3D Human Facial Traits' Analysis for Ethnicity Recognition Using Deep Learning

Noor H. Reda[1]* , Hawraa H. Abbas[2,3]

[1] College of Information Technology, University of Babylon, Babil 51002, Iraq
[2] College of Information Technology Engineering, Al-Zahraa University for Women, Karbala 56001, Iraq
[3] Department of Electrical and Electronic Engineering, University of Kerbala, Karbala 56001, Iraq

Corresponding Author Email: nhredha@gmail.com

ABSTRACT

Among different recent technologies proposed for human face classification and recognition, solutions based on analyzing the 3D geometric facial features emerged as a promising academic and practical direction. Researchers have examined both holistic and local approaches to analyzing the 3D face regions to study the impact of facial features in real-life applications such as medical and security implementations. However, a few works have investigated the relevant impact of the extracted geometric features from the descriptive local regions of the human face on identifying human ethnicity. This work proposes a classifier to categorize individuals into their distinctive ethnic groups and deeply analyzes the facial feature variations to highlight the most descriptive parts and features of the human face in race classification. The proposed ML-based classifier is preceded by extracting the 3D facial features from 3D meshes using the recent SIFT and Geodesic distance calculations. In addition, it implements and discusses the initial important preprocessing steps including, cropping the frontal parts, correcting the head pose, selecting the suitable initial key points, aligning the 3D meshes, and implementing the suitable template-based 3D registration. The proposed NN race classifiers are built and evaluated using Headspace, a well-known multi-ethnic dataset, and achieved high accuracy (90% globally, and 100% for the mouth area) especially while using the SIFT features.

1. INTRODUCTION

Analyzing the human face has a long tradition starting from ancient populations basing on their observed descriptive facial attributes and using the natural ability of human visual system perception. However, the term "Facial morphology" is commonly used to describe studying the human facial form, shape, and structure. To analyze the facial details, many face recognition directions have been suggested and implemented for various medical and social purposes. Psychology, Neurology, in addition to sociology specialists in the 20th century highlighted essential questions about the human ability to identify faces. Visual assessment or facial appearance evaluation is the process of determining a human's identity or class according to his outward facial appearance. Anthropometric measurements are used to precisely measure the human face and recognize his identity. For example, skin tone, hair color, and eye color are prevalent traits examined visually with an acceptable degree of objectivity in human face perception procedures. However, 2D imaging faced difficulties and limitations due to many factors such as the resolution impact, the techniques of capturing, head pose variations, and illumination changes. The emergence of 3D imaging techniques such as Stereophotogrammetry and laser scanners, allows obtaining real 3D facial images including their natural color and texture details, in addition to the

volumetric and depth information of the captured images. In addition to providing identity clues, the human face also conveys more demographic data such as reflecting their ethnicity, age, and gender (aka soft biometrics that influence applications of face recognition, security surveillance, and indexing facial data). Such research efforts led to a remarkable achievement in face attribute recognition using both 2D and 3D models, especially with the growing Machine Learning (ML) and deep learning (DL) abilities.

An anatomical study Zhuang et al. [1] showed that geometrical features also contain abundant information on soft biometrics. Early works such as Bertillon and McClaughry [2] created the first personal identification system for security identification using three groups of characteristics: (1) anthropometric, the measurements describe the height and length of the body parts, (2) morphological, the descriptors of the outer shape (such as anomalies of the fingers), and appearance (such as eye color), and (3) peculiar marks, the distinguishing signs seen on the body such as scars and moles. Heckathorn et al. [3] proposed a combination of personal characteristics or soft biometrics, such as race, gender, eye color, and several other obvious marks, to be utilized to accurately detect the identity. For instance, considerable differences in face dimensions, and lip distances among different ethnic groups of workers from the United States are concluded by Zhuang et al. [4]. These statistics showed that

African-American individuals have distinguishable face and lip lengths from other ethnic groups such as; Caucasians, and Hispanics. In addition, the Caucasian face width differs from both African Americans and Hispanics. Zhuang et al. [5] collected anthropometric measurements from the individuals' faces. The author observed that females can be distinguished from males by ten dimensionality features. Related to ethnic groups, the author observed that the face length feature has a significant impact on identifying the African-Americans and Hispanics races since they have longer faces than Caucasians. Han and Choi [6] showed that Africans and Americans have shallower and smaller noses compared to Caucasians by a length difference of approximately two mm. Rebar et al. [7] in 2004, noticed that Caucasians can be distinguished by their long narrow faces for both males and females. Ballihi et al. [8] extracted radial curves and several specific geometrical circulars using the AdaBoost algorithm for gender classification on the FRGCv2 dataset. Another combination of 3D faces' geodesic path features is examined by Abbas et al. [9] to implement a gender classifier on fifteen-year-old individuals from the ALSPAC medical dataset. The author observed that the nose morphology highly impacts identifying the face, especially the paths on Cupid's bow, nose ala, and the inner eye canthi. Recent research by Mohammad and Al-Ani [10] showed that attribute-based biometric systems such as ethnicity are strong supporters of explicit essential biometrics in face recognition applications.

1.1 3D facial feature descriptors and feature extraction

Typically, 3D face recognition systems include multiple steps starting with the initial capturing using devices such as; multiple cameras and laser scanners. The obtained raw faces are then preprocessed by detecting the face area, denoising, cropping, reconstructing the missing part and holes, aligning it with other faces, and correcting the 3D pose to easily match it with the targeted or templates. Noise always inevitably affects most 3D scans captured by different acquisition devices to some degree. The general first step of denoising is usually done by cropping the face region due to its ability to eliminate irrelevant face data such as neck area and hair mineral details. Then, to describe the 3D faces in lower memory space, the important and descriptive features such as edges, corners, Gabor features, HOG features, and more are detected and extracted to be saved in vectors and matrices. In addition, the interest points such as landmarking or key points are determined manually or automatically using key point detectors such as ICP, SIFT, and other types to facilitate extracting the local features, calculating distances between the interest points, and calculating the sizes of face traits. A postprocessing step is also done to select the most important features that can cause changing the class of the face while they are changing in addition to advanced dimensionality reduction to reduce the resulted feature vectors and matrices using statistical methods such as PCA. The final step in face recognition is to explore the similarity among faces and match the extracted features between templates and targeted faces from the unseen input or from specific datasets using the similarity measures and the abilities of machine learning and supervised classification approaches such as SVM and recent deep learning such as CNN and pre-trained networks. Several common main steps of 3D face recognition are summarized in Table 1.

3D face recognition also can cope the uncontrolled illumination situations since it does not solely rely on pixel-level brightness while comparing the faces. Automatic face segmentation is also one of the new advantages obtained by 3D recognition since the image background can be typically synthesized separately in the reconstruction process.

Table 1. Commonly phases in 3D face classification

| Seq.	Phase Title	Description
1	Face Capturing	2D, 3D
2	Preprocessing	cropping, alignment, pose correction, landmarking, …
3	Obtaining Features	Detection, extraction, description, …
4	Postprocessing	Dimensionality reduction, selection, normalization, PCA, …
5	Machine Learning	SVM, NN, DT, …

To simplify using the 3D features, another direction of research speeds up the features comparison by reducing the 3D features comparison into a 2D profile contour similarity measure between faces. Statistical approaches such as Principal Component Analysis (PCA), and Linear Discriminant Analysis (LDA), and that are have been used to decrease memory consumption by using a smaller-size descriptive feature. The well-known and most common PCA then has been extended to 3D feature reduction due to its high performance and simplicity with a limitation of image quality dependency [11]. A multimodal PCA approach is used by Tsalakanidou et al. [12] to enhance the recognition accuracy while implementing the proposed work in a 40-subject experiment. A comparing study by Chang et al. [13] used PCA analysis to extract 3D features over a long period and achieved near-optimal while using both 2D and 3D recognition. To solve the expression and head pose variation, many authors tried to describe the facial surface in a way that has more robustness for the isometric deformations caused by multiple expressions or head angles of the uncontrolled capturing environments.

Furtherly, maintaining the Geodesic Distances across 3D facial surfaces was another important emerged enhancing in face recognition and classification. In 3D feature extraction, researchers also tried to define point sets or profiles of each 3D face scan using surface-based global or local extraction algorithms from the initial whole 3D face features extracted by the traditional algorithms. Hybrid algorithms (using both local and global geometric surface information) also have been proposed with a limitation of complex defining of point sets. Geometrical approaches extract geometric features from 3D mesh form, including profiles, curvature, radial, iso level curves, 3D landmarks-based geodesic distance, symmetry properties, and correspondence vectors to describe the human face abstractly. The feature extraction step is often followed by the classification steps. In other words, the classification framework needs to extract facial features initially to build a good classifier including obtaining features such as demographic features, Gabor features, binary features, biologically inspired features, and local binary patterns, in addition to analyzing the extracted features using the abilities of statistical approaches and machine learning (SVM, LDA, PCA, random forest, and CNN deep learning). The main performance aspects of face descriptors include feature matching speed, discriminability power, and consumed memory. Researchers mentioned that the discriminative power

needs to be balanced by computation complexity and extraction speed. A survey by Bellavia and Colombo [14] concluded that local SIFT-like descriptors can significantly compress the length and the matching time through their proper schemes. 3D key points are interesting points in shape that are found using certain geometric data from the face's surface.

