Detecting Anatomical Landmarks for Motion Estimation in Weight-bearing Imaging of Knees

Bastian Bier¹, Katharina Aschoff¹, Christopher Syben¹, Mathias Unberath², Marc Levenston³, Garry Gold³, Rebecca Fahrig³, and Andreas Maier¹

¹ Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg ⊠bastian.bier@fau.de

² Laboratory for Computational Sensing and Robotics, Johns Hopkins University ³ Radiological Sciences Laboratory, Stanford University

Abstract. Patient motion is one of the major challenges in cone-beam computed tomography (CBCT) scans acquired under weight-bearing conditions, since it leads to severe artifacts in reconstructions. In knee imaging, a state-of-the-art approach to compensate for patient motion uses fiducial markers attached to the skin. However, marker placement is a tedious and time consuming procedure for both, the physician and the patient. In this manuscript we investigate the use of anatomical landmarks in an attempt to replace externally attached fiducial markers. To this end, we devise a method to automatically detect anatomical landmarks in projection domain X-ray images irrespective of the viewing direction. To overcome the need for annotation of every X-ray image and to assure consistent annotation across images from the same subject, annotations and projection images are generated from 3D CT data. Twelve landmarks are annotated in supine CBCT reconstructions of the knee joint and then propagated to synthetically generated projection images. Then, a sequential Convolutional Neuronal Network is trained to predict the desired landmarks in projection images. The network is evaluated on synthetic images and real clinical data. On synthetic data promising results are achieved with a mean prediction error of 8.4 ± 8.2 pixel. The network generalizes to real clinical data without the need of re-training. However, practical issues, such as the second leg entering the field of view, limit the performance of the method at this stage. Nevertheless, our results are promising and encourage further investigations on the use of anatomical landmarks for motion management.

1 Introduction

C-arm cone-beam computed tomography (CBCT) systems have been used recently to acquire 3D images of the human knee joint under weight-bearing conditions [1, 2]. Scans under weight-bearing conditions can be beneficial for the investigation of the knee health since it has been shown that the human knee joint shows different properties in a natural position under load compared to a supine acquisition [3]. Load bearing imaging requires dedicated imaging protocols. Using robotic C-arm systems driven in horizontal trajectories [1, 4, 5], it takes several seconds to acquire enough 2D projection images for a clinically satisfying reconstruction. During that time, the standing patient might move involuntarily. This motion leads to inconsistencies in the projection data, and thus, to motion artifacts in the reconstructions. Therefore, motion compensation is indispensable for achieving diagnostic reconstruction quality in weight-bearing CBCT of the knee.

In order to reduce motion induced artifacts in such scenarios, various approaches have been proposed: autofocus-based methods optimize image-quality criteria in reconstructions [6], registration-based approaches align acquired images to a static reference [7, 4, 8], while range camera-based solutions image the knee surface to estimate patient motion [5]. Another state-of-the-art method uses metallic fiducial markers externally attached to the skin of the knee [1, 4]. Due to their high attenuation, these markers are easily visible and detectable in the 2D projections. Using the detected marker locations, 3D reference marker positions can be computed. Having 2D positions as well as corresponding 3D reference positions, a refined C-arm trajectory can be computed analytically in a 2D/3D alignment step, i. e. without the need for computation-heavy optimization. Despite best-inclass performance, the usability of this method suffers: marker placement is time consuming, interrupts the clinical workflow, and must be executed carefully since markers must not overlap in the projections. Therefore, a purely image-based method similar to the fiducial marker-based approach is desirable.

A promising candidate to replace the markers are anatomical landmarks visible in projection images. Finding key points and establishing correspondences in images of the same scene is a well understood concept in computer vision. However, this concept does not translate easily to transmission imaging, where the appearance of the same landmark can vary tremendously dependent on the viewing direction. Recently, Convolutional Neuronal Network (CNN)-based sequential predication frameworks have shown promising performance in detecting anatomical landmarks in X-ray transmission images of the pelvis across a large range of viewing angles of the C-arm CT system [9].

