Measurement 182 (2021) 109686

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

# Combining machine learning and close-range photogrammetry for infant's head 3D measurement: A smartphone-based solution

Inés Barbero-García [a,b,*], Roberto Pierdicca [c], Marina Paolanti [d], Andrea Felicetti [d], José Luis Lerma [a]

[a] Department of Cartographic Engineering, Geodesy and Photogrammetry, Universitat Politècnica de València, Camino de Vera, s/n, Building 7i, 46022 Valencia, Spain
[b] Department of Cartographic and Land Engineering, Higher Polytechnic School of Avila, University of Salamanca, Hornos Caleros 50, 05003 Ávila, Spain
[c] Università Politecnica delle Marche, Dipartimento di Ingegneria Civile, Edile e dell'Architettura (DICEA), 60131 Ancona, Italy
[d] Universita Politecnica delle Marche, Dipartimento di Ingegneria dell'Informazione (DII), 60131 Ancona, Italy

ARTICLE INFO

*Keywords:*
3D data acquisition
Smartphone
Facial landmark detection
Plagiocephaly
Photogrammetry

ABSTRACT

Three-dimensional data has a wide range of applications in medicine. For the particular case of cranial deformation in infants, it is becoming a common tool for evaluation. However, there is a need for low-cost solutions that provide accurate information even with uncoll aborative infants with ultrafast movement reactions. As cranial deformation is often linked to facial abnormalities, facial information is required for comprehensive evaluation.

In this study, the integration of target-based close-range photogrammetry and facial landmark machine learning detection is carried out. The resulting tool is automatic and smartphone-based and provides 3D information of the head and face. This methodology opens a new path for the effective integration of machine learning and photogrammetry in medicine and, in particular, for overall head analysis.

## 1. Introduction

Three-dimensional (3D) models have a wide range of applications in medicine. In particular, they are becoming a common methodology for the evaluation of cranial deformation in infants, a problem that includes pathologies such as plagiocephaly or different types of craniosynostosis [1,2]. Most methodologies for the eventual creation of 3D models of moving infants are costly and not widely implemented in the clinical routine, where manual measurements, and even visual evaluation according to the Argenta's assessment tool [3], are still common [4–6]. Close-range photogrammetry proved to be reliable for the creation of trustworthy models, driving experts in the evaluation of cranial deformations; the information is complete, every kind of morphological feature can be inferred and results are not dependent on the expert [7]. In combination with abnormalities in the cranial shape, the abnormal face shape is also common for some types of deformation and has also to be considered for the right medical diagnostics.

In particular, features such as the position of the eyes [8], with hypotelorism problems (eyes closer than normal) or orbital dystopia (one eye lower than the other), are important concerns for some types of cranial synostosis [9]. Coronal synostosis can also lead to asymmetries in other areas of the face, such as the nasal root and chin [10,11]. It is thus extremely important to develop easy to use and cost-effective methods for such evaluations, as these deformations present many risks like social isolation and visual problems, just to mention some [11,12].

The 3D data provides an important advantage for the evaluation of these problems in comparison with commonly used 2D data [13]. In previous works, the authors developed a low-cost tool to measure and evaluate cranial deformation in infants called PhotoMeDAS (Spanish Patent Number P201930355). PhotoMeDAS is a tool composed of a smartphone app and a cloud-based processing system that allows non-expert photogrammetric users to obtain head 3D models for infant cranial deformation assessment using automatically detected points fitted on a coded cap [14]. The methodology uses a smartphone to detect automatic markers placed on a coded cap. These data are later used to create a 3D model and medical deformation diagnostics. As a target-based photogrammetric tool, PhotoMeDAS is able to provide accurate results even with awake, moving and uncollaborative infants, in a situation where other techniques, such as image-based photogrammetry or structure from motion (SfM), have been proved to fail or require lots of manual processing [15,16].

Despite its advantages, the main limitation of the tool is that the

* Corresponding author.

https://doi.org/10.1016/j.measurement.2021.109686
Received 24 February 2021; Received in revised form 3 May 2021; Accepted 30 May 2021
Available online 5 June 2021
0263-2241/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

obtained information is limited to the cranial vault, without considering the above-mentioned face deformation features. Towards this end and to overcome this limitation, the identification of facial landmarks represents a turnkey for obtaining missing facial data. Another advantage of having well-identified facial landmarks is that it allows the registetion of the head 3D data to a known coordinate system [17]. The registration process is required for evolution monitoring or extraction of cranial measurements. In the specific case of face landmarks, the aim is to recognise the key points that define specific features of the face like eyes, nose, mouth and chin. Therefore, face landmarks detection requires face detection as a first step. The face landmarks detection allows solving problems like the face orientation, facial expression and biometrical measures of the face. Although face detection is commonly trained using 3D models [18], no studies have been found by the authors that use automatic detection landmarks to extract/measure 3D point coordinates of these landmarks.

In this study, a new tool has been developed to add automatically detected facial landmarks to the head 3D models to improve the information contained in the cranial vault 3D model. More in deep, the tool combines a target-based photogrammetric pipeline with a machine learning algorithm, trained to fulfil facial landmark detection. The application, running in real-time on a customer-grade smartphone, proved to reach accuracy and reliability thresholds which are between 1.2 and 6.75 times larger than the ones obtained with the coded markers used for the overall 3D head and face measurements. Nevertheless, the developed system paves the way for a novel method of 3D data acquisition in medical application, proving the effectiveness of the combination between machine learning and photogrammetry.

## 2. State of the art in facial landmark detection

Face detection is the task to identify the position and dimensions of a human face from a digital image. Digital image processing techniques are used to detect the face by ignoring everything that is not a face, which is called background.

The face detection task can be considered a particular case of object detection and/or pattern recognition where the object to be identified, or pattern to be recognised is the human face. A recent review concerning face detection methods is presented by [19]. Many strategies for face detection are implemented during the last years and can be divided as follows:

- Data source to be processed (i.e., images, videos);
- Features that are analysed, such as a region that describes (punctual, defined pixel by pixel; local, defined around a point; and global; semantic content: of low level (colour, edges, corners, textures), of high level: physical structures of the face (eyes, nose, mouth);
- Algorithms used for detection: rule-based (based on image processing; the features are combined by functions defined ad-hoc: weighted sums, thresholds, conditions…), and learned-based (based on automatic learning models, that require an a priori training). This last can be distinguished between machine learning (learn from predefined features) and deep learning (learn directly from the image).

In the last years, face detection has gained greater attention in fields of biometrics [20], pattern recognition [21] and computer vision [22]. Despite this operation is already implemented in almost every device of everyday life, face detection requires better performance in terms of robustness, especially for biometrics or medical applications. Indeed, facial occlusion, uneven lighting and shadows, face orientation, background complexity, low image resolution and noise hamper a wanted adoption among experts. For these reasons, human facial landmarking is gaining increasing attention in medicine for both 2D and 3D compartments [23], and the need for automatizing landmark detection has been widely discussed in the literature [24]. The advancements in machine learning algorithms for computer vision suggest fundamental progress, which defines the basis for the development of general methods to automate landmark data detection [25].

However, for training these learning models, a huge amount of labelled data is necessary. In the medical domain, this aspect is a critical issue due to the lack of data; moreover, when available, datasets are unbalanced and challenging (e.g. tiny features, bioimages complexity and so on). To address this problem, Fu et al. [26] propose a method for extracting facial anatomical landmarks for the diagnosis of fetal alcohol syndrome. Despite the limited training samples available, they propose to maintain the feature representations of the source model on the target task data and to influence them as a further source of supervisory signals for regularizing the target model learning [26]. Rao et al. [27] developed a machine learning approach for orthodontics clinical applications. The study is conducted on 418 facial landmark points and 220 landmark measures from 22 2D facial images of volunteers. They applied “You Only Look Once” (YOLO) network to identify the landmarks [27].

