Automatic 3D Cranial Landmark Positioning based on Surface Curvature Feature using Machine Learning

Putu Hendra Suputra [a, b, 1], Anggraini Dwi Sensusiati [c, 2], Myrtati Dyah Artaria [d, 3],
Gijsbertus Jacob Verkerke [e, 4], Eko Mulyanto Yuniarno [a, f, 5, *], I Ketut Eddy Purnama [a, f, 6]

[a] *Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember*
Gedung B, C & AJ Kampus ITS, Keputih, Sukolilo, Surabaya, 60111, Indonesia

[b] *Department of Informatics, Universitas Pendidikan Ganesha*
Jalan Udayana No.11 Singaraja - Bali 81116, Indonesia

[c] *Department of Radiology, Faculty of Medicine, Universitas Airlangga*
Kampus A Universitas Airlangga, Jl. Mayjen. Prof. Dr. Moestopo 47, Surabaya 60131, Indonesia

[d] *Department of Anthropology, Universitas Airlangga*
Room 211, Building A, FISIP, Campus B, Jl. Airlangga No.4 - 6, Gubeng, Surabaya 60115, Indonesia

[e] *University Medical Center Groningen, University of Groningen*
Hanzeplein 1, 9713 GZ Groningen, Netherlands

[f] *Department of Computer Engineering, Institut Teknologi Sepuluh Nopember*
Gedung B & C, Kampus ITS Sukolilo, Surabaya 60111, Indonesia

[1] *hendra.suputra@undiksha.ac.id;* [2] *anggraini-d-s@fk.unair.ac.id;* [3] *myrtati.artaria@fisip.unair.ac.id;*
[4] *g.j.verkerke@med.umcg.nl;* [5] *ekomulyanto@ee.its.ac.id*;* [6] *ketut@te.its.ac.id*
* corresponding author

ARTICLE INFO

ABSTRACT

Cranial anthropometric reference points (landmarks) play an important role in craniofacial reconstruction and identification. Knowledge to detect the position of landmarks is critical. This work aims to locate landmarks automatically. Landmarks positioning using Surface Curvature Feature (SCF) is inspired by conventional methods of finding landmarks based on morphometrical features. Each cranial landmark has a unique shape. With the appropriate 3D descriptors, the computer can draw associations between shapes and landmarks using machine learning. The challenge in classification and detection in three-dimensional space is to determine the model and data representation. Using three-dimensional raw data in machine learning is a serious volumetric issue. This work uses the Surface Curvature Feature as a three-dimensional descriptor. It extracts the local surface curvature shape into a projection sequential value (depth). A machine learning method is developed to determine the position of landmarks based on local surface shape characteristics. Classification is carried out from the top-n prediction probabilities for each landmark class, from a set of predictions, then filtered to get pinpoint accuracy. The landmark prediction points are hypothetically clustered in a particular area, so a cluster-based filter is appropriate to isolate them. The learning model successfully detected the landmarks, with the average distance between the prediction points and the ground truth being 0.0326 normalized units. The cluster-based filter is implemented to increase accuracy compared to the ground truth. Thus, SCF is suitable as a 3D descriptor of cranial landmarks.

I. Introduction

Craniofacial reconstruction is the art of reconstructing soft tissue that is lost from the skull to present a visual appearance of a face that is lifelike. Soft tissues are estimated based on the anthropometric features of the skull. Forensic identification also requires landmark positioning for correspondence with reference facial landmarks [1][2]. An expert must determine the position of particular anatomical features (landmarks) as the soft tissue reference.

Cranial landmarks are reference points in adding soft tissue thickness. A three-dimensional perspective poses a challenge in placing landmarks compared to positioning on two-dimensional images (images superimposition). The popular facial reconstruction methods, the American [3]. and Gerasimov [4], are the most intuitive method for a computer-based approach. Both simulate soft tissues guided by the placement of landmarks.

The current developments of computerized craniofacial reconstructions generally use a three-dimensional virtual model that mimics the idea of the manual reconstruction process [5][6]. Research related to computer-based craniofacial reconstruction is widely developed. However, most attention has been paid to reconstructing facial features based on soft tissue thickness. Determination of the position of the skull landmark is still determined manually.

Knowledge and ability to detect the position of anatomical features are critical [7][8][9][10][11][12][13][14]. The golden rules in determining the position of anatomical features are based on its surface shape. The use of Surface Curvature Feature (SCF) for classifying skull surface landmarks is inspired by conventional methods of determining anthropometric reference points. The SCF is introduced as a morphometric feature model representing a local surface's shape [15]. Nonetheless, several studies of 3D automatic detection have been carried out. Anatomical landmark positioning by Jacinto [16] proposed a multi-atlas approach to orthopedic knee landmarks. The surface local shape is used as a reference for the registration of pre-defined landmark positioning. Another approach to the three-dimensional annotation system is also carried out by Lee [17] and Lindner [18]. However, they use a 2D image or a shadowed-2D image. The methods used did not measure directly to the skull model geometrically like the original conventional method. The challenge of using a three-dimensional image modality is the input dimension which involves a very large volume of voxels. Instead of employing a full 3D model, this study interprets the three-dimensional shape information in context by a 3D descriptor. In other words, context-based information retrieval in three-dimensional form.

Each landmark has a distinctive shape. By understanding a landmark's shape characteristics, the landmark's position on the skull is known. A regular surface's shape can be defined using a quantifiable model or a feature representation (3D descriptor). This modelling will ease the computation of surface shape features. The curvature model must be able to describe a surface's shape quantitatively. A surface is concave or convex by looking at the curvature value. In this work, curvature features are used as the input to obtain characteristic descriptors of the physical shape of skull surfaces as landmarks.

By solving the volumetric issue of 3D data in machine learning, the learning work can be solved with a simple multi-layer perceptron to draw relationships between surface shapes as landmarks. The network model uses an activation function that accommodates the range of SCF sequential values to maintain the three-dimensional information context. The model can classify a point on the surface of the skull, whether it is a landmark or not. A novel method was proposed in this work, oriented towards the geometrical measurement in 3D spatial space. A three-dimensional feature representation, Surface Curvature Feature (SCF), is used to describe the local morphological shape of the landmarks. Machine learning is based on detecting and classifying the skull surface landmarks. The contribution of the proposed method is as follows:

- The first study of the use of SCF as a three-dimensional feature representation (descriptor), in detecting anatomical features of the skull surface (landmarks)
- An intuitive method for recognizing landmarks based on the shape of the surface is inspired by conventional methods.
- A simple MLP is engaged in drawing correlations on the implicit characteristics of a surface shape as a skull landmark.

II. Method

A. Related Works

Working with 3D objects for shape recognition cannot be separated from the 3D shape representation process. Generally, 3D representation is done using a feature-based approach, a facade-based approach (view), and a graphical information-based approach [19]. The feature-based approach uses shape models in transactions or classifications. The view-based approach renders a

3D view into 2D at a certain point of view. The final approach is to use the information on the surface topology.