Here, we transfer the work in Bier and Unberath *et al.* [9] to view-independent anatomical landmark detection in CBCT short scans of knees under weightbearing conditions. To this end, a CNN is trained on synthetic projection images generated from 3D CBCT data. In total, twelve anatomical landmarks are manually annotated in 3D and then predicted in projection domain. The network readily establishes landmark correspondence across images suggesting that sufficiently accurate landmark detection will pave the way for "anatomical marker"based motion compensation. Our landmark detection is evaluated on a simulated short scan, and two clinical CBCT scans in supine and weight-bearing condition, respectively. The network is trained on synthetic data [9], yet, generalizes to real projection images without the need of re-training.



Fig. 1. Network architecture [9].

2 Method

2.1 X-ray invariant Anatomical Landmark Detection

Detection of anatomical landmarks irrespective of the viewing direction has been proposed recently [9]. The concept of landmark detection was derived from a sequential prediction framework, namely the Convolutional Pose Machine (CPM) [10]. This network architecture was initially developed to detect human joint positions in RGB images and provides key benefits: it combines local image features with increasingly refined belief maps to establish landmark relationships. The network processes each image independently and, for each landmark, predicts a belief map indicating the landmark position.

The network involves successive processing of the input image over several stages, see Figure 1. In the first stage, the network architecture consists of convolutional and pooling layers, which result in initial belief maps. In the following stages, these belief maps are refined using both local image features and the belief maps of the previous stage. The cost function of the network is the difference between the predicted belief maps b_t^p and the ground truth belief maps b_t^* of all landmarks $p \in \{1, ..., P\}$ and in each stage t: consequently, the l_2 norm of this difference defines the cost function f_t [10]:

$$f_t = \sum_{p=1}^{P} \|b_t^p - b_t^*\|_2^2.$$
(1)

This network structure has several properties: it has a large receptive field (160 \times 160 pixels) on the input image, empowering the network to learn characteristic global configurations over long-distances. The stage-wise manner also allows the network to resolve ambiguities due to similar local appearance. Further, accumulating the loss after each stage prevents vanishing gradients that often occur in large CNNs [10].

2.2 Training

In order to train the network, projection images and corresponding landmark positions have to be known. We follow the approach discussed in [9, 11] and



Fig. 2. Anatomical landmarks defined on the surface of the bones in the knee joint.

generate projection images and annotations synthetically by annotating twelve anatomical landmarks in CBCT volumes of the human knee, see Figure 2. The landmarks have been selected to be good visible in the projections images as well as clearly identifiable in the 3D volume. The CBCT volumes were reconstructions of scans acquired in supine position (Siemens Zeego, Siemens Healthcare GmbH, Erlangen, Germany). In total 16 CBCT volumes were available for training. After annotation of the landmark positions in the volumes, projection images and corresponding annotations were generated synthetically using CONRAD [12]. From each dataset, 1000 projection images were simulated. For data augmentation purposes, images were sampled during projection generation on a spherical segment with a range of 240° LAO/RAO and 20° in CRAN/CAUD. This range covers more than the necessary variance of a common CBCT short scan. Additionally, random translations in three Cartesian axes and horizontal flipping of the projections were used. The belief map of a particular landmark consists of a single normal distribution centered at the true landmark location. The size of the projections was 615×479 with a pixel size of $0.6 \,\mathrm{mm}$. The belief maps were downsampled eight times.

16 supine CT scans, split $14 \times 1 \times 1$ -fold in training, validation and testing data are used for the training and testing. The network was trained with six stages for 30 epochs with a constant learning rate of 0.00001 and a batch size of one. The optimization was done using Adam optimization. Figure 3 shows that convergence is reached during both training and validation.



Fig. 3. Training loss (left side) and Validation loss (right side)

2.3 Landmark Estimation

The network outputs twelve belief maps that indicate the landmark positions. The belief map after each step is accumulated, and the 2D landmark position is defined as the maximum response in the accumulated belief map.

3 Experiments and Results

Landmark detection is evaluated quantitatively on a synthetic short scan dataset as well as qualitatively on two clinical CBCT scans in supine and standing condition, respectively. In order to investigate the prediction results over the complete trajectory, detection results sampled from different directions are represented in Figure 4. Column-wise from left to right, we show results on the synthetic dataset, the real clinical data in supine and in standing conditions, respectively. Detected landmarks are highlighted in red and reference marker positions in white, wherever available.