As stated before, landmark detection has been explored even in the case of 3D outputs. A deep learning approach for the automatic localization of the anatomical landmarks on the distal femur bone in 3D medical images is proposed in [28]. The authors adopt the narrow-band graph cut optimization to carry out the 3D segmentation of the femur surface, considering the results obtained from the neural network. This method is useful for determining the position of initial geometric landmarks on the femur surface in the 3D magnetic resonance images. Another work on 3D images is proposed by Abu et al. [29]. This paper aims at automatically processing landmark detection on 3D facial images for measuring craniofacial anthropometry. The geometry information useful to identify the facial characteristics have been used to develop an automated craniofacial landmark (ACL) on a 3D facial image system. The reference model and the acceptable degree of deformation has been determined by using a machine learning model to find out the optimum solution for all faces.

Facial landmark detection is also explored with a framework that simultaneously deals with three tasks: (i) facial landmark detection, (ii) head pose estimation, and (iii) facial deformation analysis. The method is also robust to facial occlusion [30]. The authors used a cascade iterative procedure augmented with model-based pose estimation. In this way, they iteratively predict the facial occlusion, facial landmark locations, head pose angles and facial deformation. Instead, facial landmark detection was faced by using a generative adversarial module that changes original face images to style-aggregated images [31]. A Style-Aggregated Network (SAN) was proposed, that applies a cascade strategy to generate the heatmap predictions which can be robust to the large variance of image styles.

## 3. Materials and methods

This section is devoted to the description of the face landmark detection algorithm developed to be merged with the existing photogrammetric application. Given an image and the ROI (Region of Interest) that localises the object of interest, a shape predictor has to recognise the key points of the shape. Finally, a machine learning approach detects and selects the best point to be considered for its transformation in 3D coordinates. Here following, the detailed procedure and experiments are reported.

### 3.1. Setup

The methodology of this study was tested on five dummies simulating infant’s heads. For each dummy, 10 models were created. A coded cap was placed on each dummy, it was made of elastic material and covered with ArUco automatic markers [32]. Three stickers with coded markers were also placed, between the eyes and close to the ears (Fig. 1). These three stickers were eventually used for the registration of the 3D models. AruCo markers were chosen over other options, such as circular markers, as they provide four points per marker, allowing for better


**Fig. 1.** Dummy with coded cap and registration stickers.

representation of the head shape with the minimum number of markers. In addition, using a lower number of markers makes the production of the cap less expensive and allows for better head fitting, as the fabric is much more elastic without the markers on the top. Another advantage of AruCo is its integration in OpenCV. The cap and stickers were not moved during the different acquisitions. The data for the creation of the models were acquired using the smartphone application PhotoMeDAS. The smartphone used is a Samsung A50, with a resolution limited to 960 × 720 *px*. The smartphone app can detect the coded markers as well as the facial landmarks. It is also capable of selecting useful images and guiding the user during the acquisition. A more detailed explanation of this tool (without face detection) can be found in [14].

### 3.2. The data acquisition smartphone app

A smartphone app was developed to integrate the detection of ArUco markers, present on the cap and the face points (Fig. 2). The app processes each frame captured by the smartphone and carries out the detection of the markers. When a minimum number of markers is detected, the image is preselected, otherwise, the image is discarded.

After that, each preselected image is checked to detect whether it corresponds to the face area (this information is obtained by examination of the visible markers). In case the image is considered to include the face, face landmark detection is carried out. Firstly, the image is rotated to assure it is in the correct position for face detection. This step is also carried out using the information given by the markers. If a face is detected, a correctness check is carried out: the front sticker must be placed between the eyes, otherwise the face landmarks for that image are discarded. Finally, the detected points for that image are stored, including target points for every image and facial landmarks points for images with the detected face. Then the whole head is registered and the data is sent to the server for processing. The application developed was installed on a smartphone Samsung Galaxy A50, with the operating system Android 9 and a RAM of 4 GB.

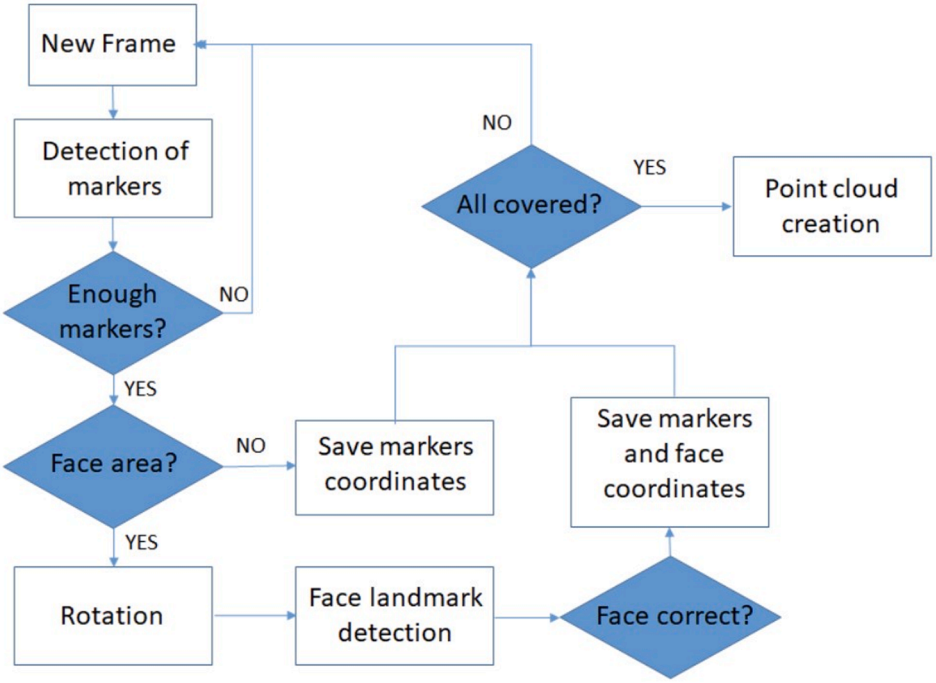


**Fig. 2.** Smartphone application workflow.

### 3.3. Face landmarks detection algorithm

The application acquires the streaming video through a smartphone camera. Each frame has to be shown on the screen with the marking of the detected objects. The time between the frame acquisition and its visualisation on the screen must be the minimum possible, to ensure good fluidity and minimum time delay. The operation executed for each frame of the streaming video can be summarised in the following steps:

1. Frame acquisition;
2. Preprocessing of the image before the face detection;
3. Face detection: localisation of the face through the bounding box;
4. Preprocessing of the image before the face landmarks detection;
5. Face landmarks detection: coordinates extraction of the landmarks in the bounding box of the face;
6. Screen visualisation.

Follows a detailed explanation of the implemented algorithm. First, when the application is started, the models for the detection of the face and landmarks are loaded. The current frame is captured from the video stream and treated as an image (Step 1).

The face detector input image was reduced by a factor of 4 in both dimensions and converted to grayscale. The reasons for scaling are to guarantee the face detector a correct processing speed with the smartphone used for the tests.

Processing small images is faster but if the images are too small, the face detector can fail (Step 2).

The algorithm used for face detection is Haar cascade, already integrated into OpenCV. In particular, for the frontal detection of the face, we use the trained model ("haarcascade frontalface default.xml"). Haar feature-based cascade classifier is an object detection method proposed by [33]. This is a machine learning-based approach where a model is trained on datasets of many face images [25]. A 24x24 size window scrolls over the input image. As in a convolutional kernel, the features of Haar are extracted. Cascade indicates that features are extracted on cascaded stages. Finally, an Adaboost classifier combines the extracted features to discriminate face and background. In Algorithm 1, this instruction is executed by the "detectMultiScale" function where, given as an input a grayscale image, returns the coordinates and dimensions of the bounding box of detected faces. It has been necessary to set different parameters as follow: "scaleFactor" that indicates how much the size of the input image is reduced to create the pyramid; this parameter is set to 1.1.

Another parameter is the "minNeighbors" which specifies how many candidate rectangles must have a face around it to be considered as such; the value of this parameter is set to 2. Another parameter set is "minSize", which defines the smallest face size that can be detected and has been set to (30,30). The output of the face detector is a list of rectangles, where the coordinates of the vertices of the rectangles identify the window that contains the face. Since this application analyses only a face in the frame, among the various possible bounding boxes, the one with the largest size is selected and considered the correct one (Step 3). The image of the current frame has been processed before putting it in input to the face landmarks detector (Step 4).

Given the region of the face, the following step is face landmarks detection. The landmark detector used is Ensemble of Regression Trees [34], already implemented in dlib library.

In particular, the model "shape predictor 68 face landmarks.dat" extracts 68 landmarks of each face (as shown in Fig. 3), belonging to the following elements of the face:

- Eye (right and left),
- Eyebrow (right and left),
- Nose,
- Mouth,
- Jaw.

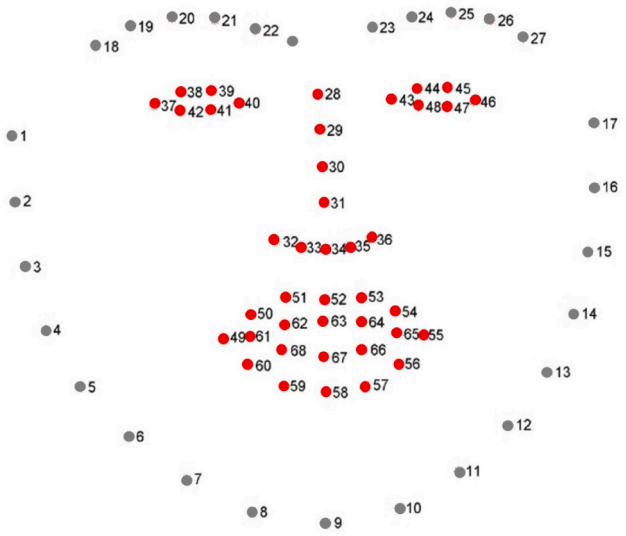

**Fig. 3.** Representation of the 68 facial landmarks extracted by the algorithm as presented for the Multi-PIE/IBUG 300-W dataset. Points used in the study in red and points discarded in grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The model is trained on iBUG 300-W dataset, which contains images of manually labelled faces with the coordinates $(x,y)$ of 68 reference points [35]. In the code, the face landmarks detection is executed by the "shape predictor" method, which has the image and the list of the bounding boxes of each face (in our case, only one bounding box is passed to the landmark prediction method) in input. Even if in the workflow a preprocessing phase is inserted (for generality) before the face landmark detection, this is not really executed in the code. The face landmarks detection directly processes the image acquired by the camera, without resizing and/or colour transformation. The reduction of the image dimension allows a greater speed of processing but a decrease in the accuracy of the extracted landmarks (Step 5).

Finally, the current frame is marked with the bounding box and the landmarks of the detected face. The marked frame is displayed on the smartphone screen (Step 6).

An encoding of the developed algorithm for face landmark detection is presented in Algorithm 1.

**Algorithm 1** Face landmark detection

1: **for** each frame of the *stream video* **do**
2: *image* ← frame acquisition;
3: $image_a$ ← preprocessing of the *image*;
4: *faceboundingbox* ← face detection: localisation of the face in the $image_a$;
5: $image_b$ ← preprocessing of the *image*;
6: *face landmarks coordinates* ←face landmarks extraction: coordinates detection of the face landmarks in the *face bounding box* on the $image_b$;
7: **end for**

The code has been developed in Java, to be executed on Android. It uses two open-source libraries: i) OpenCV 4.0.1 and ii) dlib 19.16. The dlib library is not native to Java but is written in C++. For this, the methods that use the functions of the dlib library have been implemented in C++ via the native Java JNI interface. The versions of the libraries may not be essential for the correct functioning of the application, but these two versions have been tested and guarantee correct operation. The application was developed in the Android Studio programming environment.

### 3.4. The creation of the 3D clouds

The position of the markers and face landmarks in each image is sent to a server for processing. There are no differences in format between face landmarks and markers, so no registration or data integration is

required. The use of targets instead of images drastically reduces the data to be sent to the server. This allows for nearly real-time processing and avoids taking images of infants in a real case scenario, which facilitates maintaining high data protection standards.

The 3D point clouds (Fig. 4) are created using ad-hoc software combined with MicMac [36]. Self-calibration is carried out during the creation of the 3D point cloud. Measurements with photogrammetric reproyection errors above 1 pixel are immediately discarded. Different calibration approaches were studied in a previous publication by the authors [16] from which it was demonstrated that the self-calibration approach with a significant number of images was proven to provide enough accuracy for evaluation of head deformation while maintaining the applicability of the tool for clinical (real) medical conditions.

The obtained clouds are registered to a known coordinate system (Fig. 4) using the registration points given by the three stickers placed between the eyes and in the preauricular points.

The scaling of the cloud is carried out using the known size of the markers. Although the markers are small and slight deformations can occur, the high number of measurements among coded targets assures accurate scaling. The length of two diagonals for each of the 131 markers on the cap are taking into account and an iterative least-squares adjustment is carried out, so measurements with high deformation can be discarded.

### 3.5. Comparison

For each dummy, different point clouds were registered using the points given by the stickers. For each point, in the cap or face, the coordinates were extracted for each point cloud. The variability of these coordinates was analysed to obtain the precision of each point. The number of times each point was measured is also registered and counted.

The accuracy of the tool for the 3D reconstruction of the ArUco points was known to be better than 2 mm based on tests carried out in previous works [14]. The creation of ground truth for facial landmarks was discarded due to the difficulty to manually identify the same landmarks detected by the algorithm. As a consequence, a study on the variability of those facial landmarks was identified as the best approach to test the precision in both image and object spaces.

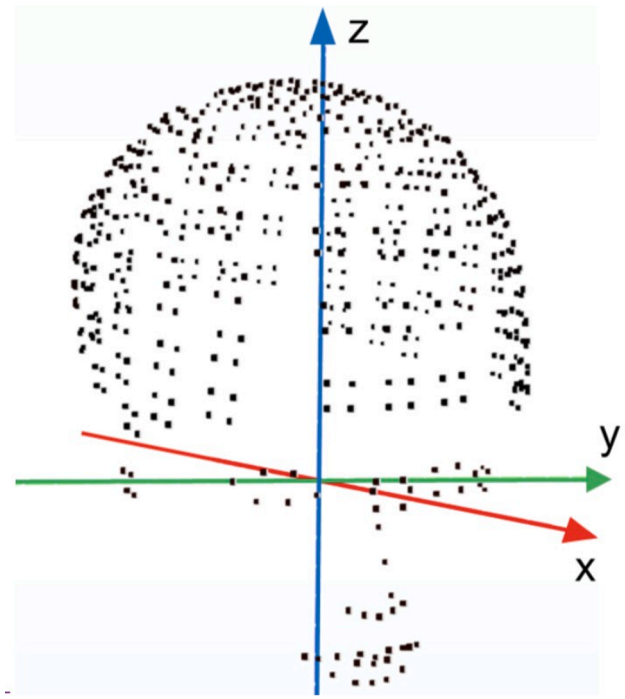


**Fig. 4.** Registered point cloud and coordinate system.

## 4. Results

To evaluate the performance of the implemented application, five 3D copies of different faces (dummies) have been considered. The 5 dummies have been labelled as G, I, J, K, L. On each dummy, the application has been executed 10 times. For each time that the application has been executed, the measurements extracted from each are processed to obtain the corresponding 3D points cloud, from which 3D models might be eventually delivered. Each set consists of 3D points saved on a text file with "PLY" format: each row represents a point and contains the three coordinates X, Y, Z and three attributes: R, G, B (Fig. 5).

