Automatic Joint Alignment Measurements in Pre- and Post-operative Long Leg Standing Radiographs

A. Gooßen1; G. M. Weber2,3; S. P. M. Dries1

1Philips Research, Hamburg, Germany; 2Philips Healthcare, Diagnostic X-Ray, Hamburg, Germany; 3Hamburg University of Technology, Vision Systems, Hamburg, Germany

Keywords
Radiography, bone, lower extremity, replacement arthroplasty, orthopedics

Summary
Objectives: For diagnosis or treatment assessment of knee joint osteoarthritis it is required to measure bone morphometry from radiographic images. We propose a method for automatic measurement of joint alignment from pre-operative as well as post-operative radiographs.

Methods: In a two step approach we first detect and segment any implants or other artificial objects within the image. We exploit physical characteristics and avoid prior shape information to cope with the vast amount of implant types. Subsequently, we exploit the implant delineations to adapt the initialization and adaptation phase of a dedicated bone segmentation scheme using deformable template models. Implant and bone contours are fused to derive the final joint segmentation and thus the alignment measurements.

Results: We evaluated our method on clinical long leg radiographs and compared both the initialization rate, corresponding to the number of images successfully processed by the proposed algorithm, and the accuracy of the alignment measurement. Ground truth has been generated by an experienced orthopedic surgeon. For comparison a second reader reevaluated the measurements. Experiments on two sets of 70 and 120 digital radiographs show that 92% of the joints could be processed automatically and the derived measurements of the automatic method are comparable to a human reader for pre-operative as well as post-operative images with a typical error of 0.7° and correlations of \( r = 0.82 \) to \( r = 0.99 \) with the ground truth.

Conclusions: The proposed method allows deriving objective measures of joint alignment from clinical radiographs. Its accuracy and precision are on par with a human reader for all evaluated measurements.

1. Introduction

Joint replacement surgery has become a standard orthopedic procedure in the past decades. The German Federal Association of Medical Technology reports 390,000 artificial hip joints and knee joints implanted in Germany each year. In the United States, according to the National Health Statistics Reports [1], surgeons even perform 780,000 hip and knee replacements per year. These primary implantations are complemented by an additional smaller number of revision surgeries, i.e. replacing an existing implant by a new one. The number of procedures per year has increased continuously over the last two decades and the ageing population is expected to further fuel this growth in the future. For surgical planning, orthopedic surgeons agree [2] that it is important to specify certain X-ray imaging based quantities, such as length and angle measurements [3]. Orthopedic surgeons perform careful pre-operative planning and templating based on exact alignment measurements prior to any treatment [2, 4, 5] and assess the outcome reading the corresponding measurements from post-operative images [6, 7]. The accurate delineation of the structures of interest required by experts in order to define such quantities, however, is often time-consuming and subject to intra- and inter-observer variability [8, 9].

In many clinics image acquisition workflow is already optimized for patient throughput. Thus, reading and evaluating the images forms the major bottleneck in the clinical routine. To cope with the increasing number of orthopedic radiographs and their assessment, accurate computer-aided segmentation may be employed, i.e. fully automatic or semiautomatic procedures [10]. Such methods have to be invariant against the large heterogeneity of images that appear in clinical practice, and have to generate consistent and reliable measurements. The major challenges when processing orthopedic radiographs are

1. poor image quality resulting from the effort to keep radiation doses as low as possible and therefore demanding for robust features,
2. a high degree of anatomical variability, especially caused by the development of bones during childhood and adolescence as well as by the malformation and misalignment during degeneration, and
3. artificial objects such as implants, fixations, or clamps, in post-operative radiographs as a result of the treatment.

This work presents a new method that addresses these challenges and automatically
determines objective and reproducible measurements. The method has been evaluated on clinical data and has been compared against the readings of experienced orthopedists.