The detection results on the synthetic dataset are in good agreement with the ground truth label positions. Visually, also the detected landmarks in the real clinical images are in agreement with the labeled locations. Note that in the supine scan also a part of the patient's feet is present in some parts of the projections. However, this does not seem to influence the landmark detections. In the projections acquired under weight-bearing conditions a second leg is also present in parts of the projection. Since there is a second knee in the field of view, the detection of the landmarks is not consistently on one leg only.

Landmark $\#$	Distance $(\mu \pm \sigma)$	Landmark $\#$	$ \text{Distance}(\mu \pm \sigma) $
1	6.6 ± 2.0	7	17.7 ± 8.6
2	10.5 ± 3.9	8	3.2 ± 1.9
3	3.8 ± 1.4	9	5.1 ± 1.6
4	8.7 ± 2.5	10	5.1 ± 1.6
5	9.5 ± 5.0	11	7.0 ± 4.6
6	18.1 ± 19.2	12	5.7 ± 3.9

 Table 1. Average distance [pixels] of the predicted landmarks to the ground truth location.

Since the reference landmark locations were known on the synthetic short scan dataset, we computed the average distance to the ground truth landmark locations as well as the detection rate. We define a landmark as detected, if the distance to its ground truth location is < 15 pixel and the maximum belief is ≥ 0.4 . The average distance of the landmark detections on the simulated short scan was then 8.4 ± 8.2 pixels and a detection accuracy of 89.16% is reached. Furthermore, we investigated the quality of the selected anatomical landmarks and computed the average distance for each landmark. The results of this are



Fig. 4. Detection results on the synthetic (left), a supine scan (center), and a standing scan (right).

shown in Table 1. Large differences between individual landmarks can be observed here. The best landmarks are the tip of the Fibula (landmark #3) and landmarks inside the knee joint. It is further noticeable that landmarks with less other neighboring landmarks, e.g. on the Patella (landmark #6), or on the Tibia (landmark #7), are detected with a much higher uncertainty.

4 Conclusion and Outlook

The presence of patient motion during CBCT scans is one of the major challenges in CBCT acquisitions acquired under weight-bearing conditions. Currently, an approach based on metallic fiducial markers is used to estimate motion. However, marker placement is time consuming and tedious. Therefore, we investigated the feasibility of using anatomical landmarks as image-based markers instead.

An X-ray invariant anatomical landmark detection approach was utilized to detect landmarks in projection images. Trained on high quality supine data of the knee, the network predicted belief maps in which the position of the anatomical landmarks can be estimated in synthetic as well as in real clinical data. These landmarks could be used to estimate motion using a 2D/3D based registration approach. The estimation of the motion with these detections is subject of future work. It also had been shown that some landmarks could be estimated more robustly than others. This might contain the potential to incorporate this information in the further processing steps. Furthermore, we believe that such approaches might be applicable to compensate other complex body motion [13], e. g., using motion models for respiratory [14] or cardiac motion [15].

Despite promising results on projection images of the knee, some limitations remain. The large angular range of short scans unavoidably implies the presence of both legs in the field of view. On the one hand, bones superimpose and hinder the detection. On the other hand, we observed "jumping" of detections from one knee to the other. These observations further motivate why landmark detection seems to visually perform better on supine than on standing data. Moreover, the method results in limited accuracy due to downsampling of the ground truth belief maps by factor of around eight. To improve the accuracy, an advanced network incorporating skip-ahead-connections might increase the performance.

Despite these limitations, this work shows that the automatically landmark detection works well for synthetically generated as well as for real X-ray projection images of knee joints. In future work, we will investigate methods to make landmark prediction more robust, particularly in presence of additional anatomy, and to use our predictions to estimate and compensate for patient motion during reconstruction.