The three steps of the keypoints-based approaches are typically interest points detection, feature extraction (or description), and face features matching. The major expert directions to find the interesting point within the 3D face are the landmarks based on facial traits and the salient points on the facial surface. These salient key points strategies include those that use the well-known SIFT and their extensions. Eye corners and nose tips are two commonly used landmarks highlighted by researchers in local approaches due to their robustness to the expression variation. Several other Local Feature-based approaches are summarized in Figure 1.
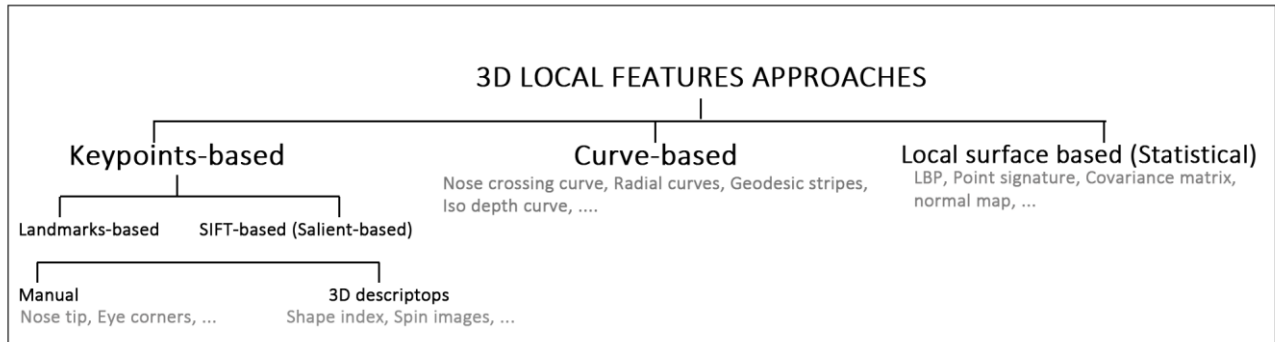
Figure 1. General steps in 3D face recognition

1.1.1 SIFT-like local descriptors

In region-based face recognition algorithms, landmarking is a key component of feature space manipulation. The success of feature extraction will be closely correlated with the precision of the landmarking. Scale Invariant Feature Transform (SIFT) was introduced to extract local features with an enhanced robustness to pose variations, it also showed a good capability in handling the partially occluded facial parts. The SIFT feature descriptor showed good robustness to uncontrolled illumination, uniform change in scaling and orientation, partial noise, and distortion. The SIFT compares the source image with different resolutions and with specific changes in the scale during computing descriptors, it avoids the traditional limitation of holistic face representations because it can select the features that are not affected by scaling, noise, and rotation. In addition to its use as a feature descriptor, SIFT can be considered a feature extraction method due to its ability to reduce the image description into a set of points used to detect similar patterns in other images. Thus, it is widely used in image matching and object detection.

As described by Lowe [15], the process of extracting SIFT features involves four key steps: (1) selecting scale-space peaks, (2) localizing the keypoints, (3) assigning orientations, and (4) creating keypoint descriptors. Initially, interest points are detected by examining the image at various locations and scales. This is achieved by building a Gaussian pyramid and locating local peaks (known as keypoints) in a sequence of difference-of-Gaussian (DoG) images. In the subsequent step, potential keypoints are pinpointed with sub-pixel precision and discarded if deemed unreliable. Then defines the orientations per each point according to its local surrounding area. The assigned orientation(s), scale and local SIFT are also key point detectors due to their ability to detect the extreme points in different scale spaces, then calculate the directions around each key point according to the attributes of the facial features, to finally calculate the local feature descriptors.

Patch-SIFT [16] is a SIFT extension example proposed to extract key point features in small patches. Mian et al. [17] identified the key points and calculated the tensor features around them, then used the features fusion at the decision level to obtain the SIFT key points from both the depth image and 3D mesh. A 3D partial face matching is proposed by Berretti et al. [18] by extracting the features from specific spherical regions of different radiuses around each detected point. The author showed that the details in small regions of the face can be better captured while using a small spherical region radius. Although implementing the SIFT feature detector on 3D images that have depth information showed its sensitivity to significant pose variations, SIFT key points have been extracted directly from 3D mesh (Mesh-SIFT) to gain more robustness. Smeets et al. [19] used slant angles and histograms of shape index, in addition to Gaussian weight to construct the feature vector. Smeets's results showed the Mish-SIFT robustness to the change in face expression, occluded and missing data, and outliers data with high recognition rates on both the FRGC v2 and the Bosphorus datasets. Lin et al. [20] calculated the third-order feature tensor per each detected salient point, and implemented the Voronoi diagram subdivision to augment the data to examine the intra and extra-face similarity using the pre-trained ResNet network. This similarity-based approach detected the salient points on the 3D mesh with an accuracy of 99.71% and 96.2% on the Bosphorus and BU3DFE databases respectively. Berretti et al. [21] used SIFT to highlight the feature points of depth images and extract specific curves combining the detected points. The authors posed side scans from the UND dataset and frontal faces from FRGC v2.0 to examine the SIFT robustness against the pose variation and detect the identical faces between these two datasets that provided multiple pose scans for each individual. Zhang et al. [22] used the shape index extrema as interesting points on the face surface. The author proposed a scale change procedure to eliminate a set of unstable key points initially, then extracted the features.

Darom and Keller [23] proposed a spin-based SIFT extension on the 3D mesh by measuring the vicinity of each detected keypoint using the depth map. The author enhanced the face rotation invariance by calculating the dominant angle using PCA. The resulting local features were experimentally robust to the scale variation and suitable for local matching of mesh segments. Additionally, proposed the local depth SIFT

(LD-SIFT) to allow extending the SIFT extraction efficiently to the 3D meshes, and augment an angle estimation scheme to make the LD-SIFT a rotation invariant descriptor. Shi et al. [24] concluded that the SIFT-extracted feature points can be translated, and scaled and can show a rotation invariance too. The author also noted that the number and positions of these detected points are typically random, and don't ensure that key points can be found in all face regions. Thus, the author proposed the ISIFT algorithm to enhance the robustness against facial expression variation.

1.1.2 Histogram of oriented gradients (HOG)

Using Histogram of Oriented Gradients (HOG) as a feature extractor on 3D facial meshes for face classification and recognition has shown promising results in various studies. HOG is a popular feature description commonly implemented in 2D models, but its adaptation to 3D facial meshes has demonstrated its potential in capturing distinctive facial characteristics. The advantages of using HOG in 3D facial mesh classification and recognition such as; its robustness to Pose Variations, and its efficient and simple implementation compared to other feature extraction methods. Thus, HOG is a suitable extraction for real-time and large-scale face classification and recognition tasks. Additionally, HOG allows focusing on capturing local shape information and texture patterns in a facial region, which makes it effective in highlighting distinctive features, such as edges, corners, and facial landmarks, which are crucial for accurate face classification and recognition. However, several challenges of using HOG for 3D facial mesh classification and recognition are also highlighted in academia such as vulnerability to occlusions due to its dependency on local gradient information, and its limited ability to capture global features due to its focus on capturing the local information.

1.1.3 Geodesic distance

The ability to extract facial curvature information from 3D imagery makes it easier to analyze the facial surface. In 3D face recognition and identification, image segmentation, and landmark localization, curvature information became a crucial part. The fact that the principle facial curvatures remain unchanged as the surface rotates or translates makes them extremely robust for feature space manipulation such as feature detection and extraction. Additionally, it has been demonstrated that principle curvature directions considerably improve identification performance because they have a strong potential to provide discriminative characteristics. Geodesic distance vector is a curve-based descriptor used as a shape descriptor that is invariant under the face shape isometric deformations. A geodesic distance matrix is a symmetric matrix whose elements (gi;j) are the geodesic distance between the points i and j. The GD vector associated with the points or landmarks will then be used as the shape descriptor for face classification and identification purposes. The geodesic path represents the shortest line between the reference point (or reference vertex on 3D mesh) which is usually considered as the nose tip, and another endpoint on the face surface. The algorithm visits all other endpoints and constructs a matrix of these calculated distances from the initial point. Then, it calculates the least distance from the start point to all other points. The resulting geodesic features matrix usually has a high dimensionality and needs to be reduced in size to allow easier analysis and machine learning phase in the next steps of face recognition and classification systems. With 3D Mesh facial form, Geodesic is usually preceded by meshes' preprocessing to clean and remove noise, outliers, and artifacts. In addition to some mesh smoothing, hole filling, and vertex alignment to ensure the quality of the data. Then, the Geodesic descriptor calculates the geodesic distance for each vertex on the face mesh using algorithms such as Dijkstra's algorithm or Fast Marching Method due to their ability to calculate the shortest path between specific pair of points of the facial mesh. Once the geodesic distance for each vertex is calculated, the facial curvature features can be extracted. A common approach to extracting the geodesic features is local features extraction in which the algorithm computes local geodesic descriptors around each vertex. This can be done by considering the geodesic distances of neighboring vertices within a certain radius. These descriptors capture local shape information and can be used for tasks like facial expression recognition or landmark detection. Considering the isometric deformation nature of human facial parts while expression is changed, based on the various surface components used in face matching, Zhao et al. [25] separated the methods into three categories: (1) Iso-metric surface method, which matches the face with the complete isometric surface. (2) The iso-geodesic stripes method, which matches faces using a set of uniformly widened iso-geodesic stripes. 3) The geodesic-based approach, also known as the iso-geodesic method, uses the geodesic curves as the facial feature for face matching. The 3D face representation such as triangle mesh allows the description of each point with its depth value over the mesh surface and the local variation on the facial surface has been extracted by selecting the rigid regions of the face to identify the individuals. This representation allows calculating multiple geometric attributes such as geodesic distances, facial shape angles, and Gaussian curvatures. Drira et al. [26] investigated the 3D geometric analysis of human nose shape under multiple expressions of faces from the FRGC dataset and measured the geodesic distances between nasal surface pairs (a set of their radial nasal curves are transformed into the shape space for comparing the faces and calculating their similarity level). The author showed that the geodesic-based recognition outperforms the baseline ICP algorithm while multiple facial expression was used. Kurtek and Drira [27] developed a framework on the BU-3DFE dataset for analyzing the elastic form of hemispherical surfaces and explored the impact of parameterization-invariant, and elastic Riemannian metrics. The scans have been registered, deformed, and compared in this work, in addition to calculating the covariance, the PCA components, and analyzing the face symmetry. Lee and Krim [28] measured the shortest distances from a reference to points lying on a specific surrounding curve using Geodesic. The author then measured the similarity after extracting the deformed curves and used the Fourier transform to reduce the dimensionality of the feature space. The author measured the geodesic circle via the closed curve around the nose tip and tracked the boundary of the upper and lower lips. Berretti et al. [29] partitioned the facial surface into eight iso-geodesic stripes, and arcs between pairs of iso-geodesic stripes, and calculated their weight to obtain the relative spatial displacement.