The feature-based approach relies on the geometric characteristics of 3D objects. Geometry characters can also be applied to global and local features, where global features are most widely used for classification work. Meanwhile, the view-based approach eliminates the volumetric load of 3D objects. The computational cost of entering 3D data is much higher than that of a two- or single-dimensional format. Therefore, several ways have been developed to convert the structure of 3D shapes into 2D. Conversion of 3D data into 2D must have information that is sacrificed as long as it maintains the desired features. The third approach emphasizes the relationships and connections of the various components of the object representation model. Computationally, the graph-based approach is considered inefficient compared to the other two approaches. If the first approach geometrically retrieves the shape of an object, the second approach must look at the object from various views to obtain the intended feature.

The work related to the local shape approach of a 3-dimensional surface is the automatic positioning proposed by Jacinto [16]. Local rigid registration between the pre-defined landmark 3D model and the patient structure model is detected of anatomical landmarks. The three-dimensional annotation system by Lee [17] and Lindner [18] uses a 2D image approach. Cephalometric annotation performed by Lee uses a shadowed-2D image approximated by the landmark coordinates using machine learning. Meanwhile, what is done by Lindner uses lateral cephalogram images. This method accurately detects landmarks' location in the 2D image. However, the detection results cannot be used directly for facial reconstruction because they do not provide 3D spatial coordinate information.

Orthopaedic landmark positioning proposed by Jacinto [16] outlines the multi-atlas approach. He proposed a method for determining the position of landmarks on a 3D orthopaedic atlas model of the patient's bone structure. There are two three-dimensional models, namely the patient and expert models. Both structure model consists of a triangulated mesh. The positioning of the landmarks is obtained by registering the patient model to the expert model. An expert identifies landmarks positions on the expert model as a pre-defined landmark model. It is then registered to the patient model.

The process includes two stages: computing the global mesh registration and local registration. The initial fitting stage of the patient model to the expert model is carried out using the Iterative-Closest-Point (ICP) algorithm. This stage aims to attach the patient model to the expert model. The next step is refining the fitting using adaptive local rigid registration. This stage places (transfers) the positions of the landmarks in the expert model to the patient model. The final position of the landmarks is refined using an automatic selection of a set of best-positioned landmarks. This approach increases the efficiency of unsupervised landmark positioning.

Lindner *et al*. [18] proposed a method of automatic landmark annotation on lateral cephalometric images. The positioning of these landmarks is a standard procedure in orthodontic diagnosis and treatment. The Fully Automatic Landmark Annotation System (FALA) was developed works by utilizing machine learning to determine the landmark position based on the landmark position data annotated manually by experts. Two stages were deployed, namely Random Forest regression, to make the system robust against various variations in the image, and the second stage was the Constrained Local Model framework (RFRV-CLM) which specifically determined the position of landmarks.

The data and the results of the landmark annotation by Lindner [18] were performed on the lateral cephalogram image. In other words, a two-dimensional image from the side of the object. The landmark detection was carried out on the bone contour and isolated the desired landmark position. The challenge is to distinguish a landmark that is on the contour of the bone or not. For this reason, adding a patch size to the training sample allows the system to learn more about the environment where the landmark is located in the learning process.

A three-dimensional modality for detecting the position of landmarks based on machine learning was carried out by Lee *et al*. [17], is driven by a shift in the trend of using three-dimensional images in surgical simulations. However, this study does not directly process three-dimensional images to detect landmarks. Shaded two-dimensional rendering of three-dimensional imagery was used as the

modality. The position of the landmark is annotated on the image to find its relationship with machine learning.

Lee sees the problem in three-dimensional processing data is the number of voxels. This challenge was overcome by working on two-dimensional shaded objects with lighting and viewing angles variations. The method uses VGG-net to draw the correlation of landmark positions based on a manually marked dataset.

Another work on cephalometric landmark detection was carried out [20]. The challenges faced are also related to the position of landmarks in three-dimensional space. They propose multistage deep reinforcement learning in detecting landmarks. The modality used is similar to Lee [17], using 3D images rendered as 2D shaded images from various views. The work consists of two stages, namely the learning stage and the inferencing stage. The use of multi-stage is due to landmarks whose position in 3D empty space (such as the Foramen magnum or Sella) cannot be depicted by simple 2D images. Both single and multi-stage approaches can detect other landmarks.

The challenge in using three-dimensional data in machine learning is the enormous data that must be processed. Several works, such as Lee [17], Jacinto [16], and Lindner [18], perform alternative three-dimensional contour processing with a two-dimensional approach. However, machine learning does not require raw image data as input. Refined or extracted information from the image can also be used if it represents the expected features. The Surface Curvature Feature method proposed by Yuniarno [15] works as a three-dimensional surface descriptor that translates the shape of the surface curvature into simpler context information for machine learning.

Unlike those works that use a view-based approach, Yuniarno [15] uses a feature-based method for point cloud registration. He proposed an iterative closest point (ICP) algorithm as used by Jacinto but by utilizing surface curvature feature estimation (SCF), is a feature-based approach that works on local features. This method performs fitting k-NN local points to the hyperbolic paraboloid equation. This method was originally used to perform point cloud matching. In general, SCF acts as a model that describes the shape of curvature. The iterative closest point-surface curvature feature (ICP-SCF) algorithm can describe a local surface so that point cloud matching can be performed.

From the related works, two issues need to be addressed in completing the task of classifying and detecting landmarks in 3D space. Unlike previous works that use two-dimensional modalities. First, this work focuses on positioning landmarks at the coordinate level. Second, apply machine learning to classify and detect the position of a landmark. The first issue is solved by representing the local surface shape context using SCF. Unlike Yuniarno, in this work, SCF is used naively as a surface curvature descriptor. The extraction results are then used as the learning input. We hypothesize that an adjacent point will have a similar surface shape for the second issue. If they are a landmark, then several points will be predicted as the same class with a similar probability level. Therefore, the top probability prediction points are taken and then filtered based on the most significant cluster.

B. Proposed Method

Considering the previous two issues, we propose a method aimed at detecting, classifying, and determining the position of cranial landmarks (Figure 1). The modality used is multislice-CT with DICOM format. Cranial sections are extracted as point clouds and stored as Cartesian coordinates. Next, the local surface context for each point cloud is extracted with SCF. This descriptor solves the problem of volumetric input by converting 3D data into feature representation values. Associations between surface shapes and landmarks are obtained by deep learning. Points with top-N probabilities are then filtered to isolate the largest cluster as a landmark area.

C. Surface Curvature Feature

Each landmark can be recognized based on the characteristics of its shape. A data representation model and a three-dimensional surface descriptor model are needed to achieve this hypothesis. The cranial is represented as a point cloud. Spatial data processing using a local coordinate system can provide a better perspective and simplify measurements. The Surface Curvature Feature (SCF) is used as the three-dimensional descriptor.

