Automatic Landmark Detection based on Deep learning to Reduce Intra- and/or Inter-Observer Variability Error in -Hip Range of Motion Analysis

Khalil Kilani

Supervisors: Dr. Nicolas Gerber, MSc Adrian Ruckli
Institution: University of Bern, School of Biomedical Precision Engineering
Examiners: Dr. Nicolas Gerber, PD Dr. Kate Gerber

Introduction

Premature osteoarthritis in the hip joint can often be linked to femoral acetabular impingement or hip dysplasia. To assess the hip range of motion (ROM) and detect impingement positions between the acetabulum and proximal femur before considering surgical treatment, clinicians use a computer-assisted tool called hip ROM analysis [1]. This requires 3D surface models of the pelvis, the proximal femur and distal femur, obtained through bone segmentation from original MRI or CT images. Traditionally, the hip ROM analysis involved manual selection of specific anatomical landmarks on these surface models to establish a reference frame for its collision detection algorithm. Due to user variance, this selection can potentially influence the analysis outcome. To address this issue, this study utilized deep learning to automate the identification of these landmarks and aimed to reduce inter-/intra-observer variability in the hip ROM analysis.

Materials and Methods

Three convolutional neural networks were trained on manually segmented bone images from the pelvis, proximal femur and distal femur to detect a total of 24 landmarks as shown in figure 1.

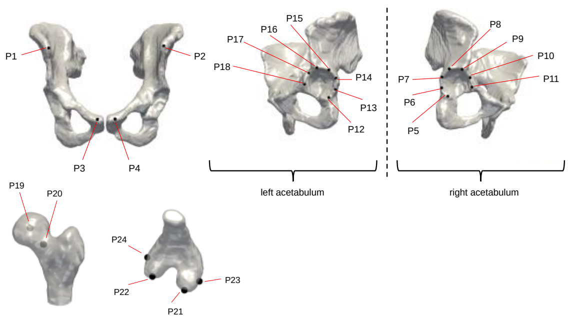

Fig. 1 All 24 landamrks used in this study on the three distinct structures (pelvis, proximal femur and distal femur)

The developed landmark algorithm takes a segmented pelvis, proximal and distal femur image as input in addition to the hip side. Using the trained networks, the coordinates of the 24 landmarks are defined in file format which inference the hip ROM analysis. To assess the algorithms ability to decrease user variability, a ROM-test commonly employed by clinicians when using the hip ROM analysis was conducted. The test involved 22 leg positions per hip/patient and was performed twice: initially using manual clinician-chosen landmarks and subsequently with automated landmarks. Using the identical data, we also evaluated intra- and inter-observer variability, as well as the difference between manually identified landmarks and the automated ones.

Results

The ROM-test was performed for 15 hips. The results showed that in 92% of the tested positions, no differences were found when using either manual or automated landmarks. In 8% the differences were minimal, with a 5° variation, except for one case with a 10° difference. The mean error between the manual and automated landmarks over all landmarks was lower with 3.5mm compared to the intra-observer (4.32mm) and inter-observer variability (4.3mm). Figure 2 shows the error for the three tested groups. The total runtime of the algorithm was 40 seconds compared to 1 minute and 19 seconds when performing manual landmark detection.

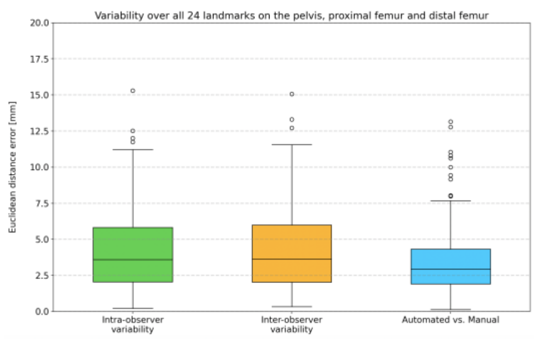

Fig. 2 Error between identical landmarks for the three distinct tested groups.

Discussion

The results indicate that the automated landmark detection algorithm reduces user variability. The runtime of the algorithm is only one-third compared to manual selection and can further be improved.

References

[1] Moritz Tannast et al., Noninvasive three-dimensional assessment of femoroacetabular impingement, Journal of orthopaedic research 25(1): 122–131, 2007.

Acknowledgements

Dr. Nicolas Gerber, MSc Adrian Ruckli, PD Dr. med. Simon D. Steppacher, Tadeus Alexander Becker and Dr. med. univ. Jose Roshardt