

AUTOMATED RADIOGRAPHIC ASSESSMENT  
OF HANDS IN RHEUMATOID ARTHRITIS

Joost A. Kauffman

De promotiecommissie:

*voorzitter en secretaris:*

Prof.dr.ir. A.J. Mouthaan      Universiteit Twente

*promotor:*

Prof.dr.ir. C.H. Slump      Universiteit Twente

*assistent promotor:*

Dr. H.J. Bernelot Moens      Ziekenhuisgroep Twente

*leden:*

Prof.dr. M.A.F.J. van de Laar      Universiteit Twente /  
Medisch Spectrum Twente

Prof.dr. E. Marani      Universiteit Twente

Prof.dr.ir. W. Philips      Universiteit Gent

Prof.dr.ir. P.P.L. Regtien      Universiteit Twente

Prof.dr.ir. G.J. Verkerke      Rijksuniversiteit Groningen /  
Universiteit Twente

This research is financially supported by the Dutch Arthritis Association.



Signals & Systems group,  
EEMCS Faculty, University of Twente  
P.O. Box 217, 7500 AE Enschede, the Netherlands



© Joost A. Kauffman, Enschede, 2009

No part of this publication may be reproduced by print, photocopy or any other means without the permission of the copyright owner.

Printed by Gildeprint B.V., Enschede, The Netherlands  
Typesetting in L<sup>A</sup>T<sub>E</sub>X<sub>2</sub>e

ISBN 978-90-365-2830-6  
DOI 10.3990/1.9789036528306

AUTOMATED RADIOGRAPHIC ASSESSMENT OF HANDS IN  
RHEUMATOID ARTHRITIS

PROEFSCHRIFT

ter verkrijging van  
de graad van doctor aan de Universiteit Twente,  
op gezag van de rector magnificus,  
prof. dr. H. Brinksma,  
volgens besluit van het College voor Promoties  
in het openbaar te verdedigen  
op 7 Mei 2009 om 13.15

door

Joost Adriaan Kauffman  
geboren op 21 April 1978  
te Den Helder

Dit proefschrift is goedgekeurd door:

De promotor: Prof.dr.ir. C.H. Slump

De assistent promotor: Dr. H.J. Bernelot Moens



# Contents

<b>Nomenclature</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Bones of the hand . . . . .	3
1.2 Rheumatic diseases . . . . .	5
1.2.1 Rheumatoid arthritis . . . . .	5
1.2.2 Osteoarthritis . . . . .	9
1.3 Radiography . . . . .	10
1.4 Analysis of joint damage in radiographs . . . . .	14
1.5 Research objective . . . . .	17
1.6 Outline . . . . .	17
<b>2 An overview of automated scoring methods for RA</b>	<b>19</b>
2.1 Introduction . . . . .	19
2.2 Methods . . . . .	20
2.3 Historical overview . . . . .	20
2.4 Image processing methods for RA assessment . . . . .	22
2.4.1 Detection and segmentation . . . . .	22
2.4.2 Joint space width (JSW) measurement . . . . .	24
2.4.3 Bone damage assessment . . . . .	25
2.5 Discussion . . . . .	25
<b>3 Quantifying joint space width</b>	<b>29</b>
3.1 Introduction . . . . .	29
3.2 Previously described methods . . . . .	30
3.3 Evaluation of methods . . . . .	32
3.3.1 Joint margin data . . . . .	33
3.3.2 Number of measurements . . . . .	33
3.3.3 JSW region . . . . .	35
3.3.4 Measurement lines . . . . .	35
3.3.5 Comparing methods . . . . .	39
3.3.6 Other measurements . . . . .	42
3.4 Discussion . . . . .	42
3.5 Conclusion and recommendation . . . . .	46

---

<b>4</b>	<b>Segmentation of bones in hand radiographs</b>	<b>47</b>
4.1	Introduction . . . . .	47
4.2	Active appearance model (AAM) of the hand . . . . .	49
4.2.1	Dataset . . . . .	50
4.2.2	Landmarks . . . . .	50
4.2.3	Shape vector . . . . .	51
4.2.4	Overall alignment . . . . .	52
4.2.5	Modeling shape . . . . .	54
4.2.6	Modeling texture . . . . .	55
4.2.7	Combining shape and texture . . . . .	57
4.2.8	Connected submodels . . . . .	59
4.2.9	Multi-model search strategy . . . . .	60
4.2.10	AAM search . . . . .	64
4.3	Results . . . . .	65
4.4	Discussion and conclusion . . . . .	68
<b>5</b>	<b>Biometrics of the hand skeleton</b>	<b>71</b>
5.1	Introduction . . . . .	71
5.2	Methods . . . . .	72
5.2.1	Data . . . . .	72
5.2.2	Biometric features . . . . .	72
5.2.3	Classification . . . . .	75
5.3	Experiments and results . . . . .	77
5.3.1	Cross verification of single hands . . . . .	78
5.3.2	Matching opposing hands . . . . .	78
5.3.3	Cross verification of combined hands . . . . .	80
5.4	Discussion and conclusions . . . . .	80
<b>6</b>	<b>Margin detection</b>	<b>83</b>
6.1	Introduction . . . . .	83
6.2	Joint margin detection . . . . .	84
6.2.1	Image data set . . . . .	85
6.2.2	Initialization . . . . .	85
6.2.3	Margin shape . . . . .	85
6.2.4	Margin detection . . . . .	88
6.2.5	Search . . . . .	90
6.2.6	Convergence . . . . .	91
6.2.7	Distance measure . . . . .	91
6.3	Experiments and results . . . . .	92
6.3.1	Margin detection . . . . .	92
6.3.2	JSW measurements . . . . .	93
6.4	OMERACT exercises . . . . .	94
6.5	Conclusions . . . . .	95

---

<b>7</b>	<b>Acquisition variability and JSW measurements</b>	<b>99</b>
7.1	Introduction . . . . .	99
7.2	Analysis of the acquisition setup . . . . .	100
7.3	Simulated projection images . . . . .	102
7.3.1	Method . . . . .	103
7.3.2	Results . . . . .	105
7.4	Conclusion . . . . .	105
7.5	Recommendation: a positioning aid . . . . .	106
<b>8</b>	<b>Revealing radiographic changes</b>	<b>113</b>
8.1	Introduction . . . . .	113
8.2	Subtraction of radiographs . . . . .	114
8.2.1	Image registration . . . . .	114
8.2.2	Intensity transformation function . . . . .	115
8.3	Results . . . . .	119
8.4	Discussion . . . . .	120
<b>9</b>	<b>Conclusions and recommendations</b>	<b>123</b>
9.1	Conclusions . . . . .	123
9.2	Recommendations . . . . .	125
	<b>Bibliography</b>	<b>127</b>
	<b>List of publications</b>	<b>137</b>
	<b>Summary</b>	<b>139</b>
	<b>Samenvatting</b>	<b>141</b>
	<b>Dankwoord</b>	<b>143</b>



# Nomenclature

## Abbreviations

AAM	Active appearance model
ANA	Antinuclear antibody
ASM	Active shape model
AUC	Area under curve
BMD	Bone mineral density
CMC	Carpometacarpal
CT	Computed tomography
DEXA	Dual energy x-ray absorptiometry
DIP	Distal interphalangeal
DMARDs	Disease-modifying antirheumatic drugs
DXA	Dual x-ray absorptiometry
DXR	Digital x-ray radiogrammetry
EER	Equal error rate
ESR	Erythrocyte sedimentation rate
FNR	False negative rate (1-sensitivity)
FPR	False positive rate (1-specificity)
HPA	Hand positioning aid
JSW	Joint space width
MCP	Metacarpophalangeal
MRI	Magnetic resonance imaging

MTP	Metatarsophalangeal
NSAIDs	Nonsteroidal anti-inflammatory drugs
OA	Osteoarthritis
OMERACT	Outcome Measures in Rheumatology Clinical Trials
PA	Posteroanterior
PCA	Principal component analysis
PIP	Proximal interphalangeal
RA	Rheumatoid arthritis
ROC	Receiver operating characteristic
ROI	Region of interest
SD	Standard deviation
SHS	Sharp/van der Heijde score [1]
SVD	Singular value decomposition
TNR	True negative rate (specificity)
TPR	True positive rate (sensitivity)

# 1

## Introduction

The first radiograph of a human body part was made by Wilhelm Conrad Röntgen of his wife's hand (Figure 1.1). His discovery of x-rays in 1895 marks the beginning of radiology, a field of medicine that would become indispensable to (non-invasive) diagnostics. It is not remarkable that Röntgen chose a human hand as a subject to demonstrate his invention. A hand appeared relatively easy to image with x-rays due to its size and the slimness of its bones and tissue. Another, and maybe an even more important aspect is that the hand is particularly appealing to one's imagination. It consists of a large number of bones and joints which together enable a complex set of functions. Our hands are our main tools, we can coordinate their movements with great precision and flexibility in combination with considerable strength. Touching, grabbing, holding and moving things around are common functions that we need while performing our daily tasks and work. Besides for practical tasks, we also use our hands for social interactions, for example when shaking hands or making gestures while we talk. Since we use our hands for so many things, they are extremely valuable to us, and any discomfort to them soon affects our daily life.

Unfortunately, taking good care of our hands and avoiding dangerous tasks does not guarantee a lifelong, problem-free use of our hands. Rheumatoid arthritis (RA) and osteoarthritis (OA) are well-known examples of rheumatic diseases that can cause pain and severe damage to joints in the entire body. Often the first signs of these diseases are noted in the joints of the hands and feet. Besides pain and swelling noted by the patient, there are also effects that can be better seen on a radiograph. As already observed by Röntgen, x-rays provide an excellent means to visualize skeletal structures. Even nowadays, with newer 3D imaging techniques

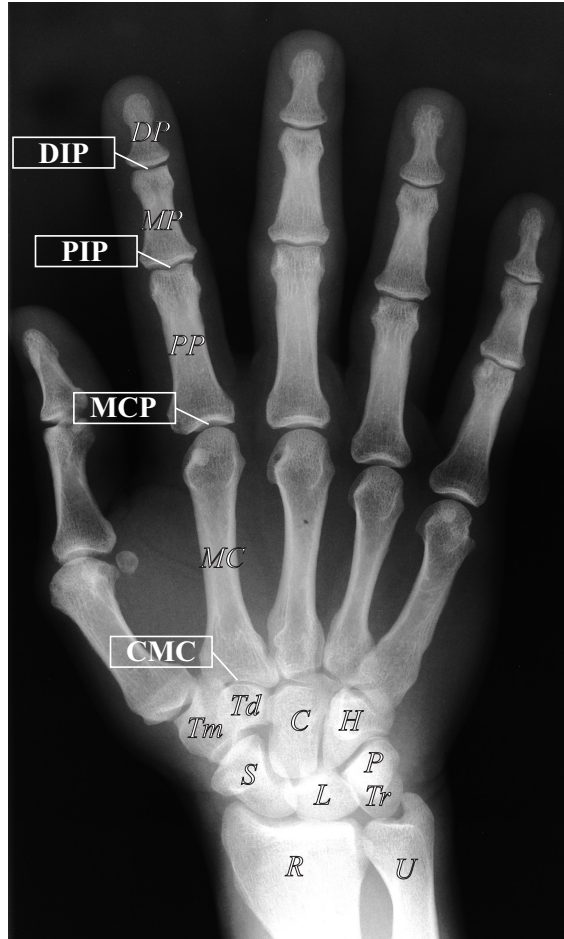
available such as MRI and CT, plain 2D radiographs play an important role in diagnosing and monitoring rheumatic diseases. The value of imaging techniques can be further increased by using the computer for image processing and visualization. By using digitized radiographs it is possible to make complex measurements and to automate time-consuming tasks. Though various efforts are being pursued, currently such techniques are not yet a common practice in rheumatology. In this thesis we investigate how various image processing techniques can be applied to assess bone damage. We have specifically focused our efforts on radiographs of the hands. However, most subjects and methods that we address in this thesis are also applicable to radiographs of the feet (and possibly also other body parts).

This introduction continues with some background information to support the topics of this thesis. In the next section, Section 1.1, we present a radiograph of the hand skeleton and list the names of the bones and joints that are relevant to hand radiography. Next, Section 1.2 provides an introduction to the rheumatic diseases RA and OA. Before going into detail about hand radiography, the basic principles of radiography are explained in Section 1.3. In the following Section 1.4 it is explained which aspects of hand radiographs are of interest for the assessment



*Figure 1.1: First radiograph of a human body part made by Wilhelm Conrad Röntgen. (Source: Reynolds Historical Library)*





*Figure 1.2: Bones and joints of the hand. The labels refer to the abbreviations of the bones and the joints listed in Tables 1.1 and 1.2*

of bone damage and disease activity. Section 1.5 and Section 1.6 present the research objectives and outline of this thesis.

## 1.1 Bones of the hand

The human hand consists of 27 bones (excluding sesamoid bones, which are described further on), 19 bones in the fingers and 8 bones in the wrist. Figure 1.2 shows a radiograph in which all hand bones are visible. Anatomically the fingers

<b>Part</b>	<b>Bone name</b>	<b>Abbreviation</b>
Fingers	Distal phalanx	DP
	Middle or intermediate phalanx	MP
	Proximal phalanx	PP
	Metacarpal bone	MC
Wrist	Trapezium	Tm
	Trapezoid	Td
	Capitate	C
	Hamate	H
	Scaphoid	S
	Lunate	L
	Triquetrum (Triangular)	Tr
Forearm	Radius	R
	Ulna	U

*Table 1.1: Names of the bones in the hand.*

<b>Joint name</b>	<b>Abbreviation</b>
Distal interphalangeal joint	DIP
Proximal interphalangeal joint	PIP
Metacarpophalangeal joint	MCP
Carpometacarpal joint	CMC

*Table 1.2: Names of joints in the hand.*

are numbered 1–5 starting with the thumb. Each finger, with exception of the thumb, consists of one metacarpal (MC) bone and three phalanges. The thumb differs in that it lacks a middle phalanx. The phalanges are named with the attributes proximal, middle and distal, indicating their location with respect to the body. The metacarpals connect the phalanges with the wrist (or carpal) bones. The carpus consists of eight small bones and is connected to the radius and ulna of the lower arm. Table 1.1 lists the names of the hand bones and their abbreviations. The joints between the phalanges are named interphalangeal joints. The knuckles, the joints between the metacarpals and the phalanges, are the metacarpophalangeal joints. The carpometacarpal joints connect the metacarpals with the carpal bones. In Figure 1.2 the locations of these bones and joints have been indicated. The abbreviations of the joints have been listed in Table 1.2.

Besides the aforementioned bones, several sesamoid bones are often visible in a hand radiograph. The number and locations of these small bones vary between persons. Usually, two can be found near the first MCP joint, one or two near MCP-2 and another near MCP-5. Sometimes they are also present near one of

the other MCP joints, near the interphalangeal joint of the thumb, or near the second DIP joint. Sesamoid bones are embedded within the tendons passing over the joint. Their function is to protect the tendon and to change its angle [2].

## 1.2 Rheumatic diseases

Rheumatism is a non-specific term referring to a variety of disorders marked by inflammation, degeneration, or metabolic derangement of connective tissue structures [3]. Especially the joints, but also organs such as the heart, kidneys, lungs and skin can be affected. The most common rheumatic disorders are RA and OA. Other examples are bursitis, fibromyalgia and ankylosing spondylitis.

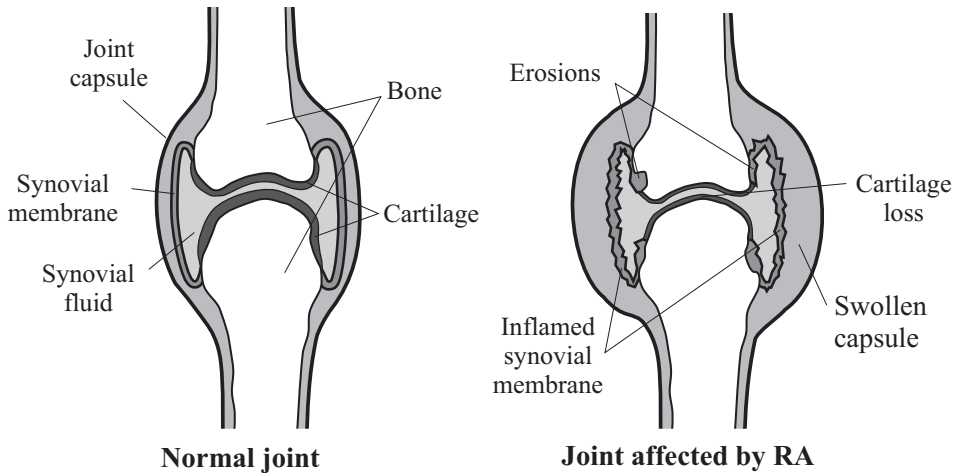
In our study the focus is on hand radiographs of patients with RA. As the joint damage caused by OA is in some aspects similar to that observed with RA, various subjects and methods discussed in this thesis are also relevant for OA. In the following two subsections both diseases are described.

### 1.2.1 Rheumatoid arthritis

RA is a chronic systemic inflammatory autoimmune disease that causes pain, swelling and stiffness in synovial joints (Figure 1.3). Multiple joints are usually affected in a symmetric pattern on both sides of the body. Commonly affected joints by RA include the hands, feet, elbows, shoulders, neck and ankles. In addition, multiple organ systems can be affected. The estimated prevalence rate is approximately 1% worldwide, with women more than twice as likely to develop the disease as men [4]. RA can occur at all ages, but often the onset is between the ages of 30 and 50. The cause of RA is still unknown, but it is suspected that genetic, environmental, hormonal and infectious factors play a role [4]. The disease activity usually changes over time, the degree of tissue inflammation decreasing and symptoms disappearing for a period of time.

**Pathophysiology** Although the generation and development of RA is still not fully understood, it is suspected that it is initiated by a T-cell reaction to an (as yet unidentified) antigen [5]. T-cells are a type of white blood cells that play an important role in the control of an immune response. These cells produce T-cell cytokines (proteins that serve as chemical messengers between cells) which lead to the recruitment of inflammatory (white blood) cells, including neutrophils, macrophages and B-cells. It is suspected that B-cells make a significant contribution to the inflammatory process, as they produce autoantibodies known as rheumatoid factor. These proteins form immune complexes which lead to a further increase of the inflammation.

Normal synovial tissue consists of an intimal lining of one to three cell layers and the synovial sublining which connects with the joint capsule. The intimal lining consists mainly of macrophages and fibroblasts. The sublining contains scattered



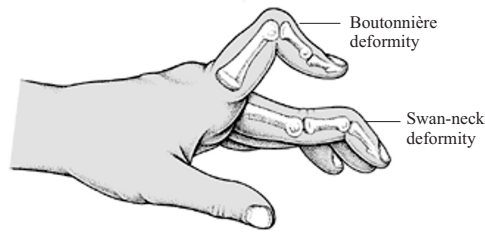
*Figure 1.3: A normal synovial joint and one affected by RA.*

blood vessels, fat cells and fibroblasts. Macrophages are large white blood cells that destroy foreign and potential harmful particles. Fibroblast cells can form connective tissue and lubrication ingredients for the synovial fluid and cartilage surface. During the inflammation, the number of cell layers (macrophages and fibroblasts) in the intimal lining of the synovium increases and new networks of small blood vessels are formed in the synovium.

In the following phase, the inflamed synovium begins to grow irregularly, and through several mechanisms between macrophages and fibroblasts bone resorptive cells named osteoclasts are formed. Osteoclasts can produce enzymes named matrix metalloproteinases, which are thought to be largely responsible for cartilage and bone degradation in RA [5]. At the synovial interface with the bone, the synovial tissue can become invasive, forming of a mass of tissue called pannus. This process leads to joint erosions (Figure 1.3).

Further joint destruction is caused by proteins released by white blood cells. Over time, also other tissues around the joint, such as ligaments, tendons and muscles can become inflamed. As the cartilage lining of a joint degrades and the bone surface erodes, the range of movement of the joint becomes impaired and deformity occurs.

Typical deformities for the hand are ulnar deviation of the fingers, Boutonnière deformity (hyperflexion at the PIP joint with hyperextension at the DIP joint), and swan-neck deformity (hyperextension at the PIP joint, hyperflexion at the DIP joint) [6]. The thumb may develop a subluxation and fixed flexion at the MCP joint, and hyperextension at the interphalangeal joint. Figure 1.4 illustrates both Boutonnière deformity and swan-neck deformity. A typical RA hand is depicted



**Figure 1.4:** Typical RA deformities: Boutonnière deformity and swan-neck deformity. (Source: Merck&Co., Inc. <http://www.merck.com>)



**Figure 1.5:** Typical appearance of a hand affected by RA: swelling and dislocations of joints, ulnar deviation of the fingers and deformity of the little finger. (Source: CH8 <http://www.ch8.ch>)

in Figure 1.5.

**Diagnosis** Commonly a diagnosis begins with a review of the history of symptoms of the patient and an examination of the joints for inflammations, deformities and the presence of rheumatoid nodules [7, 8]. Also other parts of the body are examined for inflammations. The diagnosis of RA is usually based on a combination of symptoms, including the distribution of the inflamed joints, and the blood and x-ray findings.

There are several blood tests that play a role in diagnosing RA. Some of these tests can be used to detect abnormal antibodies, such as the rheumatoid factor, which can be found in 80% of the patients [9]. Other abnormal antibodies that frequently present in RA patients are citrulline antibodies and antinuclear antibodies (ANA) [7].

The sedimentation rate (ESR) is a blood test which measures how fast red blood cells reach the bottom of a vertical test tube. The ESR is usually faster during any inflammatory activity in the body, including joint inflammation. This

method is considered to be a crude measure [7]. Another blood test for measuring the disease activity is based on the increased presence of the C-reactive protein.

The results of the aforementioned blood tests can also be abnormal in other systemic autoimmune and inflammatory conditions. Therefore these tests alone are not sufficient for a reliable diagnosis of RA.

Besides the blood, also the synovial fluid can be examined by means of arthrocentesis. In this procedure the doctor uses a needle and syringe to drain some synovial fluid out of the joint. This fluid can be analyzed to exclude other possible causes of inflammation, such as infection and gout. Sometimes arthrocentesis is also used to relieve joint swelling and pain.

In an early stage of RA, radiographs of joints may be normal or only show swelling of soft tissues. As the disease progresses, narrowing of joint space and erosions may become visible. Also the bone structure and the bone mineral density (BMD) may change. Radiographic analysis is discussed in more detail in Section 1.4.

**Treatment** Currently there is no known cure for RA and treatments are mainly based on pain relief, reduction of inflammation and restoration of function. Two classes of medication are used in treating RA: anti-inflammatory agents and disease-modifying anti-rheumatic drugs (DMARDs) [7]. These DMARDs slow down the disease progress.

The group of nonsteroidal anti-inflammatory drugs (NSAIDs), such as diclofenac and ibuprofen, belong to the first class. These drugs are ‘fast-acting’ and reduce pain and inflammation. There are more than ten NSAIDs, which may differ in effectiveness and side effects per patient. When NSAIDs are ineffective, or during severe flares of disease activity, corticosteroids are commonly used. Well-known examples are prednisone and triamcinolone. Administration of these medications is usually orally, but sometimes by injection directly into tissues and joints. Corticosteroids are very effective in reducing inflammation, and in restoring joint mobility. Unfortunately the effects last for a relatively short period and there can be serious side effects.

In the past corticosteroids were seen as part of the first line of medication. In the last decade it has been found that, especially in early RA, low dosages are effective in disease control and limit joint destruction [10]. Therefore they are now considered as DMARDs.

Other DMARD examples, which prevent joint destruction, but are not directly anti-inflammatory, are gold salts, methotrexate and hydroxychloroquine. These medications are considered to be ‘slow-acting’, as they typically take weeks or months to become effective. Furthermore, newer biologic agents are now available that block the effects of specific proteins that trigger and sustain the inflammation response. Administration of these agents is usually intravenous, and they can be combined with other medications [7].

Besides medication, also exercise is an important part of RA treatment. This

is to maintain fitness of the muscles and to preserve joint mobility and flexibility.

In an advanced stage of RA surgery may be recommended to restore mobility. Such procedures can range from tissue repair to partial or complete replacement of the joint.

### 1.2.2 Osteoarthritis

OA, also known as degenerative arthritis, is a chronic degenerative joint disease in which low-grade inflammation results in the breakdown and loss of cartilage. OA commonly affects the hands, feet, spine, and the large weight-bearing joints, such as the hips and knees. As the disease progresses, the affected joints appear larger, and become stiff and painful.

The exact cause of OA is not yet known. Often multiple members of the same family are affected, suggesting genetic factors to play an important role [11]. Also severe stress on joints due to obesity or heavy work is related to OA. Other suspected causes include repeated trauma or surgery to the joint structures, abnormal joints at birth, gout, diabetes and other hormone disorders. In the Netherlands, one in thirteen persons has OA [12]. OA can occur at all ages, but is most common at ages above 45.

**Pathophysiology** In the first stage of OA, the water content of the cartilage decreases and the protein production decreases [11]. This makes the cartilage less resilient and vulnerable to degradation. Eventually, cartilage begins to break down and small cracks are formed. When breakdown products from the cartilage are released into the synovial space, this can result in inflammation of the surrounding joint capsule. This inflammation is generally mild compared to that which occurs in RA. Over time, loss of cartilage causes friction between the bones, leading to pain and limitation of joint mobility. Often these effects are worsened by the growth of spurs near the joint margins. These bone outgrowths are induced by the inflammation of the cartilage. Examples of such spurs are Heberden's nodes and Bouchard's nodes, which are located at the distal interphalangeal joints and the proximal interphalangeal joints respectively [11].

**Diagnosis** The diagnosis of OA is usually done by reviewing the history of symptoms of the patient, followed by an examination of inflammation and deformity of the joints. Characteristic for OA is that pain in the joints increases with their use throughout the day. This distinguishes OA from RA, as with RA the pain and stiffness is usually severer in the morning. Further diagnosis can be done through x-rays, by which spurs and joint space narrowing can be detected. OA itself cannot be detected by blood tests, though often blood tests are done to exclude other causes such as RA or gout [11].

**Treatment** The damage caused by OA is irreversible, and typical treatment consists of medication or other interventions that can reduce the pain of OA and thereby improve the function of the joint. In many cases a mild analgesic (pain-reducer) is sufficient. In more severe cases, NSAIDs are often prescribed to reduce pain and inflammation. Occasionally corticosteroids are injected in the larger joints, but the benefits of this treatment do not always outweigh the risks and side effects [11].

Sometimes surgery can be used to realign deformed joints by bone removal. In severe cases joints can be fused or replaced with an artificial joint.

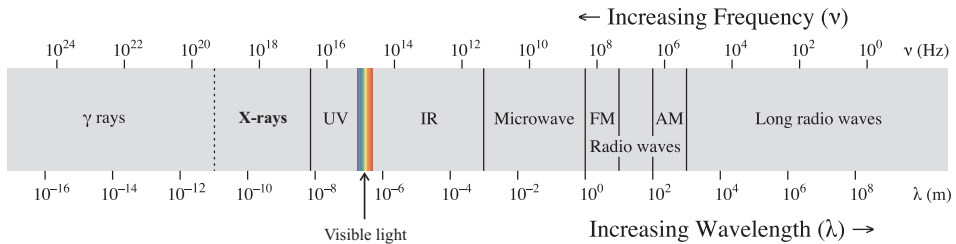
### 1.3 Radiography

X-rays, or Roentgen-rays, are generally defined as electromagnetic radiation with wavelengths between 0.01 and 10 nanometers (see Figure 1.6). This radiation can be produced by accelerating electrons with an electric field in order to collide with a metal target (the anode) such as tungsten or molybdenum. On collision with a metal atom, a bound electron from the inner shell can be knocked out. The created vacancy is subsequently filled by an electron from an higher energy level and simultaneously an x-ray photon is excited. Figure 1.7 illustrates how this process is achieved in an x-ray tube.

The energy of a photon can be calculated by:

$$E = \frac{hc}{\lambda} \quad [\text{eV}], \quad (1.1)$$

where  $h$  is Planck's constant ( $4.136 \times 10^{-15}$  eVs),  $c$  is the speed of light in [m/s], and  $\lambda$  is its wavelength in [m]. The spectrum of the excited radiation depends on the strength of the applied electric field (tube voltage  $U_2$  in Figure 1.7) and the type of metal used for the anode. Figure 1.8 shows an approximate of the spectrum for a tungsten tube with a tube voltage of 100 kVp. Evidently the maximum photon energy is limited to 100 keV. The lowest energy photons are filtered by the tube, and the highest intensity can typically be found at approximately one third of



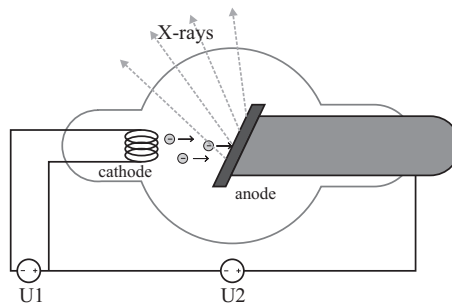
**Figure 1.6:** The electromagnetic spectrum [13].



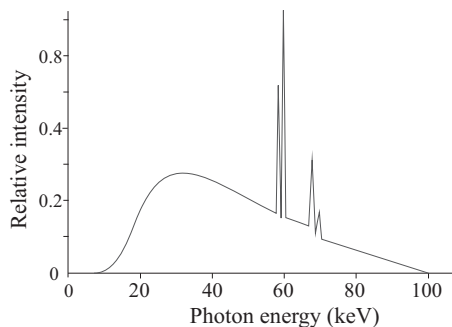
the total spectral range. The high peaks are located at the energy levels that are characteristic for the electron shells of the anode material [14]. The total intensity of the x-ray beam is determined by the electron flow (or current) from the cathode to the anode.

## Interactions

X-rays can be characterized as energetic particles or waves that are able to ionize an atom or molecule through atomic interactions. In radiography, there are two types of interaction between x-rays and matter. The first occurs primarily with lower energy x-rays and is known as the photoelectric effect. This effect takes place when the energy of an x-ray photon is transferred to an entire atom. If the photon has enough energy to eject one of the electrons from the atom's inner shells, the residual energy will be transferred to the ejected electron in the form of kinetic



**Figure 1.7:** Schematics of an x-ray tube. Source U1 controls the number of excited electrons at the cathode. Source U2 applies an electric field to accelerate electrons in the direction of the anode.



**Figure 1.8:** X-ray spectrum from a tube with a tungsten anode and an applied tube voltage of 100 kVp.

energy.

The second type of interaction is known as the Compton effect. This effect occurs when a high energy x-ray photon collides with an electron in the outer shell of an atom. The electron is freed from the atom and both particles may be deflected at an angle to the direction of the path of the incident x-ray. As the photon has transferred some of its energy, it will continue with a longer wavelength. If enough energy is left in the photon, new interactions may follow. These deflections, accompanied by a change of wavelength, are known as Compton scattering. In radiography Compton scattering can cause a decrease of image contrast and an increase of noise. Severe scattering can be reduced by using an anti-scattering grid which absorbs photons coming from other directions than from the source.

Both interaction types contribute to the overall attenuation of x-rays in a material. In general, the chance for interactions increases for higher density materials, hence the attenuation of these materials is higher. For higher energy photons the attenuation is generally less. X-ray attenuation in a material can be modeled by

$$I/I_0 = e^{-\mu t}, \quad (1.2)$$

with  $I_0$  the incident intensity (proportional to the number of photons),  $I$  the measured intensity transmitted through a layer of material with thickness  $t$  in cm and linear attenuation coefficient  $\mu$  in  $\text{cm}^{-1}$ . In literature the latter material property is often represented by the mass attenuation coefficient  $\mu/\rho$ , where  $\rho$  is the density in  $\text{g}/\text{cm}^3$ . In this case the mass thickness  $x = \rho t$  in  $\text{g}/\text{cm}^2$  is often used:

$$I/I_0 = e^{-(\mu/\rho)x}. \quad (1.3)$$

Ionizing interactions caused by x-rays can be destructive to biological organisms and can cause DNA damage in individual cells. To protect a patient from unnecessary exposure to x-rays, a thin metallic sheet is commonly placed between the source and target to filter out the lower energy ‘soft x-rays’. Soft x-rays, as opposed to ‘hard x-rays’, do not have sufficient energy to pass through the target and make it to the detector. Therefore they are not practical for imaging and only cause unnecessary dose for the patient.

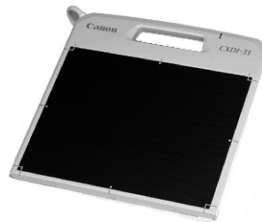
## Detectors

In order to make a projection radiograph, a subject is placed between the x-ray source and a detector. The x-rays that have not been absorbed by the subject interact with the material of the detector generating a projection image. There are several different types of detectors that are used for medical imaging.

*Photographic plates or films* are the oldest detectors, and provide a convenient and easy means of recording projection images. Since photographic films are commonly more sensitive to visible light than x-rays, they are placed between two intensifying screens (converting absorbed x-rays to visible light) and packed



*Figure 1.9: Photostimulable phosphor plates and scanner by Fujifilm.*



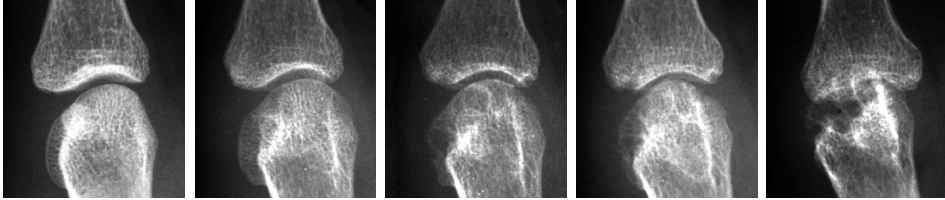
*Figure 1.10: Indirect semiconductor detector panel produced by Canon.*

in a light proof cartridge or paper envelope. After exposure, the films have to be developed chemically in a processing facility. Film radiographs can be digitized by using a transparency scanner or a digital camera. As this process is rather laborious and expensive, these detectors are losing favor.

*Photostimulable phosphor plates* are reusable detectors that contain a special class of phosphors. On interaction with x-rays, electrons are raised to a higher energy state and remain trapped in the materials crystal lattice. To read out the projection image, the detector is scanned by a small laser beam. When exposed to this beam, electrons are freed and light is emitted. This light is collected by a photomultiplier tube and converted to an electric signal which can be digitized directly. This process is also referred to as computed radiography or digital radiography. An example of this system is displayed in Figure 1.9.

Nowadays *indirect semiconductor detectors* use a scintillator screen to convert x-rays to visible light. A large array of small light-absorbing photodiodes attached to this screen converts this light to electric signals which are processed by a computer. This technique is commonly referred to as direct radiography. Figure 1.10 displays an example of this type of detector.

The exposure of the detector (and also the patient) increases approximately



*Figure 1.11: Follow up series of radiographic images of the second MCP joint of a patient with progressive RA (from left to right with approximately two years intervals).*

quadratically with the tube voltage. For a fixed tube voltage and filter, the exposure of the detector is proportional to the tube current multiplied by the time of operation. This number is commonly referred to as the mAs-number (in milliamperere-second), and its setting can be used to adjust the contrast of a radiograph.

## 1.4 Analysis of joint damage in radiographs

Radiographs are particularly suitable to visualize the shapes and structures of objects with strong density variations such as bones surrounded by soft tissue. Soft body tissues like muscles, tendons, ligaments, vessels, and also cartilage are hard to discern from one another, because of their similar densities and mass attenuation coefficients. This also means that it is difficult to identify inflamed tissue when analyzing hand radiographs of patients diagnosed with RA. Although inflammations generally manifest in soft tissue, also the bones and their mutual position become indirectly affected as the disease progresses. Figure 1.11 shows a two-year interval series of radiographs of a second MCP joint affected by RA. As explained, inflamed tissue is not visible in these images. However, one can observe that the texture of the bones gradually changes and severe erosions appear. This damage is caused by invasive pannus tissue and bone degenerative proteins, corresponding to the process described earlier in Section 1.2.1. Another noticeable effect is the mutual position of the bones. As the cartilage degrades, gradually the visible space between the bones narrows and joint luxation (dislocation) occurs. The rate of this process can differ for each joint and may vary over the years. Figure 1.12 shows a radiograph of a hand with severe joint damage in multiple MCP joints. The ulnar deviation of the fingers is typical for RA. Also the wrist has been affected with erosions and joint space narrowing.

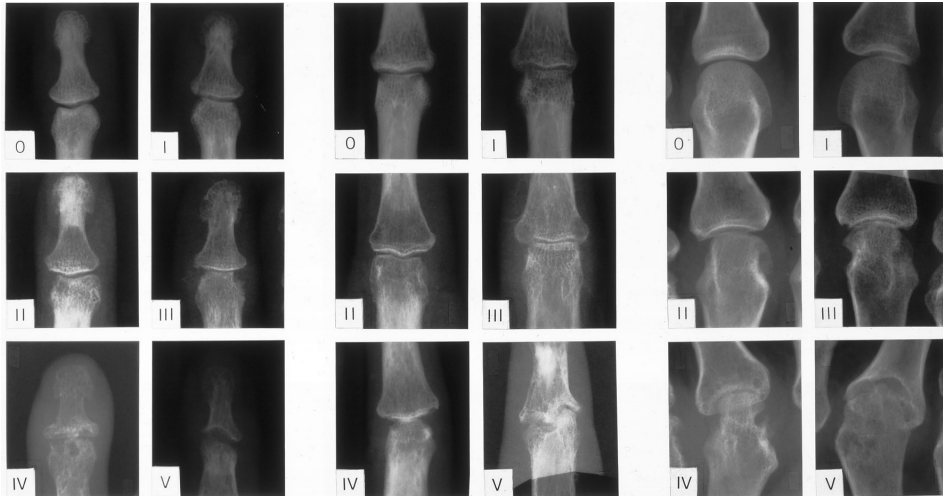
A rheumatologist uses radiographs to support his diagnoses and to examine possible joint damage. When earlier radiographs are available, he will try to estimate the disease activity in order to evaluate the effects of the treatment. Often such estimation is merely based on insight and experience. However, in large scale research, for instance when evaluating drug treatments in clinical trials, there is a



*Figure 1.12: Radiograph A displays a hand with ulnar deviation and severe joint damage caused by RA. Radiograph B displays a normal hand.*

high interest for precise quantification methods that can be used to measure disease progression and activity. For this purpose, several scoring methods have been proposed to quantify joint damage using radiographs [15]. Well-known examples are the Larsen score, the Sharp score, the Sharp/van der Heijde method and the Ratingen score [16, 17, 1, 18]. Typically these methods use a set of graphical examples displaying different disease conditions for a selection of hand and foot joints. Each disease condition is labeled with a value according to the grade of joint damage. A trained observer then evaluates the radiographs by classifying the indicated joints to the given conditions. An overall score can then be determined from the total of values. Figure 1.13 displays an example chart of normal joints that can be used to classify joint damage in finger joints to determine the Larsen score.

