2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9871184

3D Facial Landmark Localization for cephalometric analysis

Helena R. Torres, Pedro Morais, Anne Fritze, Bruno Oliveira, Fernando Veloso, Mario Rüdiger,
Jaime C. Fonseca, João L. Vilaça

Abstract— **Cephalometric analysis is an important and routine task in the medical field to assess craniofacial development and to diagnose cranial deformities and midline facial abnormalities. The advance of 3D digital techniques potentiated the development of 3D cephalometry, which includes the localization of cephalometric landmarks in the 3D models. However, manual labeling is still applied, being a tedious and time-consuming task, highly prone to intra/inter-observer variability. In this paper, a framework to automatically locate cephalometric landmarks in 3D facial models is presented. The landmark detector is divided into two stages: (i) creation of 2D maps representative of the 3D model; and (ii) landmarks' detection through a regression convolutional neural network (CNN). In the first step, the 3D facial model is transformed to 2D maps retrieved from 3D shape descriptors. In the second stage, a CNN is used to estimate a probability map for each landmark using the 2D representations as input. The detection method was evaluated in three different datasets of 3D facial models, namely the Texas 3DFR, the BU3DFE, and the Bosphorus databases. An average distance error of 2.3, 3.0, and 3.2 mm were obtained for the landmarks evaluated on each dataset. The obtained results demonstrated the accuracy of the method in different 3D facial datasets with a performance competitive to the state-of-the-art methods, allowing to prove its versability to different 3D models.**

Clinical Relevance— **Overall, the performance of the landmark detector demonstrated its potential to be used for 3D cephalometric analysis.**

I. INTRODUCTION

Cephalometric analysis refers to the assessment of the craniofacial structure to evaluate its growth and development and to diagnose cranial deformities, midline facial abnormalities, and orthodontic problems [1]. Traditionally, 2D radiographs are used for cephalometric analysis, but Computed Tomography (CT) and Magnetic Resonance (MR) imaging systems have been enabling 3D cephalometry [2]. More recently, some works proposed to use 3D digital models (e.g. laser scans) to perform the cephalometric analysis [3]. To evaluate the craniofacial anatomy in the scanned 3D models, manual identification of anatomic landmarks is performed, followed by the calculation of established measurements. However, the manual identification of the landmarks is a time-consuming task that it is also highly prone to observer variability [4], [5]. Thus, automated solutions to detect the landmarks can be useful tools to clinical practice.

Recently, with the improvement of computing capabilities, deep neural networks have been widely used for medical and computer vision tasks, such as landmark localization [4], [6]–[8]. In fact, deep learning (DL) showed superior performance over the conventional machine learning strategies or registration-based approaches [9]. Our team have already demonstrated the added-value of the DL techniques to detect some landmarks in 3D infant's head surfaces [10]. Overall, the proposed strategy is a two-stage method that includes creation of 2D maps representative of the 3D head model and detection of anthropometric landmarks in the 2D maps using a DL strategy. In this paper, inspired by our previous work, we sought to extend the proposed methodology and evaluate its performance for the detection of other cephalometric landmarks in generic facial databases with heterogenous populations (not only infants) and compare its performance against state-of-the-art landmark detection methods.

II. METHODS

A. General overview

The proposed landmark detector relies on two-stage approach (Figure 1). In the first stage, the 3D model is transformed into a 2D representation that is embedded with shape descriptors information to create 2D maps representative of the geometry of the 3D model. This stage decreases the detection complexity by decreasing the dimensionality of the data, while maintaining surface information in the shape descriptors. The second stage relies on a regression convolutional neural network (CNN) that

This work was funded by projects "NORTE-01–0145-FEDER000059, NORTE-01-0145-FEDER-024300 and "NORTE-01–0145-FEDER-000045, supported by Northern Portugal Regional Operational Programme Norte2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (FEDER). It was also funded by national funds, through the FCT–Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020 and LASI-LA/P/0104/2020 and by FCT and FCT/MCTES in the scope of the projects UIDB/05549/2020 and UIDP/05549/2020. The authors also acknowledge support from FCT and the European Social Found, through Programa Operacional Capital Humano (POCH), in the scope of the PhD grant SFRH/BD/136670/2018, SFRH/BD/136721/2018 and SFRH/BD/131545/2017.