2. Objectives

In [11] we presented a technique for delineation of the joints in the lower limbs regardless of patient age. Orthopedic measurements derived from these delineations showed high agreement with manually derived measurements. In order to fully capture the variety of images in clinical practice, this work proposes an approach to deal with artificial objects, such as implants, occurring in post-operative radiographs (Fig. 1). We address implant localization and delineation, model initialization and adaptation, and finally evaluate the quality of the delineation against the ground truth of experienced orthopedists.

Little research has been published for implant segmentation or robustness against implants. In the evaluation of their hybrid approach, based on geometric models and shape priors [12], Dong and Zheng occluded small fractions of the bone contour. However, an implant typically replaces major parts of the bone and does therefore not match their experiment. A method that directly focuses on the segmentation of prostheses after total hip joint replacements (THR) has been developed by Kotcheff et al. [13]. It performs well up to a certain degree of similarity between the trained and actual prostheses with a reported average delineation error of 0.69 mm for the bone structure and 1.73 mm for implant parts. The German Federal Association of Medical Technology, however, counts more than 200 different types of prostheses for the hip joint alone, each available in different sizes. Similarly for the other joints, there exist various types of knee and ankle, partial and full prostheses. Furthermore, implants are sometimes accompanied by internal fixation objects, such as screws, plates or nails. This highly variable appearance generally rules out any shape-based modeling technique when aiming to cover a large fraction of the possibly occurring implant variants in clinical practice.

In order to tolerate implants when delineating bone contours within the lower limbs, we developed an approach consisting of three complementing stages. The first one robustly detects the presence of implants. Subsequently, a second step segments them based on the detected seed positions. Finally, the deformable templates used for bone segmentation are adapted, avoiding the implant structure in order to precisely delineate the remaining bone contours.

3. Methods

This section presents a brief introduction on the developed template models for bone structure segmentation and subsequently describes the three stages necessary for processing post-operative images. Please note that in accordance with clinical practice in radiography all images are shown such that dark image regions correspond to high intensity, i.e. low absorption, and bright image regions correspond to low intensity, i.e. high absorption.

![Fig. 1](image1.png) Knee joint region cropped from (a) a pre-operative and (b) a post-operative radiograph of the same patient. The implant replaces the joint surfaces. Note that radiographs are generally displayed with dark gray values denoting higher X-ray intensity and vice versa.

![Fig. 2](image2.png) Knee joint model and its major mode of variation controlling the bone configuration from neutral to bow legs (genua vara) or knock knees (genua valga)
3.1 Deformable Templates for Bone Delineation

Let a shape $x = (x_1, \ldots, x_n, y_1, \ldots, y_n)^T$ consist of $n$ landmarks $(x_i, y_i)^T, i = 1, \ldots, n$. For the joint segmentation in pre-operative images we trained dedicated deformable template models on radiographs of patients without joint replacement (Fig. 2). With $\bar{x}$ denoting an average representation of a point distribution model and $P$ denoting the corresponding modes of variation we can approximate any shape $x$ by adding a linear combination of modes to the mean shape, i.e.

$$x = \bar{x} + Pb$$

(1)

with $b$ denoting the shape coefficients [14]. When approximating an unknown shape $\hat{x}$, we determine the model parameters $b$ that minimize the error

$$\Delta = (\hat{x} - (\bar{x} + Pb))^T W (\hat{x} - (\bar{x} + Pb)),$$

$$W = \text{diag}(w_1, \ldots, w_{2n})$$

(2)

between the shape target $\hat{x}$ and the shape $x$, generated by using Equation 1. The weights $w_i, i = 1, \ldots, 2n$, $w_{2i} = w_i \forall i \leq n$ control the influence of a specific landmark $(x_i, y_i)^T$ [15].

