sharp edge vertices as the candidate vertices in the specific cluster. Finally, the candidate points in the specific region are searched for points that satisfy the unique constraints of each landmark as the final result. Since each landmark has different constraints, this method also difficult to expand the landmarks because new constraints need to be defined.

In our method, we only need to select a template face mesh, and by aligning the target face and the template face, we search through the neighborhood of the nearest points on the target face of the template landmarks as candidate points to match the feature point template pattern. The candidate point with the highest matching degree is used as the final result. In our case, the extension of landmarks is very easy, only need to calculate an additional template pattern of the target landmark, because the positioning of each landmark is independent, resulting in a very good expansion of the landmarks.

5. Conclusions

In this paper, we selected 37 human facial acupoints as the facial landmarks and constructed an annotated dataset containing 846 labeled data on the FaceScape

dataset. Based on this annotation data, we propose an automatic labeling algorithm of 37 3D facial acupoint landmarks based on local geometric pattern image matching of landmarks and achieved good annotation results in the experiments. As we discussed in the algorithmic comparison section of the experiment, our automatic labeling algorithm is decoupled from the number of landmarks, so this method is a general method that can be migrated to any type and number of landmarks and has strong extensibility. In addition, the problem of large errors in landmarks far away from the middle of the face can be solved by optimizing the coarse positioning results through non-rigid registration methods, thereby improving the annotation results.

Since our prediction results are the vertices in the original mesh, thus the density of vertices in the face mesh poses a significant challenge in our study. We acknowledge that an excessively dense mesh can result in a considerable computational burden, while a sparse mesh may compromise the accuracy of our method, particularly in regions with fewer vertices, such as the flat areas near the cheeks. To address this limitation and enhance the effectiveness of our approach, we plane to focus on the following areas in our future work: 1) Exploring novel methods for optimizing mesh density and distribution to achieve a more balanced trade-off between computational efficiency and accuracy. 2) Introducing traditional Chinese medicine acupoint selection methods to assist in the optimization process and results evaluation.

Author contributions: Conceptualization, MF and WM; methodology, MF and JY; software, JY; validation, JY, MF and WM; formal analysis, MF; investigation, JY; resources, MF; data curation, JY; writing—original draft preparation, JY; writing—review and editing, MF; visualization, JY; supervision, MF; project administration, MF; funding acquisition, MF and WM. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant (Nos 62072126), in part by the Fundamental Research Projects Jointly Funded by Guangzhou Council and Municipal Universities under Grant SL2023A03J00639, and Key Laboratory of Philosophy and Social Sciences in Guangdong Province of Maritime Silk Road of Guangzhou University (GD22TWCXGC15), and Research Grants Council, Hong Kong (SAR), China (CityU 11201919).

Conflict of interest: The authors declare no conflict of interest.

References

1. Vrochidou E, Papić V, Kalampokas T, Papakostas GA. Automatic Facial Palsy Detection—From Mathematical Modeling to Deep Learning. Axioms. 2023; 12(12): 1091. doi: 10.3390/axioms12121091
2. Kim S, An H, Song M, Lee S, et al. Automated Marker-Less Patient-to-Preoperative Medical Image Registration Approach Using RGB-D Images and Facial Landmarks for Potential Use in Computed-Aided Surgical Navigation of the Paranasal Sinus. In: Advances in Computer Graphics. Cham: Springer Nature Switzerland; 2024. pp. 135–145.
3. Mercan E, Atmosukarto I, Wu J, et al. Craniofacial Image Analysis. In: Health Monitoring and Personalized Feedback using Multimedia Data. Cham: Springer International Publishing; 2015. pp. 9–29.

4. Liu F, Zhao Q, Liu X, Zeng D. Joint Face Alignment and 3D Face Reconstruction with Application to Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2020; 42(3): 664-678. doi: 10.1109/tpami.2018.2885995
5. Sharma U, Faisal KN, Sharma RR, Arya KV. Facial Landmark-Based Human Emotion Recognition Technique for Oriented Viewpoints in the Presence of Facial Attributes. SN Computer Science. 2023; 4(3). doi: 10.1007/s42979-023-01727-y
6. Gupta A, Thakkar K, Gandhi V, Narayanan PJ. Nose, eyes and ears: Head pose estimation by locating facial keypoints. arXiv. 2018.
7. Cootes TF, Taylor CJ, Cooper DH, Graham J. Active Shape Models-Their Training and Application. Computer Vision and Image Understanding. 1995; 61(1): 38-59. doi: 10.1006/cviu.1995.1004
8. Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001; 23(6): 681-685. doi: 10.1109/34.927467
9. Yang H, Zhu H, Wang Y, et al. FaceScape: A Large-Scale High Quality 3D Face Dataset and Detailed Riggable 3D Face Prediction. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Published online June 2020. doi: 10.1109/cvpr42600.2020.00068
10. Farkas LG. Anthropometry of the head and face, 2nd ed. Available online: https://api.semanticscholar.org/CorpusID:57441646 (accessed on 23 March 2023).
11. Qiao H, Tan J, Wen S, et al. De Novo Dissecting the Three-Dimensional Facial Morphology of 2379 Han Chinese Individuals. Phenomics. Published online June 8, 2023. doi: 10.1007/s43657-023-00109-x
12. Liang Y, Liu H, Gao Z, et al. Ocular phenotype related SNP analysis in Southern Han Chinese population from Guangdong province. Gene. 2022; 826: 146458. doi: 10.1016/j.gene.2022.146458
13. Ercan I, Ozdemir ST, Etoz A, et al. Facial asymmetry in young healthy subjects evaluated by statistical shape analysis. Journal of Anatomy. 2008; 213(6): 663-669. doi: 10.1111/j.1469-7580.2008.01002.x
14. Sforza C, Rosati R, De Menezes M, et al. Three-Dimensional Computerized Anthropometry of the Nose. In: Handbook of Anthropometry. New York, NY: Springer New York; 2012. pp. 927–942.
15. Agbolade O, Nazri A, Yaakob R, et al. Morphometric approach to 3D soft-tissue craniofacial analysis and classification of ethnicity, sex, and age. Wilkinson C, ed. PLOS ONE. 2020; 15(4): e0228402. doi: 10.1371/journal.pone.0228402
16. Deutsch CK, Shell AR, Francis RW, Bird BD. The Farkas System of Craniofacial Anthropometry: Methodology and Normative Databases. In: Handbook of Anthropometry. New York, NY: Springer New York; 2012. pp. 561–573.
17. Zhuang W, Chen L, Hong C, et al. FT-GAN: Face Transformation with Key Points Alignment for Pose-Invariant Face Recognition. Electronics. 2019; 8(7): 807. doi: 10.3390/electronics8070807
18. Wang X, Wang Y, Li W, et al. Facial Expression Animation by Landmark Guided Residual Module. IEEE Transactions on Affective Computing. 2023; 14(2): 878-894. doi: 10.1109/taffc.2021.3100352
19. Otberdout N, Daoudi M, Kacem A, et al. Dynamic Facial Expression Generation on Hilbert Hypersphere With Conditional Wasserstein Generative Adversarial Nets. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022; 44(2): 848-863. doi: 10.1109/tpami.2020.3002500
20. Kostinger M, Wohlhart P, Roth PM, Bischof H. Annotated Facial Landmarks in the Wild: A large-scale, real-world database for facial landmark localization. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, Nov. 2011, pp. 2144–2151. doi: 10.1109/ICCVW.2011.6130513
21. Sagonas C, Antonakos E, Tzimiropoulos G, et al. 300 Faces In-The-Wild Challenge: database and results. Image and Vision Computing. 2016; 47: 3-18. doi: 10.1016/j.imavis.2016.01.002
22. Zhu X,Ramanan D. Face detection, pose estimation, and landmark localization in the wild. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, Jun. 2012, pp. 2879–2886. doi: 10.1109/CVPR.2012.6248014
23. Le V, Brandt J, Lin Z, et al. Interactive Facial Feature Localization. In Computer Vision—ECCV 2012. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. pp. 679–692. doi: 10.1007/978-3-642-33712-3_49
24. Deng J, Roussos A, Chrysos G, et al. The Menpo Benchmark for Multi-pose 2D and 3D Facial Landmark Localisation and Tracking. International Journal of Computer Vision. 2018; 127(6-7): 599-624. doi: 10.1007/s11263-018-1134-y
25. Zhu X, Lei Z, Liu X, et al. Face Alignment Across Large Poses: A 3D Solution. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Published online June 2016. doi: 10.1109/cvpr.2016.23
26. Wu Y, Ji Q. Facial Landmark Detection: A Literature Survey. International Journal of Computer Vision. 2018; 127(2): 115-142. doi: 10.1007/s11263-018-1097-z
27. Baker S, Matthews I. Lucas-Kanade 20 Years On: A Unifying Framework. Int. J. Comput. Vis. 2004; 56(3): 221–255. doi:

10.1023/B:VISI.0000011205.11775.fd
28. Matthews I, Baker S. Active Appearance Models Revisited. Int. J. Comput. Vis. 2004; 60(2): 135–164. doi: 10.1023/B:VISI.0000029666.37597.d3
29. Gross R, Matthews I, Baker S. Appearance-based face recognition and light-fields. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004; 26(4): 449-465. doi: 10.1109/tpami.2004.1265861
30. Gross R, Matthews I, Baker S. Generic vs. person specific active appearance models. Image and Vision Computing. 2005; 23(12): 1080-1093. doi: 10.1016/j.imavis.2005.07.009
31. Cristinacce D, Cootes TF. Feature Detection and Tracking with Constrained Local Models. In: Proceedings of the British Machine Vision Conference 2006, Edinburgh, 2006, pp. 95.1-95.10. doi: 10.5244/C.20.95
32. Dantone M, Gall J, Fanelli G, Van Gool L. Real-time facial feature detection using conditional regression forests. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, Jun. 2012, pp. 2578–2585. doi: 10.1109/CVPR.2012.6247976
33. Zhang J, Hu H, Feng S. Robust Facial Landmark Detection via Heatmap-Offset Regression. IEEE Transactions on Image Processing. 2020; 29: 5050-5064. doi: 10.1109/tip.2020.2976765
34. Jeong M, Ko BC, Kwak S, Nam JY. Driver Facial Landmark Detection in Real Driving Situations. IEEE Transactions on Circuits and Systems for Video Technology. 2018; 28(10): 2753-2767. doi: 10.1109/tcsvt.2017.2769096
35. Deng Z, Liu H. Geometry-attentive relational reasoning for robust facial landmark detection. Neurocomputing. 2021; 453: 790-800. doi: 10.1016/j.neucom.2020.06.126
36. Wu Y, Ji Q. Discriminative Deep Face Shape Model for Facial Point Detection. International Journal of Computer Vision. 2014; 113(1): 37-53. doi: 10.1007/s11263-014-0775-8
37. Salazar A, Wuhrer S, Shu C, Prieto F. Fully automatic expression-invariant face correspondence. Machine Vision and Applications. 2013; 25(4): 859-879. doi: 10.1007/s00138-013-0579-9
38. Papazov C, Marks TK, Jones M. Real-time 3D head pose and facial landmark estimation from depth images using triangular surface patch features. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, Jun. 2015, pp. 4722–4730. doi: 10.1109/CVPR.2015.7299104
39. Berends B, Bielevelt F, Schreurs R, et al. Fully automated landmarking and facial segmentation on 3D photographs. arXiv. 2023.
40. Perakis P, Passalis G, Theoharis T, Kakadiaris IA. 3D Facial Landmark Detection under Large Yaw and Expression Variations. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013; 35(7): 1552-1564. doi: 10.1109/tpami.2012.247
41. Wang J, Fang S, Fang M, et al. Automatic Landmark Placement for Large 3D Facial Image Dataset. In: 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, Dec. 2019, pp. 5088–5093. doi: 10.1109/BigData47090.2019.9006310
42. Wang Z, Simoncelli EP, Bovik AC. Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, Pacific Grove, CA, USA, 2003, pp. 1398–1402. doi: 10.1109/ACSSC.2003.1292216
43. Zhou QY, Park J, Koltun V. Open3D: A Modern Library for 3D Data Processing. arXiv. 2018.
44. Ingale AK, Leema AA, Kim H, Udayan JD. Automatic 3D Facial Landmark-Based Deformation Transfer on Facial Variants for Blendshape Generation. Arabian Journal for Science and Engineering. 2022; 48(8): 10109-10123. doi: 10.1007/s13369-022-07403-2