Helena R. Torres and Bruno Oliveira are with 2Ai – School of Technology, IPCA, Barcelos, Portugal, with Algoritmi Center, School of Engineering, University of Minho, Guimarães, Portugal, with Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal, and with ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal (email: htorres@ipca.pt, boliveira@ipca.pt). Pedro Morais and João Vilaça are with 2Ai – Polytechnic Institute of Cávado and Ave, Barcelos, Portugal (email: pmorais@ipca.pt, jvilaca@ipca.pt). Anne Fritze and Mario Rüdiger are with Department for Neonatology and Pediatric Intensive Care, Children's Hospital, Medical Faculty of TU Dresden, Germany (email: anne.fritze@uniklinikum-dresden.de, mario.ruediger@uniklinikum-dresden.de). Fernando Veloso is with 2Ai – Polytechnic Institute of Cávado and Ave, Barcelos, Portugal and with Department of Mechanical Engineering, School of Engineering, University of Minho, Guimarães, Portugal (fveloso@ipca.pt). Jaime Fonseca is with Algoritmi Center, School of Engineering, University of Minho, Guimarães, Portugal (email: jaime@dei.uminho.pt).

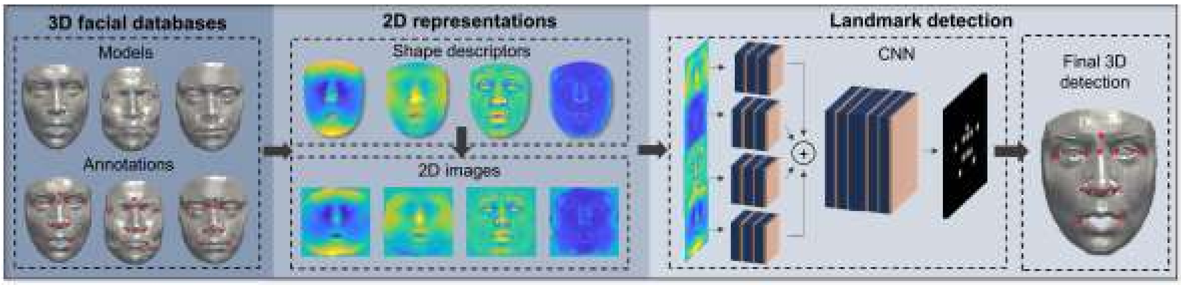

Figure 1 – Overview of the proposed landmark detector.

estimates the localization of the landmarks on the 2D representations. Here, the CNN regress landmark pixel positions by generating a probability map of them.

B. 2D representation of the 3D facial model

1) 3D shape descriptors

3D digital scanning technics allow the creation of 3D models represented by points spatially correlated. In the absence of texture information, shape descriptors can be used to collect 3D point signatures retrieved from the model's surface. In this work, four different descriptors were studied to retrieve enhanced information of the 3D facial model. The first descriptor concerns the depth map of the model, which consists in the distance between each point of the 3D model and its center. The second descriptor is represented by the radial distance between a point and the principal axis of the model. In opposite to the first two descriptors that concern the 3D localization of the model's points in respect to a given reference, the third descriptor quantifies the local curvature of the surface, being retrieved by estimating the Gaussian curvature of each point of the model. Finally, the fourth shape descriptor concerns the number of point connections needed to fully recover the details of a region of the facial model, quantifying the level of detail in the different regions. Figure **2** illustrates the different shape descriptors.