During the experimental phase, 4 sets of points have been considered:

- Cap points corresponding to 4 vertices of 131 coded markers, yielding a total of 524 points.
- Face points corresponding to 41 of 68 landmarks obtained from the face landmarks detection algorithm. All the selected points correspond to well-identified points that do not depend on the point of view. Therefore, points corresponding to face limit and eyebrows were discarded due to Their instability measured with the standard deviation (std).

- Reduced sets were made for:

– 12 points of the eyes (6 for the right eye and 6 for the left eye);
– 9 points for nose;
– 20 points for mouth.

Additional 12 points are present in the file; corresponding to 4 vertices of the 3 adhesive markers, placed near the right, left ears and base of the nose. These points are used only when registering templates; no assessments have been made on these points. It may happen that some points in the templates are missing. The causes are many: 1) they were not detected in enough "frames" during recording; 2) the convergence was bad; or 3) the marker was moved (less likely). Especially for the points of the face, it was actually possible to observe that there was a lot of variability in the landmarks detected during the measurements.

The analysis of the variability for each dummy and each set of 3D point clouds (cap, face, eyes, right and left eye, nose, mouth) was performed. Further considerations were made on the missing points for each set of points. The std of each point among the 10 executions was calculated for each dummy. Since in some executions there are missing points, the std is calculated only between those points effectively detected, as long as the point is detected in at least two sets. In particular, for each point, std is determined for each coordinate (X,Y,Z) of the point and its module $|p|$. Std of $|p|$ are calculated as Euclidean norm of the three coordinates: X,Y,Z. For this reason, std of $|p|$ is always bigger than std of each coordinate (Fig. 6). For each point, it has been evaluated the number of executions in which it was detected. To obtain the overall assessments of the std in different sets of points (after each execution), the arithmetic average between the std of each point belonging to the set of interest is made. Only points that are detected in at least two executions are included in this average. This is to exclude the points where the std is not assessable. The same arithmetic average is used for overall assessments of the std between the 5 dummies in different sets of points.

It is worth noting that, with respect to Table 1, in which the measurements are mediated across the points of each set of each dummy, Fig. 6 shows the measurements mediated across each point of each dummy. From a more in-depth analysis, it has been observed that the points detected on the cap are 100% stable: all points are detected in almost all executions. eye point detection is less stable (70%) but still more stable than other face points. Besides, from Table 1 (column valid points), it can be seen how the eye points are always detected in at least two executions for each dummy, all being available for making overall

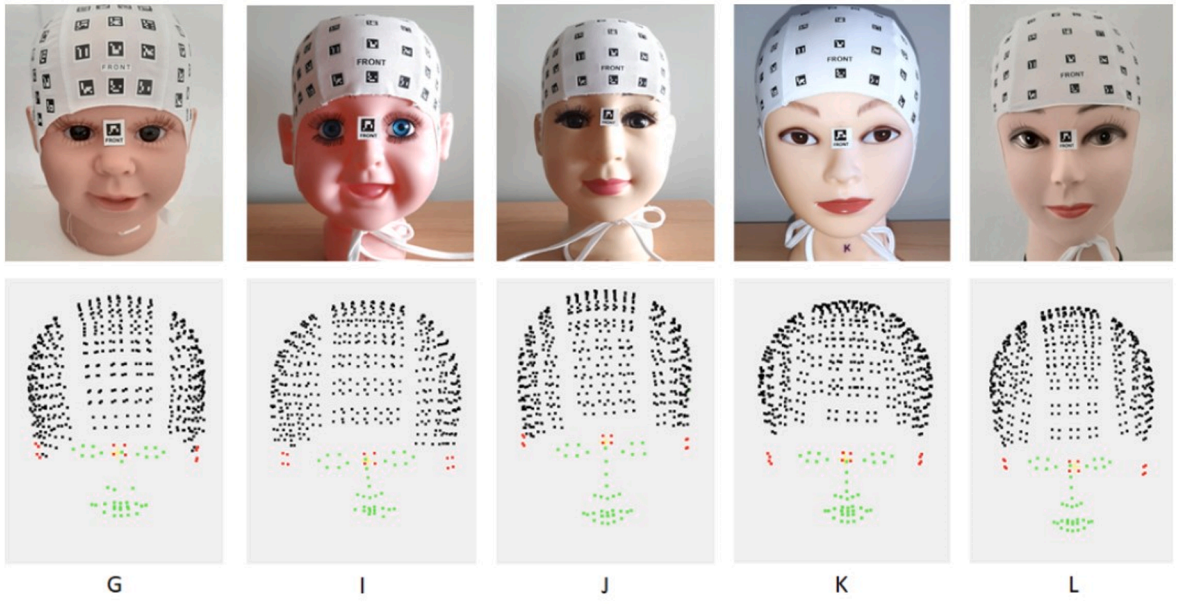


**Fig. 5.** Dummies and 3D point cloud for each one. Cap points in black, face landmark points in green and registration points in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

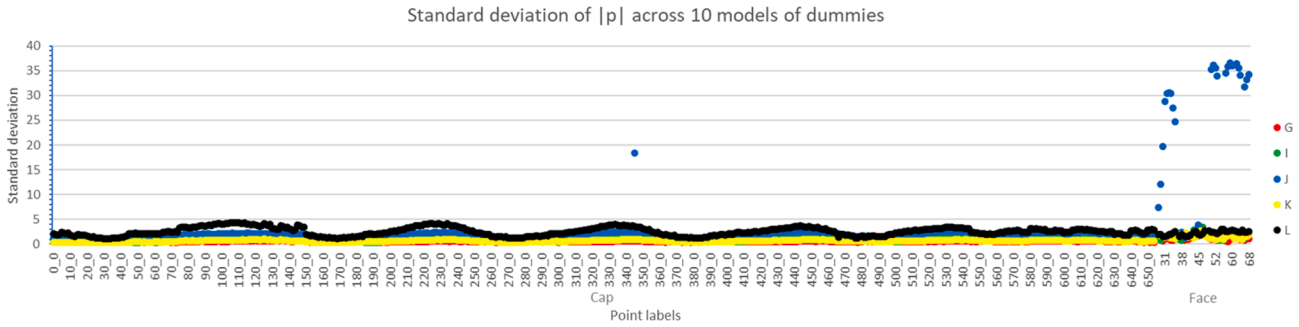


**Fig. 6.** Standard deviation of $|p|$ across 10 executions of dummies.

averages. This does not happen for the other points of the face (see Fig. 7). The std among the executions in the cap point set is always (except one) lower than the std in the face point set, in each dummy and each coordinate, a minimum std ranging from 0.4 and 2.4 mm, and an overall mean of 1.2 mm. After all, we expected it, as already from the acquisitions, greater stability of the points on the markers of the cap was observed rather than on the facial landmarks. Globally, it is observed that the std on the X coordinate is always greater than the std on the other Y and Z coordinates. It is due to the orientation of the camera during the acquisition. In our coordinate system, the X-axis follows the orthogonal direction of the front face. Since the face points can only be detected at a small angle to this X-axis, the geometry (namely the parallax angle) is worse and thus variability increases in this direction. Cap points are also affected by this problem, but much more slightly as the parallax angle at which they can be detected is greater. Consequently, they have greater variability in this X-axis direction, 14% and 33% more than in Y-axis and Z-axis, respectively.

## 5. Discussion

The results obtained during the experimental phase show an important potential of automatically detected face landmarks for overall head 3D point cloud measurement and extraction. This method is expected to be used for the evaluation of infant's cranial deformation as well as the detection of syndromes and other serious alterations and pathologies in which the face configuration and symmetry help doctors to diagnose infants. More detailed analysis is expected to be drawn from the following figures.

### 5.1. Evaluation of the obtained precision

The precision of 3D point clouds obtained from facial landmarks has been compared to cap points, generated from automatically detected target-based coded markers (Fig. 6). The precision of the cap points is close to 1 mm, as was expected from previous studies [14]. For the face points, the precision varies greatly between the different parts (Figs. 8 and 9). The eyes provide the best results with a precision that yields std values around 2 mm, while the mouth and nose have higher variability after checking their std (8.1 mm and 5.7 mm, respectively), probably due to the presence of outliers and lack of detection of the facial landmarks (Fig. 8). In order to be valuable for most applications, the methodology requires a precision similar or below the minimum difference noticeable by a layperson. According to some authors, the minimum noticeable asymmetry is 2 mm for dental maxillary shift and 4 mm for nose deviation [37], the minimum noticeable eye canting is approximately 2 mm [38]. For non-asymmetric problems, such as hypotelorism or hypertelorism, authors have defined that a layperson can detect when the intercanthal distance is 10% larger than usual [39], which results in 3 mm for the average adult [40] and above 2 mm for infants [41].

According to these thresholds, the methodology capacity to detect

**Table 1**
Results of std in different sets of points (mm).

| Dummy ID | Set of Points | N. Valid Points * | (%) Null Points ** | Standard Deviation (mm) | | | |
|---|---|---|---|---|---|---|---|
| | | | | X | Y | Z | \|p\| |
| G | Cap | 524 | 0.6 | 0.4 | 0.3 | 0.2 | 0.5 |
| | Face | 40 | 30.0 | 1.0 | 0.4 | 0.4 | 1.1 |
| | Nose | 8 | 51.1 | 0.8 | 0.4 | 0.4 | 1.0 |
| | Eyes | 12 | 15.0 | 1.2 | 0.4 | 0.5 | 1.3 |
| | Right | 6 | 0.0 | 0.8 | 0.6 | 0.4 | 1.1 |
| | Eye | 6 | 30.0 | 1.6 | 0.3 | 0.6 | 1.7 |
| | Left Eye | 20 | 29.5 | 0.9 | 0.4 | 0.4 | 1.1 |
| | Mouth | | | | | | |
| I | Cap | 524 | 0.2 | 0.4 | 0.6 | 0.4 | 0.9 |
| | Face | 22 | 75.6 | 1.3 | 0.5 | 0.6 | 1.6 |
| | Nose | 2 | 85.6 | 0.7 | 0.3 | 0.7 | 1.0 |
| | Eyes | 12 | 50.8 | 1.6 | 0.8 | 0.7 | 1.9 |
| | Right | 6 | 46.7 | 0.8 | 0.5 | 0.7 | 1.2 |
| | Eye | 6 | 55.0 | 2.4 | 1.0 | 0.8 | 2.7 |
| | Left Eye | 8 | 86.0 | 1.1 | 0.1 | 0.5 | 1.2 |
| | Mouth | | | | | | |
| J | Cap | 524 | 0.9 | 1.1 | 1.1 | 0.7 | 1.7 |
| | Face | 35 | 56.3 | 20.3 | 2.4 | 3.2 | 20.7 |
| | Nose | 9 | 53.3 | 23.2 | 2.6 | 2.0 | 23.5 |
| | Eyes | 12 | 14.2 | 1.9 | 0.7 | 0.5 | 2.1 |
| | Right | 6 | 13.3 | 1.7 | 0.5 | 0.4 | 1.8 |
| | Eye | 6 | 15.0 | 2.2 | 0.9 | 0.6 | 2.4 |
| | Left Eye | 14 | 83.0 | 34.1 | 3.8 | 6.3 | 34.9 |
| | Mouth | | | | | | |
| K | Cap | 524 | 0.2 | 0.4 | 0.3 | 0.3 | 0.6 |
| | Face | 41 | 35.1 | 1.4 | 0.3 | 0.4 | 1.5 |
| | Nose | 9 | 37.8 | 1.4 | 0.2 | 0.4 | 1.4 |
| | Eyes | 12 | 16.7 | 1.8 | 0.3 | 0.4 | 1.9 |
| | Right | 6 | 33.3 | 1.7 | 0.4 | 0.4 | 1.8 |
| | Eye | 6 | 0.0 | 1.9 | 0.3 | 0.4 | 2.0 |
| | Left Eye | 20 | 45.0 | 1.2 | 0.3 | 0.5 | 1.3 |
| | Mouth | | | | | | |
| L | Cap | 524 | 0.2 | 1.6 | 1.1 | 1.4 | 2.4 |
| | Face | 41 | 67.6 | 1.9 | 0.7 | 0.9 | 2.2 |
| | Nose | 9 | 71.1 | 1.7 | 0.5 | 0.7 | 1.9 |
| | Eyes | 12 | 44.2 | 1.7 | 0.6 | 0.4 | 1.9 |
| | Right | 6 | 38.3 | 1.2 | 0.9 | 0.6 | 1.6 |
| | Eye | 6 | 50.0 | 2.2 | 0.4 | 0.3 | 2.2 |
| | Left Eye | 20 | 80.0 | 2.1 | 0.8 | 1.3 | 2.6 |
| | Mouth | | | | | | |
| **Overall mean** | Cap | | 0.4 | 0.8 | 0.7 | 0.6 | 1.2 |
| | Face | | 52.9 | 5.2 | 0.9 | 1.1 | 5.4 |
| | Nose | | 59.8 | 5.6 | 0.8 | 0.8 | 5.7 |
| | Eyes | | 28.2 | 1.6 | 0.6 | 0.5 | 1.8 |
| | Right | | 26.3 | 1.2 | 0.6 | 0.5 | 1.4 |
| | Eye | | 30.0 | 2.0 | 0.6 | 0.5 | 2.2 |
| | Left Eye | | 64.7 | 7.9 | 1.1 | 1.8 | 8.1 |
| | Mouth | | | | | | |

* Number of valid points to calculate std across 10 executions in a point set. Only points that are detected in at least two executions are included in this number.
** Frequency, normalized, as a percentage, of the points of the set not detected in the 10 executions.

asymmetries in the eyes is similar to the capacity of a layperson for all models. For the nose and mouth, the precision is similar to the those of the eyes and to the threshold detectable by a layperson, with the exception of model J.

The high deviations of model J were found to be caused by a particular measurement with high error in the x-axis, that includes only nose and mouth. The error is caused as the images covering the face have a very small paralax angle. A further development of the tool should include a methodology to identify and remove this high error measurements.

To accurately measure any deformation or evaluate changes in the patients face it would be required a tolerance below 1/3 of the measured value, which results in precision below 0.7 mm for the eyes and mouth and below 1.3 mm for the nose. The proposed method, albeit encouraging, presents some shortcomings that are fair to discuss. An important part of the error is given by the X-coordinate (0.8 mm), with better precision in the Y-coordinate (0.7 mm) and considerably better in the Z-axis (0.6 mm). This is explained by the orientation of the model, with the face normal matching the X-axis. As the face detection algorithm is very dependent on the angle of view, the geometry of the images results in a higher error in the X-direction. This fact has to be taken into account for future medical diagnostics as the methodology will provide better results for the evaluation of facial asymmetry and worse to evaluate deepen areas of the face.