1.2 Human Race identification based on face morphology

Human Face explicitly provides information for evaluating implicit data such as age, race, and gender. Face recognition methods generally can be categorized into local and holistic

methods. Although many works are conducted using the holistic approach due to its simplicity, the local approach can be a more reliable and accurate recognition technique in applications focusing on specific parts of the human face.

The human race classification emerged as one of the most prominent face classification directions due to its omni-relevance with many social cognitive and specific perceptual tasks (emotion, belief, diseases, and more). The term "Race" is used to reflect the person's physical appearance or characteristics regarding the main human categories of the population around the world. The population is categorized into seven large distinct populations or racial groups that are most commonly encountered and accepted racial groups. These seven main classes were the most accepted they represent more than 95% of the population.
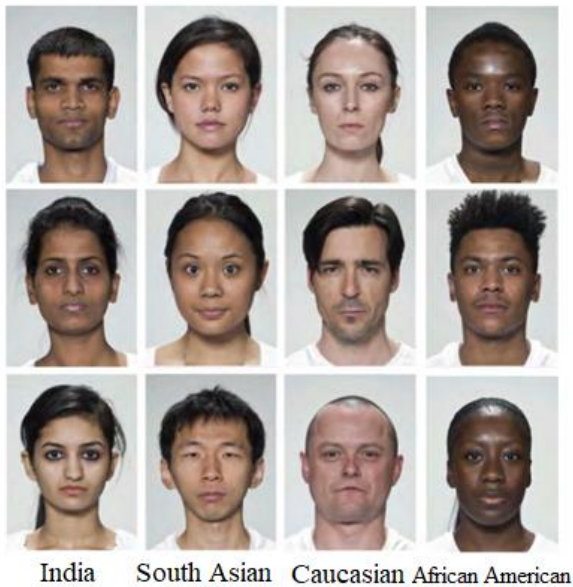

Figure 2. Several faces from large distinct groups [30]

A commonly accepted racial categorization in Fu et al. [30] includes East Asian, African/African American, Native American/American Indian, Caucasian, Pacific Islander, Hispanic/Latino, and Asian Indian (See Figure 2). Literature shows that facial features with different relative positions have a close relationship with human ethnicity and facial landmarks and key points on specific positions of the face can reflect the positions of important descriptive areas and facial features that usually impact the resulting class of individuals. While automatic ethnicity classification is a useful starting point for facial analysis applications, most ethnicity classification techniques may necessitate a laborious feature extraction and model training procedure. Initially, psychologists approached the study of human face ethnicity and gender recognition from the standpoint of cognitive science. At that time, the characteristics of a person's face and the data they extracted, such as age, gender, and ethnicity, were referred to as soft biometric features. The term 'ethnicity' is identified as a soft biometric that describes belonging the individual to a particular social group having a common nationality, cultural tradition, or geographical factors. Thus, a group of humans contains people who have similar attributes such as language, religion, and nationality. The race classification might also categorize the population into smaller sub-ethnic groups such as; east Asian, Chinese, Korean, Japanese, and Indian. In addition to traditionally reported issues of face classification, many Issues and challenges are also reported in race recognition and classification and considered as challenging specific face trait recognition research areas. Issues such as; Intra-race, inter-race, and mixed-race are discussed in several works, and a few attempts are implemented by researchers to analyze the face variation across sub-ethnic groups (Turkish, Japanese, Spanish, …). Inter-race challenges of people belonging to two racial groups with a shared skin tone and mixed race challenges of People with average looks (such as; Native American, and Mexican-American are examples of race classification challenging issues [30]. However, the task of race classification might be partially enhanced by considering facial part information extraction and training the classifier on these extracted features and parts [31]. Gen Association with Regionalized Facial Features is reported by Richmond et al. [32] as described in Figure 3.
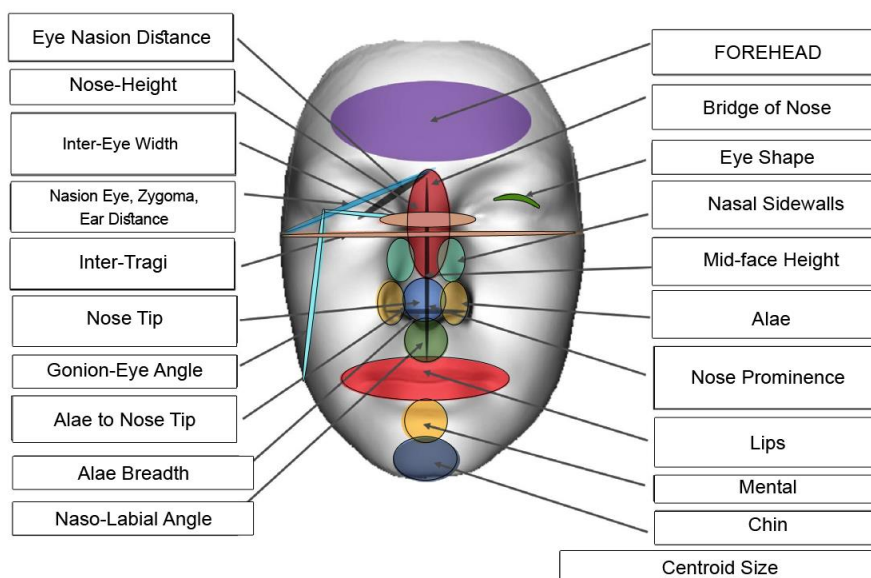

Figure 3. Facial features reported by Richmond

Because there aren't many large-scale 3D face datasets it was once thought to be extremely difficult to train discriminative deep features for 3D face recognition in comparison to 2D face datasets. To address this issue scientists attempted to modify a small number of 3D face datasets for 3D surface matching using the already-trained 2D face model. The proposed networks specifically designed for 3D face recognition and the recent capabilities of deep learning CNN encourage researchers to train massive volumes of 3D facial scans of identities more quickly and accurately. The recently proposed algorithms demonstrated excellent accuracy in both closed and open-world recognition scenarios based on the 3D datasets. The global approach component-based approach and hybrid approach - the three categories of feature extraction techniques for face recognition - are encountered once more in three-dimensional methods. Compared to 2D approaches, 3D race classification can provide better accuracy due to its robustness to the changing texture, skin tone, and appearance, in addition to its flexibility to correct the head pose and avoid the impacts of obstacles via 3D rotation.

1.3 Deep learning in 3D facial traits recognition

Many computer vision applications including 3D facial race recognition have been transformed by deep learning. Specifically useful for complex tasks like race classification from 3D facial data this technique uses multiple-layer neural networks to automatically learn relevant features from raw data. By allowing the model to learn hierarchical representations straight from the raw data, deep learning provides a considerable enhancement. The first step of the procedure could be data preprocessing which involves normalizing and formatting 3D face scans into the appropriate formats. 3D convolutional networks (3DCNNs) and CNNs approaches have been adapted to exploit the geometrical information found in 3D data in forms that hold the geometric information. The multi-layered models that extract features at various levels of abstraction gradually allow learning to distinguish between subtle yet discriminative features that may be difficult to identify using conventional techniques. This involves recording variances in the depth curvature and contours of the face that may be specific to various racial groups. The model becomes adept at recognizing patterns that are important for race classification, even when these patterns are not explicitly defined. Deep learning has been more successful in 3D facial race recognition thanks to the availability of large-scale datasets with a variety of 3D facial scans from different races. With the help of these datasets, models may be trained to identify a person's race from their facial traits and to generalize effectively to new data. Though incredibly successful, deep learning-based 3D facial race recognition is not without its difficulties. While 2D image datasets are more abundant annotated 3D facial datasets are still scarce. Thus, deep learning has allowed models to learn directly from raw data greatly improving the accuracy and dependability of 3D facial race recognition. This approach has proven effective in capturing intricate facial features that are critical for distinguishing races, and it continues to drive progress in the field of facial recognition technology. AlBdairi et al. [33] developed a DCNN-based 2D race recognition approach to determine the human race, the author used high-performance devices to build a a face recognition system and proposed a technique called field-programmable gate arrays (FPGAs). The periocular area is analyzed for race and gender

by Khellat-Kihel et al. [34] who proved that deep learning techniques in race prediction still require a large amount of labeled data, and accordingly proposed a DCNN-based predictor to solve several specific biometrics issues on the periocular part of the human faces. Periocular regions' features are extracted from 2D faces using different pre-trained architectures such as Alex-net and ResNet-50. Fan et al. [35] compared five forms of 3D data including depth images, normal maps, point clouds, DAE, and HHA (obtained from FRGC v.2 and BU-3DFE datasets). The author proposed a data augmentation approach by synthesizing multi-view 3D faces using a DAE-based deep learning model. Xia et al. [36] proposed a correlation-based 3D facial shape model, and highlighted salient features extracted from three traits, and then used the RF algorithm for ethnicity estimation on the FRGC-v2 dataset. The study comes to a conclusion demonstrating ethnicity is considerably correlated in the 3D representation of the human face.