2) Explicit functions for 2D representation

To decrease the complexity of the 3D detection, a 2D representation of the 3D facial model is used. Here, the 3D model is represented as an explicit function where one of the coordinates of the points on the model is given explicitly as a function of the remaining coordinates. Thus, one can obtain a representation function which has one less dimension than the original model. The first step to define the geometric function that maps 3D points to 2D is the definition of a coordinate system. In this work, the cylindrical coordinate system was chosen. Thus, the cartesian coordinates of each point $\boldsymbol{p} = \{x, y, z\}$ are converted in cylindrical coordinates $\boldsymbol{p} = \{\rho, \varphi, z\}$, where ρ represents the radial distance from the model principal axis to $\boldsymbol{p}$, φ is the azimuth angle, and z is the height of $\boldsymbol{p}$. The second step to formulate the explicit function is to select the coordinates that will define the explicit coordinate. Here, it was defined that the explicit coordinate ρ' is obtained as a function of the azimuth angle φ and height z. Mathematically, this can be defined as [11]:

$$g: \mathbb{R}^{n-1} \mapsto \mathbb{R}, \rho' = g(\varphi, z). \qquad (1)$$

Finally, for each (φ, z), the value of ρ' was defined to be the value of a given shape descriptor studied in this work at the corresponding (x, y, z) point, promoting the inclusion of the descriptors into the 2D representation (Figure 3). Thus, four 2D representative maps were obtained for a given 3D model, each one related to one shape descriptor.

C. Regression CNN for probability maps estimation

After creating the 2D representative maps, a regression CNN was applied to estimate probability maps for the landmarks' localization. Similar to [10], a multi-branch approach was implemented where a CNN was applied to process each 2D map individually. This generates a set of feature maps V_r, with $r \in \{1 \dots 4\}$, that are afterward concatenated into a global one. The global feature map is then feed to the second part of the CNN that is used to predict confidence maps for each landmark position. For that, a Sigmoid activation layer was added to the end of the network, considering that the confidence maps can be given by a Gaussian-like function where its maximum represents the landmark position. To guide the network training, a loss function f_{loss} that calculates the Euclidean distance between predicted maps and the ground-truth maps was used:

$$f_{loss} = \sum_{i=1}^{M} ||L_i - L_i^*||, \qquad (2)$$

where L_i and L_i^* are the prediction and ground truth maps for landmark i, respectively, and M is the number of landmarks. For each landmark, the ground-truth map was generated by applying a Gaussian-like function where the maximum of the gaussian map represents the landmark position. In the test phase, each landmark is detected by estimating the respective probability map, being the optimal position defined as the peak of the map, after a non-maximum suppression processing. In the final step, the obtained landmarks are transferred to the 3D world by reverting the transformation between the 3D model and 2D representation.

III. Experiments and results

A. 3D facial databases

The accuracy of the proposed landmark detector was evaluated in three different 3D facial surfaces benchmarks:

- Texas 3D Database: contains 1149 pairs of facial color and depth images of 105 subjects. In this work, 3D models were obtained from the depth using the specified acquisition parameters [12]–[14];
- BU3DFE Database: composed by 2500 facial expression models of 100 subjects [15];
- Bosphorus 3D Face Database**:** comprises 4666 3D faces from 105 subjects, including facial expressions, rotations,

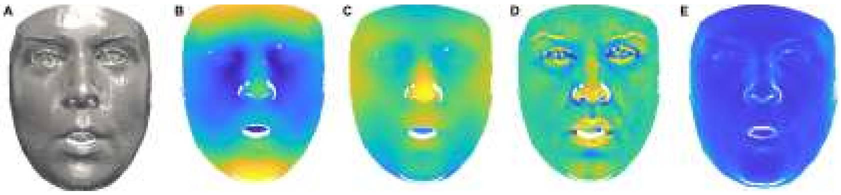

Figure 2 – Shape descriptors. (A) 3D Model; (B) Depth map; (C) Radial map; (D) Gaussian curvature map; (E) Point connections map.