### 5.2. Robustness and limitations

The 3D coordinates of the face points are successfully obtained a much lower percentage of times, in comparison to cap points. Cap points are registered more than 99% of the times while face points are registered successfully only in 48.6% of the cases (Fig. 7). An important difference between dummies is also noticed. The face landmark detection algorithm performs differently for different faces, with models of a dummy that have a much lower number of face points (missing most of the mouth and nose points). Other limitations, strictly related to the landmark detection algorithm, can be summarised as follow: first, it does not work with incomplete parts of the face; this means that the facial landmark detector is bound to extract 68 landmarks in each bounding box (even if it contains half a face). Secondly, the landmark detector does not work without face detection first. In this case, the landmarks are constrained to be extended over the entire image area.

### 5.3. Open discussion and future developments

Given the above, we can conclude that it can be foreseen to use automatic face detection algorithms to create face 3D point clouds, envisaging future lines of research. The obtainment of good results is hampered by the large variability and reliability in the detection of the points and the limitations in the angle of view, which leads to poor geometry of the images. The presence of outliers seems to be affecting greatly the overall accuracy of the points. Removing the outliers either during the real-time photogrammetric data acquisition or during post-processing with robust estimators could lead to considerably better results. But this analysis has not been developed yet and is open for future

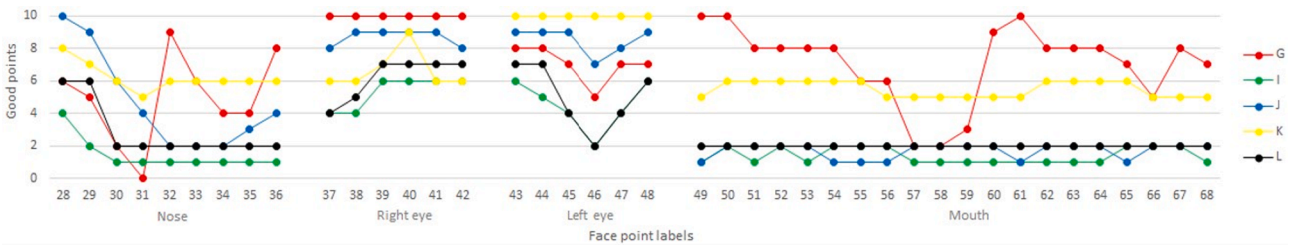


**Fig. 7.** Good points of the face across 10 executions of dummies.

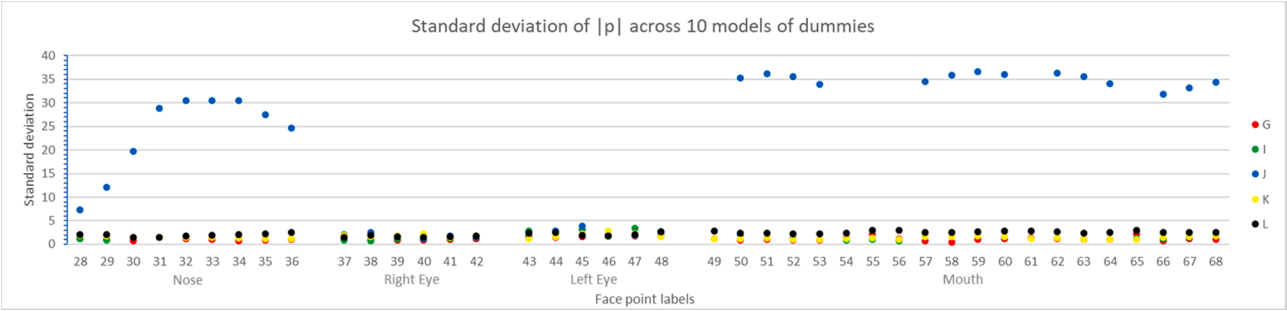


**Fig. 8.** Standard deviation of face $|p|$ across 10 executions of dummies. This figure relates to Fig. 6 by focusing on the face points.

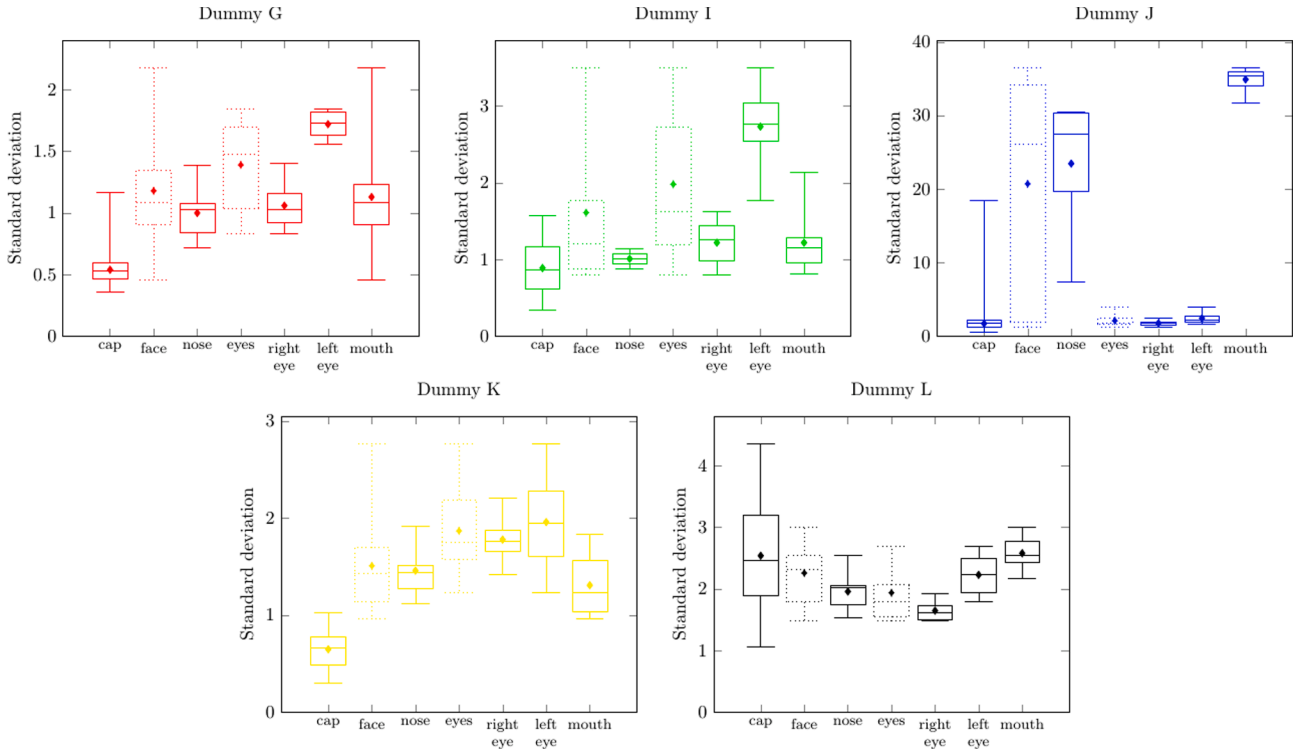


**Fig. 9.** Box plots summarizing the results of the std reported in Figs. 6 and 8. The box plots report the max and min values, boundaries are set for the first and third quartiles, central lines and central dots represent the median and mean, respectively.

experiment discussion. Besides, another improvement would consist of assuring a good geometry during the data acquisition by guiding the user, so the parallax angle between face imagery is maximized. The use of other machine learning algorithms, that provide better precision, particularly for the detection of nose and mouth points should also be explored in the future. The use of other machine learning algorithms, that provide better precision, particularly for the detection of nose and mouth points should also be explored in the future. The investigation of the scalability of the system and the development of the general framework and guidelines for easier serialization will be addressed and the training procedure will be also optimized. We aim to scale up the proposed approach in a more challenging dataset, e.g., high unbalanced setting with a huge amount of heterogeneous data. In fact, as the volume of the input increases, the continuous update of the machine learning classifier may disclose a challenge due to the problem of the "curse of dimensionality" and high in-memory search cost [42]. It is planned to optimize the features selection techniques in order to deal with this challenge while preserving the interpretability of the model. Additionally, it could be useful the design of new classifiers to model the different nature of the extracted features while decreasing the generalisation error.

