Statistics based landmark selection model for cone-beam CT derived three-dimensional cephalometry

Adrienn Dobai[1], **Zsolt Markella**[2], **Miklós Mezei**[2], **Tamás Vizkelety**[1]

[1] Department of Oral and Maxillofacial Surgery and Dentistry, Semmelweis University
[2] Kandó Kálmán Faculty of Electrical Engineering, Óbuda University, Budapest, Hungary
{dobai.adrienn_gyongyi,vizkelety.tamas}@med.semmelweis-univ.hu,
{markella.zsolt,mezei.miklos}@kvk.uni-obuda.hu

Abstract: Although some articles have already assessed the reliability of three dimensional (3D) cephalometric landmarks, but the results were questionable because most of them analized the landmarks using linear or angular measurements instead of the coordinates. Therefore, the aim of this study was to eliminate the mistakes of the 3D landmark selection by the means of statistics based landmark selection model and a practically useable decision tree. In our study three medical doctors – the "examiners" – identified 55 particular landmarks on 30 non–orthodontic Cone Beam Computed Tomography (CBCT) scans using the Cranioviewer software. The identification process has been done three times in order to increase the accuracy. Intraclass correlation coefficient and analysis of variance were applied to decrease the intra- and inter- examiner variability, while standard deviation (SD) and mean absolute difference (MAD) were used for characterization of landmark locations. Inaccurate coordinates were grouped according to both the intra- and the inter-examiner deviation of $\geq$ 1mm and the difference between the two statistical methods (SD vs. MAD). The intra-examiner identification errors were $\leq$ 1mm. The inter-examiner SD and MAD were $\geq$ 1mm except in cases of four landmarks with MAD and in cases of two with SD. Inter-examiner deviations were always higher than intra-examiner deviations. Standard deviation distorted more than mean absolute difference. Based on these result we have created a decision tree for landmark selection. Most of the coordinates belong to the landmarks can be reliably adapted to 3D cephalometric, but the statistics based decision model could be useful to eliminate mistakes in landmark selection as well. Since, the statistical rules are summarized like a decision tree it can be easily used in practice.

Keywords: Cephalometry, Planning Techniques, Computer-Generated 3D Imaging, Cone-Beam Computerized Tomography, Dimensional Measurement Accuracy

1 Introduction

Cephalometric analysis is widely used for orthodontic diagnosis and presurgical planning in orthognathic surgery. By applying CBCT three-dimensional volumes of the particular target appear. In this way almost all errors coming from two-dimensional (2D) radiography can be eliminated. These errors include the magnification, distortion and superimposition of anatomical structures [1] as well. CBCT allows more exact representation of craniofacial asymmetries. Although one disadvantage of CBCT is the relatively high radiation dose, however, thanks to low-dose modules the effective doses of modern CBCT devices have been decreased. Furthermore, for presurgical planning of orthognathic cases, many additional exposures such as periapical, occlusal and posteroanterior radiographies are needed in addition to panoramic and lateral cephalometric X-rays [2]. The sum of the effective doses of these multiple exposures can be higher than with CBCT. Therefore, in more complex cases the CBCT is a more recommended tool for the presurgical planning [3,4] than other methods with similar purposes.
Many cephalometric analytical protocols and norms have been created for 2D imaging. These norms are not proper for the three-dimensional cephalometric measurement because of methodological differences between the common lateral cephalogram and the CBCT. Thus, new cephalometric norms are required for the CBCT based 3D cephalometry. The selection of acceptable landmarks for 3D craniometry is the basis for cephalometric reliability and repeatability. While landmark identification errors have already been published in numerous articles [5] many of them use angles or lines and not factual reference points [6, 7], which could be the base for creating a three-dimensional cephalometric analysis. There are three important considerations in landmark selection, which were not assessed in the previous articles [5–7]:

1. In three-dimensional measurements not all coordinates of a landmark are as important as others from the analysis point of view. Hence, we should not only analyze the accuracy of a landmark, but rather its reliable coordinates.
2. Other decisions can be made about the reliability considering different statistical tests, such as the SD or MAD. Since, the previously published studies in this regard [5–7] have applied only SD or MAD, they did not represent the differences in the results of two calculations.
3. Many factors contribute to the reproducibility of reference points such as image quality, anatomic complexity, experience of the examiners, correct definitions and so on [8, 9]. As these effects – with the exception of the anatomical structure – can be reduced by modern technology and precise measurement methods. Namely, if we know the reasons of inaccurate landmark detection then we will be able to eliminate the errors.

Due to the purpose of this study were fourfold:

1. To assess the intra- and inter-examiner reliability of the landmarks;

2. To demonstrate the most and least determinable coordinates by the means of two statistical markers (SD and MAD);
3. To analyze the identification problem of the unreliable coordinates;
4. To use these findings in order to develop a decision tree of landmark selection for 3D cephalometric analysis.

The paper is structured as follows: first, we introduce the applied methodology including the selection process, landmark detection and applied tool for statistical analysis. After, the outcomes are presented. Then, the results are assessed in the discussion section. Finally, the conclusions of the research are presented.

2 Applied Methodology

2.1 CBCT Scan Selection

This research (number: TUKEB 2/2008) was approved by the ethical committee of Semmelweis University. Thirty (n=30) existing CBCT scans were finally selected, 19 females and 11 males (aged 18 to 30 years), who underwent CBCT scanning for non orthodontic reasons.
During the selection of CBCT images we have taken into account the ethnicity and the age of patients, the field of view, the quality of the images and the facial form. Thus, the scans have been selected by using the following inclusion and exclusion criteria:

Inclusion criteria

a) sufficient image quality
b) large field of view (FOV)
c) bilateral Class I molar relationship
d) European ethnicity

Exclusion criteria

a) filling on the contact points of molar teeth
b) diasthema or crowding
c) anatomic anomalies or skeletal asymmetries
d) evidence of previous or current orthodontic treatment

The CBCT scans were obtained by using an i-CAT Classic scanner (Xoran Technologies, Ann Arbour Michigan, USA) at 120KV and 36 mA. The FOV was 16cm (H) x 22cm (D), and the isotropic voxel size was 0.4 x 0.4 x 0.4 mm. The 30 heads were fixed with the standardized head-holder and scanned parallel to the Frankfurt horizontal.

2.2 Landmark detection

The data of CBCT were transferred as DICOM (Digital Imaging and Communications in Medicine) files into the Cranioviewer software (Cranioviewer, Budapest, Hungary) in which the landmark locations have been calculated. In the multiplanar view (MPR) three axes have been defined as follows: X moved from left to right, Y moved from front to back and Z moved from top to bottom. After the detection of the landmarks the Cranioviewer automatically determined the coordinates of each point in millimeters and it shows the distance from the center of the coordinate system. Four imaging procedures are available in the program: slice X-ray, CT, MIP (Maximum Intensity Projection), AMIP (Advanced Maximum Intensity Projection). We used the AMIP and MIP only for the rough approximation of landmarks but all points were marked accurately on the CT projection. After detection, the program is able to create a 3D frame view from the determined points. A total of 55 landmarks (11 unpaired and 22 paired) on the hard tissue have been selected on CBCT images based on the 3D definitions of each landmark – as it can be seen in Table 1. The landmark locations are also shown on a typical 3D CBCT scan in Figure 1 - 2.

The landmark identifications have been repeated three times inside one week long interval. The identification was done by three experienced medical doctors (two dentists and one radiologist). Because of the experiences and training background of the examiners were quite similar these had minimal effect on the identification error.

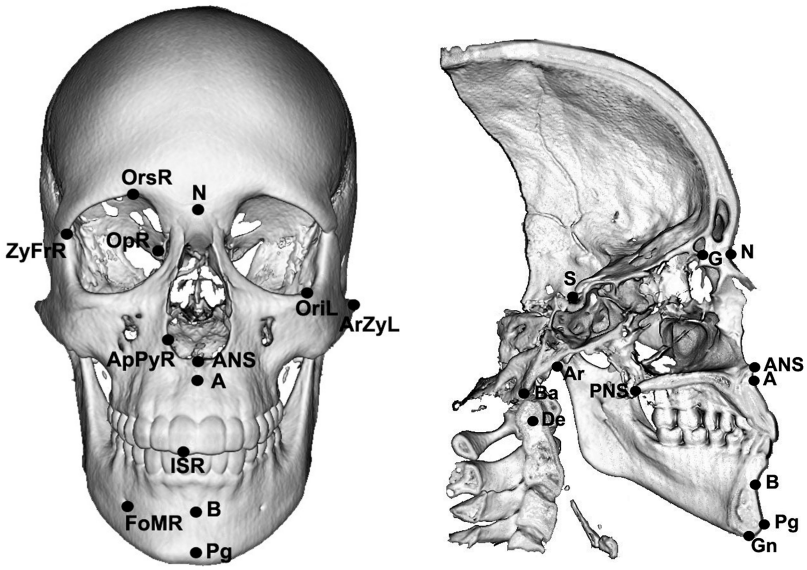

Figure 1
Position of cephalometric landmarks on the frontal and midsagittal view, respectively.