Surface curvature is the curvature of a curve on the surface that passes through a certain point [15]. Surface curvature must be able to provide consistent information about the characteristics of a curve with a certain center point p_i. The coordinate system as a parameter frame in which the
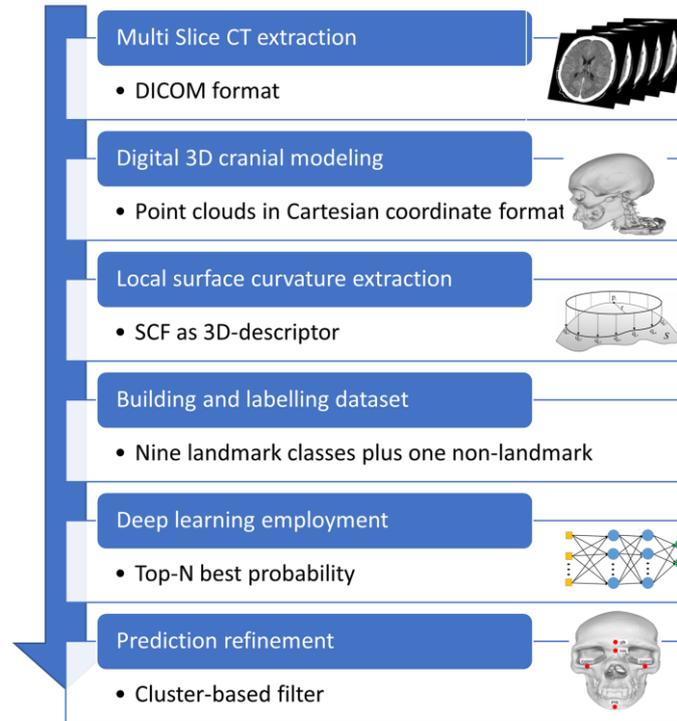
Fig. 1. Landmark Positioning Process based on Surface Curvature Feature

curvature function works must be uniform. For spatial work in geometric spaces, local coordinate systems are often used. The local coordinate system is built on a point p_i, where the normal vector n_i of the surface at point p_i is used as z coordinate axis. Meanwhile, the x and y coordinate axes are obtained from the principal direction of the p_i point surface.

D. Local Coordinate System

The local coordinate system (Figure 2) simplifies the marking and measurement process. This coordinate system is best used for smaller area extents. The conversion is done by finding a composite transformation that transforms point clouds so that the normal vector n_p of P coincides with the azimuth [0 0 1] as the local principal axis z. The transformation matrix is a composite of translation and rotation. The center of rotation must be on a point $P(x_p, y_p, z_p)$, so the translation is $\boldsymbol{t}_{x_p,y_p,z_p}$.

For the z-axis direction of the local coordinate system to be parallel to the surface's normal, the
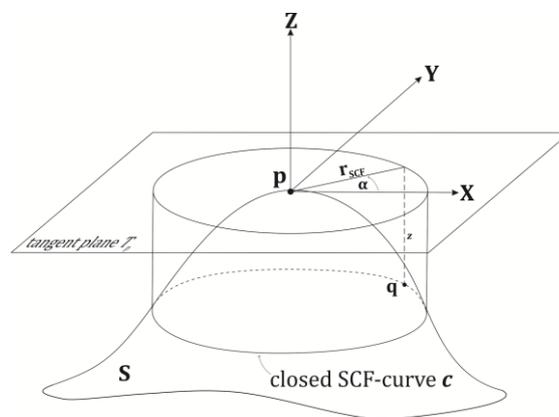

Fig. 2. Local coordinate system on . The normal vector n_p of P coincides with the azimuth as the local principal axis z

point cloud needs to be rotated. Rodrigues's rotation formula applied to find the rotation $\boldsymbol{R}_{P,\theta} = rot(\boldsymbol{n_p}, \boldsymbol{z})$, where $z = (0,0,1)$. It gives an efficient method for computing the rotation matrix $\boldsymbol{R}$ corresponding to a rotation by an angle θ about a fixed axis specified by the unit vector $\boldsymbol{n} = \boldsymbol{n_p} \times \boldsymbol{z} = (n_1, n_2, n_3)$. After the surface point cloud is transformed with a translation $\boldsymbol{t}$ and rotation $\boldsymbol{R}$, the next task is determining the principal directions.

E. Principal Directions Estimation

A regular surface has two principal directions. The principal directions of a surface are used to determine two vectors as the x and y axes of the local coordinate. They are perpendicular to each other before the z-axis, which is parallel to the normal. The principal direction of surface of p_i is a tangent direction at point p_i on a regular surface in which the normal curvature of the surface at the point p_i reaches an extremal value.

The principal directions define a direction in which the surface of the curve must travel to obtain the minimum and maximum curvature. The minimum and maximum curvature conditions are obtained from the eigenvalues of the shape operators. Meanwhile, the principal directions are obtained from eigenvectors which correlate with the eigenvalues. The maximum curvature λ_1 and the minimum curvature λ_2 of the curve is referred as the principal curvature [15].

The principal curvatures are the determinant of the Weingarten matrix:

$$\mathbf{W} = \begin{bmatrix} F_{xx} & F_{xy} \\ F_{yx} & F_{yy} \end{bmatrix} = \begin{bmatrix} A & B \\ B & C \end{bmatrix} \tag{1}$$

If λ_1 and λ_2 are the eigenvalues of Weingarten matrix $\mathbf{W}$, then the minimum curvature is λ_1 and the maximum curvature is λ_2 with $|\lambda_1| < |\lambda_2|$. The vectors associated with principal curvatures are called principal directions. A local coordinate combines the principal direction and the normal surface vector. They construct a coordinate frame (Darboux frame).

F. Surface Curvature Feature Extraction

The surface curvature feature is a sequence value that describes the curvature of a surface at a certain radius. Thirty-six sequence projection starting points were prepared (Figure 3). These points are in the planar ring perpendicular to the normal point p_i (centre) within ten degrees circular Z-axis (normal of p_i). The measurement starts from the first (zero degrees) projection point whose position is in the radius r on the X-axis. The next points are every 10 degrees counterclockwise.

One of the direct uses of SCF is to determine surface similarity. Two surfaces can be said to be similar in geometry, having similar features of the curvature of the surface. The similarity of surface geometry at two points can be compared by looking at the sequence values of the SCF. The deep learning approach can also be applied, considering the difficulty of finding a direct correlation between SCFs. By using deep learning, machines can relate the implicit definition of a classification
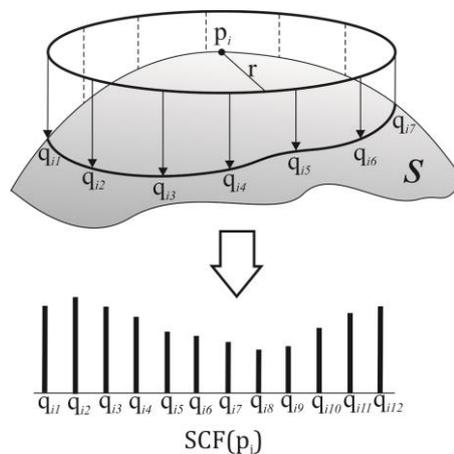

Fig. 3. Surface Curvature Feature extraction. This returns a sequential value for each surface with centre p. The sequential value is the projection distance of the planar ring to the surface. A planar ring has a radius as the specified parameter, with the centre at p