1.4 3D face analysis for identifying the race, a literature review

Lu et al. [37] used both registered range images (describe the surface shape) and intensity images for race classification (See Figure 4). The author used an SVM classifier to classify the faces and achieved 96.8% accuracy. A mixture of two frontal 3D face databases is used for evaluating the proposed schemes. Toderici et al. [38] suggested retrieving subjects from deformed 3D meshes (only global shape information) through a metric function (Harr wavelet and CW-SSIM, or structure similarity).

The author proposes four classifiers: wavelet-based learning, multi-dimensional scaling (MDS), kernelized KNN, and KNN. The FRGC v2.0 dataset was used to test these classifiers, and they produced mean accuracy values of about 99 percent. Toderici et al. [38] showed that the estimated probabilities can serve as feature importance scores, and used Gauss-Markov posterior marginals to classify the Asian and White people from the BU-3DFE and FRGC v1 tables. To prove his idea, the author compared his achieved accuracy levels with the levels achieved by Ocegueda et al. [39]. Zhong et al. [40] combined the use of 3D shape and the Boosted Local Texture) to enhance the race-based classification. To extract the local texture, the author used the Oriented Gradient Maps (OGMs) to highlight ethnicity-related local texture and used Adaboost to add an associated weight for individuals. Experiments are carried out on the FRGC-v2 dataset to categorize Asians and non-Asians with an accuracy of 98.3%.
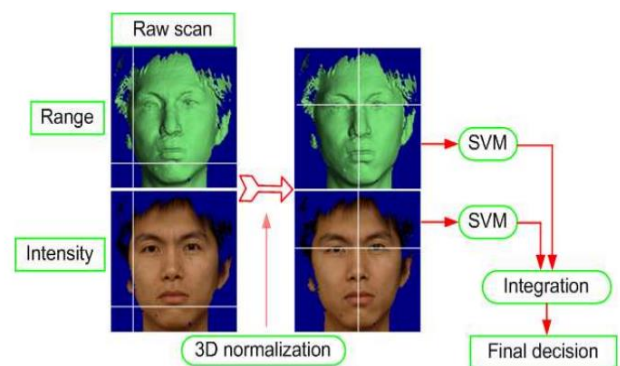

Figure 4. Integration scheme of both range and intensity [35]

Table 2. Several existing face analysis works for race classification

Author	Features	Classifier	Dataset	3D Representation	Races	Findings and Contribution
Lu et al. [37]	Global Intensity	SVM	UND and MSU	Registered range and intensity images	Asian, Non-Asian	Range image modality is useful in race recognition.
Ocegueda et al. [39]	Harr Wavelet, and Similarity Measurements	MDS, k-KNN, and wavelet.	FRGC 2.0	Retrieved meshes	Eastern, Western, Asian, Non-Asian	Gender and race can be used in mesh retrieval.
Toderici et al. [38]	Probabilities as Features	The linear classifier, LIBLINEAR	FRGC and BU-3DFE	3D meshes	European and Asian	The mouth area is important in race recognition.
Zhong et al. [40]	Gabor Filters to Build a Learned Visual Codebook (LVC)	K-means clustering	FRGC v2	Range images	Eastern, western	Modeled the ethnicity categorization as a fuzzy problem and achieved a reasonable membership degree.
Ding et al. [41]	Combination of Holistic Shape and Texture using OGMs, Adaboost and Boosting	Decision Tree	FRGC v2, and BU-3DFE	Range images	Asian, Non-Asian	Used OGMs to highlight local geometry, and observed that the eyes and nose areas are more discriminative.
Lv et al. [42]	Iso-Geodesic Measurements Between Specific Landmarks on Nose Region	Clustering according the nose measurements.	FRGC2.0 and Bosphorus3D	Range images	Asian and White	the nose is the most discriminative region for race.
Sovizi et al. [43]	Geodesic Distance	the similarity of the geodesic distance vectors of the 3D faces	SHREC08 database	3D mesh	Caucasian	proposed two landmarking methods (ICP and topological) and explore the effect GD vector size and landmark positions.

Zhong et al. [40] computed the visual codes for the eastern and western individuals using Gabor features and the learned visual codebook (LVC) method. Next, the merging and mapping distances were measured (by calculating the max distance function). The author suggested a "membership function" in the second level to determine the membership degree of the western and eastern groups for the faces taken from the FGRC v2 dataset.

Berretti et al. [29] modeled the three-dimensional faces of multiple stripes centered at the tip of the nose, in which most of the points of each stripe remain within the same stripe even when the facial expression changes. Nine stripes, each measuring one centimeter in width, were employed by the author to distinguish between various individuals. The author additionally divided each stripe into three sections—lower (L), upper-left (UL), and upper-right (UR)—concerning the coordinates of the nose tip to account for the expression deformation. This work implemented the suggested measuring on the FRGC-v2 dataset (measuring the spatial displacement between iso-geodesic stripes) and the SHREC 2008 dataset (measuring the inter-stripe and intra-stripe 3DWW) distances using the 3D Weighted Walkthroughs (3DWW) approach. Several existing related works are summarized in (Table 2) with their findings and the supported races.

1.5 Problem statement and the contribution

Although existing race classification works tried different approaches to extract the face features, insufficient research efforts have been proposed with a focus on analyzing the 3D mesh form for identifying the human race. The size limitation of the publicly available 3D facial datasets in a mesh form, the unfair racial distribution, capturing cost, and privacy aspects, are factors that negatively impacted this research direction. The data limitation forced a need for proposing efficient extractions of geometric and shape features from the well-known mesh representation, and for building accurate race classifiers that can cope with the limited data training

challenge.

The context of this work is to analyze the shape of the 3D facial scans of real individuals after extracting four types of 3D facial features implemented on different facial areas. It focuses on the mesh representation of 3D faces due to its shape details and geometrical surface information, and due to the limitations of traditional 3D representation such as depth images and z-coordinate approximations. The texture analysis, the 3D representations that had limitations in flexibility and in describing the surface geometry, and the reconstructed synthesis faces, are out of the scope of this work. The 3D face analysis of this work is extended to compare the used feature extractions for race classification due to its challenging fine and salient variation in facial details. This work is also aided by several experimental and knowledge-based secondary phases of preprocessing and 3D mesh registration that have a direct impact on the extraction and classification results. In addition, Deep Learning is used in this work for automatically landmarking and extracting the distance-based features according to specific interest key points on the facial mesh. The results of this work highlight suitable feature extraction for race classification by several important extractions, these comparing results also highlight the facial areas and traits that could highly impact identifying the human race by the face shape details.

2. THE PROPOSED METHODOLOGY FOR 3D FACIAL RACE RECOGNITION

In this work, five distinct phases are meticulously executed. The initial phase was characterized by an exhaustive survey of existing 3D facial datasets, conducted to identify the most appropriate dataset in mesh format, encompassing individuals of various racial populations. Headspace, a dataset that comes precisely aligned with the core goal is used in this work. Subsequently, in the second phase, the preprocessing of the selected 3D facial data (.obj files in triangular mesh form) took

center stage. This phase includes the precise cropping of frontal facial parts from each head complete head mesh as a critical step in isolating the target features.

Moreover, we undertook a rigorous cleaning and filtering process to eliminate the noise and ensure the purity and oriented triangular facets of the 3D faces. These steps of cropping and filtering were complemented by 3D mesh registration techniques, which harmonized the vertices and aligned the meshes, standardizing the data for the subsequent phase of feature extraction. The third phase places the feature extraction process at the core of this proposed work, in which we harnessed a spectrum of robust descriptors, including HOG, PCA, SIFT, and geodesic distance. These descriptors were thoughtfully employed to effectively capture and represent intricate details and attributes of the 3D facial meshes. They provided a multidimensional perspective, allowing for well understanding of the dataset's intricacies. In the fourth phase, our attention turned towards features post-processing, a critical endeavor in which we made use of various statistical techniques such as; the PCA, the median, the average, and the standard deviation, which are applied to enhance describing the extracted features in a reduced size.
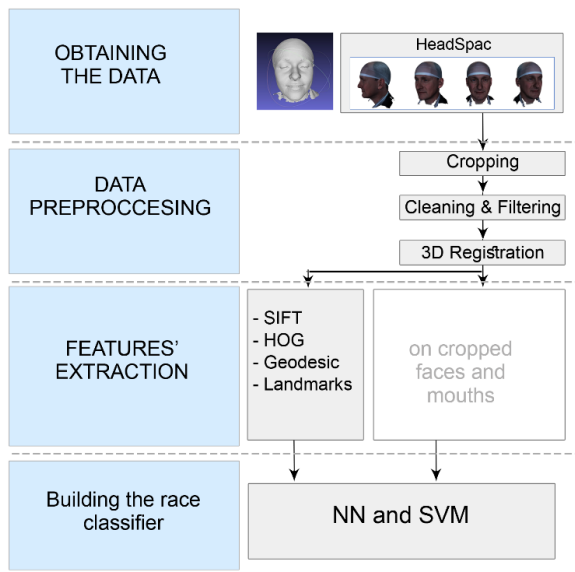

Figure 5. The proposed phases of 3D facial race classification

In the final phase, machine learning-based classification is built, and trained on the Headspace dataset, making use of the provided race labels to categorize individuals into their respective racial groups. The performance of this classifier according to its accuracy and validation measures is explored to prove its effectiveness in identifying the human race based on their 3D facial parts' geometry. The achieved high-accuracy results offer invaluable insights into the system's potential for future real-world applications and race-based implementation of computer vision technologies. The proposed phases of this work are described in Figure 5.