LANDMARK	DEFINITION
N (Nasion)	The crossing point of the sutura nasofrontalis and the sutura internasalis.
G (Galli)	Crossing point of the crista galli and the lamina cribrosa under the highest point of the crista galli.
S (Sella)	Crossing points of the sagittal and transversal diameter maximums of the cavum of the sella turcica where the transversal diameter is a constructed one determined by the proc. clinoideus ant. et post projected caudally to the level of the sagittal diameter.
Ba (Basion)	The most caudal and dorsal point of the clivus in the midplane of the skull.
De (Dens)	The crossing point of the greatest anteroposterior and transversal diameter of the dens axis.
ANS (Anterior nasal spine)	The most ventral end of the nasal floor, the tip of the spina nasalis anterior.
A (Point-A)	The deepest (most dorsal) point on the ventral surface curve of the maxilla in the sagittal midplane.
PNS (Posterior nasal spine)	The most dorsal extent of the hard palate in the midsagittal plane.
B (Point-B)	The deepest (most dorsal) point on the anterior curvature of the mandible symphisis in the midsagittal plane.
Pg (Pogonion)	The most ventral point along the curvature of the chin in the midsagittal plane.
Gn (Gnathion)	The most caudal point along the curvature of the chin in the midsagittal plane.
PtmR (Pterygomaxillare Right)	The contact point of the tuber maxillae with the pterygoid blade at the vertical level of the spina nasalis posterior on the right side.
ArR (Articulare Right)	A lateral view of the intersection of the external contour of the cranial base and the dorsal contour of the right ramus mandibulae.
ArZyR (Arcus Zygomaticus Right)	The most lateral/widest point of the zygomatic arch on the right side.
ZyFrR (Sutura Zygomaticofrontale Right)	The most ventral point of the right sutura zygomaticofrontalis.
ApPyR (Apertura Pyriformis Right)	The most lateral/widest point on the right side of the apertura pyriformis in the frontal view.
SemR (Semicircularis Right)	The most cranial point of the canalis semicircularis ant. seu sup. at the right side.
OriR (Orbitale inferius Right)	The most caudal point of the left orbita aperture.
OrsR (Orbitale superius Right)	The most cranial point of the right orbita aperture.
FmsR (Sutura Frontomaxillaris Right)	The distal end of the right sutura frontomaxillaris.
OpR (Opticus Right)	The intraorbital end of the canalis opticus at 9 o'clock at the right side of the orbit.
JR (Point-J Right)	The deepest point of the curvature formed at a junction of the anterior portion of the right ramus and corpus of the mandible.
PrCoR (Processus Coronoideus Right)	The most cranial point of the right processus coronoideus.
CmR (Condylus medialis Right)	Medial pole of the right condyle.
ClR (Condylus lateralis Right)	Lateral pole of the right condyle.
GoR (Gonion Right)	Bisection tangent of the lowest edge of the corpus and the most posterior edge of the ramus mandibulae.
IGoR (Inferior Gonion Right)	The most inferior point on the camber of the corpus mandibulae on the right side.
PGoR (Posterior Gonion Right)	The most dorsal point on the dorsal camber of the ramus mandibulae on the right side.
FoMR (Foramen mentale Right)	The most proximal/distal point of the right foramen mentale.
IIR (Incision inferior Right)	The mesiodistal midpoint of the incisor edge of the lower right incisor.
ISR (Incision superior Right)	The mesiodistal midpoint of the incisor edge of the upper right incisor.
IIAR (Incision inferior apicale Right)	The apex of the right lower incisor.
ISAR (Incision superior apicale Right)	The apex of the right upper incisor.

Table 1
Three-dimensional definitions of cephalometric landmarks.

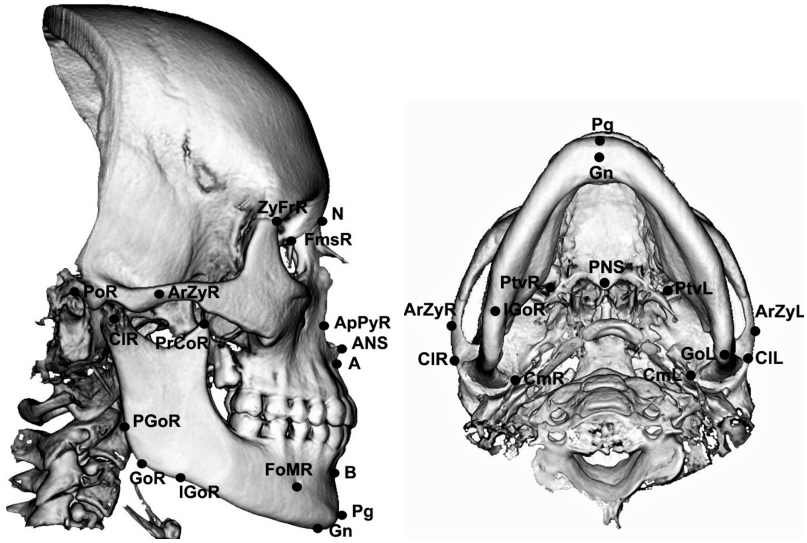

Figure 2
Position of cephalometric landmarks on the lateral and from the inferior view, respectively.

2.3 Statistical analysis

In order to establish intra- and inter-examiner reliability, values were assessed by an intraclass correlation coefficient (ICC) using SPSS statistical package 20.0 (IBM Corporation, Chicago, USA) [10]. The determination of the real position of the landmarks is impossible in case of in vivo research; therefore in order to evaluate the precision of landmark location we used descriptive statistics.

Similar to the previously published articles [2, 11, 12] in which SD or MAD have been applied, in this study we compared these two statistical estimations. Intra-examiner SD and MAD showed the repeatability of landmark identification and the inter-examiner SD and MAD characterized the reproducibility. We selected the most (SD or MAD was ≤ 0.2 mm) and less (SD or MAD was ≥ 1.0 mm) reliable coordinates in case of intra- and inter-examiner analysis, respectively. The inaccurate coordinates were divided into three groups based on the magnitude of SD of intra-examiner, inter-examiner or both deviations. By observing the differences we could determine the cause of difficulty regard to the unreliable landmark detection.The results are summarized and a decision model – including the main guidelines of landmark selection – is presented in the next section.

3 Results

In the followings we summarize the completed tests and provide a picture about the performance of the suggested technique.

