Universitat Autònoma de Barcelona

Master in Computer Vision and Artificial Intelligence

Report of the Master Project

Option: Computer Vision

Computer vision techniques for characterization of finger joints in X-ray images

Author: **Joan M. Núñez Do Rio**

Advisor: **Dr. Fernando Vilariño and Dra. Debora Gil**

Acknowledgements

I would like to express my gratitude to my supervisors F. Vilariño and D. Gil who has guided me through this research. I would also like to thank Jorge, Carles, Camp and Albert for their support and advice as well as other collegues in CVC who make the working task more comfortable.
També vull agraïr la meva familia el seu constant support, especialment els meus pares. I no oblidar, clar, a tot els meus amics que sempre ajuden a que tots els obstacles semblin més petits.
Thanks a lot - Moltes gràcies

ABSTRACT

Rheumatoid arthritis (RA) is an autoimmune inflammatory type of arthritis which mainly affects hands on its first stages. Though it is a chronic disease and there is no cure for it, treatments require an accurate assessment of illness evolution. Such assessment is based on evaluation of hand X-ray images by using one of the several available semi-quantitative methods. This task requires highly trained medical personnel. That is why the automation of the assessment would allow professionals to save time and effort. Two stages are involved in this task. Firstly, the joint detection, afterwards, the joint characterization. Unlike the little existing previous work, this contribution clearly separates those two stages and sets the foundations of a modular assessment system focusing on the characterization stage. A hand joint dataset is created and an accurate data analysis is achieved in order to identify relevant features. Since the sclerosis and the lower bone were decided to be the most important features, different computer vision techniques were used in order to develop a detector system for both of them. Joint space width measures are provided and their correlation with Sharp-Van der Heijde is verified.

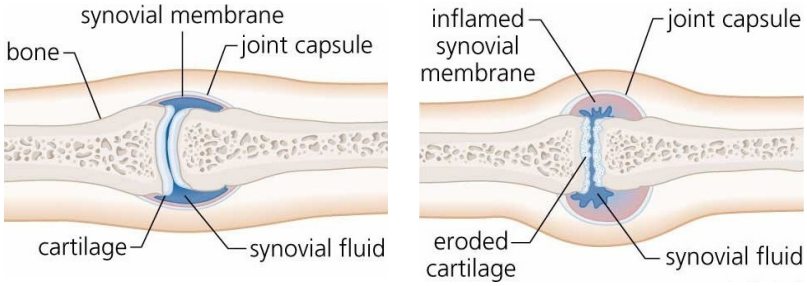
Keywords: *Rheumatoid arthritis, X-ray, Sharp Van der Heijde, joint characterization, sclerosis detection, bone detection, edge, ridge*

Contents

Chapter 1

Introduction

Rheumatoid arthritis (RA) is a form of arthritis that causes pain, swelling, stiffness and loss of function in joints. This disease is present in patients aged between 30 and 50 years [1]. Though it is thought that both genetics and environmental factors play an important role in RA affliction and severity, it is a chronic disease whose causes still remain unknown and there is not cure for it. It is widely assumed that RA is an autoimmune inflammatory type of arthritis which is triggered by the attack of immune system cells to the joints, usually synovial joints that are found between articulated bones. Cavities between bones in synovial joints are filled with the synovial fluid, encapsulated by the synovial membrane or *synovium* (Figure 1.1). Some kind of autoimmune or infectious triggering event would cause an inflammation in the *synovium*.



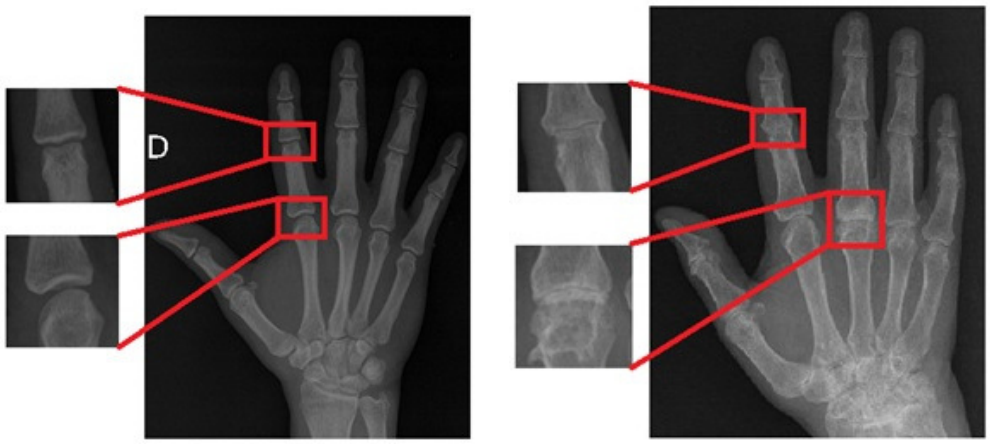
(a) (b)

Figure 1.1: (a) Healthy joint. (b) Joint affected by RA.[1]

An autoimmune or infectious triggering event may cause an inflammation in the synovium. As a consequence, swelling, joint space narrowing (JSN) and mobility reduction would affect that joint, which consequently would produce pain and damage to cartilage and bone (Figure 1.2). More severe profiles

[1] http://images.yourdictionary.com/rheumatoid-arthritis

may present bone luxation (*ankylosis*) making the joint lose its mobility or even, in more advanced stages, the disease may affect other organs and cause death.



(a) (b)

Figure 1.2: X-ray image examples (a) Healthy joints. (b) Joints affected by RA.

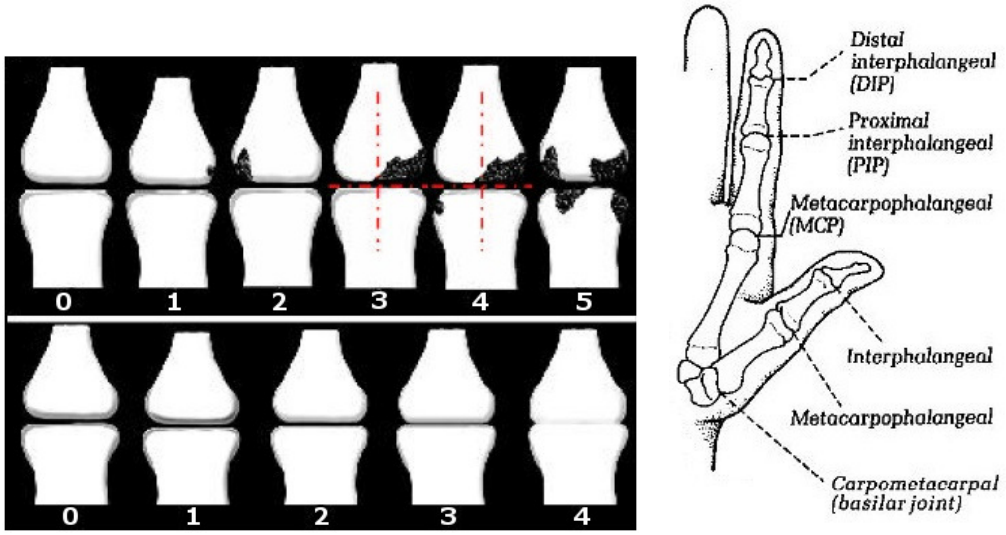
As RA is a chronic disease and there is not any cure for it, the significant deterioration produced in the patient joints can affect deeply their quality of life. Current treatments weight their utility by the success in avoiding the disease to reach advanced stages. That success can only be achieved if the therapeutic treatment is correctly planned and suits the real stage of the disease. Furthermore, research on new treatments also need accurate assessments of disease severity. Therefore, a precise assessment of illness evolution is required.

Since the joint damage produced by RA can be detected in feet and hands in early stages, several imaging methods have been implemented in order to support treatments and evaluate the deterioration of the joints [2]. Although ultrasound and magnetic resonance have been used lately for the assessment in early stages, the easier access to radiological images and its relative lower cost have made it the most widely used technique.

X-ray images expose some drawbacks such as the absence of normalized techniques to obtain them, the loss of 3D information, the lack of sensitivity against earlier non-erosive changes or the cartilage invisibility. Nevertheless, their simplicity and the fact that they allow the analysis of joint space width (JSW) and bone erosion made X-ray images become the main imaging technique in RA degree assessment. That reason has prompted the appearance of many X-ray semi-quantitative evaluation methods along the years [3, 4].

Among existing methods, Van der Heijde modification (SvdH) of the method described by Sharp in 1971 is the one most widely used nowadays [5, 6]. This method is semi-quantitative, which means that scorings are not measured straight from the image. The final scoring is based on the examination of a group of determined joints so that it represents the RA affliction of the patient. It provides separate

discrete scorings for JSW and bone erosion to every considered joint in feet and hands. The erosion scoring is provided to 16 joints in every hand and to 6 joints in every foot. Its value goes from 0 to 5 depending on the amount of bone surface affected by erosion. 0 value corresponds to a non-eroded joint and a value equal to 5 corresponds to a completely collapsed joint. The JSN scoring is provided to 15 joints in every hand and to 6 joints in every foot and its value goes from 0 to 4. In this case, 0 value corresponds to a normal joint and a value equal to 4 corresponds to complete luxation. It all means that SvdH final scoring can have a maximum value of 448. Figure 1.3a illustrates erosion and JSN scorings.



(a) (b)

Figure 1.3: (a) SvdH scorings [7], bone erosion and joint space width. (b) Hand joints.[2].

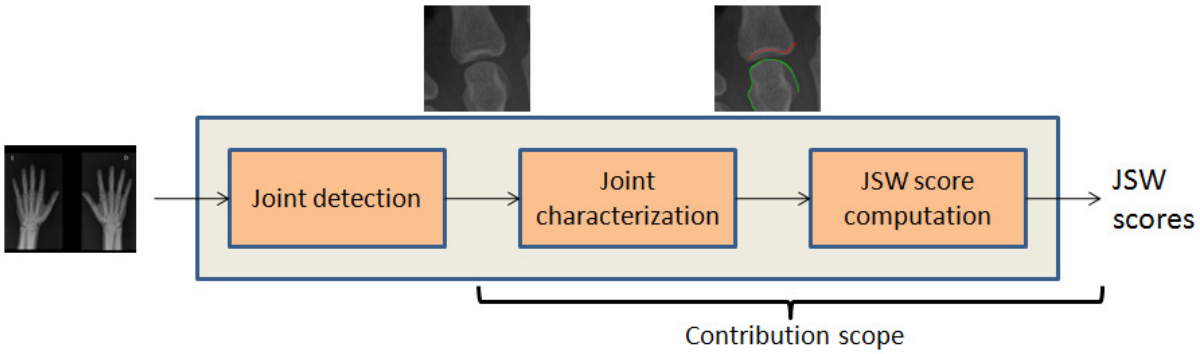
The evaluation of X-ray images in order to provide the SvdH scoring is high time consuming and requires highly trained medical personnel. Furthermore, the reliability of the scoring methods is always subject to intra and inter observer reproducibility. That is why the automation of the assessment would allow professionals to save time and effort as well as providing them with the ability of detecting small variations to give more accurate evaluations that could not be achieved by semi-quantitative methods.

Previous works have made the attempt to implement semi-automatic methods in order to solve the time cost and normalization problems but they still needed of trained personnel to provide the algorithm with the demanded inputs.

Some previous attempts have tried to develop semi-automatic method in order to extract information related to hand joints [8, 9, 10, 11]. However, few examples of attempts to develop fully automatic systems have been developed. The implementation of such algorithm involves two steps. There should be a detection module in order to extract the joints from the X-ray images so that they could be provided

[2]American Society for Surgery of the Hand (http://www.assh.org)

to a second module which would analyse the joints in order to characterize them. Afterwards, the features obtained by that module would be passed to the last stage which would provide measures correlated to the degree of RA affliction (Figure 1.4). This correlation would be represented as a correlation with the already mentioned SvdH scoring. In our case, the measures will be focused on JSW aproximations.



(a)

Figure 1.4: Hand RA assessment system based on JSW.

Zielinski (2007) and Bielecki (2008) [12, 13] tried to face the challenge of a fully automatic system for the first time. Vera (2010) [7] provided a new method improving significantly the joint detection rates. Both works tried to face the three stages altogether. Vera improved the hand bone marrow computation in order to provide more accurate joint detections. However, the joint description and modelling based on JSW computation should be improved in order to obtain measures of the patient degree of RA. This work failed in its attempt to provide with correlated measures to SvdH scorings. Moreover Vera methods need to be parametrized in order to establish the boundaries of the measuring area and an important improvement should avoid that limitation. New JSW measure measurements which verify a clear correlation to SvdH are needed.

Our work is focused on the characterization of the hand joints in order to provide relevant features. Those features will be used to compute reliable measurements that show a clear correlation with SvdH scoring. That task implies the creation of hand joint models which will lead as to identify the most important features for the characterization. Two main features were identified, the sclerosis and the lower bone contour. Afterwards, different tools were tested in order to provide sclerosis and lower bone detectors. Furthermore, a sclerosis and lower bone contour ground truths where created as well as the necessary performance metrics to evaluate the different tools. Eventually, the better sclerosis and lower bone detectors will be used to test different distance measurements and its correlation to SvdH score will be tested.

Metacarpo-phalangeal joints (MCP), proximal inter-phalangeal joints (PIP) and distal inter-phalangeal joints (DIP) (Figure 1.3b) are considered although the correlation to SvdH can only be checked for MCP and PIP since the others are not considered by SvdH method.

Chapter 2

Data analysis

2.1 Dataset

This study is focused on the same digital dataset used by Vera (2010) [7]. It provides us with DICOM format digitalized images from 20 patients. Both hands are provided in posterioanterior view. Since there are 14 joints in every hand, the whole dataset consists of 560 joint image samples. Sharp/Van der Heijde scoring method does not take into account all the hand joints so the SvdH scoring method is obviously provided for those who are scored. It means the score is provided for the four PIP and the five MCP joints in every hand. Therefore, a total scored dataset of 360 samples is available.

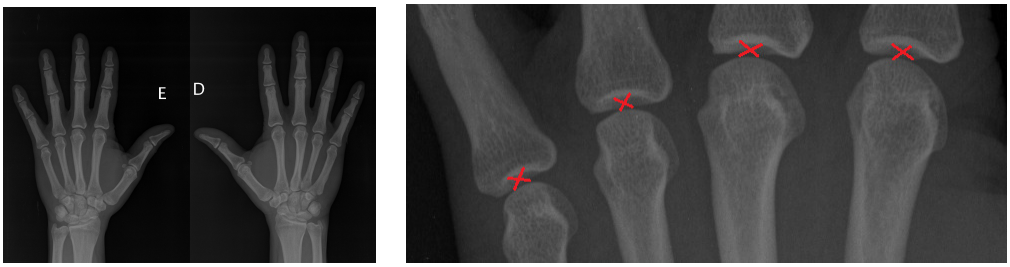
The original DICOM format images resolution is 2828x2320 and the intensity ranges from 0 to 4095 (12 bit image). In order to work with these images they had to be preprocessed to convert them into tif format images. That preprocessing also included a cropping of the images in order to remove the radius and the ulna bones present in some images and also putting both hand in the same frame.

After these steps, 20 pair of preprocessed hand images are available (Figure 2.1a). Since joint characterization is the first aim of this study, the final dataset must contain just joint samples. Those 20 images were marked by a single user who was requested to spot the position of every joint in a pre-defined order. A requested joint position is defined by the coordinates of a pixel along the corresponding finger axis which is found in the middle of the joint interspace (Figure 2.1b).

The user was requested to go through every patient's pair of hands images in a pre-defined order. The joints had to be spot from the thumb to the little finger starting from MCP and finishing with DIP. Left hand was the first to be marked for each patient.

Once the joint positions had been obtained all the images were processed again in order to obtain the joint samples. Every joint coordinate was used to crop the corresponding joint sample. Since the size of the patient images could vary, the cropping was done such as the final joint sample size could vary too. The joint samples were cropped at a size relative to the size of the original patient image size so that that the final sample center pixel was the one that the user marked as the joint position. It means

that in the final dataset the center pixel of each image is the joint position. The size of the samples range from 140x140 to 200x200.



(a) (b)

Figure 2.1: (a) Dataset patient example. (b) Joint positioning process.

In the final joint dataset each joint is tagged with its patient identifier, joint identifier, finger number, joint name and, if available, SvdH scoring. All the right hand joint sample were mirrored to only work with left joint samples.

2.2 Joint modelling

As the object of our study, hand joints must be the focus of a detailed examination and an accurate modelling. It has been previously exposed that our final dataset includes 560 joint images. Since there are 14 different joints in every hand (five MCP joints, four PIP joints, four DIP and one IP) there are 40 images for every joint type. It must be remarked that at this point it does not make any difference for us whether it is a left hand joint or a right hand joint. Therefore, for every joint type we have the left and the right sample from every of the 20 patients who are part of our dataset. The right joint samples have been mirrored to work always with left joint samples.

In order to make the task of observing the samples easier and to have a clear vision of the whole sample set, the different joint samples have been separated by its type. Every finger has three joints (MCP, PIP, DIP) except the thumb which has two joints (MCP and IP). Those 14 types of joints must be examined in order to decide whether a category or model should be developed for each of them.

The first approach to that categorization was based on the creation of a poster including the whole dataset samples organized by finger and type of joints (Figure 2.2). The poster dimensions are 140 x 100 cm and it allowed us to place 3.2 x 3.2 cm joint samples. Every column corresponds to a determined finger from the thumb (F1) to the little finger (F5). The different types of joints are organized in three rows with the exception of the thumb, since we already mentioned it only has two joints. That is why the thumb IP samples are placed in the middle of the first and the second rows. The caption of every sample includes the dataset joint index, the SvdH scoring and indicates if it is a left or right sample. The caption color is different depending on the sample SvdH scoring in order to make it easier to spot

the RA affection of every joint.

The first conclusion after this visual observation of the whole organized dataset is a clear difference between the thumb and the other fingers. Besides the anatomical differences, there is a clear impact on the samples because of the way the X-ray images have been taken. The image is taken from the hand back with the palm facing down. As a consequence of this posterioanterior X-ray view a profile view of the thumb is taken and features that can be found on the samples are significantly different (Figure 2.3). That is why the thumb was considered as a separate problem and was not included as part of this study which will therefore consider the other four fingers, each of them with theree joint (Figure 2.4).

This study should allow us to develop accurate models of the hand joints which should lead us to achieve a proper automatic feature detection in order to obtain a numeric and standard characterization of the joint RA affection. Therefore, our analysis of the finger joints has been strictly focused on the computer vision features that can be observed in the joint samples. The anatomical shape of a joint and its components and the vision of the experts is also taken into account.

Three main regions have been distinguished by the examination of joint samples: sclerosis and joint interspace, upper bone and lower bone. The upper bone and the lower bone are the bones which are linked by the joint to whom it allows movement. The sclerosis appears as a consequence of the shape of the joint bones and the properties of the X-ray images. As it has been mentioned before, the intensity level observed in a region is caused by the density of the structures which the X-ray beam had to pass through. Since the joint interspace is filled by cartilage it becomes a darker region in an X-ray image, which means that there is an intensity valley between the two bones. Those two bones conform the joint defining a concavity (Figure 2.5). It means that in the upper side of that concavity (upper bone) the X-ray beam must go through a denser bone structure which results in a brighter area in the X-ray image. It helps the medical personnel to ascertain the real joint distance in spite of the mentioned concavity. It will also help us to define the upper boundary of the joint interspace.

2.2.1 Metacarpo-phalangeal joints

MCP joints attach the base of the finger to the palm of the hand, which means that it is the joint between the second phalanges and the metacarpal bones. They are the largest joints in the hand (Figure 2.6).

The upper bone shape boundaries are described by an intensity contour all around its shape. The contours are clear as the intensity values in the bone side of the edges is significantly different from the intensity values outside the bone edges. As it has been mentioned before it is caused by the fact that the X-ray beam outside the bone edges just have to pass through skin and muscles, which are softer body tissues and result on darker regions. The shape described by the edges is remarkably similar among all the samples. An occlusion is usual in its lower part as part of the joint concavity remains hidden behind the lower bone.

The lower bone has four main features recognised as intensity edges. Its shapes and intensity values are not constant and can vary among different samples. Two of those edges describe the outer shape of

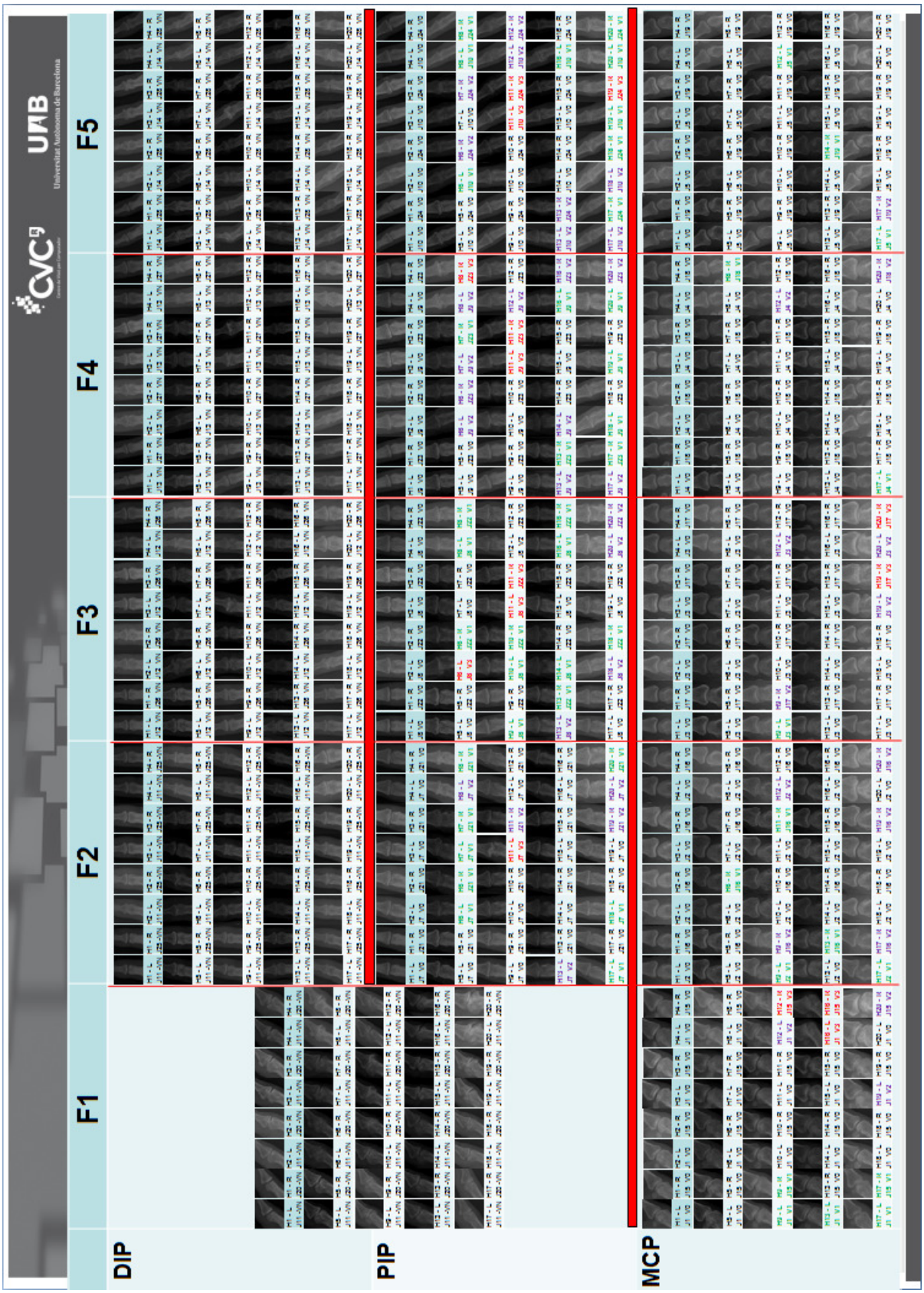
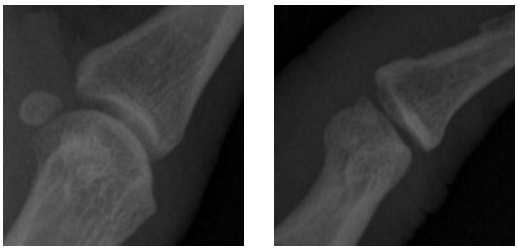


Figure 2.2: Joint sample poster (landscape view).



(a) (b)

Figure 2.3: Thumb joint examples (a) MFC. (b) IP.

the bone and are linked in the joint interspace. Its intensity values are clearly related to the presence of the two inner edges. Usually, the clearer the inner edges are the less clear the outer edges are. It is also important the fact that any of these four edges may not be present in a certain sample. It all results on intensity variability which may lead us to wrong detections on the outer shape of the lower bone.

At this point it can be already assumed that a contour detector will be suitable to extract the described edges.

The sclerosis appears as a brighter area in the lower part of the upper bone. It defines a ridge which clearly follows the concavity described by the bone shape. Its thickness vary being thicker in the middle and becoming narrower at both ends. Although it stretches to the right side of the upper bone, the intensity drops when it gets closer to the left side of the bone and the ridge becomes less clear. The fact that the upper bone is partially occluded in the lower part results on a joint interspace valley with different intensity values from the regions where the X-ray beam did not go through any bone. It means that bone intensity values can be found in the joint interspace. In this case a ridge detector will be needed in order to obtain the sclerosi segmentation.

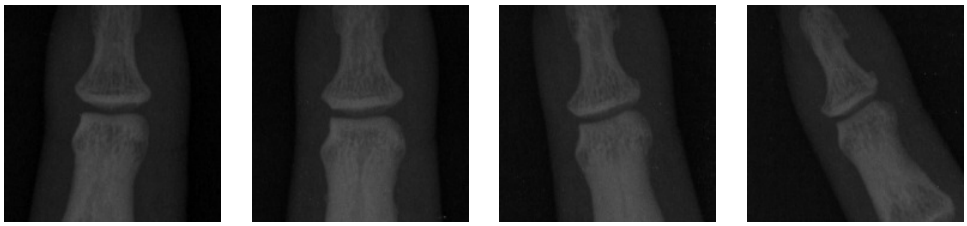
The differences between the fingers are only related to scale and a tendency of the right outer edge of the lower bone to become less clear.

2.2.2 Proximal inter-phalangeal joints

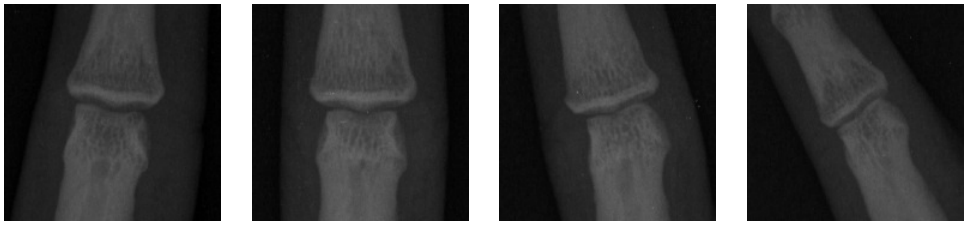
PIP joints are found between proximal and middle phalanges in fingers from second to fifth (Figure 2.7).

The upper bone shape is described by intensity edges. The shape is remarkably different to other joint type upper bones and remains similar among the PIP samples. The contours that describe the shape are clear and visible in the sides and become thicker in the upper parts. In the low part of the bone an occlusion is found again since a part of the upper bone is hidden behind the lower bone.

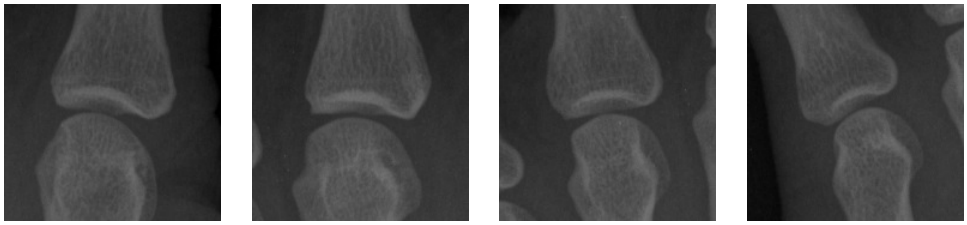
The lower bone shape is defined by two outer intensity edges and, as it happens with MCP, there are two inner edges. There is a difference, though, as in this case the two inner edges are always present. However, the outer edges are not clear in some cases and may make it harder to detect the outer shape



(a) Finger 2 - DIP (b) Finger 3 - DIP (c) Finger 4 - DIP (d) Finger 5 - DIP



(e) Finger 2 - PIP (f) Finger 3 - PIP (g) Finger 4 - PIP (h) Finger 5 - PIP



(i) Finger 2 - MCP (j) Finger 3 - MCP (k) Finger 4 - MCP (l) Finger 5 - MCP

Figure 2.4: Finger joint examples.

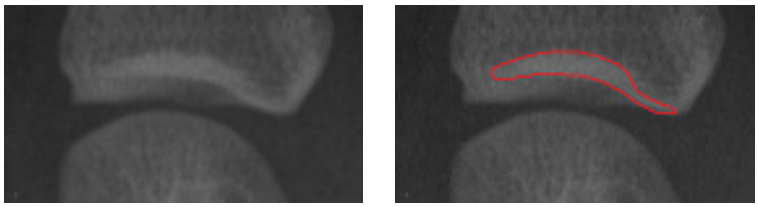
of the bone.

The sclerosis described by the brighter area in the lower part of the upper bone is clear. It is significantly thick and its constant thickness describes a clear beak-shaped ridge all along the lower part of the upper bone. Since there exists an occlusion of the upper bone we found again higher intensity values in the joint interspace that in regions with no bone presence. As a consequence the valley which it describes is not that important but still visible.

The joint feature study again leads us to the use of ridge and edge detectors.

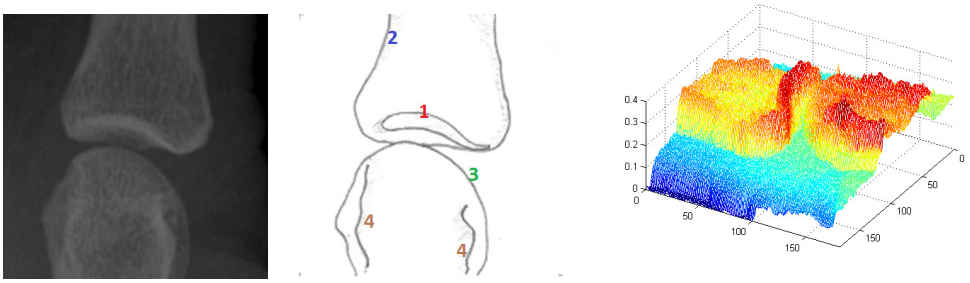
Besides all these features there is an important symmetry with respect to the finger axis. It is clear in the second phalange and remains important in the third one. It disappear in the fourth and fifth phalanges.

As well, the scale decreases from second finger to fifth and increases the tendency to fuse the lower bone edges. If it happens the lower bone shape becomes clear. Nevertheless, it is not a rule and can not be assumed as a general statement.



(a) (b)

Figure 2.5: Sclerosis detail



(a) (b) (c)

Figure 2.6: MCP model (a) Joint example. (b) Joint model schema (1: sclerosis, 2: upper bone, 3: lower bone, 4: inner edges). (c) Image 3D mesh.

2.2.3 Distal inter-phalangeal joints

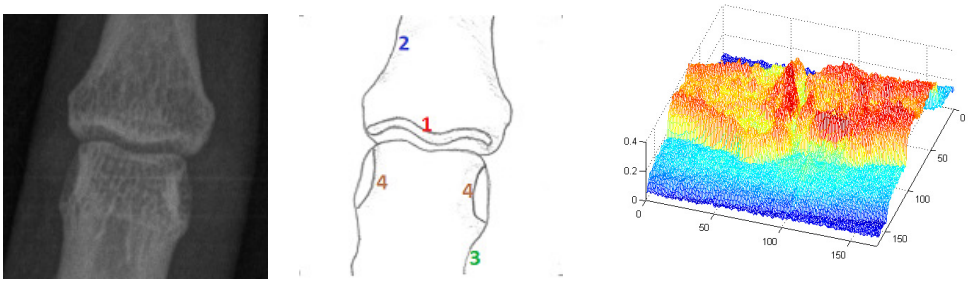
DIP joints are the smallest finger joints though their size is close to PIP joints (Figure 2.8). They are found between first and second phalanges in second, third, fourth and fifth fingers.

The upper bone shape boundaries are described by edges in the lower parts. The shape boundaries become less visible in the upper parts where it stretches at the same time the edges becomes less clear.

The lower bone shape is defined by remarkably stable intensity level edges. The intensity levels may be higher in the upper parts.

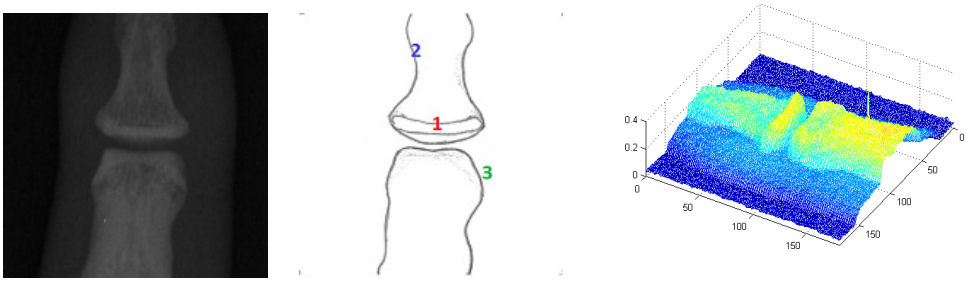
The sclerosis is again a ridge that is located in the lower part of the upper bone. There are not remarkable tendencies about its thickness. It means it can go from thick to narrow, never as thick as other sclerosis though. However, the ridge tends to be visible and almost straight with a variable degree of convexity, the same shape as the one described by the joint. As there is not upper bone occlusion the joint interspace is featured as a valley between the two bone shapes.

In spite of the mentioned differences between the three models, all of the them prompt us to use contour and ridge detectors.



(a) (b) (c)

Figure 2.7: PIP model (a) Joint example. (b) Joint model schema (1: sclerosis, 2: upper bone, 3: lower bone, 4: inner edges). (c) Image 3D mesh.



(a) (b) (c)

Figure 2.8: DIP model (a) Joint example. (b) Joint model schema (1: sclerosis, 2: upper bone, 3: lower bone). (c) Image 3D mesh.

Chapter 3

Methodology

The previously pre-defined models prompt us to use several computer vision tools. These tools should allow us to detect contours and ridges. This detection should include some features described in the model in order to allow us to characterize the joint. Eventually, that characterization would lead us to achieve measurements to score the arthritis affection (Figure 3.1).

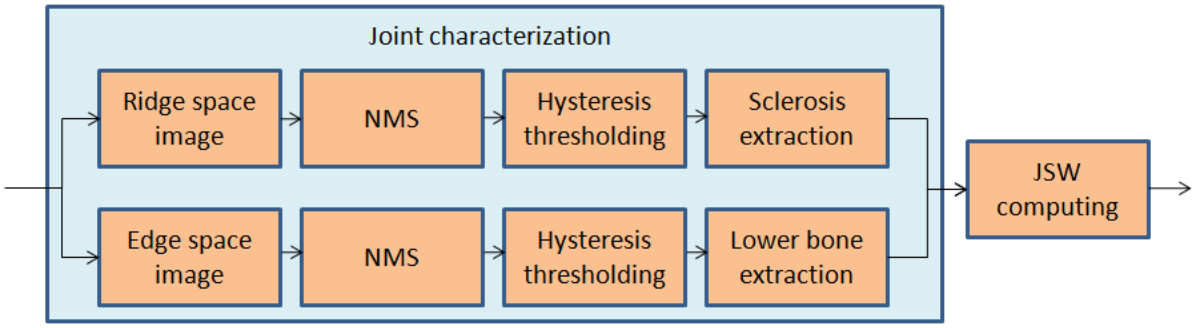


Figure 3.1: Characterization and JSW measure block diagram.

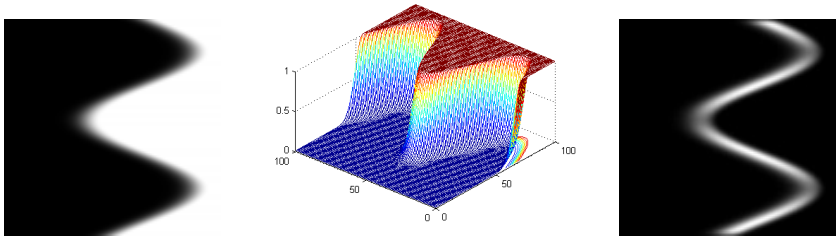
3.1 Feature space image

The basic ridge or edge detector goes through different steps from the original image to the final thin lines which define the corresponding ridges or edges. As a first step, the input image must be processed in order to obtain a feature space image. The intensity levels of that image should describe the chances of finding a ridge or edge in every pixel. It means that feature space image will be a gray level image with the same size as the input image.

A studied feature must be characterized in order to obtain the corresponding feature space image. The geometric feature represented by a concept must be established prior to its modelling and its

mathematical formalization. That formalization will determine the properties of the obtained feature space image.

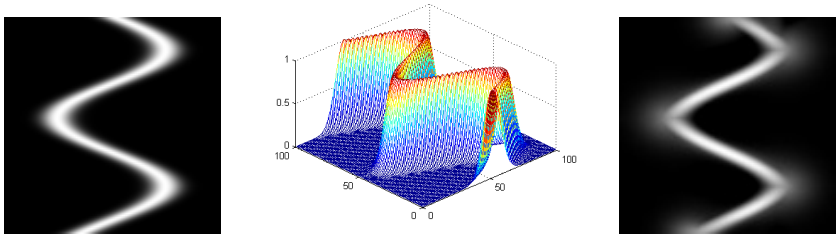
As long as edges are concerned, the intuitive concept can be described as an intensity level change between two remarkably flat regions (Figures 3.2a, 3.2b). That notion can be formalized as the image regions where the first derivative reaches both maximums and minimums. It means that the gradient module can be used as the feature space image (Figure 3.2c).



(a) (b) (c)

Figure 3.2: Example 1: edges (a) Image. (b) 3D mesh. (c) Edge space image (gradient).

The notion of ridge can be stated in several ways. A ridge can be described as the continuous line defined by the intersection of two surfaces. Those surfaces can have different degrees of irregularity (Figures 3.3a, 3.3b). Again this notion must be formalized. A ridge can be obtained as the high values in the second derivative. Following the reasoning exposed when talking about edges, in some way the ridges could be considered as the edges in the first derivative.



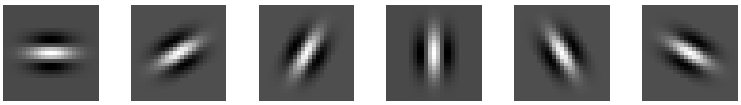
(a) (b) (c)

Figure 3.3: Example 2: ridges (a) Image. (b) 3D mesh. (c) Ridge space image (2nd derivative).

Therefore, the second derivative must be obtained as the ridge space image. This image is obtained applying to the input image a bank of filters. Each filter achieves a convolution with a second derivative of an anisotropic Gaussian with a different orientation (Figure 3.4). The output of this bank of filters will be composed as the image which takes for each pixel the maximum value of all filters outputs (Figure 3.3c).

In order to provide joint orientation to the system the method for bone marrow computation ex-

posed by Vera will be tested [7]. This method computes bone marrow as the skeleton of a previous segmentation of hand bones. Afterwords, the joint orientation can be obtained from the corresponding finger orientation.



(a) (b) (c) (d) (e) (f)

Figure 3.4: Second derivative of anisotropic gausian (6 orientations).

A different notion of ridges and valleys exposed by Lopez [14] will be also tested. This is a multilocal approach which develops a measure of *ridgeness* based on level set curves. Ridges and valleys are understood as ramified channels over a surface. The method computes level set curves over the image and looks for determined patterns. This method is highly sensitive to ridge patterns even though they have very different scales.

3.2 Non-Maximum Suppresion algorithm

As mentioned above, a feature space image is a gray level image that contains information about the chance of finding the considered feature in each pixel. That image could be directly binarized in order to obtain a final feature image. However, it may lead us to an output containing thick regions around the local maxima instead of thin and accurate edges or ridges as required. The goal of the next processing tool should be removing non-maxima pixels. That is why Non-Maximum Suppression (NMS) algorithm becomes useful.

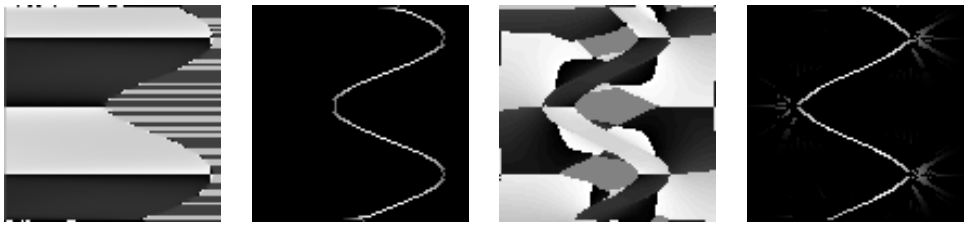
Let $I(x, y)$ be the intensity level of the pixel (x, y) in the feature space image I, where x and y are positive integers not higher than the image size for each dimension. NMS checks its neighbours just in the perpendicular directions to the corresponding edge or ridge orientation. Nearest neighbour interpolation is achieved if needed. If the initial pixel value $I(x, y)$ is a maxima it will conserve its value in the output non-maximum suppressed image $NMS(x, y)$. Otherwise, if the initial pixel value $I(x, y)$ is not a maxima its value will be set to zero in the output image:

$$NMS(x,y) = \begin{cases} I(x,y) & \theta(x_1,y_1) \leq I(x,y) \wedge \theta(x_2,y_2) < I(x,y) \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

where $\theta(x_1, y_1)$ and $\theta(x_2, y_2)$ are the intensity values of the neighbours in the perpendicular directions to the feature orientation.

In order to apply NMS algorithm to the previously exposed edge and ridge space images the orientation images are necessary. In the former case the orientation image is the gradient phase image. In the latter, the orientation for each pixel is the orientation of the determined filter from the bank of filters

which provided the output with the highest pixel value. It means the number of filters in the bank of filters will provide the orientation image resolution.

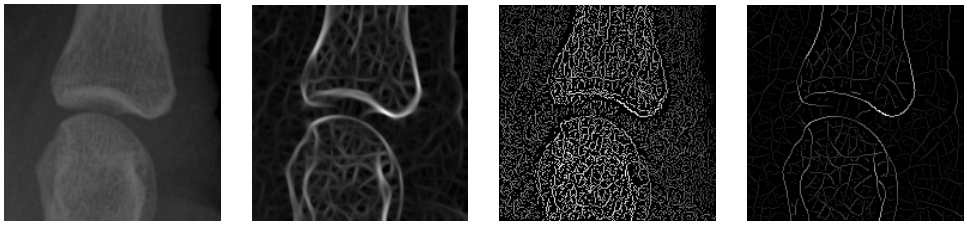


(a) (b) (c) (d)

Figure 3.5: (a) Example 1 orientations (b) NMS on example 1. (a) Example 2 orientations. (b) NMS on example 2.

The non-maximum suppressed image will contain gray level values and thin lines (Figures 3.5b, 3.5d). It tries to preserve the connectivity of the features by using information provided by the orientations. However, the lack of precision in the orientation image can cause lost of connectivity in the non-maximum suppressed image (Figure 3.5).

It must be remarked that the example followed to show edge and ridge processing only pursues to show the main steps in the process. However, it allow us to show a new issue. According to this example it may be thought that NMS algorithm should be applied in the Example 2 right to the input image and it would provide better results. Figure 3.6 shows that it is not true in more complicated problems.



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

Figure 3.6: (a) Input image. (b) Ridge feature image. (c) NMS on input. (d) NMS on ridge feature image.

3.3 Hysteresis thresholding

Since a non-maximum suppressed image is still a gray level image it should be binarized in order to obtain the final requested features. Nevertheless, in the non-maximum suppressed image there is a remarkable presence of noise, as observed in Figure 3.5. Hysteresis thresholding allows us to obtain a

binary image from the NMS output image preserving the connectivity of the features but removing weak responses (Figure 3.7).

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

Figure 3.7: Hysteresis thresholded examples (a) Example 1 (Fig. 3.2). (b) Example 1 over input. (c) Example 2 (Fig. 3.3). (d) Example 2 over input.

Hysteresis thresholding uses two thresholds τ_1 and τ_2 so that $\tau_1 > \tau_2$. Let $NMS(x, y)$ be the intensity value of a pixel (x, y) in the non-maximum suppressed image. If $NMS(x, y) > \tau_1$, then the pixel (x, y) will be forwarded to the output. If $NMS(x, y) < \tau_2$, the pixel (x, y) will be discarded. Otherwise, the pixel will be forwarded only if a pixel (x', y') so that $NMS(x', y') > \tau_1$ can be reached following the feature contour in any of both directions as long as $NMS(x_i, y_i) < \tau_2$ for every of the pixels (x_i, y_i) followed in the contour.

3.4 Feature extraction

The basic detector introduced by this study would include the steps mentioned above. As a first approach, the sclerosis and the lower bone will be the two necessary main features for any arthritis assessment system. Figure 3.8 and Figure 3.9 show edge and ridge examples of the previous steps.

Once the most accurate detection of sclerosis and lower bone have been obtained, those images must be processed one last time. That last processing achieves the final extraction of the ridge which represents the desired sclerosis or the edge which represents the lower bone.

In order to extract the sclerosis, the system chooses the longest ridge in a defined region in the center of the image. The lower bone is obtained by selecting the longest edge in a region around the lowest part of the image. The edge must also hold the higher intensity mean.

3.5 Denoising filters

It is usual to carry out a previous smoothing in Computer Vision before applying other tools. Since we aim to obtain different feature detectors, it seems the best option will be to use an smoothing tool able to preserve those features. Many options have been exposed. In this section a brief introduction to some of those methods is provided just in case they appear to be useful. The denoising module previous

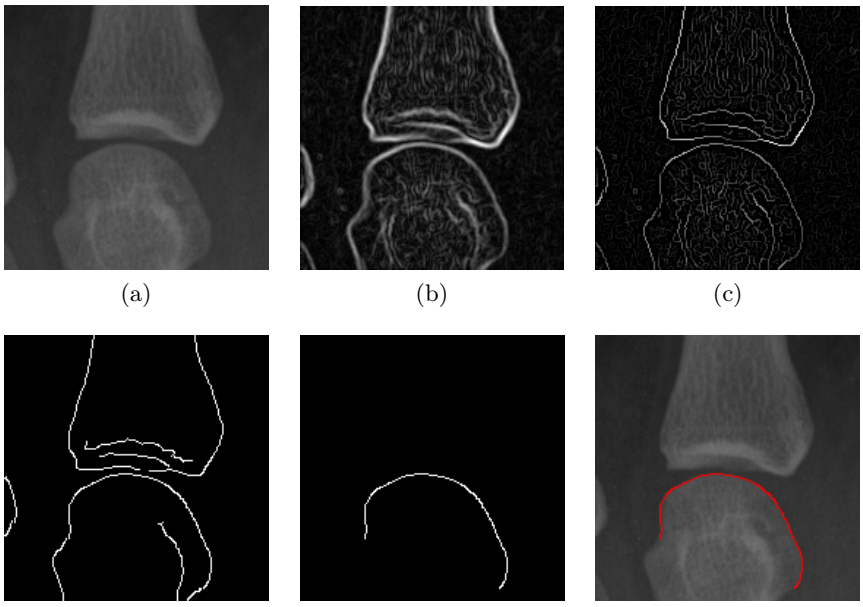


Figure 3.8: Edge detection (a) Input image. (b) Edge space. (c) NMS. (d) Hysteresis thresholded. (e) Lower bone edge extracted. (f) Final detection over input

to the feature image computing will be checked in some cases in order to assess its contribution to a better performance of the detectors.

3.5.1 Structure-preserving diffusion

An important group of feature-preserving filters are based on differential operators. These tools apply iterative methods that use partial differential equations (PDEs). They are inspired in heat diffusion theory. Diffusion is a physical process for balancing concentration changes with strong mathematical foundations. This way, smoothing filters based on diffusion interpret image intensity as that 'concentration'. Then, noise or artefacts are considered image inhomogeneities. The image intensity structure is adapted by a diffusion tensor. This tensor must be designed considering the diffusion values across structures to preserve.

Anisotropic diffusion proposed by Gil (2009) [15] will be used in this study. Non-linear anisotropic diffusion improves the adaptation of the tensor to image structure. This way diffusion values can be reduced on feature regions going around them.

All in all, anisotropic filtering will provide an output image so that the features are preserved and the noise (diffusion inhomogeneities) is reduced. Figure 3.10 shows an example where 3.10c is the vector field representing input image structure. Vectors are tangent to features and describe the structure than must be preserved.

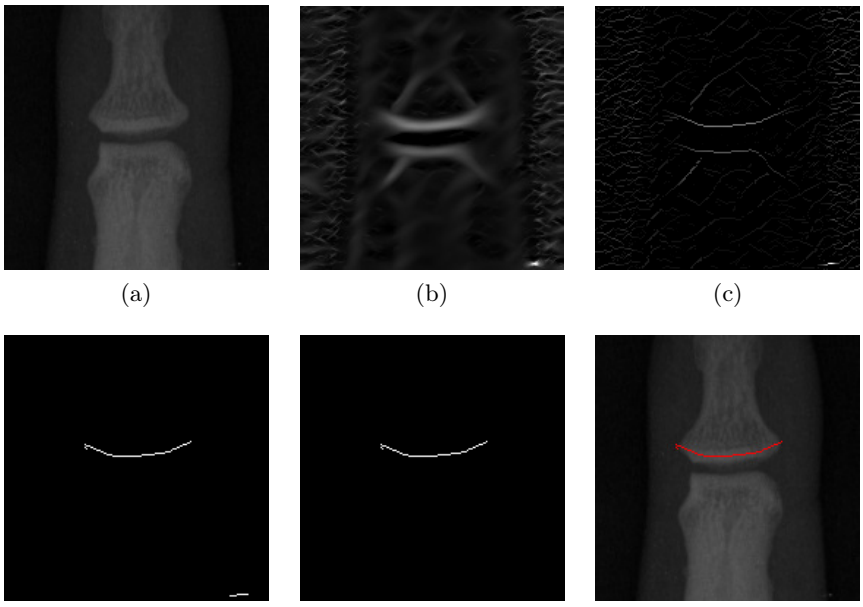


Figure 3.9: Ridge detection (a) Input image. (b) Edge space. (c) NMS. (d) Hysteresis thresholded. (e) Sclerosis ridge extracted. (f) Final detection over input

3.5.2 Bilateral filtering

Bilateral filtering is another of the many options available to smooth images while preserving edges [16]. Unlike diffusion methods, it is a non-iterative method, which make it simpler.

The basic idea of bilateral filtering consists of carrying out two different filterings. On one hand, it carries out a domain filtering, which is based on the assumption that two close pixels have similar intensity values. That assumption is not verified in regions which contain edges. Therefore, domain filtering affects to the edges present in the image. That is why bilateral filtering combines domain filtering with range filtering. Range filtering assumes that pixel similarity refers to vicinity in the range. In the same way that vicinity in domain filtering is referred by closeness.

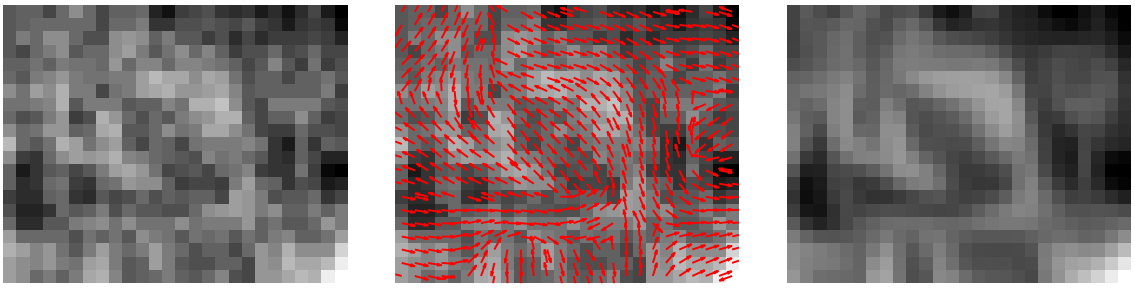
The definition of these two kind of filterings implies the necessity of a closeness function and a similarity function. In this study a shift-invariant Gaussian filtering is used in both domains so both functions are Gaussian functions of the Euclidian distance between their arguments.

3.6 Joint space width measurements

Once the main feature detector have been obtained those features are used to provide joint space width measurements. Those measurements will be tested in order to test its correlation with Sharp-Van der Heijde score.

Algorithm 1 Sclerosis extraction algorithm

```
maxlength := 0;
for Each Detected ridge j do
  Calculate Centroid(j);
  Calculate Length(j);
  if Centroid(j).x > x_left_boundary and
  Centroid(j).x < x_right_boundary and
  Centroid(j).y > y_top_boundary and
  Centroid(j).y < x_bottom_boundary and
  Lenght(j) > maxlength then
    Output := Detected ridge j;
    maxlenght := Length(j);
  end if
end for
Return Output
```



(a) (b) (c)

Figure 3.10: Diffusion example (a) Input image. (b) Vector field on input image. (c) Smoothed image.

The distance from a pixel x to a given surface S is defined as follows:

$$d_S(x) = \min_{s \in S} \|x - s\| \tag{3.2}$$

Considering that, two different sort of measurements have been obtained. The first measurement needs some kind of measure region in order to obtain the right measurement. The necessity of defining a measure region by means of parameters is something that should be avoided so these score is only used to test the correlation to SvdH. These measurement is the mean distance, d_mean, given by the following expression:

$$d_{mean} = \frac{1}{N} \sum_{x \in L'} d_S(x) \tag{3.3}$$

Algorithm 2 Lower bone extraction algorithm

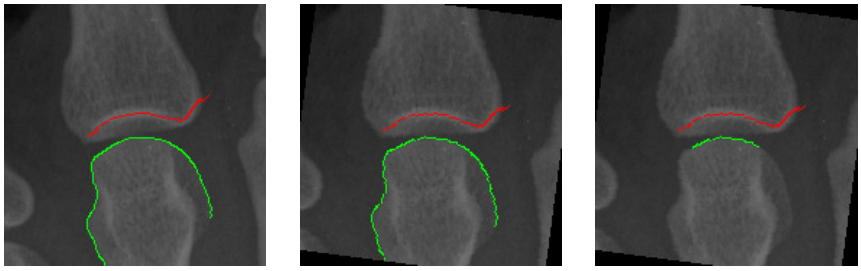
```
maxlength := 0;
maxIntensity := 0;
max_y := Image_height;
for Each Detected edge j do
  Calculate Centroid(j);
  Calculate Length(j);
  Calculate MeanIntensity(j);
  if Centroid(j).x > x_left_boundary and
  Centroid(j).x < x_right_boundary and
  Centroid(j).y > y_top_boundary and
  Centroid(j).y < max_y and
  MeanIntensity(j) > maxIntensity and
  Lenght(j) > maxlength then
    Output := Detected edge j;
    maxlenght := Length(j);
    maxIntensity := MeanIntensity(j);
    max_y := Centroid(j).y;
  end if
end for
Return Output;
```

where N is the number of samples available, S is the sclerosis surface and L' is a part of the lower bone. That part of the lower bone is obtained by using the joint orientation already computed. The joint orientation of every sample is corrected so that a vertical view of the final sclerosis and lower bone detections can be obtained. Afterwards, since the joint is centered in the sample image, the lower bone surface is cropped in order to keep only the part contained in the centered quarter of the image width, L' (Figure 3.11. This cropping is necessary as these measurements mean values and the lowest pixels of the lower bone must not be used.

The next two measurements avoid any arbitrary limitation of the measurement region. Minimum distance, d_{min}, and middle distance, d_{mid}, are defined:

$$d_{min} = \min_{x \in L} d_S(x) \qquad (3.4) \qquad\qquad d_{mid} = d_L(x') \qquad (3.5)$$

The middle distance is the distance from the sclerosis pixel x' to the lower bone. x' is the pixel at the middle of the sclerosis along the image x axis:



(a) (b) (c)

Figure 3.11: Measurement computation example. (a) System detection. (b) Orientation correction (c) Considered measurement region.

Chapter 4

Experimental setup

4.1 Ground Truth

The following experiments aim to develop different feature detectors. Those features are ridges and edges. A good ridge detection will lead us to a good sclerosis detection. In the same way, a good edge detector must allow us to obtain a good lower bone detection.

The experiments are introduced step by step in order to show the progressive improvement of the detectors based on partial results.

Eventually, the detected sclerosis and lower bone may lead us to achieve measurements and characterize the corresponding joint. That characterization should show the arthritis affection of that joint.

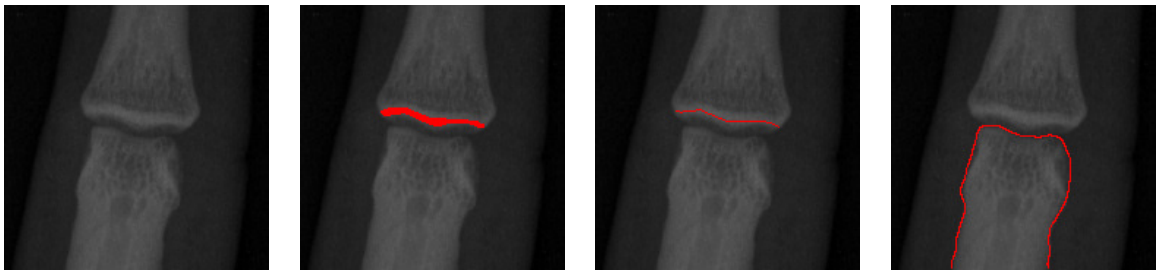
The sample dataset has already been introduced in section 2.1. 360 scored MCP and PIP samples are available. The dataset only provides SvdH scorings though. Therefore, that dataset is not useful at this stage of the study since it is firstly necessary to evaluate the quality of our detections.

Segmentation assessment is a common problem that computer vision usually have to face. Different techniques can be used [17]. The singularity of our problem and the chances of adding more samples to the ground truth in the future must be considered. That is way it was decided to compare our automatic detections to delineations provided by an expert. It means it is necessary to have at our disposal those expert delineations and some proper metrics to achieve the quantitative comparison.

A 60 sample groundtruth was created in order to help us in the task of developing our sclerosis and lower bone detector. Those 60 joint samples include 15 samples from each finger but thumbs. 5 MCP samples, 5 PIP samples and 5 DIP samples per joint.

OsiriX [18] software was used in order to develop the groundtruth. It is an image processing software which provides an environment and tools to work with DICOM images. One of those tools allows the user to define a list of points in the image. It all means our 60 samples had to be transformed to DICOM format. Afterwards, OsiriX was used to introduce the list of points defining the properly groundtruth. Finally, those point were processed again in order to obtain the final masks.

The lower bone was requested as the edge described by the lower bone contour considering the model previously exposed. However, the sclerosis delineation was asked to be the region covering the concavity that defines it. Afterwards, the sclerosis feature that has been described by a ridge present in that concavity will be obtained by skeletonization of that sclerosis mask. It helps us to minimize the difficulty of delineating a ridge and the variability it could introduce 4.1.



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

Figure 4.1: Groundtruth (a) Sample. (b) Sclerosis. (c) Sclerosi skeletonization (d) Lower bone.

4.2 Performance metrics

Once the groundtruth is available, some metrics must be developed in order to evaluate the quality of our detections.

Different metrics can be used in order to achieve the assessment of our detections. Choosing proper methods to assess the quality of our segmentations can help us to show the performance of our algorithms. Firstly, it is necessary to state what kind of information is intended to be obtained from those quality evaluation methods.

As the final objects that have to be compared are ridges and edges, it was decided to base our metrics in the average surface distance (ASD). Given a pixel p and a region R conformed by a set of pixels $S(R)$, the shortest distance from p to $S(R)$ is defined as:

$$d(p, S(R)) = \min_{s_R \in S(R)} \|p - S_R\| \tag{4.1}$$

where $\|.\|$ stands for the Euclidean distance.

Let U and V be two segmentations. Then, the average surface distance from U to V is given by

$$ASD(U, V) = \frac{1}{|S(U)|} \left(\sum_{s_U \in S(V)} d(s_U, S(V)) \right) \tag{4.2}$$

Eventually, if A denotes our automatic segmentation and M denotes the manual delineation, our two metrics are defined as:

$$Caught = ASD(A, M) \tag{4.3}$$

$$Missed = ASD(M, A) \tag{4.4}$$

Caught value is obtained as the distance from the automatic detection to the manual segmentation. Its best value would be 0. It would mean all the pixels in A are part of M. Caught metric value is related to the quality of the detector at detecting pixels that must be detected (true positive notion). The higher Catch value, the higher the accuracy of the detector and the less non-desired pixels (false positives) were detected.

Analogously, Missed value is the ASD from the manual groundtruth to the automatic detection. If it would provide a value of 0 it would mean all the pixels in the manual segmentation are part of the detection. The higher its value, the farther the manual segmentation is from the detection. It provides us with information on how many pixels which should have been detected were missed (false negatives).

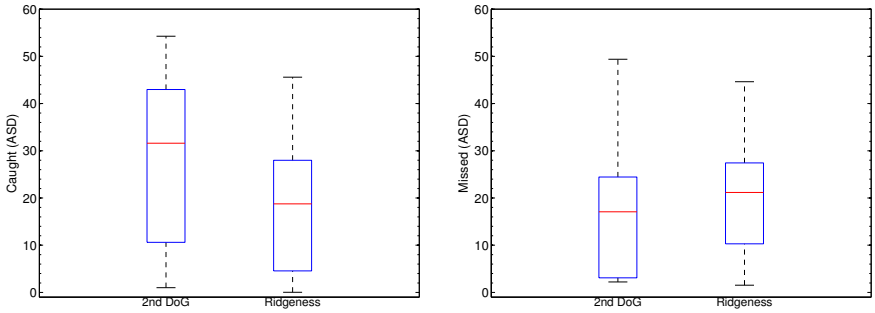
Both metrics are positive and continuous distance values in pixels. It means bad detections are more penalized if the pixels are farther. It allows us to minimize the affection of slightly shift detections.

Chapter 5

Results

5.1 Sclerosis

A proper sclerosis detection should prompt us to develop an accurate ridge detector. Our first experiment aims to compare the system performance when using the two previously introduced methods to obtain the feature space image. As shown in Figure 5.1, second derivative of Gaussian bank of filters provides better results in terms of Missed metrics, which means it misses less useful information. Caught values verifies this result as derivative of Gaussian provides higher values. Considering that, we can conclude that the second derivative of Gaussian performs better but it also introduces more non desired ridges.



(a) (b)

Figure 5.1: Comparing feature space image methods(a) Caught. (b) Missed.

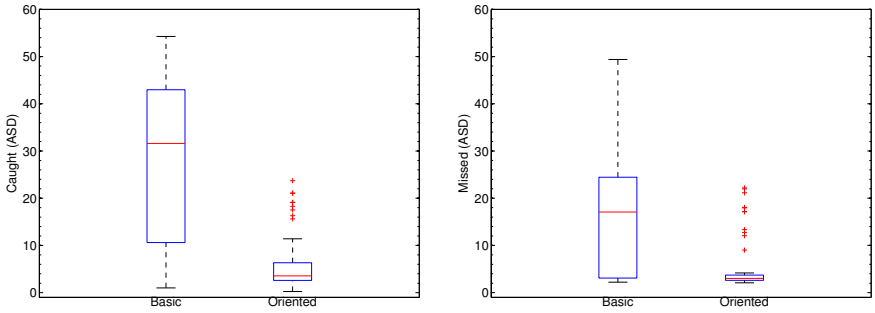
Regarding the previous result, it seems that second derivative of Gaussian method can provide us with better results. However, removing non desired detections becomes an important issue.

In that sense, an idea would be that our ridge detector should take advantage of the expert's knowledge provided by the previous model study. Our ridge and sclerosis detection can use joint orientation

information. Any joint sclerosis should be found in the orientation of the finger. This information can be obtained using the hand bone marrow computation introduced in section 3.

Once those orientations have been obtained their values are added to the available 60 sample ground truth. The orientation values are visually validated and 5 of them are found to be wrong so those samples are removed from our ground truth. Therefore, the computation of our metrics is done over the remaining 55 samples in order to show performance metrics only about our detection.

Figure 5.2 shows the box plots comparing the two systems. Adding orientation to our Basic Detector (BD) clearly increases the performance. As a matter of fact, Missed values drop considerable and Caught values fall sharply reaching pretty low rates. Considering our metrics definition, Oriented Detector (OD) is much more precise detecting sclerosis and much better than the basic one in terms of sensitivity.



(a) (b)

Figure 5.2: Comparing Basic detector and Oriented detector(a) Caught. (b) Missed.

The qualitative assessment of the results confirms the previous comments. Oriented detection of ridges helps us to obtain more accurate sclerosis (not extended) (Figure 5.3).

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

Figure 5.3: (a) BD hysteresis output. (b) BD sclerosis detection. (c) OD hysteresis output. (d) OD sclerosis detection.

Taking advantage of orientation knowledge increases our detector performance. Another important information from our model that must be considered is related to scale. Different detectors using different

filter scales have been compared in the next experiment. Figure 5.4 shows their performance where the boxplots represent our metrics and Figure 5.4c shows the amount of non-detections provided by each detector. It is important to show this information since it can not be included in the metric boxplots. These non-detections happen when a proper ridge was not found by the extraction module of the detector. From now on, if non-detections happen they will always be provided.

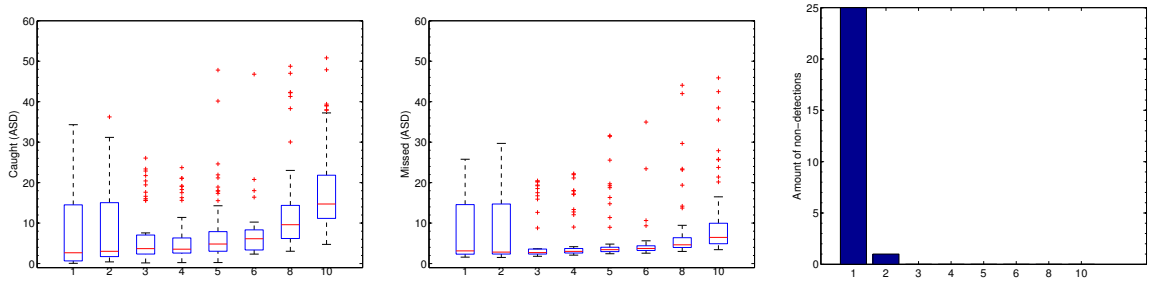


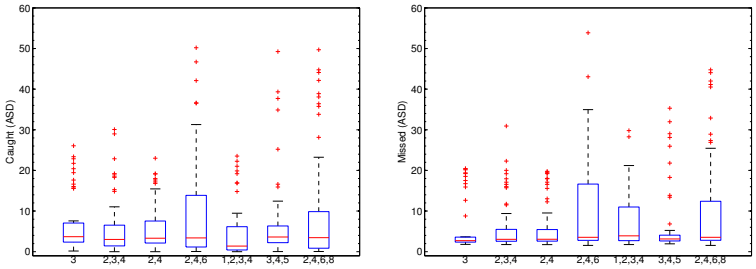
Figure 5.4: Different sigma sclerosis detector (a) Caught. (b) Missed. (c) Non-detection.

Results show that lower sigma 1 must be rejected. Regardless its performance shown by Caught and Missed values, the high non-detection rate can not be accepted. As long as the other detectors are concerned, a clear increasing trend of Caught values is observed as sigma value increases. Although sigma values of 3, 4, 5 and 6 provide close results. If Missed metric is considered, there is not a clear tend but the same values are again those which provide better results. Considering these results sigma value of 3 should be considered the best opcion.

Although some of the previous experiments seem to provide better results, an interesting idea would be to check the performance of detectors combining different sigma information. It can be achieved by computing the different feature space images for each sigma. Afterwards, those feature space images are combined before going through the NMS module. That combining operator can be the maximum of the different values of a given pixel in the different feature space images. Other operators such as minimum and average were checked always with worse results. The orientation image is also obtained as the maximum.

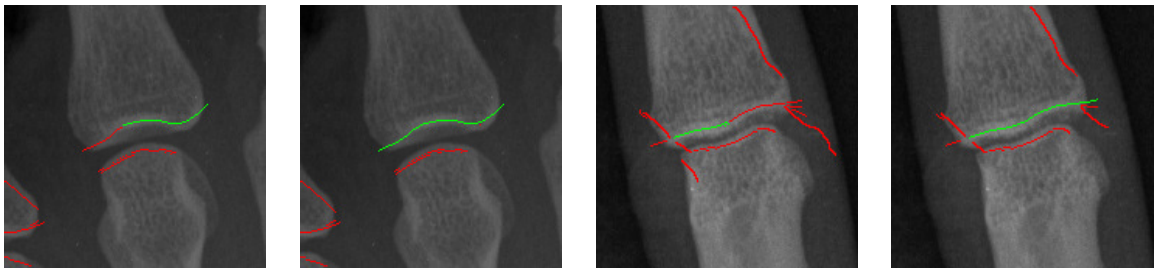
Figure 5.5 shows some results using different scales and combining them with the maximum operator. The best scale showed above is added to the plot too. Since any of the combinations seem to provide a considerable improvement the best scale will be 3.

At this point, the system can provide us with considerably good results. Its qualitative assessment let us know about some problems we must be facing. Figures 5.6a and 5.6c illustrate those results. In a remarkable number of cases the final detection appears to be partial although the missing part was also obtained. It lead us to think some kind of feature-preserving filtering could solve it. As introduced above, these methods carry out some kind of noise filtering trying to preserve edges or ridges. It may help us to obtained better and non partial sclerosis.



(a) (b)

Figure 5.5: Multiple sigma detector (a) Caught. (b) Missed. (c) Non-detection.



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

Figure 5.6: Examples about filtering (Red: hysteresis output, Green: sclerosis extraction) (a) Example 1 (no filter). (b)) Example 1 (diffusion). (c) Example 2 (no filter). (d) Example 2 (diffusion).

Some previous experiments where achieved in order to parametrize the two filters. The filters are parametrized using the same sample set. A training set can not be used since the available sample set is not big enough. Anyway, the obtained parameters help us to check these tools in order to carry out a first approach to its performance.

Diffusion filter has two parameters. The experiment results are shown in Figure 5.7. Close results are obtained for the different tested values. As a decision must be taken, the chosen values are 1 for the sigma parameter and 6 for rho. Bilateral filter parameters are obtained analogously.

The results comparing the systems which add a filtering module and the plain system are shown by 5.8. Unlike what it was expected, none of the two feature-preserving methods improved our system performance. Actually, diffusion filter metrics practically show the same performance.

The qualitative assessment of the results makes the reasons clearer. In some cases our hypothesis is satisfied and we avoid the problem of partial detections. Figure 5.6 shows two examples and confirms our assumptions. Nevertheless, in most of the cases the detection provided by the system without filtering is very similar. Filtering has a clear impact in the system. It reduces the number of detected artefacts in the shape of non-desired ridges but this impact usually becomes useless in the moment of extraction.

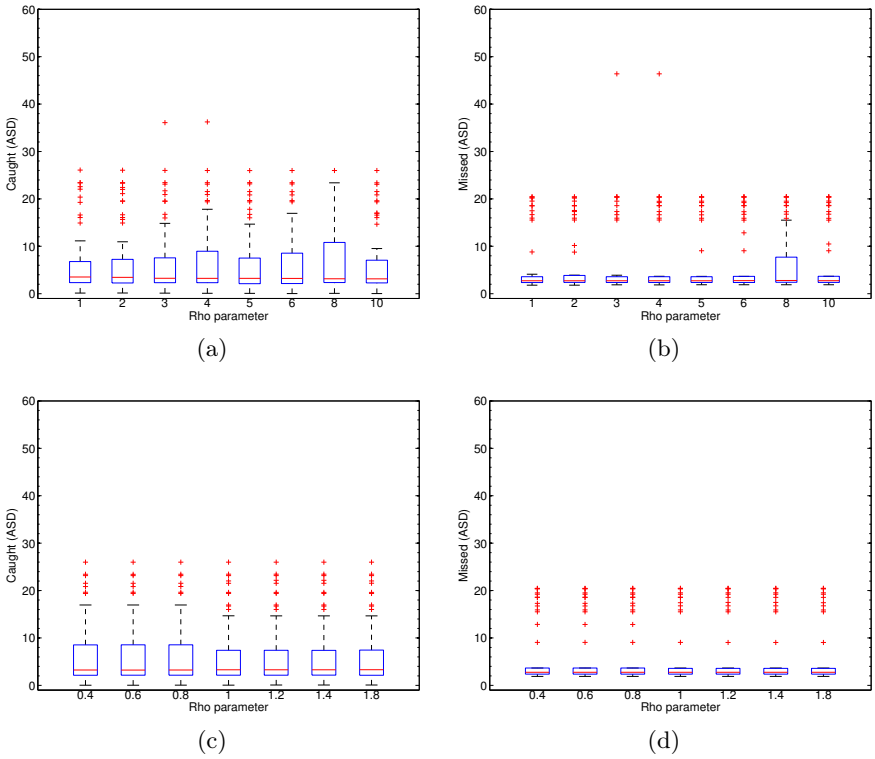


Figure 5.7: Parametrizing (a) Caught. (b) Missed.

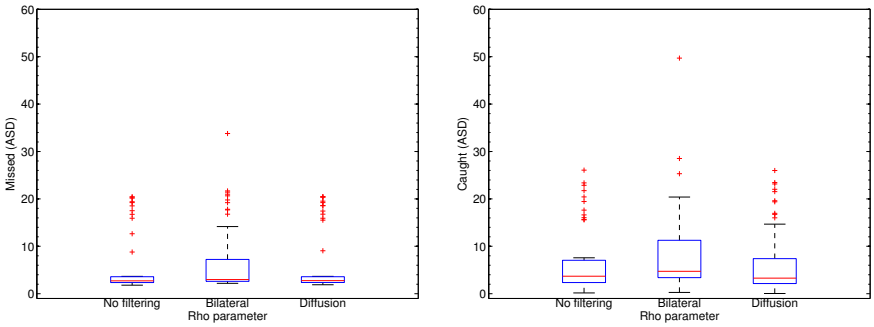
Even though the hysteresis module output presents less artefacts, the extracted ridge is nearly the same most of the times.

Another problem is observed in the qualitative assessment of the results. The extraction module provides a lot of errors. Even thought the ridge associated to the sclerosis have been detected, the extractor module does not choose it as the final output. However, 10 out of the 11 extraction errors happen with MCP joints. It can be solved parametrizing the extraction module with different parameters for MCP joints. The bottom boundary in the sclerosis extraction module (x_bottom_boundary in algorithm 1) must be lower for MCP joints.

Finally, the best sclerosis extraction does not need a previous module and the better scale is 3. Performance metrics are show in Figure 5.8. Very low values of Caught metrics, low enough values of Missed metrics and only 2 extraction errors (3.63%).

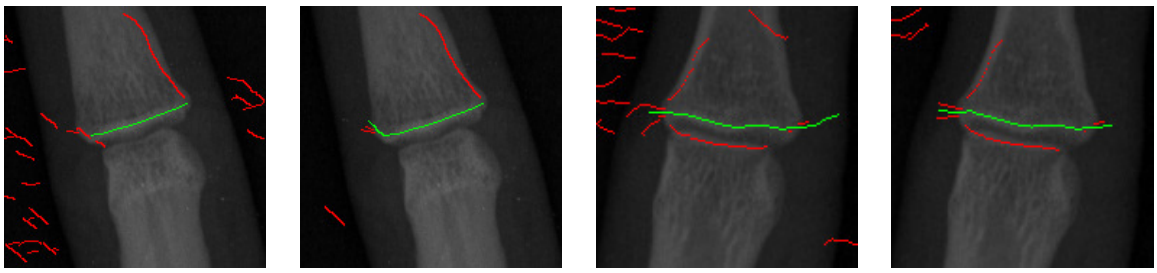
5.2 Lower bone

Analogously to the previous section, this section will show the experiments around the questions related to the lower bone detector. It does not seem useful in this case to introduce expert knowledge



(a) (b)

Figure 5.8: Comparing filters (a) Caught. (b) Missed.



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

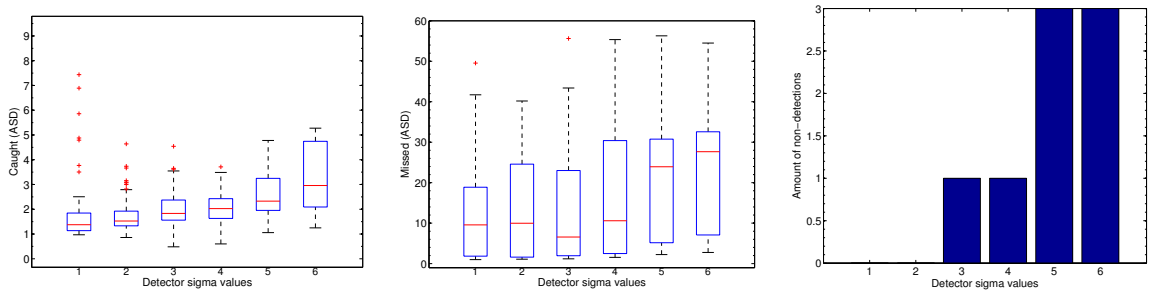
Figure 5.9: Examples about filtering (Red: hysteresis output, Green: lower bone extraction). (a) Example 3 (no filter). (b)) Example 3 (diffusion). (c) Example 4 (no filter). (d) Example 3 (diffusion).

about orientations. The lower bone shape can produce edges with multiple orientations.

The first question that can be studied is the influence of scale in the detector performance consequently leading us to the first experiment. All the modules were set in the same conditions and only scale was changed. Figure 5.10 shows the results. The first conclusion is that the lower scale is the one that provides the best results. It is an expected result considering the scale of the edges to be obtained.

These results also introduce us to a clear consideration. It can be observed that Caught values are low even in the worst cases. It means our lower bone detection generally provides us with a good detection. However, Missed values are high. It all let us to conclude that our lower bone detection is generally part of the lower bone. Even though there are some parts of it that are missed.

Obtaining incomplete edges is the main problem of the system so far. These are some samples where the detection was incomplete although the presence of the edge was clear, as our model introduced in section 2.2. Consequently, our next experiment tries to assess the impact of a preserving-edge filter on our detection. It would be expected that its abilities to remove noise preserving edges would strengthen the presence of clear and complete edges.

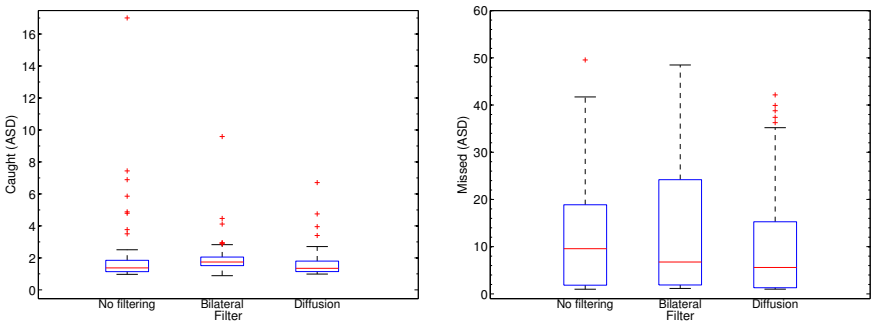


(a) (b) (c)

Figure 5.10: Different sigma lower bone detector (a) Caught. (b) Missed. (c) Non-detection.

The diffusion filter and the bilateral filter were parametrized analogously to the previous section.

The metrics obtained in the experiment are shown by Figure 5.11. The performance metrics does not show an important performance improvement. Although the impact of the filters is clear in previous stages of the system, when the amount of non-desired edge decrease, the final output provided by the extraction module does not show such important improvement. Anyway, the skills of the system if it uses diffusion filtering are higher in terms of Missed metrics which means the detector is missing less parts of the lower bone contour.

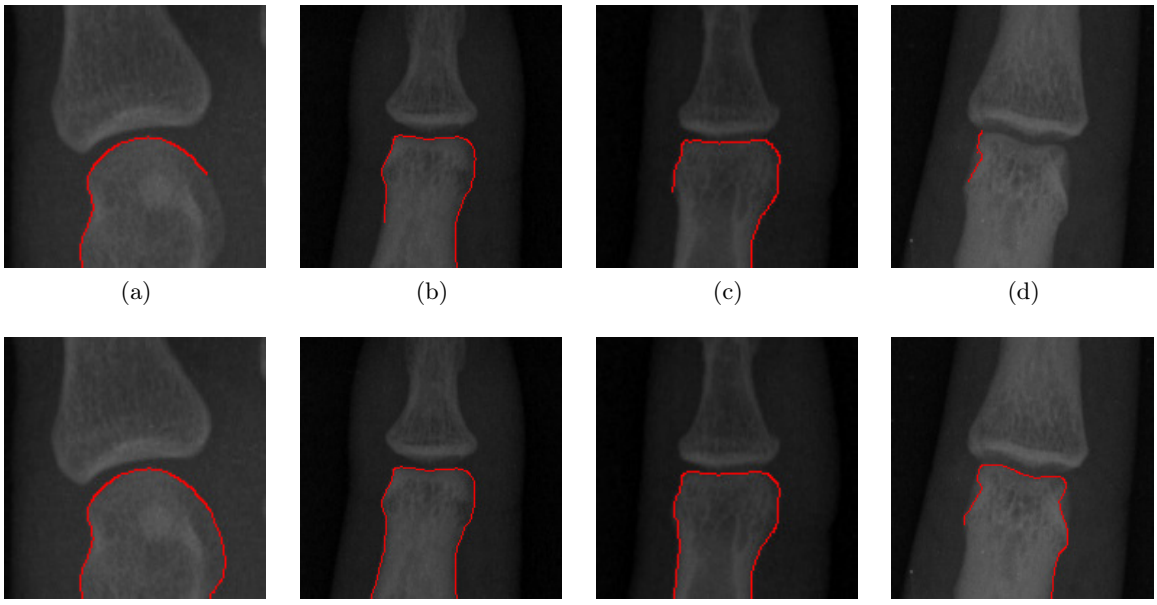


(a) (b)

Figure 5.11: Comparing filters (a) Caught. (b) Missed.

Figure 5.12 shows the final detection provided by the system when using diffusion filter compared to the detector that does not uses filtering. It illustrates the exposed problem and the expected performance improvement when adding the diffusion filter.

Even though the performance improvement is lower than expected, the use of diffusion filtering still appears to be a useful tool to obtain clearer edge detections where the presence of artefacts has been reduced. It is after extraction where the impact decreases. That is why filtering can still be a helpful tool in order to move forward in the goal of finding new landmarks to characterize the joints.



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

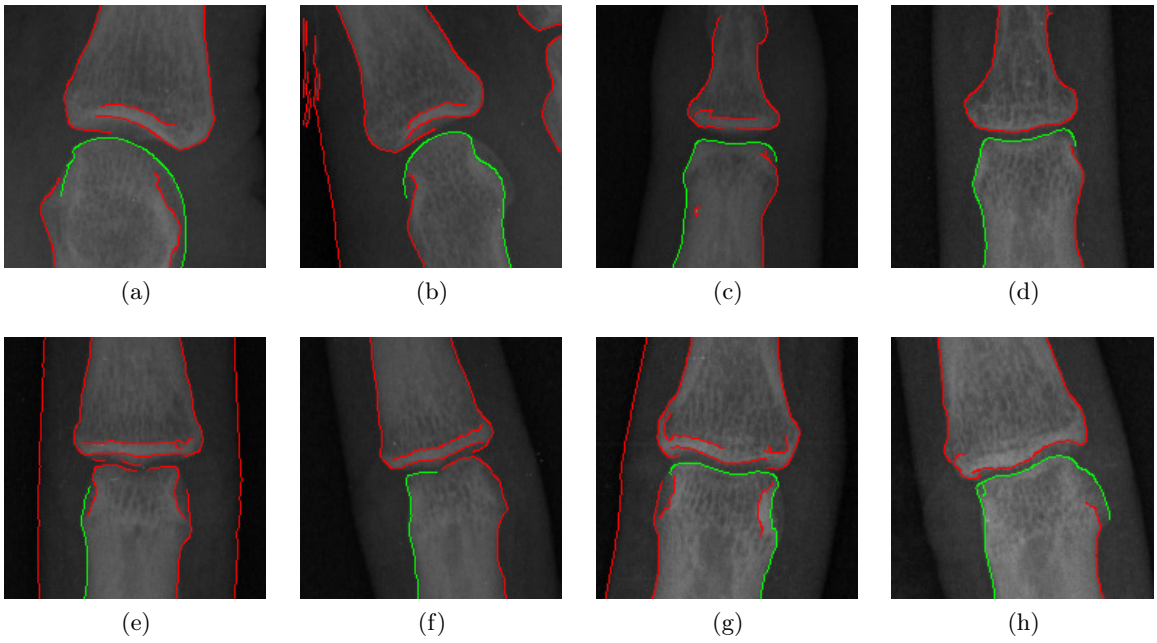
(e) (f) (g) (h)

Figure 5.12: Comparing basic detector to filtered detector. (a)(b)(c)(d) Using no filter. (e)(f)(g)(h) Using diffusion.

At this point, the qualitative assessment of the samples helps us to clarify the sources of error in our results. As far as we know, some main errors are found which are related to each other. As mentioned above, the performance of the system is fairly good in terms of precision (represented by our Caught metrics). The detection is almost in all cases right. The problem is that there are missing parts in the final extraction. It could make us think we are not detecting the required edges. Nevertheless, the lower bone edges are obtained most of the times although they are broken in different pieces. Furthermore, they are always broken in the same part of the contours, and it takes us to the problem we were expecting, which was already figured in the joint model. The inner contours in the lower bone cause error in our output.

As introduced in section 2.2, MCP and PIP joints present inner contours in the lower bone. Therefore, a clear performance difference exists between those joint types and DIP joints. Lower bone detection becomes easier in DIP joints and the errors in those cases are caused by partial detections although much less probable and only in more specific circumstances. They appear to be broken about the upper right corner of the lower bone. An edge irregularity seems to appear in that region even though there are not inner contours (Figures 5.13c, 5.13d).

As long as the other joints are concerned, they present inner edges. MCP usually present inner edges but not always. Furthermore, in some cases their intensities are lower than the outer edge intensities so they are not always a problem. MCP joint appears to be broken on the sides of the lower bone (Figures



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

(e) (f) (g) (h)

Figure 5.13: Lower bone detection problems (Red: hysteresis output, Green: lower bone extraction). (a)(b) MCP. (c)(d) DIP. (e)(f)(g)(h) PIP.

5.13a, 5.13b).

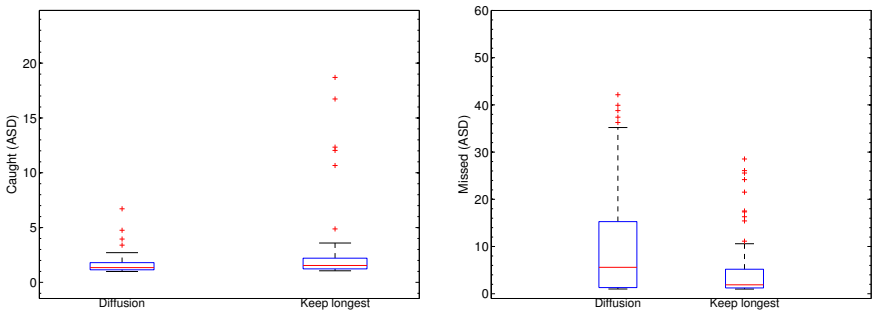
PIP joints present inner edges more often and they are even more important in healthy joints. It has much to do with the fact that PIP joints are the most difficult in terms of lower bone detection. The lower bone edges are usually broken by the sides of the bone. Furthermore, they are usually broken also in the middle of the top part of the lower bone, which means that it appears broken in the middle of the joint. It is a new problem that makes PIP joints a difficult and particular case in terms of lower bone detection (Figures 5.13e, 5.13f, 5.13g, 5.13h).

As far as we know, our detections are correct most of the times. Furthermore, adding a filter module at the beginning of our detector vary our final output although the average results keep very similar. Since our detections are correct in the different detector configurations, which means using a different filter, we could obtain those different detections and keep as the output the longer one. Therefore, our next experiment In order to obtain those different outputs 5 detector

Our next experiment is built on that basis. 6 different detectors are configured. Five of them use different filters (average, median, gaussian, bilateral, diffusion). The sixth output is the one provided by a system without filter module.

Figure 5.14 shows the results of the experiment compared to the best system so far. These results verify the expected performance. The system misses less parts of the lower bone as shown by the clear decrease in Missed values. Actually, Missed values are half of the last better system (which uses just

diffusion). The slight increase in the Caught values was also expected. That increase is small enough to make the system that keeps the longest detection the best performance system



(a) (b)

Figure 5.14: Comparing systems: using diffusion filter and using different filters and keeping longest detection (a) Caught. (b) Missed.

5.3 Correlation to Van der Heijde

In this section the better sclerosis and lower bone detectors are used in order to compute joint space width (JSW) measurements. The sclerosis detector includes the basic modules introduced in section 3 and the orientation information. The lower bone detector also includes de basic modules plus the diffusion filter and, finally, the module which chooses the longest detection module between six different systems.

As mentioned before, the SvdH score is only provided for MCP and PIP joint. Consequently, our system will have to be tested with the available 320 samples of the mentioned joints. Figure 5.15 shows the amount of samples available for every SvdH score value.

	SvdH 0	SvdH 1	SvdH 2	SvdH 3
MCP	135	11	12	2
PIP	87	35	28	10

Figure 5.15: Number of joint samples available for every SvdH score value.

Once the samples have been tested a visual qualitative assessment is necessary in order to separate wrong detections. Those sclerosis detections which does not correspond to the corresponding ridge that defines the sclerosis are considered wrong. Our system provides an 84.37% of right sclerosis detections. Those lower bone detections which does not provide the top part of the lower bone, which define the joint interspace, are considered wrong. The system provides a 47.83% of right lower bone detections.

The errors have been caused by previously mentioned situations. As far as sclerosis is concerned, the first source of error is the extraction module in the joint characterization. The right sclerosis ridge is

missed sometimes in the moment of the extraction even it was provided as a detected ridge. Furthermore, this new data set includes several degrees of RA affection, which means that those samples are different from the basic model. As the samples becomes more affected by the disease the JSW decreases but it is not the only change. The sclerosis loses its concavity and new ridges that are not considered in the model appear.

As far as lower bone is concerned, the errors are caused by the partial detections, as it has been exposed above. When applying our system to the new dataset the difficulty of the sample increases.

All the samples with both an sclerosis and a lower bone valid detection are separated in order to compute the previously introduced distance measurements.

Figure 5.16 includes box plots showing the measurements. PIP boxplots does not include samples with SvdH score of 3 as the system did not succeed in detecting the features.

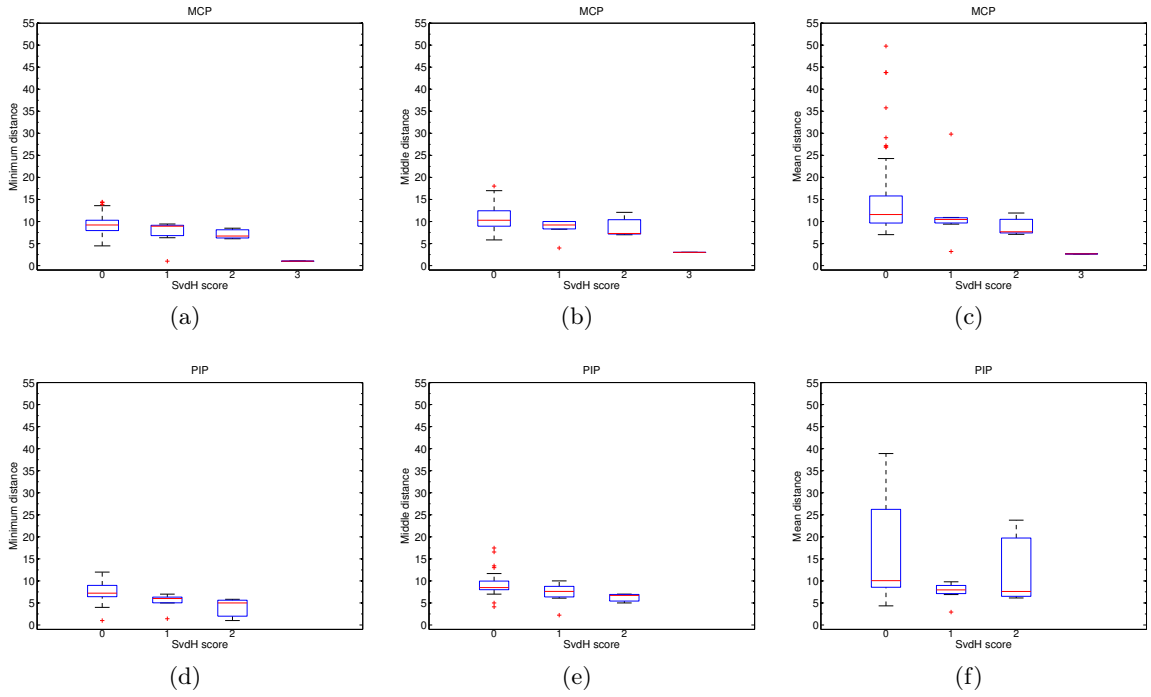


Figure 5.16: Box plots showing JSW measurements for MCP and PIP joints.

The three measurements show a clear correlation with SvdH scores. They present a decreasing trend while the SvdH value increases. The median values of every distribution are close but the decreasing trend is verified by all the distance methods in both MCP and PIP joints.

As explained before, mean distance needs to set boundaries to the area where the measures are taken. This distance approximation was tested in order to verify its correlation to SvdH values. Nevertheless, this measurement would not be the best one as it is desirable to avoid the necessity of area boundaries.

Middle distance also verifies the correlation with SvdH in both MCP and PIP joints. This measure

has other implications we wanted to test. The joints are not supposed to be symmetric. Therefore, taking as a reference the middle point of the sclerosis may lead us to wrong measurements since a joint could lose joint space in different parts. However, this results show, even though that limitation, the correlation is verified.

Minimum distance also presents a decreasing trend correlated to SvdH scores. This distance approximation would be preferable as it avoids area boundaries and have not any limitation about the considered parts of the joint.

Chapter 6

Conclusions

In this study, we have set the foundations of a modular system that aims to compute hand joint measurements which can provide standard information about Rheumatoid Arthritis disease affection. In that direction, the problem was formulated in three independent main steps: joint detection, joint characterization and JSW measurements, and we focused on the two latter stages.

A joint data set have been created and a hand joint model have been developed so that we showed that sclerosis and lower bone contours are necessary to obtain a first approach to joint characterization.

In order to develop the sclerosis and lower bone detectors, an experimental set up have been stated. This contribution includes the creation of a ground truth with part of the data set, which allows us to test the detectors, and the design of suitable performance metrics.

By means of that environment, a sclerosis detector and a lower bone detector have been obtained after testing different tools. The described systems presented pretty high rates of success in detecting sclerosis. As far as we know, the extraction step seems to be the main source of error among the different modules implied in the characterization stage. The extraction module is not accurate enough choosing the corresponding ridge even though the previous steps succeed in detecting the ridge which corresponds to the sclerosis feature described in the model.

Considering the lower bone detection, we showed that the main drawback is the detection of partial edges. Different parts of the bone have been identified as those where the edge usually appears to be broken. Furthermore, it lead us to conclude that different problems are found in the three different types of joints which should prompt us to consider them in different ways. Furthermore, PIP joints have been identified as the most difficult joints to detect the lower bone.

Preserving-edge filters were tested as a prior module in the characterization stage. Their application seemed to be interesting as they could remove non-desired ridges or edges while remarking the desired ones. Nevertheless, the detection rates did not increase. While some cases were improved because partial detections were linked, some cases did not change its performance at all. We concluded that the filtering removes non-desired artefacts but it does not have any effect in the moment that the extraction is carried

out. The image which arrives to the extraction module has less non-desired artefacts but the extracted one is usually very similar.

The better sclerosis and lower bone detectors were finally tested in the whole sample set to provide distance measurements. Considering that the top part of the lower bone is strictly necessary to carry out distance measures, the lower bone detections that did not satisfy it were discarded. We obtained high rates of sclerosis detection but lower rates for lower bone because of the reasons mentioned above.

Three distance measures have been proposed. The three of them presented a clear correlation with the SvdH score. Besides, two of them does not require to set the measuring area, which improves previous approaches. In addition, the correlation of our measures is clearer that the approaches in the bibliography.

6.1 Future work

The future work lines should involve the creation of a bigger dataset in order to test the sclerosis and lower bone systems performance. It would allow the creation of training and test sets.

Following the information provided by the model described in this study, the introduction of new landmarks would help to develop more accurate measurements. Another Computer Vision feature could be added to the model, as valleys, for example.

The extraction module in the characterization stage is the weakest step in the detection. As far as we know, the inclusion of improvements in the extraction algorithms should provide an increase in the performance rates.

The main problem in the lower bone detection are the partial detections. An interesting work line would be to tackle the task of joining different independent edges that could be part of the lower bone contour.

Expanding the scope of this study, thumbs should be included in the hand joint model as well as providing measurements of erosion, the other main symptom of RA which is visible in X-ray images.

Bibliography

[1] V. Majithia and S. Geraci, "Rheumatoid arthritis: diagnosis and management," *The American journal of medicine*, vol. 120, no. 11, pp. 936–939, 2007.

[2] A. Rojas-Villarraga and E. Páramo, "Diagnostic images in rheumatoid arthritis," *MedUNAB*, vol. 9, no. 11, pp. 108–119.

[3] A. Guerrero and C. Villaseñor, "Evaluación radiográfica del daño anatómico en la artritis reumatoide," *Revista colombiana de reumatología*, vol. 13, no. 3, pp. 214–227, 2006.

[4] S. Boini and F. Guillemin, "Annals of rheumatic diseases," *Ann Rheum Dis*, vol. 13, no. 60, pp. 817–827, 2001.

[5] J. Sharp, M. Lidsky, L. Collins, and J. Moreland, "Methods of scoring the progression of radiologic changes in rheumatoid arthritis. correlation of radiologic, clinical and laboratory abnormalities," *Arthritis & Rheumatism*, vol. 14, no. 6, pp. 706–720, 1971.

[6] D. Van der Heijde, "How to read radiographs according to the sharp/van der heijde radiological assessment in rheumatoid arthritis in long term studies," *Journal of Rheumatology*, no. 26, pp. 743–745, 1999.

[7] S. Vera, "Finger joint modelling from hand x-ray images for assessing rheumatoid arthritis," Centre de Visió per Computador, Tech. Rep., 2010.

[8] A. Pfeil, J. B
"ottcher, M. Sch
"afer, B. Seidl, M. Schmidt, A. Petrovitch, J. Heyne, G. Lehmann, P. Oelzner, G. Hein *et al.*, "Normative reference values of joint space width estimated by computer-aided joint space analysis (cajsa): the distal interphalangeal joint," *Journal of Digital Imaging*, vol. 21, pp. 104–112, 2008.

[9] M. Fabel, H. von Tengg-Kobligk, F. Giesel, L. Bornemann, V. Dicken, A. Kopp-Schneider, C. Moser, S. Delorme, and H. Kauczor, "Semi-automated volumetric analysis of lymph node metastases in patients with malignant melanoma stage iii/iv-a feasibility study," *European Radiology*, vol. 18, no. 6, pp. 1114–1122, 2008.

[10] J. Angwin, G. Heald, A. Lloyd, K. Howland, M. Davy, and M. James, "Reliability and sensitivity of joint space measurements in hand radiographs using computerized image analysis." *The Journal of Rheumatology*, vol. 28, no. 8, p. 1825, 2001.

[11] M. Wick, P. Peloschek, K. B
"ogl, W. Graninger, J. Smolen, and F. Kainberger, "The "x-ray rheumacoach" software: a novel tool for enhancing the efficacy and accelerating radiological quantification in rheumatoid arthritis," *Annals of the rheumatic diseases*, vol. 62, no. 6, p. 579, 2003.

[12] B. Zielinski, "Fully-automated algorithm dedicated to computing metacarpophalangeal and interphalangeal joint cavity widths," *Schedae Informaticae*, vol. 16, pp. 47–67, 2007.

[13] A. Bielecki, M. Korkosz, and B. Zielinski, "Hand radiographs preprocessing, image representation in the finger regions and joint space width measurements for image interpretation," *Pattern Recognition*, vol. 41, no. 12, pp. 3786–3798, 2008.

[14] A. Lopez, F. Lumbreras, J. Serrat, and J. Villanueva, "Evaluation of methods for ridge and valley detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, no. 4, pp. 327–335, 1999.

[15] D. Gil, A. Hernàndez-Sabaté, M. Burnat, S. Jansen, and J. Martínez-Villalta, "Structure-preserving smoothing of biomedical images," in *Computer Analysis of Images and Patterns*. Springer, 2009, pp. 427–434.

[16] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Computer Vision, 1998. Sixth International Conference on*. IEEE, 1998, pp. 839–846.

[17] T. Heimann, B. van Ginneken, M. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes *et al.*, "Comparison and evaluation of methods for liver segmentation from ct datasets," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 8, pp. 1251–1265, 2009.

[18] A. Rosset, L. Spadola, and O. Ratib, "Osirix: An open-source software for navigating in multidimensional dicom images," *Journal of Digital Imaging*, vol. 17, pp. 205–216, 2004, 10.1007/s10278-004-1014-6. [Online]. Available: http://dx.doi.org/10.1007/s10278-004-1014-6