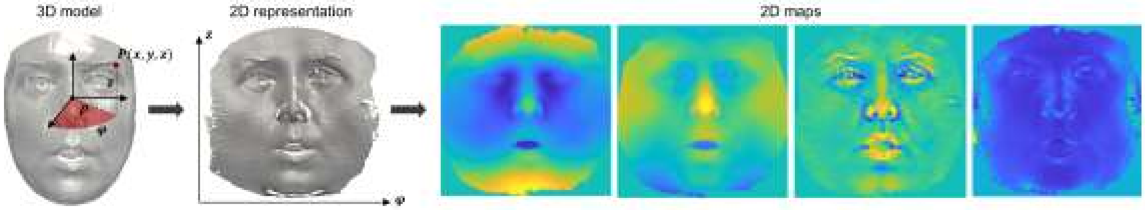

Figure 3 – Creation of 2D map representative of the 3D shape. The intensity of each 2D map concerns the values of the shape descriptors.

and occlusions [16]. In this work, 2921 models referring to facial expressions were evaluated;

For all datasets, color/texture images were discarded, being only used the 3D models. Eleven landmarks were evaluated: glabella (G), sellion (SL), right/left exocanthions (EX), right/left endocanthions (EN), right/left subalare (SN), nose tip (NP), and right/left mouth corner (MC). Once BU3DFE and Bosphorus databases do not contain labelling for GL and SL, these landmarks were estimated from the inner eyebrows and nose saddle points, respectively.

B. *Implementation details*

In the first stage of the method, 2D representation maps with a size of 368 × 368 pixels were created. To overcome overfitting problems during training, data augmentation techniques were applied to the images, namely geometric transformations (*i.e.* small rotations and scaling) and intensity-based transformations (*i.e.* brightness and contrast modification and blur). The network was trained with a mini-batch size of 10 images and using the Adam optimizer with an initial learning rate of 0.0004 and with a regularization term of 0.01. At the end of each epoch, the learning rate was updated using a polynomial learning rate decay policy. The training convergency was analyzed to select the epoch used for testing. Finally, to evaluate the accuracy of the proposed methodology in all 3D facial models, a four-fold cross validation strategy was applied for each dataset, using 75% of data for training and 25% for testing at each experiment.

C. *Landmark detector performance*

To validate the proposed method, the automatic results were compared against a manual ground-truth using the mean error, defined as the Euclidean distance between the estimated landmarks and the true positions. Table 1 summarizes the performance of the method on the different databases. The method's performance is assessed in terms of mean distance error and compared with state-of-the-art methods described in [6], [12], [17]–[20]. Figure 4 presents example results of the detector.

IV. Discussion

Analyzing Table 1, it is possible to verify the detector's accuracy since low landmark detection errors were achieved. Specifically, a mean error of 2.8 mm for all landmarks was obtained, which can be considered an acceptable distance error. The good detection results among all landmarks corroborated the feasibility of performing 3D landmark detection using 2D maps that represents the geometric properties of the 3D facial model. Thus, the results suggest that the chosen shape descriptors accurately preserve the most important features of the model. Moreover, these features represent specific characteristics for each landmark, promoting that the landmark can be correctly distinguished. Regarding the configuration of the proposed DL model, it can be concluded that the multi-branch regression network can effectively generated good predictions for the probability maps that represent the optimal landmark positions.

When compared to the remaining methods, our method obtained the best results for most of the landmarks. Moreover, the results showed that the performance of the proposed method is consistently good for different datasets, which can be also visualized in Figure 4. This corroborates the added value of the proposed methodology for different applications.

As a remark, the original landmark detection strategy was not specifically developed for the detection of cephalometric landmarks on heterogeneous facial models. Instead, the method was proposed to detect anthropometric landmarks on synthetic head models and head models retrieved from MR images. However, the extension and adaptation of the landmark detector proposed in [9] to different landmarks and different datasets, where the data is acquired with different acquisition techniques and presents different type of facial features (e.g. facial expressions or presence of hair and beard), showed to be effective, with stable results and always competitive (or even outperforming) the state-of-the art.