In order to connect the trained shape to the image data we also have to learn the local structure around each of these landmarks. For this purpose we extract a sampling vector $s_i$ perpendicular to the local shape tangent for each landmark. Similar to the training of the shape we derive the mean appearance $\bar{s}_i$ for all the models of the training set and the empirical covariance matrix, estimated by $S_i$. By help of this trained structure it is possible to measure the similarity of the model and the image content via the Mahalanobis distance

$$d = \sqrt{(\bar{s}_i - s_i)^T S_i^{-1} (\bar{s}_i - s_i)}.$$

(3)

3.2 Implant Detection and Seeding

There are plenty of different implants for various applications and many of them have a varying appearance [16, 17]. In addition, when projecting implants to a plane their image also depends on the projection angle. Shape recognition methods, like active shape models [13, 14] and template-based registration [18], would be too complex or would fail, when applied to all these different types of implants. Due to this fact an approach for the detection and segmentation of implants is needed that is independent of the shape. However, common information can still be trained for all of these types of implants using a shape-free model. This section discusses these common properties of different types of implants and derives a method for their detection.

Implants in radiographs showcase a distinct sharpness of edges and homogeneous low intensity due to the high absorption of their materials, typically stainless steel, cobalt-chromium, and titanium alloys. With less X-ray quanta contributing to the image formation in the implant regions, the fraction of scattered radiation in relation to the primary radiation increases and thus the signal-to-noise ratio (SNR) is significantly lower than in anatomic regions. Combining these properties, it is possible to locate regions in a radiograph that with a high probability belong to an implant. Therefore, for detection three seed point sets, $B_{\text{horizontal}}, B_{\text{vertical}}$, and $B_{\text{hist}}$, incorporating this knowledge, are produced. $B_{\text{horizontal}}$ and $B_{\text{vertical}}$ contain pixels enclosed by strong horizontal and vertical edges, respectively, i.e. a pixel is included as a seed point if it has a preceding strong declining edge, a successive strong inclining edge, and the variation stays below the homogeneity threshold $l_{\text{hom}}$. With $I_{\text{dx}} = \partial I/\partial x$ denoting the horizontal gradient of the image $I(x, y)$ and $I_{\text{dx}}(x', y) < l_{\text{edge}}$ and
Idy = dI/dy.

Another set, $B_{\text{hist}}$, reflects the fact that the implant components have high absorption coefficients. It is created via histogram-adaptive thresholding of the input image. Let denote the input image and let denote a window of the same width, $X$, as the original image, but limited height. Thresholding is carried out as

$$B_{\text{hist}}(x, y) = \begin{cases} 0 & I_{\text{window}}(x, y) > t_{\text{hist}}(y) \\ 1 & \text{else} \end{cases}$$

sliding the window, $\Psi$, vertically over the image frame, $\Omega$. The threshold, $t_{\text{hist}}$, is determined as a fixed percentile of the histogram of intensities of the current window $I_{\text{window}}$. Intersecting these seed point sets and morphologically eroding with a circular structure element $L$ in order to avoid leakage and remove small noise clusters yields the seed points $B_{\text{seed}} = (B_{\text{horizontal}} \cap B_{\text{vertical}} \cap B_{\text{hist}}) \ominus L$.

### 3.3 Implant Delineation

Subsequently, each seed point in $B_{\text{seed}}$ is subdued to a region growing procedure with

$$g(x, y) = \frac{1}{1 + \left| \nabla (I_{\text{smooth}}(x, y)) \right|},$$

$$I_{\text{smooth}} = G_\sigma \times I,$$

serving as the cost function [19]. Herein, $I$ denotes the original image and $G_\sigma$ denotes a Gaussian filter with standard deviation $\sigma$. Thus the segmented area continuously expands until stopped by the strong enclosing gradient, resulting in $B_{\text{grown}}$.

Because of the Gaussian smoothing the implant segmentation result of the region growing lacks precision. To maximize accuracy, the now known coarse locations of the implant borders are utilized in a further local adaptive thresholding procedure. Herein Otsu’s algorithm [20] is applied to a small sliding window which is shifted along the implant outline in $I$ to produce refined implant borders of the image $B_{\text{grown}}$. Since local image content can display several general intensity classes (implant, dense bone tissue, soft tissue, and direct radiation), it is advantageous to perform the thresholding with multiple classes. Figure 4 visualizes the segmentation displaying intermediate images and the resulting implant delineation $\gamma$ overlaid onto the original image $I$.