Fig. 4. Examples of four local surface, extracted by SCF 3D descriptor onto SCF values that are categorized as surface similar. A group of SCF categorized as similar will be trained to detect certain surface shapes.
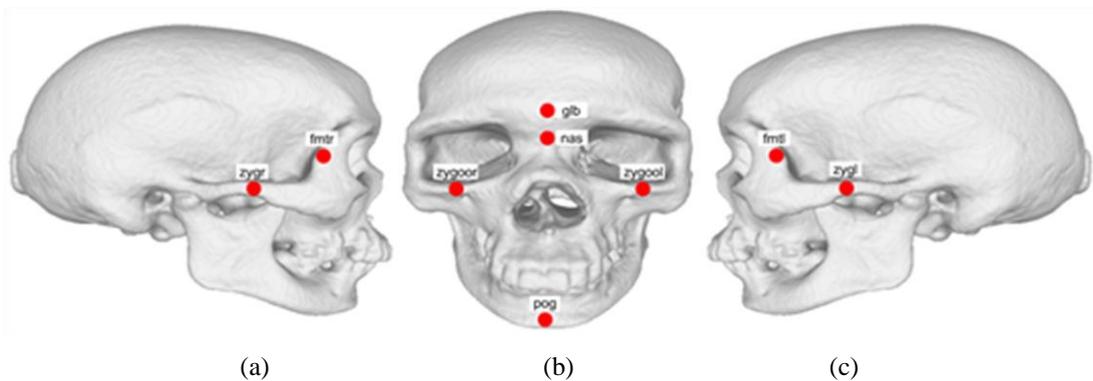

(a)   (b)   (c)

Fig. 5. Nine landmarks: (a) frontomalare temporale-right (fmtr) and zygion-right (zygr); (b) glabella (glb), nasion (nas), zygoorbitale-right(zygoor), zygoorbitale-left(zygool), and pogonion (pog); and (c) frontomalare temporale-left (fmtl) and zygion-left(zygl).

based on a given feature. The machine will detect points on the surface with a certain surface shape. The surface shape is characterized by several SCFs which are similarly categorized (Figure 4). Principal direction is applied to ensure that similar surface shapes are measured at a uniform starting point. In other words, the SCF value will appear with a similar histogram pyramid for a surface with a similar shape.

G. Data Preparation

The data sample consisted of multi-slice Computed Tomography (CT) of the skulls. The contours of the skulls were separated from soft tissue by a Hounsfield unit scale. CT is great for density-based segmentation [21]. The data are retrospective-anonymous, with an accuracy of 0.625 mm between slices of 512×512 pixels. Each skull-face sample consists of 400 to 500 slices.

Multi-slice CT of the head was then extracted from the skull surface based on tissue and soft tissue density differences. The surface is stored as a point cloud. Each local surface on the skull was extracted its surface shape using a three-dimensional descriptor as a surface curvature feature (SCF). The dataset consists of 458,000 local surfaces, extracted as SCF values as training-validating-testing data with 60:20:20 ratio.

For predicting landmark's positions, points will be taken at the part that is considered the area of the landmarks. They are frontomalare temporale-right (fmtr), zygion-right(zygr), glabella (glb), nasion (nas), zygoorbitale-right(zygoor), zygoorbitale-left(zygool), pogonion (pog), frontomalare temporale-left (fmtl), and zygion-left (zygl) [2][8][22][23]. The points are samples with a radius of 0.3 units from the centre of the desired landmarks (Figure 5).

H. Point Clouds Normalisation

A point cloud sample $PC = \{p_i\}_{i=1}^{N}$ is a collection of points extracted from multi-slice computed tomography (CT). The surface of the face and the skull are obtained by the outer contour extraction method [24] as point clouds, stored in an array of Cartesian coordinates ([x y z]). The outer points of the skull are recorded in a Cartesian format. Each skull point cloud was normalized to obtain a uniform scale. Centre point p_0 is determined as the average of the point cloud as the center of the coordinate system ([0 0 0]). Each skull is scaled so that the average point cloud distance to p_0 is one. With this normalization, the new coordinates p_i' are computed using the Equation (2) where $r_i = |p_i - p_0|$. Normalization reduces the dependence of gradients on the scale of the initial values on the training process. This approach results in higher learning rates without the risk of divergence.

$$p_i' = (p_i - p_0)\left(\frac{N}{\sum_{i=1}^{N} r_i}\right) \tag{2}$$

I. Feature Extraction

The SCF sequential value is a series of values that represent the shape of a surface by measuring the distance of the ring projection to the surface. The ring radius and the point cloud sample radius are operator-defined parameters. Measurements were made on a normalized skull point cloud. In this study, the ring radius and sample radii were 0.25 and 0.5 normalized units (normalized units after application of Equation 1). The principal direction axis determines the starting point of measurement, which moves counterclockwise every ten degrees. Figure 6 shows the measurement of the SCF's sequential value as a surface's three-dimensional descriptor.

A total of 36 user-defined projection starting points were prepared. These points are in the planar plane perpendicular to the normal point p_i within 10 degrees circular Z-axis (also normal of p_i). The measurement starts from the first (zero degrees) projection point, whose position is in the radius r on the X-axis.

There are cases where the projection of the ring does not touch the surface to reach the pre-defined *maximum projection* value. SCF is generally in a small range of values. The dominant *maximum projection* value reduces the accuracy of the training, is not an actual projection value,
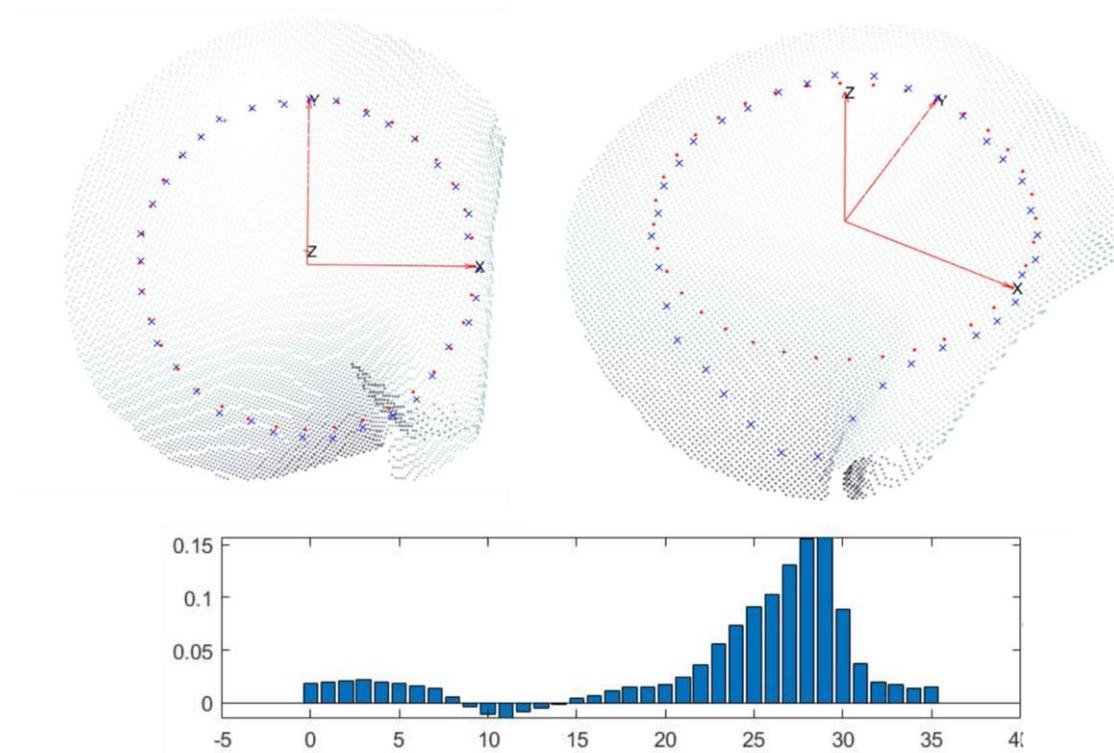

Fig. 6. Surface Curvature Feature extraction on point cloud sample. Point cloud coordinate system is converted into local coordinate. The SCF sequential value as the result of projection measurements of SCF ring to the local surface.

but a maximum given value if the surface projection is not achieved. To eliminate the dominance of large $maximum\ projection$, this is set to zero.

