Finger joint modelling from hand X-ray images for assessing rheumatoid arthritis*

Sergio Vera sergio.vera.h@gmail.com

Supervisor: Debora Gil

Abstract

Rheumatoid arthritis is an autoimmune, systemic, inflammatory disorder that mainly affects bone joints. While there is no cure for this disease, continuous advances on palliative treatments require frequent verification of patient's illness evolution. Such evolution is measured through several available semi-quantitative methods that require evaluation of hand and foot X-ray images. Accurate assessment is a time consuming task that requires highly trained personnel. This hinders a generalized use in clinical practice for early diagnose and disease follow-up. In the context of the automatization of such evaluation methods we present a method for detection and characterization of finger joints in hand radiography images. Several measures for assessing the reduction of joint space width are proposed. We compare for the first time such measures to the Van der Heijde score, the gold standard method for rheumatoid arthritis assessment. The proposed method outperforms existing strategies with a detection rate above 95%. Our comparison to Van der Heijde index shows a promising correlation that encourages further research.

Keywords: Rheumatoid arthritis, joint detection, X-ray, Van der Heijde score

∗. Parts of this work have been published in the recent C.A.R.S. congress 2010 in Geneva

1. Introduction

Depending on the degree of mobility, bone joints split in three different classes: fibrous, cartilaginous and synovial joints. Fibrous joints are formed by connective tissue, usually collagen, and provide limited or no movement at all. This type of joint can be found, for instance, on skull bones or within the tooth and the sockets in the mandible. Cartilaginous joints have cartilage as connecting element between bones and, as a consequence, they allow a higher degree of movement. An example of this joint is the one found where the ribs meet the sternum, where the cartilage allows the chest to accommodate the changes of volume due to breathing, or the discs between the dorsal spine bones. Finally, synovial joints are the most common joints, since they are located whenever articulating bones (such as the knee, the elbow or fingers) meet each other. They allow the highest mobility between bones, thanks to the presence of a membrane (the synovial membrane or *synovium*) that encapsulates the articulation filling the cavity inside the membrane with a lubricating liquid called synovial fluid. The scheme in Fig. 1(a) shows the structure of a normal synovial joint.

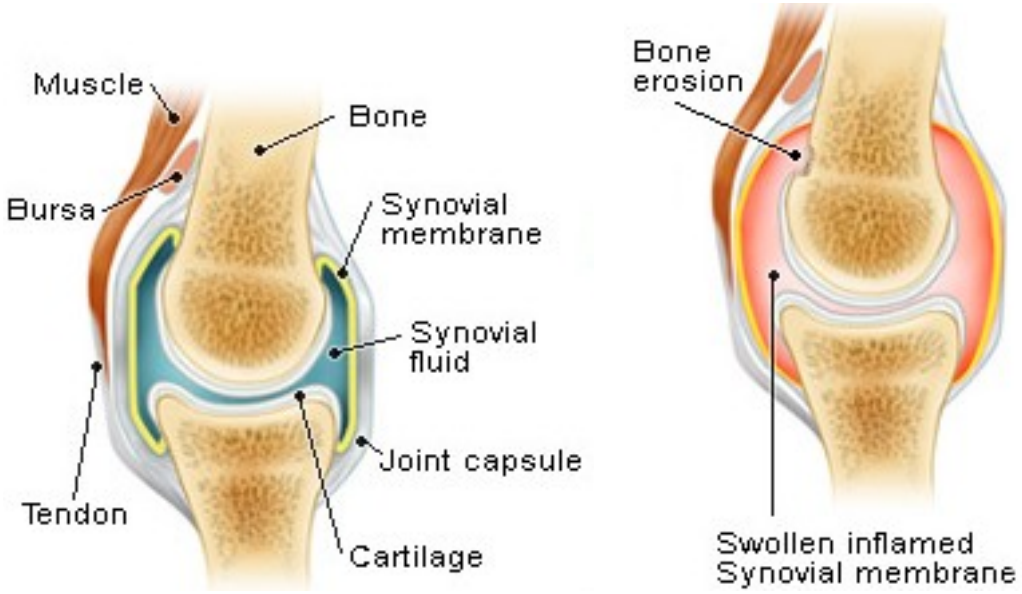


Figure 1: Synovial joint structure: (a) healthy joint and (b) joint affected by rheumatoid arthritis.

Rheumatoid arthritis (RA) is a chronic disease of unknown origin (although immune system seems to play a significant role in it and, thus, it is classified as autoimmune) that mainly affects synovial joints. At its first stages, RA causes inflammation of the synovial membrane, swelling and warming the whole joint which produces pain and reduced mobility on the affected area. In a second stage of the disease, the synovium cells start to grow and divide, thickening the membrane and releasing enzymes. Such enzymes dissolve the bone-covering cartilage (which results in a reduction in joint space) and might erode the bone

itself, (see scheme in Figure 1(b)). At severe stages of RA, the cartilage collapses and bones might mount one over the other (bone luxation) which makes the joint completely loose its function. Figure 2 shows examples of unaffected, medium and heavily affected hands.

RA mostly affects small joints such as hand finger joints (see Figure 3 for a reference on the bones and joints affected). Since it can provoke irreversible joint damage even at first stages, there is a significant decrease in the quality of life of patients affected by RA. This has motivated in the last years an active research in developing pharmacological treatments for stopping the degenerative process and preventing reaching advanced stages. Treatments against RA require medical trials for demonstrating their effectiveness, and, thus accurate methodologies for assessing the degree of RA are needed too. Although Echography and Magnetic Resonance allow a complete view of the 3D structure of the articulation (including cartilage, synovium...), the *de facto* gold standard for RA assessment is X-Ray. Despite cartilage is not visible, and the projection misses 3D information, X-Ray is preferred by its easy interpretation ,as well as availability in medical centers.

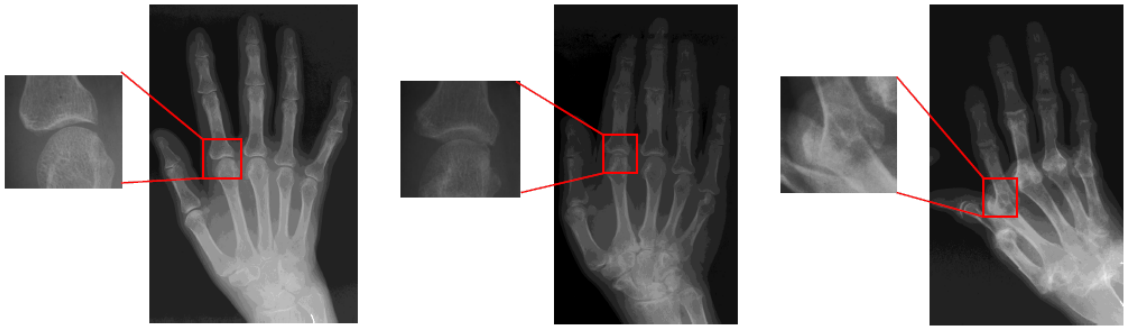


Figure 2: Several degrees of RA: Unaffected hand (left). Medium affected hand (middle). Severely affected hands (right).

Assessment of RA by X-Ray analysis is based on exploring joints in hands and feet. For each joint, two different aspects are analyzed: the separation between the bones that form the articulation and the presence of erosions. The assessment on each joint is added in a final numeric semi-quantitative result that represents the degree of affection. More information about evaluation methods of rheumatoid arthritis can be found on Jensen and Klarlund (2004) and Guerrero and Villaseñor (2006).

From all the existing methods, the most widely used is the Sharp/Van Der Heijde (VdH) method (van der Heijde (1999)). The method is semi-quantitative, so that the final score, although represents the status of the patient, is not calculated from *physical* measurements of erosion and joint space width reduction, but in relative scores based on a global examination of the joint. Van der Heijde scores range from 0 (healthy) to 5 (severely affected) in erosions and from 0 to 4 in joint space narrowing. Figure 4 illustrates the

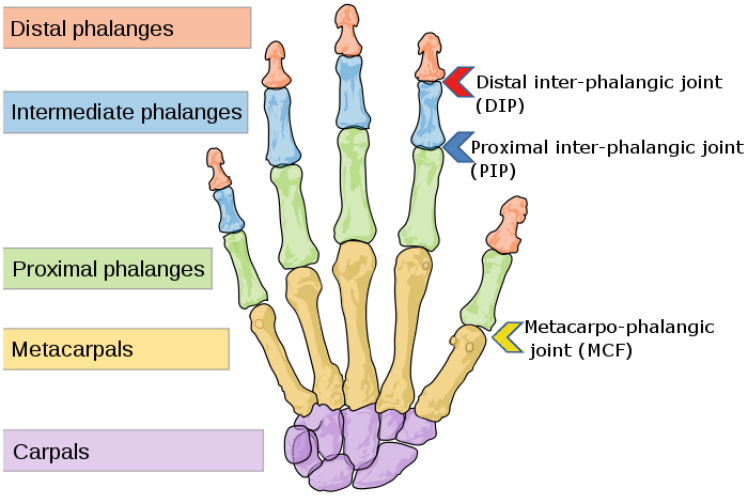


Figure 3: Hand bone structure and finger joint type.

scoring for erosions (top row images) and joint space narrowing (bottom row images). For both, erosions and joint space width, the value of zero always indicates a healthy joint. For erosions, a score of one means that there are focused small erosions. A score of two is assigned to joints presenting big erosions that affect less than half of the bone. Three is the score given to joints with big erosions, covering more than half of the bone. Four is assigned to joints with isolated erosions scattered all around the bone surfaces. Finally, five is associated to generalized erosions with complete destruction of the bone. Concerning joint space reduction, a score of one represents a small reduction in width. Two indicates a greater reduction, representing covering less than 50% of the width of a normal joint. Three means reduction of more than 50% compared to a healthy joint or with luxation. Finally, a score of four represents a total collapse of the joint space (with or without luxation). In order to calculate the final score, the individual score of the hand joints (30 in both hands) and feet (12 joints) is added. Not all the joints are evaluated. The metacarpo-phalangeal (MCP) joints, and the proximal inter-phalangic (PIP) joints are scored, while the distal inter-phanagic (DIP) joints are not analyzed (see figure 3).