Obviously, the aforementioned classification methods are subject to inter-observer and intra-observer variability. For this reason researchers have been



*Figure 1.13: Larsen score chart for the finger joints.*



*Figure 1.14: Measurement of the carpo/metacarpal ratio.*

looking for objective methods based on true measurements. An example of such measurements is the carpo/metacarpal ratio [19]. This ratio is calculated by dividing the length of the carpus, measured from the mid base of the third metacarpal to the volar-ulnar margin of the radius, by the length of the third metacarpal (see Figure 1.14). As the cartilage in the wrist degrades and the small bones become luxated under stress, the wrist becomes more compact and the carpo/metacarpal ratio decreases.

A similar, but more direct approach to determine cartilage loss is to measure joint space narrowing. This effect can already occur in an early stage of RA and is quantified by measuring the change in distance between the bones of a joint over time [20]. This distance is commonly referred to as the joint space width (JSW).

Obviously, the described methods are time-consuming, and subject to errors and subjectivity when performed by human observers. To overcome these problems, various efforts are being made to automate these methods using image processing techniques. A comprehensive overview of methods that have been developed in the past decades is presented in Chapter 2

## 1.5 Research objective

The aim of our research is to develop towards an automated system for scoring joint damage caused by RA using digitized x-rays of hands and feet. To achieve this objective, we address the following research questions:

- Is it possible and feasible to measure joint space narrowing and erosion with sufficient precision and reproducibility to replace measurement by human experts?
- What is the validity of a newly developed score compared to the current gold standard, the Sharp/van der Heijde score?
- What is the optimal combined score for joint damage in hand and feet caused by RA?
- How can an automated measurement system be applied practically within rheumatology?

Our conclusions and recommendation with respect to these questions are discussed in Chapter 9.

## 1.6 Outline

First, in Chapter 2, *Overview of automated scoring methods for RA assessment*, an overview is presented of (partially) automated scoring methods that have been developed in the past. In Chapter 3, *Quantifying joint space width*, we investigate different methods that are used to quantify the JSWs in hand radiographs. We demonstrate that measurement results depend on the applied method and offer a recommendation on which method to use. Chapter 4, *Segmentation of bones in hand radiographs*, presents a method to detect the bones of the hand skeleton in a radiograph. This image processing step is essential for the development of fully automated assessment methods and enables further radiographic analysis. In Chapter 5, *Biometric features of the hand skeleton*, we utilize the shape of the bones as biometric features to identify patients and to verify the integrity of datasets of hand radiographs.

A major challenge in automated RA assessment is JSW measurement. In Chapter 6, *Margin detection*, we present a method to detect the joint margins

in MCP and PIP joints. Subsequently we determine the JSW by calculating the average distance between two margins. As joint space narrowing is generally a slow process, it is important that measurements are precise. In Chapter 7, *Acquisition variability and JSW measurements*, we discuss how acquisition parameters and hand positioning can affect the projection image of the joint space. Besides joint space narrowing, RA can also lead to erosions and changes in bone structure. In Chapter 8, *Revealing radiographic changes*, we show how image subtraction can be used to reveal bone damage, and explain how this method can be used to quantify bone loss. Finally in Chapter 9 we present the *Conclusions and recommendations* that follow from this thesis.



# 2

## An overview of automated scoring methods for RA

### 2.1 Introduction

Rheumatoid arthritis (RA) is one of the most common autoimmune diseases. It is a chronic systemic inflammatory disorder that commonly affects the joints, particularly in the wrist, fingers and toes. Besides the joints, also other parts of the body can be affected by RA. Since there is no proven cure for RA available yet, current treatments mainly focus on pain relief, inflammation reduction, and slowing down or stopping the process of joint damage. To prevent irreversible joint damage, it is essential to detect RA at an early stage. To assess effectiveness of drug-treatment it is necessary to monitor the progression of the disease. Radiographs of hands and feet are often used to monitor the progression of joint damage caused by RA. Several scoring methods have been proposed to quantify joint damage using these radiographs [15]. Some make use of classification scores for joint erosions and deformations, for example the Larsen score, the modified Larsen score, the Sharp score, the Sharp/van der Heijde method and the Ratingen score [16, 21, 22, 17, 1, 18]. Other methods are based on relative or absolute measurements, for example by determining the carpal/metacarpal ratio, the JSW and erosion volume [17, 19]. In general these methods are time-consuming and depend on subjective visual readings [23]. In an early stage of RA it is important that the applied scoring method is sensitive to small changes over time, so the effects of medication can be monitored closely and treatments altered if necessary. Several

studies have been conducted on this subject [24, 25, 26, 27].

To eliminate observer dependency and to make the assessment procedure faster, more accurate and affordable, computer-assisted analysis may contribute to better disease treatment. In the field of rheumatology several research groups (including ours) have been inspired by these possibilities and have investigated the use of image processing techniques to analyze radiological joint damage. The aim of this chapter is to present a survey of the image processing methods that have been developed during the past 20 years in the field of joint damage assessment in radiographs of hands and feet. We consider the following image processing operations as relevant: image enhancement, segmentation, JSW measurement, erosion estimation and morphology analysis. Since all developed programs only exist in an experimental setup and are not publicly available, comparisons have to be based on published reports. In the next section we explain the applied methods to find information related to this subject. Subsequently, in Section 2.3 we present a historical overview of the topics that have been addressed by the various researchers. These topics are grouped and discussed more elaborately in Section 2.4. In the final section we discuss the importance of digitized radiographs and how to continue in future research.

## 2.2 Methods

We have consulted the following reference databases to find relevant information: PubMed, a service of the National Library of Medicine which includes citations from the Medline database (<http://www.pubmed.org>), Thomson's ISI Web of Knowledge (<http://isi4.isiknowledge.com>) and Elsevier's ScienceDirect (<http://www.sciencedirect.com>). The time span for the database searches extends from January 1985 to July 2006 and the used keywords are: rheumatoid arthritis, osteoarthritis, arthritis, computer-aided diagnoses, hand radiography, radiography, image analysis, medical imaging, joint space, scoring methods, segmentation and X-ray. Additional information was found through cross-references and with Google's internet search engine (<http://www.google.com>).

## 2.3 Historical overview

In the past twenty years, several groups have been searching for methods to analyze joint damage in RA radiographs. Various efforts have been made to automate JSW measurements for hand radiographs. Also methods for analyzing morphology and structural characteristics of bones have been investigated. In this section we present a brief chronological overview of what has been achieved in the field of computerized RA assessment during this period. Later, in Section 2.4, we categorize the various methods and discuss them in more detail.

*1986* One of the first reports of computer assistance with RA analysis originates

from Buckland-Wright et al. [28]. They use a digitizer tablet in combination with a magnified stereoscopic view of microfocal radiographs of hands and wrists to measure the erosion area. They show that measurements can be done with good accuracy.

*1987* One year later Browne, Gaydecki and colleagues describe an image processing method to measure changes in bone density and shape of the proximal phalanges [29, 30].

*1989* Dacre and colleagues introduce a new radiographic scoring system [20]. They use digital image analysis to measure the JSW in knee radiographs of patients with RA.

Michael and Nelson presented a model-based system for automatic segmentation of bones from hand radiographs [31]. The objective of this experimental study is to measure bone growth.

Allander et al. publish their research about measuring JSW of metacarpophalangeal (MCP) joints and proximal interphalangeal (PIP) joints [32]. They conclude that repeatability of measurements is better than that of manual methods and is less observer dependent.

*1993* Conrozier and Vignon et al. use a computer program to measure joint surface area and mean JSW at the hip [33]. This program has been developed over the years, and ten years later it is also used for JSW measurement of osteoarthritic knees [34].

James et al. compare computerized JSW measurements with conventional joint space narrowing scores in 1995 [35]. They show that their computerized method increases precision and sensitivity to change.

*1998* Duryea and colleagues describe a method for the segmentation of joint space and phalanx margin locations on digitized hand radiographs [36]. This method is reported to have excellent robustness and is expanded with a JSW quantification method two years later [37]. In a 2003 publication Duryea et al. expand this research to digital tomosynthesis in an attempt to measure erosion volumes [38].

*2000* Sharp et al. publish a study where they compare established scoring methods with two computer based methods; one for measuring JSW and another for erosion volume estimation [39].

*2001* Angwin et al. continue the earlier work of James et al. and further enhance their method for measuring the JSW [40, 35]. They also investigate the reliability and sensitivity for different flexed positions of the hand.

*2003* Wick and Peloschek et al. introduce a software tool for faster and more efficient quantification of RA [41]. In this same year, Langs and Peloschek

publish several papers about locating joints in hand radiographs and the detection of bone contours [42, 43, 44]. They report to have developed robust methods that are accurate and easily transferable to other anatomical structures. In 2005 they start a project to expand their software tool with their developed image processing methods.

Bird et al. use the computer to measure erosion volumes in MRI images. They report that their study demonstrates the feasibility, reliability, and validity of these measurements [45].

Jensen et al. study bone densitometry of metacarpal joints [46]. They conclude that digital x-ray radiogrammetry (DXR) is better than dual x-ray absorptiometry (DXA) for detecting and monitoring periarticular osteoporosis of the metacarpal bone.

## 2.4 Image processing methods for RA assessment

To enable automated assessment of joint damage in radiographs, one has to go through several image processing steps. First, a pre-processing step is often required to prepare the image for further analysis; for example contrast improvement, noise reduction, scaling and the removal of artifacts. Subsequently, the regions of interest have to be detected. For hand radiographs, these are the bones and their joints. This can be a difficult task when severe joint damage is present. Also, non-anatomical objects such as rings and labels may cause problems in region of interest detection. Various image segmentation and edge detection methods can be used to determine the representation of the pixels. After the objects within the image have been determined, measurements can be done such as JSW measurement, erosion estimation, classification of bone structure and morphologic assessment.

### 2.4.1 Detection and segmentation

Within the area of computerized RA assessment few publications describe a fully automated detection and segmentation method. Most implementations require operator input such as the identification of landmarks or the selection of regions of interest (ROI).

Duryea and others are well advanced in developing a fully automated method for RA assessment. They describe a method for the identification of joint space and phalanx margin locations of the distal interphalangeal (DIP), PIP and MCP joints of fingers 2–5 [36]. Their method is specifically designed for analyzing hand radiographs and is based on a priori knowledge of certain image characteristics. They report success rates (based on the number of detections within 5 mm from manual annotation) of 99%–100% on 27 pairs of hand radiographs. However, they also mention that certain radiographs were excluded from their test set, since non-anatomical structures, such as rings and labels, were present in important

parts of the images. In a later publication several improvements have been made to the previous method; by adding a neural network they succeed in detecting carpometacarpal, radiocarpal, and the scaphocapitate joints with success rates of 87%–99% for normal hands and 81%–99% for RA hands [47].

Michael et al. have developed a model-based system for the segmentation of bones [31]. They start with a preprocessing step by applying a model based histogram correction and use a threshold above the gray level of the background to find the shape of the hand. Next they use a priori knowledge to determine regions for the bones of the fingers and the palm. This step requires a standard way of positioning the hand. The bone contours are found using an adaptive contour-tracker that incorporates information about the expected shape of the particular bone. At the time of publication, the described system was under development and preliminary results were obtained from only a few experiments.

Promising methods in the bone densitometry research area have been investigated by Efford and Thodberg et al., who have used active shape models (ASM) to detect the contours of the metacarpal bones [48, 49, 50]. Thodberg et al. report a 99.5% success rate for this method (presumably these results were obtained by means of visual verification). The ASM methods are based on deformable models with statistically trained parameters that control possible shape variability. Thodberg also experimented with active appearance models (AAM), which is a more robust technique, since this also involves object texture information in the model [51]. Unfortunately this publication does not report a success rate for this method. Other research with ASMs has been done by Sotoca et al. who have developed software for computerized bone mass assessment of the metacarpals [52].

Langs et al. have developed an approach based on Gabor jets and local linear mapping nets for locating CMC, MCP and DIP joints [42]. They report success rates between 80% and 97.5% for different joints. This method was tested on a set of 10 images, whereas 30 images were used for training. Later they expand their method with an ASM driven snakes algorithm to segment the metacarpal bones [44]. In this work they note that ASMs are restricted by their training examples and the linearity of the models, which makes it infeasible to detect severe pathological changes as caused by RA. To get around these restrictions they use active contours (snakes) to find local edge structures. Their results are promising and indicate that this method can be used for quantitative assessment of bone erosions.

In our group we have developed a segmentation method based on multiple connected AAMs [93]. We are able to segment the metacarpals and phalanges in radiographs where the finger positioning variability is large. 50 radiographs were used for training the models and 30 for testing. For 73% of the images, the bone contours were found within 0.5 mm, for 93% within 1.3 mm. These results are inadequate for accurate JSW measurements; however, this method can offer a good initialization for further processing steps [98].

### 2.4.2 Joint space width (JSW) measurement

Since hand radiographs are two-dimensional projections of three-dimensional objects, their contents are dependent of positioning and projection angle. To estimate a JSW based on such projection images, one has to determine the locations of the bone edges within the joint. Dacre and others describe the development of a radiographic scoring system for measuring the JSW and joint space area in radiographs of the knee [53]. They require an operator to outline the joint space area with the mouse-pointer and subsequently measure the JSW. Positive results were found in terms of accuracy, speed and reproducibility compared with manual readings.

Allander, Forsgren and others show a similar method for the MCP and PIP joints [32, 54], but use the Sobel edge detection algorithm to detect edges in the joint space area [55]. After manual editing of false and irrelevant edges, they use a distance transform to find a medial axis of distances between the two edges. Using the distance values on the medial axis, they calculate the mean JSW.

Another method for measuring the JSW of the MCP and PIP joints is described in the publication of James and others [35]. For this method an operator has to place three markers to define a radial arc close to the proximal edge of the joint (lateral view). The proximal joint space margin is found by a local edge detection method. By scanning the image intensities radial to the arc and aligning the edge locations of the proximal joint space margin, they obtain a 'straightened out' density profile from which they calculate the mean JSW.

Sharp and others describe several image processing experiments for measuring the JSW of the MCP, intercarpal and radiocarpal joints [39]. Also, they present a method for measuring bone erosion. For the JSW measurements an operator has to select a region of interest that contains the joint to be measured. Within this region an edge finding method marks multiple points on the bone edges. As an alternative an operator can place multiple markers within the joint space region. From these initial markers a curve fitting algorithm fits a fourth-order polynomial to the edges. The average and minimum width are found by calculating the shortest distances for each point along the joint space.

A completely automated system for measuring the JSW is described in a paper by Duryea et al. [37], which was appended to the segmentation method mentioned before [36]. This software program uses features from the gradient profile as inputs for a neural network algorithm and applies multiple iterative correction steps to define the correct edge. In this work the authors report to have found a robust method that is in agreement with established scoring methods.

Angwin and others used custom software for measuring the JSW [40]. Objective was to establish the sensitivity and reliability of PIP and MCP mean joint space measurements. This method is based on that of James et al. [35] and is improved by the employment of a Gaussian distribution to uniquely locate key features in the image; tracking the features to locate continuous joint margins; and determination of mean JSW based on averaging measurements of JSW at 180 locations equally spaced across the breadth of the joint. The MCP joints

are located by positioning 3 user-inputs along the metacarpal head. The average distance is measured along the radius from the midpoint of the metacarpal head, which is similar to the technique used by Conrozier et al. for measuring the JSW in hip [56]. The PIP joints are located by selecting a rectangular region of interest. The average distance is measured by sampling parallel lines vertical across the joint (fingers pointing upward).

In our group, we have developed a margin detection method for the MCP and PIP joints based on ASMs [98]. With this method the joint margins are detected as curves defined by 25 equidistant points. Over a breadth of 6 mm we determine the average JSW by determining the point-line distance between the curves of the proximal and distal joint margin. We have found that this detection method has a higher precision considering reproducibility than manual readings.

### 2.4.3 Bone damage assessment

Browne, Gaydecki and others have focused their efforts on morphology and developed a method to detect differences in bone contours and density profiles [30, 29]. This system requires user input for segmentation and coarse edge definition. After these actions, an edge detection algorithm optimizes the bone contours and with these contours multiple features are extracted: bone area, average gray intensity, center of gravity, gravity profile, radial density and contour profile.

The computerized method for measuring erosion volumes in MRI images described by Bird et al. is based upon area measurements within each slice [45]. The erosions are outlined manually and finally the volume is estimated by multiplying the calculated area with the slice thickness. This method is comparable to the earlier technique used by Buckland-Wright et al., who used a digitizer tablet to outline erosions [28].

Jensen et al. used the X-posure System (Sectra Pronosco A/S, Vedbk, Denmark) for their research [46]. This system uses the previously mentioned segmentation method described by Thodberg to detect the shafts of metacarpals 2–4 [49]. They estimate the bone mineral density (BMD) by measuring the outer and inner diameter of the cortical bone. With this method, also known as radiogrammetry [57], the BMD can be determined with a precision of 0.65% [46].

Sotoca et al. determined the bone density of the metacarpals, proximal and middle phalanges by estimating the bone density by comparing the average bone intensity to an aluminum reference wedge placed in the image [52]. Their results show high correlation with different established measurement methods.

## 2.5 Discussion

Two-dimensional projection images of the three-dimensional joint structures are often difficult to interpret. Three-dimensional image modalities are likely to offer more possibilities for measuring erosion volumes and JSW accurately. Despite the

increased availability and quality of three-dimensional imaging techniques such as MRI and CT, plain radiographs are still indispensable. This is because of their superior resolution, the easiness and speed of the acquisition process, and also their low costs. In a comparative study it was found that there is not yet a definite advantage of MRI as compared to radiographic imaging in detecting progression of joint damage [58].

An interesting alternative approach has been demonstrated by Duryea et al. who applied digital tomosynthesis [38]. By using multiple projection images from different angles they are able to reconstruct intersecting image planes of the joints. Especially for detecting erosions this technique could become useful. For JSW measurements they comment that improvements are small as compared to the use of projection radiographs.

Looking at the current state of technology, medical practice and methodology, digitized radiographs are probably favorable for the assessment of joint damage for the upcoming years. Continuing the development of advanced radiographic analysis methods may help to extract more information from such images. To enable automatic assessment of joint damage, it is required that image segmentation is performed in a robust and accurate manner. Understanding the characteristics of the bone shapes and textures is essential for this purpose. Several reported problems with image segmentation are related to the way how images have been acquired. Between and even within datasets there is a large variability in parameters such as resolution, contrast, positioning, cropping, and presence of foreign objects.

In the past few years it has been the trend to use a model based approach using ASM or AAM techniques. A clear advantage of these methods is that they incorporate a priori information, which makes them robust to disturbances such as noise and artifacts. A negative side effect of these methods is that they generally have difficulties with detecting unusual structures such as damaged bones and joints. Several solutions have been presented to relax these statistical constraints by combining these methods with other image processing techniques [59, 44].

When segmentation is performed successfully and all bones have been identified, then regions of interest can be selected for measurements. In case of JSW measurements, the bone outline may not be sufficient to determine the joint space. The projection view of the joint space and overlapping bones may result in ambiguous and even spurious edges. So far, the choice of which edges to select for measurement and how the JSW is determined has been up to the designer of the method. For effective future validation and comparison studies, it is recommendable to define the specific characteristics of the relevant margins. The use of anatomic phantoms may help to identify these properties and can be used to set up a gold standard.

Joint damage may be detected by looking at small indentations and other irregularities in bone outlines. Deviations in the bone mass could indicate erosions, osteophytes and calcifications. Also the texture of the bone may reveal such information. Because of the variability in bone shapes between patients, it is not



straightforward to determine what is healthy or 'normal'. Comparison with previously taken radiographs of the same person may help to detect changes. By overlaying consecutive images taken over a certain period of time small changes in shape or bone density may be detected. Despite the availability of methods to detect changes that could indicate erosions, methods to quantify such effects have not been reported yet. According to several studies erosion volumes can be measured in three-dimensional modalities, but these methods cannot be applied to projection images [28, 45]. Manual, successful methods, such as the Larsen score and the Sharp/van der Heijde score, rely on classification by an expert with a set of example images. This task is difficult to automate, as the variability in erosion appearances are large and their interpretation demands a profound knowledge of hand anatomy and physiology.

Validation of the various methods is essential, to enable practical use of computerized methods in future bone damage assessment. So far many of the presented methods have been tested on small datasets from a limited number of hospitals. Because of the lack of a true gold standard, methods have to be validated with other existing methods (manual or automated). To be able to compare measurement results, it would be useful to develop a standard which defines what should be measured and how this should be done. On the other hand, it is not yet clear which measure is most discriminative for RA. To solve these problems, it is necessary that the various research groups combine their efforts by sharing their data, results and experiences. Currently, serious efforts to such collaboration are made within the special interest group on measurement of joint space and erosion of the international network of Outcome Measures in Rheumatology Clinical Trials (OMERACT; <http://www.omeract.org>) [60].



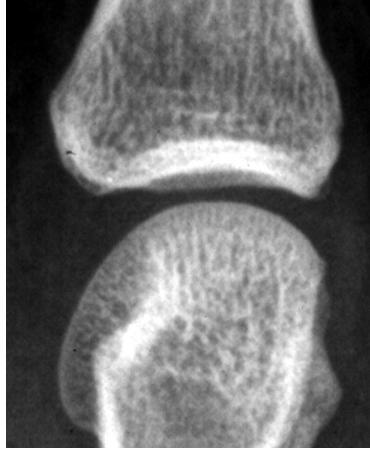
# 3

## Quantifying joint space width

### 3.1 Introduction

In RA and OA, semi quantitative scores have been used for 50 years to measure disease progression and to monitor the effectiveness of treatments [61]. Changes in the thickness of cartilage can be detected in radiographs by measuring the JSW, i.e. the distance between the opposing bones (Figure 3.1). Obviously, in reality the joint space is a 3D space between two bone surfaces. Therefore, ideally the joint space should be measured in 3D using a 3D imaging technique such as CT or MRI [38, 45, 62]. In practice this is not yet feasible, because of the high resolution requirements and the high costs of 3D imaging techniques. Also, for CT the radiation dose is relatively high compared to plain radiography. To depict the joint space in 2D projection radiographs, ideally the projection angle is chosen such that the bones do not overlap and the joint's bone surfaces are visible as sharp edges: the joint margins. Next, the JSW can be estimated by determining the distance between these margins. To be able to compare follow-up radiographs, ideally the positioning of the joint and the projection angle should be the same each time a measurement is done. For hand radiographs, postero-anterior (PA) view is most common with the palmar side of the hand positioned flat on the detector.

In conventional radiography the JSW was estimated visually, which is a time consuming task. Soon the question arose if these measurements could be done more accurately and objectively by using an automated method. Several automated methods aiming to measure JSW of hand joints in millimeters have been developed



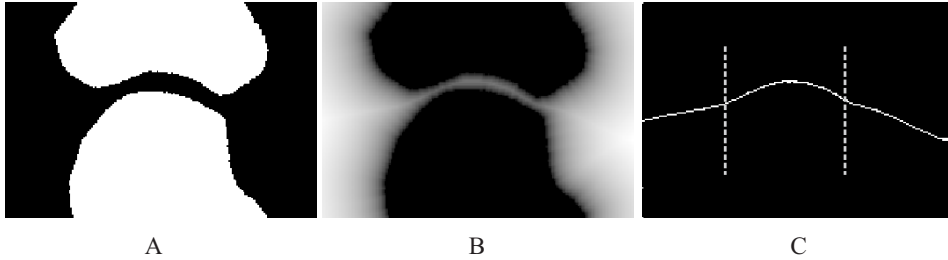
*Figure 3.1: PA projection of a metacarpal joint with the joint space clearly visible.*

with increasing precision [39, 37, 40, 98]. Some have resulted in ‘normal’ JSW values, which may differ according to age, sex and height [63]. A study with repeat radiographs by Angwin and others showed that actual physical changes in JSW of 0.11 mm ( 7%) can be detected for individual MCP and PIP joints [40]. When averaging the measurements across fingers for a single subject the detectable change improves to 0.05 mm ( 3%). According to the results of several studies in early RA the JSW in MCP and PIP joints can decrease at a rate of a few hundredths of a millimeter per year, which provides an indication of the required precision of these measurements [64, 60].

Projects to automate JSW measurement use manual or automated techniques to identify joints on radiographs, and apply an algorithm to outline the joint margins. Next, the JSW is quantified by measuring the minimum or average distance between the joint margins. All of these steps may contribute to the overall precision of a measurement system, making it difficult to compare outcomes of different systems to one another. In this chapter we focus on the final quantification step and assess whether existing methods differ with respect to the resulting JSW. To avoid the influence of variation in the preceding steps, we have used a set of digitized hand radiographs on which the joint margins were delineated manually.

## 3.2 Previously described methods

Allander, Forsgren and others [32, 54] describe a JSW measurement method based on the distance transform of a binary image of a joint. With this method a distance mapping is created where the value of a point represents the distance to the closest joint space margin. Figure 3.2 illustrates this approach. The local maxima between



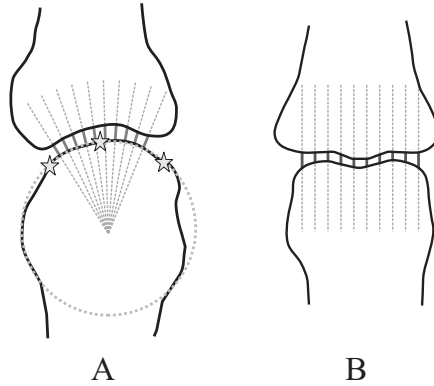
**Figure 3.2:** JSW measurement using the distance transform and medial axis. Binary image A depicts an MCP joint. Image B is the distance transform of A. C shows the local maxima of B as white pixels representing the medial axis. The dashed lines mark the measurement region.

the margins form a ridge which is called the medial axis. By calculating the average of the pixel values of the distance transform at the medial axis, the mean JSW is determined. The measurement region is limited to the points where the angle between the medial axis and the shortest path to a joint margin is less than 85 degrees.

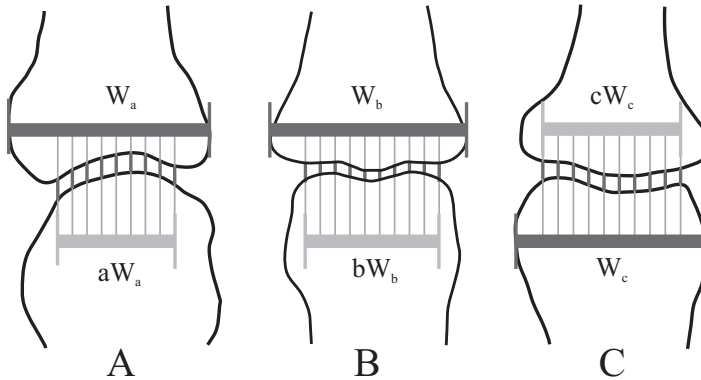
James, Angwin and others [35, 40] use a different approach to measure the metacarpophalangeal (MCP) and proximal interphalangeal (PIP) joints. For the MCP joint, three user-input points along the metacarpal head are used to define a circular arc as illustrated in Figure 3.3A. The middle point is placed at approximately the center of the metacarpal head and identifies the midpoint of the measurement arc. The exact anatomic locations of the other two landmarks are not defined in the description of this method. The mean JSW is determined by measuring the JSW along 180 equally spaced radial lines over a range of 1 radian centered on the midpoint of the measurement arc. The PIP joints are measured by measuring the JSW vertically along equally spaced parallel lines (Figure 3.3B).

Duryea et al. measure the MCP, PIP and distal interphalangeal (DIP) joint spaces [37]. First, they rotate the joints such that the joint space is approximately horizontal in the image. Then the joint is divided into columns, and for each column the distance between the margins is measured. Subsequently, the JSW is calculated by averaging these distances. By using the maximum width of the joint tips and several constants, they define measurement regions for each joint, as shown in Figure 3.4. The horizontal locations of these regions, as well as the applied constants, have been determined empirically from a set of training data.

Sharp et al. describe a method for measuring MCP, intercarpal and radiocarpal joints [39]. First they estimate the shape of the joint space by fitting two fourth order polynomials to detected margin locations (Figure 3.5). Next, the shortest JSW is measured for each point on the upper joint margin. From these measurements they calculate the mean width, the minimum width, and several other



**Figure 3.3:** Three points, marked by stars, on the proximal margin of the MCP joint (A) are used to define a circle. Next, the JSW is determined by measuring along radial lines. For the PIP joints, the JSW is measured vertically along equally spaced parallel lines.

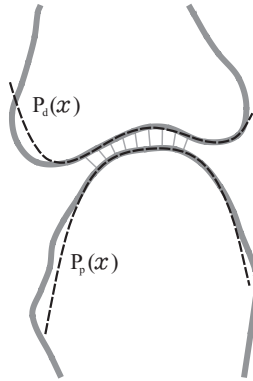


**Figure 3.4:** The size of a joint space region is determined by the maximum width of the joint tips (as indicated by the line segments  $W_a$ ,  $W_b$  and  $W_c$ ) and a multiplication constant (MCP:  $a = 0.58$ , PIP:  $b = 0.68$ , DIP:  $c = 0.74$ ).

figures which provide information about the symmetry of the joint space [39].

### 3.3 Evaluation of methods

Suppose we have detected the joint space margins correctly, then we wish to quantify the distance between these margins. Since we cannot treat the joint space margins as two parallel line segments, it is not straightforward to find an unambiguous method to measure the distance between them. In this section we evaluate various methods for JSW quantification.



**Figure 3.5:**  $P_d(x)$  and  $P_p(x)$  are fourth order polynomials fitted to both margins. For each pixel on the distal margin, the shortest distance to the proximal margin is calculated

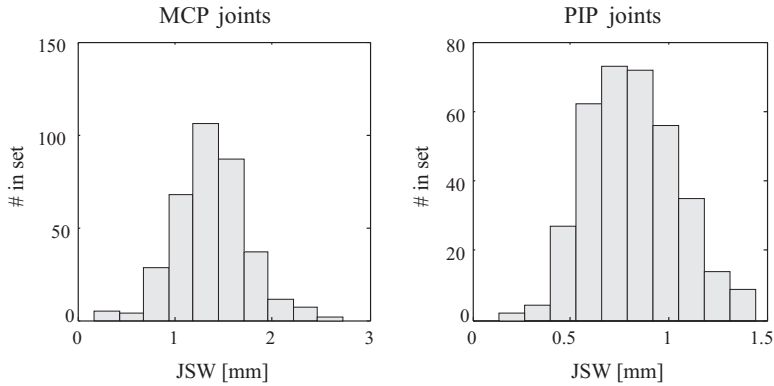
### 3.3.1 Joint margin data

For illustrative purposes and to simulate the effects of several methods, we make use of a data set of joint margins which have been obtained through manual delineation. Forty pairs of hand radiographs of RA patients with variable disease duration and damage were used. Radiographs were made by conventional radiography and scanned at a resolution of 600 dpi. From these radiographs individual images of all 2<sup>nd</sup> to 5<sup>th</sup> MCP and PIP joints were selected. Five MCP and 6 PIP joints were excluded because of severe damage with indiscernible joint margins, leaving 315 MCP and 316 PIP joints for analysis. Joint margins were outlined manually by two trained operators using a software tool developed for this purpose in Matlab. To enable precise measurements for the experiments, piecewise cubic Hermitian interpolation was used to smooth the outlines [65].

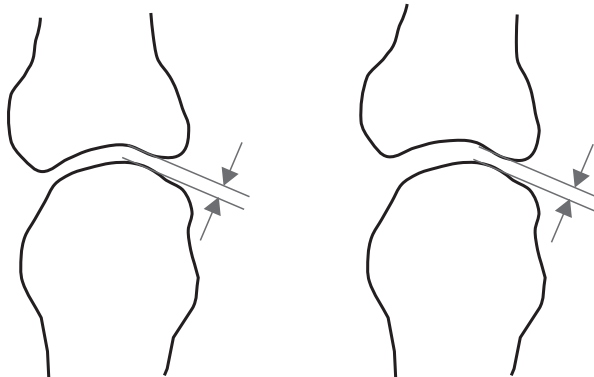
To confirm that sufficient variation in JSW was included in the dataset all joints were measured using method E which is described in Section 3.3.5. MCP JSWs vary between 0.17 and 2.7 mm (mean = 1.37 mm, standard deviation (SD) = 0.38 mm), and PIP JSWs between 0.14 and 1.44 mm (mean = 0.82, SD = 0.23). Figure 3.6 shows the histograms of the MCP and PIP JSW sizes in the data set.

### 3.3.2 Number of measurements

For manual joint space measurements it would be most practical to perform a single distance measurement, for instance to measure the minimum JSW or the width at a fixed location. A disadvantage of a single measurement is that it may not reflect the state of the whole joint space, which can result in a poor sensitivity to change. Figure 3.7 demonstrates this effect. Another disadvantage is that the precision of a single distance measurement is highly dependent on the precision of the detection of the joint margins. Small errors in this detection may result in



**Figure 3.6:** Histograms of the MCP and PIP JSWs in the evaluation data set (0.25 mm bin width).



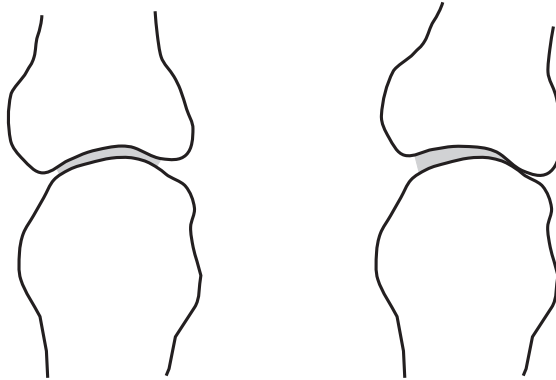
**Figure 3.7:** Two joints with the same minimum JSW.

large deviations in the measurement results.

By averaging multiple JSW measurements over a wide range of the joint space, potential margin detection errors are evened out, resulting in a higher precision. This method does not prevent certain changes from remaining undetectable, as demonstrated in Figure 3.8. Additional measurements are required to detect such changes, as for example the minimum JSW or a measure that describes the overall deviation from the average.

Generally, the precision of a mean JSW measurement increases with the number of measurements. However, the maximum number of measurements is limited by the sample resolution of the margins (which is usually limited by the resolution of the radiograph and the margin detection algorithm). Depending on the applied measurement method, it can be useful to interpolate between consecutive margin points.





*Figure 3.8: Two joints with the same average JSW.*

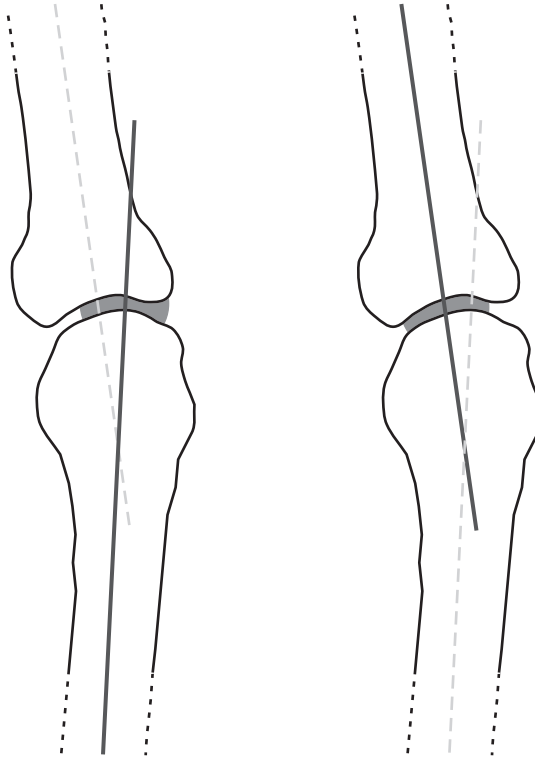
### 3.3.3 JSW region

When measuring a mean JSW, one has to define a region wherein the measurements are done. One method is to define a fixed breadth within which the mean JSW is determined. This width should be as large as possible, while still fitting all possible joint shapes and sizes. To compensate for these differences, a variable width can be used that is based on the size of the joints or the bones, similar to the method of Duryea et al. [37]. The risk of applying such method is that the required additional measure may be affected by measurement deviations, detection errors or changes over time due to bone deformations. Consequently this would decrease the measurement accuracy.

Also, the location of the joint space region has to be defined. For the MCP joints the location of the joint space depends on the degree of abduction of the fingers (PA view). This location moves along with the proximal margin of the proximal phalanx which articulates over the metacarpal head. As illustrated in Figure 3.9, the intersection of the proximal phalanx's midline with the joint space can be used to define the center. Since the PIP (and DIP) joints are rigid from a PA perspective, it makes little difference whether the midline of the proximal phalanx is used or the midline of the middle phalanx. We prefer the midline of the proximal phalanx, as it is already determined for finding the center of the MCP joint. Secondly, as the body of this bone is generally longer than that of the middle phalanx, its midline can be determined more precisely (angular deviations of the midline are likely to be smaller for longer bones).

### 3.3.4 Measurement lines

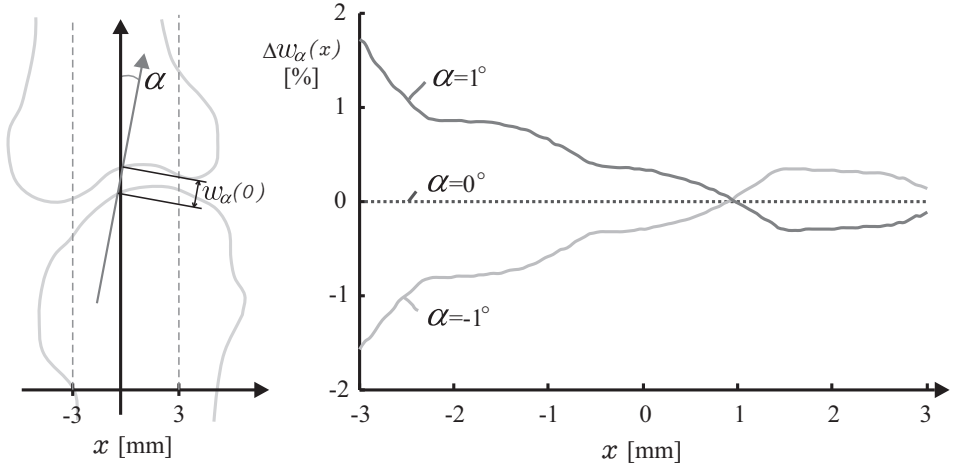
To measure a JSW, an intersecting line can be used to find two points; one on each margin. The distance between these points is then the JSW. In the following we explain how the orientation of such measurement line, with respect to the



**Figure 3.9:** In case of MCP joint abduction, the midline of the proximal phalanx (right) has a better correspondence with the center of the joint space than the midline of the metacarpus (left).

orientation of the joint, is of influence for the measured JSW.

A straightforward method is to measure multiple distances between the margins along parallel lines. If the direction of these lines is chosen nearly perpendicular to both margins, this method works well. But, since the margins are generally curved, and the direction of the parallel lines (with respect to the orientation of the joint) can change between different measurements, this can result in JSW measurement deviations. To demonstrate this effect, we have simulated such measurements on the margins of a typical MCP joint. Figure 3.10 shows the shape of the joint and the change of the measured JSW profile for angular deviations of plus and minus one degree. The joint was rotated with the midline of the metacarpus in the vertical direction. The reference measurement was done along vertical lines. The graph shows the deviations  $\Delta w_\alpha$  in the measured JSW profile over a breadth



**Figure 3.10:** JSW measurements along parallel lines become more sensitive to small angle deviations at locations where these lines are 'less perpendicular' to the joint space margins. The graph shows the change of the JSW profile for angle deviations of  $-1^\circ$  and  $+1^\circ$ .

of 6 mm along the x-axis.

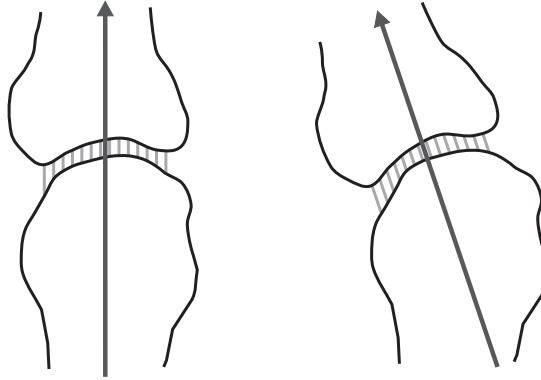
$$\Delta w_\alpha(x) = \left( \frac{w_\alpha(x)}{w_0(x)} - 1 \right) \times 100\% \quad (3.1)$$

For each measurement location a reference measurement was done along a vertical line ( $w_0$ ) and two measurements under an angle of minus and plus one degree ( $w_{-1}$  and  $w_{+1}$ ).

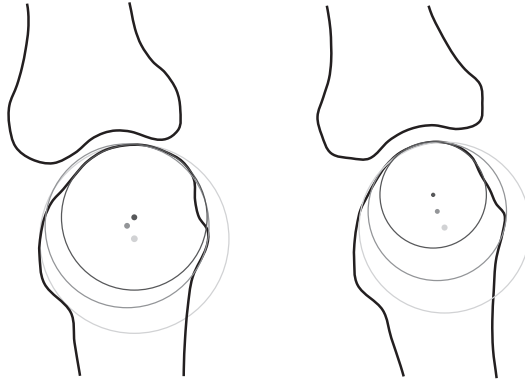
If measurements are performed along parallel lines, the direction should be independent of the positioning of the hand. In case of the MCP joints (PA projection), it is undesirable that the JSW depends on lateral flexion of the joint. To partly counter this effect, the measurement lines can be positioned parallel to the midline of the proximal phalanx (Figure 3.11).

Particularly when measuring the joint space margins of the MCPs over a wider region, precision may improve when the measurement lines are adjusted to the curvature of the joint margins. As mentioned in the previous section, James, Angwin and others [35] use radial lines originating from the center of the circle fitted to the metacarpal head. The difficulty with this approach is that fitting a circle to the margin of a metacarpal head can be ambiguous. Figure 3.12 shows that, depending on the shape of the metacarpal head, there may be different possibilities leading to different origins for the radial lines used for measurements.

Deviations in the determination of these circle origins may affect the JSW measurement. To investigate this, we applied this technique on the contour data of



**Figure 3.11:** If the measurement lines are chosen parallel to the midline of the proximal phalanx, then the JSW is less dependent on MCP abduction.

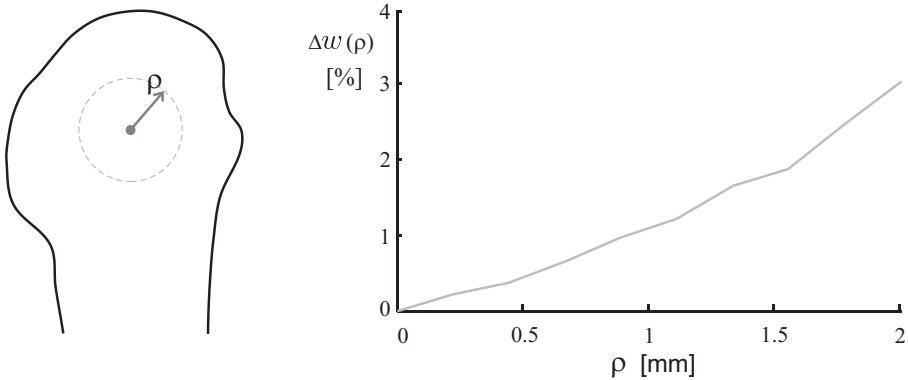


**Figure 3.12:** Two examples showing that different circles can be fitted to the margin of a metacarpal head. The centers of the circles are marked by dots.

100 MCP joints. For each sample we applied the method of James [35] to determine a reference measurement  $w_0$  of the JSW. To measure the effect of possible errors, we also performed several extra measurements  $w(\rho, \beta_i)$  by displacing the origin of the fitted circle by a distance  $\rho$  in a random direction  $\beta_i$ . Figure 3.13 shows the mean JSW deviation  $\Delta w(\rho)$  for  $N = 100$  experiments, with:

$$\Delta w(\rho) = \frac{1}{N} \sum_{i=1}^N \left| \frac{w(\rho, \beta_i)}{w_0} - 1 \right| \times 100\% \quad (3.2)$$

This graph shows that JSW measurements can change with approximately 1.5% per millimeter shift of the fitted circle. During our experiments, we found that differently fitted circles can result in different origin locations that up to four



**Figure 3.13:** Small deviations can occur with the determination of the center of the metacarpal head. As a result the measured mean JSW differs along the radial lines. The graph shows the mean absolute change of the JSW for random displacements of the center point by a distance  $\rho$ .

millimeters apart.

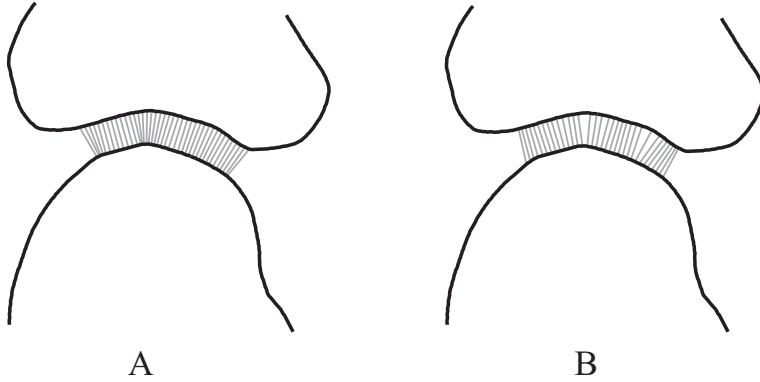
Since both joint margins generally have different curvatures, it is impossible to measure the JSW perpendicular to both margins over the entire joint space. Alternatively it is possible to measure the shortest distance between the margins at multiple locations along either of the margins. With this method the orientation of the measurement lines are solely dependent on the curvature and location of the joint space margins and independent of the orientation of the joints in the image.

Sharp and others [39] applied this method by measuring the shortest distances for locations on the distal margin. When the curvatures of the margins change gradually, this method works well. The average shortest distance can be measured from equidistant points on either margin. When the curvatures of the margins vary, the measurements can be spread with a nonuniform distribution (Figure 3.14). To correct for this problem, one could apply this method to both margins and average the results. We have used this approach in a previous study [98].

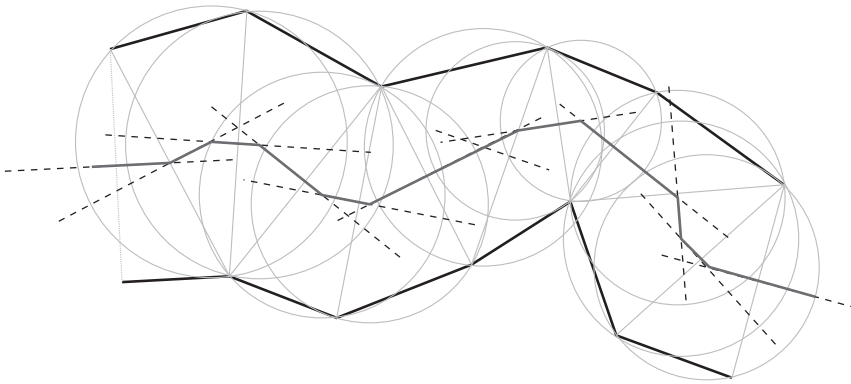
As previously referred to, Allander, Forsgren and others [32, 54] apply a similar method using a distance transform. For this method the calculations have to be made in pixel space on a discrete grid. If the margins are described by line segments, the medial axis can be found geometrically through triangulation, as demonstrated in Figure 3.15. Figure 3.16 shows how the shortest distances are measured from equidistant points on the medial axis.

### 3.3.5 Comparing methods

Six methods, varying in the abovementioned aspects were applied to all joints (315 MCPs and 316 PIPs) in the dataset:

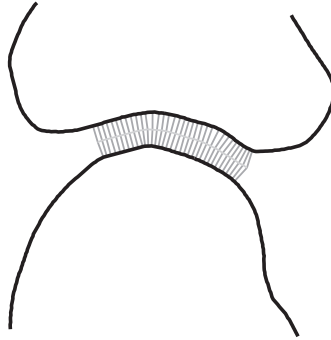


**Figure 3.14:** Measuring shortest distances from equidistant points on the distal margin (A) or the proximal margin (B). The varying curvatures result in a nonuniform distribution of the measurements.



**Figure 3.15:** The medial axis can be found by triangulation.

- A. Measurements along parallel lines in the direction of the midline of the proximal phalanx. A similar method is used by Duryea et al.[37] and Angwin et al. [40].
- B. Measurements along radial lines originating from the center of the metacarpal head (only MCPs), as used by Angwin et al. [40].
- C. Shortest distance from the proximal margin.
- D. Shortest distance from the distal margin, as used by Sharp et al. [39].
- E. Average of methods C and D, as used in our method [98].
- F. Shortest distance from the medial axis. Similar to the method presented by Allander et al. [32].



*Figure 3.16: Shortest distance measurements along the medial axis.*

For each method we calculated the average of 100 measurements at equally spaced locations. The methods described in Section 3.2 use different regions to measure JSW: some define a percentage of the joint, some are dependent on marks set by an operator, and others do not specify this aspect. For this study, we chose a fixed region of 6 mm centered on the midline of the proximal phalanx. We have found that this size fits for all MCP and PIP joints in the dataset of this and other studies.

To compare JSW outcomes between methods, we have fitted a straight line to the results of each pair of methods. Fitting was done by orthogonal regression; minimizing the sum of the squares of the perpendicular distances (offsets) between each point and the line. The outcomes of methods Y and X are modeled by:  $Y = aX + b$ . Where  $a$  is the slope and  $b$  the intercept. As a measure for the fitting residuals, we determined  $r$ , which is the standard deviation (SD) of the perpendicular offsets. Also, we calculated the mean of all JSW measurements for each method. MCP and PIP measurements were compared separately.

For the MCP joints, the comparison results between methods A–F are displayed in Table 3.17 and Figure 3.18. For the MCP joints these are displayed in Table 3.19 and Figure 3.20 (excluding method B, which only applies to MCP joints). Looking just at the means, there is little difference between methods B–F. The mean of method A is significantly larger, which is to be expected because the the measurement lines are not perpendicular to a large part of the joint margins. Compared to any of the other methods, method A also shows the largest differences in terms of slope  $a$  and intercept  $b$ . The SDs of the residuals  $r$  are the highest for comparisons with method B. The outcomes of methods C–F are almost the same for the MCP joints. For the PIP measurements the difference between these methods is also small, but slightly higher than for the MCP joints. The reason is that PIP joints generally have a narrower JSW and more irregular margins than MCP joint spaces. This results in larger differences between the shortest distance measures, as illustrated in Figure 3.14.

		A	B	C	D	E	F
A	a		1.030	1.025	1.025	1.025	1.025
	b		0.013	0.018	0.016	0.017	0.017
	r		0.026	0.016	0.014	0.015	0.015
B	a			0.996	0.995	0.995	0.995
	b			0.005	0.003	0.004	0.005
	r			0.029	0.028	0.028	0.029
C	a				0.999	1.000	0.999
	b				-0.002	-0.001	-0.001
	r				0.006	0.003	0.004
D	a					1.001	1.000
	b					0.001	0.001
	r					0.003	0.004
E	a						1.000
	b						0.000
	r						0.002
Mean JSW		1.424	1.370	1.371	1.374	1.373	1.372

**Figure 3.17:** Results comparison of JSW measurements (in millimeters) performed on 315 MCP joints using methods A–F. Orthogonal regression was used to determine the relation between the outcomes of method Y (row labels) and X (column labels):  $Y = aX + b$ . The value indicated by r, is the SD of the perpendicular offsets. The bottom row shows the means of all measurements.

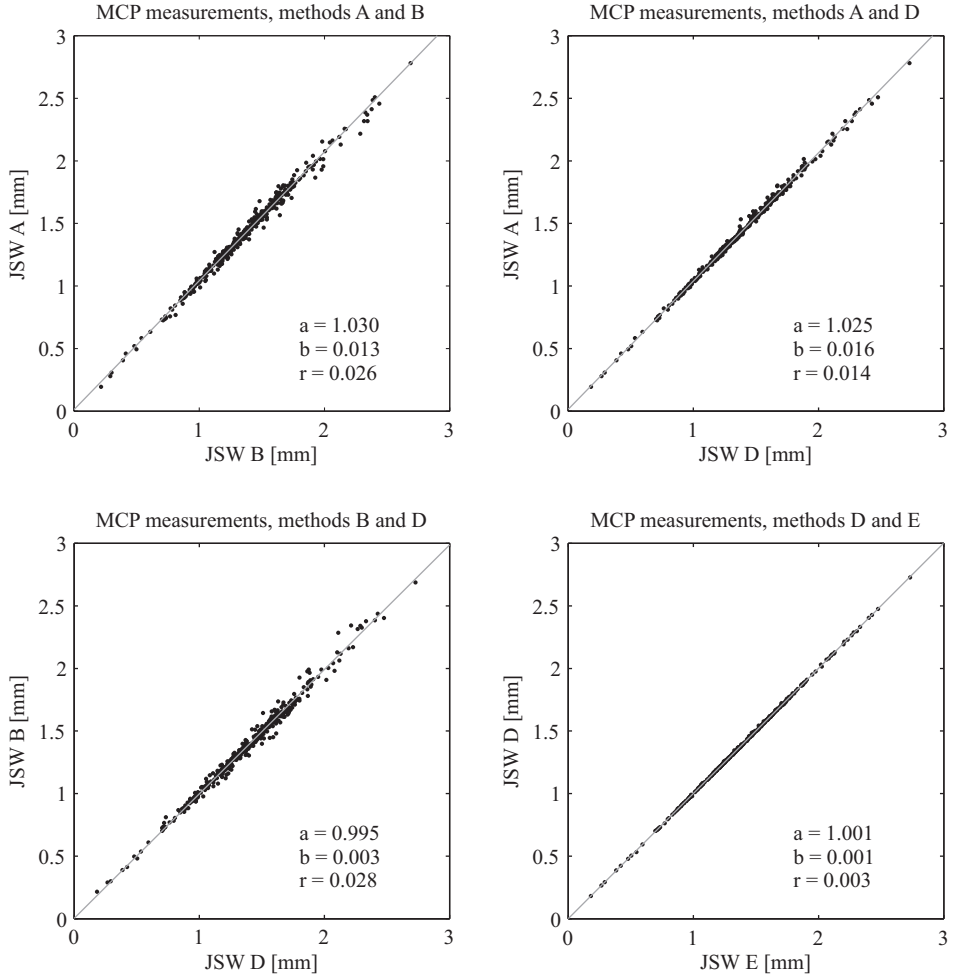
### 3.3.6 Other measurements

As noted before, it is possible that changes occur without affecting the average JSW. To detect such changes, one can build a JSW profile of each joint using one of the aforementioned methods. Next, many additional measures can be based on the change in the profile of the joint margins; e.g. the variance with respect to the mean JSW, a measure of symmetry of the joint space or the mean absolute difference between follow-up radiographs. When performing cross-validation between different margin detection methods by comparing JSW values, it is important that the same measurement method is used. Instead of comparing such calculated measures, it may be more practical to directly compare the JSW profiles.

## 3.4 Discussion

Measuring JSW using semi quantitative methods has proved successful in a large number of studies [61]. Since this requires at least two independent human observers and is time-consuming, various efforts have been undertaken to automate JSW measurement. A number of studies have presented reproducible results, ex-



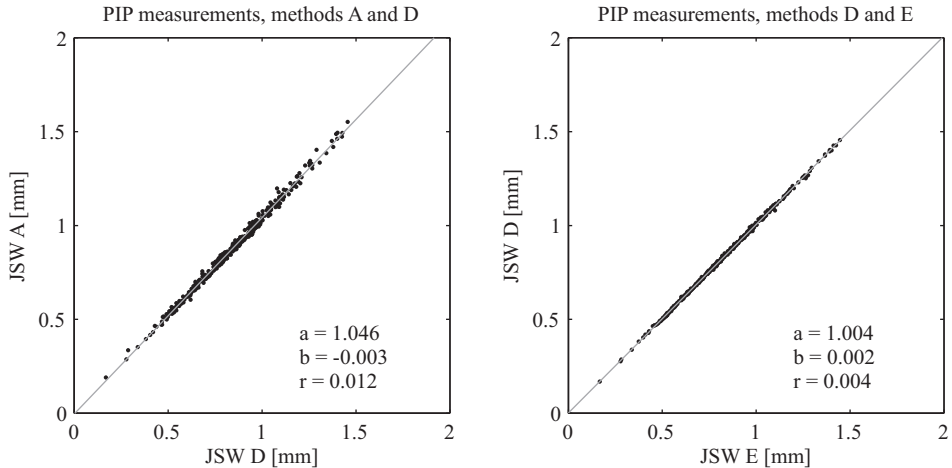


**Figure 3.18:** Selection of scatter plots with regression lines comparing the MCP JSW outcomes of 4 different method pairs. The slopes and intercepts of the regression lines are indicated by  $a$  and  $b$ , the SDs of the perpendicular offsets by  $r$ . Table 3.17 shows the results for other method pairs.

pressing JSW in millimeters. However, the methods used were different, and consensus regarding the precise way to measure JSW is lacking. Comparison of methods using the same radiographs demonstrated relevant differences in JSW [60]. In the present study we compare the proposed quantification methods, and demonstrate experiments to detect which elements lead to relevant differences. In particular, we have investigated the effects of measurement direction, region of the joint to consider, and the number of measurements to be performed on a single joint.

		A	C	D	E	F
A	a		1.053	1.046	1.049	1.044
	b		0.002	-0.003	0.000	0.005
	r		0.012	0.012	0.011	0.012
C	a			0.993	0.997	0.992
	b			-0.005	-0.003	0.003
	r			0.008	0.004	0.005
D	a				1.004	0.999
	b				0.002	0.008
	r				0.004	0.005
E	a					0.995
	b					0.006
	r					0.003
Mean JSW		0.863	0.818	0.829	0.823	0.822

**Figure 3.19:** Results comparison of JSW measurements (in millimeters) performed on 316 PIP joints using methods A–F. Orthogonal regression was used to determine the relation between the outcomes of method Y (row labels) and X (column labels):  $Y = aX + b$ . The value indicated by r, is the SD of the perpendicular offsets. The bottom row shows the means of all measurements.



**Figure 3.20:** Two scatter plots (with regression lines) comparing the PIP JSW outcomes of two different method pairs. The slopes and intercepts of the regression lines are indicated by a and b, the SDs of the perpendicular offsets by r. Table 3.19 shows the results for other method pairs.

First, in a joint the region to be measured needs to be chosen. Most studies define a fixed breadth within which the average JSW is determined. Of course, a larger sample region provides a more precise average, but it should also be small enough to fit all possible joint shapes and sizes. Duryea et al. propose a variable breadth based on the size of the joints or the bones [37]. This carries the risk of decreasing accuracy due to detection errors or changes over time by bone deformation or osteophytes (protrusions on bones). Therefore, a region with an absolute width based on consensus is preferable.

In MCP joints ulnar deviation rotates the proximal phalanx around the metacarpal head, causing an extra definition problem for the region to be measured. Angwin tries to overcome this by measuring along radial lines from the center of the metacarpal head [40]. However, this requires manual positioning of marks on the metacarpal head, which may be prone to variation. As demonstrated in our experiment, small variations in the position of the assumed center of the metacarpal head may lead to relevant changes in the resulting JSW. Alternatively, as illustrated in Figure 3.9, the intersection of the proximal phalanx's midline with the joint space can be used to define the center of the region to measure. Since the PIP (and DIP) joints are rigid from a PA perspective, it makes little difference whether the midline of the proximal phalanx is used or the midline of the middle phalanx. From a practical point of view the midline of the proximal phalanx is preferable, as it can also be used for finding the center of the MCP joint. Secondly, as the body of this bone is generally longer than that of the middle phalanx, its midline can be determined more precisely.

Once the region to be measured is defined, the distance between the joint margins can be determined in many ways. It would be most practical to perform a single distance measurement, for instance to measure the minimum JSW or the width at a fixed location. But joint space narrowing is often asymmetrical and the precision of a single distance measurement is highly dependent on the precision of the detection of the joint margins. Moreover, from our experiment on the effects of the direction of measurement it is obvious that even small variations in measurement direction have relevant effects on JSW. Therefore, it is mandatory to measure the JSW multiple times over the joint space. What should we report: the minimum or the average of these multiple measurements of JSW in the region defined in the previous steps? Obviously the minimum is based on just a single measurement and can be misleading. The average of a large number of measurements is thus more attractive, and has been implemented in most previously described methods. The precision will be limited by the resolution of the radiograph and the quality of the margin detection algorithm.

The 4 methods described in Section 3.2 use different approaches to perform multiple measurements. In this study we have applied these methods to joints with predefined joint margins, thus excluding other factors that can affect the resulting JSW. We have found that a single method (A) stands out by slightly higher values. This means that data obtained by this method cannot be directly compared with methods C-F, which appeared to be more in line with each other.

Distances between margins can be measured as the average of the shortest distances from multiple equally spaced locations on either of the margins (method C or D). The precision can be further increased, when the shortest distance is measured from both margins (method E). Corresponding results can be obtained using the medial axis (method F), but this is more complex to implement due to additional geometrical computations. We therefore recommend using method E. By determining the mean JSW not all changes in the joint space can be described. Additional measures can be based on changes in the JSW profile (the succession of local measurements) of the whole joint space region, e.g. the change in variance or minimum. In this study we have not looked at the sensitivity to detect changes in JSW over time. Therefore, we cannot exclude that a single method would perform better than the other in this respect. However, the low residual values and almost identical means of methods C–E make it unlikely that they will have uneven test characteristics. Although methods A and B could have better characteristics, they are more prone to error by small variations in hand position on radiographs or manual positioning of markers to define the region to be measured.

### 3.5 Conclusion and recommendation

To measure JSW in MCP and PIP joints, uniformity in methodology is desirable. The axis of the first proximal phalanx provides the easiest landmark to start automated analysis. The region to measure JSW can be defined by the crossing of this axis with the proximal (MCP) and distal (PIP) margin of this bone. Using a fixed size for the measurement region will be most robust when comparing follow-up images. In the dataset of this and other studies, we found that a region of 6 mm wide fits for all MCP and PIP joints. To measure the actual distance between the joint margins, the method using the average of the shortest distances measured from both the distal and the proximal margins provides the most precise results. Consensus among researchers in this field will lead to exchangeable data regarding JSW in millimeters. This may be of help in future epidemiological research and in the comparison of outcomes of interventional trials in RA or OA.

# 4

## Segmentation of bones in hand radiographs

### 4.1 Introduction

To enable automated assessment of hand radiographs, one needs to find a way to detect and identify regions of interest: i.e. joints and bones. As discussed in Section 2.4.1, thus far there are few systems that have realized this in a successful way.

One of the problems with image processing on hand x-rays is that there is a high variability in image quality. This has several causes. Some are related to the type of the x-ray system that is used, e.g. film in combination with phosphorescent plates, computed radiography, or indirect flat panel detectors. Except for differences in resolution or sensitivity of these systems, the exposure settings are generally not standardized, and therefore may differ. In case of film radiographs an additional digitization step is needed, the quality of this conversion depends on the type of image scanner or camera that is used. To be robust to this variability, an image processing algorithm needs to be robust to differences in resolution, noise, sharpness, contrast and brightness.

Depending on the type of imaging system used, there are commonly different types of non-anatomical objects in the radiographs, such as name tags, left/right labels, and digital imprints. Also, near the edges of the radiographs, the frame of the diaphragm is often visible. In case of film radiographs, written text and cutouts may appear at various locations in the image. Other objects that

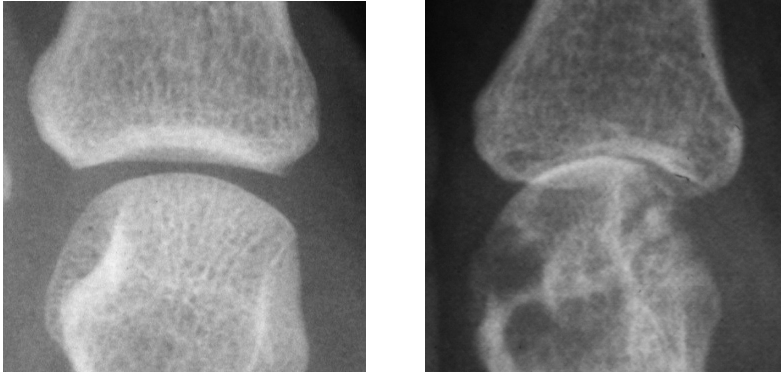


*Figure 4.1: Example set of hand radiographs of varying quality, positioning and with various non-anatomical objects present.*

are commonly present in radiographs are rings, bracelets, and implants. Figure 4.1 shows several example radiographs that are common in practice. Other image processing challenges are caused by the lack of a standard protocol on how a hand is to be positioned. Though the hand is usually placed flat on the detector, there are still several degrees of freedom in which the hand can be positioned: finger abduction/adduction, thumb flexion/extension, and wrist abduction/adduction. Furthermore, the overall orientation of the hand may be different, and sometimes a radiograph is made of both hands at the same time.

Besides the aforementioned sources of image variability there is also the hand's anatomical variation that has to be taken into account. This variation is particularly large for patients with bone degenerative diseases who may have deformities due to severely damaged bones and joints.

For future research, it would be convenient to use a fixed acquisition protocol, such that radiographs become more standardized. Further on, in Section 7.5, we



*Figure 4.2:* The left radiograph shows a healthy joint with the JSW clearly visible. The right shows a joint with severe joint damage.

present a positioning aid that can be used for this purpose. But with the current and past material, we must deal with a large diversity in radiographs.

To tackle the aforementioned problems, we have designed a segmentation algorithm that is based on an iterative search method using active appearance models (AAM) of the hand bones. An AAM is a deformable model of a certain image object that includes information about its shape and texture [66]. The modes of variation in an AAMs are commonly configured by using a set of annotated training examples that reflect the variability of objects to be modeled. By tuning the model's parameters, it is possible to control the appearance (shape and texture) of the model. To find a similar object in a new image, one has to control these parameters such that the AAM matches to the searched object.

With a properly configured model, only plausible appearances of the modeled object can be generated. Obviously this makes it difficult to obtain a good match with anomalous instances, which are likely in the case of hand radiographs of patients with RA. For instance, due to severe joint damage, the JSW can completely disappear, resulting in two bones being fused together Figure 4.2 shows an example of such severe damage. Despite that in such cases an exact fit of the model may be impossible, a partial fit can be adequate to estimate regions of interest such as the joint locations. This ability makes an AAM search robust compared to local image processing techniques. Other research with AAMs on parts of hand radiographs has been described in [67, 51] (see also Section 2.4.1).

## 4.2 Active appearance model (AAM) of the hand

The construction of an AAM of the hand begins with collecting a training set of hand radiographs that have been provided with suitable landmarks (Subsec-

tions 4.2.1 and 4.2.2). As the goal is to detect bone outlines, landmark locations are carefully chosen near the edges of the bone contours. With the training set of landmarks it is possible to describe the variability of the bone shapes using a linear model (Subsections 4.2.3, 4.2.4 and 4.2.5). Texture descriptions of the training examples can be obtained by sampling intensity values within the regions of the bones. Subsequently these textures are modeled and combined with the shape model into an AAM (Subsections 4.2.6 and 4.2.7).

As pointed out in the previous section, an important part of the variability in hand radiographs is caused by differences in hand and finger positioning. To separate the positioning variability from the anatomical variability, we make use of multiple connected AAMs [68, 69]. Each bone, except for the carpals, is modeled in a separate AAM and then combined in a single model of the hand. The carpal bones are also included in the hand model, but they are joined together in a single AAM (Subsection 4.2.8). To reduce calculation efforts in determining these general hand features, we begin the AAM search with a simplified, low-detail model. After several iterations, we switch to a high-detail model to find the bone edges (Subsection 4.2.9). Our AAM search method itself requires relatively little computation effort, as the relation between search error and model parameters are determined in advance (Subsection 4.2.10).

The main part of the AAM search algorithm has been implemented in Matlab. Due to memory and performance limitations, several functions were optimized using C++ code. We developed and tested the programs on a computer with a 2 GHz Intel Pentium IV processor and 2 GB of memory.

### 4.2.1 Dataset

We have gathered a set of 100 single hand radiographs (posteroanterior view) of patients diagnosed with RA. Both left and right hand images were present in the set, but left hand images were mirrored such that the same method could be used on both hands. For each patient in the set there were at least two radiographs available (both hands), for several patients there were multiple radiographs available from different time points.

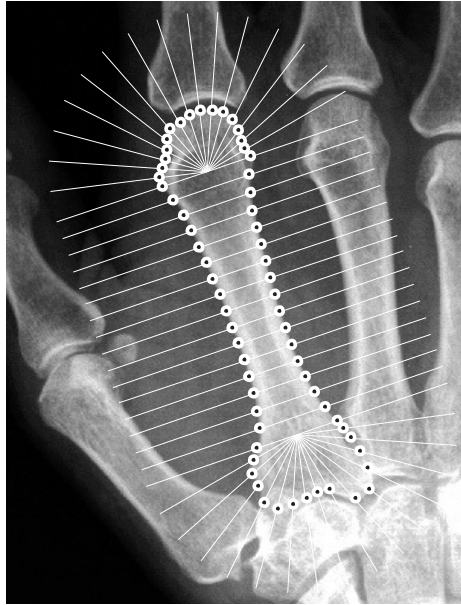
The images used in this study are digital scans from film radiographs. The scanning resolution is 600 dpi (in both directions) with 16-bit gray values (effectively 12-bit), and images were stored in lossless image formats.

The dataset has been split into a set of 50 radiographs for training and a set of 50 for testing. This has been done such that the radiographs of each particular patient are only present in either the training or the test set.

### 4.2.2 Landmarks

To record landmarks in the image data set, we have developed a custom MS Windows C++ application. This application allows an operator to accurately place landmarks by using several magnification and contrast enhancement tools. The





*Figure 4.3: A 'fish bone' shaped grid and the manually outlined contour define the landmark locations.*

metacarpals, proximal and middle phalanges have been outlined manually in each radiograph. To define landmarks on each bone contour, we place a 'fish bone' shaped grid by selecting the proximal and distal bone ends. The intersections of the grid with the (manually outlined) contour then define the landmarks as depicted in Figure 4.3. The radiating grid lines originate at 15% of the bone length from both ends. For each bone the same grid is used, consisting of 64 intersecting gridlines.

Since the distal phalanges are considered to be less important for joint damage assessment in RA, we mark them with only four landmarks. Finally, 10 landmarks indicate the carpal region. The landmark locations are displayed in Figure 4.4 and 4.5. It took between 20 and 30 minutes to provide a single hand radiograph with all described landmarks.

### 4.2.3 Shape vector

We define the term 'shape' as a set of  $L$  landmark points  $(x_j, y_j)$ . To enable statistical analysis on these points, we store the  $x$  and  $y$ -coordinates in a  $2L$  element shape vector  $(x_1, x_2, \dots, x_L, y_1, y_2, \dots, y_L)^T$ . The shape vector's unit is defined in millimeters instead of pixels, such that the shapes are independent of the resolution of the images.

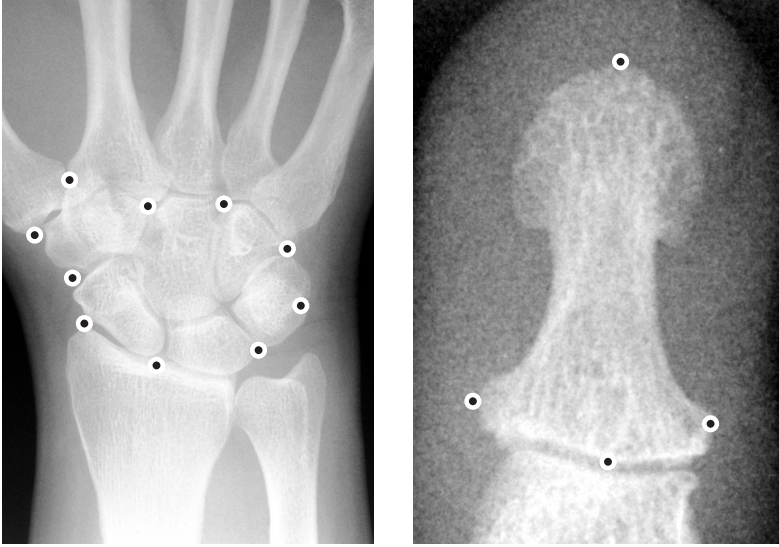


Figure 4.4: Landmark locations of wrist and distal phalanx.

#### 4.2.4 Overall alignment

To make the AAM of the hand invariant to pose variations (translation, scaling and rotation), we align the training examples to one another using partial Procrustes analysis [70]. For this overall alignment we make use of the landmarks of metacarpals 2–5, since this group of bones form a rigid block independent of finger positioning.

Firstly, we translate all shapes to the origin by subtracting the centers of the ‘metacarpal blocks’ (MBs). The center  $(\bar{x}, \bar{y})$  of a MB is calculated from its  $K$  landmarks by

$$\bar{x} = \frac{1}{K} \sum_{j=1}^K x_j, \quad \bar{y} = \frac{1}{K} \sum_{j=1}^K y_j. \quad (4.1)$$

A centered shape vector is then defined by

$$\mathbf{x}_c = (x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_L - \bar{x}, y_1 - \bar{y}, y_2 - \bar{y}, \dots, y_L - \bar{y})^T. \quad (4.2)$$

Secondly, we determine a measure for the scale  $s$  of each MB by calculating the root of the sum of square elements of  $\mathbf{x}_c$ :

$$s = \sqrt{\mathbf{x}_c^T \mathbf{x}_c}. \quad (4.3)$$

Subsequently we rescale all shapes to the mean scale  $\bar{s}$  of the training set and



*Figure 4.5: Locations of all landmarks in a training example.*

obtain

$$\mathbf{x}_s = \frac{\bar{s}\mathbf{x}_c}{s}. \quad (4.4)$$

To remove the rotational component, we select one of the MB shapes  $\mathbf{x}_{s_0}$  and rotate all other shapes  $\mathbf{x}_{s_i}$  so that the sum of distances  $D_i(\theta)$  between points is minimized:

$$D_i(\theta) = \|\mathbf{x}_{s_0} - \mathbf{x}_{s_i}\mathbf{A}(\theta)\|, \quad (4.5)$$

where  $\mathbf{A}(\theta)$  is a rotation matrix corresponding to angle  $\theta$ . Finally, we rotate all shapes to the mean angle  $\bar{\theta}$  of the training set and obtain the aligned set of shapes defined by

$$\mathbf{x}_i = \mathbf{x}_{s_i}\mathbf{A}(\theta_i - \bar{\theta}). \quad (4.6)$$

### 4.2.5 Modeling shape

Shape variation can be expressed in a statistical manner by determining the mean  $\bar{\mathbf{x}}$  and eigenvectors  $\phi_i$  of the shape vectors in the training set.

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i, \text{ and} \quad (4.7)$$

$$\mathbf{\Phi} = [\phi_1 \ \phi_2 \ \dots \ \phi_t] \quad (4.8)$$

where  $N$  is the number of examples in the training set. To reduce the number of parameters that control the shape model, we apply principal component analysis (PCA) [71], and construct matrix  $\mathbf{\Phi}$  from  $t$  eigenvectors corresponding to the largest eigenvalues. The shapes of the training set can now be approximated by

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{\Phi}\mathbf{b}, \text{ with} \quad (4.9)$$

$$\mathbf{\Phi} = [\phi_1 \ \phi_2 \ \dots \ \phi_t], \text{ and} \quad (4.10)$$

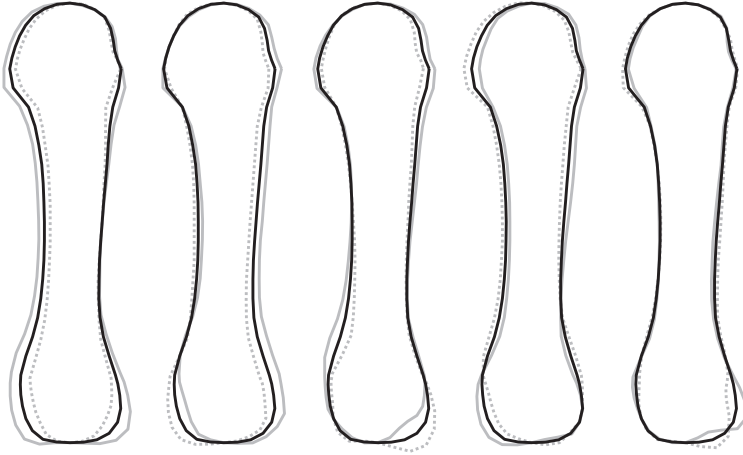
$$\mathbf{b} = \mathbf{\Phi}^T(\mathbf{x} - \bar{\mathbf{x}}). \quad (4.11)$$

The right side of equation 4.9 can be used to generate new synthetic shapes by choosing new values for the parameter vector  $\mathbf{b}$ . To make sure that only plausible shapes are generated, we set the following condition to these parameters:

$$|b_i| \leq 3\sqrt{\lambda_i}, \quad (4.12)$$

where  $\lambda_i$  is the eigenvalue of the corresponding eigenvector  $\phi_i$ .

Figure 4.6 shows the first five modes of variation (controlled by parameters  $b_1$ - $b_5$ ) of the shape model of the third metacarpal bone. For this example the pose (translation, scaling and rotation) variation was removed by aligning the proximal and distal ends of the training examples.



**Figure 4.6:** Shape model of metacarpal 3, showing the first five modes of variation (from left to right in increasing order). Black line: mean, gray line:  $+3 SD$ , dotted gray line:  $-3 SD$

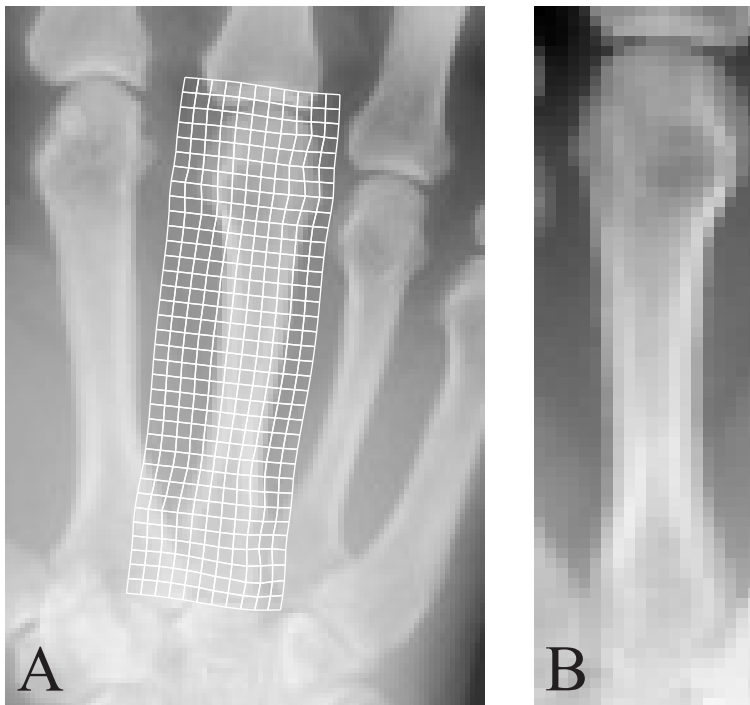
#### 4.2.6 Modeling texture

By the term 'texture', we refer to a set of intensity values that we sample from an image object using a fixed pattern. To create a model of an object's texture independent of its shape, we first warp each image from the training set to the mean of our shape model. For a single object, such as a metacarpal bone, this can be done directly by thin-plate spline warping [72], using the landmarks as source and the mean shape as target points (Figure 4.7). For partially overlapping objects this method cannot be used, because discontinuities would occur in the warped image. As bones in hand radiographs often overlap near the joints, each bone has to be warped separately.

With all image data warped to the mean shape, we define a shape-independent texture patch from which gray intensities can be sampled. As we are interested in finding the contours of the bones, we do not only sample gray values within the bone contours, but also in the near region outside of them. For this purpose we fit a bounding box to each bone contour with an additional spacing of 3 mm. Next, we define an equally spaced grid of  $M$  sample points to create a texture vector of intensity values

$$\mathbf{g}' = (g_1, g_2, \dots, g_M)^T. \quad (4.13)$$

Before sampling texture values, one has to apply an appropriate low-pass filter to avoid aliasing artifacts [73]. To compensate for differences in image contrast and



**Figure 4.7:** Example showing how the bone texture from a training example (image A) is warped to the mean shape in image B. The deformation mesh is calculated using thin-plate splines [72].

brightness we calculate the normalized texture vector by

$$\mathbf{g} = \frac{1}{\sigma_g}(\mathbf{g}' - \mu_g), \text{ with} \quad (4.14)$$

$$\mu_g = \frac{1}{N} \sum_{i=1}^N g_i, \text{ and} \quad (4.15)$$

$$\sigma_g = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (g_i - \mu_g)^2}. \quad (4.16)$$

In the same manner as for the shape vectors, the texture variation can be expressed in a statistical manner by determining the mean  $\bar{\mathbf{g}}$  and eigenvectors  $\psi_i$  of the training data. Also for the texture model we apply PCA to create matrix  $\Psi$  from  $z$  eigenvectors corresponding to the largest eigenvalues. An approximation of the textures in the training set is then given by

$$\mathbf{g} \approx \bar{\mathbf{g}} + \Psi \mathbf{d}, \text{ with} \quad (4.17)$$

$$\Psi = [\psi_1 \ \psi_2 \ \dots \ \psi_z], \text{ and} \quad (4.18)$$

$$\mathbf{d} = \Psi^T(\mathbf{g} - \bar{\mathbf{g}}). \quad (4.19)$$

The right hand side of equation 4.17 represents the statistical description of the texture in the training set. Parameter vector  $\mathbf{d}$  controls the modes of variation, and is bound to the following condition:

$$|d_i| \leq 3\sqrt{\kappa_i}, \quad (4.20)$$

where  $\kappa_i$  is the eigenvalue of the corresponding eigenvector  $\psi_i$ .

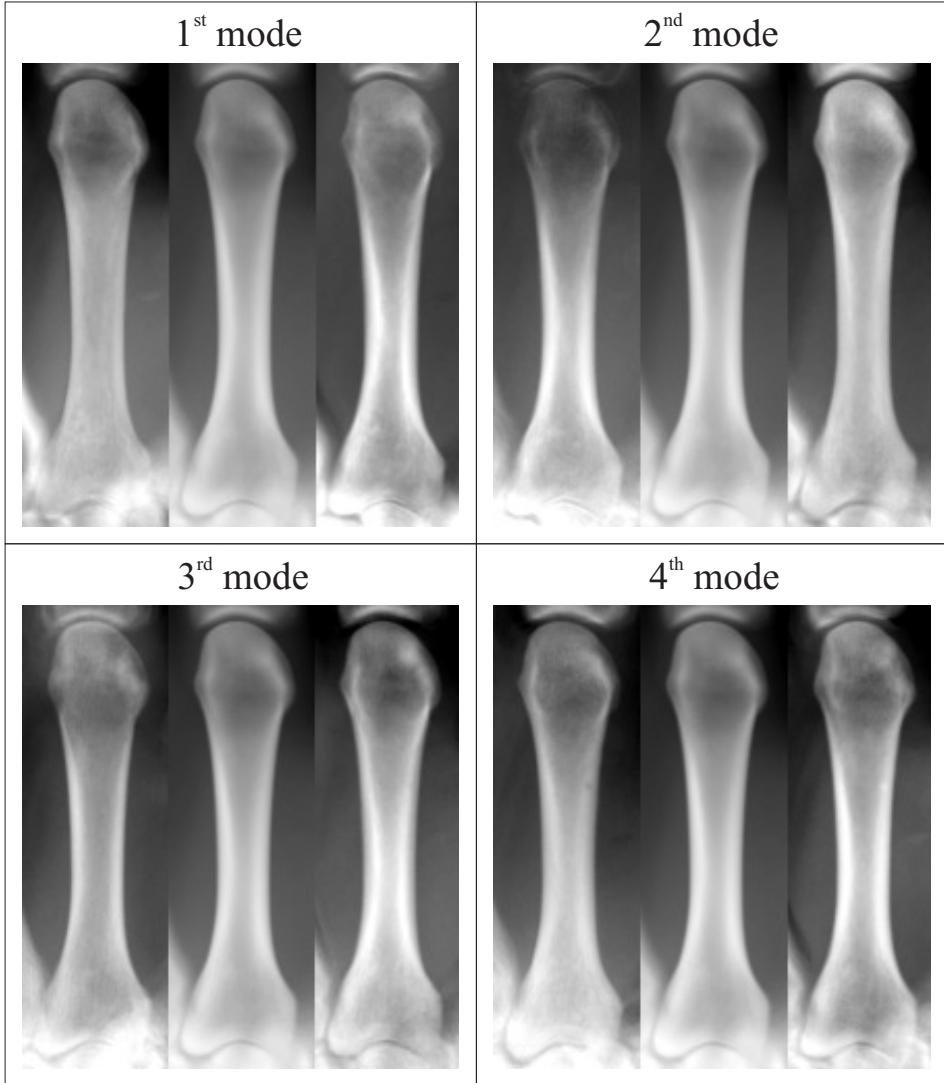
To illustrate the properties of a texture model, Figure 4.8 shows the first four modes of variation of the texture model of the third metacarpal bone. Note that the bone's shape is the same in all panels. In this example the texture's sampling resolution is 3 pixels per millimeter.

### 4.2.7 Combining shape and texture

To combine shape and texture information into a single model, we concatenate shape and texture vectors  $b$  and  $d$  (Equations 4.11 and 4.19). As shape and texture are based on different quantities, we correct the shape parameters by weighing factor  $z$  based on the total of variances in the training set.

$$\mathbf{f} = \begin{pmatrix} z\mathbf{b} \\ \mathbf{d} \end{pmatrix} \quad (4.21)$$

$$z = \frac{\text{var}(\mathbf{d})}{\text{var}(\mathbf{b})} \quad (4.22)$$



*Figure 4.8:* The first four modes of variation of the texture model. Each panel displays three instances:  $-3$  SD,  $0$  SD, and  $+3$  SD parameter change (from left to right).



As correlation may be present between shape and texture parameters, PCA is used on the vectors  $\mathbf{f}$  from the training set to reduce the number of variations in the model to the first  $k$  (principal) components. The result is matrix  $\mathbf{Q}$  with  $k$  eigenvectors  $\mathbf{q}_i$ , and since the means of  $\mathbf{b}$  and  $\mathbf{d}$  are zero over the training set,  $\mathbf{f}$  can be approximated by:

$$\mathbf{f} \approx \mathbf{Q}\mathbf{c}, \quad (4.23)$$

with parameter vector  $\mathbf{c}$  to control both shape and texture parameters.

Combining the previous equations with Equations 4.9 and 4.17, we obtain the following expression for the appearance model:

$$\begin{pmatrix} \tilde{\mathbf{x}} \\ \tilde{\mathbf{g}} \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{g}} \end{pmatrix} + \begin{pmatrix} \frac{1}{w}\Phi & 0 \\ 0 & \Psi \end{pmatrix} \mathbf{Q}\mathbf{c}. \quad (4.24)$$

### 4.2.8 Connected submodels

When making a shape model of the complete hand using the aforementioned dataset, the variability due to positioning differences dominates over anatomical variability. A second problem due to positioning differences is that non-linearities occur when a single  $(x, y)$ -coordinate frame is used. For instance, when the bones of the thumb and the little finger are combined in a single linear model, angular differences between the thumb and little finger commonly range between 0 and 90 degrees. In combination with translation and scaling components, this results in a non-linear connection between the shape (and position) descriptions of both fingers. Clearly, such connections cannot be accurately approximated in a single linear model, especially when few training examples are available. To solve this problem, we separate anatomical and positioning variability by dividing the hand model into submodels.

For posteroanterior (PA) projection of the hand we consider seven joints that can introduce lateral positioning variability: CMC 1, MCP 1–5, and DIP 1. To eliminate the variability caused by finger positioning, we subdivide the hand model at these joint locations. This results in eight different submodels:

- A base model of the carpal bones and metacarpals 2–5.
- Four submodels of phalanges 2-5 connected to the base model.
- A submodel of the first metacarpal connected to the base model.
- A submodel of the first proximal phalanx connected to the submodel of the first metacarpal.
- A submodel of the first distal phalanx connected to the submodel of the first proximal phalanx.

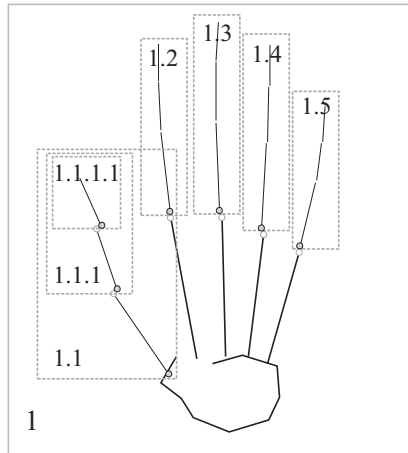
By aligning the shape examples of the training set to one another in a new sub-coordinate frame, each submodel becomes invariant to positioning (translation and

rotation) variability. The orientation of each sub-coordinate frame in the overall model is determined by the location and angle of the relevant joint in its parent's coordinate frame. Figure 4.9 displays the connected submodels in a wire frame model, and the structure of the coordinate frames. The metacarpals and the bones of the phalanges are represented by single wires which run from the proximal ends to the distal ends of the bones.

In Figure 4.10 we demonstrate the effect of the described operations on the shape variability of the training set. The left pane shows the wire frame models from the training set after overall alignment, without correcting for positioning variability. The right shows the same data set, but after aligning the landmarks in the sub-coordinate frames. Each submodel has been aligned by its proximal end to remove the translation component. The rotational component is removed by aligning the angles of the bone axes (of the proximal bone in case of phalanges 2–5).

### 4.2.9 Multi-model search strategy

With the design of an AAM and the accompanying search strategy, one has to consider the desired accuracy and the available computing power. By increasing the number of landmarks for the shape model and choosing a high resolution for the texture model, the AAM's accuracy can be improved. As a consequence, this will result in an increase of the number of model parameters, setting an extra



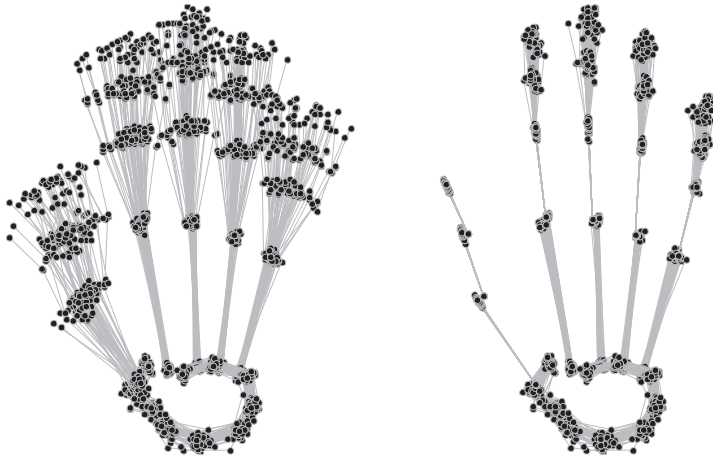
**Figure 4.9:** Wire frame model of the hand with connected submodels. The dots mark the joints that enable lateral displacement of the fingers. Seven dependent sub-coordinate frames (1.1-5, 1.1.1, 1.1.1.1) are used to make the model invariant to changes in finger positioning.

demand on the required calculation efforts.

Using pose invariant connected submodels, as described in the previous section, makes the model more accurate with regard to anatomical variations. A drawback of such approach is that the overall hand model becomes more complex and requires more calculations: To reconstruct the complete hand model, we need to add three parameters for each dependent submodel; two for translation and one for rotation. As we define 7 dependent submodels, 21 extra parameters have to be introduced for controlling the model.

On the other hand, increasing an AAM's accuracy does not necessarily mean that robustness of the AAM search algorithm improves. The more parameters are introduced to control the model, the more difficult it becomes to design a robust search strategy. This robustness is determined by how well the search method converges to global optima (in contrast to local optima), and how well the AAM must be initialized.

Taking into account these considerations, we choose a search strategy based on the succession of two different AAMs. The first is a low-detail AAM controlled by a small set of parameters, which makes it fast and robust during the search. The result of this first search is used to initialize a second, high-detail AAM. The second AAM is slower, as it is controlled by more parameters to enable an accurate match of the model. In this way we combine the robustness and speed of the first AAM with the higher accuracy of the second.



*Figure 4.10: Wire frame representation of training set before (left) and after (right) aligning the submodels in the sub-coordinate frames.*

### Low-detail model

The low-detail model consists of a single, overall shape model of the hand. The landmarks used for this model are defined by the four corners of each bone's bounding box. The carpal region is defined by a single bounding box, which makes that the shape model consists of  $20 \times 4 = 80$  points (19 bones + carpal region). To make the model invariant to pose variations, we applied the alignment procedure as described in Section 4.2.4. For the texture model, the resolution is set to 8 pixels per centimeter. After applying PCA on the data of the training set the AAM could be reduced to 30 parameter variations covering 89% of the total variance displayed in the training set. The overall pose is controlled by four extra parameters for translation ( $\times 2$ ), scaling and rotation.

Figure 4.11 shows the first three modes of variation of the low-detail model. At the locations where textures overlap the average gray intensity values are displayed.

### High-detail model

The high-detail model has its focus on anatomical variability and consists of all 920 landmarks that are available in the training set (Section 4.2.2). The AAM is made positioning invariant by aligning the overall data and using connected submodels, as explained in Section 4.2.4 and 4.2.8 respectively. The resolution of the texture model is set to 16 pixels per centimeter. After applying PCA, the AAM was reduced to 25 modes of variation covering 85% of the variance in the training set. Note that in contrast to the low-detail model, this model contains mainly anatomical variation, as positioning variability has been removed. To reestablish positioning variation, 21 parameters are added to control translation and rotation of 7 dependent submodels. Four additional parameters control the overall pose of the AAM.

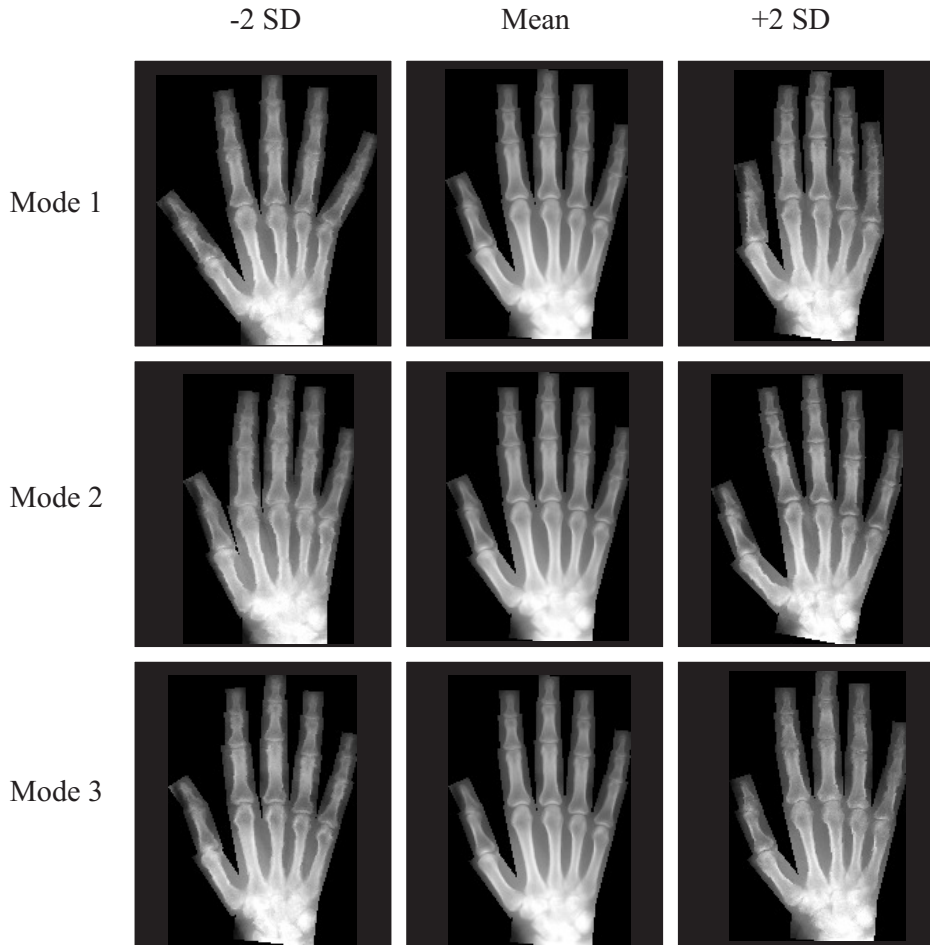
### Conversion between models

To initialize the second AAM search with the results of the first, it is necessary to find a conversion method between the parameters of both models. If both models have been created using the same training examples, a linear relationship can be deduced from the optimal model parameters.

Let matrices  $\mathbf{U}$  and  $\mathbf{V}$  be two sets of  $N$  parameter vectors of two different AAMs but corresponding the same training data. As the parameter means of the training data are zero, we can estimate the cross-covariance matrix  $\Sigma_{vu}$  by:

$$\Sigma_{vu} \approx \frac{\mathbf{V}\mathbf{U}^T}{N-1}. \quad (4.25)$$

Obviously, this estimate becomes better when more training examples are available. With the help of the cross-covariance matrix we can estimate parameter



*Figure 4.11: Synthetic images created with the low-detail model for the three largest modes of variation (rows). To demonstrate the effect of each mode, the corresponding parameter is changed with  $\pm 2$  standard deviations from the mean (columns).*

*(Remark: some minor visible image distortions are unrelated to the model, but were caused by the applied image rendering method.)*

vector  $\mathbf{v}$  based on  $\mathbf{u}$  by

$$\mathbf{v} \approx \Sigma_{vu} \mathbf{u}. \quad (4.26)$$

Since both models are different linear approximations, we can expect differences between the models' appearance with regard to shape and texture. Furthermore, the parameter reductions that have been carried out by means of PCA may result in poorly correlated features between models. Evidently it is impossible to estimate features of the second model if these are not correlated with any features of the first. In these cases the cross-covariance is close to zero and consequently the corresponding features will be initialized by their mean.

#### 4.2.10 AAM search

Different methods can be used for the implementation of the AAM search algorithm. Our method is based on the example described in [66]. The goal of an AAM search is to minimize the difference between the model's gray-level appearance and the targeted object in the actual image by tuning the model parameters  $\mathbf{p}$ . The difference  $\mathbf{r}(\mathbf{p})$  is found by generating an instance of the model, and then comparing the model's texture levels  $\mathbf{g}_{\text{model}}$  with the corresponding pixel values  $\mathbf{g}_{\text{image}}$  in the real image:

$$\mathbf{r}(\mathbf{p}) = \mathbf{g}_{\text{image}} - \mathbf{g}_{\text{model}}. \quad (4.27)$$

Furthermore, we define the minimization criterion by calculating the sum of square differences

$$e^2(\mathbf{p}) = \mathbf{r}(\mathbf{p})^T \mathbf{r}(\mathbf{p}). \quad (4.28)$$

To simplify our calculations, we make use of the first order Taylor expansion of Equation 4.27, which is given by

$$\mathbf{r}(\mathbf{p} + \delta\mathbf{p}) = \mathbf{r}(\mathbf{p}) + \frac{\partial \mathbf{r}(\mathbf{p})}{\partial \mathbf{p}} \delta\mathbf{p}. \quad (4.29)$$

Next, to find a solution for  $\delta\mathbf{p}$  that minimizes the squared difference  $e^2$ , we need to minimize  $|\mathbf{r}(\mathbf{p} + \delta\mathbf{p})|^2$ . This is done by equating Equation 4.29 to zero, which results in the following solution:

$$\delta\mathbf{p} = -\mathbf{R}\mathbf{r}(\mathbf{p}), \text{ with} \quad (4.30)$$

$$\mathbf{R} = \left( \frac{\partial \mathbf{r}(\mathbf{p})}{\partial \mathbf{p}}^T \frac{\partial \mathbf{r}(\mathbf{p})}{\partial \mathbf{p}} \right)^{-1} \frac{\partial \mathbf{r}(\mathbf{p})}{\partial \mathbf{p}}^T. \quad (4.31)$$

Note that according to the previous solution, one needs to know matrix  $\mathbf{R}$  for all possible parameter instances  $\mathbf{p}$ . In general, this is computationally too demanding to determine during search, therefore we estimate  $\mathbf{R}$  beforehand, by assuming that it is approximately the same for all instances. To find an estimate

for  $\mathbf{R}$ , we determine the average of  $\frac{\partial r(\mathbf{p})}{\partial \mathbf{p}}$  for the examples of the training set:

$$\frac{dr(\mathbf{p})_i}{dp_j} = \sum_k \zeta(\epsilon_{jk})(r_i(\mathbf{p} + \epsilon_{jk}) - r_i(\mathbf{p})), \text{ with} \quad (4.32)$$

$$\zeta(\epsilon_{jk}) \propto \frac{\exp(-\epsilon_{jk}/2\sigma_j^2)}{\epsilon_{jk}}, \quad (4.33)$$

where  $\zeta(x)$  is a Gaussian weighing function,  $r_i$  an element from the difference vector (Equation 4.27),  $\sigma_j^2$  the variance of the pertinent parameter  $p_j$ , and  $k$  the number of parameter deviations tested. The parameter changes  $\epsilon_{jk}$  used for our experiments are  $\pm 0.2\sigma_j$ ,  $\pm 0.5\sigma_j$  and  $\pm 0.7\sigma_j$ .

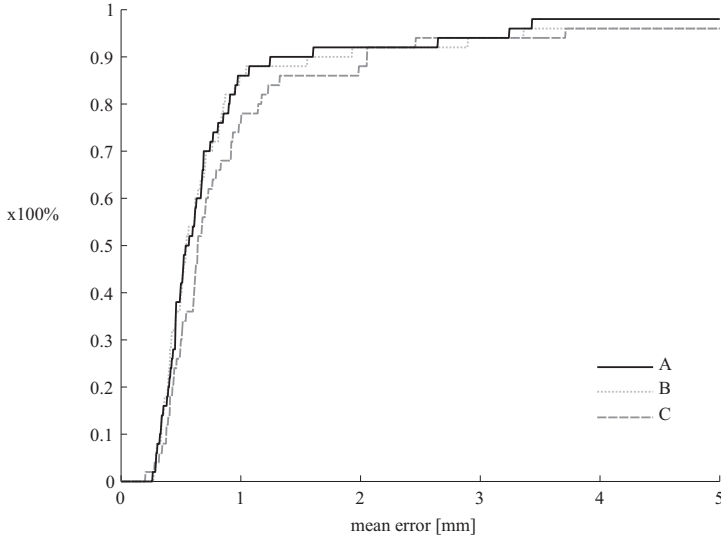
The linearity assumption works best for small parameter differences. For larger differences the estimate of  $\delta\mathbf{p}$  will be poor. Therefore, during the search it is necessary to use several iterations to converge to a (local) solution. As there is a risk of repeatedly overestimating  $\delta\mathbf{p}$ , which would result in poor convergence (due to oscillation), we multiply  $\delta\mathbf{p}$  with a scalar attenuation coefficient. Initially this coefficient is set to one, but as the overall difference  $e$  (Equation 4.28) does not decrease between successive iterations, this coefficient is decreased by 30%. The iterative process is stopped when there has not been a decrease of  $e$  for three following iterations with decreasing step size.

We initialize the first AAM search with the low-detail model by setting the shape and texture parameters to their mean (zero). The model's pose parameters for rotation and scaling are also set to their mean, which are calculated from the training set. The overall translation parameters are calculated by aligning the centers of gravity of the model to that of the image. For this computation, a 3 cm margin is omitted in the test image to compensate for possible artifacts, such as name tags in the image.

## 4.3 Results

Experiments were performed on the dataset of 100 radiographs described in Section 4.2.1. This dataset was split in a training set of 50 images for the AAMs and a test set of 50 images for testing.

Firstly, we performed experiments with the succession of the low-detail and high-detail AAMs. The first search step with the low-detail model detected all bones correctly in 38 images of the test set. A result was considered correct, when each bone had been enclosed in the corresponding bounding box of the low-detail AAM. This was determined by using the available contour landmarks of the dataset (Section 4.2.2). On average, convergence was reached within 8 iterations, taking about 0.3 seconds per iteration in our development environment. Eight images showed small failures with the detection of the phalanges; the detected locations were off by less than 0.5 cm. In the other four images larger search failures were caused by poor initialization; the initial model had started too far off



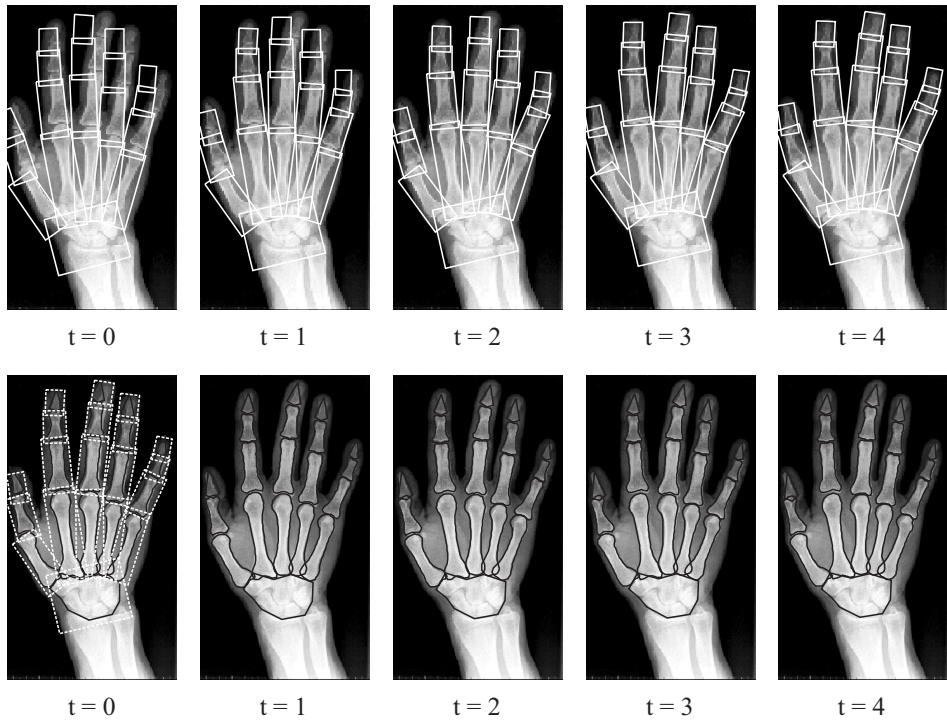
**Figure 4.12:** Cumulative histograms of the mean error for three different AAM search methods. Method A: low-detail AAM followed by high-detail AAM using pose invariant submodels. Method B: High-detail AAM using pose invariant submodels. Method C: Same as Method A, but without pose invariant submodels.

from the center of the hand and therefore did not find the fingers correctly. After reaching convergence, the search algorithm continued with the high-detail model. As mentioned earlier in Section 4.2.9, occasionally small segmentation differences were visible just after the model transition. In most cases this became corrected after one or two further iterations.

The second search step with the high-detail model converged within 7 iterations on average, taking 2.5 seconds for each iteration. The 38 images that were correctly segmented in the first search also succeeded in the second. Also, five of the images that initially showed small failures were segmented successfully in the second search. For these 43 images, 70% of the landmarks were found within 1.3 mm distance from manual segmentation. In Figure 4.12 the results are displayed by the curve marked as method A. This graph shows the cumulative histogram of the mean error of the landmarks with respect to the manual outlining of the bone contours. Figure 4.13 shows the results of one of the experiments after several iterations during the two search levels. Several gross errors occurred when one or more fingers of the model converged to the wrong location, Figure 4.14 shows two examples.

Secondly, we also tested two alternative AAM search configurations to determine the effects of the multi-level search strategy and the pose invariant submodels. In the first alternative, method B, we used the described high-detail model without

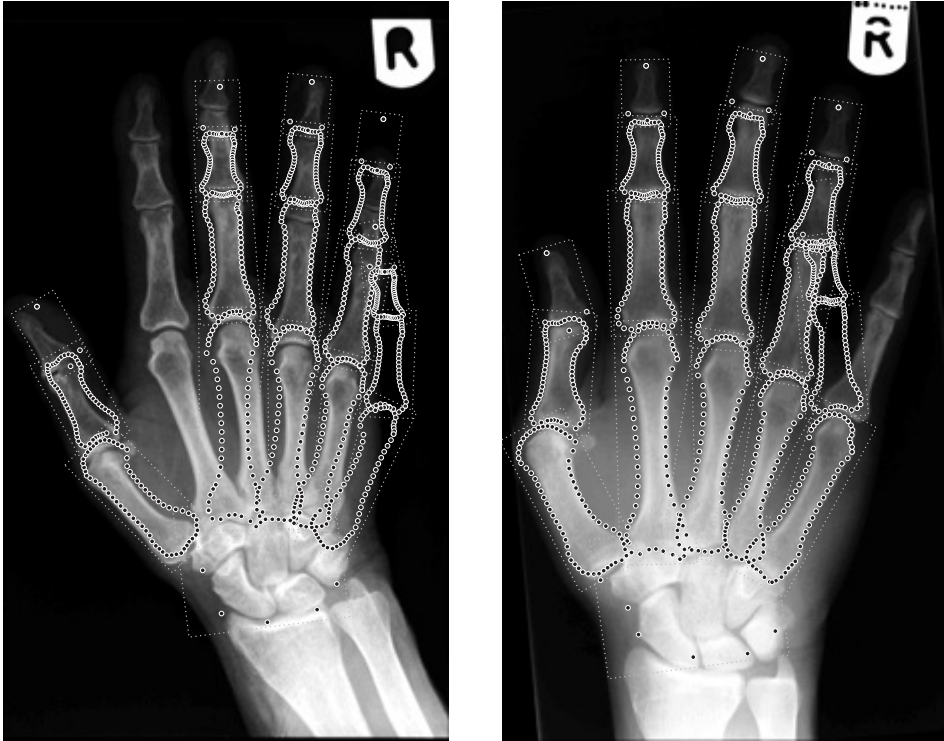




*Figure 4.13: Example of a successful AAM search initialized with the low-detail model (upper series in white) and followed by the high-detail model (lower series in black) after  $t$  iterations. The landmarks of the submodels have been connected and are depicted as contours. The models' textures have not been displayed in these images.*

the initialization by the low-detail model. The segmentation results were similar as with method A, as is shown in Figure 4.12. In two cases additional errors occurred in the detection of one or two of the fingers. As we did not make any changes to the high-detail model, this may indicate that the preceding low-detail model leads to a better initialization (on visual inspection of the iterative process, this seemed to be the case for these images). Furthermore, the execution time increased: on average it took 16 iterations before convergence was reached at an average 2.5 seconds for each iteration. Overall, the single AAM search method is about two times slower than the multi-level method.

In the second alternative, method C, we modified method A by leaving out the alignment step for the submodels the high-detail model. This makes the model variant for pose variabilities. This configuration performed worse than the other methods. Figure 4.12 shows the results of these tests. There was no noticeable change in execution time compared to method A.



*Figure 4.14: Two examples of errors that occurred during the experiments. In the left the AAM cannot converge correctly, as three fingers have converged to the wrong locations. In the right pane the little finger is ‘stuck’ onto the ring finger.*

## 4.4 Discussion and conclusion

The results of our experiments show that in most cases the AAM search method is able to find the specified landmarks within a few pixels from manual segmentation by a trained person. The maximum texture resolution of 16 pixels per centimeter in the second search step sets a limit to the accuracy of the results. Increasing the resolution will improve accuracy, but will also demand extra processing efforts.

Using two search levels of increasing model quality showed to be effective in reducing the calculation time. The search with the combination of the two models performs about twice as fast as the one with the single high-detail model without negatively affecting the results. In a few cases results improved, seemingly due to a better initialization of the high-detail AAM. As both models are different linear approximations, switching between the two models sometimes results in minor deviations between their appearances. Usually this deviation is corrected within

one or two iterations.

To compensate for large variabilities caused by differences in finger positioning, we made several submodels of the fingers invariant to rotation and translation. Our results show that this improves detection without affecting the algorithm's execution time.

Some small segmentation failures may be caused by a too poorly trained model and a limited number of permitted variations. This may be improved by increasing the size of the training set. On the other hand, since we are trying to segment hands affected by RA, it can be expected that rare appearances occur. Although the training set consists of RA hands, it is (even with a larger data set) unlikely that all possible variability can be incorporated in the model.

Some other failures may be caused by a poor initialization. As a result of this the AAM search can converge to a false local minimum. Except for altering the initialization method, one could also try multiple different initializations (for instance by using small initial displacements) and subsequently pick the best of the results.

The current method's accuracy is suitable for detecting regions of interest such as bones and joints. If we want to measure JSWs, we will need to be able to detect the contours more accurately. In Chapter 6 we continue this work with a local image processing algorithm to detect the joint margins.



# 5

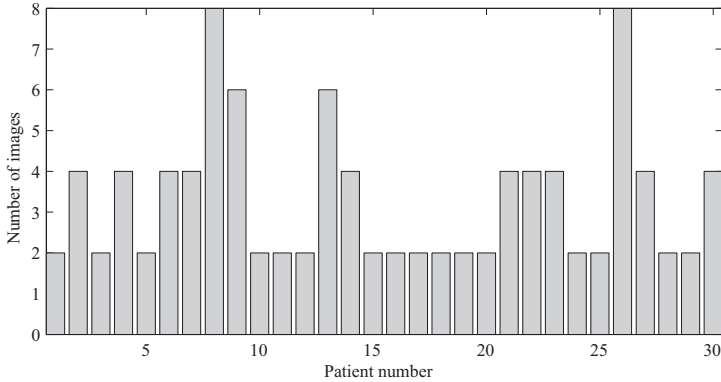
## Biometrics of the hand skeleton

### 5.1 Introduction

Large databases of patient records used in clinical trials may contain a significant number of false entries and other inconsistencies. In the process of data acquisition, data analysis and administration there are several stages where patient data can be filed incorrectly. Particularly when information is passed through by paper forms (illegible handwriting) or manual entry, there is a potential risk for errors, such as misspelled names, swapped records, and incorrect numbers. As a result the outcome of a medical trial may be compromised.

There exist several statistical methods to search through databases for unusual deviations in numbers, but this is less straightforward for other information in patient records, such as radiographs. In our case, we are interested in verifying databases of hand radiographs. Our goal is to identify possible errors, such as double entries (one patient filed under more than one name), wrong patient labels (different patients filed under the same name), and mirrored images (left and right mixed up). To accomplish this, we look for characteristic features in the shapes of the hand bones. Next we use a classification method to compare these 'biometric features' of the images in our dataset. Hereafter we determine for each image how likely it is that it has been filed correctly.

Using the shape of hands for biometric verification and identification is not new. Similar methods exist for using the outer geometry of the hand [74, 75]. In [51] it is suggested that hand radiographs can be used for this purpose, but thus far we are not aware of any other research on this subject.



*Figure 5.1: Graph displaying the number of available images for each patient entry in the dataset.*

Besides detecting patient database inconsistencies and preventing faulty entries, biometric identification of bones may be useful for forensic applications. Other applications may be found in the security field. Though x-rays are seldom used for this purpose, sometimes low dose x-ray scanners are used for searching people for weapons and contrabands. A biometric identification system could be a valuable extension to such systems.

## 5.2 Methods

### 5.2.1 Data

For our experiments we have used a set of 100 posterior anterior single hand radiographs. This set consists of 50 pairs of hands originating from 30 patients diagnosed with RA. Figure 5.1 shows the number of available radiographs per patient, which varies between one and four image pairs that have been taken with several years in between. Though most images have been labeled ‘left’ or ‘right’, we remove this distinction by mirroring all (apparent) left hand images, such that all thumbs appear on left side of the hand. By this all images can be treated the same, which also makes it possible to compare both hands with each other.

### 5.2.2 Biometric features

To be able to compare the radiographs with one another, we make use of contour descriptions of the metacarpals, proximal and middle phalanges. To obtain these descriptions, we use the same method of landmark selection as used for the AAMs as described in Section 4.2.2. This results in 14 contour descriptions of 64 points

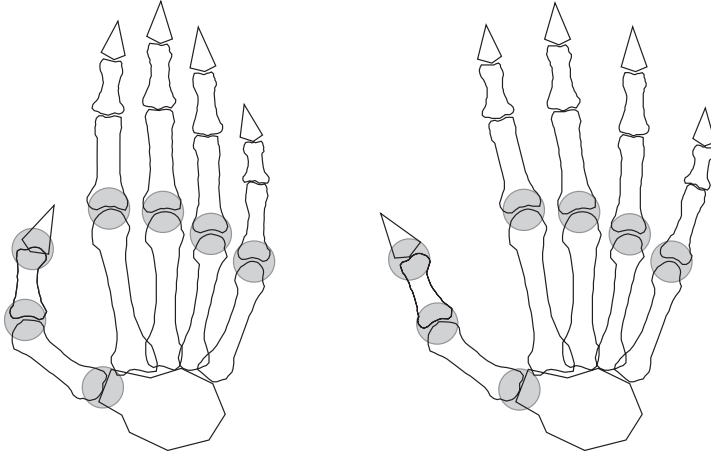


**Figure 5.2:** The contours of 14 bones (metacarpals, proximal and middle phalanges) are described by 64 points each.

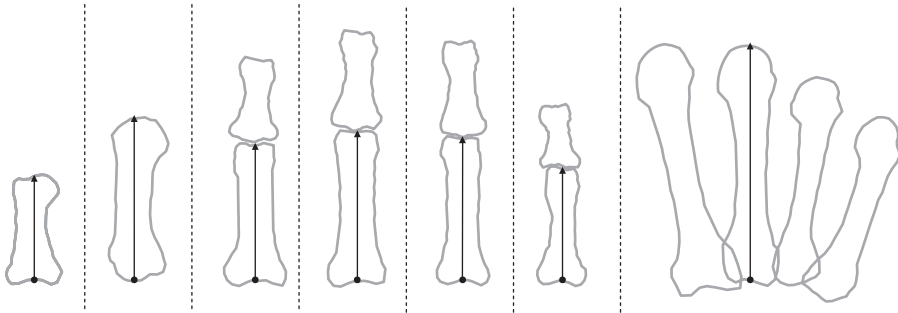
each, Figure 5.2 shows the locations of these points in one of the radiographs.

As we are merely interested in anatomical, biometric features, we have to compensate for the variability that is caused by differences in overall pose and finger positioning. Since generally the hand is placed with the palmar side flat on the table, the mobility of most joints is considerably reduced. The joints and bone groups that are relevant to positioning variability are displayed in Figure 5.3. Note that the phalanges of a finger are mutually in a fixed position when the hand is placed on a flat surface, and that therefore their mutual position can be considered as a characteristic feature. This is similar for the fixed block of metacarpals 2–4. Subsequently, by ‘dissecting’ the articulating parts we obtain seven groups of landmarks that consist of one or more bones (Figure 5.4). We have left out the distal phalanges and the carpal landmarks, as in our data set they have too few landmarks for adequate contour descriptions.

To remove positioning variability, we subsequently translate and rotate each group to a common  $(x, y)$ -coordinate frame. To enable these operations we define an alignment axis for each group, as depicted in Figure 5.4. The direction of these axes are defined by the midline that runs from the proximal to the distal end of the (most proximal) bone. The midline of the third metacarpal defines the reference axis of the metacarpal block. The landmarks are translated and rotated such that



**Figure 5.3:** Variability in finger positioning due to articulation of the joints. The gray circles indicate the articulating joints.



**Figure 5.4:** The selected landmarks of each movable part is centered and rotated to a uniform position. The arrows indicate the medial axes of the bones used for alignment. The black dots mark the rotation origins.

their alignment axis is parallel to the  $y$ -axis and centered to the origin.

Using the aligned  $(x, y)$ -coordinates as characteristic features, we can now describe the bone contours (of a single hand radiograph) by a feature vector  $\mathbf{x}$  of  $N$  elements ( $N = 14 \times 64 \times 2 = 1792$ ).

$$\mathbf{x} = [x_1, \dots, x_N, y_1, \dots, y_N]^T \quad (5.1)$$



### 5.2.3 Classification

To decide whether or not an extracted feature vector originates from a certain patient, we make use of a likelihood-ratio classifier for Gaussian probability densities [76]. The likelihood ratio  $\Lambda$  can be expressed by

$$\Lambda = \frac{p(\mathbf{x}|q)}{p(\mathbf{x}|\bar{q})}, \quad (5.2)$$

where  $p(\mathbf{x}|q)$  is the likelihood that feature vector  $\mathbf{x}$  belongs to patient  $q$ , and  $p(\mathbf{x}|\bar{q})$  the likelihood that it does not. If we have a sufficiently large dataset of different patients, such that we may assume that the distribution of a single patient (intra-patient distribution) does not significantly influence the total distribution, then  $p(\mathbf{x}|\bar{q}) = p(\mathbf{x})$ . The likelihood ratio then becomes:

$$\Lambda = \frac{p(\mathbf{x}|q)}{p(\mathbf{x})}. \quad (5.3)$$

The probability density function for a multivariate normal distribution  $X \sim \mathcal{N}_N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  is given by

$$f(\mathbf{x}) = (2\pi)^{-\frac{N}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad (5.4)$$

where  $|\boldsymbol{\Sigma}|$  denotes the determinant of the covariance matrix. Now let us assume that the feature vectors of patient  $q$  and the total dataset are both multivariate normal distributed with means  $\boldsymbol{\mu}_q$  and  $\boldsymbol{\mu}_t$ , and covariance matrices  $\boldsymbol{\Sigma}_q$  and  $\boldsymbol{\Sigma}_t$ , respectively. We can rewrite the likelihood ratio from Equation 5.3 to

$$\Lambda = \frac{|\boldsymbol{\Sigma}_q|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_q)^T \boldsymbol{\Sigma}_q^{-1}(\mathbf{x} - \boldsymbol{\mu}_q)\right)}{|\boldsymbol{\Sigma}_t|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_t)^T \boldsymbol{\Sigma}_t^{-1}(\mathbf{x} - \boldsymbol{\mu}_t)\right)}. \quad (5.5)$$

The likelihood-ratio can be used to classify any feature vector  $\mathbf{x}$  by comparing it to a predefined threshold: If  $\Lambda$  exceeds this threshold, then  $\mathbf{x}$  is classified as being of patient  $q$ . If not, then it is from another patient. To simplify calculations with regard to this classification method, we reformulate Equation 5.5 using the log-likelihood ratio ( $LLR = -2 \ln \Lambda$ )

$$LLR = (\mathbf{x} - \boldsymbol{\mu}_q)^T \boldsymbol{\Sigma}_q^{-1}(\mathbf{x} - \boldsymbol{\mu}_q) - (\mathbf{x} - \boldsymbol{\mu}_t)^T \boldsymbol{\Sigma}_t^{-1}(\mathbf{x} - \boldsymbol{\mu}_t) + \ln \frac{|\boldsymbol{\Sigma}_q|}{|\boldsymbol{\Sigma}_t|}. \quad (5.6)$$

Note that the last term is a constant that can be accounted for in the selected threshold. Thus, what remains is the difference of two squared Mahalanobis distances [77].

The means and covariance matrices are unknown, and therefore have to be estimated from our data set. But, for an accurate estimation of the covariance

matrices, this would require many more training examples ( $\gg N$ ) than there are available in the data set.

To overcome this problem, we apply principal component analysis (PCA) on all samples of the data set [71] and reduce the dimensionality of the transformed feature vectors by removing the dimensions that express the lowest variance. For this purpose we use the feature vectors of  $M$  examples of the dataset ( $M$  depends on the type of experiment; see Section 5.3) to create matrix  $\mathbf{X}$  with their overall mean  $\boldsymbol{\mu}_T$  subtracted.

$$\mathbf{X} = [\mathbf{x}_1 - \boldsymbol{\mu}_T, \dots, \mathbf{x}_M - \boldsymbol{\mu}_T] \quad (5.7)$$

Through the singular value decomposition (SVD) of  $\mathbf{X}$ , solving

$$\mathbf{X} = \mathbf{U}_X \mathbf{S}_X \mathbf{V}_X^T, \quad (5.8)$$

we obtain  $N \times N$  orthonormal matrix  $\mathbf{U}_X$ ,  $M \times M$  orthonormal matrix  $\mathbf{V}_X$ , and  $N \times M$  non-negative matrix  $\mathbf{S}_X$  with decreasing singular values on the first diagonal. By taking the first  $N_{\text{PCA}}$  columns of  $\mathbf{U}_X$ , we can create an  $M \times N_{\text{PCA}}$  transformation matrix  $\tilde{\mathbf{U}}_X$ . The number of dimensions  $N_{\text{PCA}}$  is chosen such that the variance displayed in the training set is covered for 95% (in our experiments  $N_{\text{PCA}} = 25$ ). With  $\tilde{\mathbf{U}}_X$ , we can transform the data in  $\mathbf{X}$  to a new feature space with  $N_{\text{PCA}}$  dimensions:

$$\mathbf{Y} = \tilde{\mathbf{U}}_X^T \mathbf{X}. \quad (5.9)$$

Since the number of radiographs available per patient is insufficient to estimate patient specific covariance matrices ( $\boldsymbol{\Sigma}_q$ ), we estimate the average intra-patient variance. This is done under the assumption that the intra-patient variations are similar between patients. First, within the reduced feature space, we define matrix  $\mathbf{W}$  by

$$\mathbf{W} = [\mathbf{Y}_1 - \bar{\mathbf{Y}}_1, \dots, \mathbf{Y}_K - \bar{\mathbf{Y}}_K], \quad (5.10)$$

where  $\mathbf{Y}_i$  is the set of feature vectors of a single patient with mean  $\bar{\mathbf{y}}_i$  ( $\bar{\mathbf{Y}}_i$  is a matrix of the same width as  $\mathbf{Y}_i$  with  $\bar{\mathbf{y}}_i$  as columns), and  $K$  is the number of patients. Next, by SVD of  $\mathbf{W}$  we obtain two orthogonal matrices  $\mathbf{U}_W$  and  $\mathbf{V}_W$ , and a non-negative diagonal matrix of singular values  $\mathbf{S}_W$ :

$$\mathbf{W} = \mathbf{U}_W \mathbf{S}_W \mathbf{V}_W^T \quad (5.11)$$

To be able to normalize the intra-patient variance in the  $N_{\text{PCA}}$ -dimensional feature space, we can now determine a second transformation matrix  $\mathbf{U}_{\overline{W}}$ :

$$\mathbf{U}_{\overline{W}} = \sqrt{N-1} \mathbf{U}_W \mathbf{S}_W^{-1}. \quad (5.12)$$

Next, we can maximize the discrimination between the total distribution and the (average) intra-patient distribution by applying linear discriminant analysis (LDA) [78]. First, we create a submatrix  $\tilde{\mathbf{U}}_{\overline{W}}$  from the first  $N_{\text{LDA}}$  (principal) column vectors of  $\mathbf{U}_{\overline{W}}$  (in our experiments  $N_{\text{LDA}} = 20$ ). If we apply transformation

matrix  $\tilde{\mathbf{U}}_{\tilde{W}}$  to the feature vectors of  $\mathbf{Y}$ , we obtain a set of features  $\mathbf{Z}$  within an  $N_{\text{LDA}}$ -dimensional feature space:

$$\mathbf{Z} = \tilde{\mathbf{U}}_{\tilde{W}}^{\text{T}} \mathbf{Y} \quad (5.13)$$

If we apply our classification model (Equation 5.6) in this new feature space, the estimated intra-patient variance will be unity in all dimensions. The total distribution, on the other hand, will still show correlations. To be able to remove these correlations, we calculate the SVD of  $\mathbf{Z}$ :

$$\mathbf{Z} = \mathbf{U}_Z \mathbf{S}_Z \mathbf{V}_Z^{\text{T}} \quad (5.14)$$

This gives us an orthogonal transformation matrix  $\mathbf{U}_Z$  and a non-negative diagonal matrix of singular values  $\mathbf{S}_Z$ . The latter can be used to calculate the (diagonal) covariance matrix

$$\Sigma_Z = \frac{1}{N-1} \mathbf{S}_Z^2. \quad (5.15)$$

The described sequence of transformations and dimension reductions can be combined in a single transformation matrix

$$\mathbf{T} = \mathbf{U}_Z^{\text{T}} \tilde{\mathbf{U}}_{\tilde{W}}^{\text{T}} \tilde{\mathbf{U}}_X^{\text{T}}. \quad (5.16)$$

Now, let  $\mathbf{u}$  and  $\mathbf{v}$  be the transformed input-feature vectors with subtracted means of the patient distribution and the total distribution:

$$\mathbf{u} = \mathbf{T}(\mathbf{x} - \boldsymbol{\mu}_q), \quad (5.17)$$

$$\mathbf{v} = \mathbf{T}(\mathbf{x} - \boldsymbol{\mu}_t), \quad (5.18)$$

where  $\boldsymbol{\mu}_p$  is the mean feature vector of a patient to be matched and  $\boldsymbol{\mu}_t$  the mean of the total dataset. Subsequently, we can define a classification score  $s$  based on the LLR (Equation 5.6) by calculating the Mahalanobis distance

$$s(\mathbf{x}) = \mathbf{u}^{\text{T}} \mathbf{u} - \mathbf{v}^{\text{T}} \Sigma_Z^{-1} \mathbf{v} - \ln |\Sigma_Z|. \quad (5.19)$$

In the ideal case a threshold value should be chosen for the classification score such that a perfect discrimination can be made. In practice, both distributions can overlap, making it necessary to determine the desired balance between the classifier's sensitivity and specificity.

## 5.3 Experiments and results

Three different experiments have been performed with the described classification method to investigate the uniqueness of a hand's skeletal shape as shown in a radiograph. The goal of the first experiment is to determine how well single hands can be discriminated from others in the dataset (including the patient's other

hand). Besides the discriminative properties between patients, this experiment also indicates how well left and right hands can be discriminated. The opposite has been investigated in the second experiment by matching hands of one side to the other in order to find the correct pairs. For this we have assumed that both hands are highly symmetrical. In the last experiment the distinction between left and right has been left out to discriminate between patient features only.

### 5.3.1 Cross verification of single hands

For this experiment we cross-verify all radiographs in the data set to possibly detect whether images have been filed incorrectly. Errors may be caused by incorrect patient naming or false left/right labeling. The overall distribution is computed using all feature vectors of the dataset. As each hand is classified separately, the intra-patient distribution is computed from the patient data where multiple images of the same hand were available. For 16 patients only a single pair of radiographs was available and therefore these could not be used in this computation. Each radiograph is tested against both (left and right) classes of each patient in the dataset. To avoid any direct bias of the evaluated feature vector, the vector being tested is excluded from the computed mean of the corresponding class.

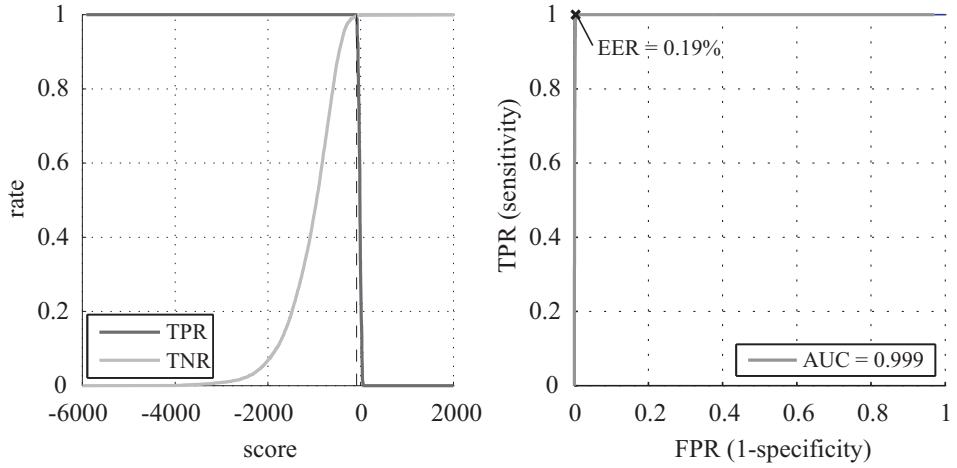
During the first run of this experiment, we found that one pair of radiographs did not match the others of this specific patient. Instead we found that this pair showed a good match to the radiographs of another patient. After inquiry in the originating hospital's records, our findings were confirmed: it appeared that the radiographs had been filed incorrectly.

After correcting the dataset for these errors, we repeated the experiment. The graph on the left of Figure 5.5 displays the resulting true-positive-rate (TPR) and the true-negative-rate (TNR) in relation to the calculated similarity scores. The graph on the right shows the receiver operating characteristic (ROC) with the equal error rate (EER) at 0.19% and the area under curve (AUC) close to one. The EER is the rate where the false positive rate (FPR) equals the false negative rate (FNR).

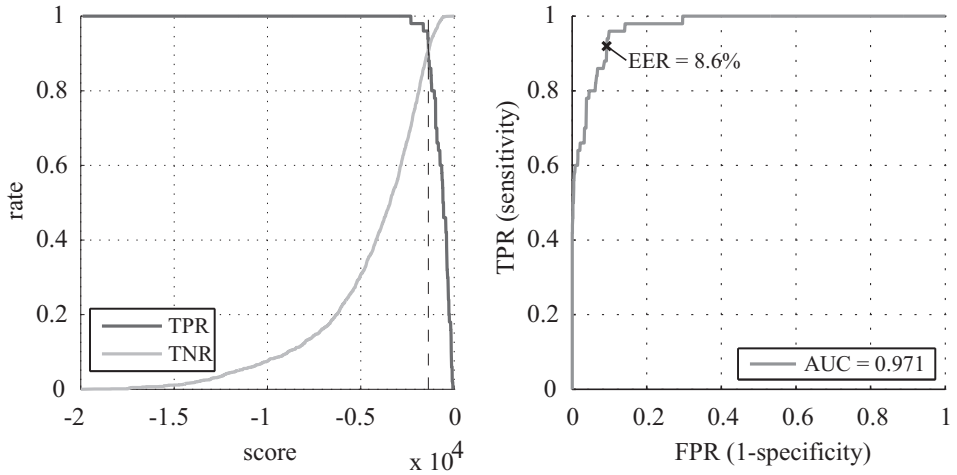
### 5.3.2 Matching opposing hands

In this experiment we have tested if it is possible to identify patients by their left hand using the features of their right hand and vice versa. To investigate this, we first included only the 50 right hand images in the training set for determining the average intra-patient variability and the individual patient templates. The remaining 50 left hand images have been tested one by one to fit any of the patient classes of the training set. The same experiment is repeated with the left hand images in the training set and the right hands in the test set.

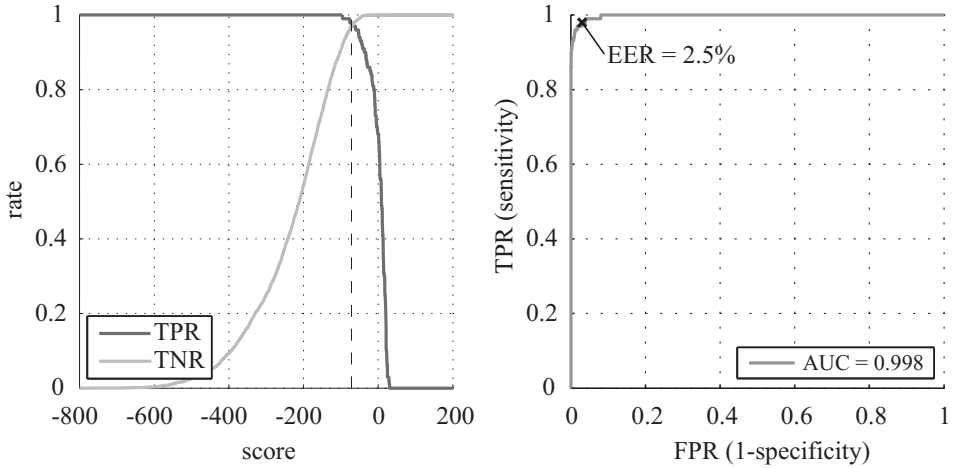
Figure 5.6 shows the combined results of these experiments. Again the left graph shows the TPR and the TNR, and the right graph the ROC with the EER at 8.6% and an AUC of 0.971.



*Figure 5.5: First experiment: cross verification of single hands. The graph on the left displays the TPR and the TNR with respect to the similarity score. The graph on the right displays the ROC with the AUC and the EER.*



*Figure 5.6: Second experiment: matching opposing hands. The left graph displays the TPR and the TNR with respect to the similarity score. The right displays the ROC with the AUC and the EER.*



*Figure 5.7: Results of third experiment: cross verification of combined hands. The left graph displays the TPR and the TNR with respect to the similarity score. The right graph displays the ROC with the AUC and the EER.*

### 5.3.3 Cross verification of combined hands

For this final experiment, intra-patient classes have been formed by omitting the distinction between left and right hands. By doing this, more examples become available for the estimation of the average intra-patient distribution, as also patient data with only one pair of radiographs can be included. This approach is expected to increase the accuracy of the estimated intra-patient variance at the cost of the accuracy of the mean of the class. Each radiograph has been cross verified with all patient classes. As with the first experiment, any direct bias was avoided by not including the test feature in the calculation of its mean class.

The resulting TPR and the TNR with respect to the similarity score are displayed in the left graph of Figure 5.7. The right graph displays the ROC showing an EER of 2.5% and an AUC of 0.998.

## 5.4 Discussion and conclusions

In this chapter we have demonstrated a method to extract biometric features from bone shapes in hand radiographs. To anticipate variability in hand and finger positioning, we made this method invariant to positioning variability. By applying a classifier based on the likelihood ratio we were able to verify a patient's identity and to detect possible inconsistencies in a patient database with high certainty. As a proof of concept, we have found an error in our dataset.

Based on the first experiment (Section 5.3.1) we can conclude that there is

---

sufficient discriminative information in the bone shapes to distinguish (mirrored) left and right hands of the same patient. On the other hand, the second experiment (Section 5.3.2) shows that in many cases (8.6% EER) both hands are sufficiently symmetrical to match one hand with the other. This latter property may be particularly useful in certain forensic applications. For the third experiment (Section 5.3.3) we combined the properties used in the first two experiments, and showed that the classifier can be trained such that it is possible to recognize a patient by any of his hands.

In practice the number of examples per subject is usually small for radiographic datasets, therefore it is necessary to make use of the average intra-patient variance instead of the patient specific intra-class variance. Though, we expect that results improve when a larger dataset is used with more available radiographs per patient. As with a larger number of examples, the estimations of both the overall distribution and the average intra-patient distribution become more accurate.





# 6

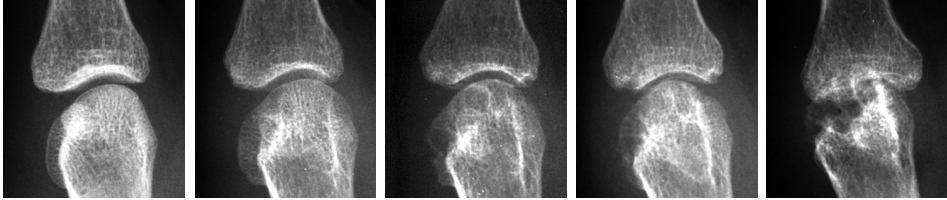
## Margin detection

### 6.1 Introduction

It is essential that joint degenerative diseases, such as RA, are treated in an early stage, ideally before any irreversible joint damage occurs. To assess effectiveness of drug-treatment it is necessary to precisely monitor the progression of the disease. Currently there exist several visual scoring methods [15] to quantify joint damage in radiographs of hands and feet. In general, these methods are time-consuming and depend on subjective measurements. The sensitivity to change [24, 26, 27] of these methods is highly dependent on intra-observer and inter-observer variability. As of this reason, in the past two decades researchers have been looking for automated methods to measure joint damage in a more objective manner (see also Section 2.4).

An important measurable effect of RA is joint space narrowing. Already in an early stage of RA, the loss of cartilage in the joints can be determined indirectly by measuring the joint space width (JSW) [20]: the distance between the joint margins. Figure 6.1 shows an example of how the JSW decreases with the progression of RA. It is important to note that the absolute value of the JSW is not a measure for RA. It is the decrease in JSW over time that provides an indication of the disease's activity and progression.

Several researchers have experimented with (partially) automated methods [32, 35, 37, 15] to measure the JSW in hand radiographs and have shown that the sensitivity of such methods is higher than manual methods [40]. Though progress is made in this research, there are currently no methods available for



*Figure 6.1: Follow up series of radiographic images of the second MCP joint of a patient with progressive RA (from left to right with approximately two years intervals).*

practical use. Except for various limitations within these methods, validation is also difficult due to the absence of a gold standard. In an attempt to work towards a clinically accepted standard, we work with other researchers in a special interest group of the international network of Outcome Measures in Rheumatology Clinical Trials (OMERACT) [79]. In this group experiences with different methods are exchanged and results are compared in order to reach international consensus on future measurement methods (Section 6.4).

Our goal is to develop an automated method to measure joint space narrowing in hand radiographs with a high sensitivity to change. In previous research, see Chapter 4, we have presented a segmentation algorithm to detect the bones in hand radiographs [93]. We have found that this method was precise enough for detecting regions of interest such as the joints, but not for quantitative JSW measurements. In this chapter we proceed with the joints and present a robust and precise method for detecting joint margins and measuring the distance between them.

## 6.2 Joint margin detection

For JSW measurements we wish to measure the distance between the surfaces of the bones that make the joint. Since radiographs are two-dimensional projection images of three-dimensional structures, the definition of the joint margins is rather ambiguous. On the other hand this is not truly essential for relative measurements, providing that the precision and reproducibility of the detection method is sufficient. The reliability of the measurement may also be affected by the positioning of the hand. Other research [40] has shown that JSW measurements are highly reproducible, even when there are small variations in the hand positioning.

In the following subsections we present an iterative search method using modified active shape models [50] which are based on statistical properties of a set of example images. The detection of joint margins is done within several regions of interest which are determined from multiple initialization points. We demonstrate how different shape models can be created for the various joint margins, and how these models can be applied to detect plausible margin shapes. To find the margins, we look for nearby joint margins by scanning along lines perpendicular to an

initial estimate of the margin. For each scan point we use its neighboring pixels to determine a likelihood score based on a statistical distance measure. Next we apply a dynamic programming method to find a new margin location which can be translated to a plausible margin corresponding to the shape model. This process is repeated for several iterations until the difference between results in consecutive iterations reaches a minimum or until a maximum number of iterations is reached. Finally we determine the distance between each margin pair.

### 6.2.1 Image data set

For this investigation we have used 100 plain single hand radiographs in posteroanterior view. These are from both left and right hands of 40 different patients diagnosed with RA. For several patients there were multiple image pairs available of different time points. To be able to use the same method for both hands, the left hand images were mirrored.

The radiographs have been digitized on a 12-bit grayscale scanner at a resolution of 600 dots per inch. Linear contrast enhancement was applied such that the full intensity range was used. The images were stored in a lossless 8-bit grayscale format.

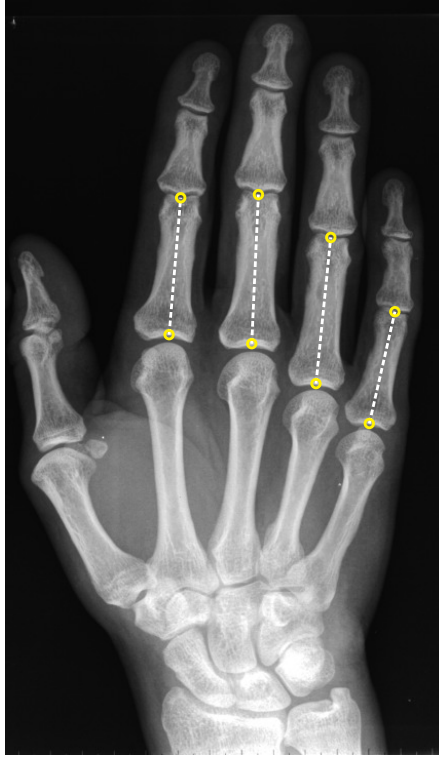
The set of images was split randomly into two independent sets of 50 images of 20 patients; a training set for the extraction of features for setting up the parameters of the applied statistical methods, and a test set for the experiments.

### 6.2.2 Initialization

For each hand we measure the JSW at eight locations: at metacarpophalangeal (MCP) joints 2–5 and at proximal interphalangeal (PIP) joints 2–5. For the detection of the joint locations, we use the method described in Chapter 4 and [93]. With this method we are able to detect the phalanges and metacarpal bones in most of the images used in this research. A few detection errors were corrected by manually indicating the initialization points. For a correct working of the JSW measurement algorithm eight initialization points are required. These points are defined by the proximal and distal ends of the proximal phalanges and are on the central axis of the bone (Figure 6.2). As the definition of the location of these points may be ambiguous, the method allows small deviations up to 1 mm in any direction.

### 6.2.3 Margin shape

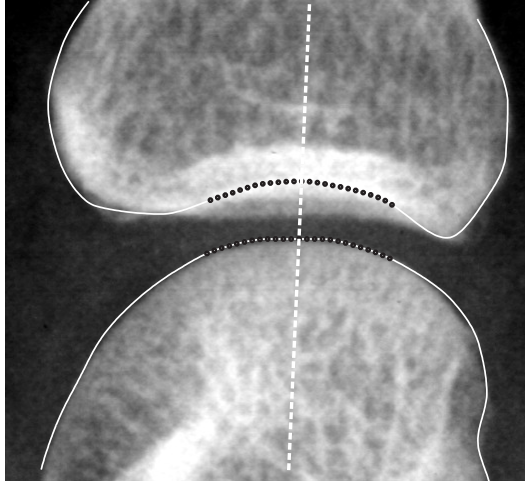
To prevent the algorithm from finding false edges, we constrain the shapes of the detectable margins to statistical shape models based on the training set of 50 example radiographs. In these sample images the joint space margins were manually outlined by an expert. Since it is difficult to define where the joint space begins and ends, the JSW measurement is limited to a region of 6 mm around the



*Figure 6.2:* The proximal and distal ends of the proximal phalanges are used for the initialization of the joint locations and margin angles.

joint's center. The central axis of the proximal phalanx, which is defined by the initialization points (Figure 6.2), is used to determine the centers of the margins. Others have used different methods [37, 39] to define this region based on the size and shape of the joint, but we expect better robustness using a fixed region size. Also, the locations of the joint margins are less clear near the sides of the joint space, due to the projection view of the joint.

The shape of each joint margin is characterized using  $N$  equidistant points (landmarks). In the shown examples and for our tests we use 25 points (Figure 6.3). Separate shape models [50] are created for the proximal and distal margins of all eight joints (MCP 2–5 and PIP 2–5). To remove translation and rotation variability, all 50 example margin shapes are aligned to a common co-ordinate frame. This is done by first translating all shapes such that their center of gravity is on a common origin. Next the shapes are rotated such that the central axis of the proximal phalanx aligns with the  $y$ -axis. The  $x$  and  $y$ -coordinates of the



**Figure 6.3:** The set of training images with manually outlined margins is used to create the shape model. The proximal and distal margins are both characterized by  $N$  equidistant points over 6 mm around the center of the joint ( $N = 25$ ).

landmarks of margin shape  $i$  are stored in a  $2N$  element vector  $\mathbf{x}_i$ .

$$\mathbf{x} = (x_1, x_2, \dots, x_N, y_1, y_2, \dots, y_N)^T \quad (6.1)$$

The mean shape  $\bar{\mathbf{x}}$  is calculated for each margin:

$$\bar{\mathbf{x}} = \frac{1}{50} \sum_{i=1}^{50} \mathbf{x}_i. \quad (6.2)$$

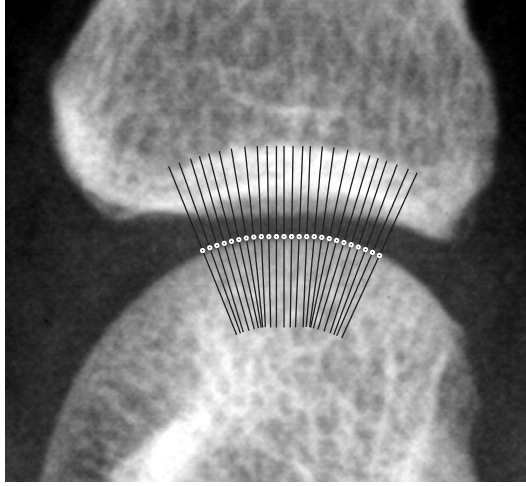
Next, data matrix  $\mathbf{X}$  is created with vectors  $\mathbf{x}_i$  as columns, and an equal size matrix  $\bar{\mathbf{X}}$  with vectors  $\bar{\mathbf{x}}$ . We use the singular value decomposition (SVD) [71] of  $\mathbf{X} - \bar{\mathbf{X}}$  to find an orthogonal matrix  $\Phi$  of eigenvectors  $\phi_i$  and the corresponding eigenvalues  $\lambda_i$  (with  $\lambda_i \geq \lambda_{i+1}$ ). We truncate  $\Phi$  to the first  $Z$  eigenvectors, such that the total variance covered in the truncated matrix  $\Phi_Z$  is 99%. In our case six eigenvectors were sufficient for the MCP joints and 8 for the PIP joints.

Approximations  $\tilde{\mathbf{x}}$  of  $\mathbf{x}$  can now be generated using parameter vector  $\mathbf{s}$ :

$$\tilde{\mathbf{x}} = \bar{\mathbf{x}} + \Phi_Z \mathbf{s}. \quad (6.3)$$

With this parametric description of the margin shapes we can determine the closest possible shape to a set of newly found co-ordinates  $\mathbf{y}$ . To achieve this,  $\mathbf{y}$  is projected into the parameter space to obtain  $\mathbf{s}$ :

$$\mathbf{s} = \Phi_Z^T (\mathbf{y} - \bar{\mathbf{x}}). \quad (6.4)$$



**Figure 6.4:** For each landmark an intensity profile is extracted along a 6 mm line perpendicular to the margin shape.

To make sure that only plausible shapes are found, we truncate the shape parameters of  $\mathbf{s}$  such that  $|s_i| \leq 3\sqrt{\lambda_i}$ . Following, a new estimate of the margin shape can be calculated through Equation 6.3.

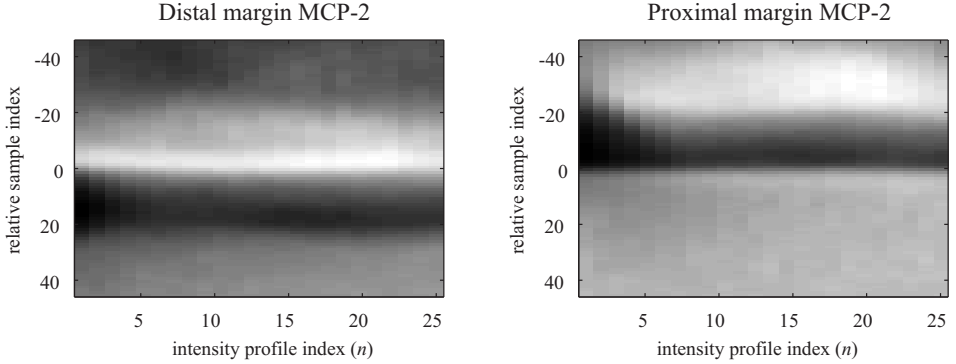
## 6.2.4 Margin detection

To detect plausible margin locations, we use a probability score based on the Mahalanobis distance [77]. To achieve this we use the intensity profiles along lines perpendicular to an estimated joint margin (Figure 6.4). In the first iteration this estimate is the mean  $\bar{\mathbf{x}}$ . Each intensity profile is sampled along a line of 6 mm using bilinear interpolation at  $L$  points with equal spacing  $s_L$ . To compensate for differences in image level and contrast settings, we remove any offset and normalize the intensities with the standard deviation of all intensity values sampled for the pertinent joint. For each point of the margin shape the resulting intensity profile is stored in a vector

$$\mathbf{g} = (g_1, g_2, \dots, g_L)^T. \quad (6.5)$$

In our tests we sample the profiles over 6 mm with a spacing of 15 points per millimeter (thus  $s_L \approx 0.067\text{mm}$  and  $L = 91$ ). Next, we determine the mean profile of the 50 examples that we have in our training set

$$\bar{\mathbf{g}} = \frac{1}{50} \sum_{i=1}^{50} \mathbf{g}_i. \quad (6.6)$$



**Figure 6.5:** The left image shows the average intensity profiles of the distal margins for the MCP-2 joint. The right image shows the corresponding profiles for the proximal margins.

Figure 6.5 shows all mean profiles of the proximal and the distal margins of the second MCP joint. The Mahalanobis distance for the profile at margin point  $n$  is then defined as

$$D_n(\mathbf{g}_t) = \sqrt{(\mathbf{g}_t - \bar{\mathbf{g}}_n)^T \boldsymbol{\Sigma}_{g_n}^{-1} (\mathbf{g}_t - \bar{\mathbf{g}}_n)}, \quad (6.7)$$

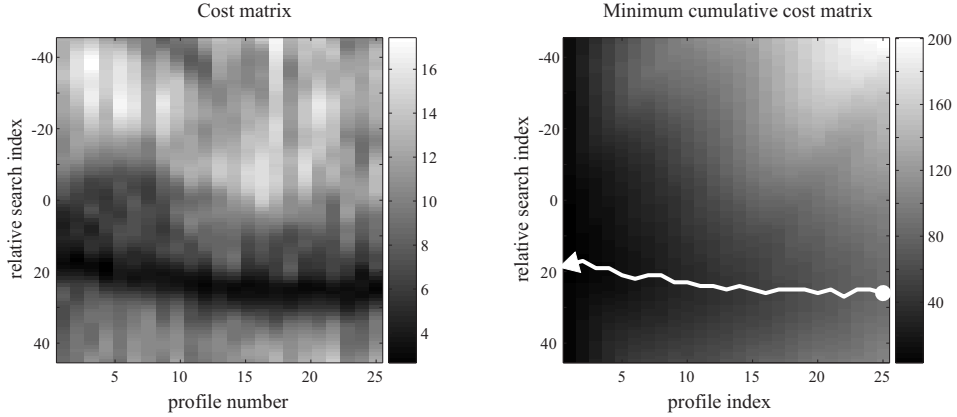
where  $\boldsymbol{\Sigma}_{g_n}$  is the covariance matrix of the sample profiles in the training set. For a good estimate of the inverse covariance matrix in principle we need a large training set. Therefore we reduce the number of dimensions in a similar manner as we do for the margin description in Section 6.2.3. For each margin point a data matrix  $\mathbf{G}_n$  is created with sample profiles  $\mathbf{g}_{n,i}$  as columns, and an equal sized matrix  $\bar{\mathbf{G}}_n$  with vectors  $\bar{\mathbf{g}}_n$  as columns. Through the SVD of  $\mathbf{G}_n - \bar{\mathbf{G}}_n$  we find  $N$  orthogonal matrices of eigenvectors  $\tilde{\boldsymbol{\Psi}}_n$  with the eigenvalues  $\lambda_{n,i}$ . Next, we truncate  $\tilde{\boldsymbol{\Psi}}_n$  to the first principal  $W$  eigenvectors covering 98% of the variance in the example set. In our case 20 eigenvectors are sufficient for both the MCP joints and the PIP joints.

Suppose we have obtained test profile  $\mathbf{g}_t$ , we can transform this to a point  $\mathbf{g}_r$  in the reduced parameter space by

$$\mathbf{g}_r = \tilde{\boldsymbol{\Psi}}_n^T (\mathbf{h}_n - \bar{\mathbf{g}}_n). \quad (6.8)$$

Since there is no correlation between the elements of  $\mathbf{u}_n$ , we can now determine a probability score by calculating the normalized Euclidean distance in the parameter space:

$$\hat{D}_n(\mathbf{g}_r) = \sqrt{\sum_{w=1}^W \frac{g_{r,w}^2}{\lambda_{n,w}}}. \quad (6.9)$$



**Figure 6.6:** The left image shows the cost matrix. The right image shows the minimum cumulative cost matrix and minimum cost path, which is traced back from the position of the minimum in the last row of the matrix.

## 6.2.5 Search

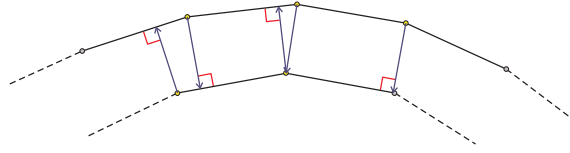
The search process for possible margin point candidates goes along the direction of the profile lines. This is done by sampling  $M$  profiles that are shifted with step size  $s_M$  in both directions. In the tests we use a search length of 6 mm with a resolution of 15 pixels per millimeter (thus  $M = 6 \times 15 + 1 = 91$  and  $s_M = 6/91 \approx 0.067$  mm).

The extracted profiles are scored with the distance measure of Equations 6.8 and 6.9, which results in a  $M \times N$  cost matrix  $\hat{\mathbf{D}}$ . This matrix may indicate multiple possible margin candidates per row. Therefore, to ensure that the correct margin points are found and that they are connected, we use a dynamic programming method [80]. The minimum cost path is found by first calculating the minimum cumulative cost matrix as

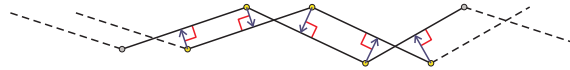
$$\begin{aligned} C_{m,1} &= D_{m,1}, \\ C_{m,n} &= \min(C_{m-k,n-1}, \dots, C_{m+k,n-1}) + D_{m,n}, \end{aligned} \quad (6.10)$$

where  $k$  is an integer specifying the connectedness (the number of rows a path is allowed to travel per subsequent column). For our experiments we used  $k = \lceil \frac{s_L}{s_M} \rceil = 4$  which is approximately the ratio between the spacing of the margin points ( $s_L$ ) and the size of the search steps ( $s_M$ ). Next, the path is traced back along the minimum cost value gradient. Figure 6.6 shows the cost matrix and the traced minimum cost path in the minimum cumulative cost matrix. With the minimum cost path we find  $N$  new margin points, which we convert using the method of Section 6.2.3 to a plausible margin shape that corresponds with the shape model.





**Figure 6.7:** The JSW is measured by determining the point-line distances from the margin points to the closest point on the opposite margin.



**Figure 6.8:** The mean absolute point-line distances are calculated to determine the difference between two differently estimated margins.

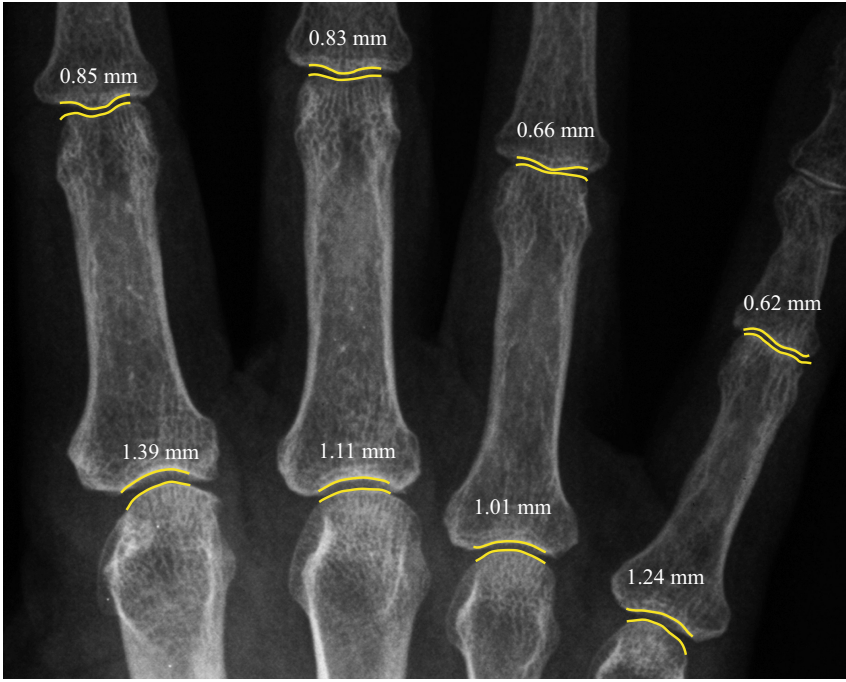
## 6.2.6 Convergence

As the search process operates along a series of lines, that are dependent of the initial margin shape, it is unlikely that the correct margin is found in a single iteration. Furthermore, the margin detection method works best when the intensity profiles are perpendicular to the actual margin. Therefore several iterations are necessary to converge to a solution where the change between detected margins in subsequent iterations is minimal. Due to restrictions in the parametric shape model, is unlikely that an exact match is found between the fitted margin and the minimum cost path. This can result in an endless loop between multiple (similar) solutions. Therefore a maximum is set to the number of iterations. From the tests we found that five iterations are sufficient for reaching convergence.

## 6.2.7 Distance measure

To measure the JSW, we determine the average distance between the margins. Though several proposals have been done by others [35, 39, 54], there is currently no standard method for measuring the average distance between the joint margins. A detailed discussion about this topic can be found in Chapter 3. For our measurements we calculate the average of the point-line distances from all landmark points on the proximal margin to the line segments describing the distal margin and vice versa. Since the points on the margins are equidistant, we can omit their spacing in the calculation of the average. Figure 6.7 demonstrates how the point-line distances are determined.

To measure the difference between two detected margins, we use the mean absolute point-line difference. This is done instead of using the point to point distance, to allow small differences along the direction of the actual margin. An example is shown in Figure 6.8.



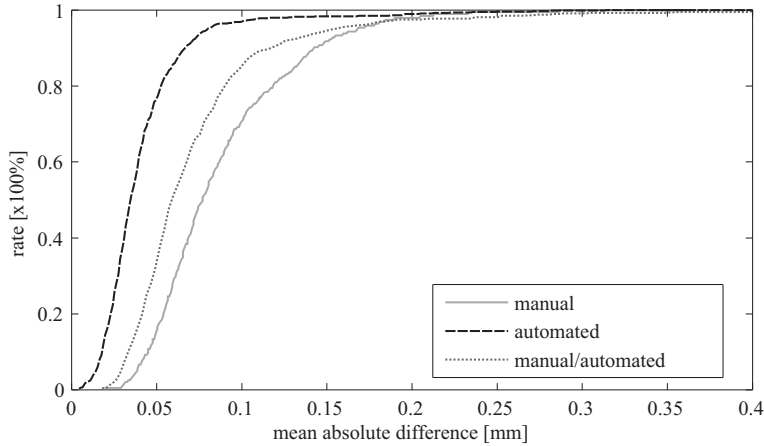
*Figure 6.9: Example showing the results from the margin detection algorithm. The JSW measurements have been printed near the joints.*

## 6.3 Experiments and results

The described method was tested on the 50 images of the test set described in Section 6.2.1. Figure 6.9 shows the results for one of the test images. Almost all margins were found within a short distance of their assumed locations. In one of the images the fifth PIP joint was excluded from the results, since it had been completely deformed and did not have a visible joint space.

### 6.3.1 Margin detection

For all joints we have tested how well the margins are detected compared to manual outlining. Firstly, we have determined the precision of the manual method by letting a trained person indicate the joint margins in the 50 test images. This exercise was repeated a second time by the same person shortly afterwards. As a measure for the intra-observer variability, we calculated the mean absolute point-line differences of the indicated margins. We found that the difference was within 0.14 mm for 90% of the margins. Secondly, we simulated repeated measurements



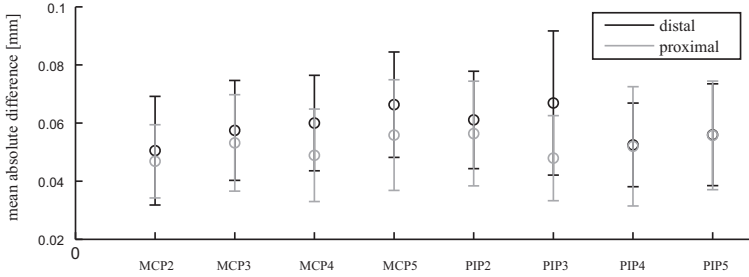
**Figure 6.10:** The precision of the margin detection methods as an estimated probability density function of the mean absolute difference between repeated measurements. The depicted results are for the manual method, the automated method and the mean absolute difference between both methods.

for the automated method by inducing small changes in the initialization of the detection algorithm (Section 6.2.2). The initialization was altered by shifting the initialization points by 0.5 mm in an arbitrary direction. Also for this method, we calculated the difference between the margins. For 90% of the margins this was within 0.071 mm. Thirdly, the difference between the manual and the automated detection method was determined. This was within 0.12 mm for 90% of the margins. Figure 6.10 shows the overall results for these three experiments.

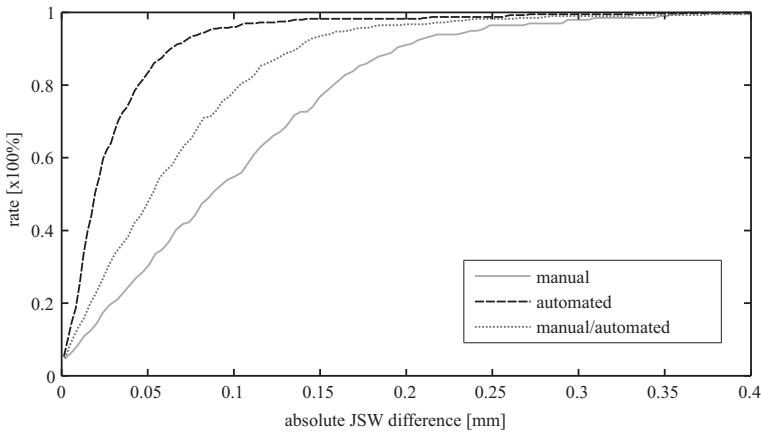
The described results are the averages for all detected joint margins. Minor differences were found between the precisions per joint; see Figure 6.11 showing the precision of the automated method for the distal and proximal margin of each joint. Overall the proximal margin is detected slightly more precisely than the distal margin.

### 6.3.2 JSW measurements

The repeatability of the JSW measurements was determined for both the automated and the manual method. The mean differences were close to zero ( $\sim 0.002$  mm). The absolute differences were within 0.065 mm for the automated method and 0.20 mm for manual readings for 90% of the measured joints. Between the two methods the difference was 0.14 mm. Figure 6.12 shows the overall results. Also for the JSW measurements, there were no considerable differences in precision between the joints, as can be seen in Figure 6.13.



**Figure 6.11:** The median ( $\pm 1$  SD) of the mean absolute point-line difference of the distal and proximal margins per joint for the automated method.

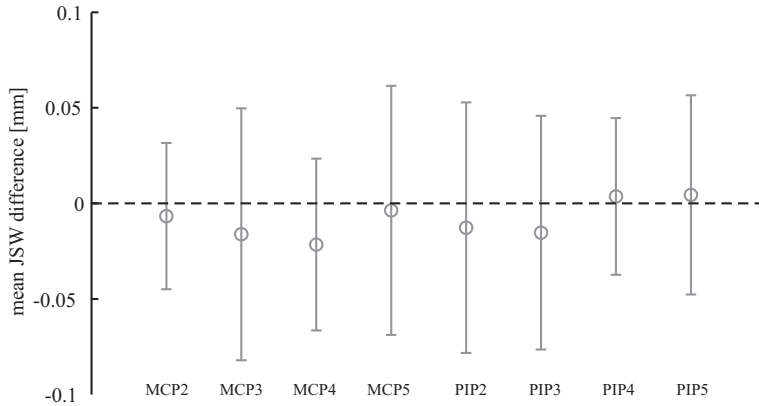


**Figure 6.12:** Repeatability of the JSW measurements as estimated probability density functions of absolute JSW differences between repeated measurements. Results are shown for the manual method, the automated method and the absolute difference between both methods.

## 6.4 OMERACT exercises

Two exercises were performed in collaboration with other research groups taking part in a subcommittee of OMERACT. Comprehensive results of these exercises have been published in [60].

The first exercise involved a radiographic data set of 4 patients, 3 with 2 different time-points and one with 4. A comparison was made between 4 different JSW measurement methods: the method of Angwin [40, 64], Sharp [15], Duryea [81, 37] and ours (Kauffman [93, 98]). The level of automation differed between methods; a summary of these methods can be found in Section 2.4. The radiographs of both



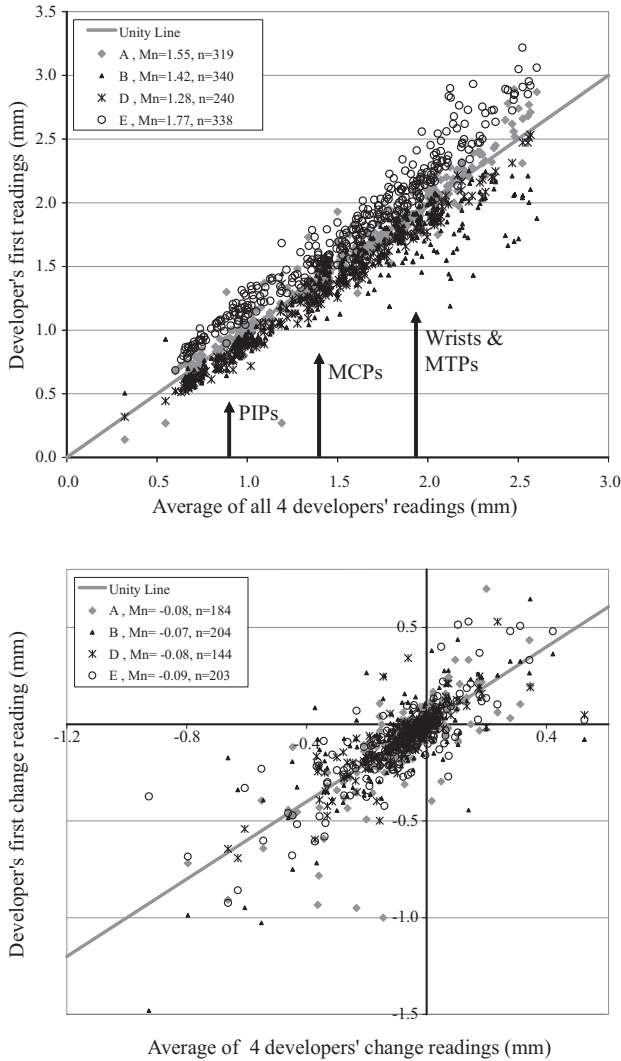
**Figure 6.13:** Mean difference ( $\pm 1$  SD) between automated and manual JSW measurements per joint.

hands and feet were analyzed for each time-point. We measured MCP joints 2–5, PIP joints 2–5 and 4 wrist joints. For the feet all metatarsophalangeal (MTP) joints were measured. To be able to measure the wrist and MTP joints, we extended our margin detection algorithm using the same method as applied for the MCP and PIP joints (Section 6.2). Figure 6.14 shows a comparison between the measurement results of the four different methods. Each point in the graph represents a measurement for a single joint.

For the second exercise a dataset of digitized radiographs was made available by the investigators of the COBRA trial [82]. In this trial two different patient treatment groups had been evaluated using the Sharp/van der Heijde score (SHS) which is often regarded as a gold standard. A selection was made of 107 patients and 428 time-points: baseline, 6 months, 1 year and 18 months. To obtain a single JSW score for one time-point, all measurements were averaged. For our method we measured MCPs 2–5, PIPs 2–5 and MTPs 1–5. The results showed that manual scoring by the SHS (joint space narrowing + erosion score) outperformed all automated methods in discriminating between treatment groups. However, with regard to the joint space narrowing component of the SHS, the automated methods outperformed the manual readings.

## 6.5 Conclusions

We have described and tested a method for detecting joint margins within a predefined region of interest. This method is robust and accurate in cases where the joint margin is clearly visible. For cases where the joint space has disappeared due to severe joint damage, the algorithm is unable to estimate the margins correctly.



**Figure 6.14:** Results from [60], showing a comparison between the measurements of methods A (Sharp), B (Kauffman), D (Duryea) and E (Angwin). The top graph shows the absolute measurements, the bottom graph shows the measured JSW change over time. The graph legends also report the mean ( $Mn$ ) and number of performed measurements ( $n$ ) for each method. Remark: The MTP measurements were not performed with the method D. As methods A and E show systematically higher readings than methods B and D, this may have affected the locations of the MTP readings with respect to the other measurements.

Such cases may be detected by measuring the difference between the found margins in the last two iterations of the search algorithm (a large difference indicates a poor convergence of the detection algorithm). From this stage on it would be necessary to use a different method to quantify joint damage.

In Section 6.3.1 we determined that the precision of the automated detection method is higher than the precision of the manual method. Also we have found that the differences between the results of both methods were smaller than the precision of manual readings. This suggests that the automated method is more consistent in detecting the margins. Considering this, it can be beneficial to recalculate the means and eigenvector matrices of the intensity profiles using the newly, more consistently, detected margins (Section 6.2.4).

Figures 6.11 and 6.13 indicate that the absolute precision for each joint is similar. However, if we consider relative precision, the JSW measurements of the PIP joints are less precise than those of the MCPs. This is because the JSW of a PIP joint is on average 30% smaller than that of an MCP joint.

The constraints set by the shape models prevent the detection of false edges, but also limit the possibility of detecting unusual margin shapes and small erosions. Therefore several margins were poorly detected. Fortunately, this does not always have to affect the JSW measurement, since this is determined by averaging and therefore small aberrations may be leveled out. This effect is noticeable in the results, as the precision of the JSW measurements is better than that of the detected margins.

We took part in two exercises to compare our method with other methods. The first exercise showed some systematic differences between JSW measurements by automated methods. However, change measured in serial radiographs showed good agreement, with no observable systematic differences [60]. The results of the second exercise show that automated JSW measurements can outperform manual JSW measurements by the SHS. However, with erosion scoring included, the SHS performs better than automated JSW measurements.





# 7

## Acquisition variability and JSW measurements

### 7.1 Introduction

The radiographic representation of an object is for a major part determined by its orientation with respect to the x-ray device. Therefore, when using projection radiographs for medical analysis and measurements, one has to pay careful attention on how a body part is positioned. With repeated radiographic acquisitions, especially when there is a time lapse, it is likely that differences occur in the positioning of the subject with regard to the x-ray source and the detector. Consequently, if we investigate a series of consecutive radiographs, we should not only consider the differences caused by disease activity, but also differences caused by changes in positioning during the acquisition of the radiograph. With regard to this latter aspect we can differentiate between two sources of variability. The first is caused by changes within the x-ray setup, for example the location of the x-ray source with respect to the detector. The second is caused by the positioning of the subject, such as differences in orientation and articulation of the joints.

In practice, at least according to our experiences with hand radiography, acquisition settings with regard to positioning are seldom recorded in detail and it is possible that radiographers deviate from the applicable standard acquisition protocol (if present). To determine the influence of such deviations, we analyze how variations in the acquisition setup can affect the outcome of JSW measurements. In the next section (Section 7.2) we will study the properties of a common acquisi-

tion setup and identify the variations that may occur. In the following Section 7.3 we will use simulated projection images to determine how these variations can affect JSW measurements.

Research on the effects of hand positioning variability has been done by Angwin et al. in order to determine the precision of radiographic JSW measurements [40]. They investigated the effect of small positioning changes and found that with their method JSW changes of more than 7% can be detected in individual joints. Averaging the results across fingers for a single subject decreases the detectable change to 3%. Their results show that careful positioning is essential to obtain precise measurements. Because of this, and the results of our study, we conclude that it makes sense to develop a method to standardize hand positioning in order to increase the precision of JSW measurements. In Section 7.5 we present a method to realize this by introducing a positioning aid for hand radiography.

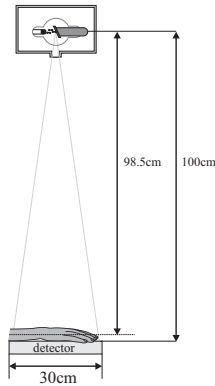
## 7.2 Analysis of the acquisition setup

To investigate the effects resulting from variations in a typical radiographic acquisition setup, we first analyze a generic model of an x-ray source and detector. In the case of hand radiography for the purpose of JSW measurements and joint damage assessment, the hand is generally placed with the palmar side down, flat on the detector. The x-ray source is positioned roughly one meter from the detector surface above the center of the hand. In this way the entire hand including the wrist can be captured in a single image using a rectangular diaphragm which is preferably kept as small as possible to prevent unnecessary exposure to the radiation. When an acquisition is made from both hands at the same time, the center point is generally chosen between both hands. As the size of an x-ray source's focal spot is relatively small (typically less than 1 mm [83]) with regard to the distance between the source and target/detector (typically 1 m), we consider that the x-rays originate from a point source, which leads to a perspective projection.

**Source-detector distance** Since the hand is placed between the source and the detector, a minor magnification of the projected image occurs due to the diverging bundle of x-rays. In this situation the magnification can be calculated as follows:

$$M = \frac{d_{td}}{d_{sd} - d_{td}} \times 100\%, \quad (7.1)$$

with  $M$  the percentage of magnification,  $d_{sd}$  the distance between the source and the detector, and  $d_{td}$  the distance between the target and the detector. Since the target is three dimensional, this latter distance depends on the projected location within the target's volume. To simplify our model, we therefore consider the center of the volume. If the hand is placed flat on the detector, the mean location of the MCP and PIP joints will be at approximately 1.5 cm distance from the detector's



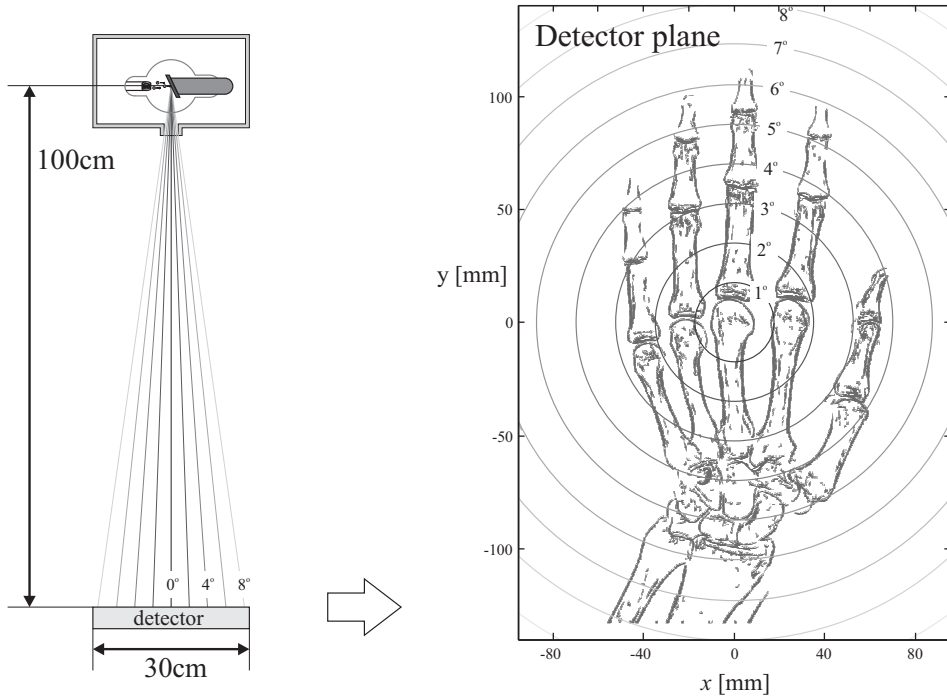
*Figure 7.1: Schematic model of the acquisition setup.*

surface. Consequently, with a source–detector distance of 100 cm (and a target–detector distance of 1.50 cm), the magnification is 1.52%. Figure 7.1 depicts the model of the setup.

In practice, radiographic measurements are rarely corrected for this type of magnification, as this effect occurs with each recording and is expected to remain constant over time when the same setup is used. Since the source–detector distance is large compared to the object–detector distance, small deviations in the source–detector distance have no significant effect on the magnification. For example, if the source–detector distance is set to 90 cm instead of 100 cm, the magnification becomes 1.69%. The difference in magnification, 0.17%, results in deviations of about 2.6 micron (for typical JSWs of 1.5 mm), which is small compared to the pixel spacing (42 micron for a typical scanning resolution of 600 dpi).

The distance between the hand and the detector may also vary between acquisitions. This can be a result of differences in positioning, swelling of joints or modifications of the detector’s cassette. Such differences are generally in the order of several millimeters, and have no significant effect on the magnification. For example, if the joints are located at 2.0 cm above the detector surface instead of 1.5 cm, the magnification increases by 2.04%. As explained in the previous example, such a magnification increase has no significant effect on JSW measurements for commonly used image resolutions.

**Projection angles** By the term *projection angle* we refer to the angle of the path of an x-ray with regard to the orientation of the exposed object. Obviously, rotation of the object with regard to the source and detector can affect the projection angle and therefore the depicted projection. In addition, since the x-rays originate from a point source their angles vary throughout the exposed area. Figure 7.2 illustrates for a typical x-ray setup, how the angles of the incoming x-rays vary in the entrance plane of the detector. This illustration shows that not only



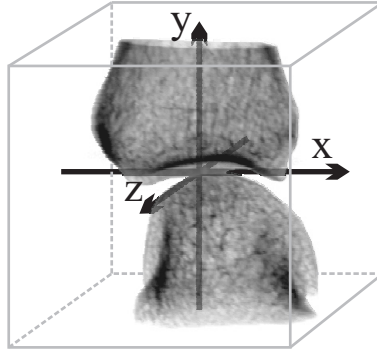
*Figure 7.2:* The left illustration shows an x-ray point source and the path of the diverging rays to the detector. The right illustration shows the angles (in degrees) of the x-rays with the normal of the detector's surface.

rotational variability in the positioning of the subject affects the projection angle, but also translations within the detector plane. For example, a shift of 1 cm from the center of the exposed area, the projection angle changes with  $\arctan\left(\frac{1}{100}\right) \approx 0.6$  degrees.

How differences of the projection angle affect the radiographic projections of a finger joint is difficult to predict. In the next section we investigate this aspect by analyzing simulated projection images of three dimensional volume data of several MCP joints.

### 7.3 Simulated projection images

By rendering perspective projection images from three dimensional CT data we can simulate how changes of the projection angle affect the representation of a joint space and its width.



*Figure 7.3: Axis definition with regard to the 3D volume data.*

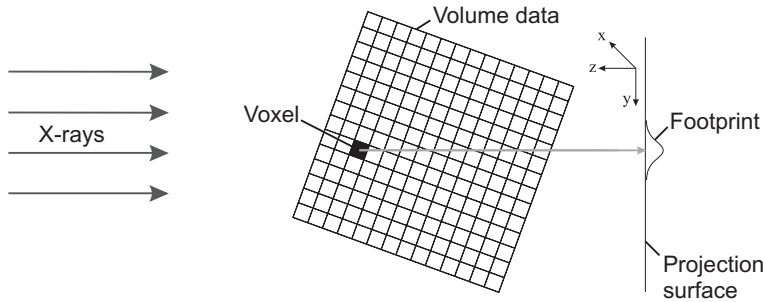
### 7.3.1 Method

To investigate the influence of the projection angle on the visualization of a joint space, we use four sets of volume data of single MCP-joints. Each set exist of a collection of *voxels* which are defined in a 3D grid. These voxels are parallel to the pixels in a 2D image. The value of a voxel corresponds to the Hounsfield unit [84], which is a linear transformation of the linear x-ray attenuation coefficient. As the required resolution for this exercise is not within reach of common CT scanners,  $\mu$ CT-scans have been used. These 3D scans have a voxel size of  $66 \mu\text{m}$ , which is similar to the pixel size of a digitized radiograph. These  $\mu$ CT-scans originate from a study by Duryea et al. [85] and were kindly provided to us.

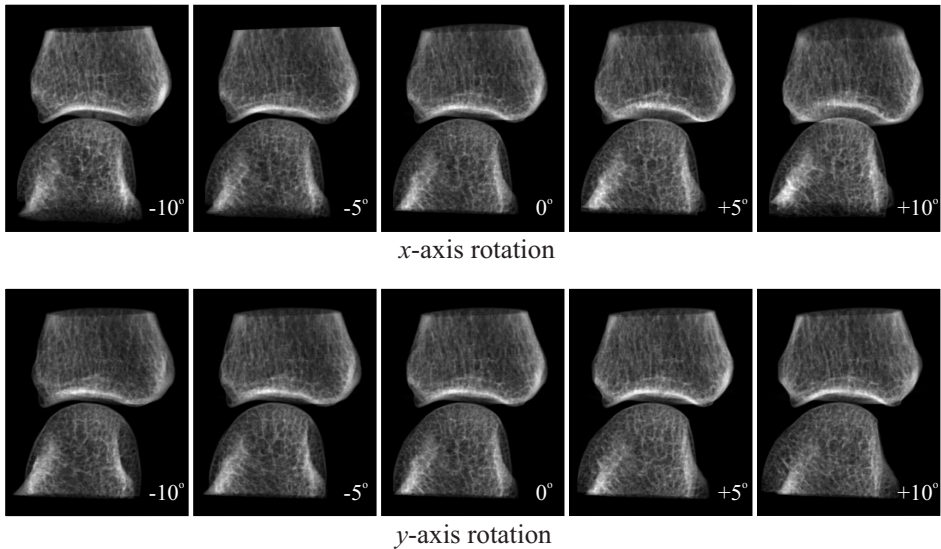
Before rendering projection images, we define the spatial axes within the volume according to the example in Figure 7.3. For 3D structures there are six independent degrees of freedom in 3D space; three translations and three rotations corresponding to the axis of each dimension. To simplify the projection model, we only consider the orientation of the joint and presume that the detector is parallel to the  $x$ - $y$  plane with the source location at 1 m distance on the  $z$  axis. The center of the volume containing the joint is located on the  $z$  axis at 1.5 cm above the detector plane.

As we are interested in orientation differences that may cause a change in the joint's appearance, we limit this analysis to rotations around the  $x$  and  $y$  axes. Rotations around the  $z$  axis occur perpendicular to the plane of the projection, and are therefore considered irrelevant. The same is the case for translations in the  $x$ - $y$  plane. Translations in the  $z$  direction cause a magnification of the projection as discussed in the previous section.

To generate a projection image from the volume data, we make use of a rendering method called *splatting* [86]. For this method each voxel is 'splatted' onto a virtual projection plane according to the path of the passing x-ray's. This process leaves a footprint for each voxel depending on the voxel's value and projection path. The sum of all footprints results in a simulated projection image. The



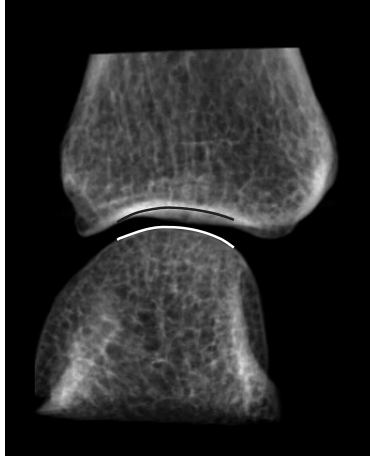
*Figure 7.4:* Illustration of the splatting volume rendering method. Each voxel is ‘splat-  
ted’ onto the projection surface leaving a footprint according to the direction of the x-rays.  
The combination of footprints forms the projection image.



*Figure 7.5:* Generated projection images of an MCP joint originating from micro-CT data. The upper and lower series show projections for varying rotations over the x-axis and y-axis respectively.

working of this method is illustrated in Figure 7.4.

For each set of volume data we have generated projection images for rotations between  $-10$  and  $10$  degrees around the  $x$  and  $y$  axis in steps of 1 degree. Figure 7.5 shows an example series of projections of one of the  $\mu$ CTs. The JSW is detected and measured in the generated projection images using the margin detection method described in Chapter 6. As only a single joint is visible in each image and not the entire hand, we have initialized the margin detection algorithm



*Figure 7.6: Detected margins in a generated projection image.*

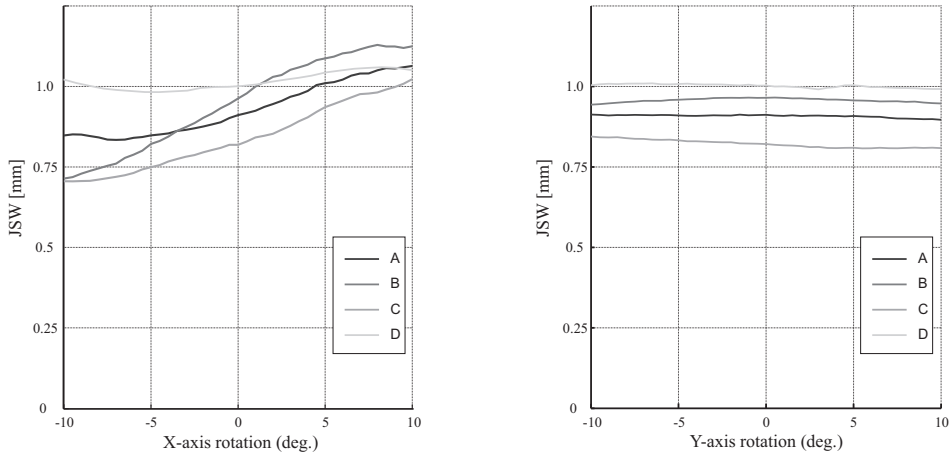
manually by indicating the center of the joint margin. For each dataset this was done once for the projection image with zero degree rotation. The initializations of the subsequent rotated projections have been derived successively from one another by the previously detected margins. Figure 7.6 shows the location of the detected margins for one of the projections.

### 7.3.2 Results

The JSW measurement results are displayed in the graphs of Figure 7.7. These results show that for these joints, there are minor JSW variations for  $y$  axis rotations. For the  $x$  axis the difference in JSW between rotations of  $-5$  and  $+5$  degrees can be up to  $0.5$  mm (volume data set B), resulting in a difference of  $50$   $\mu\text{m}$  per degree rotation.

## 7.4 Conclusion

We have noticed that positioning and system setup are rarely recorded and that it is possible to deviate from a standard acquisition protocol. As a consequence the analysis of subsequent radiographs becomes more complicated because of variations in positioning and image appearance. Besides this aspect, we also found that measurement precision can be affected by positioning and acquisition setup variability. In the analysis of Section 7.2 we demonstrate that the projection angle varies over the detector surface, meaning that deviations of several degrees can occur for different positions of the hand.



**Figure 7.7:** Results of the JSW measurements with varying projection angles for four different volume data sets (A, B, C and D).

The results of Section 7.3.2 show that particularly  $x$  axis rotations of the projection angle can affect the visualized joint and its JSW. The measurement outcomes of Figure 7.7 indicate that the JSW can vary with  $50 \mu\text{m}$  per degree rotation. As this variation differs for each joint, it is impossible to develop a generic method to compensate measurements based on a known projection angle.

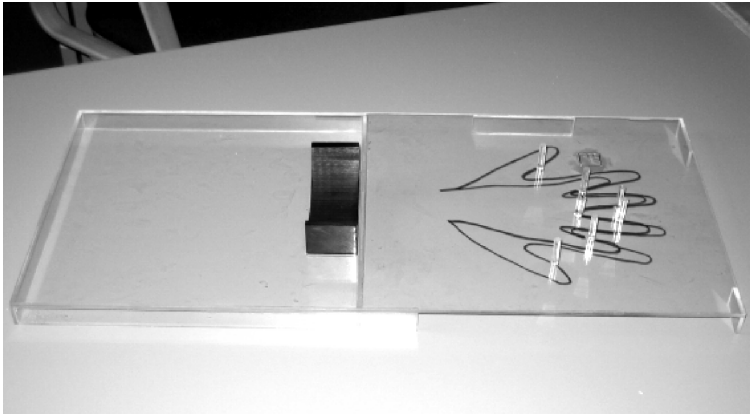
Other studies [39, 40] have shown that a precision higher than  $0.1 \text{ mm}$  can be achieved for JSW measurements. Our analysis indicates that several degrees of deviation in the projection angle can possibly result in a few tenths of a millimeter deviation in the measured JSW. Considering this, we infer that the outcome of JSW measurements can be significantly affected by differences in positioning and acquisition methods.

## 7.5 Recommendation: a positioning aid

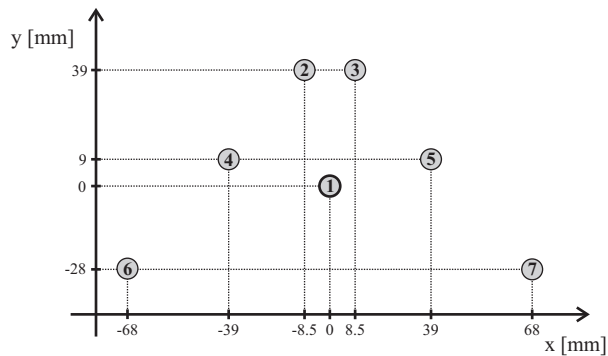
According to the conclusions of the previous section, the variability in hand positioning and acquisition setup can affect the measurement precision of the JSW. This is not the first time that we notice a negative contribution of positioning variability; also for our segmentation method (Chapter 4) we find that positioning variability complicates the detection of bones. Considering these two issues, it would be practical to find a way to minimize this variability.

In general, positioning variability is caused by the absence of well defined and observed positioning guidelines, and not by limitations of a patient's freedom of movement (although this can be a factor in cases of severe joint damage). Obviously, it would be desirable to develop a method to standardize the way a hand is positioned during radiographic acquisitions. In order to realize this, a custom





*Figure 7.8: Photo of the positioning aid.*

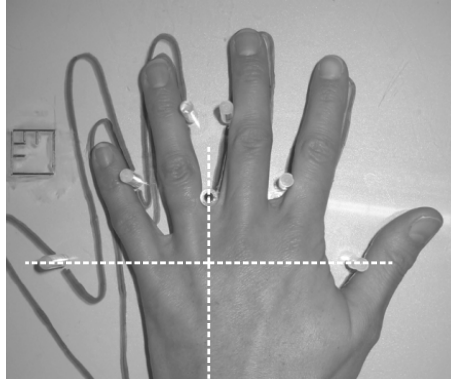


*Figure 7.9: Peg locations on the HPA.*

hand positioning aid (HPA) has been developed.

### First prototype

Figure 7.8 shows a photo of the first prototype HPA. The idea for this design has been inspired by a system that is used to position hands for a hand geometry-based verification system [74]. The development of the HPA is done in collaboration with the departments of rheumatology, clinical physics and radiology of the local hospital, Ziekenhuisgroep Twente (ZGT), in Hengelo. The HPA consists of a perspex plate with nine pegs at fixed locations. The locations of the pegs have been chosen symmetrically such that the HPA can be used for both left and right hands. Figure 7.10 shows the location of the pegs and demonstrates how a hand is positioned. The hand is moved up until it touches the center peg between the middle and ring finger. When these fingers are adducted such that they touch



*Figure 7.10: Placement of the hand on the HPA. The dashed cross indicates where the x-ray beam should be centered.*

pegs two and three, metacarpals 2–5 will become fixed to a certain extent by their limited mutual lateral mobility. The gentle abduction of the remaining fingers up to the other pegs results in a stable positioning of the rest of the hand. To ensure a straight and stable positioning of the wrist, the HPA is also provided with a forearm support.

The bottom plate of the HPA is constructed such that it fits tightly over the x-ray semiconductor detector (Canon type CXDI 31 BG7-2350) that is used in the local hospital (see also Section 1.3). In this manner the HPA is always positioned correctly with regard to the detector. The thickness of the perspex is 4 mm, which causes some attenuation of the radiation. As the attenuation is homogeneous over the entire surface of the detector, this is not visible in the radiograph after overall contrast adjustment. Figure 7.11 shows the appearance of a radiograph taken with the HPA.

An additional advantage of using the HPA is that it can be equipped with objects for calibration and quality assessment. The first prototype HPA has been equipped with an aluminum step wedge and a special peg that can be used to determine the projection angle. The applied aluminum step wedge varies in thickness in 5 steps (0.7, 1.4, 2.1, 2.8 and 3.5 mm). In combination with the x-ray device settings such as (wavelength and exposure settings), this is a reliable indicator for image contrast adjustment and calibration. Figure 7.12 shows a drawing of the special peg that is used to determine the angle of the incoming x-rays. The peg has been made of perspex which is highly translucent for x-rays. On top of this peg a small steel ball has been attached and underneath is a steel ring. When the incoming x-rays are parallel to the peg, the projection image will show the location of the ball in the center of the ring as depicted in Figure 7.12. Any angle deviation will make the ball move out of the center of the ring. In practice it is not necessary that the radiograph is taken such that the ball is perfectly in the ring's



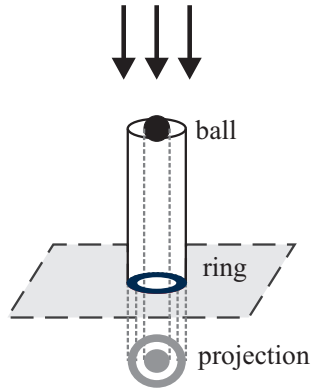
*Figure 7.11: Radiograph with use of positioning aid.*

center. More important is that the location of the ball with regard to the ring is the same for each radiograph. This ensures that the projection angle is constant over time with respect to the position of the hand.

### **Second prototype**

During three months the first prototype HPA has been tested for practical use at the ZGT hospital. The radiodiagnostic assistants evaluated the HPA as a helpful instrument and guide for positioning the hands of patients. The HPA proved to be easy to use for the patient and assistant. Also the radiographic quality was evaluated positive by radiologists and rheumatologists. Unfortunately, the construction of the HPA turned out to be too weak for daily use. The bottom frame was damaged and several pegs were missing. Also the contours indicating the hand positions had been partly wiped off after cleaning with alcohol.

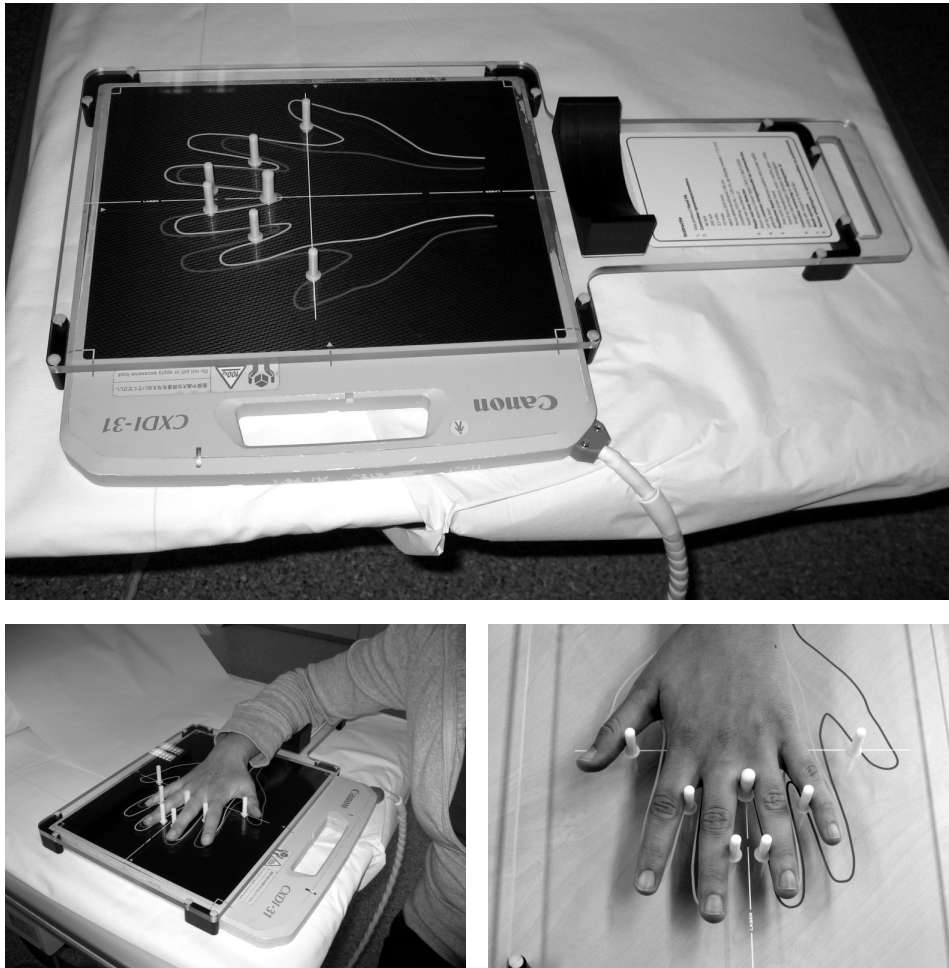
For the second prototype, named the ZGT FingerFIX (Figure 7.13), several



*Figure 7.12: Angle detection pin.*

improvements have been made. The construction of the aid is stronger and more durable. The pegs have been attached by screw connections (instead of glue), and can now be replaced by spares. Furthermore the printed hand contours are now resistant against alcoholic solutions, and also a center cross has been added to assist with focusing the x-ray beam. For completeness, the FingerFix provides a short instruction list describing the acquisition protocol.

The ZGT FingerFix will be part of new studies that will be conducted in cooperation with several other hospitals to collect new data sets. The goal is to validate that the use of this aid improves image quality resulting in robust automated analysis and more measurements.



*Figure 7.13: Images of the ZGT FingerFIX.*



# 8

## Revealing radiographic changes

### 8.1 Introduction

As presented in the overview in Chapter 2, several scoring methods have been proposed to quantify joint damage in hand radiographs [15]. Some make use of classification scores for joint erosions and deformations, for example the Larsen score [16], and the Sharp/van der Heijde method [1]. Other methods are based on relative or absolute measurements, for example determining the carpal/metacarpal ratio [19], joint space width measurement and erosion volume estimation.

With the abovementioned methods disease activity is determined *indirectly* by looking at changes in scores, rather than detecting differences *directly* by comparing the successive radiographic images. However, when two follow-up radiographs are displayed side by side, it is easy to overlook small erosions or differences in bone density. A comparative analysis becomes even more difficult when images have been acquired through different devices or with altered settings. Even a change in the contrast and brightness settings may affect a reader's opinion.

As nowadays more and more hospitals are working with digitized radiographs, new possibilities for radiographic analysis are becoming available. Radiographic analysis on a computer screen has several advantages; for example, regions of interest can easily be magnified, the contrast and brightness settings can be changed and special filters can be applied to enhance certain image features. In this chapter we show how some of these new possibilities can be applied to make a direct comparison between radiographic images.

The method that we propose is based on image subtraction in order to reveal

changes in bone structures between follow-up hand radiographs. This is accomplished by aligning a region of interest within both images and calculating the difference between the pixel intensities. In the field of image processing, this alignment procedure is referred to as ‘image registration’ [87]. The large variability in hand positioning makes it difficult to register two images entirely. Overlapping bone parts near the joints and differences in projection angle cause interfering artifacts when stringent elastic transformations are applied to one of the images. Therefore the analysis is restricted to rigid regions of interest, such as the individual bones.

The images to be compared have been taken at different time instances, typically with several years in between. It is not uncommon with such data that the radiographs have been acquired differently: the hospital’s equipment may have been renewed or its settings changed, the patient might have gone to another hospital or the acquisition protocol may have been altered. This makes it difficult to compare radiographs directly, since illumination settings such as contrast and brightness may be different. We compensate for such differences by determining an intensity transformation function based on the joint histogram of the relevant images. Finally, the difference is determined through image subtraction and displayed to the operator by means of a color overlay in the radiograph that is examined.

## 8.2 Subtraction of radiographs

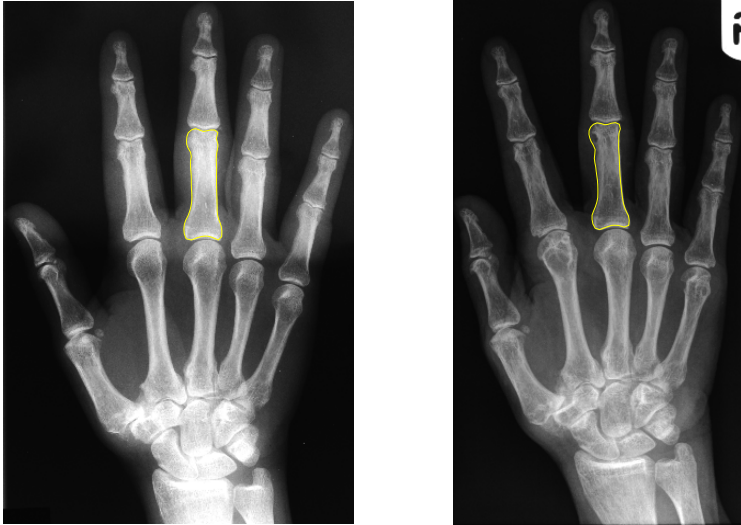
To illustrate our method we take two follow-up hand radiographs that have been made with several years in between. The third proximal phalanx is selected for this example, since it shows a clearly visible erosion at its distal end in the second radiograph. For the selection of the bone, we apply the algorithm described in Chapter 4 to detect the approximate bone outline in both images. Figure 8.1 shows the detected outline of the selected bone in both images.

Next, we extract a rectangular region of interest (ROI) that fits the bones with 3 mm of extra space around the outlines. A rigid transformation is used to warp these ROIs to two new images  $A$  and  $B$ , as shown in Figure 8.2. Both images are now roughly aligned. Due to small deviations in the detected contours, a small error may still be present. To further improve alignment, we apply an image registration method.

### 8.2.1 Image registration

A medical image registration algorithm [88] is used to register the first image to the second. Although bones are mostly rigid, we allow subtle elastic transformations in the registration algorithm. This is necessary to correct for small, but smooth, contour variations caused by differences between the projection angles during acquisitions. By smoothing the deformation field we prevent the registra-





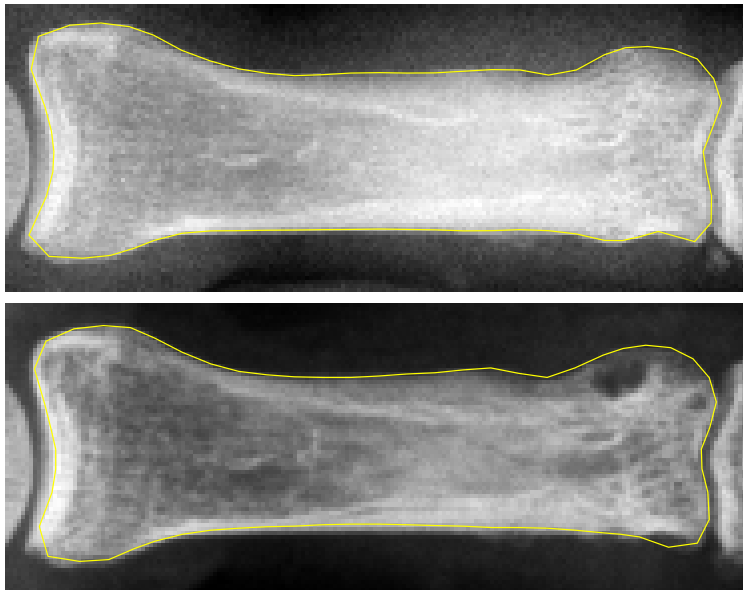
*Figure 8.1: Two radiographs of the same patient; the left is the baseline image, the right has been made several years later. The detected outline of the third proximal phalanx is shown in both images.*

tion algorithm from applying strong local deformations that might conceal erosions. Figure 8.3 shows the effect of the calculated deformation field on a mesh. The result of the registration is displayed in the checkerboard image of Figure 8.4.

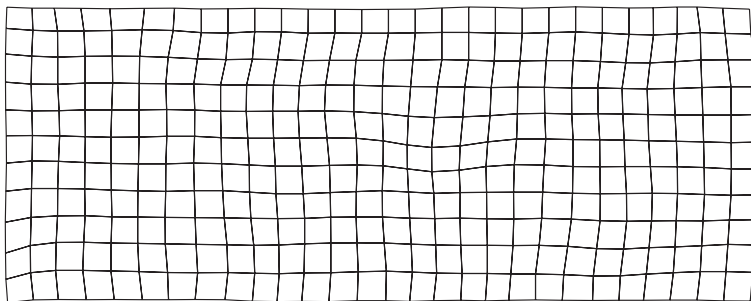
## 8.2.2 Intensity transformation function

After registration of the images, we want to compare the image intensities through subtraction. Since we do not have any information on how the images have been acquired, the relation between their intensity values is unknown. If there is only a difference in contrast and brightness, this relation would be linear. However, it is likely that this relation is non-linear, as a result of differences in device setup. For example, a different kV-setting would result in an exponential relation (see Section 1.3).

The goal is to find an intensity transformation function that changes the pixel values of image  $B$  such that we can compare it to the intensities of *reference image*  $A$ . Since both images have been aligned, we can make use of the joint distribution  $f_{A,B}(a,b)$ , where  $a$  and  $b$  are values of corresponding pixels in image  $A$  and  $B$ . Figure 8.5 shows the joint distribution of the example images in a gray-scale image. The bit depth of the used images is 8-bit, therefore the axes have been divided in 256 equal size bins. The higher the displayed intensity in the distribution, the higher the occurrence.

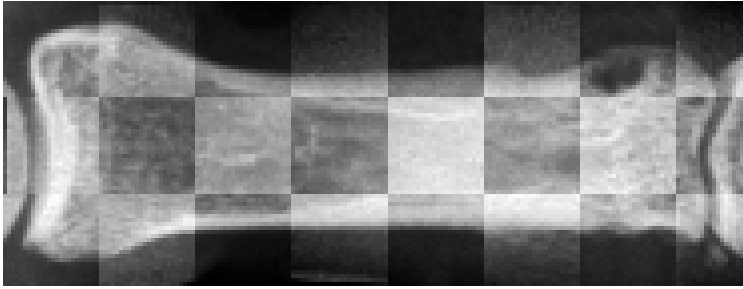


*Figure 8.2:* The upper image shows reference image  $A$ , the baseline image, and the lower shows  $B$  made several years later.

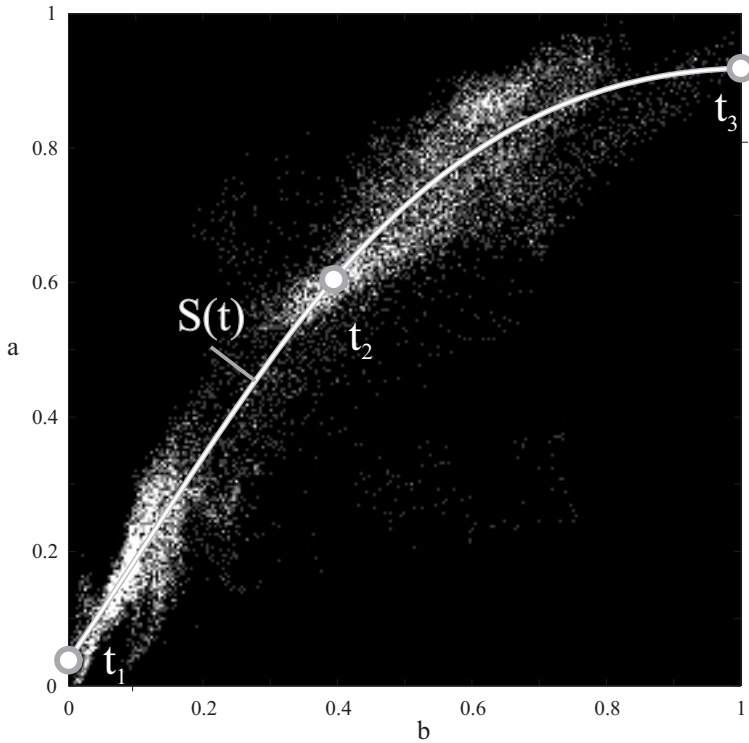


*Figure 8.3:* The deformation field applied to a mesh.

We estimate the intensity transfer function  $S(t)$  by fitting a piecewise polynomial form of a cubic spline interpolant [89] to the data of the joint distribution. Interpolation is done between three points defined by the knot vector  $\mathbf{t}$  with the



*Figure 8.4: Checkerboard view of registered source and target images.*



*Figure 8.5: The joint distribution in gray intensities and the fitted transformation function  $S(t)$  displayed by the curve. The circles are the locations of the knots in vector  $\mathbf{t}$ .*

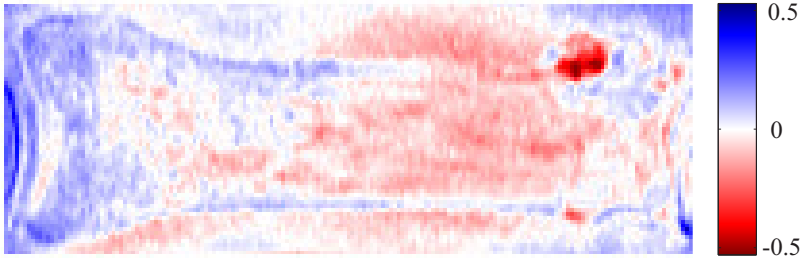
following conditions:

$$t_1 = 0 \quad (8.1)$$

$$0 < t_2 < 1 \quad (8.2)$$

$$t_3 = 1 \quad (8.3)$$

$$S(t_1) < S(t_2) < S(t_3) \quad (8.4)$$



**Figure 8.6:** Difference image normalized by the average intensity of the bone in image  $A$

The begin and end conditions  $S'(t_1)$  and  $S'(t_3)$  (where  $S'(t)$  is the first derivative of  $S(t)$ ) are constrained to positive values, such that the curve is increasing for  $t = [0, 1]$ . The function is fitted to the joint distribution in a least squares manner using a reflective Newton method [90]. The fitted spline of the example is also displayed in Figure 8.5.

Next, we apply the intensity transformation to image  $B$  and calculate difference image  $D$  by subtracting image  $A$ :

$$D = S(B) - A \quad (8.5)$$

Since we do not have any information which we can use to quantify the image intensities, we normalize the difference image by dividing it by the average intensity of the bone in image  $A$ . We note that this average may be inaccurate, since the intensity scale is likely to be slightly nonlinear. The resulting difference image is shown in Figure 8.6. A color mapping is used such that a bone density decrease is colored red and a bone density increase blue.

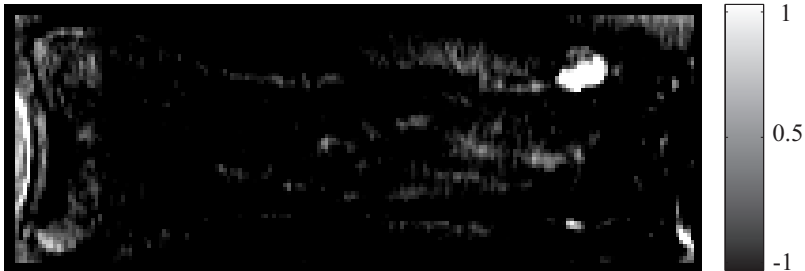
In this example the erosion is clearly visible, but in other cases this representation may still be difficult to interpret: the reader is easily distracted by small deviations caused by noise and differences in illumination. To enhance the readability of the results we add an alpha channel (transparency map) to the colored difference image and use image  $B$  as background [91]. The alpha channel image  $\alpha_D$  is constructed from the difference image  $D$  by taking the absolute value and mapping the values between a lower and upper threshold ( $\tau_1$  and  $\tau_2$ ) to values between 0 and 1:

$$\alpha_D = \begin{cases} 0 & \text{if } |D| \leq \tau_1 \\ \frac{1}{\tau_2 - \tau_1} (|D| - \tau_1) & \text{if } \tau_1 < |D| < \tau_2 \\ 1 & \text{if } |D| \geq \tau_2 \end{cases} \quad (8.6)$$

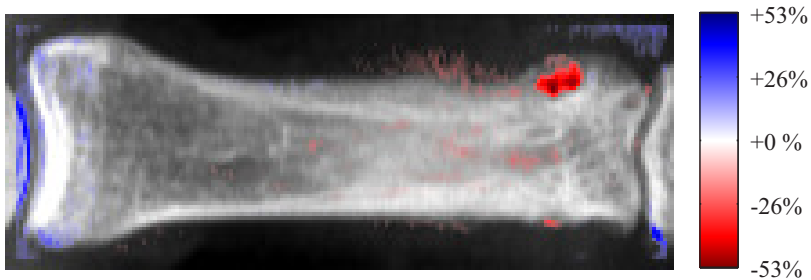
The thresholds  $\tau_1$  and  $\tau_2$  are free to choose and can be set to any value desired by the operator.

Figure 8.7 displays the alpha channel for described example. This alpha channel is used to mix  $A$  and  $D$  by means of the following equation:

$$M = A(1 - \alpha_D) + \alpha_D D \quad (8.7)$$



*Figure 8.7: Alpha channel  $\alpha_D$ .*



*Figure 8.8: Mixed image: image A and D with applied alpha channel. The difference is displayed as the percentage of the average intensity of the bone at baseline. Applied thresholds for alpha channel:  $\tau_1 = 0.1$  and  $\tau_2 = 0.3$ .*

The resulting mixed image  $M$  of our example is displayed in Figure 8.8.

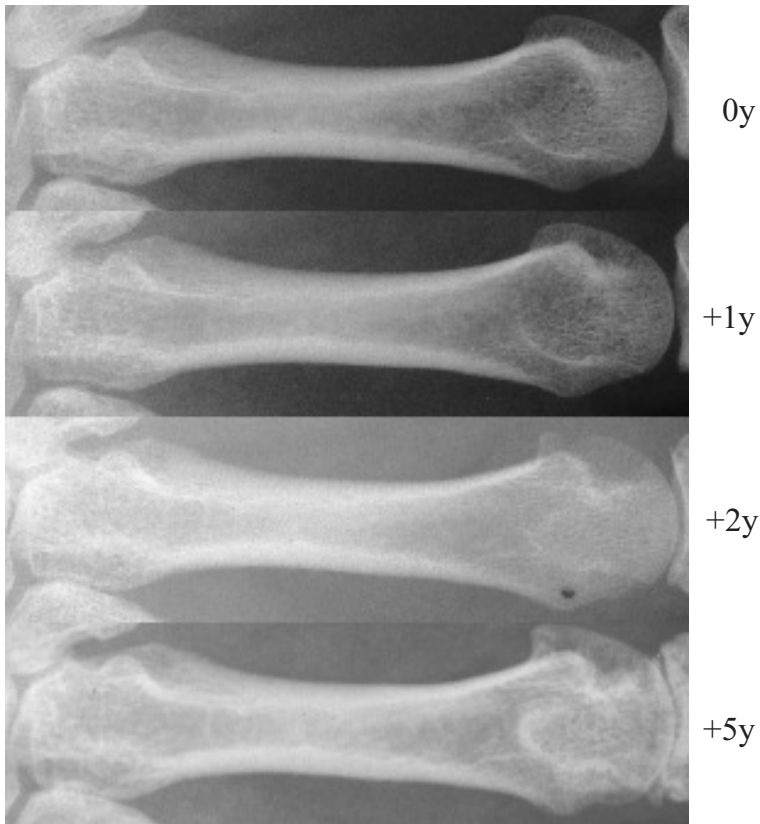
As can be seen in the result, there are some strong differences visible at the adjacent bones. Optionally, we can remove these regions by applying a mask using the initially detected contours. But this also means that possible changes near the edges of bone become invisible.

One may also note the thin blue edge at the metacarpal head. It is likely that this is the result of joint space narrowing, which indicates the loss of cartilage in the joint.

## 8.3 Results

Experiments with several follow-up radiographs show that erosions and changes in bone mineral density can be visualized. Figure 8.9 shows four follow-up radiographs of a third metacarpal bone. The top image is the baseline image and is used as the reference image. The following three images have been made after 1, 2 and 5 years. The pixel size is 0.25 mm for all images.

The described method was applied as follows. Firstly, we register the baseline image to the other three images. Secondly, we correct the intensities of the latter



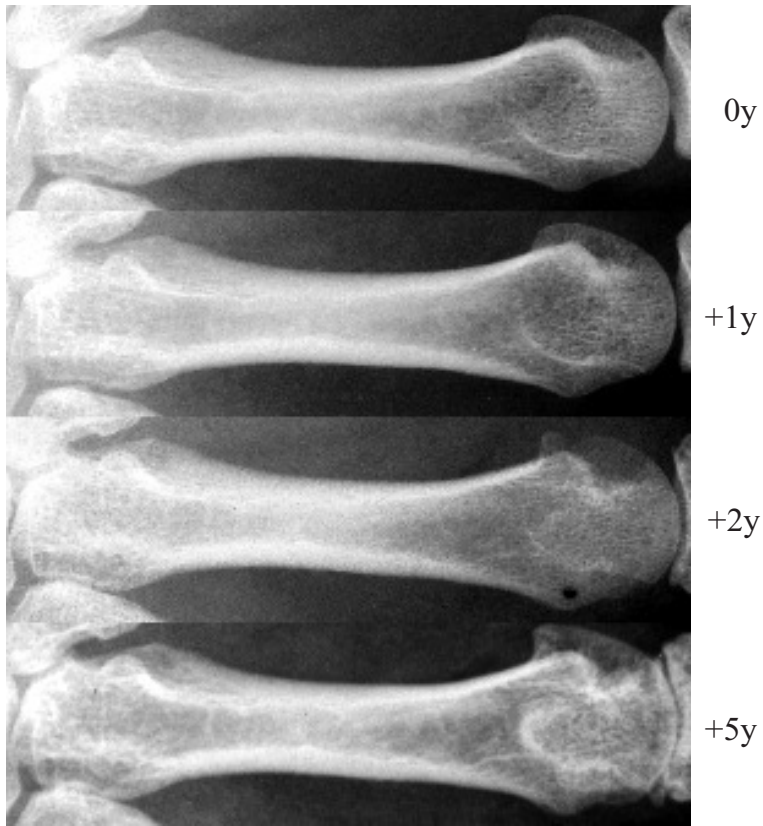
*Figure 8.9: Original series of follow-up images of a third metacarpal bone. The baseline image is at the top, followed by images made after 1 year, 2 years and 5 years.*

three images to match the baseline image; see Figure 8.10. Finally, for each image we calculated the difference with the baseline image. Figure 8.11 shows the differences as overlays. The alpha channel thresholds are:  $\tau_1 = 0.1$  and  $\tau_2 = 0.4$ .

## 8.4 Discussion

Our experiments show that image subtraction techniques can be used to visualize local changes in bone density. Such changes are important indicators of disease activity and may reflect the effects of treatments. The presented method of displaying intensity differences by means of a colored overlay offers a useful aid to the reader during radiographic analysis.

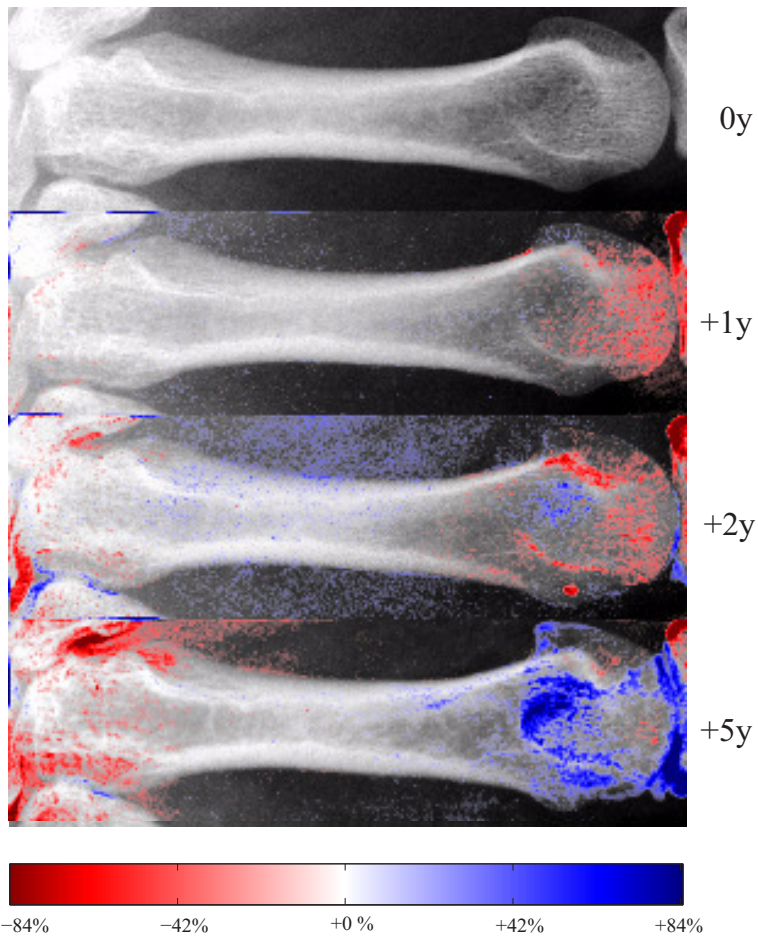
Within the current method the joint distribution is used to estimate the intensity transformation function for matching the illumination settings between both



**Figure 8.10:** The same images as in Figure 8.9, but with applied intensity transformation to match the intensities of the baseline image.

images. With this method it is difficult to detect uniform changes in bone density. Also, large changes between consecutive images may result in a poor estimation of the intensity transformation function. Ideally a calibration object should be placed in the image area during image acquisition. Matching the illumination properties of the calibration object instead of the bones may solve this problem. In future research we plan to use an aluminum wedge for this purpose.

Currently, intensity differences have been quantified with a relative measure: as a percentage of the average bone intensity at baseline. For absolute measurements there should be an accurate bone density measurement for at least one of the time instances. For example, this could be done with a DEXA scan.



*Figure 8.11: Images as in Figure 8.10 with difference overlay. The difference is in percentage of the average intensity of the bone at baseline.*



# 9

## Conclusions and recommendations

### 9.1 Conclusions

This thesis describes various aspects with regard to joint damage assessment in hand radiographs. The developed methods for segmentation, joint space width (JSW) measurement and erosion detection are generic and can also be applied for the feet and other body parts. By creating anatomical models based on projection images of the hand skeleton, we have found robust solutions for image processing problems related to positioning variability and image quality. In addition, biometric features obtained from the bone contours proved to be useful for detecting errors in x-ray data sets.

With regard to JSW measurements, we have investigated several factors that influence the reliability and precision of measurement outcomes. Supported by analysis and simulations, we have demonstrated that, apart from the used image processing method, also the applied quantification method and the x-ray acquisition protocol are relevant.

In the following paragraphs we evaluate our findings by answering the research questions listed in Section 1.5.

*Is it possible and feasible to measure joint space narrowing and erosion with sufficient precision and reproducibility to replace measurements by human experts?*

As presented in this thesis, extensive automation of radiographic analysis, such as measuring joint space narrowing and quantifying bone damage, is possible

(Chapters 6 and 8). Although the success rate of the developed JSW measurement method is promising, verification of the results by an operator is advisable. However, the level of expertise needed for this task is considerably less compared to what is required for established scoring methods such as the Sharp/van der Heijde score (SHS). Besides that manual JSW measurements are very inefficient considering the amount of time and effort needed, our results indicate that automated measurements have a better reproducibility and precision (Section 6.3). In addition, the results of automated measurements minimize the influence of subjectivity.

In practice, erosions are seldom truly quantified, because erosion volumes are difficult to estimate using projection radiographs. Instead current methods (such as the SHS) rely on scores assigned through visual inspection by an expert. Generally, these scores are based on a complex combination of visible features, such as erosions, joint space narrowing, deformations and bone density differences. With the presented image subtraction technique it is possible to detect the development of bone erosions (Section 8.4). Furthermore, by standardizing the radiographic acquisition protocol and using calibration aids, it is possible to quantify bone loss.

In spite of the many advantages offered by automated methods, it is unlikely that they can replace assessment by human experts completely. Automated methods are bound to the specific conditions and tasks for which they have been designed. This can possibly result in false measurements and unnoticed points of importance, especially in cases of abnormal and severe deviations.

***What is the validity of a newly developed score compared to the current gold standard, the Sharp/van der Heijde score?***

To determine the performance of our JSW measurement method, we have worked together with several other research groups that are developing similar methods. A comparison between four automated methods showed some systematic differences between JSW measurements. However, change measured in serial radiographs showed good agreement, with no observable systematic differences (Section 6.4 and [60]).

The developed JSW measurement method has also been tested on a large dataset of radiographs that had been assessed according to the SHS. With regard to the joint space narrowing component of the SHS, in most cases our automated method shows better results than manual readings. However, the combined SHS (joint space narrowing + erosion score) outperforms the automated method [60].

***What is the optimal combined score for joint damage in hand and feet caused by RA?***

As the extent of joint space narrowing and erosion formation varies between patients, it is difficult to come to an overall and optimal combined measure. The established SHS offers a well validated combination between JSW measurements and erosion scoring. However, with regard to JSW measurements, automated

measurements perform better than manual readings [60]. This suggests that the SHS can be improved by replacing the manual JSW readings with the automated JSW measurements. This improvement is particularly relevant in cases where erosions are not predominant.

If a positioning aid is used for radiographic acquisition, it is possible to make quantitative measurements of bone loss using the described image subtraction technique (Chapter 8). Whether such measurements can be used as a supplement or replacement for the erosion score of the SHS, is open for future research.

## 9.2 Recommendations

### *How can an automated measurement system be applied practically within rheumatology?*

With regard to automated measurement systems, there are two different application areas within rheumatology. The first is in clinical trials, where large quantities of radiographic image data require objective analysis. Automated measurement systems shorten the time that is spent on radiographic assessment, thereby reducing the required validation period for new treatment methods. The second application area is within the rheumatology clinic. A rheumatologist or radiologist may be served by having additional tools to objectively monitor the disease progression of his patients. This also enables the follow-up of therapeutic interventions. Obviously, the second application area involves a larger user group, setting higher demands on the implementation of a measurement system than within a controlled research area. Such a system must be robust and user friendly. A practical implementation will also demand seamless integration and coexistence with existing software that is used in the clinic.

For automated assessment it is important that the radiographic acquisition protocol is standardized. This can be effectively achieved by using a positioning aid (Section 7.5). In this way positioning variability can be minimized, which will improve the segmentation results and the precision of JSW measurements. An additional advantage is that quality indicators and reference objects can be included, which can be used to verify and assess imaging parameters later on. As already indicated, the use of a positioning aid also enables image subtraction techniques, thereby creating new research opportunities for quantitative measurements of bone loss.

The image data that has been used for this thesis, originated from film radiographs scanned at a high resolution (42 micron pixel size). The new generation of digital x-ray detectors have a lower resolution (typically 100 micron pixel size), but generally display a higher image contrast due to digital filtering techniques. How this affects the performance of our image processing algorithms is not yet clear. To investigate this, it is important that over a period of several years a large set of new image data is acquired. Preferably a new training set of manually segmented image data should be created, in order to optimize our segmentation

and margin detection methods. For a reliable validation, it is important that the joint margins are indicated independently by a panel of experienced radiographic experts.

# Bibliography

- [1] D. van der Heijde, “How to read radiographs according to the Sharp/van der Heijde method,” *The Journal of Rheumatology*, vol. 27, no. 1, pp. 261–263, 2000.
- [2] A. F. Moore, K.L.D., “Introduction to clinically oriented anatomy,” in *Clinically Oriented Anatomy* (P. J. Kelly, ed.), p. 15, Philadelphia: Lippincott Williams and Wilkins, 4th ed., 1999.
- [3] Dorland’s, “Dorland’s medical dictionary.” Website, 2007. <http://www.mercksource.com>.
- [4] H. Smith, “Rheumatoid arthritis.” Website, 2006. <http://www.emedicine.com/med/topic2024.htm>.
- [5] M. Buch and P. Emery, “The aetiology and pathogenesis of rheumatoid arthritis,” *Hospital Pharmacist*, vol. 9, pp. 5–10, 2002.
- [6] J. Rindfleisch and D. Muller, “Diagnosis and management of rheumatoid arthritis,” *American Family Physician*, vol. 72, no. 6, pp. 1037–47, 2005.
- [7] W. Shiel, “Rheumatoid arthritis.” Website, 2007. [http://www.medicinenet.com/rheumatoid\\_arthritis/article.htm](http://www.medicinenet.com/rheumatoid_arthritis/article.htm).
- [8] L. Carmona, I. Gonzalez-Alvaro, A. Balsa, M. Belmonte, X. Tena, R. Sanmarti, and E. Group, “Rheumatoid arthritis in Spain: occurrence of extra-articular manifestations and estimates of disease severity,” *Annals of the Rheumatic Diseases*, vol. 62, no. 9, pp. 897–900, 2003.
- [9] R. Goldbach-Mansky, J. Lee, A. McCoy, J. Hoxworth, C. Yarboro, J. Smolen, G. Steiner, A. Rosen, C. Zhang, H. Menard, Z. Zhou, T. Palosuo, W. Van Venrooij, R. Wilder, J. Klippel, J. Schumacher, H.R., and H. El-Gabalawy, “Rheumatoid arthritis associated autoantibodies in patients with synovitis of recent onset,” *Arthritis Research*, vol. 2, no. 3, pp. 236–43, 2000.
- [10] J. Bijlsma, J. Hoes, A. Van Everdingen, S. Verstappen, and J. Jacobs, “Are glucocorticoids DMARDs?,” *Annals of the New York Academy of Sciences*, vol. 1069, no. 1, pp. 268–274, 2006.

- 
- [11] W. Shiel, "Osteoarthritis (degenerative arthritis)." Website, 2007. <http://www.medicinenet.com/osteoarthritis/article.htm>.
- [12] Reumafonds, "Reumafonds." Website, 2007. <http://www.reumafonds.nl>.
- [13] G. Verkerk, J. Broens, W. Kranendonk, F. van der Puijl, J. Sikkema, and C. Stam, "Elektromagnetisch spectrum," in *Binas*, pp. 32–33, Groningen: Wolters-Noordhoff, 1992.
- [14] A. Bos, F. Draaisma, W. Okx, and C. Rasmussen, "Stralingsbronnen," in *Inleiding tot de Stralingshygiene*, Maarssen, The Netherlands: Elsevier gezondheidszorg, 2000.
- [15] J. Sharp, "An overview of radiographic analysis of joint damage in rheumatoid arthritis and its use in meta-analysis," *The Journal of Rheumatology*, vol. 27, no. 1, pp. 254–260, 2000.
- [16] A. Larsen, "How to apply Larsen score in evaluating radiographs of rheumatoid arthritis in long-term studies," *The Journal of Rheumatology*, vol. 22, no. 10, pp. 1974–1975, 1995.
- [17] 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 and Rheumatism*, vol. 14, no. 6, pp. 706–720, 1971.
- [18] R. Rau, S. Wassenberg, G. Herborn, G. Stucki, and A. Gebler, "A new method of scoring radiographic change in rheumatoid arthritis," *The Journal of Rheumatology*, vol. 25, no. 11, pp. 2094–2107, 1998.
- [19] D. Trentham and A. Masi, "Carpo:metacarpal ratio. A new quantitative measure of radiologic progression of wrist involvement in rheumatoid arthritis," *Arthritis and Rheumatism*, vol. 19, no. 5, pp. 939–944, 1976.
- [20] J. Dacre, J. Coppock, K. Herbert, D. Perrett, and E. Huskisson, "Development of a new radiographic scoring system using digital image analysis," *Annals of the Rheumatic Diseases*, vol. 48, no. 3, pp. 194–200, 1989.
- [21] J. Kirwan, "Using the Larsen index to assess radiographic progression in rheumatoid arthritis," *The Journal of Rheumatology*, vol. 27, no. 1, pp. 264–268, 2000.
- [22] S. Young-Min, S. Shakhapur, N. Marshall, I. Griffiths, T. Cawston, and A. Grainger, "Modified Larsen scoring of digitized radiographs in rheumatoid arthritis," *The Journal of Rheumatology*, vol. 30, no. 2, pp. 238–240, 2003.

- [23] D. Scott, "Radiographs in rheumatoid arthritis," *International Journal of Advances in Rheumatology*, vol. 1, no. 1, 2003.
- [24] D. van der Heijde, T. Dankert, F. Nieman, R. Rau, and M. Boers, "Reliability and sensitivity to change of a simplification of the Sharp/van der Heijde radiological assessment in rheumatoid arthritis," *Rheumatology*, vol. 38, no. 10, pp. 941–947, 1999.
- [25] D. van der Heijde, A. Boonen, M. Boers, P. Kostense, and S. Der Linden, "Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials," *Rheumatology*, vol. 38, no. 12, pp. 1213–1220, 1999.
- [26] M. Lassere, D. van der Heijde, K. Johnson, M. Boers, and J. Edmonds, "Reliability of measures of disease activity and disease damage in rheumatoid arthritis: implications for smallest detectable difference, minimal clinically important difference, and analysis of treatment effects in randomized controlled trials," *The Journal of Rheumatology*, vol. 28, no. 4, pp. 892–903, 2001.
- [27] K. Bruynesteyn, D. van der Heijde, M. Boers, M. Lassere, A. Boonen, J. Edmonds, H. Houben, H. Paulus, P. Peloso, A. Saudan, and S. van der Linden, "Minimal clinically important difference in radiological progression of joint damage over 1 year in rheumatoid arthritis: preliminary results of a validation study with clinical experts," *The Journal of Rheumatology*, vol. 28, no. 4, pp. 904–910, 2001.
- [28] J. Buckland-Wright, I. Carmichael, and S. Walker, "Quantitative microfocal radiography accurately detects joint changes in rheumatoid arthritis," *Annals of the Rheumatic Diseases*, vol. 45, no. 5, pp. 379–383, 1986.
- [29] M. Browne, P. Gaydecki, R. Gough, D. Grennan, S. Khalil, and H. Mamtora, "Radiographic image analysis in the study of bone morphology," *Clinical Physics and Physiological Measurement*, vol. 8, no. 2, pp. 105–121, 1987.
- [30] P. Gaydecki, M. Browne, H. Mamtora, and D. Grennan, "Measurement of radiographic changes occurring in rheumatoid arthritis by image analysis techniques," *Annals of the Rheumatic Diseases*, vol. 46, pp. 296–301, 1987.
- [31] D. Michael and A. Nelson, "HANDX - A model-based system for automatic segmentation of bones from digital hand radiographs," *IEEE Transactions on Medical Imaging*, vol. 8, no. 1, pp. 64–69, 1989.
- [32] E. Allander, P. Forsgren, H. Pettersson, and P. Seideman, "Computerized assessment of radiological changes of the hand in rheumatic diseases," *The Scandinavian Journal of Rheumatology*, vol. 18, no. 5, pp. 291–296, 1989.

- [33] T. Conrozier, A. Tron, J. Balblanc, P. Mathieu, M. Piperno, G. Fitoussi, M. Bochu, and E. Vignon, "Measurement of the hip-joint space using an automatic digitalized-image analyzer," *Revue du Rhumatisme*, vol. 60, no. 2, pp. 137–143, 1993.
- [34] E. Vignon, M. Piperno, M.-P. Le Graverand, S. Mazzuca, K. Brandt, P. Mathieu, H. Favret, M. Vignon, F. Merle-Vincent, and a. Conrozier et, "Measurement of radiographic joint space width in the tibiofemoral compartment of the osteoarthritic knee: comparison of standing anteroposterior and Lyon schuss views," *Arthritis and Rheumatism*, vol. 48, no. 2, pp. 378–384, 2003.
- [35] M. James, G. Heald, J. Shorter, and R. Turner, "Joint space measurement in hand radiographs using computerized image analysis," *Arthritis and Rheumatism*, vol. 38, no. 7, pp. 891–901, 1995.
- [36] J. Duryea, Y. Jiang, P. Countryman, and H. Genant, "Automated algorithm for the identification of joint space and phalanx margin locations on digitized radiographs," *Medical Physics*, vol. 26, no. 3, p. 453, 1998.
- [37] J. Duryea, Y. Jiang, M. Zakharevich, and H. Genant, "Neural network based algorithm to quantify joint space width in joints of the hand for arthritis assessment," *Medical Physics*, vol. 27, no. 5, pp. 1185–1194, 2000.
- [38] J. Duryea, J. Dobbins, and J. Lynch, "Digital tomosynthesis of hand joints for arthritis assessment," *Medical Physics*, vol. 30, no. 3, pp. 325–333, 2003.
- [39] J. Sharp, J. Gardner, and E. 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*, vol. 43, no. 6, pp. 1378–1386, 2000.
- [40] 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, pp. 1825–1836, 2001.
- [41] M. Wick, P. Peloschek, K. Bogl, 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, pp. 579–582, 2003.
- [42] G. Langs, P. Peloschek, and H. Bischof, "Locating joints in hand radiographs," in *Proceedings of the Computer Vision Winter Workshop* (O. Drbohlav, ed.), (Valtice, Czech Republic), pp. 97–102, Czech Pattern Recognition Society, 2003.



- [43] G. Langs, P. Peloschek, and H. Bischof, "Determining position and fine shape detail in radiological anatomy," in *Pattern Recognition*, vol. 2781 of *Lecture Notes in Computer Science*, pp. 532–539, Berlin: Springer Berlin / Heidelberg, 2003.
- [44] G. Langs, P. Peloschek, and H. Bischof, "Asm driven snakes in rheumatoid arthritis assessment," in *Image Analysis*, vol. 2749 of *Lecture Notes in Computer Science*, pp. 454–461, Berlin: Springer Berlin / Heidelberg, 2003.
- [45] P. Bird, M. Lassere, R. Shnier, and J. Edmonds, "Computerized measurement of magnetic resonance imaging erosion volumes in patients with rheumatoid arthritis: a comparison with existing magnetic resonance imaging scoring systems and standard clinical outcome measures," *Arthritis and Rheumatism*, vol. 48, no. 3, pp. 614–624, 2003.
- [46] T. Jensen, M. Klarlund, M. Hansen, K. Jensen, J. Podenphant, T. Hansen, H. Skjodt, and L. Hyldstrup, "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*, vol. 63, no. 1, pp. 15–22, 2004.
- [47] J. Duryea, S. Zaim, and F. Wolfe, "Neural network based automated algorithm to identify joint locations on hand/wrist radiographs for arthritis assessment," *Medical Physics*, vol. 29, no. 3, pp. 403–411, 2002.
- [48] N. Efford, "Knowledge-based segmentation and feature extraction of hand radiographs," Tech. Rep. 94.31, University of Leeds, 1994.
- [49] H. Thodberg and A. Rosholm, "Application of the active shape model in a commercial medical device for bone densitometry," in *Proceedings of British Machine Vision Conference* (T. F. Cootes and C. J. Taylor, eds.), (Stockport, UK), pp. 43–52, 2001.
- [50] T. Cootes, A. Hill, C. Taylor, and J. Haslam, "Use of active shape models for locating structure in medical images," *Image and Vision Computing*, vol. 12, no. 6, pp. 355–365, 1994.
- [51] H. Thodberg, "Hands-on experience with active appearance models," in *Medical Imaging 2002: Image Proceedings* (F. Sonka and, ed.), vol. 4684, pp. 495–506, SPIE, 2002.
- [52] J. Sotoca, J. Inesta, and M. Belmonte, "Hand bone segmentation in radioabsorptiometry images for computerised bone mass assessment," *Computerized Medical Imaging and Graphics*, vol. 27, no. 6, pp. 459–467, 2003.
- [53] J. Dacre and E. Huskisson, "The automatic assessment of knee radiographs in osteoarthritis using digital image analysis," *The British Journal of Rheumatology*, vol. 28, no. 6, pp. 506–510, 1989.

- [54] P. Forsgren and P. Seideman, "An interobject distance measure based on medial axes retrieved from discrete distance maps," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 4, pp. 390–397, 1990.
- [55] D. Ballard and C. Brown, *Computer vision*. Englewood Cliffs: Prentice-Hall, 1982.
- [56] T. Conrozier, C. Jousseau, P. Mathieu, A. Tron, J. Caton, J. Bejui, and E. Vignon, "Quantitative measurement of joint space narrowing progression in hip osteoarthritis: a longitudinal retrospective study of patients treated by total hip arthroplasty," *The British Journal of Rheumatology*, vol. 37, no. 9, pp. 961–968, 1998.
- [57] J. Dequeker, "Quantitative radiology: radiogrammetry of cortical bone," *The British Journal of Radiology*, vol. 49, no. 587, pp. 912–920, 1976.
- [58] P. Bird, B. Kirkham, I. Portek, R. Shnier, F. Joshua, J. Edmonds, and M. Lassere, "Documenting damage progression in a two-year longitudinal study of rheumatoid arthritis patients with established disease (the DAMAGE study cohort): Is there an advantage in the use of magnetic resonance imaging as compared with plain radiography?," *Arthritis and Rheumatism*, vol. 50, no. 5, pp. 1383–1389, 2004.
- [59] T. Cootes and C. Taylor, "Combining elastic and statistical models of appearance variation," *Computer Vision - ECCV 2000*, vol. 1842, pp. 149–163, 2000.
- [60] J. Sharp, J. Angwin, M. Boers, J. Duryea, G. von Ingersleben, J. Hall, J. Kauffman, R. Landewe, G. Langs, C. Lukas, J. Maillefert, H. Bernelot Moens, P. Peloschek, V. Strand, and D. van der Heijde, "Computer based methods for measurement of joint space width: update of an ongoing OMERACT project," *The Journal of Rheumatology*, vol. 34, no. 4, pp. 874–83, 2007.
- [61] J. Kellgren and J. Lawrence, "Radiological assessment of rheumatoid arthritis," *Annals of the Rheumatic Diseases*, vol. 16, no. 4, pp. 485–93, 1957.
- [62] F. McQueen, M. Lassere, J. Edmonds, P. Conaghan, C. Peterfy, P. Bird, P. O'Connor, B. Ejbjerg, M. Klarlund, and a. Stewart et, "OMERACT rheumatoid arthritis magnetic resonance imaging studies. Summary of OMERACT 6 MR imaging module," *The Journal of Rheumatology*, vol. 30, no. 6, pp. 1387–1392, 2003.
- [63] E. Goligher, J. Duryea, M. Liang, F. Wolfe, and A. Finckh, "Radiographic joint space width in the fingers of patients with rheumatoid arthritis of less than one year's duration," *Arthritis and Rheumatism*, vol. 54, no. 5, pp. 1440–3, 2006.

- [64] J. Angwin, A. Lloyd, G. Heald, G. Nepom, M. Binks, and M. James, "Radio-graphic hand joint space width assessed by computer is a sensitive measure of change in early rheumatoid arthritis," *The Journal of Rheumatology*, vol. 31, no. 6, pp. 1050–1061, 2004.
- [65] F. Fritsch and R. Carlson, "Monotone piecewise cubic interpolation," *The SIAM Journal on Numerical Analysis*, vol. 17, no. 2, pp. 238–246, 1980.
- [66] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.
- [67] M. Stegmann, R. Fisker, B. Ersboll, H. Thodberg, and L. Hyldstrup, "Active appearance models: Theory and cases," *9th Danish Conference on Pattern Recognition and Image Analysis*, vol. 1, pp. 49–57, 2000.
- [68] M. Roberts, T. Cootes, and J. Adams, "Linking sequences of active appearance sub-models via constraints: an application in automated vertebral morphology," in *Proceedings of British Machine Vision Conference 2003*, vol. 1, pp. 349–358, 2003.
- [69] M. Bukovec, R. Truyen, B. Likar, R. Bernard, and F. Pernus, "Image segmentation by connected parametrical models," in *Medical Imaging 2004: Image Processing*, vol. 5370, pp. 398–409, SPIE, 2004.
- [70] C. Goodall, "Procrustes Methods in the Statistical-Analysis of Shape," *The Journal of the Royal Statistical Society. Series B-Methodological*, vol. 53, no. 2, pp. 285–339, 1991.
- [71] J. Gerbrands, "On the relationships between SVD, KLT and PCA," *Pattern Recognition*, vol. 14, no. 1-6, pp. 375–381, 1981.
- [72] F. Bookstein, "Principal warps - thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 567–585, 1989.
- [73] F. Crow, "The aliasing problem in computer-generated shaded images," *Communications of the ACM*, vol. 20, no. 11, pp. 799–805, 1977.
- [74] A. Jain, A. Ross, and S. Pankanti, "A prototype hand geometry-based verification system," in *Proceedings of the second International Conference on Audio- and Video-Based Personal Authentication (AVBPA)*, (Washington), pp. 166–171, 1999.
- [75] R. Veldhuis, A. Bazen, W. Booiij, and A. Hendrikse, "A comparison of hand-geometry recognition methods based on low- and high-level features," in *15th Annual Workshop on Circuits Systems and Signal Processing (ProRISC)*, (Veldhoven, The Netherlands), STW Technology Foundation, 2004.

- [76] A. Bazen and N. Veldhuis, "Likelihood-ratio-based biometric verification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 86–94, 2004.
- [77] P. Mahalanobis, "On the generalized distance in statistics," *Proceedings of the National Institute of Sciences of India*, vol. 2, pp. 49–55, 1936.
- [78] H. van Trees, *Detection, estimation, and modulation theory: Part I*. New York: Wiley, 1968.
- [79] J. Sharp, D. van der Heijde, J. Angwin, J. Duryea, H. Bernelot Moens, J. Jacobs, J. Maillefert, C. Strand, and O. Group, "Measurement of joint space width and erosion size," *The Journal of Rheumatology*, vol. 32, no. 12, pp. 2456–2461, 2005.
- [80] A. Amini, T. Weymouth, and R. Jain, "Using dynamic programming for solving variational problems in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 9, pp. 855–867, 1990.
- [81] J. Duryea, J. Li, C. Peterfy, C. Gordon, and H. Genant, "Trainable rule-based algorithm for the measurement of joint space width in digital radiographic images of the knee," *Medical Physics*, vol. 27, no. 3, pp. 580–591, 2000.
- [82] M. Boers, A. Verhoeven, H. Markusse, M. van de Laar, R. Westhovens, J. van Denderen, D. van Zeben, B. Dijkmans, A. Peeters, P. Jacobs, H. van den Brink, H. Schouten, D. van der Heijde, A. Boonen, and S. van der Linden, "Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis," *The Lancet*, vol. 350, no. 9074, pp. 309–18, 1997.
- [83] Philips Medical, "Radiography Components - Technical Specifications." Website, 2007. <http://www.medical.philips.com/main/products/xray/products/radiography/components/index.html>.
- [84] J. Hsieh, *Computed tomography: principles, design, artifacts, and recent advances*. SPIE Press Monograph (Vol. PM114), Bellingham: SPIE, 2003.
- [85] J. Duryea, G. Neumann, H. Yoshioka, and J. Dobbins Iii, "Digital tomosynthesis rendering of joint margins for arthritis assessment," in *Medical Imaging 2004: Physics of Medical Imaging*, vol. 5368, (San Diego, CA, USA), pp. 12–18, SPIE, 2004.
- [86] K. Mueller and R. Yagel, "Fast perspective volume rendering with splatting by utilizing a ray-driven approach," in *Proceedings of the 7th conference on Visualization '96*, (San Francisco, California, United States), pp. 65–72, IEEE Computer Society Press, 1996.

- 
- [87] B. Zitova and J. Flusser, “Image registration methods: a survey,” *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.
- [88] S. Periaswamy and H. Farid, “Elastic registration in the presence of intensity variations,” *IEEE Transactions on Medical Imaging*, vol. 22, no. 7, pp. 865–874, 2003.
- [89] J. Ahlberg, E. Nilson, and J. Walsh, *The theory of spline functions and their applications*. New York: Academic Press, 1967.
- [90] T. Coleman and Y. Li, “A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables,” *The SIAM Journal on Optimization*, vol. 6, no. 4, pp. 1040–1058, 1996.
- [91] T. Porter and T. Duff, “Compositing digital images,” in *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, ACM, 1984.



## List of publications

- [92] J. A. Kauffman, C. H. Slump, and H. J. B. Moens, “Segmentation of radiographs of hands with joint damage using customized active appearance models,” in *ProRISC 2004 15th Annual Workshop on Circuits, Systems and Signal Processing*, vol. 15 of *Proceedings of ProRISC*, (Veldhoven, The Netherlands), pp. 261–265, Technologistichting STW, November 2004.
- [93] J. A. Kauffman, C. H. Slump, and H. J. B. Moens, “Segmentation of hand radiographs by using multi-level connected active appearance models,” in *Medical Imaging 2005: Image Processing* (J. M. Fitzpatrick and J. M. Reinhardt, eds.), vol. 5747 of *Proceedings of SPIE*, (San Diego, California USA), pp. 1571–1581, SPIE International Society for Optical Engineering, May 2005.
- [94] J. A. Kauffman, C. H. Slump, and H. J. B. Moens, “Matching hand radiographs,” in *ProRISC 2005 16th Annual Workshop on Circuits, Systems and Signal Processing*, (Veldhoven, The Netherlands), pp. 629–633, Lemma, November 2005.
- [95] H. J. B. Moens, J. A. Kauffman, and C. H. Slump, “Automated analysis of hand radiographs by using multi-level connected active appearance models,” in *Annual European Congress of Rheumatology EULAR*, vol. 64, Suppl 3 of *Annals of the Rheumatic Diseases*, (Vienna, Austria), p. 1244, BMJ Publishing Group Ltd, June 2005.
- [96] T. T. Withaar, J. A. Kauffman, and H. J. B. Moens, “Analyzing the precision of JSW measurements using 3D scans and statistical models,” in *Proceedings of SPS-DARTS 2006, the second annual IEEE Benelux/DSP Valley Signal Processing Symposium* (W. Philips, ed.), (Antwerp, Belgium), pp. 123–126, IEEE Benelux EMBS, March 2006.
- [97] L. Segerink, T. T. Withaar, M. D. I. Lansbergen, J. A. Kauffman, M. L. Schimmelpenninck-Scheiffers, and H. J. B. Moens, “Standardisation of hand x-rays for arthritis: the relevance of hand positioning,” in *Annual European Congress of Rheumatology EULAR*, vol. 65, Suppl 2 of *Annals of the Rheumatic Diseases*, (Amsterdam, The Netherlands), p. 593, BMJ Publishing Group Ltd, June 2006.

- [98] J. A. Kauffman, C. H. Slump, and H. J. B. Moens, "Detection of joint space narrowing in hand radiographs," in *Medical Imaging 2006: Image Processing* (M. Reinhardt and J. W. P. Pluim, eds.), vol. 6144 of *Proceedings of SPIE*, (San Diego, USA), pp. 1332–1342, SPIE International Society for Optical Engineering, February 2006.
- [99] H. J. B. Moens, J. A. Kauffman, and C. H. Slump, "Detection of joint space narrowing in hand radiographs," in *Annual European Congress of Rheumatology EULAR*, vol. 65, Suppl 2 of *Annals of the Rheumatic Diseases*, (Amsterdam, The Netherlands), p. 579, BMJ Publishing Group Ltd, June 2006.
- [100] J. T. Sharp, J. Angwin, M. Boers, J. Duryea, G. von Ingersleben, J. R. Hall, J. A. Kauffman, R. Landewe, G. Langs, C. Lukas, J.-F. Maillefert, H. J. B. Moens, P. Peloczek, V. Strand, and D. van der Heijde, "Computer based methods for measurement of joint space width: update of an ongoing omeract project," *The Journal of Rheumatology*, vol. 34, pp. 874–883, 2007.
- [101] J. A. Kauffman, C. H. Slump, and H. J. B. Moens, "Revealing bone damage using radiographic image registration," in *Proceedings of SPS-DARTS 2007, the third annual IEEE Benelux/DSP Valley Signal Processing Symposium* (W. Philips, ed.), (Antwerp, Belgium), pp. 23–27, IEEE Benelux/DSP, March 2007.
- [102] C. Lukas, J. T. Sharp, J. Angwin, M. Boers, J. Duryea, J. R. Hall, J. A. Kauffman, R. Landewe, G. Langs, H. J. B. Moens, P. Peloschek, and D. van der Heijde, "Automated measurement of joint space width in small joints of patients with rheumatoid arthritis," *The Journal of Rheumatology*, vol. 35, pp. 1288–1293, June 2008.
- [103] J. A. Kauffman, C. H. Slump, and H. J. B. Moens, "Revealing bone damage in radiographs of hands and feet using image subtraction," in *Annual European Congress of Rheumatology EULAR*, vol. 67, Suppl 2 of *Annals of the Rheumatic Diseases*, (Paris, France), p. 643, BMJ Publishing Group Ltd, June 2008.



# Summary

Rheumatoid arthritis (RA) is a chronic inflammatory disease leading to severe joint damage, particularly in the wrist, fingers and toes. To prevent irreversible joint damage, it is crucial that RA is treated in an early stage. The effect of treatment methods can differ per patient. Therefore it is important that disease activity is carefully monitored, such that a treatment can be altered if necessary. In order to visualize joint damage, radiographs of hands and feet are often used. In the past fifty years several methods have been developed to express visible joint damage in a quantifiable score. In general these methods are time-consuming and depend on subjective visual readings.

With the introduction of the computer and the availability of digitized radiographs, researchers have been looking for new automated methods to measure the progression of joint damage. A measurable effect in early RA is joint space narrowing which is caused by degradation of the cartilage between the bones. A considerable part of this thesis deals with this subject.

Firstly, different methods are evaluated on how to quantify the joint space width in an accurate and robust manner. Hereby methods developed by others have been evaluated. Subsequently, we present several image processing methods for analyzing radiographs. The first is a segmentation method based on a model of the hand skeleton. With this model it is possible to detect the locations of the bones and the joints in radiographs. The second part describes a method that accurately detects the bone margins within the joint. Finally, the joint space width is calculated using the detected joint margins. This method is evaluated by comparing measurement results with both manual measurements, and automated measurements developed by others. It is shown that automated measurements are more precise than manual measurements.

It was observed from the image data, that a large variability of hand positioning is allowed during x-ray acquisition. This does not only complicate automated analysis, but can also have an effect on the projection angle within a specific joint. In successive radiographs, a change in the projection angle can negatively affect the precision of measurements. By using projection simulations it is demonstrated that this effect is relevant. Therefore, we recommend to standardize the protocol for radiographic acquisitions. A special positioning aid has been designed, which can be used to place a hand in a standard position.

Another measurable effect of RA is the development of bone damage in the form of erosions. To be able to measure this effect, we investigated the possibility

to make subtraction measurements between successive radiographs that have been taken with some time in between. It is demonstrated that different radiographs of the same bone can be compared using an image registration algorithm. By determining the difference between two images developing erosions can be revealed. However, with the current available radiographs it is not yet possible to quantify bone damage. We expect that in the future this will become feasible by applying the recommended positioning aid during x-ray acquisition.

# Samenvatting

Reumatoïde artritis (RA) is een chronische ontstekingsziekte die kan leiden tot ernstige gewrichtsschade en verminderde mobiliteit. Behandeling in een vroeg stadium van deze ziekte is belangrijk, zodat onomkeerbare gewrichtsschade voorkomen kan worden. De effectiviteit van behandelmethoden kan per patiënt verschillen. Daarom is het noodzakelijk om de ontwikkeling van het ziektebeeld nauwkeurig te volgen, zodat behandelmethoden eventueel tijdig aangepast kunnen worden. In het geval van RA wordt vaak gebruik gemaakt van röntgenfoto's van handen en voeten, om een beeld te krijgen van veranderingen in het skelet. In de afgelopen vijftig jaar zijn verschillende methoden ontwikkeld om de in röntgenfoto's waarneembare gewrichtsschade te kunnen vertalen in een score. Deze methoden zijn gewoonlijk gebaseerd op visuele analyse door geoefende radiologen of reumatologen. Het gebruik van deze methoden neemt over het algemeen veel tijd in beslag, en bovendien is de nauwkeurigheid en objectiviteit van resultaten afhankelijk van de persoon die de metingen uitvoert.

Met de komst van de computer en de mogelijkheid om röntgenfoto's te digitaliseren, is men op zoek gegaan naar nieuwe automatiseerbare methoden om de ontwikkeling van gewrichtsschade te kunnen meten. Een meetbaar effect in een vroeg stadium van RA, is de versmalling van gewrichtsspleten als gevolg van de afname van tussenliggend kraakbeen. In dit proefschrift wordt uitgebreid aandacht besteed aan dit onderwerp.

Allereerst is onderzocht op welke manier gewrichtsspleten robuust en nauwkeurig gemeten kunnen worden. Hierbij is gekeken naar de verschillen tussen reeds ontwikkelde methoden van diverse onderzoekspartijen. Vervolgens is een beeldverwerkingsmethode ontwikkeld om röntgenfoto's te analyseren. Een eerste onderdeel bestaat uit een segmentatiemethode, welke is gebaseerd op een model van het skelet van de menselijke hand. Hiermee is het mogelijk om botten en gewrichten in een röntgenfoto te lokaliseren. Een tweede onderdeel bestaat uit een algoritme waarmee met hoge nauwkeurigheid de botranden van gewrichten gedetecteerd kunnen worden. Met behulp van de gevonden botranden wordt vervolgens de breedte van de gewrichtsspleet bepaald. Door meetresultaten te vergelijken met zowel handmatige metingen als met de resultaten van methoden ontwikkeld door derden, is aangetoond dat automatische gewrichtsspleetmetingen nauwkeuriger resultaten opleveren dan handmatige.

Bij de analyse van röntgenfoto's is geconstateerd dat tijdens de röntgenopname een grote variabiliteit in handpositionering is toegestaan. Behalve dat dit de au-

tomatische analyse bemoeilijkt, kan dit ook van invloed zijn voor de projectiehoek van een specifiek gewricht. In opeenvolgende foto's kan een verandering van de projectiehoek de nauwkeurigheid van een meting negatief beïnvloeden. Met behulp van projectiesimulaties is aangetoond dat dit effect relevant is. Daaruit volgt de aanbeveling om het protocol voor het maken van röntgenfoto's te standaardiseren. Om dit te kunnen realiseren is een hulpmiddel ontworpen, waarmee een hand eenvoudig in een standaard positie geplaatst kan worden.

Een ander meetbaar effect van RA is het ontstaan van schade aan het bot in de vorm van erosies. Om dit effect te kunnen meten is onderzocht of gebruik gemaakt kan worden van verschilmetingen tussen röntgenfoto's die met tussenperioden zijn gemaakt. Door het toepassen van een beeldregistratiemethode is gebleken dat het mogelijk is om verschillende röntgenopnamen van een bot met elkaar te vergelijken. Door het bepalen van het verschil kan ontstane schade worden gedetecteerd. Met de huidige beelddata is het nog niet mogelijk om botschade te kwantificeren. De verwachting is dat dit in de toekomst wel mogelijk wordt door toepassing van het aanbevolen positioneringshulpmiddel bij de röntgenopname.

# Dankwoord

Dit promotieonderzoek had niet tot stand kunnen komen zonder de financiële steun van het Reumafonds. Daarom wil ik graag iedereen bedanken die bijdraagt aan het belangrijke werk en streven van deze organisatie.

Minstens zo belangrijke steun ontving ik in de vorm van hulp en betrokkenheid van mensen om mij heen. In de hoop dat ik niemand vergeet, wil ik enkelen hiervan met name noemen.

Allereerst wil ik mijn beide begeleiders bedanken voor het mogelijk maken van mijn onderzoek: Mijn promotor, Kees Slump, voor het geven van alle vrijheid en vertrouwen tijdens het uitvoeren van mijn onderzoek. Maar ook voor die heerlijke verse pot (echte) koffie, die hij elke ochtend –speciaal voor mij– zet. Mijn assistent promotor, Hein Bernelot Moens, die ik heb leren kennen als een reumatoloog met een enorm enthousiasme en interesse voor techniek en innovaties in zijn vakgebied. Iemand die tijdens drukke periodes voor je klaar staat om de handen uit de mouwen te steken. Zelfs als het gaat om saaie en tijdrovende klussen, zoals het nauwkeurig aanbrengen van landmarks in honderden röntgenfoto's. Kees en Hein, bedankt!

Meerdere interessante (onverharde) zijpaden binnen dit onderzoek zouden niet zijn bewandeld zonder de hulp van een aantal studenten. Hiervoor ben ik hen erg dankbaar. Zo heeft Thijs Withaar mij geholpen met het verkennen van 'active appearance modellen' en de analyse van röntgen projecties, Jelmer van Schoonhoven heeft onderzoek gedaan naar alternatieven voor botdichtheidmetingen, en Loes Segerink heeft een mal met gebruikersprotocol ontwikkeld om in de toekomst betrouwbaardere metingen te kunnen doen.

Graag wil ik de leden en betrokkenen van de OMERACT subcommissie (aangaande geautomatiseerde gewrichtsspleetmetingen) bedanken voor de samenwerking en het delen van informatie en resultaten. In het bijzonder noem ik daarbij Jane Angwin, Jeff Duryea, prof. Maarten Boers en wijlen John Sharp voor het beschikbaar stellen van onderzoeksdata.

Binnen de Ziekenhuisgroep Twente is Hengelo ben ik dank verschuldigd aan verschillende medewerkers. Waaronder Catherine Luiken-MacLaren van de afdeling radiologie, die ons geduldig geholpen heeft met, zelfs in mijn ogen, ongebruikelijke experimenten. Van de afdeling klinische fysica, Sanne Vaartjes en Michaël Lansbergen voor hun hulp en inzet bij het verder ontwikkelen van een hand- en voetmal.

Verder wil ik Dr. R. Hildebrand van het Institute of Anatomy te Muenster bedanken voor zijn hulp bij mijn onderzoek. Dankzij hem en prof. Enrico Marani

heb ik onder andere de kans gekregen om de bovenste extremiteit ook eens in kleur (en geur) te onderzoeken.

De prettige werksfeer gedurende al die jaren, heb ik te danken aan al mijn collega's van de leerstoel Signalen en Systemen. In het bijzonder wil ik Anneke, Geert-Jan en Henny bedanken voor al die –soms onopgemerkte– zaken, die het werk merkbaar gemakkelijker maken. Ook wil ik mijn kamergenoten bedanken, voor de vele interessante discussies en fijne gesprekken, maar ook voor de rust wanneer dat even nodig was. Ron, Marnix, René en Dirk-Jan bedankt!

Ik bedank mijn familie en vrienden, voor hun interesse in mijn werk, maar ook voor hun bijdrage aan de fijne en gezellige momenten buiten werktijd. Vooral mijn ouders, die mij en mijn zussen altijd alle denkbare vrijheid en kansen gegeven hebben.

Tot slot, aan mijn bron van liefde en geluk: Lieve Tanja, bedankt voor je hulp, alle aanmoedigingen en geduld. Met jou trouwen is zonder twijfel het slimste wat ik ooit heb gedaan!