J. Neural Network Deployment

Input data is in the form of SCF, which is already a feature. Naturally, it can be processed directly using a multi-layer perceptron. A model is developed to detect landmarks into ten classes (glb, fmtr, fmtl, zygr, zygl, zygoor, zygool, pog, nasion, and non-landmark). It consists of layers which include an input layer, hidden layers, and output layer, is trained to acquire the probability value of membership of an SCF into ten landmark classes. The probability value shows how certain the data is predicted to be a class member.

The SCF sequential value is the projection distance of the ring with radius r on the surface of the skull, work using 36 equidistant projection points. For example, on a convex surface, all SCF sequences will be positive in extreme cases. However, with variations in surface shape, they could be in the range of negative to positive. It takes an appropriate activation function to accommodate the range of values. The $tanh$ activation function is considered because it can maintain a negative magnitude in the training input. Using the $tanh$ activation function in combination with the $relu$ gives significantly better results than the linear function. The loss function used in this classification model is $Categorical - CrossEntropy$, as it is commonly used in multi-label classification.

III. Results and Discussions

An artificial neural network model has been created to classify nine landmarks into ten classes (Figure 7). Non-landmark points are treated as separate labels beside the nine landmarks. The dataset consists of SCF values of 458.000 local surfaces labeled with nine landmark classes plus one non-landmark class. A non-landmark class is included to increase confidence in the prediction results. Most of the skull surface is not a landmark. So, adding a class that says "not a landmark" really helps improve accuracy. The neural network model was trained for 200 epochs to ensure the training results (Figure 8). It can be seen that around the 22nd epoch, there has been convergence. In some epochs, insignificant spikes occur, and the learning returns to convergence.

An untrained cranial point cloud was tested on the neural network model. The point cloud consists of 200,000 evenly distributed samples (local surfaces). Predictions for each landmark are scattered in several areas but mostly clustered in the designated target. The class prediction in Figure 9 was fired by the top hundred highest predicted probability values for each landmark class.
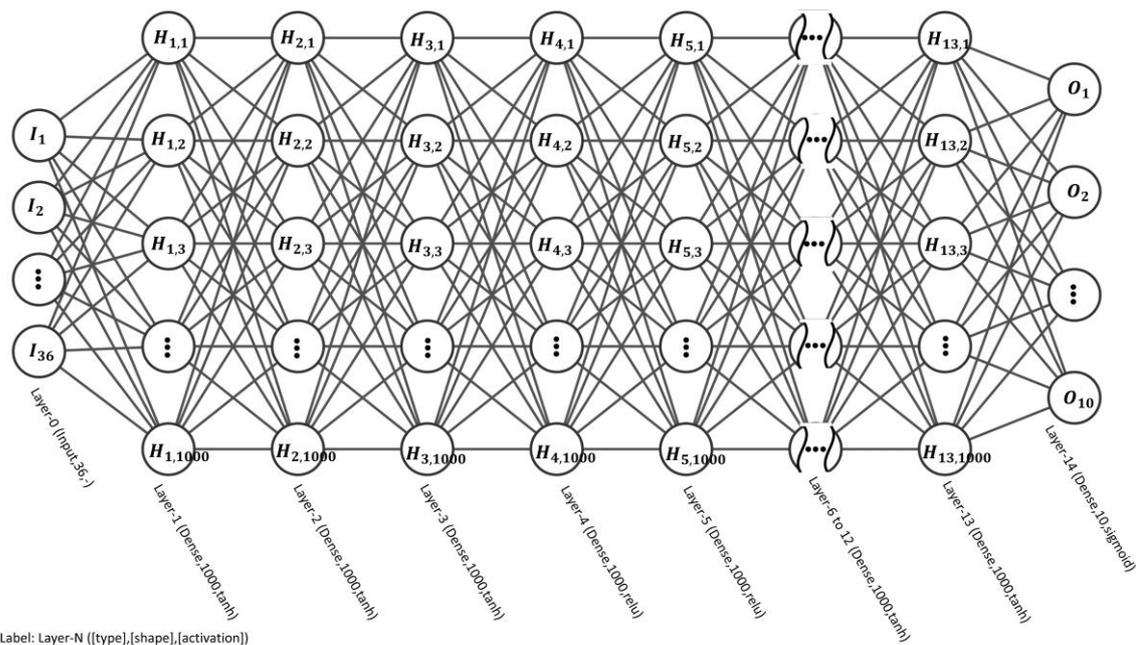

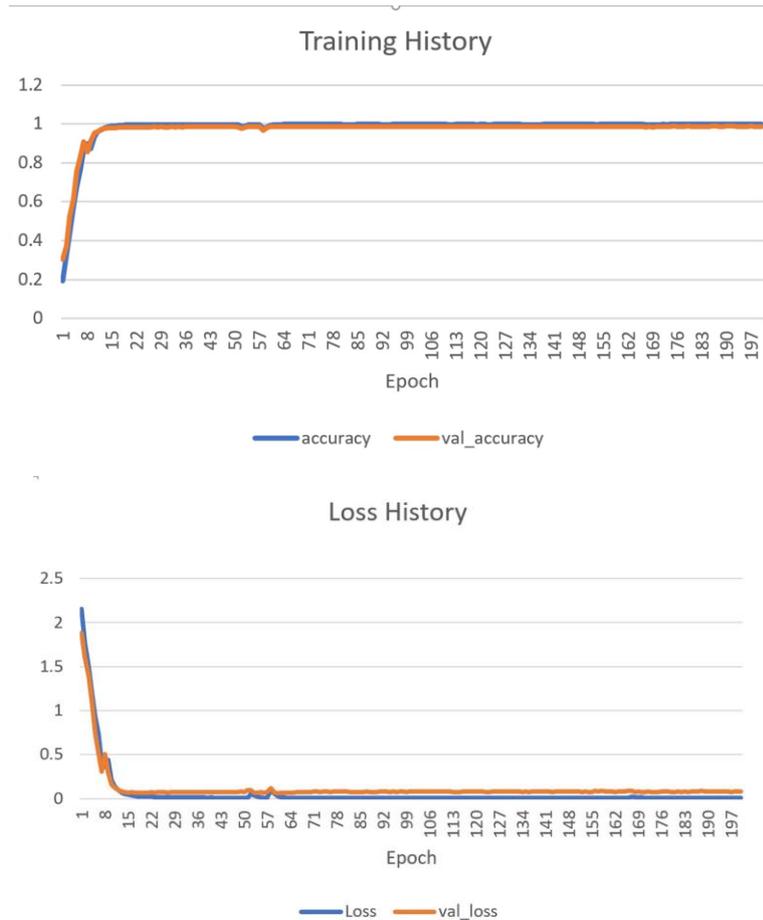Fig. 7. Model summary with categorical cross entropy loss

Fig. 8. Training (top) and lost (bottom) history of the learning process with 200 epochs. In approximately the 22nd epoch, there has been convergence. In some epochs, insignificant spikes occur and returns to convergence.