2.1 Dataset and preprocessing

In this work, the Headspace Liverpool-York Headspace Model (LYHM) [44] a multiple-race dataset is used to investigate the impact of the face trait geometrical shape on identifying the human race without using more complex

texture information of the facial regions. The Headspace dataset includes 3D models of human heads (shape and texture information) introduced by a collaboration between the Craniofacial Unit at Liverpool's Alder Hey Children's Hospital and the Computer Science Department of York University. Using a 3dMD five-camera system, a high-resolution texture image and a 3D triangular surface consisting of approximately 180K vertices connected into 360K triangles were generated for every subject [45]. This dataset consists of full human head 3D images (OBJs) of 1519 subjects (~1212 subjects after exclusion), in addition to their texture information that is stored as bitmap (BMP) files. The subjects are wearing tight-fitting latex caps to reduce the effect of hairstyles and show the shape of the cranium (See Figure 6).

Figure 6. The latex cap used in Headspace data

Headspace data allows the possibility of multiple images per subject and uses Five five-digit strings as a subject identifier and twelve-digit string as an image identifier. The authors depend on a balanced number of males and females and further exclude several 3D faces due to poor conditions such as; improper fitting of latex cap, hair bulge under it, unintended captured noise, or missing partial areas of the 3D scan. The dataset is provided with subject-based and capture-based metadata. Each face of the data is described using 15 attributes that are mentioned inside a separate text file. Gender, stated ethnic group, age, eye and hair color, beard and mustache descriptors (none, low, medium, high), and a spectacles flag are among the subject information. A quality descriptor (free text, such as "data spike right upper lip"), a hair bulge flag (hair bulge under latex cap distorting the apparent cranial shape), a cap artifact flag (due to poor fitting, the cap has a ridge at its apex), a cranial hole flag (a missing part in the data scan at the cranium), and an under chin hole flag (missing part under chin) are all included in the capture information. The Vertices and Faces of each 3D facial mesh are described with their corresponding x,y, and z coordinates in addition to the norm information of each vertex. The race label is mentioned in a text file attached with each facial scan, the racial groups of this data include (1012 scans of White-British, 48 of White-European, 43 of White-Irish, 22 of White-Welsh, 15 Chinese, and smaller numbers of several other sub-races. The training set used in this work contains three races (White, Asian, and African) limited by the least number of them.
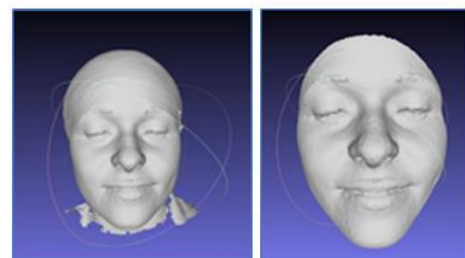

Figure 7. A Headspace scan and its cropped face area

The used faces are cropped using "MeshEditor", a public GitHub code by Harry Matthews to crop the frontal Face Region from the 3D facial mesh data. The extracted frontal face region contains the relevant facial features and eliminates any irrelevant information, such as hair, neck, or ears, which could otherwise interfere with the classification process. The meshes' poses are corrected, and the frontal view of each 3D face is extracted (See Figure 7) to be ready for the next preprocessing step.

## 2.2 3D registration to map the facial meshes

Then, to align and fit each 3D facial mesh to other meshes (to align the targets to a template), a 3D registration step is implemented to find the best possible transformation that maps one mesh onto the other, and ensures that corresponding points on both meshes are aligned. This alignment is essential for various applications, including 3D face recognition and classification, facial expression analysis, and facial animation. In this phase, we implement the deformation procedure proposed by Amberg et al. [46] to extend the ICP over nonrigid registration while keeping its original advantages. This framework could allow different regularizations due to its adjustable parameter (decreasing stiffness weights used in deforming the template towards the target gradually.
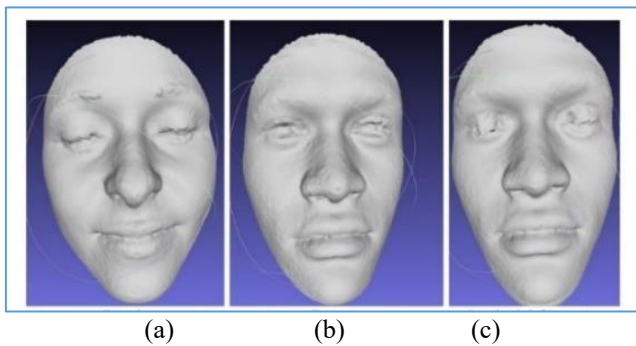
Figure 8. The 3D registration on Headspace data
(a)Template, (b)Target, (c)The result

ICP is then implemented in this approach to find the optimal deformation among the others created iteratively via calculation of the correspondences by a nearest-point search. This implementation works well over the 3D facial meshes of the Head Space dataset, giving each vertex an affine transformation and minimizing the difference in the transformation of neighboring vertices. especially while choosing a template of an approximate similar number of vertices with the average number of the targeted meshes. A template, a sample target, and the resulting deformation are shown in Figure 8, the human perception comparison shows clearly how this algorithm kept the details of the target mesh and modified the number of vertices to be ready for the next steps of feature extraction and classification. Implementing this registration procedure while using a template of a vertice number that extensively differs from that of targets is also examined. However, the results show that it yields higher distorted facial traits due to a higher difference in vertices number between the template and the target. Another proposed registration procedure in the study by Audenaert et al. [47] is also investigated in this work and showed a degraded quality and distorted facial traits on Head Space data as shown in Figure 9.
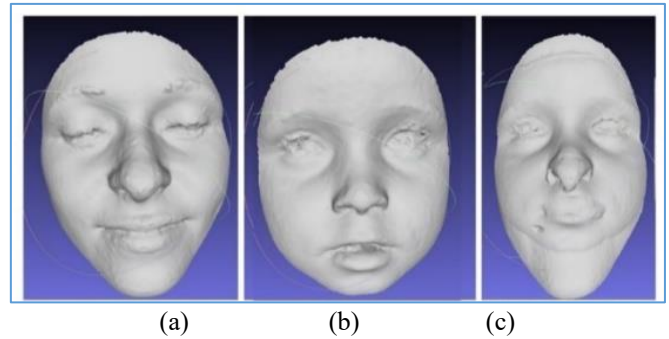
Figure 9. The impact of inappropriate registration

## 2.3 Extracting the descriptive features from 3D faces

The extractions in this work are 3D feature extractions extended from several well-known approaches proposed in the literature. due to their essential advantages mentioned in the literature, in addition to the experimental enhancement in classification accuracy achieved in this work, we extend these feature extractions to 3D mesh scans of real facial data of Headspace. The LD-SIFT is a SIFT extension specified for mesh representation with high robustness to pose variations, uniform change in scaling and orientation, and partial noise. A HOG extension called the 3D Voxel HOG algorithm originally implemented on furniture and handcraft edges, has been extended to facial data due to its robustness to pose variations, and discriminating feature representation. Aided by deep learning automatic landmarking, Geodesic distance is used as the third feature extraction due to its ability to capture the curvature information and 3D dimension of the mesh areas.

After the initial steps of preprocessing such as denoising and cropping to remove the irrelevant details, and registration to align the vertices and unify their number, the important and descriptive features need to be extracted and represented efficiently. Features such as; edges, corners, Gabor features, HOG features are detected, extracted, and saved in numerical vectors or matrices. In addition, the interest point (i.e., landmarking, keypoints) is determined manually or automatically using keypoints detectors such as; ICP, SIFT, and other types to facilitate extracting the local features. In this work, Scale Invariant Feature Transform (SIFT) is used to extract local features due to its pose variations and scaling robustness, and its capability of handling the partially occluded facial parts and uncontrolled illumination conditions Thus, SIFT can cope with the traditional limitation of holistic-based face representations by selecting features that are not affected by scaling, noise, and rotation. The SIFT allows obtaining facial features in the small regions of the human face. In this phase, SIFT is Implemented directly on 3D mesh Head Space data, and the procedure proposed by Darom and Keller [48] is borrowed to obtain the local features of each 3D Face. The author of implemented SIFT on Google 3D Warehouse data that is consisted of different categories of general images. We borrowed this procedure to implement of real 3D facial meshes of Headspace, the resulting numerical array size proportional to the vertices' resolution of the 3D mesh representation. Experimentally, a Headspace 3D facial mesh of (18215) vertices yielded approximately a (1×1104) size features array. The extracted features' vectors are then reduced in dimensionality using statistical representations like PCA to allow efficient machine learning in the next training and classification steps. The second scenario of feature extraction