### 3.4 Bone Delineation in Presence of Implants

Segmentation of the bone structure consists of i) prepositioning the anatomical models close to the target structures and ii) adapting the models to the patient-specific anatomy. Either of these two steps has to deal with implants within the image and thus has to be altered compared to an approach dealing only with non-artificial image structure.

#### 3.4.1 Model Initialization

Similar to Prakash et al. [21] we make use of profiles perpendicular to long bone dia-
These profiles show characteristic symmetrical properties in their intensity. We exploit these characteristics to extract the anatomical axes of the femur and tibia. Thus, the initialization problem shrinks to one dimension, namely a search along the axes. Upon closer inspection of long bone appearance in digital radiographs, one recognizes a generic pattern. Typically the human body is surrounded by air and X-radiation passing through hits the detector almost unattenuated. Moving into the body, the absorption within the soft tissue areas of the leg produces the first contrast step. Further inside, the bone, especially the compact bone as its outer shell, produces a strong edge and the highest absorption. Finally, the interior part of the bone, the medullary cavity, containing the bone marrow, again has a lower density which results in less absorption. Since the bone as well as the leg is approximately cylindrical, these stages repeat on the opposite side. Refer to Figure 5 and the work of Prakash et al. [21] for a depiction of long bone contrast in radiographs together with a typical profile of a horizontal cut through a long bone. The long bone appearance is symmetric to an axis passing through the centre of the medullary cavity, which exactly represents the anatomical axis of the bone. To extract this axis we hence compute symmetrical properties with a dedicated operator.

Patients are radiographed from top to bottom, either standing upright or lying on a table, i.e. the legs are oriented vertically. Without loss of generality we restrict the search for symmetry to horizontal axial symmetry using the operator

$$\psi(x, y) = \sum_{l=1}^{L} |I(x + l, y) - I(x - l, y)| \cdot e^{-\frac{l}{\sigma}}$$

with $L$ denoting the capture range of the symmetry operator and $\sigma$, set to the average bone diameter (Fig. 5), controlling the strength of the decay, i.e. the weight of the differences depending on their distance to the assumed symmetry axis. This method for bone symmetry detection does not require any preliminary segmentation. Applied to the complete image, the operator efficiently generates a map of horizontal axis symmetry, with a local minimum corresponding to a candidate of the anatomical axis.

In order to avoid artificial objects, such as implants, but especially radiation pro-
tions from producing high symmetry properties, they are detected and delineated via a prior implant delineation step and the symmetry measure \( \psi(x, y) \) inside each of the artificial objects is set to \( \psi(x, y) = \infty \) (Fig. 6).

Once all candidates are identified it is straightforward to extract the axes themselves by linear regression and inter- as well as extrapolation of missing segments. In addition to the axis we also compute a local bone diameter by locating the first significant edges in either direction from the symmetry axis. Finally, we derive the location and orientation of the anatomic bone axis, together with a local bone thickness. These parameters can directly be used to set up a similarity transformation and transform the mean shape representation to its possible location, with the appropriate scale and rotation.

With the anatomical bone axis known, the initialization problem is reduced to a one dimensional search along the axis until the model "snaps in", i.e. fits with minimum energy, at a certain position. The important side effect is that not only the translation of the initial model is computed, but also an initial tilt angle is derived from the orientation of the symmetry axis and the scaling factor is derived from the previously extracted local bone thickness. Thus, instead of a full four dimensional search for all possible model translations on a two-dimensional image with evaluating different angles and different model scales, the number of model positions to examine shrinks to the number of pixels in the vertical image dimension.

A correlation analysis of initial angle and scale compared to the reference angle and scale derived from manual annotation reveals that the proposed initialization scheme well predicts these two parameters with correlations between 0.92 and 0.99 for the orientation angle and between 0.82 and 0.97 for the scale (Fig. 6).

First of all this strategy results in a much smaller computational complexity, but it also robustly avoids local optima. To rate the quality of a certain initial model position the trained structure of the model is reused and compared to the local structure of the current position via the Mahalanobis distance (Eq. 3). The minimum distance therefore corresponds to the best initialization, which is selected as the first iteration for the subsequent model adaptation. Figure 6 depicts an example of the localization via bone symmetry and subsequent segmentation for the three joints of a radiograph after total knee replacement.

### 3.4.2 Model Adaptation

Generally, operated legs still contain major bony parts with only small regions, especially the joint surfaces, influenced by the implant. Thus, the introduced shape models still hold for the preserved bone, whereas the joint areas are partially or fully replaced by implants (Fig. 7). However, due to their materials, implants produce strong contrast in the radiograph on their interface to soft tissue or bones. The corresponding strong gradients attract the trained model features and thus hamper the segmentation of the remaining bone structure of post-operatively examined legs. Hence, without modification the segmentation algorithm for pre-operative legs fails for artificial objects within the capture range of the template model and it thus has to be adapted to reliably segment joints with implants. To avoid the implant edges from attracting the shape model a modified approach checks whether one of the search vectors \( \psi \) overlaps the segmented implant region \( \gamma \). For any landmark with such a deteriorated search vector the weight, \( w_i \), is lowered to 0 in order to avoid the corresponding gradient to influence the deformation of the fitted shape model. As we use a coarse-to-fine approach with increasing image resolution, landmarks that have been disabled on a coarser level might get a valid weight, \( w_i \), on a finer resolution and contribute to the delineation. Thus, it is possible to achieve maximum accuracy while maintaining the robustness against artificial objects. After the deformable template model has converged, the bone shape is merged with the implant contour. In order to do this, intersection points between the fitted shape and the implant contour need to be identified. The bone contour and the implant outline might not necessarily intersect at all, i.e. the bone segmentation encloses the implant. In that case the points on the medial and lateral side of the implant with minimum distance to the bone contour are identified and used as intersection points instead. In a last processing step, the portions of the fitted shape in between the intersection points are replaced by the respective implant borders (refer to Fig. 7 for a depiction of the fusion of contours).
4. Experiments

The performance of a method deriving bone morphometry from digital radiographs is characterized by two quantities: i) The rate of automatically processed images and ii) the accuracy and precision of the derived measurements. In order to evaluate the former we determined the number of successful joint localizations on a set of 70 long leg standing digital radiographs with a pixel spacing of 0.143 mm containing a total of 291 joints. We compared our approach to an initialization developed by Ruppertshofen et al. [22] based on the generalized Hough transform (GHT) using the same images. As both methods generated different outputs, i.e. translation, rotation, and scaling for the proposed method and translation only for the GHT, localization was considered successful if the subsequent segmentation converged with a mean curve-to-curve error of less than 2.0 mm.

In order to evaluate accuracy and precision of the method we compared the clinically most relevant angular measures of knee joint alignment. These measures represent the mechanical leg axis and joint line alignment as depicted in Figure 8. A population of 60 pairs of pre- and post-operative long leg standing radiographs with a pixel spacing of 0.300 mm was drawn from an implant evaluation study [23] along with the angles originally documented using the clinical PACS system within the study as ground truth. A second set of angle measurements was generated by the last author (MD, former specialty registrar in orthopedics) using a dedicated, proprietary, manual measurement application. The third set of angle measurements was generated from a prototype version of the algorithm as introduced above that included a manual shape model positioning correction step (Fig. 9). This was necessary because the images in this evaluation came from a different acquisition chain than the ones used for training of the algorithm, exhibiting other features that hampered automatic prepositioning but not automatic segmentation and measurement derivation, which was the subject of this evaluation part.

Each set of measurements (clinical ground truth, second medical observer, algorithmic result) was generated independently from each other on the same set of 60 pre- and post-operative images, respectively, of 60 legs with the knees treated for osteoarthritis by implantation of a bicellular, posterior cruciate ligament retaining total knee arthroplasty [23]. The results of the second observer and the algorithm, respectively, were each compared with ground truth per combination of angle (Fig. 8) and pre- or post-operative stage. The comparison was carried out by descriptive statistics of the absolute deviations (mean, standard deviation, and range) and by correlation analysis.