V. Conclusion

In this work, a landmark detection method for 3D facial models was proposed. The results obtained by the proposed method in different facial datasets demonstrated its high accuracy and competitiveness with state-of-the-art landmark detectors. Overall, the proposed method can be used in clinical practice for 3D cephalometric analysis. Moreover, craniofacial pathologies, *e.g.* midfacial abnormalities, can also be evaluated using the proposed method.

Table 1 – Landmark detection errors (in mm) on 3D facial datasets. Best results for each landmark and dataset are presented in bold.

		GL	R-EX	R-EN	SL	L-ED	L-EX	R-SN	NT	L-SN	R-MC	L-MC	Mean of pairs			
													EX	EN	SN	MC
Texa	Proposed	**2.55**	**2.49**	**1.98**	**2.06**	**2.09**	**2.52**	2.38	**2.13**	**2.33**	**2.15**	**2.33**	**2.50**	**2.03**	2.35	**2.24**
	Gupta et al.[12]	-	4.36	2.38	-	2.16	4.13	**2.25**	-	2.33	5.19	4.97	4.25	2.27	**2.29**	5.08
BU3DFE	Proposed	**3.32**	**2.94**	**2.17**	**2.69**	**2.16**	**3.04**	3.29	3.47	**3.30**	3.41	3.69	**2.99**	**2.16**	3.30	3.55
	Salazar et al. [17]	-	8.49	6.14	-	6.75	9.63	7.17	5.87	6.47	-	-	9.06	6.45	6.82	-
	Gilani et al. [6]	-	3.30	2.40	2.90	2.20	3.80	-	**2.50**	-	-	-	3.55	2.30	**2.30**	4.60
	Sun et al. [18]	-	3.02	2.63	-	2.77	3.22	**3.19**	-	3.30	3.43	**3.32**	3.12	2.70	3.25	**3.38**
	Fanelli et al. [19]	-	4.00	2.80	-	2.60	3.60	4.10	-	3.90	**3.10**	4.41	3.80	2.70	4.00	3.76
Bosphorus	Proposed	**2.09**	**3.32**	**2.04**	4.48	3.74	**2.13**	3.10	2.60	5.81	**3.45**	**2.90**	**2.73**	2.89	4.45	3.18
	Gilani et al. [7]	2.63	-	-	-	-	-	-	**2.24**	-	-	-	2.98	2.68	**2.68**	**2.76**
	Vezzetti et al. [20]	-	-	-	3.74	-	-	-	2.62	-	-	-	5.38	4.36	4.83	-
	Gilani et al. [6]	-	4.01	2.40	**2.32**	**2.35**	3.57	**2.99**	2.82	**2.5**	4.91	4.85	3.79	**2.38**	2.75	4.88

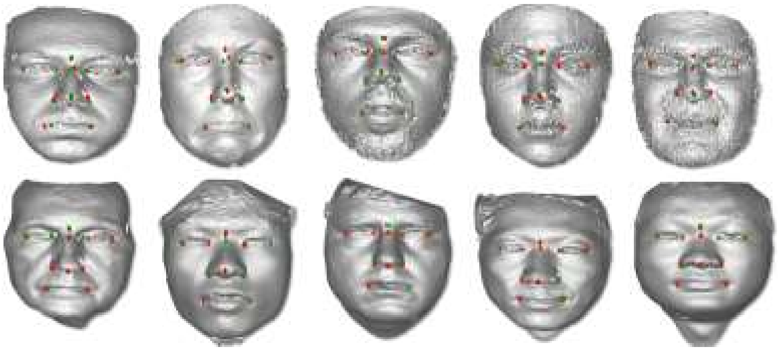

Figure 4 – Example of detection results (red) and manual labelling (green). First row - Bosphorus database; Second row – Texas database.

VI. References

[1] H. J. Cho, “A three-dimensional cephalometric analysis,” *J. Clin. Orthod.*, no. June, 2014.