in this work is done using the HOG feature extractor on Headspace data to obtain the pros of its robustness to pose variations, simplicity in computation, its ability to capture the local shape information of the facial region and traits, and its low dimensionality of the resulting feature space. The effectiveness of implementing HOG as a feature extractor is affected by the feature representation nature and the chosen type of parameters, such as the size of the histogram cells and the number of orientation bins. Thus, selecting appropriate parameters is crucial to achieving optimal performance while using HOG in 3D face analysis. For each vertex in the mesh described as v(x, y, z), the gradient information is calculated to capture the information about local shape variations. Then, the 3D facial mesh is divided into fixed-size small cells (local regions) and the gradients of the neighboring vertices are analyzed to compute their orientations. The orientation is then quantized into a fixed number of bins, typically by dividing the 360 degrees of rotation into equal intervals. Each cells histogram as well as each bins gradient orientations are computed. The local gradient orientation distribution within the cell is well-represented by this histogram. Subsequently adjacent cells are categorized into blocks and the histograms within each block are combined to generate a combined histogram that depicts the local structure. The final feature vector is formed by concatenating all the histograms from different blocks or cells to summarize the spatial distribution of local gradients throughout the 3D facial mesh. Mathematically, the HOG feature extraction can be represented as follows: Let (M) be the 3D facial mesh with (V) vertices, and for each vertex ($v_i \in M$), let G_i represent the gradient at vertex v_i. Divide the mesh into cells C_j, and within each cell, calculate the histogram of gradients H_j with K orientation bins. The concatenated histograms from all cells and blocks from the final feature vector (F). The HOG feature extraction can be expressed as ($F=[H_1, H_2, ..., H_N]$), where (N) is the total number of histograms from all cells and blocks. The resulting feature vector (F) can then be used as input for various machine learning algorithms in 3D facial mesh classification, recognition, or other related tasks. In this work, we extend the HOG algorithm proposed by Dupre et al. [49] to be implemented as a feature extractor on Headspace data using MATLAB R2022b. This 3D Voxel HOG descriptor is made especially to be appropriate for local feature recognition while taking the density of an object into account. The normalized combination of gradient vectors from a specified number of pixels is used by the conventional HOG. However, this 3D Voxel HOG uses voxels and 2D histograms for adapting the 3D models. It is initiated by separating the voxel volume into feature spaces (*f*) containing some cubic 3D cells (c), which in turn is contain voxels (*v*). A filter mask [-1; 0; 1] is applied to the neighboring voxels of each selected voxel within a cell (in all three dimensions) to calculate the gradient vector (*g*). After determining the gradient vector's magnitude, particular angles are used to express the vector's orientation. Furthermore, each voxel (w) has a defined weight that is used to scale its contribution to the 2D histogram of that cell. The voxels inside each cell are then binned into a 2D histogram (h) based on their angles once these values have been determined. The 3D VHOG allows obtaining the faces of a mesh in addition to the information of the area within it, which decreases the impact of artifacts. In addition, the density of an object is also considered by the 3DVHOG to support accurate medical imaging, since it builds one 2D histogram (visualized in 3D) per cell. This histogram allows for visualization of the empty

and full gap within the 3D mesh. In summary, using HOG as a feature extractor on 3D facial meshes for face classification and recognition offers significant advantages, such as robustness to pose variations, efficiency, and discriminative feature representation.

For more investigation into the relation between facial features and human ethnicity, the Geodesic distance is also used in the third scenario as a feature extraction to calculate the distances between the interest points on the facial mesh. Specific landmarks have been assigned to the HeadSpace 3D scans using a deep learning-based pre-trained network [49] on BU-3DFE and DTU datasets. A Specific Matlab code is designed to map the landmarks indices with the vertices coordinates of the 3D mesh. The resulting landmarks on a HeadSpace sample are shown in Figure 10.
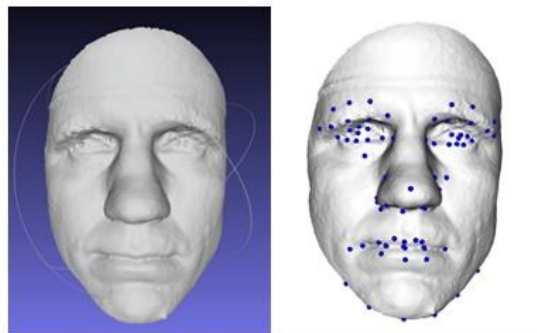

Figure 10. Landmarking the HeadSpace data

The Geodesic Distance allows extracting facial curvature information from 3D imagery efficiently while the principle facial curvatures remain unchanged as the surface rotates, translates, and changes in scales. Each curve-based descriptor represents the minimum Geodesic distance that separates the facial points *i* and *j* (the reference vertex and the end vertex on the facial surface) such as Dijkstra's algorithm or Fast Marching Method. A sample of extracting the Geodesic features globally from a Headspace facial mesh of (93829) vertex using the Fast Marching algorithm is shown in Figure 11. Similarly to the *Dijkstra algorithm,* Fast Marching [50] calculates the shortest paths on graphs. Using a gradient descent of the distance function D. These computations are done in a "mex" file to fasten the distance calculations. A distance map from multiple start points using an arbitrary isotropic metric is then created to compute geodesic paths from any point that joins the closest starting point. The calculated shortest geodesic distances finally allow extracting the information of the facial curvature (extracting the geodesic feature).

A common approach to extracting the geodesic features is to compute local geodesic descriptors around the nose tip. The resulting geodesic features matrix then is represented with a suitable format for further analysis and next machine learning tasks according to the specific required task and the nature of the used data. The resulting geodesic features matrix usually has a high dimensionality and needs to be reduced in size to allow easier analysis and machine learning phase in the next steps of face recognition and classification systems.

Additionally, a local feature experiment is done to extract the SIFT features from a cropped local part of each facial scan. This knowledge-based selection of the mouth area as a local part was according to the conclusions of several existing

related works. The classifier is then trained on these Mouth-SIFT features to predict the individual ethnicity.
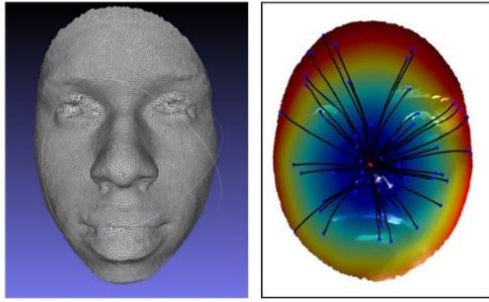
Figure 11. The geodesic paths on a facial scan

2.4 Building the race classifier

A neural network model often is trained using backpropagation, and optimization methods like stochastic gradient descent (SGD) or its variations are used in backpropagation to train neural networks (NN). Iterative training of the weights and biases enables the network to learn how to classify faces. Several factors influence the accuracy of neural network classification, such as the network architecture including the number of neurons per layer, and the number of layers, the activation functions selected, learning rate, batch size, and regularization strategies like L2 and dropout regularization. Five-fold cross-validation is implemented to evaluate the NN race classifier performance.

On the other hand, to train the SVM classifier, the best hyperplanes for classifying the data are identified. This is accomplished by using a kernel function for non-linear separation to optimize the margin between the classes. The regularization parameter, kernel-specific parameters, and the kernel function all affect the classification accuracy. The SVM finds the hyperplane that can maximize the distance between any class and the hyperplane, given multiple points that belong to a pair of distinct classes. This hyperplane allows separating the largest possible fractions of points of each class on the same separate side.

Because it is reliable, accurate, and efficient even with limited training data, supervised learning in this work employs SVM with a linear kernel. After extracting the feature descriptors from 3D facial meshes, the features have been classified using SVM into multiple classes (racial groups used in this work). The classifier performance has been evaluated using 5-fold cross-validation. The observations have been divided into k subsets. Every time, a test set is created from one of the k subsets, and a training set is created by combining the remaining subsets. The average accuracy throughout the k trials is then determined.

The SVM maps the input sample to a high-dimensional feature space in trials for locating the optimal hyperplanes and decreasing the classification error. The largest margin between the two classes is the best hyperplane for the SVM. The margin means the maximal width of the slab parallel to the hyperplanes that have no interior data points and the support vectors are the width constraint of that margin. The classifiers have trained over numerical form features resulting from using different feature extraction such as; SIFT, HOG, and Geodesic. The numerical form of features also decreases the privacy and ethical concerns within intelligent systems, particularly when data is collected without informed consent which is more suitable for human data such as explicit facial details and

implicit ethnicity information, since such information might raise ethical and privacy issues.

These representations are then organized and combined into abstract sets of facial information. In the race classification task, the network is architecture to have an input layer that matches the dimensionality of the numeric features and an output layer with units corresponding to the number of classes or race categories aimed to be classified. The intermediate layers (also known as hidden layers), can be customized based on the complexity of the required tasks and the considered facial details. Once the network has been designed, training involves feeding the feature data into the network and adjusting the model's internal parameters (weights and biases) to minimize the prediction errors. The used network is defined with a specific input layer and a specific number of features according to the fed input set. This network's input layer is likewise set up to use Z-score normalization to normalize the data. Then a fully linked layer with an output size of 50 is added, a batch normalization layer, and a ReLU layer in that order. Another completely connected layer is set up with an output size equal to the number of classes (the racial groups utilized in each experiment) for the final classification. MATLAB R2022b application is used to implement deep learning functions and the network is trained to recognize patterns of the numerical features that are associated with different race categories. In order to train a network using categorical features, the names of all the categorical input variables are specified in a string array, and the categorical features are transformed to numeric by converting the categorical predictors to categorical using the "Convertvars" function. Then, a specific function is used to loop over the categorical input variables by which for each variable; the categorical values are converted to one-hot encoded vectors using the "one-hot-encode" function, and then the one-hot vectors are added to the table using another function. To ensure that the model is generalized well, the features set is divided into three parts: a training set of (70%), a validation set of (15%), and a testing set of (15%). The training set is used to update the model's parameters during training. Every epoch the data is shuffled and mini-batches of size 16 are used to train the network. Next by providing validation data the network accuracy is tracked during training. During training the validation set is used to monitor the models performance and adjust hyperparameters. The testing set evaluates how well the model predicts outcomes based on hypothetical data. Additionally, depending on the results of the evaluation, fine-tuning is also done via adjusting hyperparameters like learning rates, batch sizes, or the number of hidden layers.