Intra- and inter-examiner investigation: the reliability of the examiners was determined by calculating the ICC values for all landmarks of the X, Y and Z coordinates. To get a more detailed picture about the reliability of landmark positions, separate X, Y, Z coordinate values were used instead of three dimensional distances. The intraclass correlation coefficient was greater than 0.9 for all axes and for each landmark in intra-examiner measurements and it was greater than 0.9 in inter-examiner measurements.

Intra-examiner test for repeatability: in case of the intra-examiner test the standard deviation and the mean absolute difference of each landmark were assessed for each examiner trial. In most cases the values of SD were low ($\leq$1 mm), only by the following cases were between 1.03 and 2.00 mm: Orbitale Inferior (ob3:X-axis), bilateral Inferior Gonion (ob1,2,3:Y-axis), Orbitale Inferior (ob3:Y-axis), bilateral Posterior Gonion (ob1,2,3:Z-axis), J-point (ob2,3:Z-axis), Apertura Piriformis (ob3:Z-axis). The mean absolute differences were lower than 1 mm except the Orbitale Inferior (ob3:X-axis), bilateral Inferior Gonion (ob2,3:Y-axis), Apertura Piriformis (ob3:Z-axes) and bilateral Posterior Gonion (ob 1,2, 3:Z-axis).

Inter-examiner test for reproducibility: by using the inter-examiner test the SD and MAD were determined from the averaged intra-examiner trials of the 3 medical staff members (see Table 2.).

The inter-examiner standard deviation corresponded each landmark for X, Y and Z coordinates were evaluated from the average coordinates of the three investigators. For the most part the deviation was lower than 1mm. Only in case of the following structures was higher than 1 mm:

- On the X axis: Orbitale inferior right (3.44 mm) and left (3.56 mm), J-point right (1.55 mm) and left (1.77 mm).
- On the Y axis: Orbitale inferior right (2.44 mm) and left (2.23 mm), J-point right (1.8 mm) and left (1.78 mm), Inferior Gonion right (2.9 mm) and left (3.38 mm), Sutura Zygomaticofrontale right (1.32 mm) and left (1.28 mm).
- On the Z axis: Orbitale inferior right (2.27 mm) and left (2.25 mm), J-point right (3.37 mm) and left (3.34 mm), Sutura Frontomaxillare right (1,07 mm) bilateral Apertura Piriformis (2.59 mm; 2.56 mm) and Posterior Gonion (2.34 mm; 1.99 mm).

The measurement errors were also determined as the average of mean absolute differences for all coordinates for each landmark. The average mean differences on the X, Y and Z axis were bigger than 1 mm in these cases:

- On the X axis: Orbitale inferior right and left (2.6 and 2.64 mm), J-point right and left (1.15mm and 1.3mm)
- On the Y-axis: inferior Gonion right and left (1,68 and 2.8mm) and J-point right and left (1.34mm and 1.32mm), Orbitale inferior right and left (1.88mm and 1.8 mm)
- On the Z-axis: J-point right and left (2.14mm and 2.1mm), inferior Gonion right and left (1.71mm and 1.91mm)

Landmark	SD_X (mm)	SD_Y (mm)	SD_Z (mm)	MAD_X (mm)	MAD_Y (mm)	MAD_Z (mm)
N	0,21	0,24	0,33	0,16	0,18	0,20
G	0,29	0,86	0,58	0,22	0,64	0,54
S	0,38	0,24	0,29	0,28	0,17	0,20
Ba	0,23	0,39	0,27	0,16	0,26	0,28
De	0,18	0,23	0,36	0,13	0,16	0,21
ANS	0,21	0,31	0,26	0,16	0,22	0,21
A	0,19	0,15	0,55	0,14	0,11	0,27
PNS	0,23	0,28	0,31	0,17	0,18	0,24
B	0,23	0,18	0,52	0,17	0,12	0,22
Pg	0,26	0,16	0,58	0,19	0,13	0,24
Gn	0,30	0,39	0,24	0,23	0,31	0,22
PtmR	0,47	0,40	0,76	0,34	0,26	0,49
PtmL	0,44	0,43	0,78	0,33	0,28	0,52
ArR	0,31	0,22	0,49	0,24	0,18	0,32
ArL	0,36	0,35	0,48	0,27	0,27	0,36
ArZyR	0,14	0,66	0,54	0,10	0,47	0,44
ArZyL	0,15	0,84	0,68	0,11	0,62	0,48
ZyFrR	0,62	1,32	1,00	0,46	0,76	1,00
ZyFrL	0,44	1,28	0,84	0,33	0,74	1,00
ApPyR	0,42	0,33	2,59	0,31	0,21	0,77
ApPyL	0,34	0,43	2,56	0,25	0,28	0,83
SemR	0,29	0,30	0,24	0,21	0,22	0,19
SemL	0,32	0,28	0,31	0,24	0,22	0,19
OriR	3,44	2,42	2,27	2,60	1,88	1,61
OriL	3,56	2,23	2,25	2,64	1,80	1,57
OrsR	0,96	1,00	0,46	0,70	0,83	0,59
OrsL	0,83	1,00	0,47	0,60	1,00	0,69
FmsR	0,58	0,92	1,07	0,53	0,74	0,57
FmsL	0,65	0,79	0,89	0,48	0,62	0,51
OpR	0,68	0,89	0,88	0,64	0,69	0,72
OpL	0,51	0,97	0,97	0,38	0,70	0,84
JR	1,55	1,85	3,37	1,15	1,34	2,14
JL	1,77	1,78	3,34	1,30	1,32	2,10
PrCoR	0,26	0,33	0,28	0,19	0,20	0,23
PrCoL	0,28	0,28	0,31	0,21	0,20	0,19
CmR	0,21	0,32	0,55	0,15	0,25	0,33
CmL	0,20	0,33	0,55	0,14	0,23	0,33
ClR	0,16	0,32	0,58	0,12	0,23	0,35
ClL	0,15	0,27	0,54	0,11	0,20	0,29
GoR	0,27	0,65	0,68	0,19	0,49	0,47
GoL	0,33	0,70	0,71	0,24	0,53	0,53
IGoR	0,89	2,90	0,91	0,66	1,68	1,71
IGoL	1,00	3,38	1,00	0,74	2,08	1,91
PGoR	0,39	0,81	2,34	0,29	0,47	1,00
PGoL	0,39	0,66	1,99	0,29	0,47	0,93
FoMR	0,56	0,55	0,86	0,41	0,31	0,52
FoML	0,59	0,63	0,93	0,43	0,37	0,57
IIR	0,61	0,28	0,35	0,46	0,20	0,27
IIL	0,63	0,24	0,33	0,47	0,17	0,25
ISR	1,00	0,21	0,16	0,80	0,16	0,14
ISL	0,89	0,25	0,17	0,67	0,19	0,15
IIAR	0,33	0,46	0,72	0,24	0,36	0,46
IIAL	0,24	0,51	0,74	0,18	0,36	0,54
ISAR	0,25	0,30	0,41	0,18	0,24	0,25
ISAL	0,29	0,43	0,43	0,21	0,30	0,30