Another aspect is the improvement of the machine learning algorithms performance by integrating more complex deep learning architectures combined with a generalisation procedure to test the trained networks in other classes.

The presented computer vision and photogrammetric tool that makes use of both natural and artificial marks, has the potential to become a useful methodology for the creation of 3D point clouds of the overall head that could be used to detect and quantify infants' abnormalities in proportions or symmetry (as soon as better estimations are confirmed for the face landmarks). The methodology has important advantages, inherited from the previous head modelling tool PhotoMeDAS. The tool is designed to work well with moving people (a particular feature of infants and children with psychological alterations). Furthermore, anybody can carry out the data acquisition and the anonymity of the people measured is granted and the requirements for sending the data to a server for processing and storage are low. Nevertheless, it is still to be tested how changes in facial expressions during the acquisition will affect the creation of a comprehensive infant's head point cloud.

## 6. Concluding remarks

This paper presents a novel approach that integrates machine learning facial landmarks and close-range photogrammetry coded targets in real-time during data acquisition to measure head (cranium vault plus face) 3D point clouds.

The precision of the points varies depending on the nature (natural and artificial) of the marks and on the part of the face. Without any doubt, artificial coded targets yield the best results with a precision of around 1 mm in the three axes. As regards the second-best, the right eye yields an overall precision of 1.4 mm. The inclusion of the left eye degrades slightly the results, in a way that eyes provide a precision of 1.8 mm, which is below the threshold for a layperson detecting the facial abnormality. In conclusion, the 3D coordinates of the eye points are able to identify facial abnormalities from an aesthetical point of view. On the contrary, the points in the nose and mouth present a higher number of outliers and a considerable worse precision. However, the nose points precision would be enough to improve the recognition of abnormality by a layperson in most cases. These results should be confirmed on real infants. The largest error is in the direction of the X-axis, due to the limitations in the angle of view for the facial landmarks detection algorithm. In general, it can be stated that the 3D coordinates of facial landmark points were not successfully obtained in an important percentage of cases (52.9%) and their precision is similar to a layperson detection capacity for the eye points and slightly worse for the nose points. However, further improvements either in the data acquisition stage or in the facial landmarks detection algorithms are required in order to increase the reliability of the computer vision and photogrammetric integration and the precision of the overall methodology. The machine learning and close-range photogrammetric methodology adaptation presented in this paper is an extension of a previously patented automatic tool for the creation of 3D models of infant's heads. The addition of facial landmarks to automatic cranial deformation measurement will provide relevant information to precisely quantify facial abnormalities and will also improve the evaluation and diagnosis of related cranial deformation issues, which are very difficult to measure on dynamic infants using conventional devices during consultation.

## CRediT authorship contribution statement

**Inés Barbero-García:** Software, Methodology, Investigation, Writing - original draft. **Roberto Pierdicca:** Software, Data curation, Writing - original draft, Writing - review & editing. **Marina Paolanti:** Software, Data curation, Writing - review & editing. **Andrea Felicetti:** Software, Formal analysis, Visualization. **José Luis Lerma:** Conceptualization, Validation, Writing - review & editing, Supervision, Funding acquisition.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

*Acknowledgement*

This work was supported by the Instituto de Salud Carlos III and European Regional Development Fund (FEDER), project number PI18/00881.

The authors would like to thank Gaspar Mora-Navarro for his invaluable effort in allowing 3D models computation.

## References

[1] G. de Jong, E. Bijlsma, J. Meulstee, M. Wennen, E. van Lindert, T. Maal, R. Aquarius, H. Delye, Combining deep learning with 3D stereophotogrammetry for craniosynostosis diagnosis, Sci. Rep. 10 (2020) 15346, https://doi.org/10.1038/s41598-020-72143-y.

[2] H. Schaaf, C.Y. Malik, P. Streckbein, J. Pons-Kuehnemann, H.-P. Howaldt, J.-F. Wilbrand, Three-dimensional photographic analysis of outcome after helmet treatment of a nonsynostotic cranial deformity, J. Craniofac. Surg. 21 (2010) 1677–1682, https://doi.org/10.1097/SCS.0b013e3181f3c630.

[3] L. Argenta, Clinical classification of positional plagiocephaly, J. Craniofac. Surg. 15 (3) (2004) 368–372, https://doi.org/10.1097/00001665-200405000-00004.

[4] M.H. Siegenthaler, Methods to diagnose, classify, and monitor infantile deformational plagiocephaly and brachycephaly: a narrative review, J. Chiropr. Med. 14 (3) (2015) 191–204, https://doi.org/10.1016/j.jcm.2015.05.003.

[5] E. Ballardini, M. Sisti, N. Basaglia, M. Benedetto, A. Baldan, C. Borgna-Pignatti, G. Garani, Prevalence and characteristics of positional plagiocephaly in healthy full-term infants at 8–12 weeks of life, Eur. J. Pediatr. 177 (10) (2018) 1547–1554, https://doi.org/10.1007/s00431-018-3212-0.

[6] S. Nahles, M. Klein, A. Yacoub, J. Neyer, Evaluation of positional plagiocephaly: Conventional anthropometric measurement versus laser scanning method, J. Cranio-Maxillofacial Surg. 46 (1) (2018) 11–21, https://doi.org/10.1016/j.jcms.2017.10.010.

[7] J.L. Lerma, I. Barbero-García, Á. Marqués-Mateu, P. Miranda, Smartphone-based video for 3D modelling: application to infant's cranial deformation analysis, Measurement. 116 (2018) 299–306, https://doi.org/10.1016/j.measurement.2017.11.019.

[8] J. van der Meulen, Metopic synostosis, Child's Nerv. Syst. 28 (9) (2012) 1359–1367, https://doi.org/10.1007/s00381-012-1803-z.

[9] G. Maltese, P. Tarnow, C.G. Lauritzen, Spring-assisted correction of hypotelorism in metopic synostosis, Plast. Reconstr. Surg. 119 (3) (2007) 977–984, https://doi.org/10.1097/01.prs.0000252276.46113.ee.

[10] M. Hansen, B.L. Padwa, R.M. Scott, P.E. Stieg, J.B. Mulliken, Synostotic frontal plagiocephaly: anthropometric comparison of three techniques for surgical correction, Plast. Reconstr. Surg. 100 (6) (1997) 1387–1395, https://doi.org/10.1097/00006534-199711000-00002.

[11] C. Linz, A.B.M. Gerdes, P. Meyer-Marcotty, U. Müller-Richter, H. Böhm, R.-I. Ernestus, A. Kübler, G.W. Alpers, T. Schweitzer, Perception of children's faces with unilateral coronal synostosis—an eye-tracking investigation, Child's Nerv. Syst. 32 (1) (2016) 135–141, https://doi.org/10.1007/s00381-015-2798-z.

[12] F. Samra, J.T. Paliga, Y. Tahiri, L.A. Whitaker, S.P. Bartlett, B.J. Forbes, J. A. Taylor, The prevalence of strabismus in unilateral coronal synostosis, Child's Nerv. Syst. 31 (4) (2015) 589–596, https://doi.org/10.1007/s00381-014-2580-7.

