Analysis of Dynamic Contrast Enhanced MRI Datasets

by

Olga Kubassova

Submitted in accordance with the requirements
for the degree of Doctor of Philosophy.

The University of Leeds
School of Computing

October 2007

The candidate confirms that the work submitted is her own and that the appropriate credit has been given where reference has been made to the work of others.

Abstract

The purpose of this research is to perform automated analysis of 4D dynamic contrast enhanced MRI datasets (DCE-MRI) of the hand and wrist relating to rheumatoid arthritis (RA) studies. In DCE-MRI, sequences of images are acquired from the joints over time, during which a contrast agent pre-injected into a patient enhances disease affected tissues. Measurement of this enhancement, which is specific to voxels representing particular tissue types, allows assessment of the patient's condition.

Currently, analysis of DCE-MRI data is performed using semi-automated or manual techniques, which are time-consuming and subjective. These approaches involve no pre-processing techniques that can compensate for patient motion and hardware instability, or locate the tissue of interest.

In this thesis we present a solution for fully automated objective assessment of DCE-MRI data acquired from RA patients. Analysis begins with application of a registration technique that permits compensation for patient motion. Secondly, independent automatic algorithms for accurate segmentation of both bone interiors, joint exteriors, and blood vessels from data volumes of the metacarpophalangeal joints are introduced.

Performance of the segmentation algorithms is evaluated with both state-of-the art and novel techniques developed as a part of this thesis. We have utilised and enhanced a supervised approach and developed a family of unsupervised metrics for automated evaluation of segmentation outputs.

Lastly, the datasets are interpreted using a model-based approach, which permits understanding of the behaviour of tissues undergoing the medical procedure, and allows for a robust and accurate extraction of various parameters that quantify the extent of inflammation in RA patients.

The algorithms proposed have been demonstrated on datasets acquired with both low and high field scanners, from different joints, using various pulse sequences. They are user-independent, time efficient, and generate easily reproducible and objective results. Expert observers found our results promising for possibly future diagnosis and monitoring of RA.

Acknowledgements

Foremost, I would like to express my gratitude to my family, friends, and colleagues in the School of Computing, who shared their expertise and were always available to talk about new ideas and results. A special thanks to my supervisor Prof. Roger D. Boyle, who guided me with discussion and advice, but always left room for my own ideas. His scrupulous attention to details and thoroughness were extremely valuable in the preparation of this thesis.

This work would never have been accomplished without generous help and encouragement provided by medical experts and experienced observers: Dr. Steven F. Tanner and Dr. A. Radjenovic [Department of Medical Physics, Leeds General Infirmary, Leeds, UK], Dr. Mikael Boesen, Dr. Karl E. Jensen and Prof. Henning Bliddal [Frederiksberg Hospital, Parker Institute, Copenhagen, Denmark], and Dr. Marco A. Cimmino [Clinica Reumatologica, Genoa, Italy].

I also would like to acknowledge the support received from a Dorothy Hodgkin Postgraduate Award.

It is impossible to list all the individuals who assisted me in this endeavor, but an especially big thank-you to my mother Dr. Irina Kubassova for her love and forbearance, my husband Paul Robb for learning about the metacarpophalangeal joints more than he even wanted to know, Prof. Roger D. Boyle, Prof. David Cope, Prof. Reinhold Behringer, Matthew Birtwisle, and Paul Robb for wise counsel regarding the nuances of the English language. Thanks to all the friends I made during my time at University of Leeds: Dr. Debbie and John Elliott, Dr. Mark Everingham, Dima Al Damen, Prit Rehal, and others.

Declarations

This work won Leeds University's Inaugural Sir Peter Thompson Award for Enterprise. Some parts of the work presented in this thesis have been published in the following articles.

O. Kubassova, R. D. Boyle, and A. Radjenovic. Quantitative analysis of dynamic contrast-enhanced MRI datasets of the metacarpophalangeal joints. *Academic Radiology*, 14(8): 1189–1200, 2007.

O. Kubassova, M. Boesen, R. D. Boyle, M. A. Cimmino, K. E. Jensen, H. Bliddal, and A. Radjenovic. Fast and robust analysis of dynamic contrast enhanced MRI datasets. *In Proceedings of Medical Image Computing and Computer Assisted Intervention*, 2007. To appear.