An expert on the cranial surface annotates ground truths. The coordinates of the annotation points are extracted to be compared with the predicted results. The accuracy of each prediction point is measured as a three-dimensional Euclidean distance (Equation 3) between the ground truth (μ_x, μ_y, μ_z) and the prediction point (x_i, y_i, z_i). Seven of the nine landmarks give promising results compared to the ground truth. The prediction for zygoor and zygool have multiple clusters, with the largest are closer to ground truth. Accuracy is then improved with the implementation a distance radius-based filter and cluster-based filter.

$$\text{distance}_i = \sqrt{(x_i - \mu_x)^2 + (y_i - \mu_y)^2 + (z_i - \mu_z)^2} \qquad (3)$$

Radius-based filters aim to select the points closest to the centre of mass (μ_x, μ_y, μ_z) of the prediction points. The mass centre is determined by the average coordinates of the predicted points. Each prediction points (x_i, y_i, z_i) is then calculated its distance (Equation 3) from the centre (μ_x, μ_y, μ_z) and sorted in ascending order. The filter aims to leave candidate prediction points close to the ground truth. The hypothesis is that the prediction points should converge close to the ground truth position.

The radius-based filter assumes that the center of mass is at the ground truth position. This filter works well on several landmark classes such as FMTL, GLB, POG, NAS, ZYGL, and ZYGL. But for other classes, this filter does not give consistent results. The problem is that the prediction points are spread over a wide range. This filter is also not able to handle multi-cluster predictions.

Filter by cluster is carried out to cope with multi-cluster prediction. This filter is a simplification of neighborhood-based filtering techniques by Han [25]. This relies on the number of neighbors per prediction point within a normalized 0.05-unit radius. The idea is to eliminate isolated points or
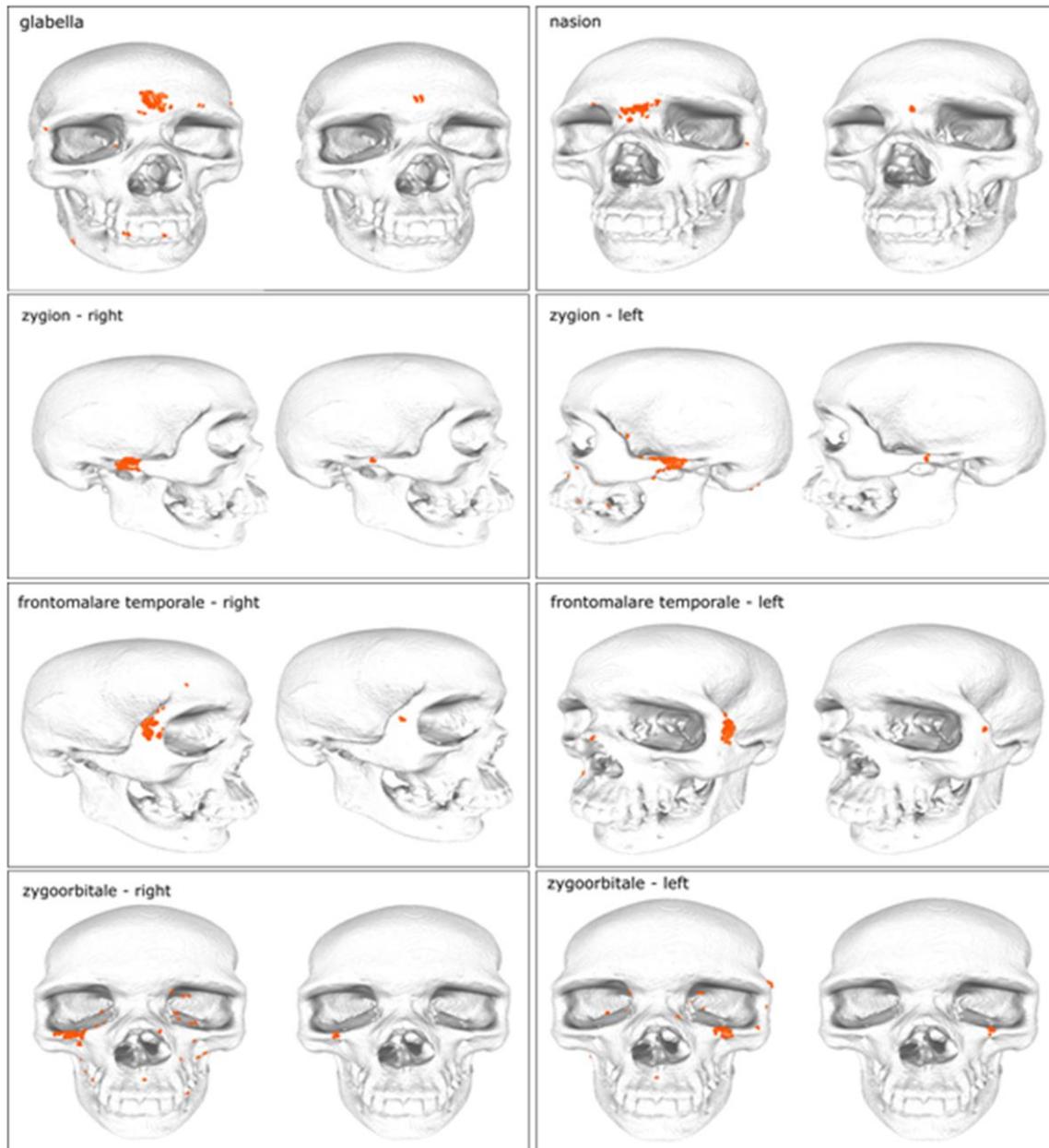
Fig. 9. Results of top-hundred landmark detection, compared with cluster-based filtering side by side.

small clusters (number of neighbors). Thus, the tighter the filter, the larger the cluster points remain. In the case of multi-cluster prediction, radius-based filters are not optimal. The average distance from the ground truth widened in the tighter filter. Otherwise, cluster-based filter consistently provides better results than radius-based filters (Table 1). On zygool and zygoor landmarks, the prediction points are spread over a wide range. Especially in zygool landmarks, there is more than one cluster (Figure 9). Radius-based filters failed to improve accuracy for finding top-20 points, while cluster-based filters cope this problem. The performance of cluster-based filters is generally better than mass-radius-based filters (Figure 10 and Figure 11). Referring to the position of the actual landmark, Figure 9 shows visual prediction of landmark and the filtered results. The mean distance from the ground truth is reduced (better). It improves the accuracy by eliminating points scattered too far from the center of mass.

The results shown in Figure 10, Figure 11, and Table 1 confirm that the filter can eliminate less relevant points. The cluster-based filter consistently provides better results. It also strengthens the hypothesis that the prediction of a landmark will be clustered because it has a similar SCF. With neighborhood-based filtering techniques [25], they can be clustered.

