Reviewer's report

Title: The accuracy of a designed software for automated localization of craniofacial landmarks on CBCT images

Version: 2
Date: 3 May 2014

Reviewer: Bassam Hassan

Reviewer's report:

Dear authors:

With interest I read your manuscript titled “The accuracy of a designed software for automated localization of craniofacial landmarks on CBCT images”. Please find below my comments:

General impression:

The study is relatively well-written although several grammatical errors and typos persist throughout the manuscript. The objective is well stated and relevant literature is cited. The applied methodology is valid though I do have my reservations and many remarks can be stated regarding the authors’ choice for the applied registration technique. The results are well presented yet some details are lacking. The discussion cites relevant literature but does not really satisfy the curiosity of the reader about the applicability of this technique or the potential for future improvements.

Major Compulsory Revisions:

1) I have deep concerns regarding the registration approach between the reference and the test dataset. CBCT scans are usually matched either by using iterative closest point (ICP) for segmented isosurfaces or better yet, using direct volumetric registration using the maximization of mutual information (MMI) which provides sub-voxel matching accuracy (Maes et al., 1997). One can argue that matching of surfaces is threshold dependent and therefore it’s inaccurate especially in large FoV CBCT scans where too many artefacts are present. However, matching based on voxel data is far more accurate and is readily accessible. A detailed yet concise explanation of registration based on principal axis and vector should be provided and also a justification should be provided for electing this technique and not the more familiar techniques.

2) If the study as a retrospective case control was indeed approved by a recognized local ethics committee, then a study number should be provided. I’m not sure that I understand the sample collection. It was mentioned that inclusion criteria included patients prior, during and post orthodontic treatment. Does that apply to reference and test datasets? How many were pre, during and post? And by during I presume that the patients must have been wearing braces.
during making the scan. That would definitely influence the segmentation and registration accuracy. Please elaborate.

I’m concerned about the age distribution of the subjects in this study. Were large field of view scans made for 5 years old just for cephalometric analysis? How many patients were like this? And what was the indication? The age distribution of the test datasets should be mentioned as well.

Minor Essential Revisions:

Abstract:

1) Results: ICC results should be mentioned.
2) Conclusions: The Principal Axis Registration (PAR) term should be mentioned. Only PAR abbreviation is mentioned.

Introduction:

1) It is stated in line 114 “For 2D, there is (should be are) …..”. I would like more information (briefly) with regard to the accuracy of the four mentioned approaches.
2) In line 120 The results of the projects of Mestiri and Pan Zheng should be briefly discussed and compared to the results obtained in this study. This could also be done in the discussion section.

Materials and methods:

Software:

1) A figure should be provided regarding the MPR and the 3D view illustrating the manual selection of the landmarks.
2) Were the measurements independent of the 3 experts or in consensus? was there any pre-calibration for the observers in the study?
3) It doesn’t sound very assuring that the only criteria for selecting a reference ‘dataset’ is patient’s age instead of some image based features. Were there no image features used for recognition?

Automated landmark detection:

1) What were the values of the segmentation? Was it only binary thresholding?
2) What does the mass point theory entail? For the clinical readership of the journal some relevant information should be provided?
   I do not understand this sentence "Identified vector (i) was defined based on manual detected landmarks"

3) What does it mean that the length of the vector is not identical? Does it mean that the datasets had different scale?
   Line 219: I presume “rigid affine” and not only “affine"
Results:

1) I presume that the ‘standard deviation’ provided in table 2 is actually the ‘standard error of the mean’. Is that correct?

2) Standard deviation results of the 6 manual measurements per point should be provided.

3) A figure with the matching between the reference and the test image should be supplied as well.

I’m not sure that providing the description of the points in table one based on their visibility on the MPR images is quite informative. I don’t believe it’s quite necessary to have it.

Discussion:

1) Again, as mentioned in the comments in the introduction. The discussion should go in details comparing the results here with previously obtained results of 3D matching from previous studies. Also, the differences in the methodology should be emphasized.

Study limitations should be mentioned. Think about registration inaccuracy, the transfer technique of landmarks between test and reference image, artifacts, ethnic group limitation, etc..

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests.