Automatic Knee Joint Space Measurement from Plain Radiographs

Tuomas Frondelius

University of Oulu

Biophysics

2020

ABSTRACT

Knee osteoarthritis is a common joint disease and one of the leading causes of disability. The disease is characterized by loss of articular cartilage and bone remodeling. Tissue deformations eventually lead to joint space narrowing which can be detected from plain radiographs. Joint space narrowing is typically measured by an experienced radiologist manually, which can be time consuming and error prone process. The aim of this study was to develop and evaluate a fully automatic joint space width measurement method for bilateral knee radiographs.

The knee joint was localized from the x-ray images using template matching and the joint space was delineated using active shape model (ASM). Two different automatic joint space measurement methods were tested and the results were validated against manual measurements performed by an experienced researcher. The first joint space width measurements were done by binarizing the joint space and measuring the local thickness of the binary mask using disk fitting. The second method classified bone pixels to tibia and femur. Classification was based on the ASM delineation. Nearest neighbors between femur and tibia were then used to find the joint space width. An automatic method for tibial region of interest (ROI) selection was also implemented. The algorithms used in this thesis were also made publicly available online.

The automatically obtained joint space widths were in line with manual measurements. Higher accuracy was obtained using the disk fitting algorithm. Automatic Tibial ROI selection was accurate, although the orientation of the joint was ignored in this study. An open source software with a simple graphical user interface and visualization tools was also developed. Computationally efficient and easily explainable methods were utilized in order to improve accessibility and transparency of computer assisted diagnosis of knee osteoarthritis.

TIIVISTELMÄ

Polvinivelrikko on eräs yleisimpiä niveltauteja sekä yksi merkittävimmistä liikuntavammojen aiheuttajista. Nivelrikolle ominaisia piirteitä ovat nivelruston vaurioituminen ja muutokset nivelrustonalaisessa luussa. Kudosten muutokset ja vauriot johtavat lopulta niveltilan kaventumiseen, mikä voidaan havaita röntgenkuvista. Tavallisesti kokenut radiologi tekee niveltilan mittaukset manuaalisesti, mikä vaatii usein paljon aikaa ja on lisäksi virhealtis prosessi. Tämän tutkielman tavoitteena oli kehittää täysin automaattinen niveltilan mittausmenetelmä bilateraalisille polven röntgenkuville.

Polvinivel paikallistettiin röntgenkuvista muotoon perustuvalla hahmontunnistuksella ja nivelväli rajattiin käyttämällä aktiivista muodon sovitusta (active shape model, ASM). Nivelvälin mittaukseen käytettiin kahta eri menetelmää, joita verrattiin kokeneen tutkijan tekemiin manuaalisiin mittauksiin. Ensimmäinen nivelvälin mittausmenetelmä sovitti ympyränmuotoisia maskeja niveltilasta tehtyyn binäärimaskiin. Toinen mittausmenetelmä luokitteli luuhun kuuluvat pikselit sääri- ja reisiluuhun. Luokittelu perustui aikaisemmin tehtyyn automaattiseen nivelvälin rajaukseen. Nivelvälin mittaukseen käytettiin lähimpiä naapuripikseleitä sääri- ja reisiluusta. Työssä kehitettiin myös menetelmä automaattiseen sääriluun mielenkiintoalueiden (region of interest, ROI) valintaan. Käytetyt algoritmit ovat julkisesti saatavilla verkossa.

Automaattiset nivelväli mittaukset vastasivat manuaalisia mittauksia hyvin. Parempi tarkkuus saatiin käyttämällä ympyrän sovitusta hyödyntävää algoritmia nivelvälin mittaukseen. Sääriluun mielenkiintoalueet onnistuttiin määrittämään automaattisesti, tosin nivelen orientaatiota ei huomioitu tässä työssä. Lisäksi kehitettiin avoimen lähdekoodin ohjelmisto yksinkertaisella graafisella käyttöliittymällä ja visualisointityökaluilla. Työssä käytettiin laskennallisesti tehokkaita ja helposti selitettäviä menetelmiä, mikä edesauttaa tietokoneavusteisen menetelmien käyttöä polvinivelrikon tutkimuksessa.

Table of Contents

1. Introduction

Osteoarthritis (OA) is a degenerative joint disease with no known cure and it is one of the leading causes for disability [1]. Symptoms of the disease include pain and loss of motion in the affected joint, which can lead to reduced mobility and ability to work. OA is also associated with increased risk of mortality, due to reduced physical activity [2]. Furthermore, OA carries significant economic impact from expensive treatments and reduced participation to work force [2,3]. Detection of OA progression and identification of risk factors are important steps in determining the intervention methods and treatments for the disease [1,4,5]. OA treatment focuses on slowing the disease progression and improving the patients quality of life, as there is no known cure for the disease [1,5]. Identifying the risks and detecting progression of the disease early is important for effective treatment.

OA is characterized by loss of cartilage and bone remodeling in the affected joint [3,4,5]. Joint space narrowing is a strong indicator of OA progression [1]. Joint space narrowing can be detected and measured from plain radiographs, by measuring the distance between femoral head and tibial plateaus [1,5]. OA induced changes in the subchondral bone can also be detected from radiographs using numerical analyses [1]. Bone analyses can be done from regions selected from tibial subchondral bone. Manual annotation of images and joint space width (JSW) measurements are subject to intra- and inter-observer variability and human error [6,7]. In order to alleviate the measurement errors and speed up the analyses, it is necessary to develop automatic methods for delineation of the joint space.

In the past 20 years, various computer vision and machine learning algorithms have been developed for automatic segmentation and analysis of medical images [8]. Region growing, wavelets, deformable models, statistical shape models, and atlas based methods have been utilized in image segmentation and delineation of anatomical regions of interest (ROI) [7,8,9,10]. Feature extraction methods, such as histograms of oriented gradients (HOG), local binary patterns (LBP) and deep convolutional neural networks (CNN) have also been utilized in segmentation and classification tasks for medical images [1,11,12,13,14].

Aforementioned segmentation algorithms can be utilized in the JSW measurement. By delineating the femur and tibia, joint space can be measured by finding the distance between appropriate regions from the segmented shapes. Chen et al used HOG to locate landmarks around bones on hip, knee and cranial radiographs [11,12]. HOG features were used to construct a descriptor for small randomly selected regions around the landmarks. Corresponding regions were then identified from test image and a voting based method was used to find the landmark points. A statistical shape model was then fit to the bone landmarks selected by the algorithm.

Podsiadlo et al. used active shape model (ASM) to delineate cortical bone plates from tibia and to select bone ROIs automatically [15]. ASM was used to delineate the cortical bone plate from a preprocessed region around the joint. Results of the ASM were then used to automatically select tibial ROIs for further analyses. Woloszynski et. al. used similar tibial ROIs to predict OA progression based on the numerical analysis of the bone in a follow up study [16]. OA progression was predicted based on roughness and anisotropy measurements. Hirvasniemi et al also used numerical analyses from tibial ROIs to determine OA progression using LBP [1,17].

Modern approaches in image segmentation and analysis typically utilize CNNs [13,14]. Ambellan et al utilized U-net CNN architecture to segment knee bone and cartilage from magnetic resonance imaging (MRI) images [13,18]. Zhou et al used a 2D CNN for multi-class tissue segmentation from knee MRIs [14]. CNN based approaches usually achieve highest accuracy in image analyses and can even perform at human level [19]. However, the drawback of CNNs is that they require large datasets and powerful workstations to learn the appropriate features from the data [20]. Furthermore, the features extracted by CNNs are not easily explained in human terms, and even visualization of the regions used in decision making require advanced computational algorithms [19,21].

Objective of this study is to develop an automatic method for knee JSW measurement from plain radiographs. Fully automatic JSW measurement was performed using template matching and ASM. The automatic JSW measurement pipeline is described in Figure 1. Joint was first localized using template matching, followed by bone delineation using ASM. The

delineation results were used to automatically place tibial ROIs to the radiographs. Finally, JSW was measured using two different methods.
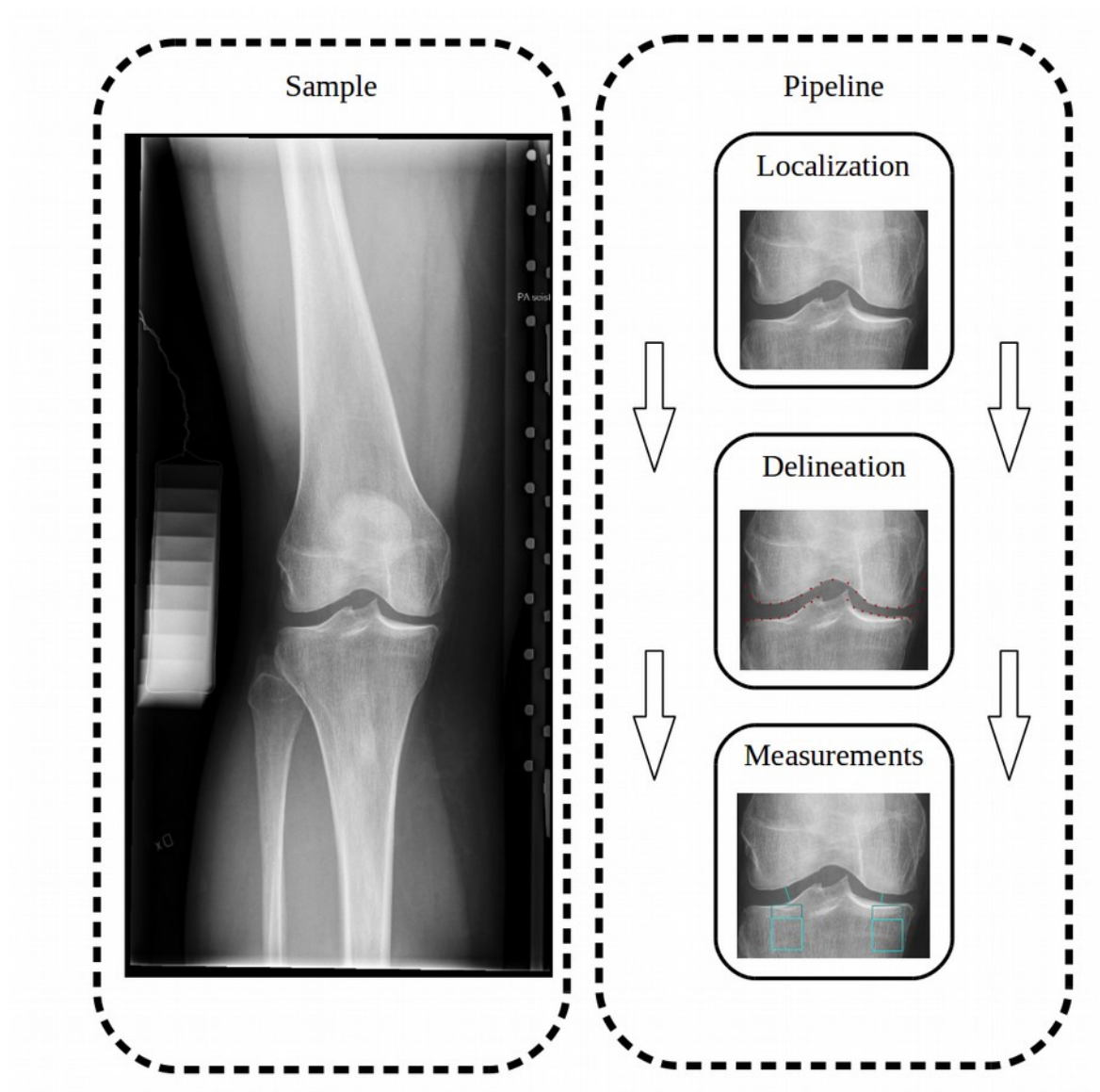

Figure 1. Visual representation of the joint space width (JSW) measurement pipeline. At first, the joint is localized using template matching, followed by joint delineation. The joint space is delineated using ASM. Finally, the JSW is measured and tibial ROIs are placed on the subchondral bone.

2. Osteoarthritis

Osteoarthritis (OA) is a degenerative joint disease characterized by loss of cartilage and remodeling of bone in the affected joint [1,2,22]. Symptoms of the disease include pain, stiffness and loss of motion in the joint [2,22,23]. Intermittent pain is often among the first symptoms of the disease, and the pain may become constant over time [2,23]. OA can significantly reduce the individual's quality of life due to pain and reduced mobility [1,2]. Higher rates of mortality have also been associated with the disease, most likely due to decreased physical activity and the resulting elevated risk of cardiovascular problems [2]. Common joints affected by the disease are knees, hip and hands.

Osteoarthritis of knee is the most common form of OA in large load bearing joints, most often affecting the medial compartment of the joint [24]. Knee OA is among the leading causes of pain and physical disability among adults [1,22,23]. Knee OA affects approximately 5% men and 7% percent women in Finland and the disease prevalence is expected to increase in the future as the population ages [1,3]. Knee OA has a significant economic impact, due to treatment costs and reduced participation to the work force [3]. There is no cure for the disease, and the existing treatments focus on pain management using medication and improvement of mobility with physical therapy [1,2,5,22]. Total knee replacement (TKR) is one of the most effective treatments and approximately 108 TKR surgeries are performed per 100 000 people in Finland alone [3]. Furthermore, approximately 6% of disability pensions are granted based on diagnosed knee OA in Finland [3]. Knee OA related costs add up approximately 1 billion Euros annually in Finland. Since the disease has no cure and requires costly treatments, early detection and diagnosis are important [1,22]. Treatment outcomes are improved if the diagnosis is acquired before joint degeneration is advanced, and the disease progression can be slowed [2].

Diagnosis of OA is usually done based on symptoms (symptomatic OA), patient history, physical examination and radiographs (radiographic OA) [1,2,23]. Native x-ray imaging is typically used to diagnose radiographic OA, due the low cost and speed of the imaging modality [1,25]. Radiographic features of OA include joint space narrowing, bone remodeling and changes in the bone texture [1,2,25]. Joint space narrowing is correlated with the loss of

articular cartilage (AC), which may be completely eroded in advanced cases of OA [1,5,26,27]. OA induces changes in the bone include osteophytes, also known as bone spurs, cysts and subchondral bone sclerosis (bone thickening) [1,2]. Furthermore, numerical analysis of the bone texture reveals OA changes not visible or difficult to interpret with the naked eye [1,25]. Radiographic OA may be present in patients who display no symptomatic OA, as the tissue changes may occur without pain or stiffness [1,25]. Symptomatic OA may also occur without radiographic features, although the tissue degeneration may be visible using different imaging modality [25].

Risk factors of OA include age, obesity, female sex, genetic factors and previous history of injury [1,2,22]. Risk of developing OA increases with aging, possibly due to decreased regenerative abilities of the tissue [1,2]. Other risk factors, such as obesity and sedentary lifestyle, also increase with aging, leading to higher risk of OA [1,2,28]. Obesity has been linked to higher tissue degeneration in load-bearing joints, such as knees, and higher OA risk in non-load-bearing joints [2,29]. OA of hands and knees has higher occurrence in female patients, and the hip OA progresses more rapidly in women [5]. Joint injuries are also linked with higher OA risk [1,2,30]. Meniscal and anterior cruciate ligament (ACL) injuries correlate with post-traumatic knee OA development, although the exact reason for the increased risk is not known [30]. Potential explanations may be damaged cartilage when the injury is received and mechanical changes in the injured joint.

2.1 Osteoarthritis induced tissue degeneration

OA progression induces biochemical and mechanical changes in AC [16,17,27,31,32]. AC is a specialized joint tissue, which provides a smooth surface for the articulations and aides in load transmission in the joint [31]. AC is comprised of extracellular matrix (ECM) and sparsely distributed cells called chondrocytes [27,31]. AC is devoid of blood vessels and relies on synovial fluid for metabolism [32]. As the disease progresses, ECM becomes disorganized and chondrocyte disorganization and clustering becomes more prominent [27]. OA induced biochemical changes in the AC lead to changes in the mechanical properties of the tissue. The elastic and shear moduli decrease, resulting in greater deformations of the tissue under load [32]. Furthermore, AC degenerates as the disease progresses, manifesting as

discontinuities and tears in the tissue [27]. In advanced cases, the tissue might be completely lost, leaving the subchondral bone exposed.

The disease also changes the properties of subchondral bone (SB) [1,16,17,27,32,33]. SB includes subchondral trabecular bone and subchondral plate, found immediately beneath the cartilage [4,18,32,33]. SB has a porous structure which houses blood vessels and nerves [33]. Trabecular bone and AC are separated by subchondral plate, which allows diffusion of nutrients into the AC [31,33]. Although the subchondral plate is less porous than trabecular bone, some blood vessels and nerves penetrate the structure [31]. The micro-architecture of the bone changes as a result of OA progression, leading to thinning of subchondral bone plate and increased porosity [1,2,33]. However, in advanced cases, the bone density is increased, although the mineralization and stiffness of the bone decreases [1,32]. The increased density may occur as a compensation for the decreased mineralization and loss of bone stiffness [4]. The mechanical properties, such as elastic and shear moduli, are decreased in osteoarthritic bone [32].

2.2 Knee anatomy and osteoarthritis

Knee joint is a modified hinge joint and one of the largest weight bearing joints in the human body [24,34,35]. The complex structure of the joint allows both anterior-posterior translation as well as some axial rotation [34,35]. The joint consists of two distinct articulations, the tibio-femoral joint and patello-femoral joint [35]. Tibial and femoral head delineate the tibio-femoral articulation. The bones are separated by meniscus and AC, which help facilitate the relative motion between the bones [34]. Stabilization and motion of the tibio-femoral is facilitated by surrounding muscles and ligaments. Muscles, such as quadriceps and the hamstring group muscles, facilitate locomotion, with ligaments, such as ACL and PCL, acting as antagonists to the muscles and providing stability [35].

Patello-femoral joint is delineated by patella and femur and the articulation acts as an extensor mechanism in anterior-posterior translation [34,35]. Patella is connected to the joint via capsular ligaments and patellar tendons. Tibia is separated from patella via fat pad and infrapatellar tendon bursa [35]. Patello-femoral joint is also subject to OA. Patello-femoral

OA carries similar symptoms as other forms of OA, and is characterized by similar tissue degeneration [36].

Injuries to the supportive and connective tissues of the knee increase the risk of developing knee OA [24,30,34,35]. Meniscal tears and ACL injuries carry especially high risk of developing post-traumatic OA [24,30]. Furthermore, OA may impact the supportive tissue function and the disease has been associated with ACL damage [2,37]. Patients with knee OA walk with greater muscle contraction, which may change the load distribution in the joint and cause further tissue degeneration [37].

2.3 Radiography and knee osteoarthritis

Native x-ray images are used in knee OA diagnosis along with the patient history, physical examination and symptoms [1]. X-ray imaging was developed based on the discovery made by Wilhelm Roentgen in 1895 [38]. First medical imaging applications were developed by the year 1900. The discovery awarded Roentgen the first Nobel prize in physics. X-ray images are 2D projections of the subject, where contrast is a result of the different X-ray attenuation between bone and soft tissue [1,38]. The high calcium content of bone gives it higher attenuation constant than soft tissue, such as muscles and fat [39]. As x-ray photon pass through an object, the intensity of radiation is attenuated according to

$$I_x = I_0 \times e^{-\mu x} \qquad (1),$$

where I_x is the intensity of the radiation at depth x, μ is the attenuation constant and x is the thickness of the material. Attenuation constant μ is proportional to the atomic number of the material, density of the tissue and the energy of the radiation [1].

X-rays are produced by accelerating electrons from a cathode filament to heavy metal anode. Electrons are emitted from cathode material via thermoionic emission and the emitted electrons are accelerated to the anode using kilo Volt (kV) range voltage [38,40]. As the electrons interact with the anode, they are either decelerated or absorbed, resulting in a

spectrum with characteristic intensity spikes at certain photon energies [40]. The kinetic energy of the photons is emitted as high energy radiation. X-rays are detected after passing through the subject, and the intensity of the radiation is measured by a 2D array of detectors. The detector is either made using material capable of measuring high energy electromagnetic radiation, or with a scintillator which converts the radiation to visible part of the electromagnetic spectrum [41].

Radiographic knee OA is evaluated using extended-knee radiograph, where the subject is standing and with both knees flexed and a bilateral anterio-posterior radiograph is acquired [25]. Tissue specific features, such as osteophytes, bone sclerosis and cysts, are visible from plain radiographs with JSW narrowing [1,25]. The severity of radiographic OA is described using semi-quantitative metrics, such as Kellgren-Lawrence (KL) scale [1,25,42]. The grading scale ranges from 0 (no OA features) to 4 (severe OA) [25,43]. The KL grading scale is commonly used as a metric of OA progression, but it's criticized for assuming OA to progression to be linear [25]. Furthermore, because KL grading is semi-quantitative metric, it suffers from intra- and inter-observer variability [1].

3. Computer vision in Medical imaging

Computer vision is a cross-disciplinary field attempting to obtain high level understanding of images using mathematical and numerical methods, such as signal processing and pattern recognition algorithms [44,45]. Common problems in computer vision include object detection and tracking, image classification and segmentation. Computer vision problems are difficult, since they are inverse problems, where parameters which characterize the system are unknown [44,46]. Humans and animals are able to accomplish these tasks effortlessly due to the complex neurological systems involved in vision, while computers may struggle with these problems [44]. However, modern computer vision methods can achieve human level performance and even outperform humans in certain simple tasks [47].

Designing computer vision applications and algorithms requires extracting non-redundant information from an image or a set of images, fine-tuning the parameters of the model and model validation. Feature extraction can be done using hand-crafted features, such as HoG, or LBP features, or using artificial neural networks, which learn the relevant features from the annotated data [48,49,50]. Computer vision model performance can usually be improved by fine tuning the model parameters, such as number of bins used to extract HoG features, or changing the hyper-parameters of a neural network [48,50]. Selecting appropriate validation metrics for the model performance is also important. Statistical analyses of the performance metrics can reveal systematic errors and potential edge cases from the dataset, which should be taken into account [47].

Computer vision has multitude of applications in medical imaging ranging from image registration and segmentation to automatic diagnosis and abnormality detection. Manual analysis of medical images can be difficult due to noisy images, unclear boundaries and poor contrast between tissues and varying size and shape of the organs [8,51]. Image registration can be used in surgical planning and atlas based segmentation tasks [8,52]. Registering an annotated image with a test image yields an approximation of the tissue segmentation for the test image [52]. This can be done automatically, by detecting and matching landmarks from the labeled image and the test image. Automatic tissue segmentation can be done using

feature extraction, deformable models or CNNs [11,12,13,14,15,51]. Tissue segmentation can be used to examine the geometry of the organs and to find important anatomical landmarks. Computer vision and machine learning can also be utilized in automatic diagnosis and disease prediction [1,16,17,53]. Medical images contain a large amount of data, which is invisible to the naked eye. Numerical analyses of the tissues can reveal changes is tissue microstructure and composition [1,16,17,53].

3.1 Digital image processing and medical imaging

Digital image processing (DIP) has been utilized in the analysis of natural images since 1960s, when computers became powerful enough to process images in a meaningful way [54]. Images are represented either as 2D matrices where graylevel intensity is a function of the coordinates. The image is then represented by a function f(x,y), where f is the graylevel intensity and x and y are coordinates. Natural images are often represented as a 3 dimensional tensor, where the third dimension contains red, green and blue channels, while the first two dimensions remain as spatial coordinates.

DIP techniques have been used in medical imaging since late 1970s, when advent of computed tomography (CT) lead to digitization of the images [41]. Plain radiographs are 2-dimensional matrices, while CT images and MRI images are 3-dimensional tensors. Color channels are typically not utilized, since regular x-ray detectors cannot bin the photons based on their energy [41]. DIP techniques are utilized in medical imaging from low-level operations, such as denoising and sharpening, all the way to higher level operations such as computer vision tasks discussed in the previous chapter. The complexity of DIP operations can be characterized by the output of the operation. Output of the low-level operations are slightly modified versions of the input images and output of the mid-level operations are heavily modified images, such as segmentation masks of objects or gradient images, while high-level operations attempt to perform cognitive functions associated with vision and generally fall under computer vision [54].

Low level DIP applications in medical imaging include noise reduction and contrast enhancement. Low-level noise reduction is typically performed using digital filters, such as

mean, median and Gaussian filters. Example Gaussian kernel is shown in figure 2. Images are filtered by sliding a moving filtering kernel over the image, as shown in figure 2. This can be expressed mathematically using convolution or correlation, although correlation and convolution are often identical operations since the low-level kernels tend to be symmetrical [54]. Image sharpening can be done using unsharp masking, where a smoothed image is subtracted from the original image, and the result is added to the original image (see equation 2). Unsharp masking can be expressed as

$$I_{sharp} = I + (I - G) \qquad (2),$$

where I_{sharp} is the sharpened image, I is the original image and G is smoothed image. Contrast enhancement can be performed using histogram equalization, or gamma-correction. Histogram equalization is performed by mapping the original distribution of the graylevel values into a flat distribution. Gamma correction is done by raising the graylevel values to some power k (see equation 3). Gamma correction can be expressed as

$$I_\gamma = I^k \qquad (3),$$

where I_γ is the gamma-corrected image, I is the original image and k is some, typically small, value. Whether k increases or decreases the luminance, depends on the graylevel values. If the graylevel values are scaled between 0 and 1, $k<1$ increases the luminance, and $k>1$ decreases the luminance. If graylevel values are integers, e.g. 8 or 16-bit typically, $k<1$ decreases luminance and $k>1$ increases luminance.

Mid-level DIP operations in biomedical imaging include image stitching, edge detection, and simple segmentation methods. Stitching can be used to combine multiple images into a single image, which allows high resolution imaging of large samples [55]. Edge detection can be performed computing image gradients, which are then thresholded to obtain a binary edge image [54,56]. Simple binary segmentation tasks can be performed by thresholding the image [54]. However, medical images can be noisy, and tend to have poor contrast between tissues, making thresholding a poor choice for medical image segmentation. Decent segmentation can be achieved if the thresholding is applied locally, rather than globally (see section 3.3).

Segmentation quality also depends on the selected threshold. Appropriate threshold can be selected based on graylevel histogram by selecting Nth percentile or by using Otsu's method [57]. Automatic binary segmentation can also be done using the following algorithm:

1. Compute mean graylevel
2. Threshold, and compute mean graylevels for the segmented regions
3. Compute the average of the mean graylevels
4. Threshold image using the average mean graylevel from 3.
5. Repeat 3. and 4. until the average of the two mean graylevels converges.

Binary segmentation results can be improved using morphological operations [54]. Morphological operations are based on set theory and include dilation, erosion, opening and closing [54]. Dilation and erosion are performed by sliding a structuring kernel over a binary image. Dilation of binary image is performed by moving a dilation kernel over the image, which add non-zero pixels to the edges of the foreground object. Erosion is performed by sliding a similar kernel over the binary image, but pixels are removed from the foreground object. Opening is performed by first eroding the image followed by dilation. Closing is performed in a similar fashion, but with order of morphological operations reversed, i.e. dilation is done before erosion.
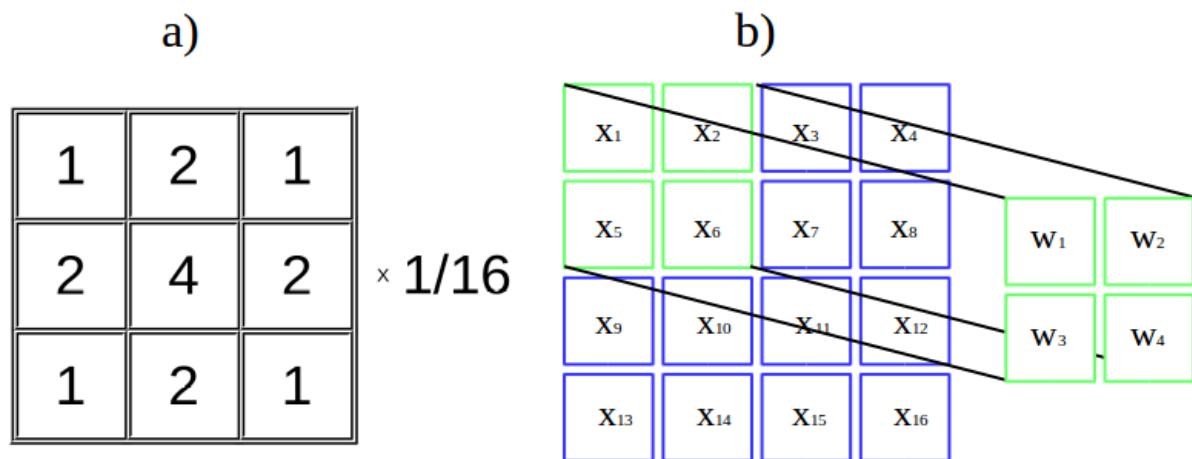

Figure 2. Digital image processing kernel visualization. Filtering an image with a Gaussian kernel (figure a) yields a blurred image, as the kernel outputs a weighted sum of the area under the receptive field of the kernel (figure b). Kernel is moved across the image and the output is recorded in a new image.

3.2 Feature extraction

Feature extraction is used in computer vision to find non-redundant information from the input image. Extracted features are then used in the image analyses. Hand crafted features, such as HoG, scale-invariant feature transform (SIFT) and LBP have been used in landmark detection, object detection, image classification and segmentation tasks [1,11,12,55,58]. Feature descriptors are visualized in figure 3. HoG features are extracted by computing image gradients, and then binning the gradients by their orientation. The HoG descriptor has been utilized in object detection from natural images and landmark detection from plain radiographs [11,12,48]. The input image is split to cells, and the image gradients from each pixel belonging to a cell are binned according to their orientation. The cells are then split to spatially connected blocks, and the gradient bins belonging to a block are normalized. The normalization process reduces the effect of illumination. Chen et al used HoG descriptor to find landmarks from hip, knee and cranial radiographs [11,12].
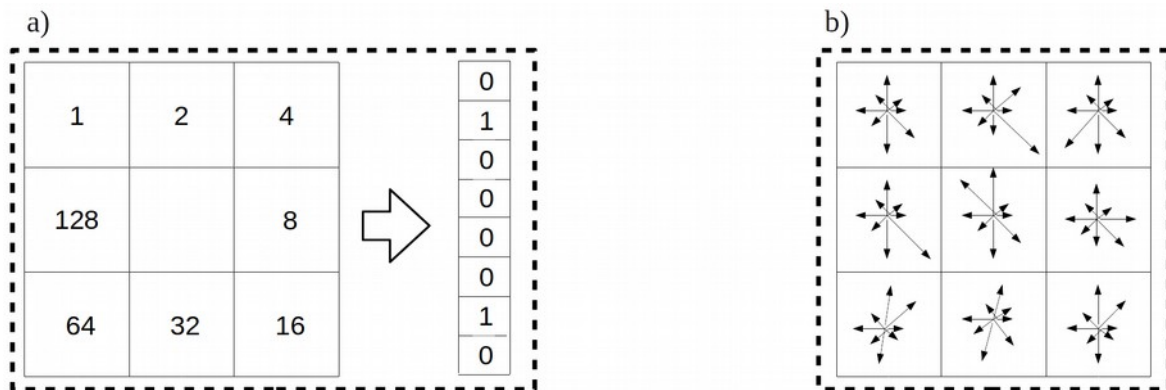

Figure 3. Common feature descriptors. Figure a) shows LBP descriptor. Neighboring pixel values equal or greater than the center pixel are set to 1, and a value based on their position is assigned to the non-zero pixels. Finally, the binary vector is converted to a numerical value. Figure b) shows a visualization of binned gradient features used in SIFT and HoG descriptors. Both of the aforementioned descriptors compile image gradient information under a small window, which describes the local image structure efficiently.

SIFT and it's more modern variant SURF, have been utilized in landmark detection [58,59]. SIFT descriptor uses difference of Gaussian (DoG) images to identify landmarks [54]. DoG images are computed according to equation 4. DoG computation can be expressed as

$$D = G_1 - G_2 \qquad (4),$$

where D is the DoG image and G_1 and G_2 are smoothed versions of the input image. The input image is smoothed by two Gaussian kernels, constructed using different standard deviation [58]. Several DoG images are computed with increasing smoothing using multiple spatial resolutions. Keypoints are selected from the adjacent DoG images, followed by removal of low-contrast points and edges. Gradient based descriptor similar to HoG descriptor is then computed around the keypoints. Landmarks from different images can then be matched using Euclidean (see equation 5.) distance between the feature vectors extracted from the images. Euclidean distance is given by

$$S = \sqrt{\left(\left(x_1^1 - x_1^2\right)^2 + \left(x_2^1 - x_2^2\right)^2 + ... + \left(x_n^1 - x_n^2\right)^2\right)} \quad (5),$$

where S is the distance, x_n^1 and x_n^2 are the nth element of the feature vector. Rister et al used a 3D variant of SIFT descriptor to register brain MR images and abdominal CT scans [60]. SIFT based descriptor SURF has been also utilized in biomedical image analyses [59]. Yang et al used a modified version of the SURF descriptor to perform automatically stitch high-resolution rat tissue samples [55].

LBP descriptor is constructed by comparing center pixel of a small radially samples neighborhood of pixels to its neighbors [49]. The center pixel is subtracted from the neighbors, and results are concatenated in a 1-dimensional vector. Elements below zero are set to 0, and rest of the elements are set to 1. Each element of the vector is assigned a value between 2^0 and 2^N, where N is the number of neighbors. Hirvasniemi et al used LBP features to analyze subchondral bone regions knee radiographs [1,17].

Modern feature extraction algorithms often utilize CNNs to find optimal features for computer vision tasks [13,14]. CNNs have been used in computer vision tasks since early 2000s, but they used to lack mainstream appeal in computer vision and machine learning community [50]. CNNs became the dominant approach in computer vision after the success of Krizhevsky et al in ImageNet competition, who used a CNN to beat the competitive algorithms by a large margin [50,61]. CNNs are trained on a manually labeled dataset. Optimization is done using variants of gradient descent algorithm [62]. The algorithm can be

used to learn a new set of optimal features, or a pre-trained network can be used in feature extraction, with a classifier trained on top of the pre-trained feature extractor [63]. CNNs trained for image segmentation typically use end-to-end training where new features are learned directly from the data.

3.3 Active shape models

Active shape model (ASM) was developed by Cootes et al in 1995 [64]. ASM is a deformable model, which can be deformed to fit data in a manner, which is consistent with its training data. ASM is trained by selecting landmarks from training dataset, which are then aligned using Procrustes transformation [51,64]. Principal component analysis (PCA) is then used to find the allowed modes of deformation from the aligned training shapes.

Landmarks are typically selected from distinctive locations, such as contour end-points and corners. Landmark selection is described in figure 4. Intermittent landmarks are selected from less distinct points, typically from half-way points of the more distinct landmarks selected initially. Intermittent points are used in order to capture more robust features from the shape. Training shapes are aligned by minimizing the energy function, described in equation 6.

$$E = \left(x_i - M(s,\theta)[x_j] - t_j \right)^T W \left(x_i - M(s,\theta)[x_j] - t_j \right) \qquad (6),$$

where x_i is the ith shape vector in format $x = [x_0, y_0, x_1, y_1, \ldots, x_{N-1}, y_{N-1},]^T$, t_j is a translation vector $t = [t_x^0, t_y^0, t_x^1, t_y^1, \ldots, t_x^{N-1}, t_y^{N-1}]$. W is a diagonal matrix of weights and $M(s_j, \theta_j)[x_j]$ is rotated and scaled *jth* shape vector. Rotation and scaling is done according to equation 7. Rotation and scaling is performed on *jth* coordinate matrix according to

$$M_j = \begin{pmatrix} s\cos(\theta) & \text{-s }\sin(\theta) \\ s\sin(\theta) & s\cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \qquad (7),$$

15

where x and y are vectors of coordinates, s is a scaling factor and θ is the angle of rotation. The output is formatted as $M_j = [M_{jx1}, M_{jy1},..,M_{jx(N-1)},,M_{jy(N-1)}]$. Shapes are aligned iteratively, according to following:

    1. Align training shapes with the first shape

    2. Compute mean shape

    3. Align shapes with the mean shape

    4. Repeat 2. and 3. until convergence.

Once the shapes are aligned, PCA is used to find k unit eigenvectors P and eigenvalues λ which cover ~97% of the total shape variance. Any allowed deformation of the model can be expressed using equation 8. Allowed deformations are given by

$$x = x_{mean} + Pb \qquad (8),$$

where x is a shape, x_{mean} is the mean shape, P is matrix of k eigenvectors and b is a vector of weights. The weights in equation 8 are constrained to a range $-3\sqrt{\lambda_k} \leq b_k \leq 3\sqrt{\lambda_k}$.
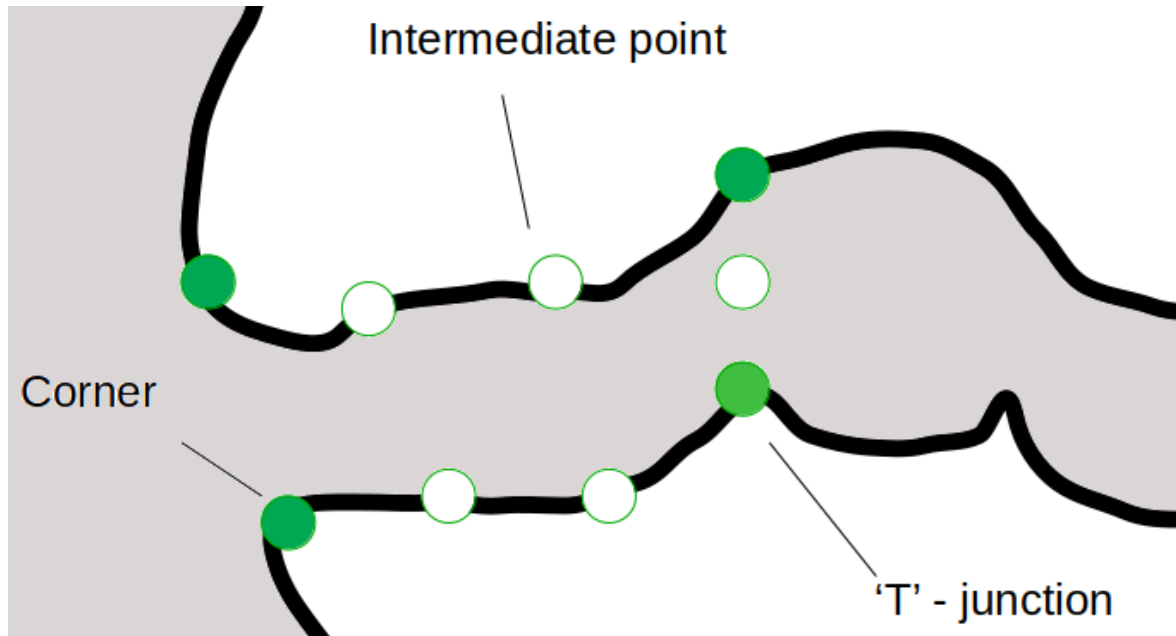

Figure 4. Active shape model (ASM) landmark selection. Landmarks are typically selected from distinct region such as corners and structure extrema. Intermediate points between more distinct landmarks can be used to create more robust representation of the object shape.

New landmarks are selected from test image by inspecting the grayvalues of the test image along the contour normals of the ASM. An appearance model or grayscale model can be constructed from the local grayvalues or their gradients sampled along the contour normal around the training points. Landmark points are sampled from a new image along the contour gradients after selecting an initial location for the model points. Best matching landmark points are selected from the image, and the model is transformed to fit the new landmarks, using equation 6. Residual adjustments are computed by subtracting the fitted model from the new landmarks, and shape constraints imposed by the eigenvalues are enforced using equation 8. Shape constraints are enforced by selecting weights b from a range set by the eigenvalues. Equation can written as

$$b = P^T \left(x - x_{mean} \right) \qquad (9).$$

Since the columns of P are mutually orthogonal and unit length, equation 9 holds as $P^T = P^{-1}$. The landmark search is repeated until the model converges to a location.

4. Aim of the Thesis

Main goals of this thesis were:

1. Development of automatic joint space delineation algorithm using template matching and ASM.
2. Development and comparison of two automatic joint space width measurement algorithms.
3. Implementation of the algorithms in an open source software, with robust visualization tools for the results.

Plain radiography is an important tool in osteoarthritis diagnosis due to the low cost and wide availability. Manual analyses of radiographs are susceptible to human error and intra- and inter-observer variability, which can be alleviated by automating the analysis process. Automatic segmentation and measurement results are repeatable as well as faster than manual analyses. In this study, ASM based algorithm was used to delineate the joint space. The acquired landmarks were then used in JSW measurements. ASM can be trained using a relatively small dataset, and the model does not require high-end computer for the analyses, making the model accessible to larger number of users. Secondary goal of this thesis includes automatic ROI placement. A simple ROI selection method, which uses the automatically obtained landmarks, was implement in the software. Tibial ROIs are selected below the approximate middle point of the delineated tibial contours. More refined approach to ROI placement is beyond the scope of this study.

The algorithm was implemented using python programming language, and graphical user interface was developed for the analyses (see figure 5). Furthermore, data visualization was implemented in the software. Radiographs can be viewed using the software, and the images can be automatically analyzed. The software can also visualize the delineation, JSW measurement and ROI placement results. The source code of the software is published on github (*https://github.com/jfrondel/KneeAnalyzer*).
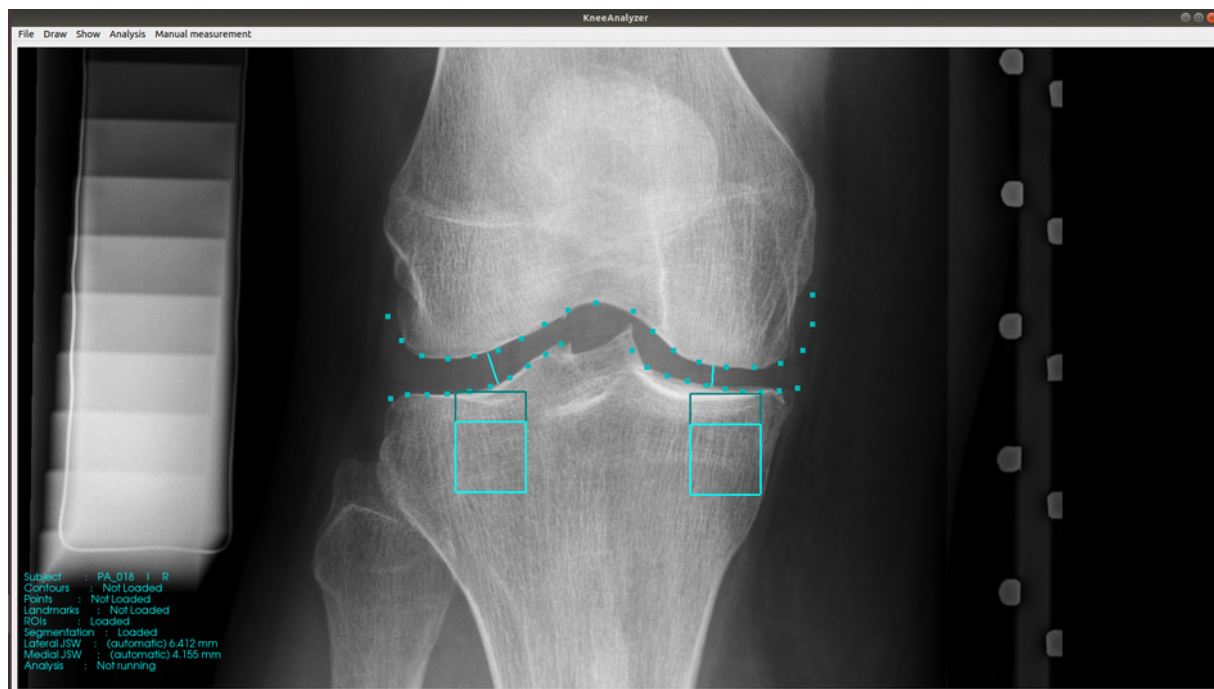
Figure 5. Picture of the graphical user interface for the analysis software. The software allows user to manually annotate knee radiographs and to train segmentation models. Multiple images can be later selected for sequential automated analyses, which will run in the background.

5. Dataset and Methods

Bilateral weight bearing posterior-anterior radiographs with knees in semi-flexion were acquired (DigitalDiagnost, Philips Medical Systems, 60 kVp, automatic exposure) from 107 subjects (66 female, 43 male) with and without OA. Spatial resolution of the radiographs was 0.148x0.148 mm^2. OA severity was assessed using KL grading scale. 4 knees were omitted from the analyses, due to TKR prostheses. Experienced researcher performed manual JSW measurements and tibial ROI selection from the images. Tibial and femoral contours were delineated for this study using custom python software. The images were then split to training and test set for the analyses. The knees were stratified based on their KL grading, and training dataset and test dataset were generated automatically.

Automated JSW analyses and delineation of the joint contours was performed with a custom made software written in python 3.5. Graphical user interface was made using Qt based PyQt5 library [65]. Visualization was handled using Visualization Toolkit (VTK) library for python [66]. VTK was also used to implement the manual delineation to the software. Basic image processing tasks were performed using OpenCV library [67]. Automatic joint localization algorithm and JSW measurement algorithms were implemented in python and integrated into the software.

5.1 Manual bone contour delineation and ASM construction

Manual delineation of the knee contours was performed using python software with visualization and drawing tools implemented using VTK library. Samples were delineated in 4 parts, as shown in figure 6. Lateral and medial tibia contours were drawn staring from the neck of the tibia, to the tibial spines. Starting point from the neck was selected based on the underside of tibio-femoral joint. Lateral and medial femur contours were drawn from the femoral neck to the midpoint of the joint. Starting location was selected based on the top of the patella.

Each of the four contours was split into 20 landmarks using k-means clustering [68]. The full knee model was created by concatenating the 20 point models into a single 80 point model. Separate 40 point tibial and femoral models were also created. Finally 20 points were selected from the tibial and femoral heads, and concatenated into a 40 point joint model. The four models were used to automatically delineate the joint space in 3 stages. Semi-automatically generated landmarks were used due to small amount of distinct landmarks present around the knee joint.

ASM was constructed for each of the models as described in section 2.3. A modification was made to the alignment process, where a constant size was enforced for the model during the alignment. Height and width of the model was recorded from the first mean shape by measuring the distance between the maximum and minimum x- and y-coordinates. The mean shape was scaled to same height and width during the alignment. This modification was necessary to prevent the model from shrinking during the alignment process. PCA was performed to find the unit eigenvectors and eigenvalues for the shape constraints. N eigenvectors were selected, as they covered > 97% of the variance.
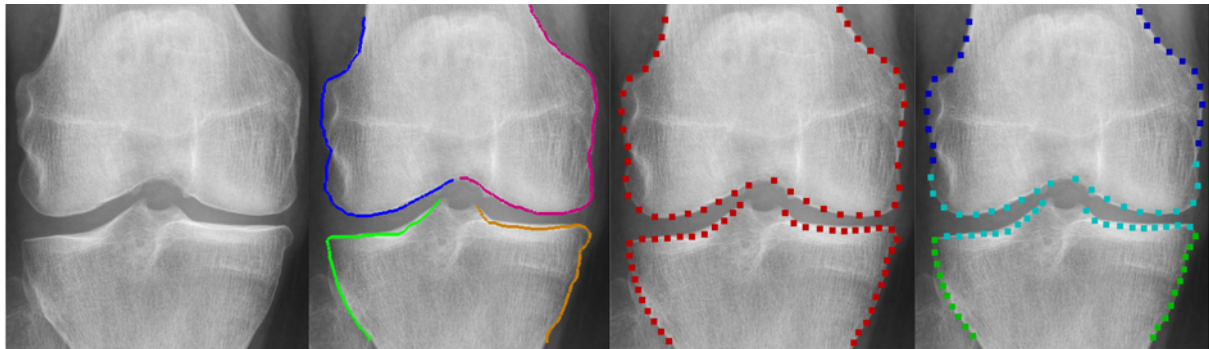

Figure 6. Knee annotations and bone ASM. The joint was annotated in 4 separate parts (lateral tibia (green line), medial tibia (orange line), lateral femur (blue line) and medial femur (purple line)). The annotation contours were clustered to 80 points (red points) and points from the training dataset were used to train the ASM. Separate tibial (green points), femoral (blue points) and JS (teal points) models were also constructed from the training dataset, which were used to annotate the test dataset in steps.

5.2 Automatic delineation and joint space measurement

At first, the joint center was localized by binarizing the test image, taking the cross-correlation of the binarized image and a black and white image of the knee ASM mean shape. DoG and Canny edge detection algorithm were used in the binarization. DoG was computed

using Gaussian filters with 5x5 kernel size and standard deviations 1 and 5. DoG image was then binarized by thresholding the pixels in the top 20th percentile to 1, and rest to 0. Morphological closing was then applied to the resulting image, using a 3x3 kernel. Binary objects with area less than 20 pixels were removed from the image. Second binary image was constructed using Canny edge detection followed by morphological dilation with a 3x3 kernel. The intersection of the two images was then used to localize the joint center using the cross-correlation of the binary image and the knee model image. Joint localization is described in figure 7.

The joint center location was then refined by cropping a 120x120 mm^2 area around the initial joint coordinate. The cropped image was thresholded, by setting the bottom 40th percentile of the pixels to 0 and rest to 1. A sliding window was used to find lateral and medial edges of the knee. Window size was 120x11.84 mm^2, and the edge of the joint was determined to be at the location where more than 50% of the pixels were 1. Center of the found knee edges was then used as x-coordinate for the joint center. Vertical location, or the y-coordinate, was measured by taking the sum 5.92 mm wide window around the x-coordinate along x-axis, and finding the strongest gradient of the resulting vector. The vector was down-sampled with a factor of 15, in order to reduce noise for the gradient computation.
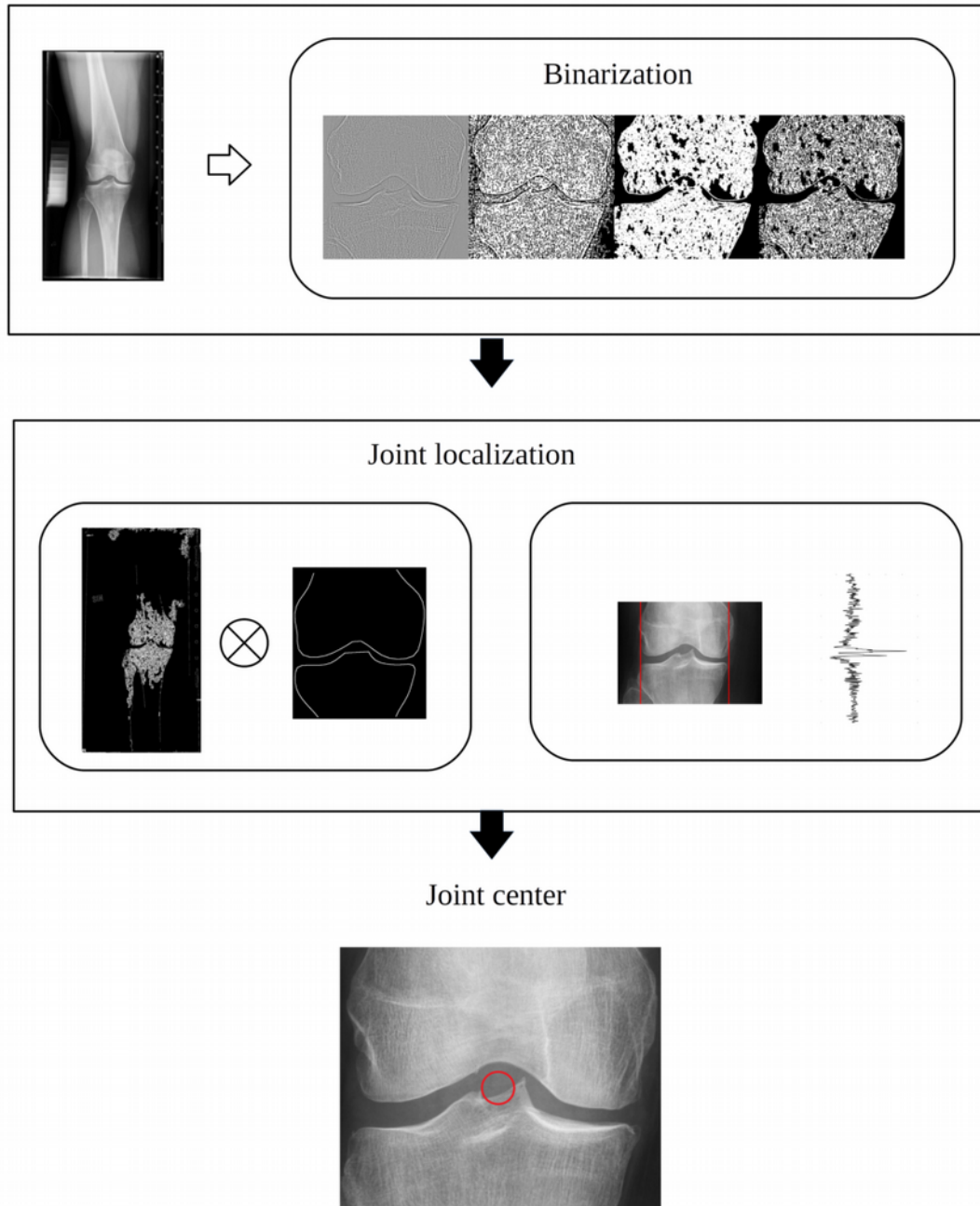
Figure 7. Joint localization pipeline. The image is first binarized by computing DoG image and using Canny edge detection. ASM mean shape is used to produce initial localization of the joint center, which is then refined using image processing and signal analysis methods. Analysis region is then selected around the joint center.

The mean shape of the ASM was centered around the refined joint center, and new landmarks were found as described in section 2.3. Once the full knee model converged, the landmarks were used as an initial location for tibial and femoral ASMs. Finally, the femoral and tibial head landmarks from the previous step were used as an initial location for the joint ASM.

23

Joint ASM landmark locations were refined as described in section 2.3. Joint delineation progression is shown in figure 8.
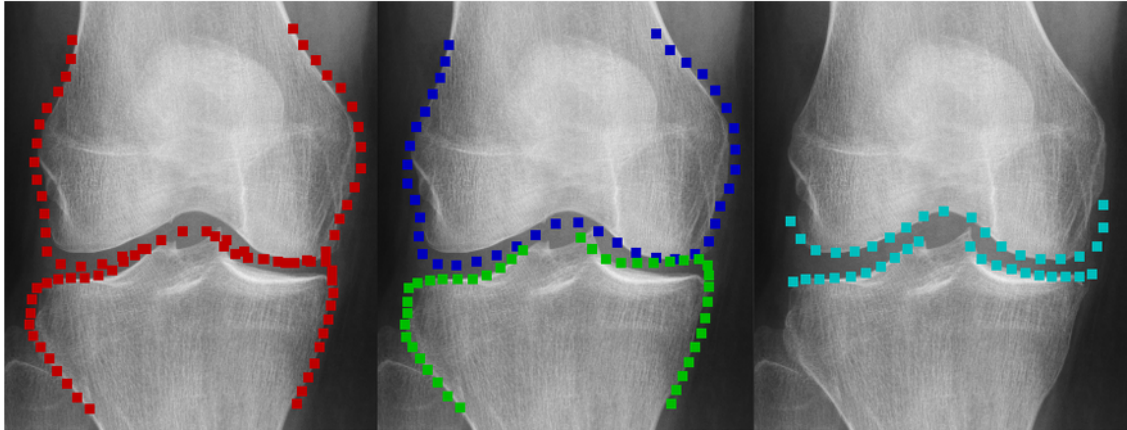

Figure 8. ASM delineation progression. Full knee model is aligned initially with the joint image. Tibia and femur contour locations are then refined by aligning separate models for the different bones. Finally, a joint model is aligned to the image, producing the final delineation result.

JSW was measured around the automatically obtained landmarks. JSW measurements are visualized in figure 9. The input image was first cropped around the segmentation. The cropped area was selected from the minimum and maximum coordinates of the joint landmarks, with 7.4 mm added to the top and bottom of the area. Height of the area was increased in order to ensure that the full joint space and the surrounding bone were included in the image used in the measurements, in case of delineation errors. The cropped image was then binarized, using the automatic thresholding method described in section 2.1. The image was split to three regions along its width, and each third was thresholded. Binary disks with decreasing radius were fitted to the resulting binary image, and the maximum radius of the disks were recorded at each pixel. Minimum, maximum, mean and median radii were obtained from the area under the delineated joint.

Second JSW measurement was performed by binarizing the image and finding the edges of the resulting binary mask. Edge pixels were then classified to lateral tibia, medial tibia, lateral femur and medial femur using the joint delineation. A contour corresponding to each anatomical region was selected from the joint landmarks, and edge pixels were classified using nearest neighbor search. Finally, the JSW was measured by finding nearest neighbors between the tibial and femoral pixels.

Subchondral bone ROIs were selected automatically using the joint delineation. The center of each ROI was placed below the 5th landmark of lateral and medial tibial contour. Smaller ROI (6.1×14.1 mm^2) was set to 1mm below the landmark, and larger ROI (14.1×14.1 mm^2) was set immediately below the smaller ROI.
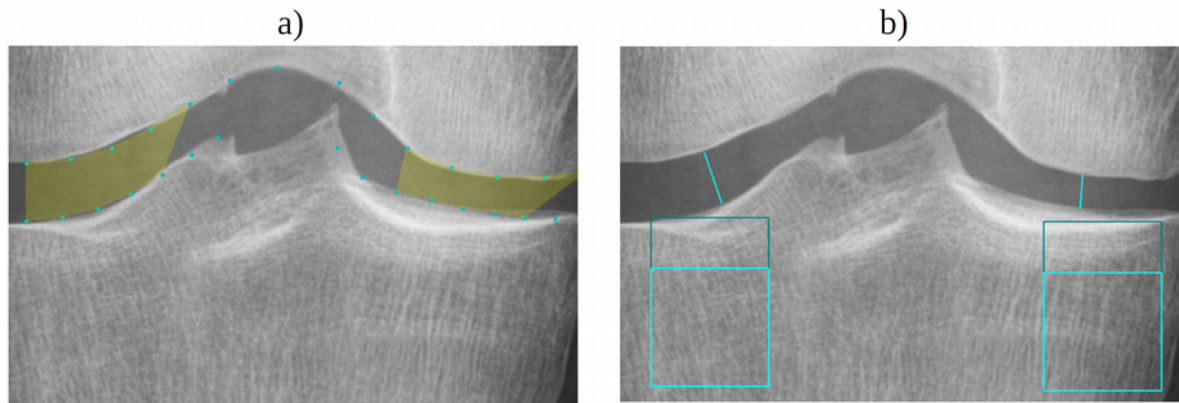

Figure 9. JSW measurements and ROI placements. First JSW measurement method selects lateral and medial regions where spherical binary masks are fitted, as shown in figure a). Largest radius of a masks that fits to the region is recorded at each pixel, and JSW metrics are extracted within the masks. Second method binarizes the joint and classifies the binary pixels to lateral and medial femur and tibia. Nearest neighbors between the contours are then used to measure the JSW, as shown in figure b). Figure b) also shows automatic tibial ROI placements.

5.3 Data analyses

Model performance was analyzed by measuring the Euclidean distance and Sørensen-Dice coefficient between manual and automatic annotations [69,70]. Automatic JSW measurement was compared to the manual measurement using Pearson correlation coefficient and root mean squared error. Automatic ROI placement was evaluated by measuring the Sørensen-Dice coefficient as well as the Euclidean distance between the center locations of the ROIs.

6. Experiments and results

Samples were split to test and training datasets before the experiments. Each dataset contained 50% of the samples (103 training samples, 103 test samples). Landmark search radius was optimized experimentally before the automatic JSW measurements. Search radius optimization was done by fitting the ASM to the images and varying the landmark search radius. Best performing parameters were then used in the JSW measurements.

JSW was measured after optimizing the landmark search radius using two different methods. At first, JSW was measured by fitting circles in a binarized joint space image cropped around the ASM landmarks. Minimum, maximum, mean and median radii of the circles were recorded under the are delineated by the ASM landmarks. The second JSW measurement was done by detecting the bone contours from the cropped joint space image. The contour pixels were then classified to tibia and femur using the ASM landmarks. JSW was then measured by finding nearest neighboring points between the tibia and femur contour pixels. Finally, subchondral bone ROIs were placed on the images based on the ASM delineation. Automatic ROI placements were evaluated against manual placements by using Soerensen-Dice coefficient and distance between the ROI centers.

6.1 Optimizing landmark search radius

The landmark search radius was optimized by delineating the knees using different radii for the landmark search. Experiments were performed at 10 (1.48 mm) pixel increments ranging from 10 pixels (1.48 mm) to 120 pixels (17.76 mm). Performance of the full knee model as well as tibial and femoral models at different search radii is shown in figure 10. Automatic alignment of the models were evaluated against the full ground truth (GT) model of the knee. Best automatic alignment of the full knee ASM was achieved with 60 pixels (8.88 mm) search radius for the test dataset and 70 pixel (10.36 mm) search radius for the training dataset. RMSEs were 6.05 mm and 5.23 mm respectively. The 60 pixel search radius was used in the rest of the experiments, as it yields slightly faster performance due to smaller amount of computations needed for the model alignment and performs better with the test dataset.

Tibial and femoral models were aligned to the radiographs using the automatically aligned full ASM of the knee as initial location and aligning separate models of the tibial and femoral contours to the image. The 60 pixel search radius was used in the full model alignment, and search radius experiments were conducted using the same 10 pixel increments for the tibial and femoral models as in the previous experiments. Alignment was evaluated against the full GT contour by concatenating the tibia and femur models into a single set of points. Best alignment was obtained using 10 pixel (1.48 mm) search radius for both the training and test datasets. RMSEs were 5.84 mm for the test dataset and 5.11 mm for the training dataset.
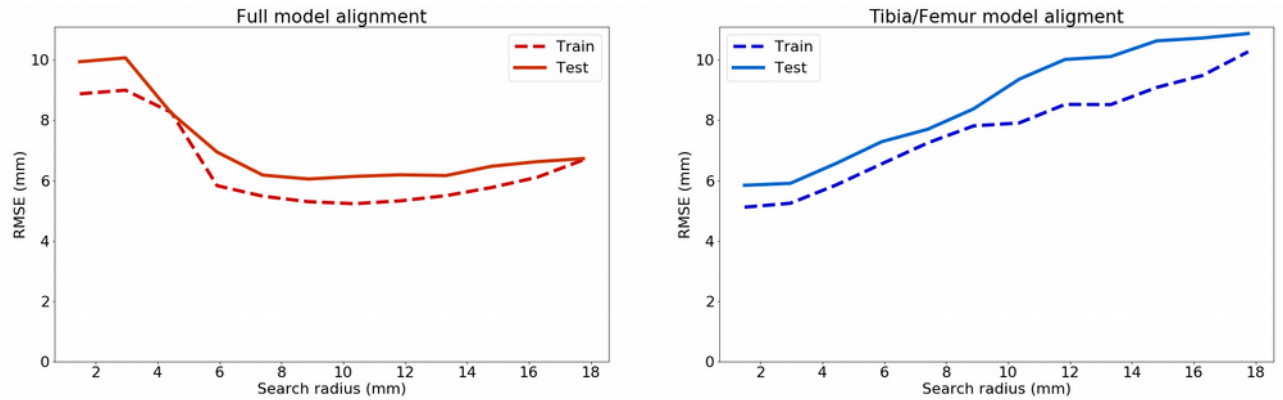

Figure 10. RMSE between aligned ASM and GT models against landmark search radius. Full model alignment performed best when the landmark search radius was 10.4 mm (70 pixels) for training dataset (RMSE 5.23 mm). Best performing search radius was 8.88 mm (60 pixels) for test dataset (RMSE 6.05 mm). Best performing landmark search radius was 1.48 mm (10 pixels) when separate tibia and femur models were fitted to the image. Same search radius achieved best performance for both test and training datasets (RMSEs 5.11 mm and 5.84 mm respectively).

Finally, the landmark search radius was optimized for the joint model. Joint ASM used joint space landmarks from the aligned tibial and femoral models as initial locations. Joint ASM was aligned to the radiograph after aligning the tibial and femoral models using search parameters obtained from the previous experiments. Joint RMSE was also evaluated for the full knee model as well as tibial and femoral models during the landmark search. Performance of the joint ASM as well as the other models around the joint space, is shown in figure 11. Best joint ASM alignment was obtained using search radius of 20 pixels (2.96 mm) for both training and test datasets. RMSEs were 3.78 mm and 4.63 mm. RMSEs are significantly lower than ones obtained from aligned full knee model (training dataset RMSE

27

4.65 mm, test dataset RMSE 5.46 mm) or tibial and femoral models (training dataset RMSE 4.33 mm, test dataset RMSE 5.09 mm).
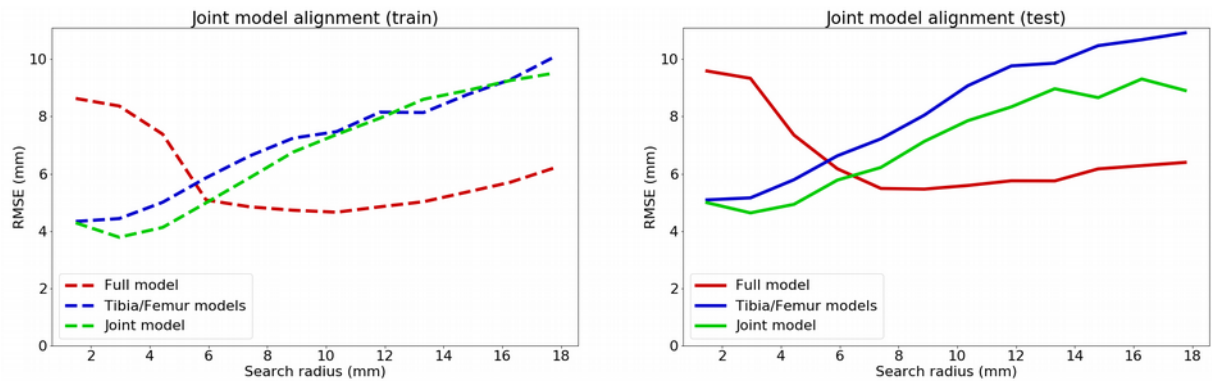

Figure 11. Joint ASM and GT model RMSE against landmark search radius. Best performing landmark search radius was 2.96 mm (20) for training (RMSE 3.78 mm) and test dataset (RMSE 4.63 mm) for the joint ASM. Model performance was evaluated from 40 landmarks around the joint space. Evaluation was done for full knee ASM as well as tibial and femoral models. Best performance was acquired when landmark search radius was 10.36 mm (70 pixels) for full knee ASM training dataset (RMSE 4.65 mm) and 8.88 mm (60 pixels) for test dataset (RMSE 5.46 mm). Best joint space landmark alignment from tibial and femoral models was obtained when the search radius was 1.48 mm (10 pixels) for both training (RMSE 4.33 mm) and test datasets (RMSE 5.09 mm).

6.2 Automatic JSW measurement from disks

Automatic JSW measurements using disk fitting were validated against manually measured JSWs. JSW was measured from a region located within the joint delineation. Measurements were done by fitting disks to a binarized image, and recording the maximum radius of the disk at each pixel in the region. Minimum, maximum, mean and median radii were then obtained from the pixels within the automatically selected area. Pearson correlation coefficient and RMSE were measured between the automatically obtained values and manually measured ground-truth JSW. RMSE, Pearson correlation coefficient and corresponding p-values between automatically obtained JSWs and GT JSWs for the training dataset are reported in table 1. Visual representation of the results for training dataset is provided in figure 12 and for the test dataset in figure 13. Automatically measured mean JSW had the smallest RMSE (1.39 mm on lateral side, 1.57 mm on medial side) and best correlation (Pearson R 0.68 on lateral side and 0.61 on medial side, p-values < 0.05) with the GT JSW on the training dataset. Mean JSW also had smallest RMSE (1.69 mm on lateral side, 1.46 mm on medial side) with the test dataset measurements. Highest correlation

28

(Pearson R 0.58) was obtained from automatic mean JSW on lateral side measurements from the test dataset. On medial side, highest correlation was obtained from automatic median JSW (Pearson R 0.59) although the RMSE was slightly higher (1.60 mm). Results for the test dataset are shown in table 2. The visual representation of the results is shown in figure 13.

Table 1. Automatic JSW measurements vs. GT JSW measurements for training dataset.

Measurements	Lateral			Medial		
	RMSE (mm)	Pearson	P-value	RMSE (mm)	Pearson	P-value
Minimum	4.56	0.08	0.41	3.33	0.13	0.20
Maximum	2.00	0.68	<0.05	2.97	0.44	<0.05
Mean	1.39	0.64	<0.05	1.57	0.61	<0.05
Median	1.65	0.57	<0.05	1.71	0.59	<0.05

Table 2. Automatic JSW measurements vs. GT JSW measurements for test dataset.

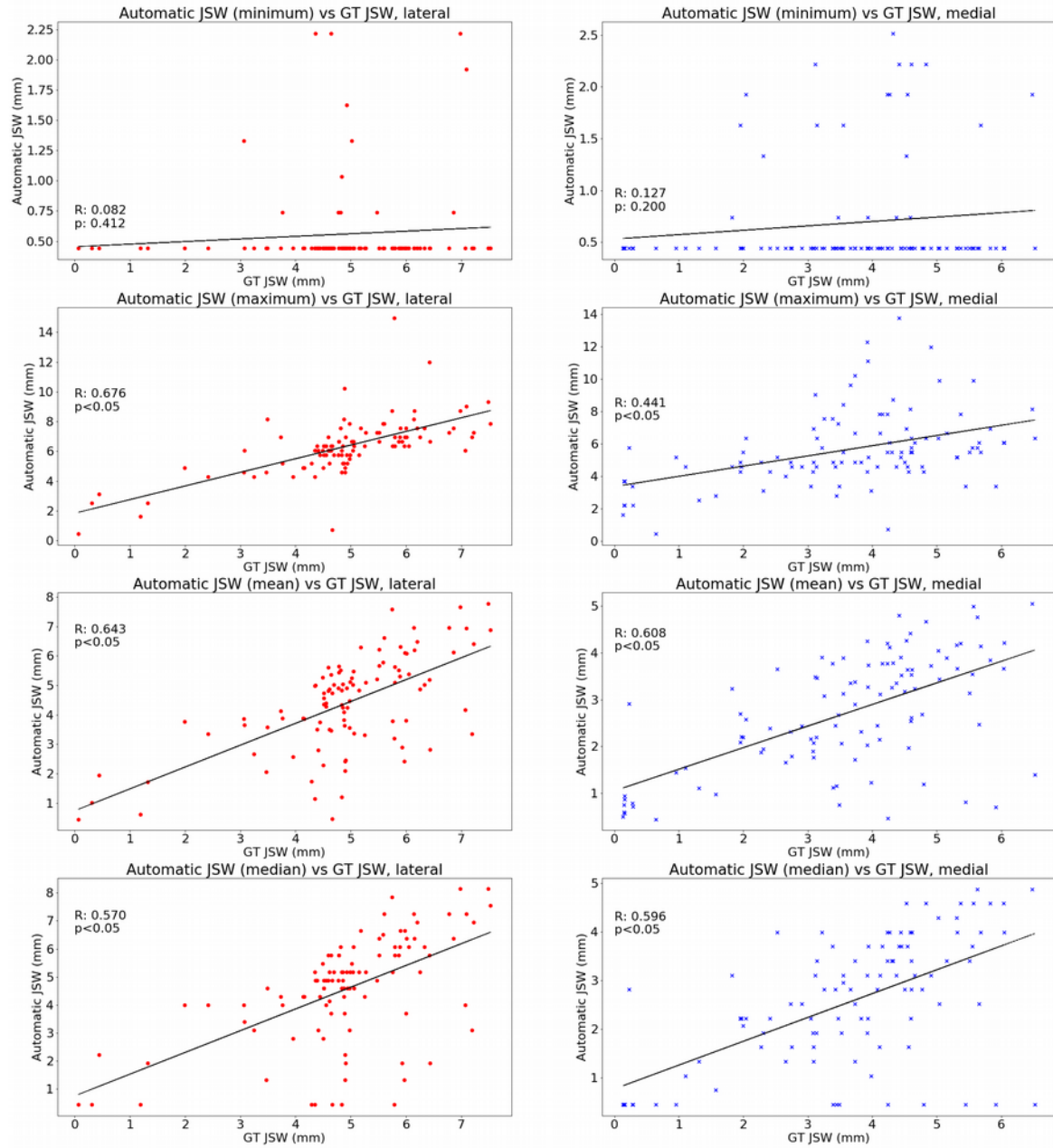| Measurements	Lateral			Medial		
	RMSE (mm)	Pearson	P-value	RMSE (mm)	Pearson	P-value
Minimum	4.67	0.12	0.20	3.35	0.12	0.21
Maximum	2.35	0.41	<0.05	3.17	0.32	<0.05
Mean	1.69	0.52	<0.05	1.46	0.57	<0.05
Median	1.97	0.46	<0.05	1.60	0.59	<0.05

Figure 12. Correlations between automatic JSW and GT JSW for training dataset. Mean and median JSWs have the strongest correlation with the GT JSW on lateral and medial side respectively. On the medial side, mean and median JSWs have comparable correlations to the manually measured JSW.
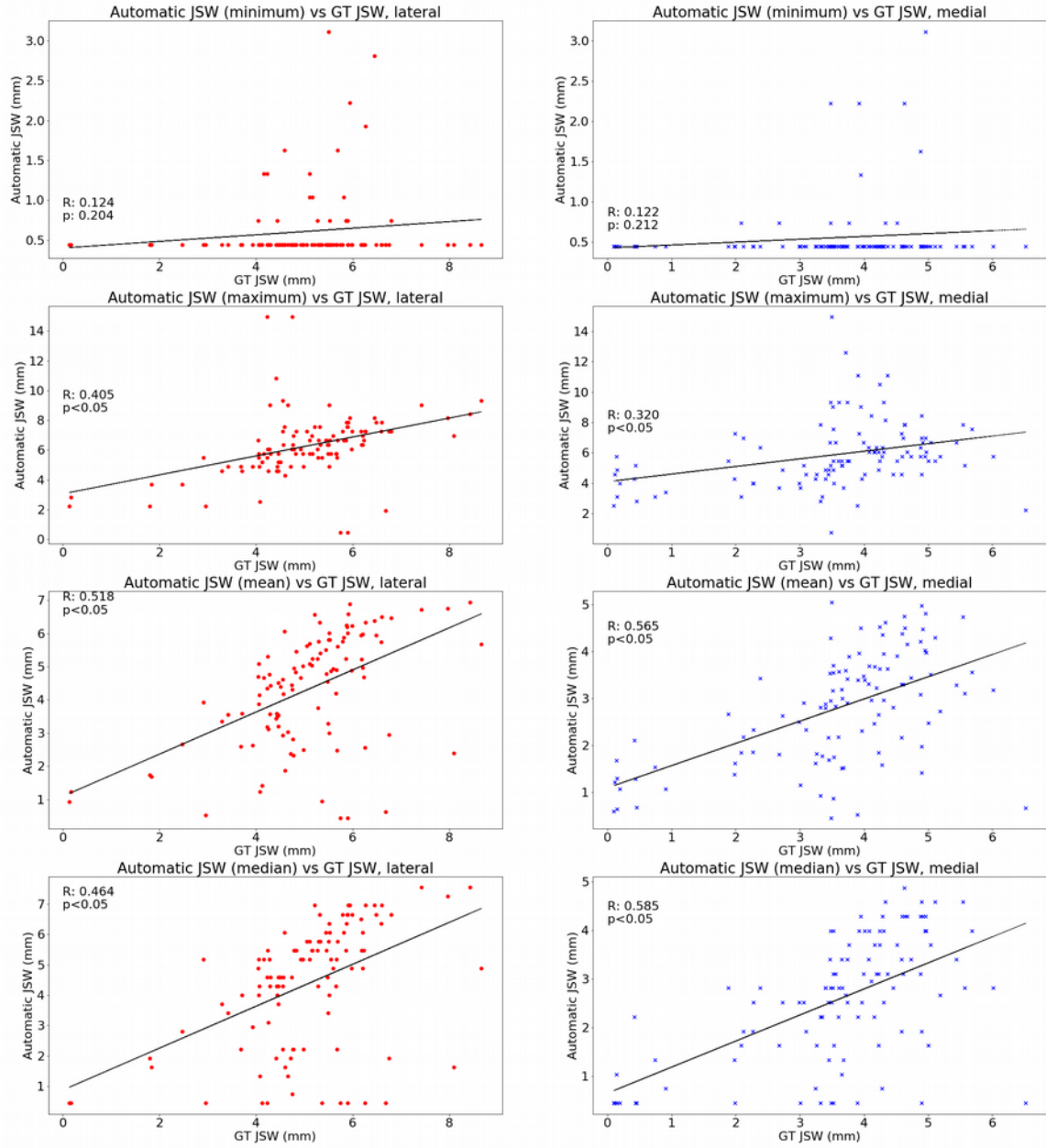
Figure 13. Correlations between automatic JSW and GT JSW for test dataset. Mean and median JSWs have the strongest correlation with the GT JSW on lateral and medial side respectively. On the medial side, mean and median JSWs have comparable correlations to the manually measured JSW.

31

6.3 Automatic JSW measurement using nearest neighbors

JSW was also measured from the binarized joint images by classifying the binarized pixels to femur and tibia. Classification was obtained by finding nearest neighbors from the automatically delineated joint space landmarks for the binary image edge pixels at the joint space. Binarization was done the same way as in the previous measurements, followed by edge detection. Edge detection was done by computing gradient image from the binarize image, and setting bottom 80 percentile of the pixels to 0. After the classification, JSW was measured by finding nearest neighbors between femur and tibia edges. Separate measurements were performed on lateral and medial sides. Results for training dataset are shown in figure 14, and results for test dataset are shown in figure 15.
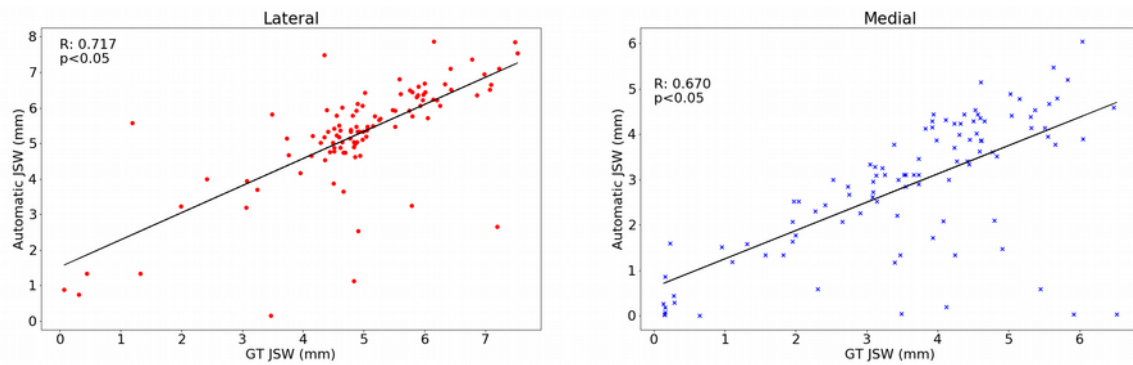

Figure 14. Automatic JSW measurement from edge pixels for training dataset. Pearson correlation coefficient on lateral side was 0.717 and 0.670 on medial side. P-values were <0.05.
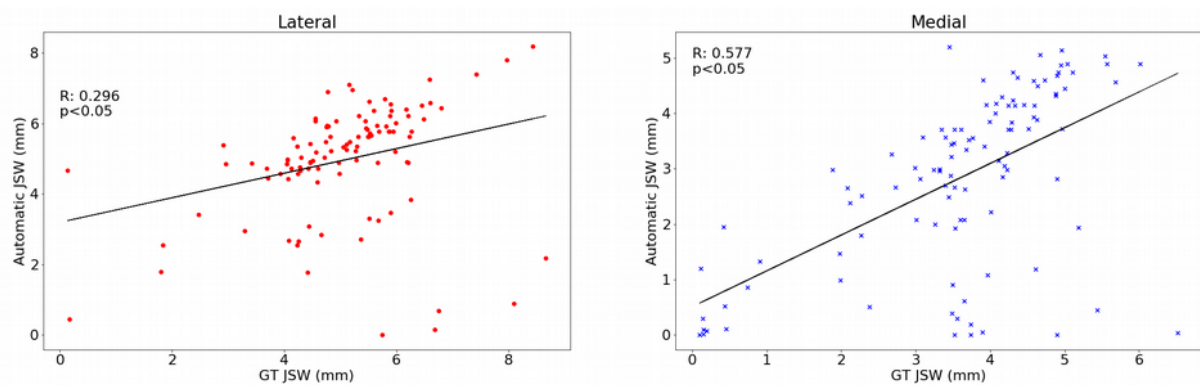

Figure 15. Automatic JSW measurement from edge pixels for test dataset. Pearson correlation coefficient on lateral side was 0.296 and 0.577 on medial side. P-values were <0.05.

6.4 Automatic ROI placements

Tibial ROIs were automatically selected based on the delineation of the joint space. Smaller subchondral bone ROIs were placed 1.48 mm (10 pixels) below the middle landmarks. The landmark placements were validated using Soerensen-Dice coefficient, and by measuring the distance between automatically and manually placed ROIs. Center distance measurements were used since the images were rotated for the manual ROI placements, which has an impact on the Dice coefficient. Mean scores for the training and test datasets are presented in table 3. Smaller tibial ROIs achieved higher Dice coefficients across the experiments (0.716, 0.700, 0.699 and 0.678 for lateral train, lateral test, medial train and medial test dataset respectively) than the larger Subchondral bone ROIs (0.590, 0.580, 0.559 and 0.520 for lateral train, lateral test, medial train and medial test datasets). ROI center placement differed less between tibial and subchondral bone ROIs. Mean distance between automatically and manually placed ROIs were 4.16 mm and 4.29 mm for tibial and subchondral ROIs for the train dataset on lateral side. On medial side distances, were 4.59 mm and 4.61 mm between automatically and manually placed tibial and subchondral ROIs. With the test dataset, corresponding distances were 4.28 mm, 4.38 mm, 5.32 mm and 5.06 mm.

Table 3. Automatic ROI placement metrics.

	Dice (train)	Center distance (mm) (train)	Dice (test)	Center distance (mm) (test)
Subchondral, lateral	0.590	4.16	0.580	4.28
Tibial, lateral	0.716	4.29	0.700	4.38
Subchondral, medial	0.559	4.59	0.520	5.32
Tibial, medial	0.699	4.61	0.678	5.06

7. Analysis software

Automatic JSW measurements and ROI placement methods were implemented in custom made open source python software. Software usage is demonstrated in figure 15. Source code is published on github (*https://github.com/jfrondel/KneeAnalyzer*). Python language was used to implement the software, due to its open source nature and the high availability of external libraries. Open source libraries were used implement the data visualization and the user interface. Many of the necessary mathematical operations were also imported from the external libraries.

 The data was read from DICOM format using PyDicom library [71]. Image pre-processing was performed using OpenCV and NumPy libraries, which were also used to implement the analysis scripts, along with Scikit-Learn library [67,72,73]. The data and results were visualized using VTK library [66]. Graphical user-interface was developed using PyQT5 library [65]. Results were saved in csv format for the sake of accessibility using pandas library [74]. The software includes ASM annotation tools which allow the user to draw continuous bone contours manually. The software then selects landmarks from the annotations using k-means clustering. Tools for manual ROI placements and JSW measurements are included. ASM can be trained once the data is annotated. The software uses 50% of the annotated data to train the model. Stratified split is generated automatically from a table containing the KL grades for the dataset. The user must specify a path to a table containing the grades. Automatic delineations and measurements can be generated either for a single knee or multiple knees from all of the selected radiographs with a single button press. The data is analyzed in a separate thread running while the software is running, allowing user to visualize the results for the already analyzed samples. The status of the analyses is provided to the user in the graphical user interface (see figure 16.).
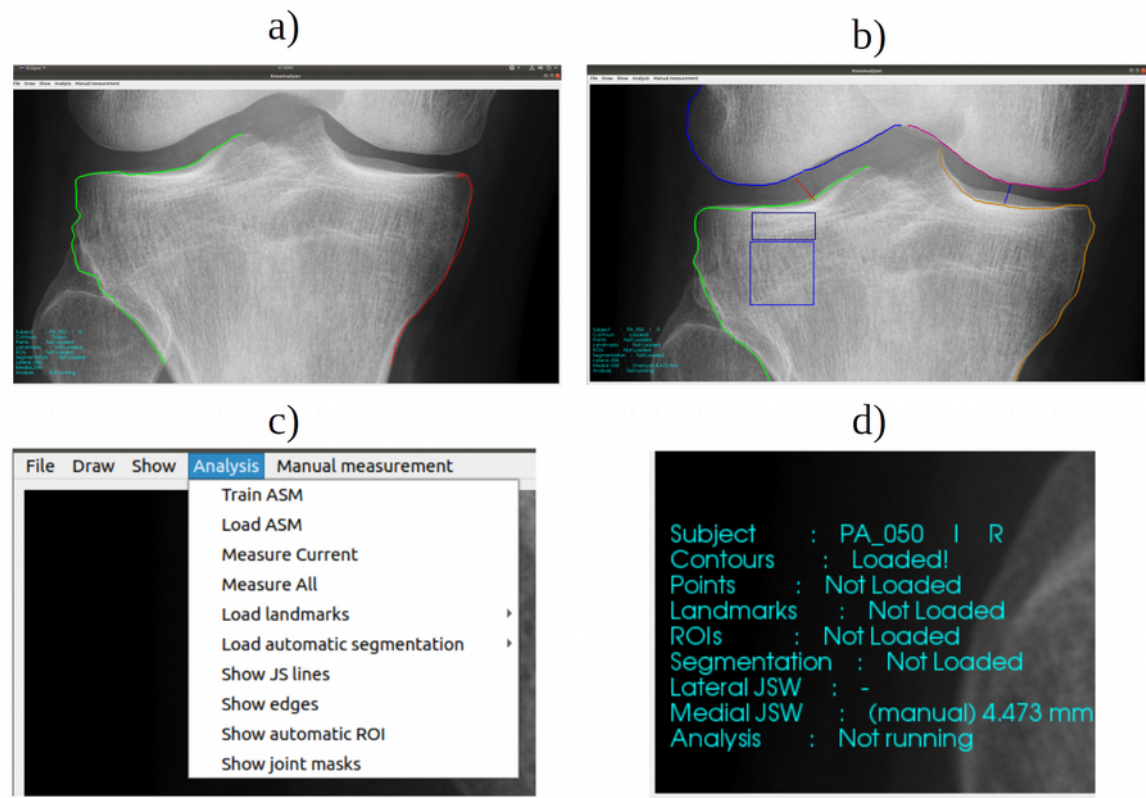
Figure 16. Knee analysis software usage. The software has tools for manually annotation of the bone contours (figure a)) as well as manual measurement of JSW (figure b)). Furthermore, the tibial ROIs can be placed manually on the bone. The software has minimalistic menu based user interface as shown in figure c). Useful information about the analyses are shown to the user (figure d)) at the bottom left corner. The software displays sample name, leg which is currently viewed, and the JSW measurement results including whether they were obtained automatically or manually.

8. Discussion

Detection of OA induced changes in the joint is an important part of OA diagnosis. In the future, automatic methods for image analyses may be required in clinical and research setting, due to the disease prevalence and increasingly larger amounts of available data. In this study, fully automatic methods for joint localization and joint space measurements were implemented using image processing and computer vision methods. Computational cost was kept low by using more traditional methods, allowing the algorithms to be run on low-end workstation, and thus making the models available to larger number of users. Furthermore, the algorithms can be easily scaled for multiple processors.

Two different methods for joint space measurement were implemented and compared after the joint delineation. The joint delineation was performed using ASM with a simple gradient-based feature for selecting new landmark locations for the model. Automatic joint space measurements were done using pixel classification and fitting circular masks to a binarized joint space image. The results correlate well with the manually obtained GT joint space widths. Joint space measurements were generally more accurate when the measurements were performed by fitting disks to the binarized image. Furthermore, a simple method for automatic tibial ROI placement was implemented. The full joint analysis pipeline was implemented in a python software, and the source code was made publicly available. Model performance will be discussed in greater detail in following sections.

8.1 ASM Performance

Accurate delineation of JSW was obtained using multi-stage ASM. The landmark search radius was optimized before the actual measurements. Optimization was done by measuring the error between the points obtained from manual annotations and the automatic delineation. Optimization improved the results significantly. The model could be further optimized by tuning other parameters, such as the number of landmarks and the number of principal components used in the ASM adjustments. However, these optimization steps were beyond the scope of this study. The number of landmarks was selected to be 80 for the full knee.

Selection was based on initial experiments made during the development, since it provided good accuracy with low computational cost. The number of landmarks could have been set to ~100, while still maintaining the stability of the model, since the training dataset included similar amount of samples. Changing the number of principal components changes the allowed deformations in the model alignment, which impacts the performance of the model. The number of principal components for each model was selected automatically based on the criteria that the corresponding eigenvalues cover >97% of the total variance. The 97% threshold was obtained from the literature [64]. A lower threshold has been used in other studies using ASM to delineate the knee contours [15]. Future work should also include stratification by KL grade when aligning the models. OA induced changes in the varus-valgus alignment of the joint can impact the model performance, especially when using a model of the full joint. The allowed deformations of the model are constrained by the variance in the dataset. If the training data does not include enough deformed joints, the performance of the model may suffer.

The full knee model was relatively rigid and could not capture the shape of the joint space as well as the smaller models. However, it provided a good initial for the joint landmark. The results were improved using separate models for tibia and femur. Separating the two contours was helpful for cases, where the bones were incorrectly aligned due to the disease progression. The greatest improvement to the model accuracy was obtained using the smaller model comprised of joint space landmarks. The joint space model was able to better follow the contours of the joint, as the landmarks outside the joint space did not impact the alignment. The multi-stage model also allowed using very simplistic model for the landmark appearance. A simple gradient magnitude was used to determine a new landmark during the alignment, which requires a good starting location for the model. Good initial locations for the joint landmarks were obtained using accurate joint localization and multiple bone delineations with different models. Example results are given in figure 17.
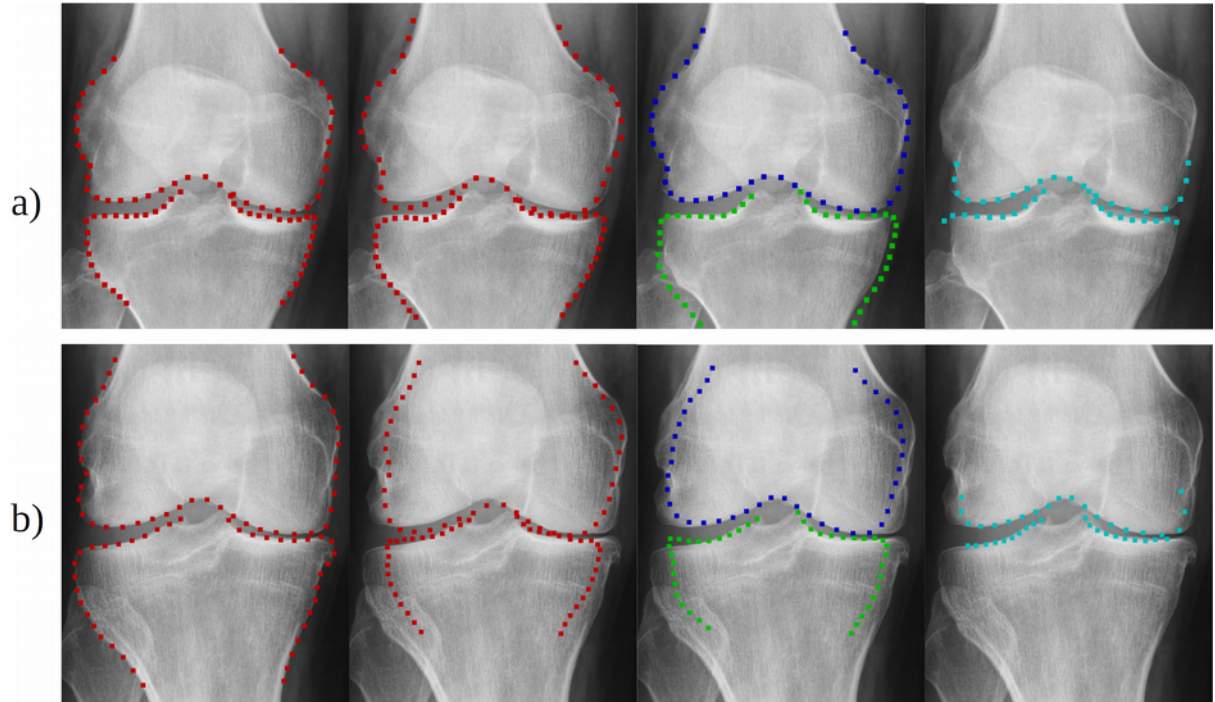
Figure 17. Example ASM delineation results. Figure a) shows a well aligned model. Ground truth landmarks are shown in the left-most image. Second image from the left shows aligned full knee model. The full model delineates the contours of the bones well, but the joint landmark alignment requires improvement. The third image from the left shows tibial and femoral model alignment. Now the models are able to follow the bone contours better at the joint region. The medial tibia contour is slightly mis-aligned. Finally, the right-most image show joint model alignment. Tibial landmarks are now correctly aligned in both lateral and medial regions. Figure b) show poorly aligned model. Manual annotation, full model, tibial and femoral models and the joint model are in the same order as in the figure a). The model was now aligned to the small structures in the bone, yielding a poor overall alignment. However, this is compensated by the joint model alignment, which follow the contours of the tibia and femur, making measurements still feasible from the delineation.

Drawback of selecting new landmarks from strongest image gradient locations is the susceptibility of the model to noise and poor contrast. If the bones do not have good contrast against the surrounding tissue, aligning the model to the correct location could be difficult. Furthermore, strong edges in the bone can cause the model to converge to a wrong location. Using multiple model alleviated these issues, since the joint space was typically clearly visible from the images in the dataset. Even if the initial alignment of the larger models was sub-optimal, the joint space model was able to converge to the appropriate location.

The automatic delineation results are comparable to landmark detection method implemented by Chen et al in [11]. Chen et al achieved higher accuracy by combining a sparse shape model with their automatic landmark detection algorithm. The ASM based joint delineation

method utilized here is significantly less complex. Combining multiple different statistical models for the joint space delineation was beyond the scope of this study. The delineation accuracy could be further improved by using more advanced features in landmark selection and aligning a new model after the initial landmark localization using the gradient feature.

8.2 JSW measurements

Two separate methods were tested for the JSW measurements. The first method utilized disk-shaped binary kernels to find the distance between the bones and the second method searched nearest neighbors between the bone contours. Overall performance of the disk method was better than the nearest neighbor method, although the nearest neighbor method achieved better performance with the training dataset. All of the JSW metrics collected under the joint segmentation regions had significant correlation with the ground truth JSW measurement, excluding minimum automatic JSW. Minimum JSW measured from the JS mask typically consisted of one pixel, thus lacking any useful information. The maximum JSW under the mask typically over estimated the distance between tibia and femur, most likely due to noisy binarization of the joint. Automatically obtained mean and median JSWs correlated with the manual measurements quite well, mean JSW having better overall correlation with the manual measurements.

Nearest neighbor JSW measurement had good correlation with the manual measurements when the measurements were done with the training dataset. However, the performance on the lateral side of the joint was quite poor when the measurements were done with the training dataset. The issue arises from the relatively low tissue contrast of the lateral femoral condyle, which results in poor binarization of the joint tissues. When the JSW is measured using the disk method, sufficient amount of the joint is properly binarized, leading to more accurate results when compared to the nearest neighbor method. The nearest neighbor method performance could be improved by more accurate binarization of the joint space. Fine tuning the binarization could improve the results significantly. Example results are shown in figure 18 with ROI placements.
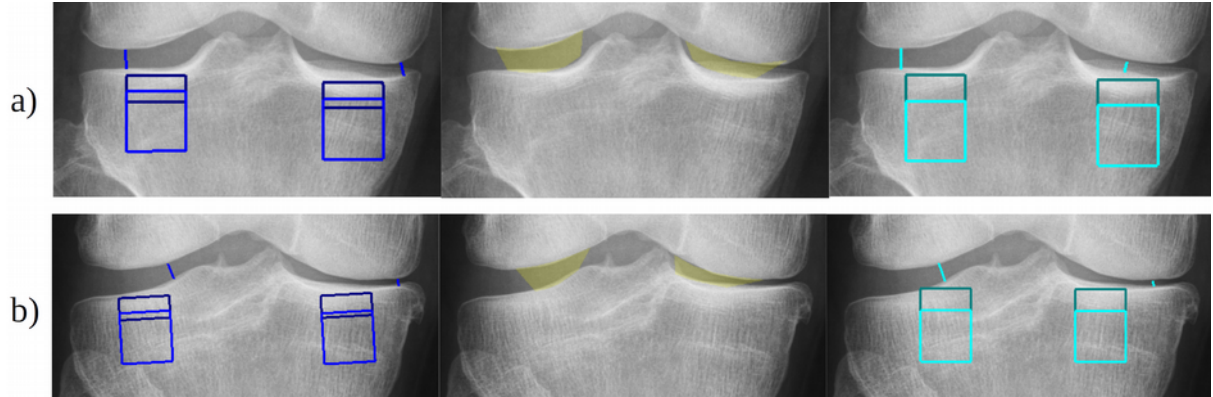
Figure 18. Automatic JSW measurements. Figure a) shows results from well aligned model. Left-most image shows manual JSW measurement and the manually placed tibial ROIs. Middle image shows joint masks obtained from the binarized joint image with the ASM delineation. JSW was measured by fitting disk shaped kernels to the area denoted by the yellow mask. Right-most image show JSW measurements from binary contours obtained from the joint image. The binary pixels were classified using the ASM delineation. Lateral JSW was measured to be 4.516 mm (manual), 4.740 mm (mean result from the mask) and 4.884 mm (nearest neighbor pixels from the binary contours). Corresponding medial JSW values are 3.083 mm, 1.898 mm and 2.621 mm. Figure b) shows results from the poorly aligned model (see figure 16.). Despite the lacking accuracy of the ASM alignment, the JSW measurements are consistent with the manual measurement. Lateral JSW was measured to be 4.515 mm (manual), 4.559 mm and 5.420 mm. Corresponding JSWs on the medial side were 1.953 mm. 2.701 mm and 1.635 mm.

8.3 Automatic ROI placement

The ROI placements were done automatically based on the delineation results. The centers of the ROIs were close to the GT placements, while the Dice coefficients were quite low. The Dice coefficients suffered due to the misalignment of the ROIs, as the GT ROIs were placed on rotated images. The ROI placement could be further improved by using landmarks from the tibial ASM. Using multiple landmarks to select the ROI location, for example by voting for the center location, could yield significantly higher accuracy. However, this was beyond the scope of this study. Example ROI placements are shown in figure 18 and closer examination of the ROI overlap is shown in figure 19. The automatic ROI selection could be improved by utilizing the method used by Podsiadlo et al in [15]. Since the automatic ROI placement was a secondary goal of this study, implementation of the more advanced method was omitted.
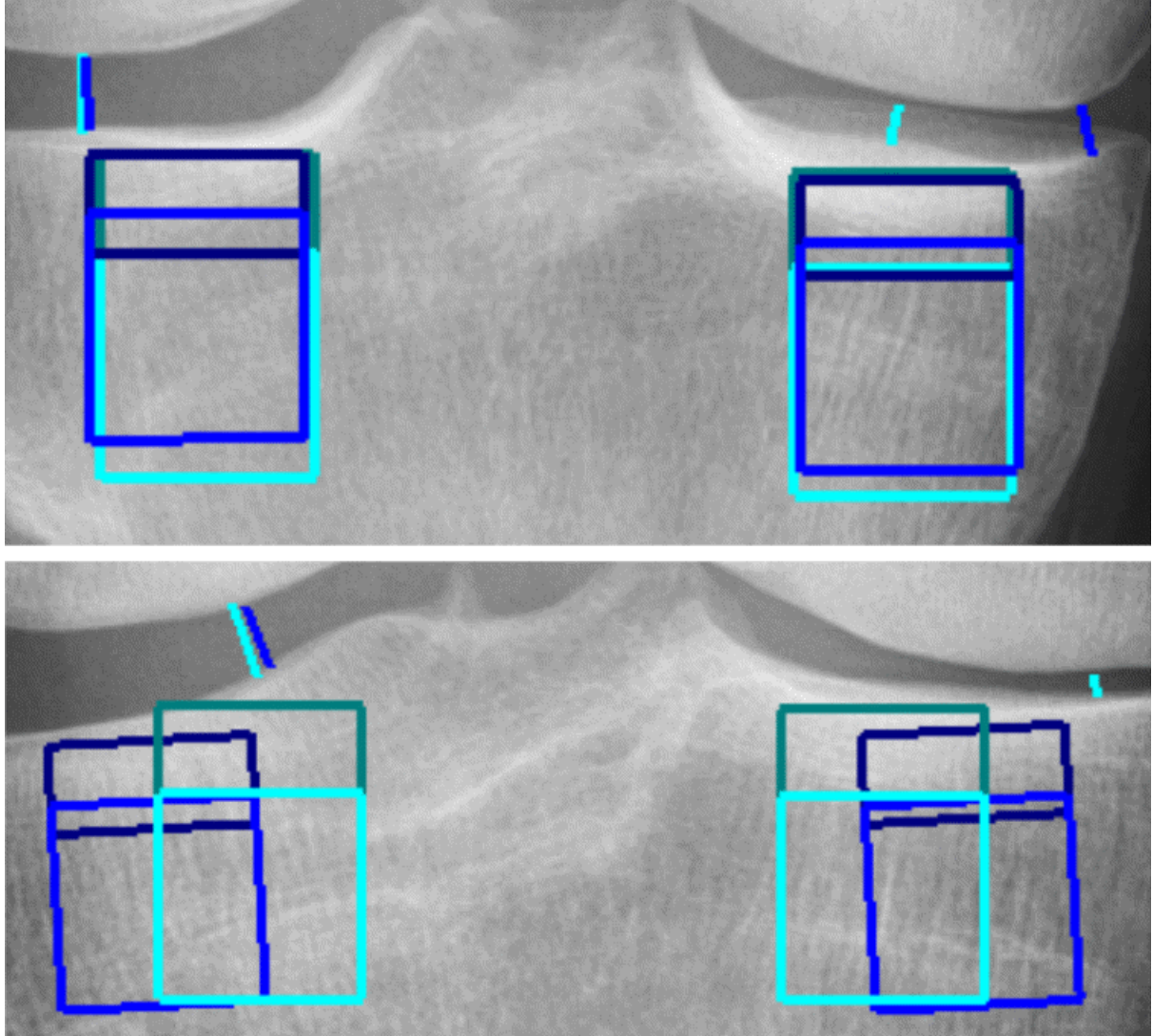
Figure 19. Comparison between the manually (blue) and automatically placed ROIs (cyan). The images were rotated when the during the manual ROI placements, which impacts the similarity metrics. Good agreement between the manual and automatic ROI placements was still achieved when the ASM alignment was good (top image). If the automatic delineation performed poorly, the ROI placement suffered as well (bottom image). The automatic method places the larger subchondral ROI immediately below the smaller ROI, which also impacts the performance, since it does not take the possible overlap into the account.

8.4 Limitations of the study

The biggest limitation of this study was the use of single dataset obtained using the same x-ray machine and imaging parameters. Evaluating the models using different dataset would

make it possible to test how well the models generalize to images obtained using different imaging parameters. The joint localization and joint space delineation methods will likely require the use of different model parameters, since the contrast between the different tissues and background may change depending on x-ray machine and imaging parameters used in the data acquisition. However, the model parameters can be easily adjusted for new datasets and controls for these can be added to the software in the future. Furthermore, combining more advanced models with the delineation algorithm may improve the results further. Constructing a more advanced landmark feature descriptors could provide more accurate results and faster converge for the ASM, although this may increase the computational cost. The automatic ROI placement should be improved, for example by utilizing the same methods used by Podsiadlo et al in [15]. Another option would be to use multiple landmarks from the ASM to select the ROI localization and orientation.

Further improvements to the performance could be obtained by using CNNs to segment tibia and femur. CNNs can achieve human level performance in many image analysis tasks, although they require large amounts of training data. Training neural networks also requires significant amount of computational power not required by the methods utilized in this study. Despite this, the CNNs may be more attractive alternative to traditional image processing and computer vision methods in the future for automatic analyses as availability of data increases and computer become faster.

Traditional image processing and computer vision methods can be used in semi-automatic annotation of images. Manually obtained GT datasets typically suffer from intra- and inter-observer variability, which can be alleviated using local image features. Features, such as image gradients, can be obtained from a small region around a manually selected location and used to refine the manual annotation. This would reduce the variability, as the final annotation would be constructed using automatic and repeatable methods. Automated analyses can also be used to obtain initial approximation of the desired results, which can later be manually refined. Future improvements to the radiograph software developed here should include refinement of the manual annotations using image gradients, as well as the option to manually tune the obtained results. These features were beyond the scope of this study. The source code for the algorithms implemented here will be freely available for other researchers on github (*https://github.com/jfrondel/KneeAnalyzer*).

9. Conclusions

Decrease in the joint space width is an indicator of knee OA and accurate diagnostic tools are important in early detection of the disease. A fully automated ASM based method for joint space delineation and measurement from plain radiographs was developed and implemented in this study. The automatically obtained results agree well with manual joint space measurements. Furthermore, a method for tibial ROI selection was included in the algorithms. Algorithms developed here were implemented in open source software with robust visualization tools for the results. The low computational cost and easily explainable model features make the automatic tools used here accessible to larger research community.

10. References

1. Hirvasniemi J (2015) "Novel X-Ray-Based Methods for Diagnostics of Osteoarthritis." University Of Oulu, Institute of Clinical Medicine and Department of Applied Physics.

2. Arden N, Blanco F, Cooper C, Guermazi A, Hayashi D, Hunter D, Javaid MK, Rannou F, Roemer F & Reginster J (2014) "Atlas of osteoarthritis", Springer.

3. Heliövaara M, Slätis P & Paavolainen P (2008) "Nivelrikon esiintyvyys ja kustannukset."

4. Oksendahl H, Gomez N, Thomas C, Badger G, Hulstyn M, Fadale P & Fleming B (2009) "Digital radiographic assessment of tibiofemoral joint space width–a variance component analysis." The journal of knee surgery 22(03): 205-212.

5. Zhang W, Ouyang H, Dass CR & Xu J (2016) "Current research on pharmacologic and regenerative therapies for osteoarthritis." Bone research 4(1): 1-14.

6. Goker B, Haznedaroglu S, Tufan A & Block J (2016) "Observer Variability of Joint Space with Measurements Is Subject Related in Medial Knee Osteoarthritis." Arthritis & Rheumatology.

7. Iglesias J & Sabuncu M (2015) "Multi-atlas segmentation of biomedical images: a survey." Medical Image Analysis 24(1): 205-219.

8. Weese J & Lorenz C (2016) "Four challenges in medical image analysis from an industrial perspective." Medical Image Analysis 33(8): 44-49

9. Jiang H, He B, Fang D, Ma Z, Yang B & Zhang L (2013) "A region growing vessel segmentation algorithm based on spectrum information." Computational and Mathematical Methods in Medicine 2013.

10. Liu H, Chen Z, Chen X & Chen Y (2006) "Multiresolution medical image segmentation based on wavelet transform." 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. , IEEE: 3418-3421.

11. Chen C, Xie W, Franke J, Grutzner P, Nolte L & Zheng G (2014) "Automatic X-ray landmark detection and shape segmentation via data-driven joint estimation of image displacements." Medical Image Analysis 18(3): 487-499.

12. Chen C, Wang C, Huang C, Li C & Zheng G (2014) "Fully-Automatic Landmark detection in Skull X-Ray images." submitted to Automatic Cephalometric X-ray Landmark Detection Challenge .

13. Ambellan F, Tack A, Ehlke M & Zachow S (2019) "Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative." Medical Image Analysis 52: 109-118.

14. Zhou Z, Zhao G, Kijowski R & Liu F (2018) "Deep convolutional neural network for segmentation of knee joint anatomy." Magnetic resonance in medicine 80(6): 2759-2770.

15. Podsiadlo P, Wolski M & Stachowiak G (2008) "Automated selection of trabecular bone regions in knee radiographs." Medical Physics 35(5): 1870-1883.

16. Woloszynski T, Podsiadlo P, Stachowiak G, Kurzynski M, Lohmander L & Englund M (2012) "Prediction of progression of radiographic knee osteoarthritis using tibial trabecular bone texture." Arthritis & Rheumatism 64(3): 688-695.

17. Hirvasniemi J, Thevenot J, Immonen V, Liikavainio T, Pulkkinen P, Jämsä T, Arokoski J & Saarakkala S (2014) "Quantification of differences in bone texture from plain radiographs in knees with and without osteoarthritis." Osteoarthritis and cartilage 22(10): 1724-1731.

18. Ronneberger O, Fischer P & Brox T (2015) "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention, Springer: 234-241.

19. Lundervold AS & Lundervold A (2019) "An overview of deep learning in medical imaging focusing on MRI." Zeitschrift für Medizinische Physik 29(2): 102-127.

20. Justus D, Brennan J, Bonner S & McGough AS (2018) "Predicting the computational cost of deep learning models." 2018 IEEE International Conference on Big Data (Big Data). , IEEE: 3873-3882.

21. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D & Batra D (2017) "Grad-cam: Visual explanations from deep networks via gradient-based localization." Proceedings of the IEEE international conference on computer vision. : 618-626.

22. Chen D, Shen J, Zhao W, Wang T, Han L, Hamilton J & Im H (2017) "Osteoarthritis: toward a comprehensive understanding of pathological mechanism." Bone research 5(1): 1-13.

23. Hunter D, McDougall J & Keefe F (2008) "The symptoms of osteoarthritis and the genesis of pain." Rheumatic Disease Clinics of North America 34(3): 623-643.

24. Vincent K, Conrad B, Fregly B & Vincent H (2012) "The pathophysiology of osteoarthritis: a mechanical perspective on the knee joint." PM&R 4(5): S3-S9.

25. Braun H & Gold G (2012) "Diagnosis of osteoarthritis: imaging." Bone 51(2): 278-288.

26. Moisio K, Chang A, Eckstein F, Chmiel JS, Wirth W, Almagor O, Prasad P, Cahue S, Kothari A & Sharma L (2011) "Varus–valgus alignment: reduced risk of subsequent cartilage loss in the less loaded compartment." Arthritis & Rheumatism 63(4): 1002-1009.

27. Pritzker K, Gay S, Jimenez S, Ostergaard K, Pelletier J, Revell P, Salter D & Van den Berg W (2006) "Osteoarthritis cartilage histopathology: grading and staging." Osteoarthritis and cartilage 14(1): 13-29.

28. Kim HT, Kim HJ, Ahn H & Hong Y (2016) "An analysis of age-related loss of skeletal muscle mass and its significance on osteoarthritis in a Korean population." Korean Journal of Internal Medicine 31(3): 585-593.

29. Bliddal H, Leeds A & Christensen R (2014) "Osteoarthritis, obesity and weight loss: evidence, hypotheses and horizons–a scoping review." Obesity reviews 15(7): 578-586.

30. Thomas A, Hubbard-Turner T, Wikstrom E & Palmieri-Smith R (2017) "Epidemiology of posttraumatic osteoarthritis." Journal of athletic training 52(6): 491-496.

31. Fox S, Bedi A & Rodeo S (2009) "The basic science of articular cartilage: structure, composition, and function." Sports health 1(6): 461-468.

32. Peters A, Akhtar R, Comerford E & Bates K (2018) "The effect of ageing and osteoarthritis on the mechanical properties of cartilage and bone in the human knee joint." Scientific reports 8(1): 1-13.

33. Li G, Yin J, Gao J, Cheng TS, Pavlos N, Zhang C & Zheng M (2013) "Subchondral bone in osteoarthritis: insight into risk factors and microstructural changes." Arthritis research & therapy 15(6): 223.

34. Kazemi M, Dabiri Y & Li L (2013) "Recent advances in computational mechanics of the human knee joint." Computational and Mathematical Methods in Medicine 2013.

35. Flandry F & Hommel G (2011) "Normal anatomy and biomechanics of the knee." Sports medicine and arthroscopy review 19(2): 82-92.

36. Kim Y & Joo Y (2012) "Patellofemoral osteoarthritis." Knee surgery & related research 24(4): 193-200.

37. Kumar D, Manal K & Rudolph K (2013) "Knee joint loading during gait in healthy controls and individuals with knee osteoarthritis." Osteoarthritis and cartilage 21(2): 298-305.

38. Mikla V (2014) "1 - Advances in Imaging from the First X-Ray Images" Medical Imaging Technology, Elsevier: 1-22

39. Chen H, Rogalski M & Anker J (2012) "Advances in functional X-ray imaging techniques and contrast agents." Physical Chemistry Chemical Physics 14(39): 13469-13486.

40. Mikla V (2014) "5 - X-Ray Detectors" Medical Imaging Technology, Elsevier: 1-22

41. Seibert J (2004) "X-ray imaging physics for nuclear medicine technologists. Part 1: Basic principles of x-ray production." Journal of nuclear medicine technology 32(3): 139-147.

42. Kellgren J & Lawrence J (1957) "Radiological assessment of osteo-arthrosis." Annals of the rheumatic diseases 16(4): 494-502.

43. Petersson I, Boegard T, Saxne T, Silman A & Svensson B (1997) "Radiographic osteoarthritis of the knee classified by the Ahlback and Kellgren & Lawrence systems for the tibiofemoral joint in people aged 35-54 years with chronic knee pain." Annals of the rheumatic diseases 56(8): 493-496.

44. Szeliski R (2010) "Computer vision: algorithms and applications." Springer Science & Business Media.

45. Jähne B & Haußecker H (2000) "Computer vision and applications. A Guide for Students and Practitioners." Elsevier

46. Tarantola A (2005) "Inverse problem theory and methods for model parameter estimation." siam.

47. Borji A (2018) "Negative results in computer vision: A perspective." Image and Vision Computing 69: 1-8.

48. Dalal N & Triggs B (2005) "Histograms of oriented gradients for human detection." 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05): 886-893.

49. Pietikäinen M, Hadid A, Zhao G & Ahonen T (2011) "Computer vision using local binary patterns." Springer Science & Business Media.

50. LeCun Y, Bengio Y & Hinton G (2015) "Deep learning." Nature 521(7553): 436-444.

51. Cootes T, Baldock E & Graham J (2000) "An introduction to active shape models." Image processing and analysis : 223-248.

52. Risholm P, Golby AJ & Wells W,3rd (2011) "Multimodal image registration for preoperative planning and image-guided neurosurgical procedures." Neurosurgery clinics of North America 22(2): 197-206, viii.

53. Stachowiak G, Wolski M, Woloszynski T & Podsiadlo P (2016) "Detection and prediction of osteoarthritis in knee and hand joints based on the X-ray image analysis." Biosurface and Biotribology 2(4): 162-172.

54. Gonzalez R, Woods R, & Masters B (2008). "Digital image processing third edition." Pearson International Edition.

55. Yang F, Deng Z & Fan Q (2013) "A method for fast automated microscope image stitching." Micron 48: 17-25.

56. Canny J (1986) "A computational approach to edge detection." IEEE Transactions on pattern analysis and machine intelligence (6): 679-698.

57. Otsu N (1979) "A threshold selection method from gray-level histograms." IEEE transactions on systems, man, and cybernetics, 9(1): 62-66.

58. Lowe D (2004) "Distinctive image features from scale-invariant keypoints." International journal of computer vision 60(2): 91-110.

59. Bay H, Tuytelaars T & Van Gool L (2006) "Surf: Speeded up robust features." European conference on computer vision. Springer: 404-417.

60. Rister B, Horowitz M & Rubin D (2017) "Volumetric image registration from invariant keypoints." IEEE Trans Image Process 26(10): 4900-4910.

61. Krizhevsky A, Sutskever I & Hinton G (2012) "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems: 1097-1105.

62. Ruder S (2016) "An overview of gradient descent optimization algorithms." arXiv preprint arXiv:1609.04747 .

63. Tan C, Sun F, Kong T, Zhang W, Yang C & Liu C (2018) "A survey on deep transfer learning." International conference on artificial neural networks. Springer: 270-279.

64. Cootes T, Taylor C, Cooper D & Graham J (1995) "Active shape models-their training and application." Computer vision and image understanding 61(1): 38-59.

65. P. Thompson (2016), "PyQt5", https://pypi.org/project/PyQt5/, [Online; accessed 10.10.2019]

66. Schroeder W, Lorensen B, & Martin K (2004) "The visualization toolkit: an object-oriented approach to 3D graphics." Kitware.

67. Bradski G, & Kaehler A (2008) "Learning OpenCV: Computer vision with the OpenCV library." O'Reilly Media, Inc.

68. Lloyd S (1982) "Least squares quantization in PCM." IEEE transactions on information theory 28.2 (1982): 129-137.

69. Sørensen T (1948) "A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons."

70. Dice L (1945) "Measures of the amount of ecologic association between species." Ecology 26(3): 297-302.

71. Mason D (2011) "SU-E-T-33: pydicom: an open source DICOM library." Medical Physics, 38(6): 3493-3493.

72. Oliphant T (2006) "A guide to NumPy (Vol. 1, p. 85)." Trelgol Publishing.

73. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O & Vanderplas J (2011) "Scikit-learn: Machine learning in Python." Journal of machine learning research, 12(10)