Table 2
Standard deviation and mean absolute difference characteristics for each landmark (dark grey - the most unreliable reference points, light grey - the most reliable reference points).

The most reliable points (SD $<$ 0.2 mm) are quoted below:

- On the X axis: Dens (0.18 mm) and Point-A (0.19 mm), Arcus Zygomaticus left and right (0.14 mm; 0.15 mm), Lateral Condyle left (0.15 mm) and right (0.16 mm).
- On the Y axis: Point-A (0.15 mm), Pogonion (0.16 mm) and Point-B (0.18 mm)
- On the Z axis: Incision Superior right and left (0.16 mm; 0.17 mm)

By the calculation of average MAD more reference points had lower dispersion (< 0.2 mm) than in case of SD, such as Nasion, Sella, Basion, Posterior Nasal Spine, Coronoid Process, Medial Condyle, Semicirculare, Incision Inferior, Incision Inferior and Superior Apicale.

Unreliable landmark characterization: the MAD values were lower than the SD values in case of the same coordinates which proves the fact that the standard deviation distorts more - because of the outliers - than the absolute difference. Therefore, we divided the unreliable landmarks into different groups in accordance with the high intra-or inter-examiner standard deviation (see Table 2.). The inter-examiner differences dependent mainly on the training and experience of the examiners [13] and definitions of landmarks contrary to the intra-examiner analysis, in which the image quality and the position of the anatomical structures played important roles (see Table 2.). The investigators were close or equally qualified. Those coordinates - where only the intra-examiner deviation was high – are not usable in the future for 3D cephalometry, because the anatomical structure was fixed and the image – provided by the iCAT machine – had high quality. Those points which had problem with the deviation only in the inter-examiner test can be improved by using better definition or training courses. The most problematic coordinates were those which had unreliable detection by both tests, these coordinates are not suggested for use in 3D measurement.

Decision tree: when starting to plan surgical treatment, it is important to know which reference points can be used. Therefore in the evaluation of the two statistics (SD and MAD) of the repeatability and reproducibility of landmark detection we focused to the most unreliable coordinates. These conclusions are relevant for the selection of the proper points for the 3D cephalometry:

1. The coordinates of the same landmarks may exhibit a high degree of variance. Since not every landmark plays important role in presurgical planning we should focus not on the landmarks themselves, but on their coordinates.

2. Standard deviation can be more distorted than mean absolute difference.

3. Inter-examiner SD or MAD is mainly higher than intra-examiner value.

Figure 3. shows the suggested method for landmark selection. By considering the results and these conclusions, the following method is suggested when planning:

1. One must respect those coordinates, which are necessary for the planning of the treatment.

2. Looking at the inter-examiner standard deviation of the questionable coordinate:

    a) If the inter-examiner SD is low in those coordinates which are necessary then the landmark can be used safely.

    b) If the inter-examiner SD is high (> 1mm), it is recommended to check the MAD.