[2] A. Juerchott *et al.*, “In vivo comparison of MRI- and CBCT-based 3D cephalometric analysis: beginning of a non-ionizing diagnostic era in craniomaxillofacial imaging?,” *Eur. Radiol.*, vol. 30, no. 3, pp. 1488–1497, 2020.

[3] S. H. Kim and H. S. Shin, “Three-dimensional analysis of the correlation between soft tissue and bone of the lower face using three-dimensional facial laser scan,” *J. Craniofac. Surg.*, vol. 29, no. 8, pp. 2048–2054, 2018.

[4] H. R. Torres *et al.*, “Deep learning-based detection of anthropometric landmarks in 3D infants head models,” in *SPIE Medical Imaging*, 2019, no. March, p. 112.

[5] B. Oliveira *et al.*, “Automatic strategy for extraction of anthropometric measurements for the diagnostic and evaluation of deformational plagiocephaly from infant’s head models,” in *SPIE Medical Imaging. International Society for Optics and Photonics*, 2019, no. June, p. 9.

[6] S. Z. Gilani, A. Mian, F. Shafait, and I. Reid, “Dense 3D Face Correspondence,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 7, pp. 1584–1598, 2018.

[7] S. Z. Gilani, A. Mian, and P. Eastwood, “Deep, dense and accurate 3D face correspondence for generating population specific deformable models,” *Pattern Recognit.*, vol. 69, pp. 238–250, 2017.

[8] Helena R. Torres *et al.*, “Deep Learning-based Detection of Anthropometric Landmarks in 3D Infants Head Models,” in *SPIE Medical Imaging*, 2019.

[9] K. Khabarlak and L. Koriashkina, “Fast Facial Landmark Detection and Applications: A Survey,” 2021.

[10] H. R. Torres *et al.*, “Anthropometric Landmark Detection in 3D Head Surfaces using a Rotation-Invariant Deep Learning-based Approach,” *IEEE J. Biomed. Heal. Informatics*, vol. 2194, 2020.

[11] Q. Duan, E. D. Angelini, and A. F. Laine, “Real-time segmentation by Active Geometric Functions,” *Comput. Methods Programs Biomed.*, vol. 98, no. 3, pp. 223–230, 2010.

[12] S. Gupta, M. K. Markey, and A. C. Bovik, “Anthropometric 3D face recognition,” *Int. J. Comput. Vis.*, vol. 90, no. 3, pp. 331–349, 2010.

[13] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik, “Texas 3D Face Recognition Database,” *Proc. IEEE Southwest Symp. Image Anal. Interpret.*, pp. 97–100, 2010.

[14] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik, “Texas 3D Face RecognitionDatabase.” [Online]. Available: https://live.ece.utexas.edu/research/texas3dfr/.

[15] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, “A 3D facial expression database for facial behavior research,” *FGR 2006 Proc. 7th Int. Conf. Autom. Face Gesture Recognit.*, vol. 2006, no. August, pp. 211–216, 2006.

[16] A. Savran *et al.*, “Bosphorus database for 3D face analysis,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5372 LNCS, pp. 47–56, 2008.

[17] A. Salazar, S. Wuhrer, C. Shu, and F. Prieto, “Fully automatic expression-invariant face correspondence,” *Mach. Vis. Appl.*, vol. 25, no. 4, pp. 859–879, 2014.

[18] J. Sun, D. Huang, Y. Wang, and L. Chen, “Expression robust 3D facial landmarking via progressive coarse-to-fine tuning,” *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 15, no. 1, pp. 1–23, 2019.

[19] G. Fanelli, M. Dantone, and L. Van Gool, “Real time 3D face alignment with Random Forests-based Active Appearance Models,” *2013 10th IEEE Int. Conf. Work. Autom. Face Gesture Recognition, FG 2013*, 2013.

[20] E. Vezzetti, F. Marcolin, S. Tornincasa, L. Ulrich, and N. Dagnes, “3D geometry-based automatic landmark localization in presence of facial occlusions,” *Multimed. Tools Appl.*, vol. 77, no. 11, pp. 14177–14205, 2018.