O. Kubassova. Automatic segmentation of blood vessels from dynamic MRI datasets. *In Proceedings of Medical Image Computing and Computer Assisted Intervention*, 2007. To appear.

O. Kubassova, R. D. Boyle, and A. Radjenovic. Novel method for quantitative evaluation of segmentation outputs for dynamic contrast-enhanced MRI data in RA studies. *In Proceedings of Joint Disease Workshop, Medical Image Computing and Computer Assisted Intervention*, 1:72–79, 2006.

O. Kubassova, R. D. Boyle, and A. Radjenovic. Improved parameter extraction from dynamic contrast-enhanced MRI data in RA studies. *In Proceedings of Joint Disease Workshop, Medical Image Computing and Computer Assisted Intervention*, 1:64–71, 2006.

O. Kubassova, R. D. Boyle, and A. Radjenovic. Statistical analysis of DEMRI data of MCP joint. Technical Report 2006.3, University of Leeds, Leeds, UK, 2006. `http://www.comp.leeds.ac.uk/research/pubs/reports/2006/2006_03.pdf` (last access on 20.09.07).

O. Kubassova, R. D. Boyle, and M. Pyatnizkiy. Bone segmentation in metacarpophalangeal MR data. *In Proceedings of 3rd International Conference on Advances in Pattern Recognition* , 2:726–735, 2005.

O. Kubassova and R. D. Boyle. Segmentation of 4D natural MR images based upon morphological image analysis and image geometry. *In Proceedings of EPSRC Conference on Postgraduate Research in Electronics, Photonics, Communications and Software (PREP'05), 1:186–188, 2005.*

Contents

List of Figures

List of Tables

Table of abbreviations

CT	Computer tomography
DCE-MRI	Dynamic contrast-enhanced magnetic resonance imaging
DOF	Degree of freedom
EM	Expectation maximisation
fMRI	Functional magnetic resonance imaging
FOV	Field of view
Gd-DTPA	Gadolinium diethylene triamine pentacetic acid
GRE	Gradient echo
GT	Ground truth will be used to indicate manual segmentation or gold standard or correct result
IRE	Initial Rate of Enhancement
MCP	Metacarpophalangeal
MCPJs	Metacarpophalangeal joints
ME	Maximum Enhancement
MO	Mutual overlap
MRI	Magnetic resonance imaging
MRSI	MR spectroscopic imaging
MSE	Mean square error
PCA	Principal component analysis
PET	Proton emission tomography
RA	Rheumatoid arthritis
RF	Radio-frequency
ROI	Region of interest
RMS	Root mean square
TE	Echo time
TR	Repetition time
SE	Spin echo
SPECT	Single proton emission computer tomography
SWM	Support vector machines
US	Ultrasound
WM	Weighted majority

Chapter 1

Introduction

In today's world, where medicine and technology collide, the use of computer science is one more tool in the medical professionals' arsenal [23, 92, 118]. The involvement of machine vision in medical image analysis assists doctors in the data acquisition, reconstruction, evaluation, and ultimately disease diagnosis.

For years, medical imaging modalities such as X-ray, computer tomography (CT), and magnetic resonance imaging (MRI) have allowed doctors to perform non-invasive human body examinations. However, the results obtained with these techniques are limited by their reliability, reproducibility, and subjective interpretation.

This thesis attempts to solve one of the many problems associated with the interpretation of such data. We have chosen to focus our attention on dynamic datasets pertaining to rheumatoid arthritis (RA) acquired with MRI scanners.

The prime objective of this work is to deliver algorithms that assist medical experts in the interpretation and evaluation of datasets acquired from RA patients; specifically, to overcome problems associated with the methods currently used for the data assessment. A second objective of this work is to demonstrate the value of pre-processing techniques which compensate for a patient's movement and allow for the location of tissues of interest. Lastly, we discuss the value of evaluation techniques, which are used to assess the reliability of the results produced by the pre-processing techniques.

1.1 Monitoring rheumatoid arthritis

Rheumatoid arthritis is an autoimmune condition that causes swelling, pain, stiffness, and redness in the joints, which often become unstable resulting in deformities, for example of the hand. Any joint may be affected, but it is commonly the hands (metacarpophalangeal (MCP) joints), feet, and wrists.

In RA, the synovium, which is the smooth lining of a joint, becomes inflamed and changes its character causing cartilage destruction and secondary joint damage. The normally thin synovium becomes thick and makes the joint swollen and puffy to the touch. Figure 1.1 illustrates normal and arthritic joints[1].

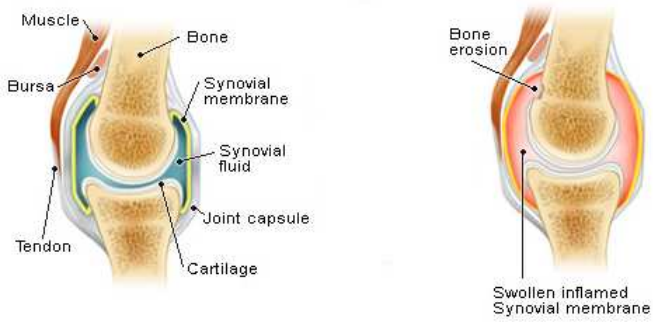


Figure 1.1: Left: Normal joint. Right: Arthritic joint. The inflammation of synovial tissue is shown in red.

RA is a progressive illness that can result in joint destruction and severe disability. The cause of RA is unknown and its course varies from patient to patient. When tissues are inflamed the disease is called active; when symptoms of the disease disappear and the tissue inflammation subsides, the disease is inactive (in remission). Remission can occur spontaneously or with treatment, and might last for weeks. However, patients often relapse after a short period.

RA can begin at any age, but most often starts after the age of forty and before the age of sixty [7]. More than two percent of the world-wide adult population suffer from RA, corresponding to several million people in the USA and about 0.5 million in the UK [225].

There is no known cure for RA and the treatment normally involves a combination of exercise, medications, and occasionally surgery. To date, the goal of the treatment has been to reduce the inflammation and pain, and prevent joint destruction. Early medical intervention has been shown to be important in improving outcomes and preventing serious disability [129].

[1] Taken with permission from [223].

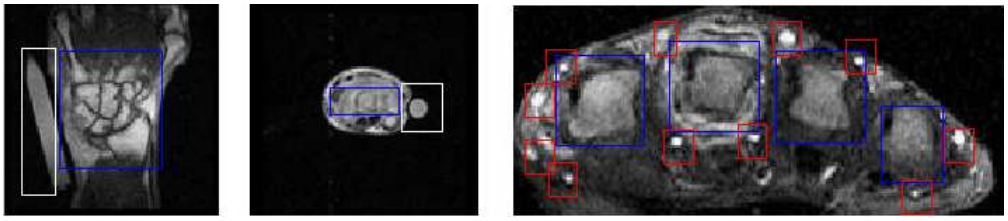


Figure 1.2: Left: MR image of wrist acquired by a low-field scanner. Middle: MR image of the MCP joints acquired by a low-field scanner. Right: MR image of the MCP joints acquired by a high-field scanner. In the images markers are marked by white rectangles, bones by blue rectangles, blood vessels by red rectangles.

1.2 Diagnosis of rheumatoid arthritis and the value of MRI

The diagnosis of RA begins with acquiring information about the severity of symptoms such as pain, morning stiffness, fatigue, etc. This information is used as a baseline for the future evaluation of the disease progression. Secondly, the patient's joints are examined for indication of warmth, swelling, and limitation of motion. Then, laboratory tests that include the measuring of a rheumatoid factor (an antibody produced in response to RA) in the blood and analysis of the synovial fluid, are performed. Lastly, the joints are examined with a conventional X-ray.

The inexperienced clinician may have difficulty in diagnosing RA. The main reasons are inter-patient variability of the disease patterns and the similarity of RA to other diseases such as lupus, osteoarthritis, and gout. Often, the tests are inconclusive; in early RA, the blood tests and X-ray may be normal.

Recently, MRI has emerged as a promising technique for the assessment and monitoring of RA [6], and became an alternative diagnostic tool to the conventional clinical examination and radiography [77]. MRI can identify all kinds of tissue, poses minimal risk to health and has no limit to the number of images that can safely be taken. Patients require no preparation (unless the contrast agent is used), and there is no recovery time. MRI is non-invasive and does not utilise radiation, and produces three-dimensional images with a high tissue contrast [90]. Figure 1.2 illustrates MR images of the wrist and MCP joints, where markers are marked white white rectangular, bones – by blue, blood vessels – by red.

Numerous studies have shown that MRI is more sensitive than conventional clinical examination and radiography for the detection of early inflammatory and destructive changes [77, 83, 128, 240]. It allows for the detection of RA bone erosions, inflammatory

soft tissue changes such as synovitis, tenosynovitis and enthesitis[2] earlier than conventional radiography [173]. Østergaard et al. have demonstrated the predictive value of MRI in the detection of bone oedema and bone erosions with respect to the subsequent radiographic progression [173].

The diagnostic value of MRI is still being investigated. Some studies [232, 234] suggest that the incorporation of MRI assessment of synovitis increases the accuracy of the examinations and allows for earlier RA diagnosis.

MRI has been enhanced by the introduction of contrast agents, which allow for even better distinction between normal and abnormal tissues [180, 250]. In the presence of the contrast agent, the data is acquired in sequential slices over a period of time during which the intensity of the inflamed tissues in MR images changes in response to the injected contrast agent. This technique is known as dynamic contrast-enhanced MRI (DCE-MRI).

DCE-MRI is proven to be a valuable tool in the assessment of RA patients [175]. It appears to provide a sensitive measure of the disease progression. However, the widespread use of DCE-MRI is limited by the need for efficient techniques for data processing, interpretation, and visualisation.

1.3 Problems associated with DCE-MRI data

Acquisition of a temporal slice takes approximately 4 minutes and a dataset acquired from a patient suffering from rheumatoid arthritis might contain up to 15 temporal slices. An examination might result in up to 300 images, where some pixels, normally located in the disease affected areas, are enhancing in response to the contrast agent.

To acquire datasets used in this work, patients were expected to hold a hand still for up to 24 minutes (acquisition of up to 6 temporal slices). The patients affected by RA often cannot hold their hand still, therefore images are often corrupted by artefacts due to the patient movement.

Assessment of the patient's condition is performed via quantification of the volume of inflamed synovium in each rheumatoid joint. However, manual measurement of the volume is time-consuming and highly subjective. Moreover, the presence of the noise, sparse location of the enhancing pixels, and subtle intensity changes make interpretation of the DCE-MRI data difficult.

To assess tissue condition, it is normally assumed that each voxel represents a particular tissue type. However, problems with patient motion during imaging might render this

[2] Enthesitis is an inflammation of the entheses, the location where a bone is joint to a tendon or a ligament [260].

assumption invalid and introduce artefactual enhancement in some tissues.

The data acquisition settings are particular for the scanning hardware – therefore, an objective comparison of the data acquired with different equipment is not trivial.

Hence, we are dealing with several problems. Firstly, there is a need for efficient pre-processing techniques that can compensate for patient motion, locate tissue of interest, and reduce artefactual enhancement. Secondly, an objective quantitative technique, that would not be influenced by the noise, scanning equipment, acquisition parameters, and time-course of arrival of the contrast agent, is required to allow for the assessment and interpretation of DCE-MRI datasets.

1.4 DCE-MRI data analysis

In this thesis we present a solution for a fully automated and objective evaluation of DCE-MRI data acquired from RA patients. The data analysis begins with the application of a registration technique that compensates for patient's motion. Secondly, we introduce segmentation techniques that allow for the elimination of irrelevant tissues such as bone interior, which includes trabecular and cortical bone cross-sections and certain blood vessels, which can complicate the interpretation of the data.

Performance of the algorithms is assessed with state-of-the art and novel evaluation techniques, which were developed as a part of this thesis. We used and enhanced a supervised metric and developed a family of unsupervised metrics for the automated evaluation of the segmentation outputs.