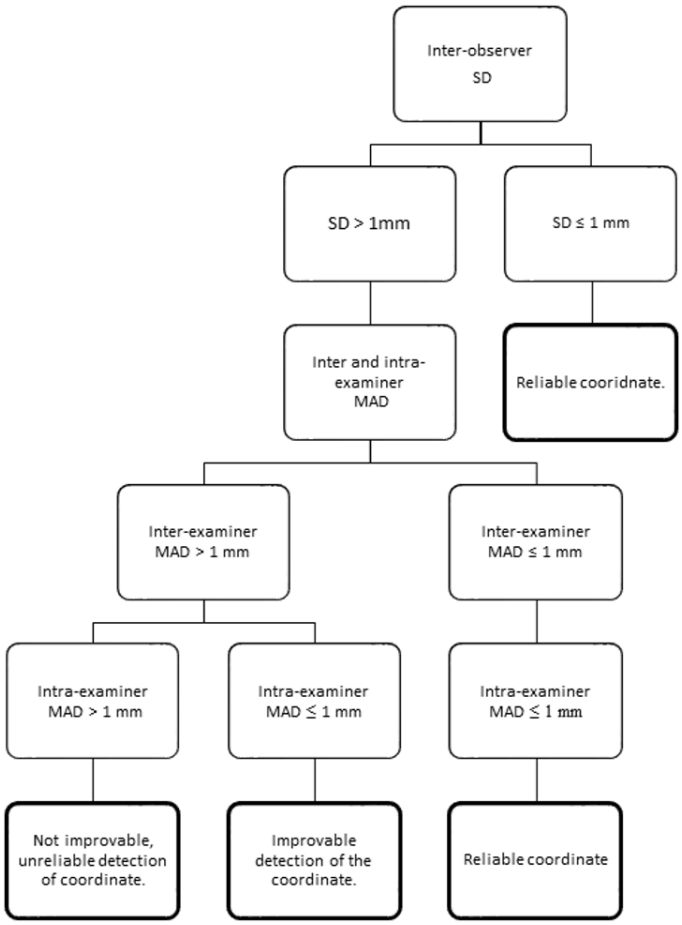

Figure 3
Decision tree for landmark selection.

3. Looking at the inter- and intra-examiner mean absolute difference of the coordinate:

   a) When inter- and intra-examiner MAD are high ($>$ 1mm), many factors, such as image quality, facial anatomy, incorrect definition of the landmark, and the examiner could influence the identification error. Due to the various sources of the problem the accuracy of the detection is not improvable.

   b) When inter-examiner MAD is high ($>$ 1mm), but the intra-examiner MAD is low ($\leq$1 1mm) this can be useful but during the detection paying attention to the source of the problem is needed. The lack of precise definition of landmarks and different training level of examiners can cause the non unambiguous marking. In these cases the problem can only be solved with a new, more accurate definition or with training of

the examiners at a higher quality level.

c) If the inter-examiner MAD is low, the intra-examiner MAD will also be low, so these coordinates are adaptable in 3D accurately.

4 Discussion

Three-dimensional cephalometry plays an important role in the planning of the clinical treatment especially in the field of orthognathic surgery. Therefore in the present study the repeatability and reproducibility of the cephalometric landmarks were analyzed for creating a selection method. Other studies used dry skulls, lateral cephalograms or CBCTs to evaluate landmark identification, but most of them used only lines and angles [14–18], and not the reference points. Previously published studies concluded which reference points were not accurate. However, in 3D cephalometry often happens that we need only one direction of the landmark during the analysis, thus we need to focus the coordinates and not lines or angles.

30 patients with normal occlusion were included in this study and 55 landmarks were located by three examiners three times using Cranioviewer cephalometric software. The members of the examined group had normal occlusion, thus these data can be useful for creating cephalometric analyses in the future, which characterize the ideal state of the face in the presurgical planning.

Some studies have already contrasted the precision of landmark location based on multiplanar reconstructed images and volume rendered 3D cephalometry with the in vitro examination based on dried skulls. Fernandes et al. in 2014 and Neiva et al. in 2015 have reported that measurements performed on multiplanar reconstructed images were more accurate than measurements in volume rendering compared by using digital calipers [19, 20]. Another study stated, that the inter-raters ICCs were higher for multiplanar (ICC $\geq$ 0.90 in 82.16 %) than 3D reconstruction (ICC $>$ 0.90 in 67.76 %) [20]. Besides, Katkar et al. stated that the surface-rendered method increased the evaluation time without improving the precision of identification [21]. Therefore in this study the landmarks were investigated using surface-rendered methods.

According to Bookstein the basic problem with the landmark detection is the lack of suitable, practical definitions of 3D landmarks [22], therefore in this research a reference guide was defined accurately allowed for each plane.

In this study the intra-and inter-examiner intraclass correlation coefficients of each landmark were high (ICC $\geq$ 0.9) in every axis, similar results were reported by Zamora et al. [2]. The results of Oliveria et al. showed less correlations than we calculated. The ICC $\geq$ 0.9 was found for 85 % of the intra-examiner and for 65.5 % of the inter–examiner measurements [23] . To sum up we showed that the ICCs in the 3D cephalometry are mainly high ($\geq$ 0.9) and the inter-examiner reliability is generally higher than intra–examiner [24].