5. Results

Using the proposed initialization scheme achieves a localization performance of 92.1% compared to 87.6% using the GHT [22]. The failed localizations can be divided into three groups according to the corresponding joint, namely 5.2% hip, 1.0% knee, and 1.7% ankle joints for the proposed method and 7.2% hip, 1.0% knee, and 4.1% ankle joints for the GHT.

For the considered knee implants we achieved a detection rate of 99.2% out of 132 total implants. The average true positive rate of the delineation was 99.1%.

The evaluation on the clinical study data set showed comparable deviations for both the algorithm and second observer from ground truth with an average accuracy in-

Table 1 Precision, accuracy, and correlation of knee alignment measurement – mechanical lateral distal femur angle (mLDFA), mechanical medial proximal tibia angle (mMPTA), mechanical axis deviation (MAD) – derived using our method.

<table>
<thead>
<tr>
<th>Method</th>
<th>Pre-operative</th>
<th>Post-operative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Deviation</td>
<td>Correlation</td>
</tr>
<tr>
<td>mLDFA</td>
<td>0.8 ± 0.6</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.4 ± 0.7</td>
<td>0.99</td>
</tr>
<tr>
<td>mMPTA</td>
<td>1.3 ± 1.1</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9 ± 0.9</td>
<td>0.90</td>
</tr>
<tr>
<td>mTFA</td>
<td>1.2 ± 1.0</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0 ± 0.7</td>
<td>0.93</td>
</tr>
</tbody>
</table>
Table 1 reflects the performance after correcting the corresponding angles.

6. Discussion

The results in terms of accuracy and precision are sufficiently well below the variability induced by patient positioning [24]. Furthermore, they do not exceed clinically relevant deviations [25]. The high correlation of $r = 0.99$ and $r = 0.97$ (pre- and post-operatively) of the algorithm with the ground truth data for the mechanical tibiofemoral angle matches the performance of our second observer and even a typical intra-observer correlation of $r = 0.98$ reported in literature [26].

Still the maximum deviations warrant further consideration. The inherent objectives of automating the initialization of long leg morphometry measurements are increased workflow efficiency and decreased rate of error. Measuring times of 11 min 46 s and 6 min 34 s per leg for a comprehensive set of angles [8] and 4 min 54 s and 1 min 5 s for the mechanical axis alignment only [27] have been reported, using manual and computer-assisted measurements, respectively. In comparison the proposed method significantly speeds up the workflow with an automated processing time of 10 s.

Potential sources of error for manual measurements in the clinical setup comprise result transcription, i.e. wrong patient (study ID), wrong leg (laterality), wrong angle name (entity), and result generation, i.e. wrong angle value (inaccuracy of placing the measurement tool), wrong angle sign (mis- taken side of angle opening). The reported outliers in the ground truth study did not adversely affect the original results of the study [23] in three of the four instances, because the absolute value of the angular deviation, invariant of the opening side, was evaluated there. Still both automatic analysis as well as manual reading using the proprietary workflow were able to detect the outliers. In case of the secondary manual reading this is attributed to the fact that the measurement setup directly links images to numbers and the data is acquired in an integrated workflow without any manual transcription, thus removing the potential sources of error that are given above for transcription and the wrong angle sign error.

7. Conclusions

The proposed method allows deriving objective measures of joint alignment from clinical radiographs. Its accuracy and precision are on par with experienced readers trained in orthopedics for the evaluated measurements of knee morphometry.

Acknowledgments

The authors would like to thank the clinical study team and the clinical study sponsor for providing us with the image and radiographic result data, namely Dr. Anusch Sufi, Dr. Frank Lampe, Dr. Karina Bohlen and Prof. Dr. Ekkehard Hille of Schön Klinik Eilbek, Hamburg, Germany, as well as Dr. Andrea Weiler and Mr. Norbert Pfeiff of B. Braun Aesculap AG, Tuttinglen, Germany. We also thank Mr. Thomas Pralow and Mr. Eugen Hermann of Philips Healthcare, Hamburg, Germany, for guiding and supporting the prototype development.

References