The evaluation of a patient is time consuming because of the number of joints to be considered, and the complexity to determine the score at joints moderately affected. Indeed, it is not uncommon to spend more than forty minutes to give the final score of a patient. Such a lengthy process is not useful on the daily clinical practice. In fact, in this scenario simplified versions of the Van der Heijde score are used in assessment and diagnosis. In addition, inter observer variability is high, except between experts who have been trained together for long periods of time (months). Besides being time consuming, the Van der Heijde grading can not detect small variations of the disease. The scores assigned to the joints are assigned based on relative criteria with little granularity. This is undesirable in scenarios like clinical trials, which require a higher sensitivity for evaluating the progress of the illness. Besides, modern ethics in drug development and testing forbid prescripting

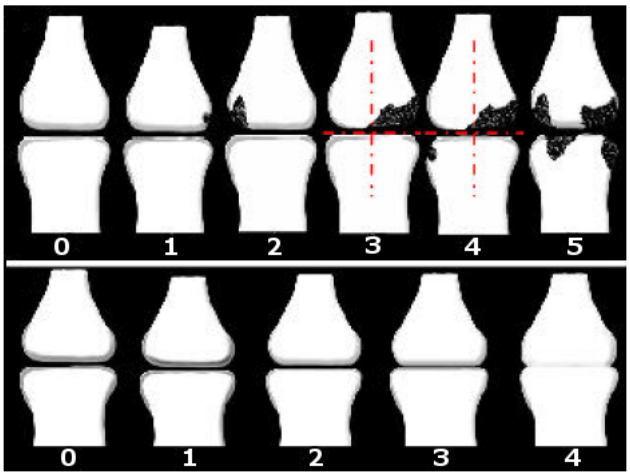


Figure 4: Van der Heijde score grading for erosions (top row) and joint space narrowing (bottom row).

placebo substances to RA patients as a single year might severely degrade the state of the patient. In a study of a year, a RA patient with a placebo substance can lose an important degree of movement when there are substances that can reduce this degradation. Also, the time devoted to such research has been reduced over time, so it is important to be able to have finer measurements that can detect the evolution of the illness over a shorter period. Therefore, the availability of automated methods to perform accurate and objective measurements would greatly benefit not only the diary diagnosis of RA but also the pharmacological and clinical research environments.

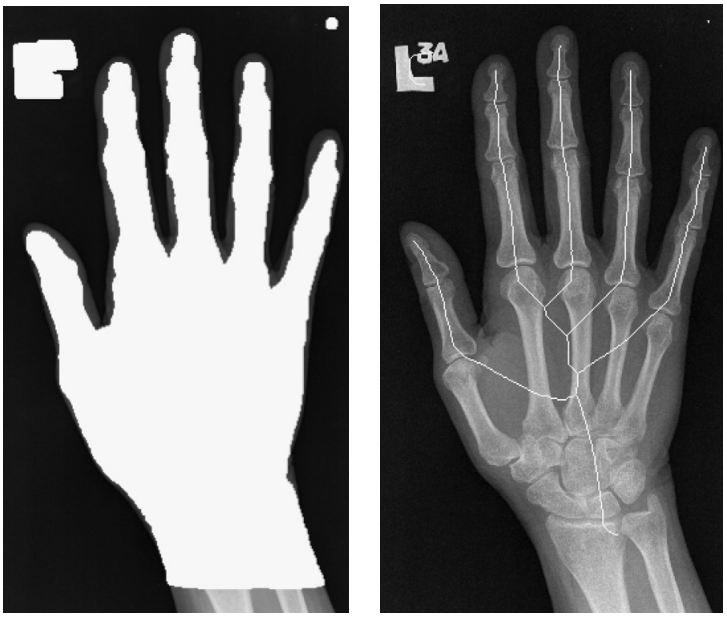
1.1 State of the art

Although there is a significant amount of work on X-Ray image processing, to our knowledge, the study of RA assessment has mainly been carried out at the medical side (Sharp et al. (2000), Jensen and Klarlund (2004), Angwin et al. (2001)) among others. Existing works focus on addressing the feasibility of using computer programs to measure radiographic joint space width and the degree of erosion, rather than automatic computation of RA scores. A reference work is Sharp et al. (2000) since one of the authors contributed to the development of the gold standard Van der Heijde score. In his work, Sharp and his collaborators use a combination of semi-automatic methodologies with the help of medical imaging software (US National Institutes of Health's NIH Image software). Joints are manually cropped and the resulting images are manually re-oriented to set the joint space horizontal in the image. After pre-processing, edges of the upper and the lower bone are located, either automatically using edge detectors or manually, fitting a fourth order polynomial to control points distributed along edges and manually selected. Several distance measures are considered: minimum and average distances between edges, in addition to the distance function along the lower bone. The correlation and slope of the regression for the

later provide information about asymmetry of the joint space. The authors perform a series of tests to compare the reproducibility of automatic measurements against manual ones on a trial of placebo and gold treatments. Although experiments show the discriminant power of the computerized measurements, no attempts to correlate them to Sharp/Van der Heijde scores are made.

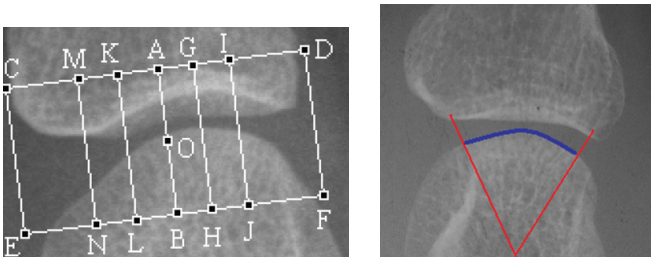
Little to no research has been done by the image processing community: Peloschek et al. (2007) presents a method for automatic joint space detection and measurements for MFC joints. In this work, an initial training set of joints is segmented interactively with a live-wire algorithm and an Active Shape Model (ASM) is generated with the segmentations. A Neural network locates the position of the metacarpo-phalangeal joints in the hand domain, and an ASM segments joint contours. Once segmented, distance measures are computed from the upper contour to the lower one. As a whole, the method is too focused on a single subset of joints, and perform well because has tight control on the input images. We consider that, the method presents two main shortcomings.First, only MCP joints are considered in this scheme, resulting in four joints analysed per hand. Thus, it cannot be used to automatize VdH scoring as the four joints represent a tiny subset of the joints required for a complete evaluation. Additionally the method discards joints that are too damaged, with luxation, severe reduction of the joint spaces or images with low quality. This is a main limitation for assessing pathological cases in clinical practice.

The pioneering works by Bielecki et al. (2008) and Zieliński (2007) extended the detection and measurement to all finger joints. The overall idea is that the joint is a structure perpendicular to the bone marrow. Joint detection is a twofold process: first, bone marrow is computed and, then, a specific pattern along it is computed. Bone marrow is computed by applying otsu thresholding to a dilation of the image. As the difference of foreground and background has been boosted by the dilation, otsu effectively separates hand from background. Since otsu threshold includes flesh and bone (see figure 5(a)), the skeleton of the otsu mask is only an approximation to the true bone marrow. In particular, it looses the metacarpal bones located inside the palm of the hand, especially for thumbs and little fingers as shown in figure 5(b). In order to overcome this systematic artifact, the author's use an heuristic approach by projecting the skeleton of the finger by a length relative to the proportions of the fingers. However, this skeleton projection is prone to fail if the metacarpal and the proximal phalanx bones are not aligned, either because of luxation or the posture of the hand. As a result, there is a significant drop in detection rates for MFC joints in thumbs and little fingers. Joint detection is based on searching along the skeleton path, and drawing a profile plot while searching for the maximum bright-dark difference. Searching only in a pixel wide area is too restrictive and might drop its efficiency at DIP and PIP joints as such joints have a narrow separation at the center, prone to collapse at early stages of RA. Regarding joint space width, a series of distance measurements along lines parallel to the skeleton and covering the whole joint width are computed. We consider that this approach has two main limitations for its application in clinical practice. First, measurements are not normal to the upper bone as physicians do. Secondly measurements cover areas of the bone that are not considered by medical doctors, who restrict to a cone area as shown in figure 6(b).



(a) (b)

Figure 5: Current approach performance. Skeleton computation using otsu method: (a) otsu mask including flesh and bone and (b) thinning-based skeleton.



(a) (b)

Figure 6: Current approach performance. Measurements in joint areas: (a) measurements along M and I line will give erroneous joint space width measurements, (b) area of measurement should limited to the triangle area.

1.2 Objective and contributions

The goal of this master's thesis is to develop suitable methodologies to help automatizing the evaluation of the RA using Van der Heijde standards in X-Ray images. In particular we will focus on joint space width measurements.

This work contributes to existing works in three aspects:

- ***Joint detection***. We present a synthetic model of joints based on a second derivative of an anisotropic Gaussian oriented across bone marrow that takes into account information in an area of each pixel in order to define a jointness measure
- ***Joint space characterization***. We define absolute and relative measures in terms of the curve distance between the two joint bones and the asymmetry in joint space reduction
- ***Correlation to VdH index***. We explore for the first time the relation between joint space descriptors and the VdH score.

The remains of the work are organized as follows. Section 2 is devoted to joint detection and section 3 to joint characterization, including the joint space reduction measures. Section 4 details the experimental settings and section 5 reports the results. Finally, section 6 exposes the discussion and further lines of research.

2. Joint detection

In order to locate the joints of the hand, we must perform an initial extraction of the region of interest where we expect to find the joints, to speed up the whole detection time. Naturally in our case we want to detect the region of interest (ROI) covering the bones of the hand, so initial steps of the whole process involve segmentation of the bones. The process of locating the position of the joints inside the image is based on finding a good synthetic model of an ideal joint. Examination of the images show that a joint is a ridge (sclerosi of the bone) oriented perpendicular to the direction of the marrow, followed by a valley (joint space) equally oriented (figure. 7(a) and (b)). Therefore, we model a joint by means of the second derivative of an anisotropic Gaussian oriented perpendicular to bone marrow (figure. 7(c) and (d)). In order to apply our model inside our own bone ROI, we need to determine the direction of the finger so we can use the correct orientation in which the model has to be applied. Therefore, joint detection is split in two main steps: bone marrow computation and joint pattern detection.

2.1 Bone marrow computation

Segmenting the bone from the rest of the image is a challenging task since the bone structures present a great variation in image intensities. Around the edges of the bone, where the x-ray traverses mostly cortical bone tissue, the resulting pixel is brighter while in the middle of the bone, where the x-ray traverses less cortical bone and more travecular bone and marrow the pixel value is lower. In fact, background, flesh and bone can have the same gray level value at some points.

There are multiple methods for bone segmentation in X-Ray images. Focusing specifically on hand bone segmentation, significant amount of work dealing with X-ray images has been related with bone age assessment: the task of determine the age of a patient from the structure and distribution of the hand bone structure, which during childhood, presents an evolution over the number of bones and their position. Examples tackling this problem with different approaches are Niemeijer et al. (2003) using graph based approaches and active shape models, Han et al. (2007) and de Luis et al. (2003) using snakes and Hsien et al.

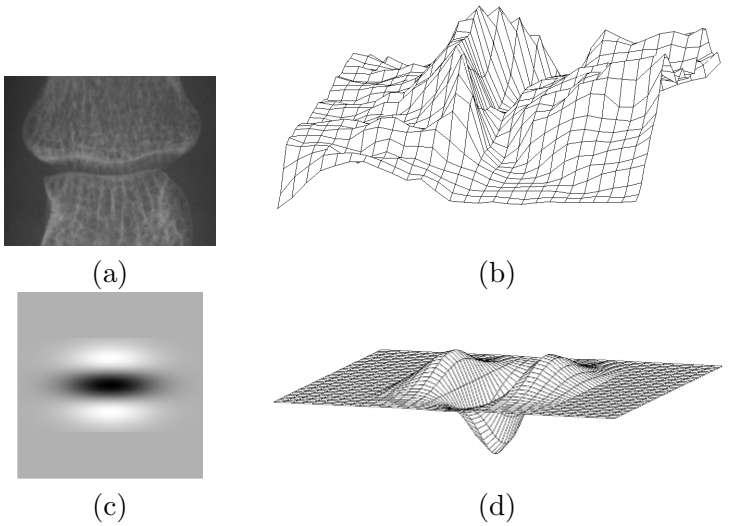


Figure 7: Synthetic joint model: (a) bone joint appearance in X-Ray images, (b) height map of the joint, (c)second derivate of an anisotropic Gaussian as an image, (d) 3D view of the anisotropic Gaussian.

(2007), using canny edge based segmentation. Given the reduced number of images that we had, methods that require training or supervised learning were discarded. Also, we aimed to generate a fully automatized method, not requiring any kind of human intervention, either in the way of training or initialization (initial seed points for example). Both circumstances limited our choice of methodology used for segmentation to unsupervised methods

Given the distribution of gray level values in the images, thresholding methods on image intensity are not able to correctly segment the hand bones (Figure 5(a)). In order to properly detect bones by thresholding we will compute first a bone descriptor. Given that bone local appearance is higher than the flesh, we will use a blob detector. While scale invariant blob detectors (Lindeberg (1993)) are prone to give a response to both flesh and bone, the Laplacian blob detector is biased towards the smaller isotropic scale, which reduces the response to flesh. We compute the Laplacian blob detector as a filter bank of second derivatives of bidimensional isotropic Gaussians. The range of the scale of the filter bank is set by the size of the average phalanx.

Let $I = I(x, y)$ be a two dimensional image, then the Laplacian of I is defined as:

$$\Delta I = \partial_{xx} I_\sigma + \partial_{yy} I_\sigma \tag{1}$$

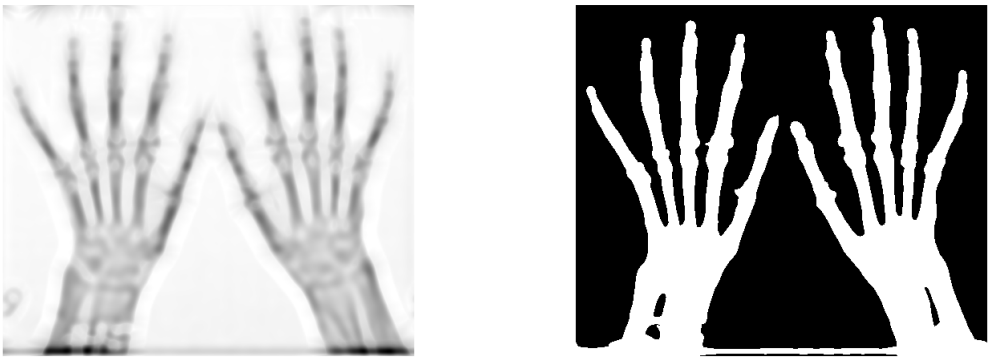
where $\partial_{xx} I_\sigma$, $\partial_{yy} I_\sigma$ are second derivatives defined by the convolution:

$$\partial_{xx} I_\sigma = \partial_{xx} g_\sigma * I \text{ and } \partial_{yy} I_\sigma = \partial_{yy} g_\sigma * I \tag{2}$$

for g_σ an anisotropic Gaussian kernel of variance σ. The laplacian operator has strong positive responses for dark blobs of size $\sqrt{\sigma}$ and strong negative responses for bright blobs

of the same size. Figure 8(a) shows the maximum responses to the laplacian filter bank. Note that the highest negative response (dark areas in the image) is attained at hand bones.

The bone region is obtained by binarizing the negative responses to the Laplacian filter bank. In order to obtain a bone mask as complete as possible (i.e. without gaps), we use a k-means with three clusters and we correct the subsequent oversegmentation of the image by setting the cluster with bigger area as the background, and merging the results of the other two. We use three clusters instead of two, in order to to achieve a better clusterization in regions with lower blob response. Such regions are located mainly in blobby areas bigger than the scales used in the filter bank. These (carpal region, or epiphysis of bones). The k-means clusterization gives a region of interest that may contain some small gaps, that can be selectively filled if they are smaller than a certain size by performing area calculations on the holes of the mask and marking them as foreground. This hole filling has to be constrained to a maximum value or we may end up by closing a wrong hole, formed between legit bone and image noise or the border of the image. Holes are created because the high variance of intensity inside bones, creates areas that are in fact dark blobs or that have very low response to the filter bank, so they are not detected by the Laplacian detector that specifically aims to detects bright blobs. Figure 8(b) shows the finalized clusterization of the blob response in Figure 8(a).



(a) (b)

Figure 8: Blob detection and binarization, a) blob response, b) final mask.

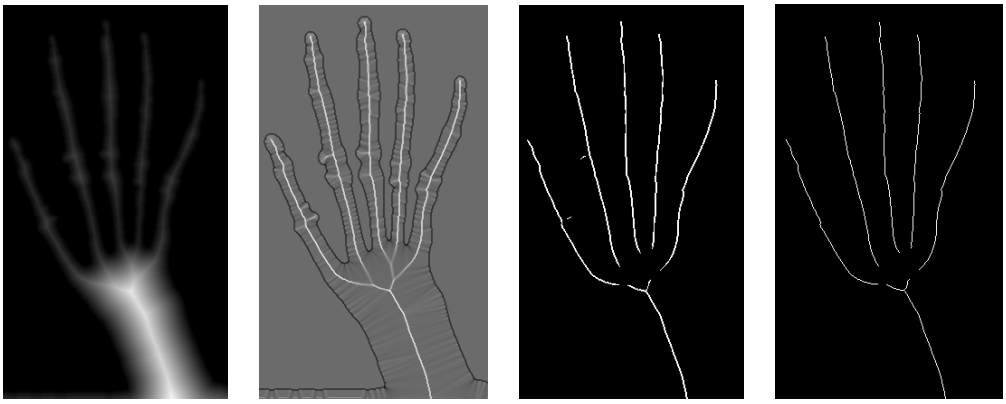
Bone marrow is computed as the skeleton of a segmentation of hand bones. The method selected to generate the skeleton significantly affects the calculation of the direction of the bone marrow. As a mater of fact, an ideal skeleton should outline the general direction that the fingers have and present as less spikes (spurious branches) as possible. Skeletons generated by morphological image operations such as iterative thinning, generate *noisy* skeletons, with spikes due to changes in the curvature of the mask boundary. Such spikes can corrupt the finger direction calculation so they have to be removed. Instead of using morphological-based skeletonization methods, we use a method to generate smoother skeletons with a reduced number of spikes, based on the ridges of the distance map to the bone mask.

The distance map is generated by computing the Euclidean distance transform of the binary contours. For each pixel in the mask, the distance transform assigns a number that is the distance between that pixel and the nearest edge pixel of the mask (see fig 9(a)). The centerlines (skeleton) can be computed as the ridges of the distance map. We compute ridges using the normalized creaseness measure based on the structure tensor described in López et al. (1999). Let D denote the distance map to the mask boundary and $\nabla D = (\partial_x D_\sigma, \partial_y D_\sigma)$ its gradient, where image derivatives are computed using gaussian kernels, g_σ, of variance σ (differentiation scale) as: $\partial_x I_\sigma = \partial_x g_\sigma * I$ and $\partial_y I_\sigma = \partial_y g_\sigma * I$. Then, the structure tensor is given by the convolution

$$ST(\rho,\sigma) = g_\rho * \left[\begin{pmatrix} \partial_x D_\sigma \\ \partial_y D_\sigma \end{pmatrix} (\partial_x D_\sigma, \partial_y D_\sigma) \right] = \begin{pmatrix} g_\rho * \partial_x D_\sigma^2 & g_\rho * \partial_x D_\sigma \partial_y D_\sigma \\ g_\rho * \partial_x D_\sigma \partial_y D_\sigma & g_\rho * \partial_y D_\sigma^2 \end{pmatrix} \quad (3)$$

for g_ρ a Gaussian kernel of variance ρ (integration scale). Let V_1 be the eigenvector of principal eigenvalue and consider its reorientation $V = (P, Q)$ as $V = sign(< V_1 \cdot \nabla D >) \cdot V_1$ for $< \cdot >$ the scalar product. Then, the creaseness measure is given by the divergence:

$$R = div(V) = Px + Qy \quad (4)$$



(a) (b) (c) (d)

Figure 9: Bone marrow computation: a) distance map to bone mask, b) ridges and valleys of distance map, c) thresholded ridges and d) final skeleton representing bone marrow

The ridge and valley image is thresholded to obtain the stronger ridges. The thresold value is fixed for all the images as the ridges and valleys are normalized values. The generated thresholded image is then thinned to generate one pixel-wide lines and then we remove small unconnected lines. Most of the spikes that may be generated as small ridges often get disconnected from the main branches and hence, are eliminated in this step.

The skeleton is further prunned below the wrist in order to save computational time and avoid errors at the wrist where the complex structure makes the skeleton unable to track all

the joints between small bones. The wrist area is located under the assumption that finger branches of the skeleton start at a central point on the wrist area. We detect this multiple junction point by computing the local topology of the complementary of the skeleton mask. The number of connected components in a sliding windows gives the local topology of the skeleton complementary. Note that for windows at the center of the palm/wrist, we should expect to find a higher amount of regions than in other areas of the image as illustrated in figure 10. In order to perform a better localization of the wrist, several sliding windows of different size are tested in a coarse to fine strategy.

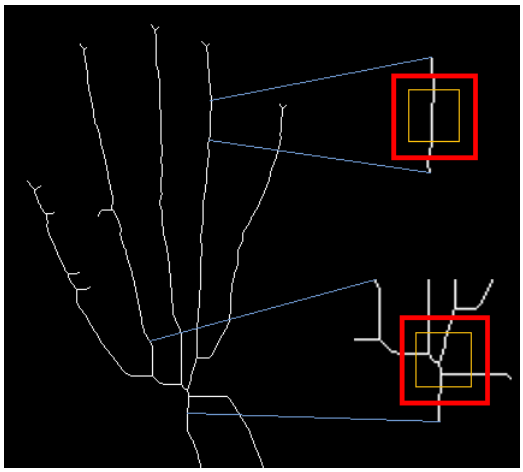


Figure 10: Coarse to fine wrist detection scheme

2.2 Joint detection

Joints are detected by convolving the image with our template of a joint given by a second derivative of an anisotropic gaussian (see equation 5) oriented perpendicularly to the bone marrow. Let θ be the direction perpendicular to the bone marrow and $\sigma = (\sigma_1, \sigma_2)$ the anisotropic scale with $\sigma_1 > \sigma 2$. Consider the oriented anisotropic gaussian:

$$g_{\sigma,\theta} = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{\tilde{x}^2}{2\sigma_1^2} - \frac{\tilde{y}^2}{2\sigma_2^2}} \tag{5}$$

where $\tilde{x}, \tilde{y}$ are the x and y axes rotated θ radians in order to be aligned to the bone marrow:

$$\begin{aligned} \tilde{x} &= -x\sin\theta + y\cos\theta \\ \tilde{y} &= x\cos\theta + y\sin\theta \end{aligned} \tag{6}$$

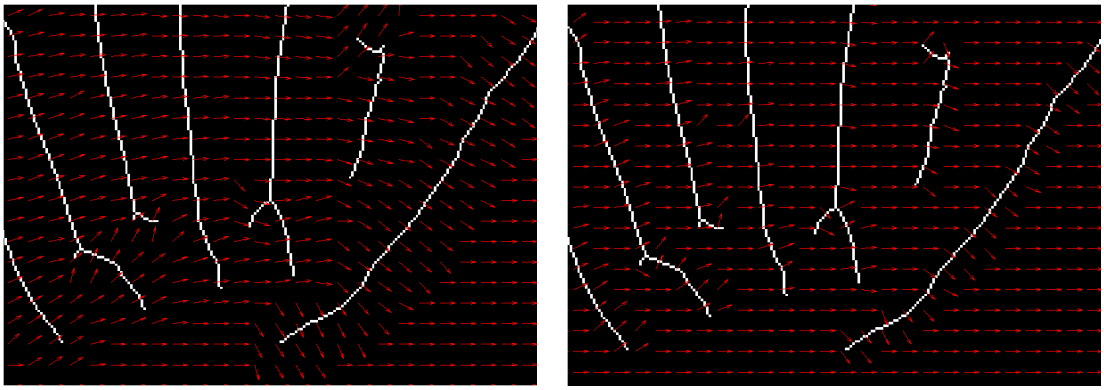
Then, the convolution of the image with the second derivative along $\tilde{y}$:

$$\partial_{\tilde{y}\tilde{y}} g_{\sigma,\theta} = -\frac{1}{\sigma_2^2}(g_{\sigma,\theta} + (-\tilde{y}^2 \frac{g_{\sigma,\theta}}{\sigma_2^2})) \tag{7}$$

gives a measure of the local similarity between the image intensity and our pattern. Large positive responses correspond to valleys (figure 12(a)) and large negative ones to rigdes

(figure 12(b)). We will note by J_p the image of positive responses and by J_n the image corresponding to the negative ones.

The direction of the bone marrow, θ, is computed using the principal eigenvector of the structure tensor given in equation (3) computed for the skeleton image. By using the appropriate integration scale in the structure tensor, the principal eigenvector direction is more robust with respect to bifurcations or gaps on the skeleton, as shown on figure 2.2. In order to reduce computational time and the impact of bone edges not belonging to joints, we restrict convolutions to the branches of the pruned skeleton.



a) b)

Figure 11: Gradient obtained using different sigma integration values. $\sigma = 6$ (left) and $\sigma = 2$ (right)

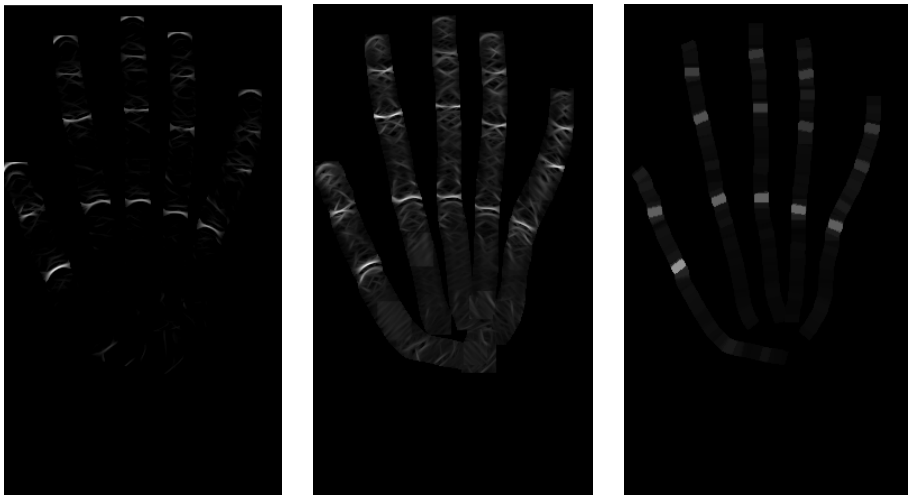
The convolution images J_p and J_n, need to be combined in a single score representing the likeliness of having a joint. Traversing the skeleton from fingertip to wrist, a pixel of the skeleton will be a joint if negative responses in an area -oriented perpendicular to the marrow- above its position are large and positive responses in an equally oriented are below it are large too. Upper and lower oriented areas at a point, (x_0, y_0), are given by the positive and negative values of a first derivative of an oriented anisotropic Gaussian kernel centered at (x_0, y_0) with scales chosen to cover the width of the bone:

$$\partial_{\tilde{y}} g_{\sigma,\rho} = -\tilde{y} g_{\sigma,\rho} / \sigma_2^2 \tag{8}$$

with $\tilde{x} = -(x - x_0)\sin\theta + (y - y_0)\cos\theta$ and $\tilde{y} = (x - x_0)\cos\theta + (y - y_0)\sin\theta$. In order to minimise noisy spurious responses and increase responses at collapsed joints, our jointness measure is given by the integral:

$$J = \int_{\{\partial_y g_{\sigma,\rho} < 0\}} Jp(x, y)dxdy - \int_{\{\partial_y g_{\sigma,\rho} \geq 0\}} Jn(x, y)dxdy \tag{9}$$

Figure 12(c) shows the jointness measure desired from the positive and negative responses shown respectively in figure 12(a) and figure 12(b). While positive responses sharply discriminate cartilage, the negative ones include the joint sclerosis and other part of the finger bones. By combining both responses in an structured way, our jointness measure attains significant discriminant power.



(a) (b) (c)

Figure 12: Decoupling of the response to the synthetic joint model: (a) Positive, (b) Negative and (c) final jointness result.

2.3 Anatomy-based selection of joints

This jointness response is low for noise and high where the shape of a joint is found. However, in some cases, where the joint is almost collapsed and has lost partially the shape we are modelling, we might get a low jointness value. This makes the decision of what minimum values of jointness really corresponds to a joint a difficult task. We need to discard any kind of global decision. For instance, setting a threshold of minimum response to separate real joints from false ones or using either otsu or k-means to select true joint responses from noise, fails to detect the thinnest of the joints. Any selection scheme has to work locally if we want to locate the joints correctly.

The best approach is to use some a priori information about hand structural anatomy to improve the detection ratio. We propose a finger by finger approach based on the fact that, for each finger, the three (two in the case of thumbs) highest responses should correspond to joints. The method works by splitting the skeleton in individual branches, and trying to locate which branches belong to fingers. A combination of size and position is used to obtain this information: finger branches are usually the longest branches of the skeleton, while thumb branches are additionally located near the center of the image. Figure 13 shows joint detections for different degrees of RA types.

3. Joint space characterization

Joint space is given by the pixels lying between the sclerosis and the edge of the lower bone. The brightest surface in the upper bone represents a concavity in the bone called *sclerosis* formed when the X-Ray traverses the more dense cortical bone. Detecting the sclerosis in the joints is crucial because distances have to be calculated from the sclerosis surface in

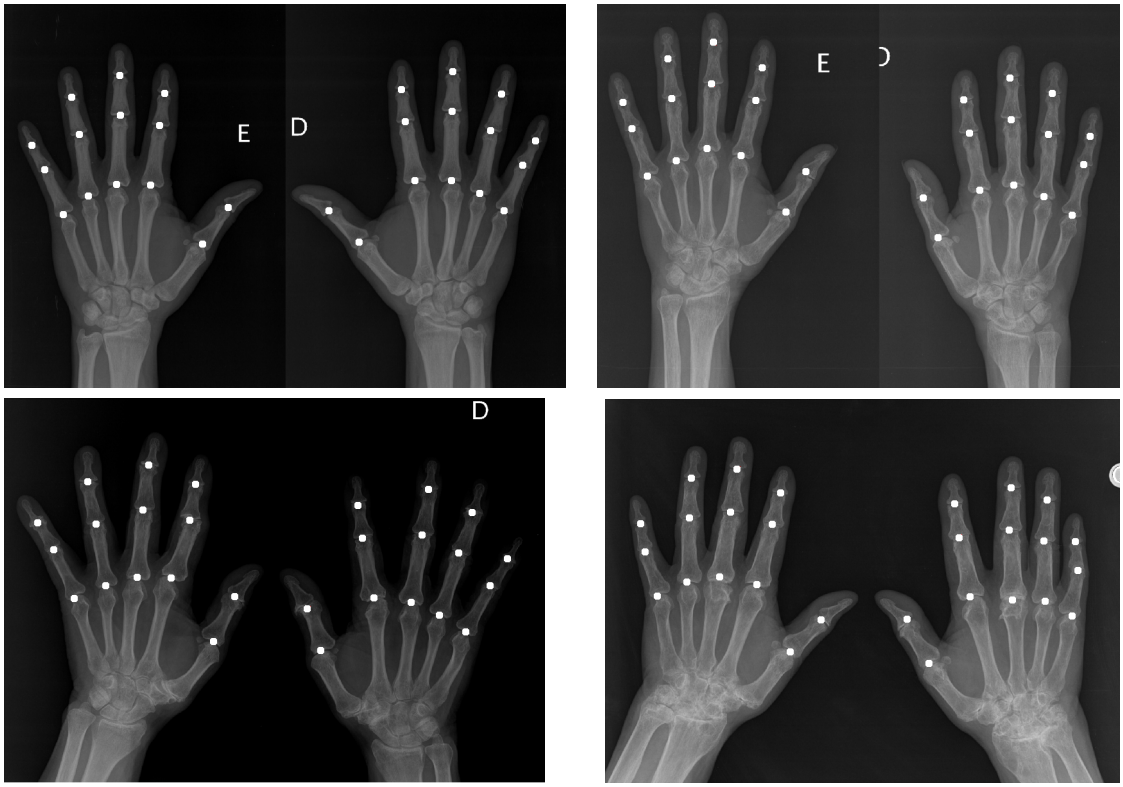


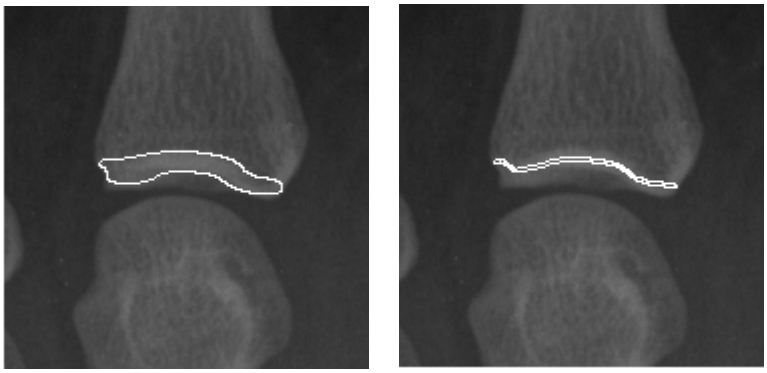
Figure 13: Joint detection for healthy hands (top row) and hands with affected joints (bottom row).

the upper bone to the edge of the lower bone. The space that separates the two bones of the joint in X-Ray images, is not really empty as it seems in the images, but occupied by the cartilage that protects the bones from friction. However, the cartilage tissue does not absorb any noticeable amount of X-Ray radiation so its completely transparent in X-ray images. This fact makes that the joint space in an radiography can be seen as a dark valley. In order to characterize and measure joint spaces in the detected areas, we will use our definition of joint in terms of the oriented ridge and valley pattern.

The sclerosis is defined as the ridge of the positive response (computed using the normalized correlation) J_n in the joint area as shown in figure 14(a) and (b). The upper contour of the lower bone is given by the lower extreme of the binarized response J_p in the joint area, (as shown in figure 15(a)).

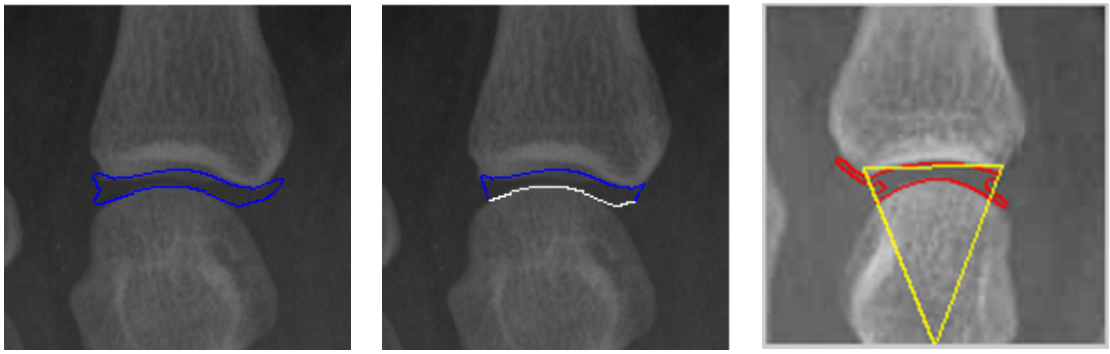
The scope of the lower bone line has to be limited depending of the bone shape in order to be consistent with what physicians do (see figure 15). According to the restriction in movements that each joint has, we can distinguish two types of joints:

- Condyloid joints: they are located in the metacarpo-phalangeal joints, and they allow two degrees of freedom of movement. The bones in this joint have curved surfaces, resembling a ball and socket joint.



(a) (b)

Figure 14: Sclerosis computation: (a) sclerosis detection and (b) refined segmentation



(a) (b) (c)

Figure 15: Lower bone computation: (a) lower bound computation: (b) measurement zone delimited, (c) another example of delimitation

- Hinge joints: they are located at the interphalangeal articulations, and its movement is limited to one degree of freedom. The bones constituting the joint are almost planar surfaces.

By their curved pattern, condyloid joints require a specific protocol. In order to capture the valley representing the joint space width in this kind of joints, we have to search for a valley oriented in the correct angle with some variation in the inclination. The curvature of condyloid joints makes the detected valley to extend outside the desired measurement area. The valley has been pruned to an area defined by the triangle beetween the extremes of the sclerosis and the inferior extreme of the image (taking into account the direction of the finger). A clearer example can be seen in Figure 15(c). We see that the detected valley (red contour) is too far from the upper bone at some points. By cutting the valley by the zone limited by the area we are easily approximating the actual region of measurement with a simple methodology.

3.1 Learning the Van der Heijde score for joint narrowing

According to physicians, measurements between the sclerosis and the lower bone should be performed not in a parallel way to the finger direction but taking into account that both surfaces have curved shapes. Therefore, although measuring the distances in straight parallel lines may give accurate enough values, we consider that the correct approach in order to mimic physicians is taking into account the distance to the sclerosis curve. The distance of an arbitrary pixel y to a given surface X, is defined as:

$$d_X(y) = min_{x \in X} \|y - x\| \tag{10}$$

Our approach to compute the above distance is given by computing the distance map from the sclerosis and then, evaluate it at the lowermost pixels of the valley that represents the joint space.

For all distances, $d_X(y)$, along the lower bone, we consider the following scores:

- ***Score1: Minimum distance***. It is given by minimum of $d_X(y)$ over each joint. The minimum should be a good indicator of the degree of joint space width reduction.
- ***Score2: Average distance***. It is given by the average of $d_X(y)$ along each joint. The mean should also be a good discriminant of affected joints. which should have lower values.
- ***Score3: Distance standard deviation***. It is defined as the standard deviation of all the values $d_X(y)$ on the joint. This score should give information of the homogeneity of the measurements. Higher standard deviations should be associated with asymmetry in the loss of cartilage and thus it should increase too with VdH score.

In order to obtain relative measurements independent of resolution or size of the images and joint type, we compute the following additional scores:

- ***Score4: Average relative distance***. We compute a relative distance based on the ration between the mean distance and the maximum distance along the lower bone line.
- ***Score5: Coverage***. It is defined as the ratio between the number of pixels detected in the sclerosis, N_s, and lower bone line, N_b, as $min(N_s/N_b, N_b/N_s)$. The coverage always ranges from zero to one, and indicates the quality of the segmentation as well as severe asymmetry in the joint.

Figure 16 shows our scores for Van der Heijde Index assessment for different joint types.

4. Experimental settings

4.1 Data sets

In the context of medical image processing aimed at helping diagnosis, there are no public databases and images have to be gathered (and manually labelled) for each particular application. Our data set consists in analogical and digital images. The first data set was used to learn the model of a joint, while the second one was used in the validation of the system.

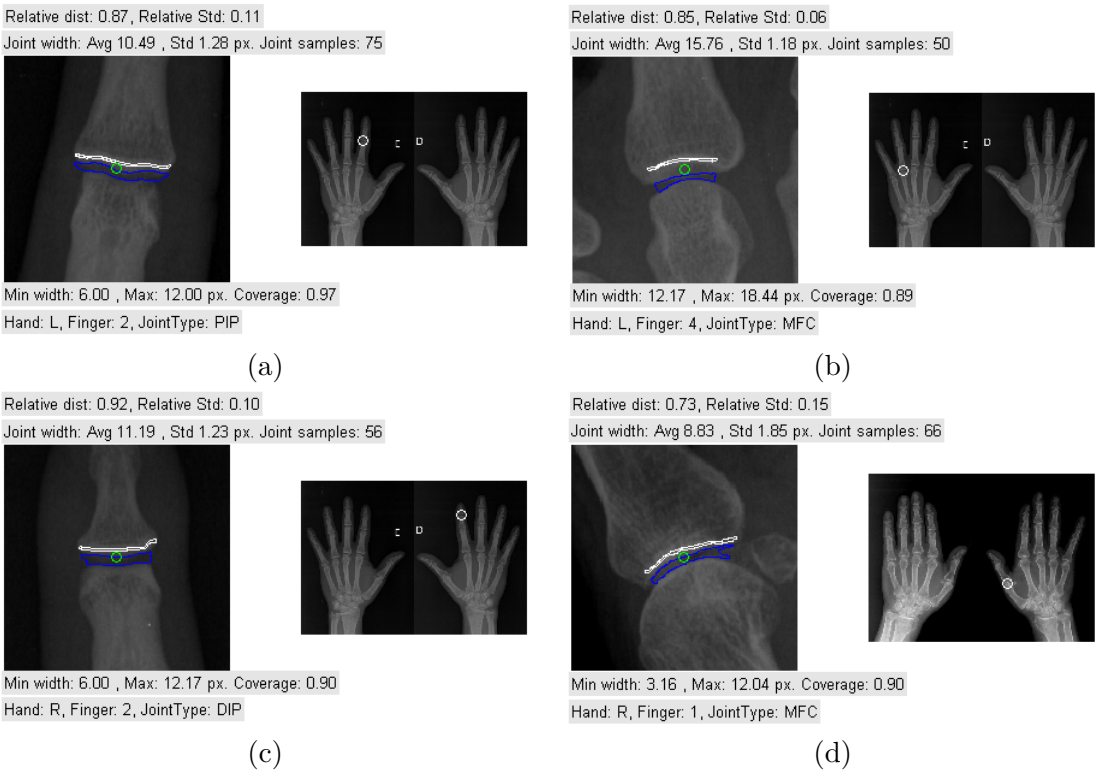


Figure 16: Joint space measurements for different kind of joints: (a) Index PIP, (b) Ring MCF, (c) Index DIP, and (d) Thumb MCF

4.1.1 Analogical dataset

Initially, we were provided with a set of seven analogical X-Ray images from five different patients at several stages of RA. Although the images had no associated score for each joint nor a global Van der Heijde score, they were tagged with a very roughly score, ranging from 1 to 4 depending on the level of affection of the hands. Images had been acquired using different X-Ray devices so that they had a different contrast and brightness. In order to ensure the maximum homogeneity across them, we tested three different digitalization settings:

- Automatic selection: the scanner's software selects the intensity values of maximum and minimum intensity that optimize the visualization.
- Fixed values: the two values are set to two fixed values, and those values are kept the same for all the input images.
- Manual selection: the maximum and minimum intensities are set by hand so that the maximum value of intensity is picked by selecting a subjectively the brightest bone region and the lowest value picked from the background of the image.

Figure 17 shows some examples of the sample radiographies and the different outputs of the three different variations of scanning. Often the third technique produces too saturated images (last row of images in figure 17) where some of the information of the image is lost due to an excessive "exposure". The other two methods produce similar image quality but we prefer to use the automatic one since it produces better results for low contrast images (see the second column of images in (last row of images in figure 17).

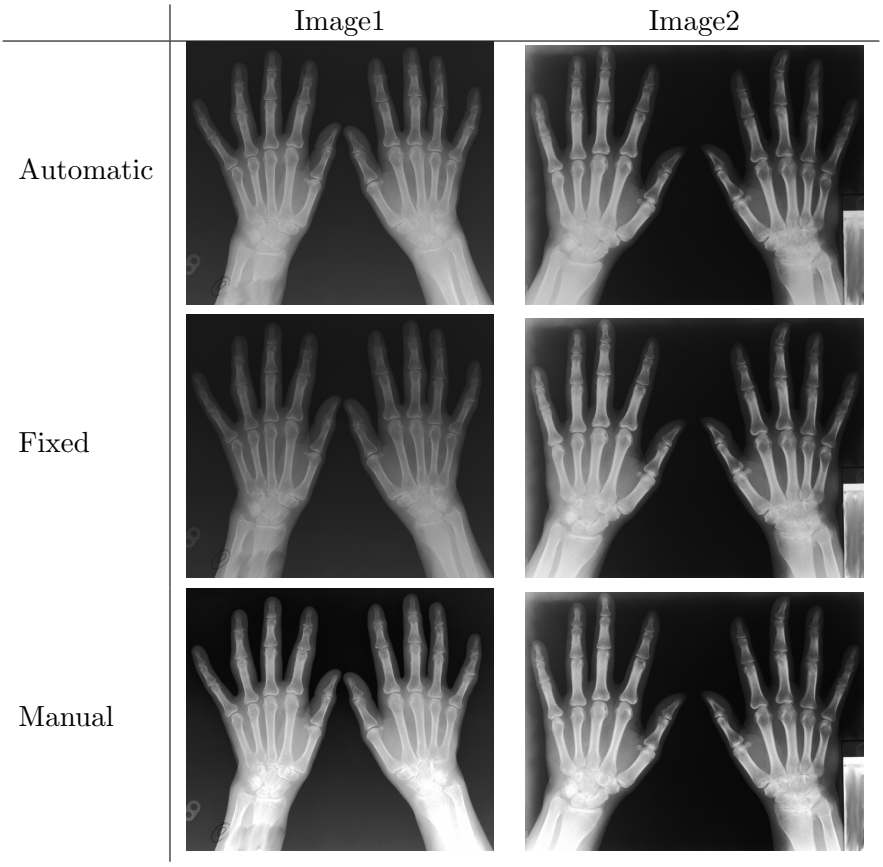


Figure 17: X-Ray digitalization for the three scaner settings (rows) for a standard X-Ray quality image (Image1) and alow contrast one (Image2). Automatic settings in top row, fixed settings in middle row and manual settings in lower row.

4.1.2 Digital dataset

Several months after we received a second set of X-Ray images, in digitised DICOM format with scores of joint-space narrowing for each joint. The scores of joint space reduction only ranged from 0 to 3. We specifically excluded the most pathological cases (score of 4) as they are easily identified by the physician and because current palliative treatments prevent such advanced stages are. The real interest lies in the intermediate cases where the damage of the joint is not so severe and the correct cure may stop the degenerative process. This

second group of images consisted on 20 images of both hands, for a total of 560 joints. The joints that are used in the Van der Heijde score had their associated joint space narrowing score, making a total of 360 scored joints. Six of the images came from healthy patients, and the rest ware patients with several degree of RA affection. The ratios of joints according to their Van der Heijde score are the following:

- 251 healthy joints (69.7%)
- 51 joints with score of 1 (14.2%)
- 43 joints with score of 2 (11.9%)
- 15 joints with score of 3 (4,2%)

The dataset is clearly unbalanced: healthy joints are much more abundant than the rest, and we can consider the number of samples of medium affected joints lower than desired. This is because typically RA does not affect all the joints with the same intensity, and even patients who have joints scored with 3 may have several unaffected joints.

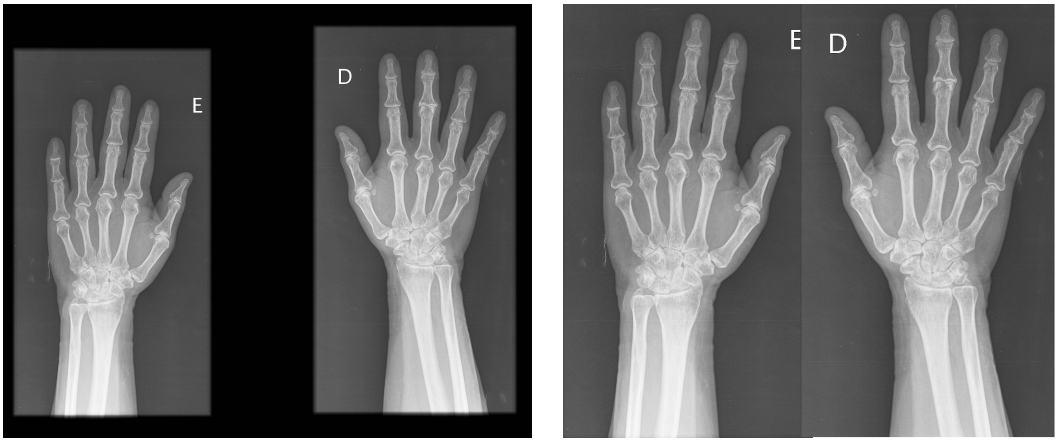
The preprocessing of this set of images was significantly different from the analogical one. Firstly, this images are in DICOM format with 2828x2320 resolution and with intensities ranging from 0 to 4095 (16 bit image). Images in DICOM format can not to be used directly by our algorithm and had to be preprocessed. The preprocessing included conversion to tif format and application of the same window-level transform to all the images. Also, some images had both hands separated in two different frames inside the image (see Figure 18(a)). This artificial partition impacted negatively on our method,as it introduced artificial contours which misleaded the blob detection stage, making the whole algorithm non functional. Therefore, the images had to be manually cropped to discard the frames. Also, the images included large portions of the radius and the ulna, that have no interest to RA and are prone to deviate bone marrow computation. diagnose so the images were cropped also to exclude most of the unwanted radius and ulna.

4.2 Validation protocol

Three different experiments have been carried out:

1. ***Joint detection rate***. We have considered a global score and a joint-type-wise one. The global joint detection rate is measured for every image by dividing the number of accurate located joints by the number of hand joints (14) and calculating the range for all hands (given by the average $\pm$ standard deviation. The joint-wise one is measured by dividing the number of accurate located joints for a given type by the number of hands (40 in our case).

2. ***Joint characterization rate***. We have also considered a global and a joint-wise score. The joint characterization rate. Both are given in terms of the number of accurate characterized joints like the joint detection rates. We consider that a joint has been accurately characterised if both the sclerosis and the lower bone lines visually match the expected position in the image. For this experiment only joints correctly detected have been taken into account.



(a) (b)

Figure 18: Example of digital dataset image: (a) original and (b) and it's manual fixing.

3. ***Towards correlation to VdH score***. Since the goal is to explore the correlation between our scores and VdH, joints not successfully detected or characterized have been excluded. We have compared the features extracted at the joints with the manual label. As a preliminary approximation, we have considered boxplots grouped by the VdH manual label. Since absolute distances depend of the type of joint, we have considered different boxplots for each joint type.

5. Experiments

5.1 Joint detection rate

The global detection ratio range is 95.54 ± 4.16 %, which compares to the numbers given in Bielecki et al. (2008). The first column in table 1 reports the joint-wise detection error for all joints, and Figure 27 graphically represents such errors. In general, we achieve an accurate joint-wise detection rate (above 90%). Thumb joints are the most rarelly located (with 10 and 13% of error). For the remaining fingers, we would like to note that, the detection rate mainly depends on the type of joint rather than on the finger (Bielecki et al. (2008)). DIP and MCP are the best located joints (from 0% to 3% of incorrect locations). There is a significant increase in PIP location error (up to 8%) due to bone marrow comptation artifacts. See the discussion section 6 for further details about the failing cases.

5.2 Joint characterization rate

Joint characterization rate has been computed only for joints correctly detected. The global characterization ratio range is $88.79\% \pm 12.1\%$. The second column in table 1 reports the joint-wise characterization error for all joints, and figure 20 graphically represents such errors. Thumb and little finger joints are the worst characterized with errors above 15%. This is mainly attributed to a deviation from our synthetic model of joint (caused either by finger position or severe RA affection) and will be discused in the next section. For the

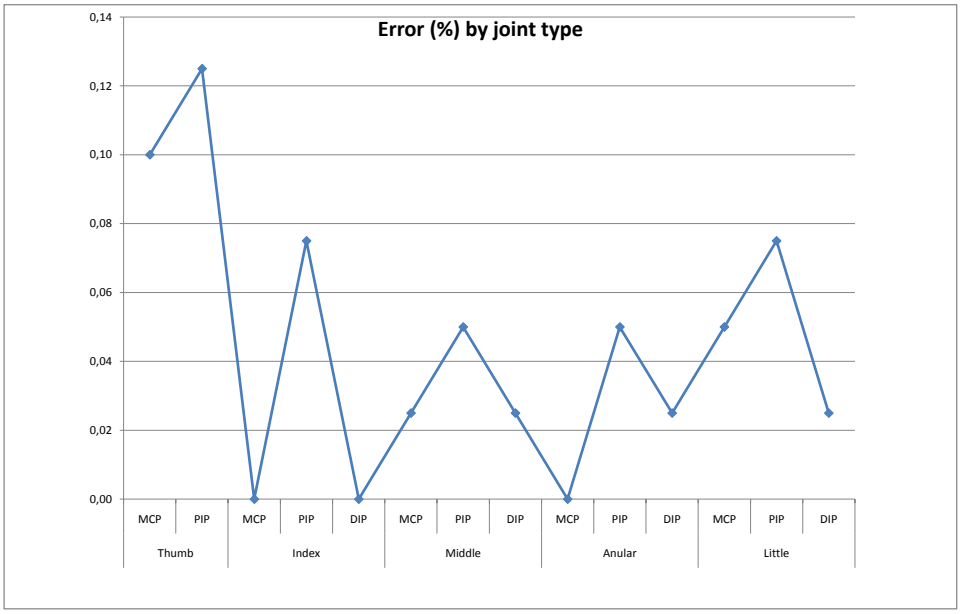


Figure 19: Joint location error for every joint type.

Joint name	Detection error	Characterization error
Thumb-PIP	13%	19%
Thumb-MCP	10%	29%
Index-DIP	0%	3%
Index-PIP	8%	5%
Index-MCP	0%	3%
Middle-DIP	3%	5%
Middle-PIP	5%	5%
Middle-MCP	3%	3%
Annular-DIP	3%	5%
Annular-PIP	5%	8%
Annular-MCP	0%	8%
Little-DIP	3%	16%
Little-PIP	8%	16%
Little-MCP	5%	33%

Table 1: Detection and characterization error per joint

remaining fingers there is not a significant difference among joint types with an average error of 5%.

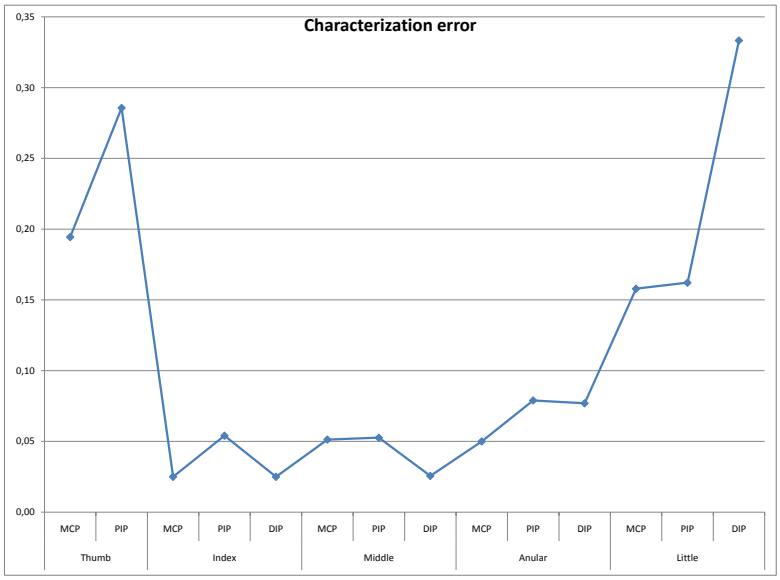


Figure 20: Joint characterization errors for every joitn type

5.3 Towards correlation to VdH score

We recall that joints badly characterized have been excluded, and that only PIP and MFC are considered for VdH scoring.

Figure 21 shows boxplots for the absolute scores for each joint type: minimum distance (score1) in the first row, average distance (score2) in second row and standard deviation (score3) in the third row. The first column shows plots for PIP joints and the second one for MFC joints. In general, our scores present the expected tendency for PIP joints. The average of the minimum distance clearly decreases as VdH increases and, in the case of healthy joints, it is above 75% of the patological samples. Mean distances have the highest correlation to VdH scores and present the strongest decreasing statistics across VdH score (grade 3 omitted). Finally, the trend for the standard deviation is also good since it increases as a function of VdH score and takes characteristic high values for VdH scores equal to 3. For MFC joints the correlation is not so clear and it biased by the low number of samples in the group VdH 3. However, even discarding the boxes corresponding to VdH 3, minimum distances increase with VdH score and mean distances keep invariant. Only the standard deviation presents an incresing dependency for scores 0 to 2. This generalized anomalous tendency in the case of MFC joints is mainly attributed to a bad definition of the triangular area limiting the scope of our measurements (see Discussion for further details).

Figure 22 shows boxplots for the relative scores (average relative distance (score4) in left boxes and coverage (score5) in right boxes) for all the joints: the average relative distance fails to give the expected correlation with a non monotonous tendency. However, the coverage shows a strong decreasing tendency in the average, and takes characteristic high values for healthy joints (VdH equal 0).

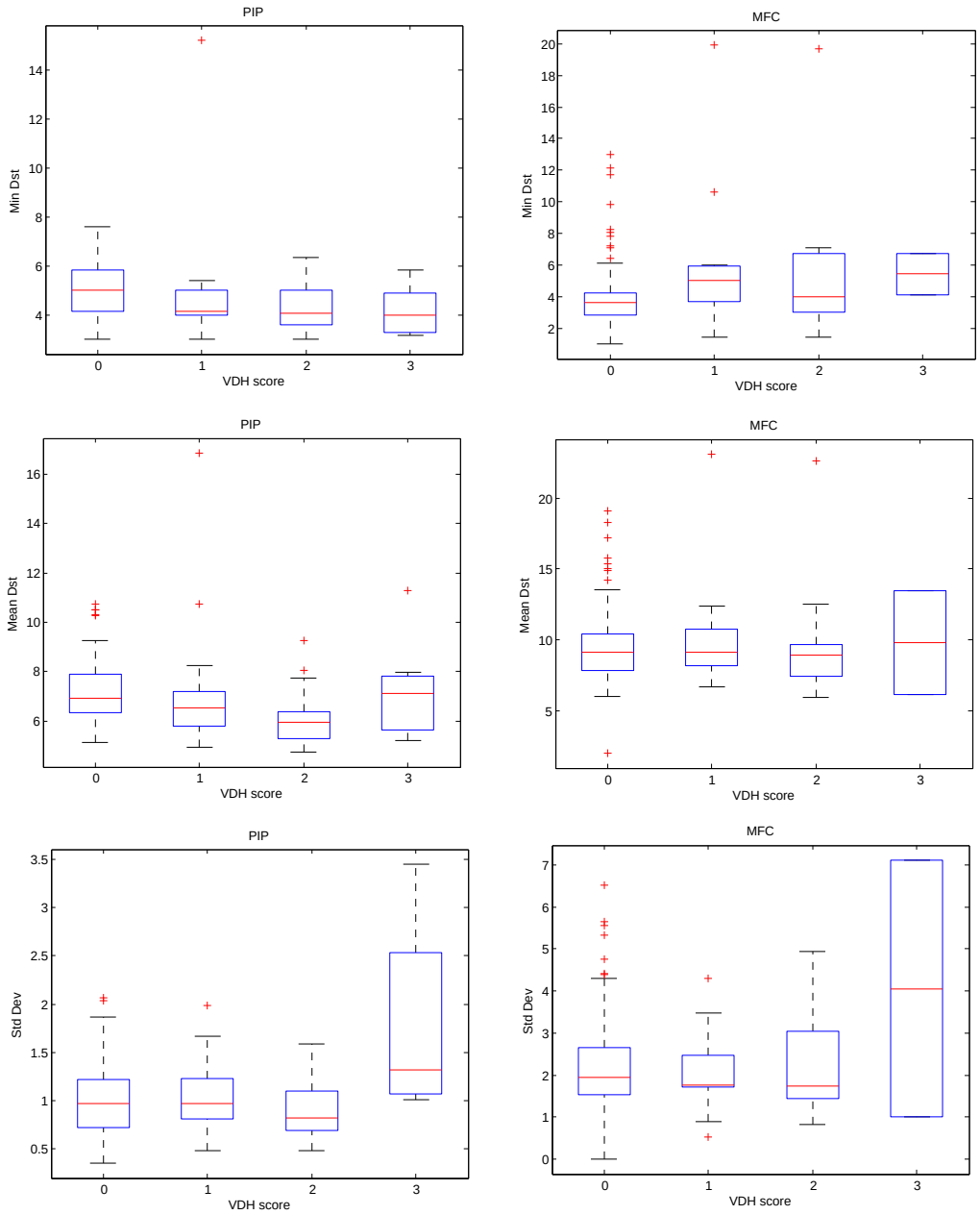


Figure 21: Box plots for absolute scores for PIP and MFC joints. From top to bottom: minimum distance, average distance and standard deviation

6. Discussion and conclusions

The method presented in this master's thesis has two main contributions. On one hand, a fully automatic method for detection and characterization of hand joints has been developed.

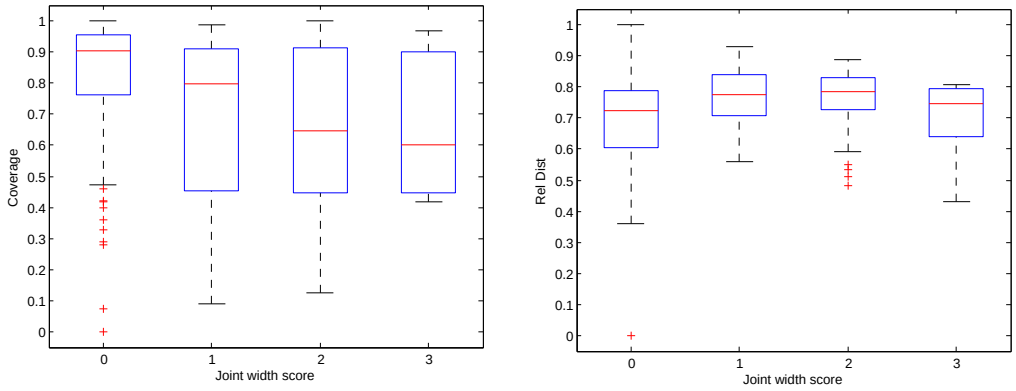


Figure 22: Box plots for relative scores: coverage (left) and relative distance (right).

On the other one, the correlation between VdH score and joint narrowing measurements has been explored for the first time.

6.1 Comparison to state of the art

Our joint detection strategy outperforms the work by Bielecki et al. (2008) in two aspects. First, in general, the detection rates per finger and joint type significantly improve, with a reduction in detection error of up to 50%. Such reduction is achieved by a more accurate computation of bone marrow. Second, by using a jointness measure instead of a pixel-wise approach, our detection shows the same behavior across index, middle and little fingers. This is a significant step forward since evaluation of these fingers is crucial for computation of VdH score (the thumb only contributes with MCF joint). Unlike Bielecki, our performance does not decrease for severely affected joints, which validates our methodology for clinical practice.

Experiments show that thumbs are harder to detect, specially the inter-phalangical joint. This is because when the hand lies over the X-Ray plaque, thumb bones appear in the image as seen from the side, and can overlap or cover each other. This reduces the strength of the valley representing the cartilage and thus, our jointness measure drops. The MFC joint on the thumb although it is situated on a lateral position, shows less tendency to reduce its visibility because his condyloid shape is similar disregarding the angle of view. In the case of MFC thumb joints, often the detection lies in the joint between the carpals and the metacarpal, as shown on figure 24. Failures at thumb DIP joint are not relevant since this joint is not evaluated in VdH narrowing score. Another conflictive joint-type are proximal-inter-phalangical joints. This lower detection ratio has two explanations: first, the PIP joint can present (by its anatomical shape), lower jointness value than DIP joints. However, the main cause for these missdetections is related to issues in the bone marrow calculation. In the pruning phase, when the skeleton is analyzed and spurious branches are removed, two separated finger branches may become fused on a single branch, as shown on figure 25. In such cases there are six joints covered by the same branch, but the method will still search only for three.

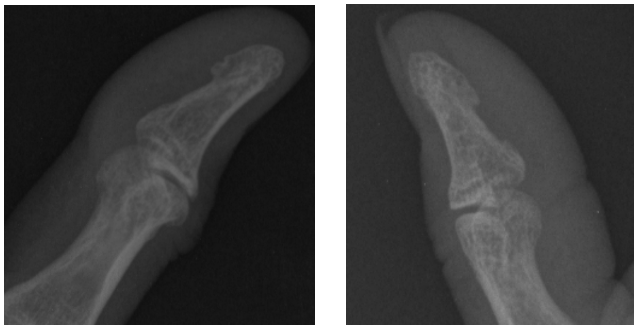


Figure 23: Thumb position on the X-Ray plaque. The detection of the valley representing the cartilage fails.

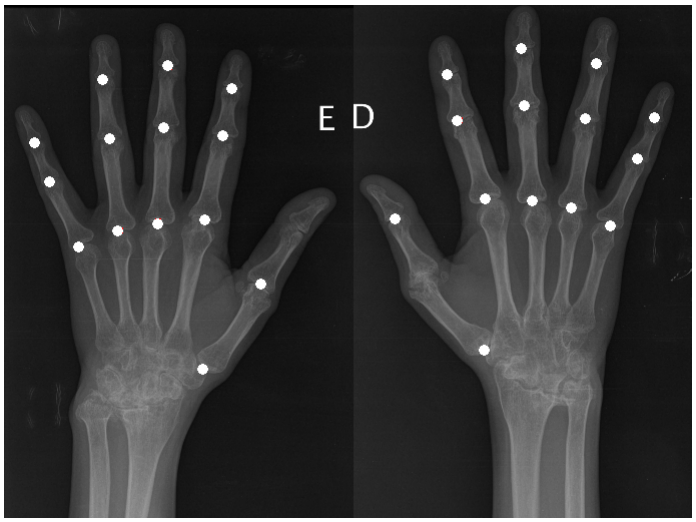


Figure 24: Two typical thumb detection errors. Left hand: the DIP joint has bad visibility. Right hand: the MCP is collapsed and has low jointness. In both cases the algorithm searches for maximum responses, that are found on the carpal-metacarpal joint.

Concerning joint characterization rates, we have to note that our quality measures cannot be numerically compared to the one reported in Bielecki et al. (2008), since they are based on manual delineation of sclerosis and lower bone, and ours are a simple visual match. However the trend of our joint characterization rate error indicates if there is any systematic bias. As in Bielecki et al. (2008), we observe a similar decrease in characterization for thumbs and little fingers and equal behaviour for the rest of the joints. Difficulties with the characterization of the joints are related with several causes. First, the scales we use when searching for oriented valleys may be too big to detect small articulations such as DIP

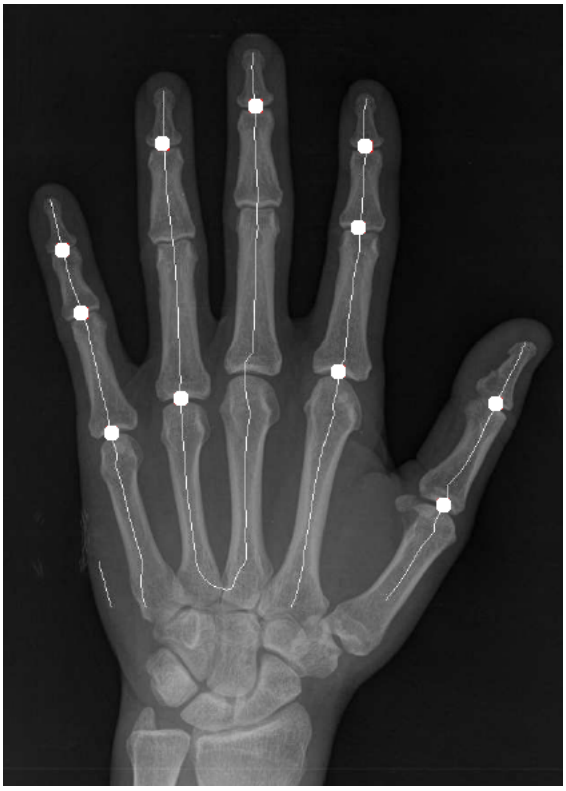


Figure 25: After pruning, the branches of the middle and the annular finger have been merged, and the algorithm searches the three stronger jointness responses, in a branch that has six joints.

joints. Secondly, the lateral position of the thumb may interfere with the characterization of its joints increasing the chances of failure. Finally, on joints that are affected by RA to the extent that the joint space is no longer observable, the characterization may fail. However, neither the thumb PIP nor the little finger DIP are used in the VdH score and this failures have low practical impact.

Figure 27 shows a linear regression test made with the data of this experiment, that shows some correlation with the number of characterization errors with the narrowing score of the hand. A unilateral Spearman's rho test has confirmed also correlation between the two variables with a p-value of $9.9037 \cdot 10^{-5}$.

6.2 Correlation to Van der Heijde Index

Two different kinds of measures are considered: absolute (score 1, 2 and 3) and relative ones (scores 4 and 5). The first group might take different values depending on the joint and thus, they are analyzed separately. The second ones, are designed to be independent of joint type and can be analyzed considering all the joints. Our preliminary study shows

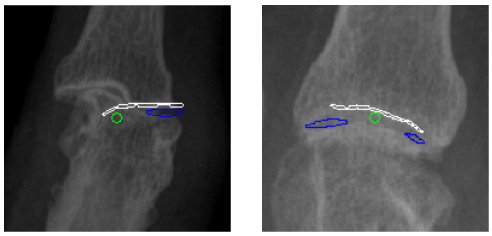


Figure 26: Failure to detect the valley or the sclerosis on collapsed joints.

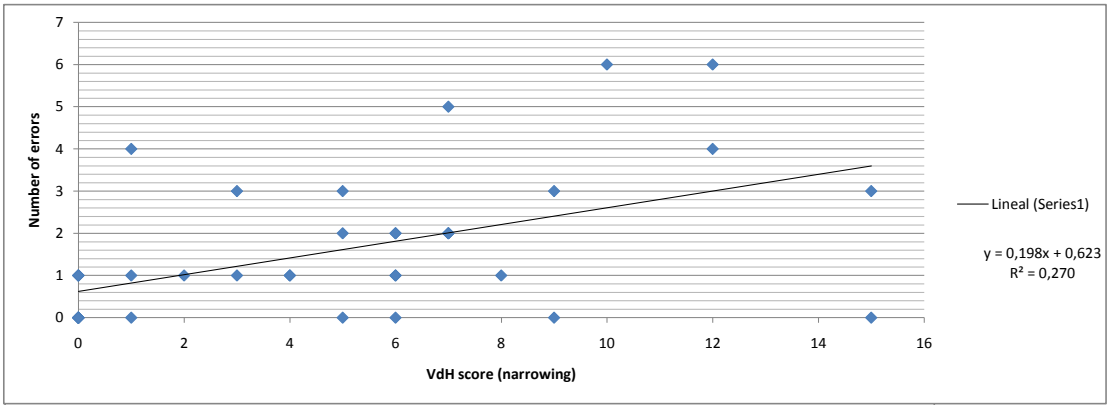


Figure 27: Joint characterization error versus VdH score

a promising relation between some of the proposed measures and VdH score, especially for PIP joints. The absolute average and minimum distances in PIP joints present the expected behaviour, decreasing as the VdH score increases but for the joints with VdH score 3. Joints with a score of 3 deviate from this tendency. The anomalous behaviour for VdH 3 box might be attributed to a low number of samples. Since there are only 7 valid joints labelled with VdH score of 3 no solid conclusions can be inferred. For MFC joints the correlation between absolute measures and VdH scores is not so clear. Average and minimum distances show unexpected results with no clear tendency to decrease as score increases. We suspect that this is due to a bad computation of the triangular area used to restrain the measures. The only absolute score preserving an encouraging tendency is the standard deviation. Again, the box corresponding to VdH 3 should be excluded because of the low number of samples.

Concerning relative scores, the average relative distance does not present a monotonous tendency, probably due to a wrong formulation based on a normalization by the maximum distance along the joint. We consider that a constant maximum width set for each joint type should significantly improve the results. The coverage shows a great characterization of healthy joints and a decreasing tendency in average as score increases.

6.3 Future lines of research

Improvements can be made to the presented method, that also opens new lines of work and opportunities.

- **Improvements on the skeleton**. Either a new skeletonization algorithm or a new pruning algorithm may improve the detection ratios if we get better skeletons.
- **Improve measurement on damaged joints**. Our model is not valid with this articulations.
- **Extend the model to the feet**. Only six articulations on the feet are considered, the five metatarso-phalangical ones and the distal of big toe. The method would require minor fixings in order to work with the foot, and mainly on the segmentation of the foot bones.
- **Carpals**. The carpals should be detected and segmented to perform also measurements on them. However, their complex structure needs a completely different approach.
- **Erosion measurements**. A problem not tackled yet by the medical imaging community.
- **From scores to Van der Heijde**. Our results indicate correlation between our scores and the VdH scores, but more work is required in order to find the optimum measurements that allow us to translate measurements to Van der Heijde Scores.

7. Acknowledgements

We would like to thank Dr Sara Marsal from the Unitat d'investigació de Reumatologia of the Hospital de la Vall d'Hebron and Dr Jaume Maymó from the Hospital del Mar, for their expertise and for providing us with images to develop this master thesis. Special thanks also to Fernando Vilariño for his help, and to Debora Gil for his guidance and patience.

References

Jane Angwin, Geoff Heald, Andrew Lloyd, Kate Howland, Maria Davy, , and Michael F. James. Reliability and sensitivity of joint space measurements in hand radiographs using computerized image analysis. *The Journal of Rheumatology*, 28(8):1825–1836, 2001.

Andrzej Bielecki, Mariusz Korkosz, and Bartosz Zieliński. Hand radiographs preprocessing, image representation in the finger regions and joint space width measurements for image interpretation. *Pattern Recogn.*, 41(12):3786–3798, 2008. ISSN 0031-3203. doi: http://dx.doi.org/10.1016/j.patcog.2008.05.032.

R. de Luis, M. Martin-Fernandez, M. Martin, and C. Alberola-Lopez. A fully automatic algorithm for contour detection of bones in hand radiographs using active contours. In *Proceedings of the IEEE International Conference on Image Processing (ICIP'03)*, volume III, pages 421–424, Barcelona, Spain, September 14–17 2003.

Angélica Vargas Guerrero and Carlos Pineda Villaseñor. Evaluación radiográfica del daño anatómico en la artritis reumatoide. *REVISTA COLOMBIANA DE REUMATOLOGÍA*, 13(3):214–227, 2006.

Chin-Chuan Han, Chang-Hsing Lee, and Wen-Li Peng. Hand radiograph image segmentation using a coarse-to-fine strategy. *Pattern Recogn.*, 40(11):2994–3004, 2007. ISSN 0031-3203. doi: http://dx.doi.org/10.1016/j.patcog.2007.01.010.

Chi Wen Hsien, Tai Lang Jong, Tiu, and Chui Mei. Bone age estimation based on phalanx information with fuzzy constrain of carpals. *Med Biol Eng Comput*, 45(3):283–95, 2007. ISSN 0140-0118.

T Jensen and M Klarlund. Bone loss in unclassified polyarthritis and early rheumatoid arthritis is better detected by digital x ray radiogrammetry than dual x ray absorptiometry: relationship with disease activity and radiographic outcome. *Annals of the Rheumatic Diseases*, 63(11):15–22, 2004.

Tony Lindeberg. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *International Journal of Computer Vision*, 11(3):283–318, 1993.

Antonio López, Felipe Lumbreras, Joan Serrat, and Juan J. Villanueva. Evaluation of methods for ridge and valley detection. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 21(4):327–335, 1999.

M. Niemeijer, B. van Ginneken, C. Maas, F.J.A. Beek, and M.A. Viergever. Assessing the Skeletal Age From a Hand Radiograph: Automating the Tanner-Whitehouse Method. In M. Sonka and J.M. Fitzpatrick, editors, *SPIE Medical Imaging*, volume 5032, pages 1197–1205. SPIE, SPIE, 2003.

P. Peloschek, G. Langs, M. Weber, J. Sailer, M. Reisegger, H. Imhof, H. Bischof, and F. Kainberger. An automatic model-based system for joint space measurements on hand radiographs: initial experience. *Radiology*, 245(3):855–862, 2007.

John T. Sharp, Jill C. Gardner, and Earl M. Bennett. Computer-based methods for measuring joint space and estimating erosion volume in the finger and wrist joints of patients with rheumatoid arthritis. *Arthritis and Rheumatism*, 46(6):1378–1386, 2000.

D. van der Heijde. How to read radiographs according to the sharp/van der heijde method. *The Journal of rheumatology*, 26(3):734–745, 1999.

Bartosz Zieliński. A fully-automated algorithm dedicated to computing metacarpophalangeal and interphalangeal joint cavity widths. *Schedae Informaticae*, 16:47–67, 2007.