The repeatability and reliability of reference points were characterized with SDs and MADs because of two considerations: *(i)* with the help of the two statistics we could compare the results also to those published articles in which only SD or only MAD was applied and *(ii)* this calculation serves for providing the difference between the

two statistical parameters.
If we compare the results from viewpoint of SD and MAD values, then we can see that the standard deviation was always higher than mean absolute difference, which can be, because the standard deviation is more sensitive to the outliers. Many articles report only the SD values [2]. Due to this fact we could not do exact comparison with the existing literature. However, we compared the results of Zamora et al. with the calculation of this study in the case of midplane landmarks including: Na, S, Ba, Me, Gn, A and B-point, and we found lower MAD than SD in each case. In this study the unreliable coordinates (SD $\geq$ 1 mm or MAD $\geq$ 1 mm) were the same by the two statistics except for Apertura Piriformis and J-Point, whereby the SDs were higher than 1mm, but the MAD was lower. So a landmark selection based on only MAD or only SD can result maybe two different decisions. Therefore during the selection of the landmarks for the presurgical planning we cannot ignore this fact.
If we look at the difference between the intra and inter-examiner results, then we note the same situation as with the ICC, so the intra-rater SD or MAD was lower than inter-rater values. Lagrave et al. analyzed the inter and intra reliability and results showed that inter-examiner mean measurement differences were higher than intra-examiner [11,12,25]. In this study the intra–examiner MADs (0.04 - 1.58 mm) were less than inter-examiner errors, which have been between 0.09 - 2.64 mm.
However, many studies determined the imprecise landmarks, but this problem was not explored, therefore we have created groups based on the type of deviation (Table 3). With the help of Table 3. we can conclude that by some coordinates of the following landmarks – J-point, Zygomaticofrontal suture, Frontomaxillary suture and Infrerior orbitale – the landmark identification error can be reduced with correct definition or more training. With the Y coordinate of the Zygomatic Arch the anatomy and the image quality play important role in the identification. Regarding image quality, Katkar et al. reported a study, in which they compared the landmark identification error with two difference CBCT machines. They stated, that machine differences were significant for almost all landmarks, but some landmarks were more accurate on Galileos and some of them were more accurate on i-CAT machines [21]. Therefore, by those coordinates which had high intra- and low inter-examiner SD, the identification error can be hardly reduced.

Axis	Intraexaminer test	Interexaminer test	Intra–and interexaminer test
X	–	J-point	Orbitale Inferior
Y	Zygomatic Arch	J-point Sutura Zygomaticofrontale	Inferior Gonion Orbitale Inferior
Z	–	Orbitale inferior Sutura Frontomaxillare	J-point Apertura Piriformis Posterior Gonion

Table 3
Anatomical structures, which have had high standard deviation in the intra-examiner, in the inter-examiner or by both test.

Conclusions

In this study most coordinates of the landmarks were reliably adapted to the three-dimensional measurement based CBCT scans. We suggest caution in selection of the coordinates of the following landmarks: Zygomatic Arch, J-point, Zygomaticofrontal suture, Frontomaxillary suture, Inferior Orbitale, Inferior Gonion, Posterior Gonion and the Piriform Aperture.
The characterization of coordinates by SD and MAD showed that the enrollment decision of landmark for 3D cephalometry depends strongly on the statistics.
The analysis of the unreliable coordinates revealed the source of the identification problem, we could improve the detection.
The decision tree can be useful to eliminate the mistakes for the presurgical planning in the practice.

Acknowledgement

M. Mezei thankfully acknowledges the support of the Robotics Special College and the Applied Informatics and Applied Mathematics Doctoral School of Óbuda University.

References

[1] M. A. Papadopoulos, C. Jannowitz, P. Boettcher, J. Henke, R. Stolla, H. F. Zeilhofer, L. Kovacs, W. Erhardt, E. Biemer, and N. A. Papadopulos. Three-dimensional fetal cephalometry: an evaluation of the reliability of cephalometric measurements based on three-dimensional CT reconstructions and on dry skulls of sheep fetuses. *J Craniomaxillofac Surg*, 33(4):229–237, 2005.

[2] N. Zamora, J. M. Llamas, R. Cibrian, J. L. Gandia, and V. Paredes. A study on the reproducibility of cephalometric landmarks when undertaking a three-dimensional (3D) cephalometric analysis. *Med Oral Patol Oral Cir Bucal*, 17(4):e678–e688, 2012.

[3] W. De Vos, J. Casselman, and G. R. Swennen. Cone-beam computerized tomography (CBCT) imaging of the oral and maxillofacial region: a systematic review of the literature. *Int J Oral Maxillofac Surg*, 38(6):609–625, 2009.

[4] P. M. Cattaneo and B. Melsen. The use of cone-beam computed tomography in an orthodontic department in between research and daily clinic. *World J Orthod*, 9(3):269–282, 2008.

[5] O. Lisboa Cde, D. Masterson, A. F. da Motta, and A. T. Motta. Reliability and reproducibility of three-dimensional cephalometric landmarks using CBCT: a systematic review. *J Appl Oral Sci*, 23(2):112–119, 2015.

[6] O. J. van Vlijmen, T. Maal, S. J. Berge, E. M. Bronkhorst, C. Katsaros, and A. M. Kuijpers-Jagtman. A comparison between 2D and 3D cephalometry on CBCT scans of human skulls. *Int J Oral Maxillofac Surg*, 39(2):156–160, 2010.