[13] K.S. Gabrick, R.T. Wu, A. Singh, S.P. Bartlett, J.A. Taylor, J.A. Persing, M. Alperovich, Assessing facial asymmetry in postoperative patients with unilateral coronal craniosynostosis, J. Craniofac. Surg. 31 (2020) 1000–1005, https://doi.org/10.1097/SCS.0000000000006355.

[14] I. Barbero-García, J.L. Lerma, G. Mora-Navarro, Fully automatic smartphone-based photogrammetric 3D modelling of infant's heads for cranial deformation analysis, ISPRS J. Photogramm. Remote Sens. 166 (2020) 268–277, https://doi.org/10.1016/j.isprsjprs.2020.06.013.

[15] I. Barbero-García, J.L. Lerma, Á. Marqués-Mateu, P. Miranda, Low-cost smartphone-based photogrammetry for the analysis of cranial deformation in infants, World Neurosurg. 102 (2017) 545–554, https://doi.org/10.1016/j.wneu.2017.03.015.

[16] I. Barbero-García, M. Cabrelles, J.L. Lerma, Á. Marqués-Mateu, Smartphone-based close-range photogrammetric assessment of spherical objects, Photogramm. Rec. 33 (162) (2018) 283–299, https://doi.org/10.1111/phor.2018.33.issue-16210.1111/phor.12243.

[17] M. Martini, A. Klausing, M. Messing-Jünger, G. Lüchters, The self-defining axis of symmetry: a new method to determine optimal symmetry and its application and limitation in craniofacial surgery, J. Cranio-Maxillofacial Surg. 45 (9) (2017) 1558–1565, https://doi.org/10.1016/j.jcms.2017.06.023.

[18] Y. Li, B. Sun, T. Wu, Y. Wang, Face detection with end-to-end integration of a convNet and a 3D model, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Springer Verlag, 2016: pp. 420–436. https://doi.org/10.1007/978-3-319-46487-9_26.

[19] A. Kumar, A. Kaur, M. Kumar, Face detection techniques: a review, Artif. Intell. Rev. 52 (2) (2019) 927–948.

[20] C.-W. Wang, C.-T. Huang, M.-C. Hsieh, C.-H. Li, S.-W. Chang, W.-C. Li, R. Vandaele, R. Maree, S. Jodogne, P. Geurts, C. Chen, G. Zheng, C. Chu, H. Mirzaalian, G. Hamarneh, T. Vrtovec, B. Ibragimov, others, Evaluation and comparison of anatomical landmark detection methods for cephalometric x-ray images: a grand challenge, IEEE Trans. Med. Imaging. 34 (9) (2015) 1890–1900.

[21] Z.-H. Feng, J. Kittler, M. Awais, P. Huber, X.-J. Wu, Face detection, bounding box aggregation and pose estimation for robust facial landmark localisation in the wild, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Work., 2017, pp. 160–169.

[22] V.V. Khryashchev, A.A. Lebedev, A.L. Priorov, Enhancement of fast face detection algorithm based on a cascade of decision trees, Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XLII-2/W4 (2017) 237–241.

[23] L.F. Porto, L.N.C. Lima, M.R.P. Flores, A. Valsecchi, O. Ibanez, C.E.M. Palhares, F. de Barros Vidal, Automatic cephalometric landmarks detection on frontal faces: An approach based on supervised learning techniques, Digit. Investig. 30 (2019) 108–116.

[24] R. Vandaele, J. Aceto, M. Muller, F. Peronnet, V. Debat, C.-W. Wang, C.-T. Huang, S. Jodogne, P. Martinive, P. Geurts, others, Landmark detection in 2D bioimages for geometric morphometrics: a multi-resolution tree-based approach, Sci. Rep. 8 (2018) 1–13.

[25] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis, Deep learning for computer vision: a brief review, Comput. Intell. Neurosci. 2018 (2018) 1–13.

[26] Z. Fu, J. Jiao, M. Suttie, J.A. Noble, Cross-Task Representation Learning for Anatomical Landmark Detection, in: Int. Work. Mach. Learn. Med. Imaging, 2020: pp. 583–592.
[27] G.K.L. Rao, A.C. Srinivasa, Y.H.P. Iskandar, N. Mokhtar, Identification and analysis of photometric points on 2D facial images: a machine learning approach in orthodontics, Health Technol. (Berl) 9 (2019) 715–724.
[28] D. Yang, S. Zhang, Z. Yan, C. Tan, K. Li, D. Metaxas, Automated anatomical landmark detection ondistal femur surface using convolutional neural network, in: 2015 IEEE 12th Int. Symp. Biomed. Imaging, 2015: pp. 17–21.
[29] A. Abu, C.G. Ngo, N.I.A. Abu-Hassan, S.A. Othman, Automated craniofacial landmarks detection on 3D image using geometry characteristics information, BMC Bioinformatics 19 (2019) 65–80.
[30] Y. Wu, C. Gou, Q. Ji, Simultaneous facial landmark detection, pose and deformation estimation under facial occlusion, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 3471–3480.
[31] X. Dong, Y. Yan, W. Ouyang, Y. Yang, Style aggregated network for facial landmark detection, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 379–388.
[32] S. Garrido-Jurado, R. Muñoz-Salinas, F.J. Madrid-Cuevas, M.J. Marín-Jiménez, Automatic generation and detection of highly reliable fiducial markers under occlusion, Pattern Recognit. 47 (2014) 2280–2292, https://doi.org/https://doi.org/10.1016/j.patcog.2014.01.005.
[33] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition. CVPR 2001, 2001, pp. I–I.
[34] V. Kazemi, J. Sullivan, One millisecond face alignment with an ensemble of regression trees, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2014, pp. 1867–1874.
[35] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, M. Pantic, 300 Faces In-The-Wild Challenge: database and results, IMAVIS 47 (2016) 3–18, https://doi.org/10.1016/j.imavis.2016.01.002.
[36] M. Pierrot Deseilligny, I. Clery, Apero, an open source bundle adjusment software for automatic calibration and orientation of set of images, ISPRS – Int.Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XXXVIII-5/ (2012) 269–276, https://doi.org/10.5194/isprsarchives-XXXVIII-5-W16-269-2011.
[37] B. Pereira Silva, E. Jimenez-Castellanos, R. Martinez-de-Fuentes, J. Greenberg, S. Chu, Laypersons' perception of facial and dental asymmetries, Int. J. Periodontics Restorative Dent. 33 (2013) e162–e171, https://doi.org/10.11607/prd.1618.
[38] Kyoung-Ho Kwak, Yong-Il Kim, Hyung-Jin Nam, Seong-Sik Kim, Soo-Byung Park, Woo-Sung Son, Differences among deviations, genders, and observers in the perception of eye and nose asymmetry, J. Oral Maxillofac. Surg. 73 (8) (2015) 1606–1614, https://doi.org/10.1016/j.joms.2015.02.009.
[39] S. Naran, A.M. Wes, D.M. Mazzaferro, S.P. Bartlett, J.A. Taylor, More than meets the eye, J. Craniofac. Surg. 29 (2018) 40–44, https://doi.org/10.1097/SCS.0000000000004072.
[40] Roberto L. Barretto, Robert H. Mathog, Orbital measurement in black and white populations, Laryngoscope. 109 (7) (1999) 1051–1054, https://doi.org/10.1097/00005537-199907000-00007.
[41] Garrett M. Pool, Ryne A. Didier, Dianna Bardo, Nathan R. Selden, Anna A. Kuang, Computed tomography-generated anthropometric measurements of orbital relationships in normal infants and children, J. Neurosurg. Pediatr. 18 (2) (2016) 201–206, https://doi.org/10.3171/2016.2.PEDS15475.
[42] C. Yang, X. Yu, Y. Liu, Continuous KNN join processing for real-time recommendation, in: 2014 IEEE Int. Conf. Data Min., 2014, pp. 640–649.