References

 Choi, J.H., Fahrig, R., Keil, A., Besier, T.F., Pal, S., McWalter, E.J., Beaupré, G.S., Maier, A.: Fiducial marker-based correction for involuntary motion in weightbearing C-arm CT scanning of knees. Part I. Numerical model-based optimization. Medical Physics $\mathbf{41}(6)$ (2014) 061902

- Choi, J.H., Maier, A., Keil, A., Pal, S., McWalter, E.J., Beaupré, G.S., Gold, G.E., Fahrig, R.: Fiducial marker-based correction for involuntary motion in weightbearing C-arm CT scanning of knees. II. Experiment. Medical Physics 41(6) (2014) 061902
- Powers, C.M., Ward, S.R., Fredericson, M.: Knee Extension in Persons With Lateral Subluxation of the Patella : A Preliminary Study. Journal of Orthopaedic and Sports Physical Therapy 33(11) (2013) 677–685
- Berger, M., Müller, K., Aichert, A., Unberath, M., Thies, J., Choi, J.H., Fahrig, R., Maier, A.: Marker-free motion correction in weight-bearing cone-beam CT of the knee joint. Medical Physics 43(3) (2016) 1235–1248
- Bier, B., Ravikumar, N., Unberath, M., Levenston, M., Gold, G., Fahrig, R., Maier, A.: Range Imaging for Motion Compensation in C-Arm Cone-Beam CT of Knees under Weight-Bearing Conditions. J. Imaging 4(4) (2018) 561–570
- Sisniega, A., Stayman, J., Yorkston, J., Siewerdsen, J., Zbijewski, W.: Motion compensation in extremity cone-beam CT using a penalized image sharpness criterion. Physics in Medicine and Biology 62(9) (2017)
- Unberath, M., Choi, J.H., Berger, M., Maier, A., Fahrig, R.: Image-based compensation for involuntary motion in weight-bearing C-arm cone-beam CT scanning of knees. In: SPIE Medical Imaging. Volume 9413. (mar 2015) 94130D
- Ouadah, S., Jacobson, M., Stayman, J.W., Ehtiati, T., Weiss, C., Siewerdsen, J.H.: Correction of patient motion in cone-beam CT Correction of patient motion in cone-beam CT using 3D 2D registration. (2017)
- Bier, B., Unberath, M., Zaech, J.N., Fotouhi, J., Armand, M., Osgood, G., Navab, N., Maier, A.: X-ray-transform invariant anatomical landmark detection for pelvic trauma surgery. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2018) to appear
- Wei, S.E., Ramakrishna, V., Kanade, T., Sheikh, Y.: Convolutional pose machines. In: CVPR. (2016) 4724–4732
- Unberath, M., Zaech, J.N., Lee, S.C., Bier, B., Fotouhi, J., Armand, M., Navab, N.: Deepdrr-a catalyst for machine learning in fluoroscopy-guided procedures. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2018) to appear
- Maier, A., Hofmann, H.G., Berger, M., Fischer, P., Schwemmer, C., Wu, H., Müller, K., Hornegger, J., Choi, J.H., Riess, C., Keil, A., Fahrig, R.: CONRAD - A software framework for cone-beam imaging in radiology. Medical Physics 40(11) (2013) 111914
- Müller, K., Maier, A., Schwemmer, C., Lauritsch, G., De Buck, S., Wielandts, J., Hornegger, J., Fahrig, R.: Image artefact propagation in motion estimation and reconstruction in interventional cardiac c-arm ct. Physics in Medicine & Biology 59(12) (2014) 3121
- Geimer, T., Birlutiu, A., Unberath, M., Taubmann, O., Bert, C., Maier, A.: A Kernel Ridge Regression Model for Respiratory Motion Estimation in Radiotherapy. In Maier-Hein, K., Deserno, T., Handels, H., Tolxdorff, T., eds.: Bildverarbeitung für die Medizin 2017, Heidelberg, Berlin (2017) 155–160
- Unberath, M., Geimer, T., Höhn, J., Achenbach, S., Maier, A.: Myocardial Twist from X-ray Angiography. In Maier, A., Deserno, T., Handels, H., Maier-Hein, K.H., Palm, C., Tolxdorff, T., eds.: Bildverarbeitung für die Medizin 2018 - Algorithmen - Systeme - Anwendungen, Berlin, Heidelberg (2018) 365–370