[7] C. Michele, A. Federica, D. G. Roberto, and S. Alessandro. Two-Dimensional and Three-Dimensional Cephalometry Using Cone Beam Computed Tomography Scans. *J Craniofac Surg*, 26(4):e311–e315, 2015.

[8] M. Fuyamada, M. Shibata, H. Nawa, K. Yoshida, Y. Kise, A. Katsumata, E. Ariji, and S. Goto. Reproducibility of maxillofacial landmark identification on three-dimensional cone-beam computed tomography images of patients with mandibular prognathism: Comparative study of a tentative method and traditional cephalometric analysis. *Angle Orthod*, 84(6):966–973, 2014.

[9] W. J. Houston, R. E. Maher, D. McElroy, and M. Sherriff. Sources of error in measurements from cephalometric radiographs. *Eur J Orthod*, 8(3):149–151, 1986.

[10] R. N. Landers. Computing Intraclass Correlations (ICC) as Estimates of Interrater Reliability in SPSS. https://thewinnower.com/papers/1113-computing-intraclass-correlations-icc-as-estimates-of-interrater-reliability-in-spss. Accessed: 2017-06-20.

[11] M. O. Lagravere, J. M. Gordon, I. H. Guedes, C. Flores-Mir, J. P. Carey, G. Heo, and P. W. Major. Reliability of traditional cephalometric landmarks as seen in three-dimensional analysis in maxillary expansion treatments. *Angle Orthod*, 79(6):1047–1056, 2009.

[12] M. O. Lagravere, C. Low, C. Flores-Mir, R. Chung, J. P. Carey, G. Heo, and P. W. Major. Intraexaminer and interexaminer reliabilities of landmark identification on digitized lateral cephalograms and formatted 3-dimensional cone-beam computerized tomography images. *Am J Orthod Dentofacial Orthop*, 137(5):598–604, 2010.

[13] W. Schlicher, I. Nielsen, J. C. Huang, K. Maki, D. C. Hatcher, and A. J. Miller. Consistency and precision of landmark identification in three-dimensional cone beam computed tomography scans. *Eur J Orthod*, 34(3):263–275, 2012.

[14] P. W. Major, D. E. Johnson, K. L. Hesse, and K. E. Glover. Landmark identification error in posterior anterior cephalometrics. *Angle Orthod*, 64(6):447–454, 1994.

[15] E. L. Delamare, G. S. Liedke, M. B. Vizzotto, H. L. da Silveira, J. L. Ribeiro, and H. E. Silveira. Influence of a programme of professional calibration in the variability of landmark identification using cone beam computed tomography-synthesized and conventional radiographic cephalograms. *Dentomaxillofac Radiol*, 39(7):414–423, 2010.

[16] S. A. Stratemann, J. C. Huang, K. Maki, A. J. Miller, and D. C. Hatcher. Comparison of cone beam computed tomography imaging with physical measures. *Dentomaxillofac Radiol*, 37(2):80–93, 2008.

[17] D. R. Periago, W. C. Scarfe, M. Moshiri, J. P. Scheetz, A. M. Silveira, and A. G. Farman. Linear accuracy and reliability of cone beam CT derived 3-dimensional images constructed using an orthodontic volumetric rendering program. *Angle Orthod*, 78(3):387–395, 2008.

[18] A. A. Brown, W. C. Scarfe, J. P. Scheetz, A. M. Silveira, and A. G. Farman. Linear accuracy of cone beam CT derived 3D images. *Angle Orthod*, 79(1):150–157, 2009.

[19] T. M. Fernandes, J. Adamczyk, M. L. Poleti, J. F. Henriques, B. Friedland, and D. G. Garib. Comparison between 3D volumetric rendering and multiplanar slices on the reliability of linear measurements on CBCT images: an in vitro study. *J Appl Oral Sci*, 23:1, 2014.

[20] M. B. Neiva, A. C. Soares, O. Lisboa Cde, V. Vilella Ode, and A. T. Motta. Evaluation of cephalometric landmark identification on CBCT multiplanar and 3D reconstructions. *Angle Orthod*, 85(1):11–17, 2015.

[21] R. A. Katkar, C. Kummet, D. Dawson, L. Moreno Uribe, V. Allareddy, M. Finkelstein, and A. Ruprecht. Comparison of observer reliability of three-dimensional cephalometric landmark identification on subject images from Galileos and i-CAT cone beam CT. *Dentomaxillofac Radiol*, 42(9):20130059:1–11, 2013.

[22] F. L. Bookstein. Morphometric Tools for Landmark Data. *Cambrige University Press*, page p453, 1991.

[23] A. E. de Oliveira, L. H. Cevidanes, C. Phillips, A. Motta, B. Burke, and D. Tyndall. Observer reliability of three-dimensional cephalometric landmark identification on cone-beam computerized tomography. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, 107(2):256–265, 2009.

[24] J. F. Gravely and P. M. Benzies. The clinical significance of tracing error in cephalometry. *Br J Orthod*, 1(3):95–101, 1974.

[25] Gy. Eigner, J.K. Tar, I. Rudas, and L. Kivács. LPV-based quality interpretations on modeling and control of diabetes. *ACTA Pol Hung*, 13(1):171–190, 2016.