Table 1. The average distance between the top-n prediction points and the ground truth points

| Landmark	Top100	Top-90	Top-80	Top-70	Top-60	Top-50	Top-40	Top-30	Top-20	Top-10	Top-5
GLB											
Cluster filter	0.1420	0.0615	0.0457	0.0406	0.0384	0.0348	0.0316	0.0308	0.0287	0.0259	0.0228
Radius filter	0.1420	0.0531	0.0449	0.0417	0.0379	0.0353	0.0318	0.0289	0.0251	0.0280	0.0246
FMTR											
Cluster filter	0.0973	0.0382	0.0336	0.0294	0.0258	0.0223	0.0199	0.0163	0.0134	0.0104	0.0125
Radius filter	0.0973	0.0387	0.0372	0.0361	0.0378	0.0387	0.0390	0.0418	0.0478	0.0611	0.0651
ZYGR											
Cluster filter	0.1933	0.0483	0.0435	0.0393	0.0371	0.0359	0.0324	0.0277	0.0243	0.0206	0.0162
Radius filter	0.1933	0.0484	0.0440	0.0417	0.0379	0.0354	0.0326	0.0334	0.0337	0.0322	0.0324
POG											
Cluster filter	0.2150	0.1444	0.0623	0.0582	0.0554	0.0503	0.0506	0.0525	0.0527	0.0553	0.0512
Radius filter	0.2150	0.1093	0.0623	0.0545	0.0474	0.0412	0.0357	0.0305	0.0249	0.0185	0.0136
FMTL											
Cluster filter	0.0917	0.0395	0.0350	0.0338	0.0311	0.0301	0.0278	0.0244	0.0227	0.0184	0.0218
Radius filter	0.0917	0.0400	0.0378	0.0359	0.0336	0.0310	0.0287	0.0267	0.0229	0.0225	0.0291
ZYGL											
Cluster filter	0.1717	0.0944	0.0489	0.0414	0.0388	0.0353	0.0316	0.0315	0.0313	0.0331	0.0304
Radius filter	0.1717	0.0798	0.0492	0.0452	0.0433	0.0384	0.0397	0.0421	0.0429	0.0467	0.0562
ZYGOOR											
Cluster filter	0.4486	0.3882	0.3394	0.2567	0.1774	0.0501	0.0350	0.0230	0.0206	0.0240	0.0316
Radius filter	0.4486	0.3385	0.2580	0.1953	0.1205	0.0781	0.0671	0.0806	0.1153	0.1968	0.2798
ZYGOOL											
Cluster filter	0.7019	0.6982	0.6783	0.6244	0.5983	0.5412	0.4521	0.2040	0.0219	0.0177	0.0212
Radius filter	0.7019	0.6237	0.5780	0.5062	0.5040	0.5786	0.6806	0.7681	0.7726	0.6747	0.6177
NAS											
Cluster filter	0.3305	0.2705	0.2427	0.1393	0.0666	0.0708	0.0719	0.0731	0.0721	0.0717	0.0695
Radius filter	0.3305	0.2150	0.0900	0.0634	0.0584	0.0551	0.0505	0.0494	0.0511	0.0675	0.0700
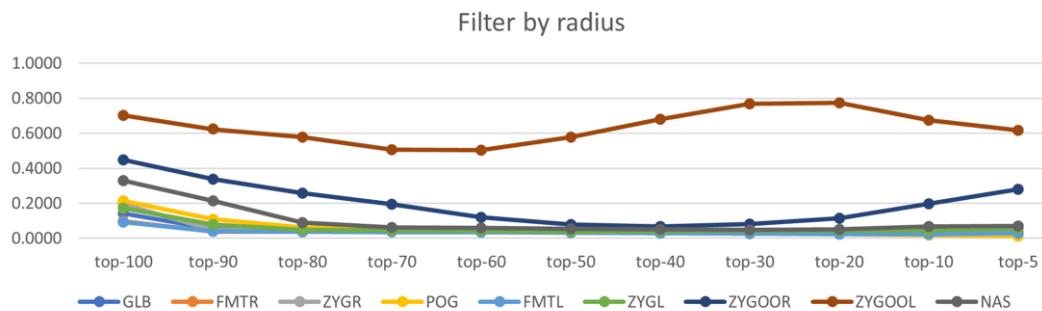
Fig. 10. Filter landmark prediction points by distance radius. With the center of mass of the predicted points as the center of the radius.

Using SCF with the Iterative Closest Point (ICP-SCF) successfully registered point clouds based on local surface SCF [15]. A similar idea is used in this work with a machine learning approach. Learning runs convergently, and the prediction results show that SCF can act as a surface descriptor that can be recognized for its landmark characteristics. The main challenge in using 3D modalities is many voxels [17]. This work addresses this issue by using three-dimensional descriptors to draw the shape context of a three-dimensional surface. Input learning is information in the form of a 3D surface shape context. Thus, the workload of the learning process will be much reduced.
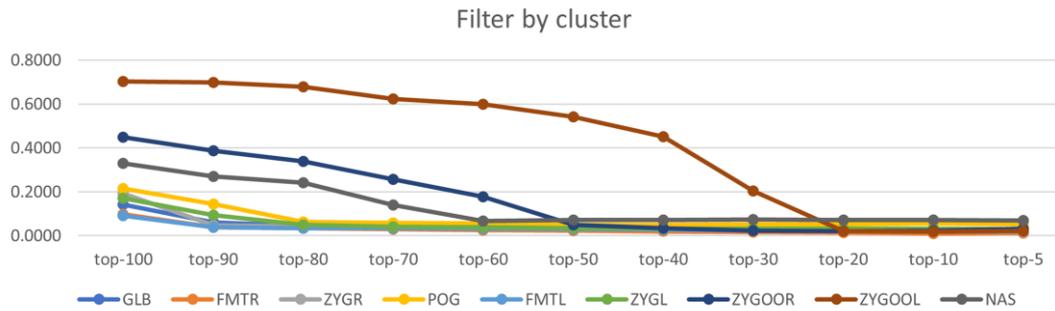
Fig. 11. Filter based on the largest cluster

The feature-based approach is not only good for classification based on global features. However, a good understanding of local surface features can also be used to detect the unit shape of a landmark. So far, such detection and prediction have only been carried out using a view-based approach, such as works by Lindner [18], Lee [17], Jacinto [16], or Kang [20]. The method of automatically detecting the position of landmarks is necessary in craniofacial reconstruction. Especially in computer-assisted reconstruction, the morphological approach is a computational opportunity because it can be simulated by a computer based on available medical modalities (CT, MRI, USG).

IV. Conclusions

The hypothesis that inspires this work is that each landmark has a unique surface shape characteristic. SCF is used as a feature representation of the shape of the landmark surface. The neural network model has a convincing performance, with the average distance to the ground truth is 0.0326 normalised units. Several prediction points with the highest probability values are taken for each landmark. They are scattered but tend to be clustered in the desired area. Cluster-based filters are better than mass-radius-based filters, consistently giving better pinpoint accuracy, especially in multi-cluster cases. Precision is measured based on the average distance of the top predictions to the ground truths. The cluster-based filter is needed for multi-cluster distribution to isolate the largest cluster. It is successfully coping with the multi-cluster case. Based on these results, using SCF was successful as a 3D descriptor representing local surface features. With the success of this work, the next research is its implementation as part of the craniofacial reconstruction framework. Not only for forensics, but the detection of landmarks is also helpful in simulating medical reconstruction and plastic surgery. Future research based on this work is organ damage reconstruction research with template transfer based on landmarks. With a similar concept, research on implant reconstruction is also possible.