3. RESULTS AND DISCUSSION

The proposed classifiers of this work are trained on the extracted features in the previous phase and evaluated according to their accuracy and validation calculations of the used ML classifier. The collected SIFT features from the cropped frontal face are used to train the NN and SVM classifiers. The experiment is run using a single Core-i7 CPU, 16 G of random-access memory, and a Windows 10 operating system. Three main racial categories have been used; African, Chinese, and White. In the second and third experiments, the HOG and Geodesic features have been extracted from the cropped face area for classifying the same three races. Additionally, a local feature-based experiment is investigated

to measure the classification performance while training the algorithm on just cropped mouth parts of Headspace facial scans to explore the feasibility of using the mouth area in identifying the race. The Mouth-SIFT features showed a high accuracy (100%) in addition to decreasing the feature space size. Comparing the obtained results showed that the use of the SIFT features on the complete facial area of the Headspace dataset outperforms both the HOG and Geodesic performance with an accuracy of 90%. While the HOG-based classification achieved the least accuracy (70%) in distinguishing the three used classes. A comparison of the achieved NN classifier accuracy of the used extractions is shown in Figure 12. The results of the SVM classifier are shown in Figure 13. For further evaluation, the confusion matrix parameters have been used to evaluate the positive and negative errors of the actual and the predicted class labels as shown in Figure 14.
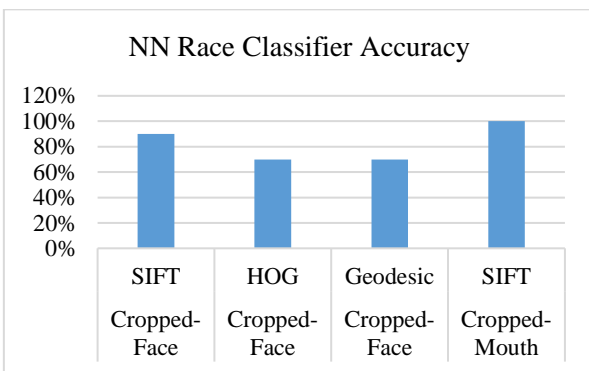

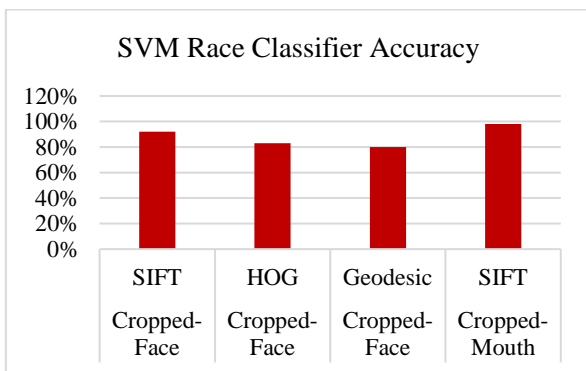**Figure 12.** Comparing the achieved NN classifier accuracies

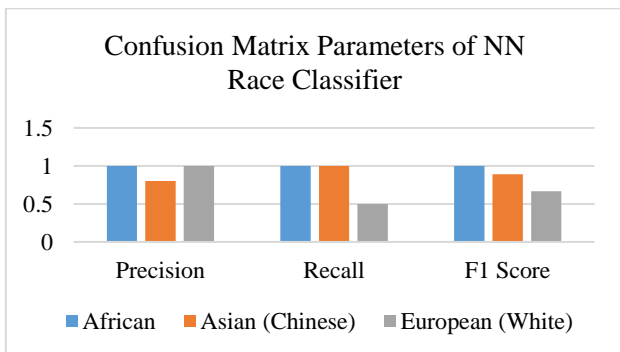**Figure 13.** Comparing the achieved SVM classifier accuracies

Figure 14. Confusion matrix parameters of the NN classifier

Extraction of the SIFT features from the cropped mouth area enhanced the prediction accuracy in addition to reducing the memory consumption by reducing the consumed time and the search area of the used extraction and reducing the size of the resulting feature vectors. The high performance of 3D SIFT was consistent with the literature of traditional 2D works that highlight SIFT's ability to capture the local facial details with high robustness against noise and scale variation. In addition, SIFT descriptors outperform the HOG and Geodesic in the feature vectors' size due to their role in reducing the image description into a relatively acceptable size set of points used later to measure the similarity in the patterns of other scans in the dataset.

4. CONCLUSIONS AND FUTURE WORK

This work proved the high performance of implementing the SIFT as a feature descriptor on 3D facial scans. In addition, it highlighted the mouth area as one of the most descriptive areas within the human 3D facial scans in identifying human ethnicity. The limitation of this work came from the limitation in the available 3D mesh datasets especially the availability of African race 3D scans. This constrained the implementation of the proposed feature extraction and race classification on three racial groups due to the unfairly distributed samples of races. Combining two available facial mesh datasets of Asians and Europeans can be useful to solve this limitation. However, the inconsistent data representation such as representing the norm of vertices and faces, and indexing the vertices can be a challenging task in extracting the features and automatic approaches of landmarking and cropping. In the future, this work is aimed to be extended to explore more 3D datasets with fair numbers of each group or combine Asian data with the Headspace to increase the training set size. In addition, it aims to implement the proposed classifier on the eyes and nose areas to investigate the correlation with the predicted racial class.

REFERENCES

[1] Zhuang, Z., Landsittel, D., Benson, S., Roberge, R., Shaffer, R. (2010). Facial anthropometric differences among gender, ethnicity, and age groups. Annals of Occupational Hygiene, 54(4): 391-402. https://doi.org/10.1093/annhyg/meq007

[2] Bertillon, A., McClaughry, R.W. (1896). Signaletic instructions including the theory and practice of anthropometrical identification. Werner Company. https://doi.org/10.1038/054569a0

[3] Heckathorn, D.D., Broadhead, R.S., Sergeyev, B. (2001). A methodology for reducing respondent duplication and impersonation in samples of hidden populations. Journal of Drug Issues, 31(2): 543-564. https://doi.org/10.1177/002204260103100209

[4] Zhuang, Z., Guan, J., Hsiao, H., Bradtmiller, B. (2004). Evaluating the representativeness of the LANL respirator fit test panels for the current US civilian workers. Journal of the International Society for Respiratory Protection, 21: 83-93.

[5] Zhuang, Z., Landsittel, D., Benson, S., Roberge, R., Shaffer, R. (2010). Facial anthropometric differences among gender, ethnicity, and age groups. Annals of Occupational Hygiene, 54(4): 391-402. https://doi.org/10.1093/annhyg/meq007

[6] Han, D.H., Choi, K.L. (2003). Facial dimensions and predictors of fit for half-mask respirators in Koreans. Aiha Journal, 64(6): 815-822. https://doi.org/10.1080/15428110308984877

[7] Rebar, J.E., Johnson, A.T., Russek-Cohen, E., Caretti, D.M., Scott, W.H. (2004). Effect of differing facial characteristics on breathing resistance inside a respirator mask. Journal of Occupational and Environmental Hygiene, 1(6): 343-348. https://doi.org/10.1080/15459620490447956

[8] Ballihi, L., Amor, B.B., Daoudi, M., Srivastava, A., Aboutajdine, D. (2012). Boosting 3-D-geometric features for efficient face recognition and gender classification. IEEE Transactions on Information Forensics and Security, 7(6): 1766-1779. https://doi.org/10.1109/tifs.2012.2209876

[9] Abbas, H., Hicks, Y., Marshall, D., Zhurov, A.I., Richmond, S. (2018). A 3D morphometric perspective for facial gender analysis and classification using geodesic path curvature features. Computational Visual Media, 4: 17-32. https://doi.org/10.1007/s41095-017-0097-1

[10] Mohammad, A.S., Al-Ani, J.A. (2018). Convolutional neural network for ethnicity classification using ocular region in mobile environment. In 2018 10th Computer Science and Electronic Engineering (CEEC), Colchester, UK, pp. 293-298. https://doi.org/10.1109/ceec.2018.8674194

[11] Kaushik, V.D., Pathak, V.K., Gupta, P. (2010). Geometric modeling of 3D-face features and its applications. Journal of Computers, 5(9): 1305-1314. https://doi.org/10.4304/jcp.5.9.1305-1314

[12] Tsalakanidou, F., Tzovaras, D., Strintzis, M.G. (2003). Use of depth and colour eigenfaces for face recognition. Pattern Recognition Letters, 24(9-10): 1427-1435. https://doi.org/10.1016/s0167-8655(02)00383-5

[13] Chang, K.I., Bowyer, K.W., Flynn, P.J. (2003). Face recognition using 2D and 3D facial data. In Workshop in Multidimonal User Authentication, pp. 25-32.