Declarations

Conflict of interest

The authors declare no known conflict of financial interest or personal relationships that could have appeared to influence the work reported in this paper.

Additional information

Reprints and permission information are available at http://journal2.um.ac.id/index.php/keds.

Publisher's Note: Department of Electrical Engineering - Universitas Negeri Malang remains neutral with regard to jurisdictional claims and institutional affiliations.

References

[1]   P. T. Jayaprakash, "Conceptual transitions in methods of skull-photo superimposition that impact the reliability of identification: A review," *Forensic Sci. Int.*, vol. 246, pp. 110–121, 2015.

[2]   M. A. Lenza, A. A. De Carvalho, E. B. Lenza, M. G. Lenza, H. M. De Torres, and J. B. De Souza, "Radiographic evaluation of orthodontic treatment by means of four different cephalometric superimposition methods," *Dental Press J. Orthod.*, vol. 20, no. 3, pp. 29–36, 2015.

[3]   C. C. Snow, B. P. Gatliff, and K. R. McWilliams, "Reconstruction of facial features from the skull: An evaluation of its usefulness in forensic anthropology," *Am. J. Phys. Anthropol.*, vol. 33, no. 2, pp. 221–227, 1970.

[4]   H. Ullrich and C. N. Stephan, "Mikhail Mikhaylovich Gerasimov's Authentic Approach To Plastic Facial Recpnstruction," *Anthropologie*, vol. LIV/2, no. October 2014, pp. 97–107, 2016.

[5]   P. Claes, D. Vandermeulen, S. De Greef, G. Willems, J. G. Clement, and P. Suetens, "Computerized craniofacial reconstruction: Conceptual framework and review," *Forensic Sci. Int.*, vol. 201, no. 1–3, pp. 138–145, 2010.

[6]   P. H. Suputra, A. D. Sensusiati, E. M. Yuniarno, M. H. Purnomo, and I. K. E. Purnama, "3D Laplacian Surface Deformation for Template Fitting on Craniofacial Reconstruction," in *ICCCM'20: Proceedings of the 8th International Conference on Computer and Communications Management*, 2020, pp. 27–32.

[7]   M. de Buhan and C. Nardoni, "A facial reconstruction method based on new mesh deformation techniques," *Forensic Sci. Res.*, vol. 3, no. 3, pp. 256–273, 2018.

[8]   T. Gietzen *et al.*, "A method for automatic forensic facial reconstruction based on dense statistics of soft tissue thickness," *PLoS One*, vol. 14, no. 1, pp. 1–19, 2019.

[9]   P. Guyomarc'h, F. Santos, B. Dutailly, P. Desbarats, C. Bou, and H. Coqueugniot, "Three-dimensional computer-assisted craniometrics: A comparison of the uncertainty in measurement induced by surface reconstruction performed by two computer programs," *Forensic Sci. Int.*, vol. 219, no. 1–3, pp. 221–227, 2012.

[10] L. Jiang, J. Zhang, B. Deng, H. Li, and L. Liu, "3D Face Reconstruction with Geometry Details from a Single Image," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 4756–4770, 2018.

[11] A. Lodha, M. Mehta, M. N. Patel, and S. K. Menon, "Facial soft tissue thickness database of Gujarati population for forensic craniofacial reconstruction," *Egypt. J. Forensic Sci.*, vol. 6, no. 2, pp. 126–134, 2016.

[12] B. Rosario Campomanes-Álvarez *et al.*, "Modeling Facial Soft Tissue Thickness for Automatic Skull-Face Overlay," *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 10, pp. 2057–2070, 2015.

[13] L. J. Short, B. Khambay, A. Ayoub, C. Erolin, C. Rynn, and C. Wilkinson, "Validation of a computer modelled forensic facial reconstruction technique using CT data from live subjects: a pilot study," *Forensic Sci. Int.*, vol. 237, pp. 147.e1-147.e8, 2014.

[14] Y. Zhao *et al.*, "Laplacian Musculoskeletal Deformation for Patient-Specific Simulation and Visualisation," in *2013 17th International Conference on Information Visualisation*, Jul. 2013, pp. 505–510.

[15] E. Mulyanto Yuniarno, M. Hariadi, and M. Hery Purnomo, "Point cloud registration for a non-deformable object using surface curvature features," *J. Theor. Appl. Inf. Technol.*, vol. 51, no. 3, pp. 506–514, 2013.

[16] H. Jacinto, S. Valette, and R. Prost, "Multi-atlas automatic positioning of anatomical landmarks," *J. Vis. Commun. Image Represent.*, vol. 50, no. November, pp. 167–177, 2018.

[17] S. M. Lee, H. P. Kim, K. Jeon, S. H. Lee, and J. K. Seo, "Automatic 3D cephalometric annotation system using shadowed 2D image-based machine learning," *Phys. Med. Biol.*, vol. 64, no. 5, 2019.

[18] C. Lindner, C. W. Wang, C. T. Huang, C. H. Li, S. W. Chang, and T. F. Cootes, "Fully Automatic System for Accurate Localisation and Analysis of Cephalometric Landmarks in Lateral Cephalograms," *Sci. Rep.*, vol. 6, no. August, pp. 1–10, 2016.

[19] H. Kanaan and A. Behrad, "Three-dimensional shape recognition and classification using local features of model views and sparse representation of shape descriptors," *J. Inf. Process. Syst.*, vol. 16, no. 2, pp. 343–359, 2020.

[20] S. H. Kang, K. Jeon, S. H. Kang, and S. H. Lee, "3D cephalometric landmark detection by multiple stage deep reinforcement learning," *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, 2021.

[21] N. Chowdhury *et al.*, "Concurrent segmentation of the prostate on MRI and CT via linked statistical shape models for radiotherapy planning," *Med. Phys.*, vol. 39, no. 4, p. 2214, 2012.

[22] M. Bayome, J. Hyun Park, and Y. A. Kook, "New three-dimensional cephalometric analyses among adults with a skeletal class I pattern and normal occlusion," *Korean J. Orthod.*, vol. 43, no. 2, pp. 62–73, 2013.

[23] A. H. Ross, D. E. Slice, and S. E. Williams, "Geometric Morphometric Tools for the Classification of Human Skulls. Research Report," pp. 1–59, 2010.

[24] M. A. Ulinuha, E. M. Yuniarno, S. M. S. Nugroho, and M. Hariadi, "Outer contour extraction of skull from CT scan images," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 185, no. 1, 2017.

[25] X. F. Han, J. S. Jin, M. J. Wang, W. Jiang, L. Gao, and L. Xiao, "A review of algorithms for filtering the 3D point cloud," *Signal Process. Image Commun.*, vol. 57, no. May, pp. 103–112, 2017.