[14] Bellavia, F., Colombo, C. (2020). Is there anything new to say about SIFT matching? International Journal of Computer Vision, 128(7): 1847-1866. https://doi.org/10.1007/s11263-020-01297-z

[15] Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60: 91-110. https://doi.org/10.1023/b:visi.0000029664.99615.94

[16] Hema, D., Kannan, S. (2021). Patch-SIFT: Enhanced feature descriptor to learn human facial emotions using an Ensemble approach. Indian Journal of Science and Technology, 14(21): 1740-1747. https://doi.org/10.17485/ijst/v14i21.2261

[17] Mian, A.S., Bennamoun, M., Owens, R. (2008). Keypoint detection and local feature matching for textured 3D face recognition. International Journal of Computer Vision, 79: 1-12. https://doi.org/10.1007/s11263-007-0085-5

[18] Berretti, S., Del Bimbo, A., Pala, P. (2011). 3D partial face matching using local shape descriptors. In Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding, New York, United States, pp. 65-71. https://doi.org/10.1145/2072572.2072591

[19] Smeets, D., Keustermans, J., Vandermeulen, D., Suetens, P. (2013). meshSIFT: Local surface features for 3D face recognition under expression variations and partial data. Computer Vision and Image Understanding, 117(2): 158-169. https://doi.org/10.1016/j.cviu.2012.10.002

[20] Lin, S., Liu, F., Liu, Y., Shen, L. (2019). Local feature tensor based deep learning for 3D face recognition. In 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), Lille, France, pp. 1-5. https://doi.org/10.1109/fg.2019.8756616

[21] Berretti, S., Del Bimbo, A., Pala, P. (2012). Sparse matching of salient facial curves for recognition of 3-D faces with missing parts. IEEE Transactions on Information Forensics and Security, 8(2): 374-389. https://doi.org/10.1109/tifs.2012.2235833

[22] Zhang, C., Gu, Y.Z., Hu, K.L., Wang, Y.G. (2015). Face recognition using SIFT features under 3D meshes. Journal of Central South University, 22(5): 1817-1825. https://doi.org/10.1007/s11771-015-2700-x

[23] Darom, T., Keller, Y. (2012). Scale-invariant features for 3-D mesh models. IEEE Transactions on Image Processing, 21(5): 2758-2769. https://doi.org/10.1109/tip.2012.2183142

[24] Shi, Y., Lv, Z., Bi, N., Zhang, C. (2020). An improved SIFT algorithm for robust emotion recognition under various face poses and illuminations. Neural Computing and Applications, 32: 9267-9281. https://doi.org/10.1007/s00521-019-04437-w

[25] Zhao, J., Pan, Z., Duan, F., Lv, Z., Li, J.H., Zhou, Q., Shang, X.G., Sun, J., Wang, P.P. (2019). A Survey on 3D Face Recognition based on Geodesics. Journal of Information Hiding and Multimedia Signal Processing, 10(2): 368-383. https://doi.org/10.1016/b978-012088452-0/50017-0

[26] Drira, H., Amor, B.B., Srivastava, A., Daoudi, M. (2009). A Riemannian analysis of 3D nose shapes for partial human biometrics. In 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, pp. 2050-2057. https://doi.org/10.1109/iccv.2009.5459451

[27] Kurtek, S., Drira, H. (2015). A comprehensive statistical framework for elastic shape analysis of 3D faces. Computers & Graphics, 51: 52-59. https://doi.org/10.1016/j.cag.2015.05.027

[28] Lee, D., Krim, H. (2017). 3D face recognition in the Fourier domain using deformed circular curves. Multidimensional Systems and Signal Processing, 28: 105-127. https://doi.org/10.1007/s11045-015-0334-7

[29] Berretti, S., Del Bimbo, A., Pala, P. (2010). 3D face recognition using isogeodesic stripes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(12): 2162-2177. https://doi.org/10.1109/tpami.2010.43

[30] Fu, S., He, H., Hou, Z.G. (2014). Learning race from face: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(12): 2483-2509. https://doi.org/10.1109/tpami.2014.2321570

[31] Khan, K., Ali, J., Uddin, I., Khan, S., Roh, B.H. (2021). A Facial Feature Discovery Framework for Race Classification Using Deep Learning. arxiv preprint arxiv:2104.02471. https://arxiv.longhoe.net/abs/2104.02471

[32] Richmond, S., Howe, L.J., Lewis, S., Stergiakouli, E., Zhurov, A. (2018). Facial genetics: A brief overview. Frontiers in Genetics, 9: 412703. https://doi.org/10.3389/fgene.2018.00462

[33] AlBdairi, A.J.A., Xiao, Z., Alkhayyat, A., Humaidi, A.J.,

Fadhel, M.A., Taher, B.H., Alzubaidi, L., Santamaria, J., Al-Shamma, O. (2022). Face recognition based on deep learning and FPGA for ethnicity identification. Applied Sciences, 12(5): 2605. https://doi.org/10.3390/app12052605

[34] Khellat-Kihel, S., Muhammad, J., Sun, Z., Tistarelli, M. (2022). Gender and ethnicity recognition based on visual attention-driven deep architectures. Journal of Visual Communication and Image Representation, 88: 103627. https://doi.org/10.1016/j.jvcir.2022.103627

[35] Fan, Y., Zhao, Q., Wang, D. (2020). On 3D face attributes analysis using deep learning: A preliminary case study on gender and ethnicity recognition. In Proceedings of the 2020 2nd International Conference on Image Processing and Machine Vision, Bangkok Thailand, pp. 69-73. https://doi.org/10.1145/3421558.3421569

[36] Xia, B., Amor, B.B., Daoudi, M. (2017). Joint gender, ethnicity and age estimation from 3D faces: An experimental illustration of their correlations. Image and Vision Computing, 64: 90-102. https://doi.org/10.1016/j.imavis.2017.06.004

[37] Lu, X., Chen, H., Jain, A.K. (2005). Multimodal facial gender and ethnicity identification. In Advances in Biometrics: International Conference, ICB 2006, Hong Kong, China, pp. 554-561. https://doi.org/10.1007/11608288_74

[38] Toderici, G., O'malley, S.M., Passalis, G., Theoharis, T., Kakadiaris, I.A. (2010). Ethnicity-and gender-based subject retrieval using 3-D face-recognition techniques. International Journal of Computer Vision, 89: 382-391. https://doi.org/10.1007/s11263-009-0300-7

[39] Ocegueda, O., Fang, T., Shah, S.K., Kakadiaris, I.A. (2012). 3D face discriminant analysis using Gauss-Markov posterior marginals. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(3): 728-739. https://doi.org/10.1109/TPAMI.2012.126

[40] Zhong, C., Sun, Z., Tan, T. (2009). Fuzzy 3D face ethnicity categorization. In Advances in Biometrics: Third International Conference, ICB 2009, Alghero, Italy, pp. 386-393. https://doi.org/10.1007/978-3-642-01793-3_40

[41] Ding, H., Huang, D., Wang, Y., Chen, L. (2013). Facial ethnicity classification based on boosted local texture and shape descriptions. In 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, pp. 1-6. https://doi.org/10.1109/fg.2013.6553815

[42] Lv, C., Wu, Z., Zhang, D., Wang, X., Zhou, M. (2019). 3D Nose shape net for human gender and ethnicity classification. Pattern Recognition Letters, 126: 51-57. https://doi.org/10.1016/j.patrec.2018.11.010

[43] Sovizi, J., Rai, R., Krovi, V. (2014). Draft: 3D face recognition under isometric expression deformations. Proceedings of the ASME 2014 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE 2014, Buffalo, USA, pp. 1-6. https://doi.org/10.1115/DETC2014-34449

[44] Dai, H., Pears, N., Smith, W., Duncan, C. (2020). Statistical modeling of craniofacial shape and texture. International Journal of Computer Vision, 128(2): 547-571. https://doi.org/10.1007/s11263-019-01260-7

[45] Dai, H., Pears, N., Smith, W.A., Duncan, C. (2017). A 3D morphable model of craniofacial shape and texture variation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, pp. 3085-3093. https://doi.org/10.1109/iccv.2017.335

[46] Amberg, B., Romdhani, S., Vetter, T. (2007). Optimal step nonrigid ICP algorithms for surface registration. In 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA pp. 1-8. https://doi.org/10.1109/cvpr.2007.383165

[47] Audenaert, E.A., Van Houcke, J., Almeida, D.F., Paelinck, L., Peiffer, M., Steenackers, G., Vandermeulen, D. (2019). Cascaded statistical shape model based segmentation of the full lower limb in CT. Computer Methods in Biomechanics and Biomedical Engineering, 22(6): 644-657. https://doi.org/10.1080/10255842.2019.1577828

[48] Darom, T., Keller, Y. (2012). Scale-invariant features for 3-D mesh models. IEEE Transactions on Image Processing, 21(5): 2758-2769. https://doi.org/10.1109/tip.2012.2183142

[49] Dupre, R., Argyriou, V., Greenhill, D., Tzimiropoulos, G. (2015). A 3D scene analysis framework and descriptors for risk evaluation. In 2015 International Conference on 3D Vision, Lyon, France, pp. 100-108. https://doi.org/10.1109/3dv.2015.19

[50] Gabriel Peyre. (2023). Toolbox Fast Marching https://www.mathworks.com/matlabcentral/fileexchange/6110-toolbox-fast-marching.