Lastly, the datasets are interpreted using a model-based approach, which permits the understanding of the behaviour of tissues undergoing the medical procedure, and allows for a robust and accurate extraction of various parameters that quantify the extent of inflammation in RA patients. Additionally, with this method it is possible to locate tissues that did not fully absorbed the contrast agent by the end of imaging procedure.

The experiments were performed on datasets acquired from different joints using scanning hardware of both high and low field strengths, confirming the portability of the approach. The process takes into account many variable factors that affect the clarity of these datasets and their interpretation, such as the physical tremor associated with rheumatoid arthritis and the time taken for the diseased tissue to absorb the contrast agent.

This combination of techniques permits the automated analysis of DCE-MRI data, which should reduce the amount of time radiologists current spend on data assessment and increase the efficiency and reproducibility of the RA evaluation with MRI.

1.5 Overview of this thesis

In this chapter we outlined the problems associated with monitoring RA and highlighted the techniques that are going to be introduced in this thesis. The remaining chapters are organised as follows:

Chapter 2: Background discusses the fundamental aspects and background research in various areas of medical image analysis such as data registration, segmentation, evaluation, and quantitative analysis.

Chapter 3: MRI data firstly, highlights advantages and drawbacks of the various imaging modalities for the monitoring of rheumatoid arthritis and discusses MRI scanning in more detail. Secondly, we introduce the DCE-MRI datasets involved in this research, discuss their properties, and acquisition parameters.

Chapter 4: Image registration introduces our modification to a publicly available registration algorithm and its application to DCE-MRI data.

Chapter 5: Image segmentation discusses the automatic algorithms for the joints' envelope and bone interiors segmentation from the DCE-MRI data.

Chapter 6: Evaluation techniques presents supervised and unsupervised evaluation metrics and shows how they can be employed to assess the results produced by the segmentation algorithms.

Chapter 7: Quantitative analysis of DCE-MRI datasets introduces a model-based algorithm that permits the objective analysis of the DCE-MRI data and visualisation of the activation events.

Chapter 8: Blood vessel segmentation presents a technique for the automatic segmentation of the blood vessels from the DCE-MRI datasets.

Finally, the conclusions and possible extensions of this work are presented in Chapter 9.

Chapter 2

Background

2.1 Medical image analysis

Medical image analysis is often perceived as a set of techniques that allow us to reconstruct, display, analyse and interpret the data acquired from a human body or parts thereof for disease detection and diagnostic purposes. Generally, the steps involved in assessment of medical data include image pre-processing (registration and segmentation) and image interpretation (via perception, understanding, evaluation, and visualisation) [203, 273].

Registration involves finding a transformation that brings different images of the same object into strict congruence. Segmentation approaches allow for accurate recognition and delineation of important objects in the image. Interpretation includes display and manipulation of the data. Various evaluation approaches are used to assess the algorithms' performance.

In this chapter we aim to provide an overview of these steps, to outline the main components of image processing algorithms, and to discuss the recent and most successful techniques for registration, segmentation and its evaluation, and visualisation in application to DCE-MRI datasets.

2.2 Registration techniques

Registration is a fundamental task in medical image processing which seeks to match two data sets that were acquired at different time or view points or by different image modalities [71, 152]. It is required when, for example, an atlas image is needed to be aligned to the patient's data for automatic identification of tissue anatomy and lesion location, or when pre- and post- treatment images need to be matched in order to monitor disease progression [61, 150].

Generally, registration is the determination of a transformation (rigid or non-rigid) that aligns pixels from one image or volume with pixels from another [108]. To detect disease progression or a treatment effect, patients are often re-scanned after a long period of time. Registration of such 4D datasets is another more complicated and challenging problem.

Rigid methods [109, 183] are applied when alignment of two images can be done via rotation and translation. Today rigid registration is often extended to include affine transformations, which incorporates scale factors and shears [108].

However, deformation of the human body can only be poorly approximated by rigid models. Therefore, some contemporary work concentrates on developing non-rigid methods, which can model complex motions caused by inconsistencies in a patient's posture or differences in an organ's shape and volume [13, 57]. Obviously, there is often a trade-off between the complexity of the transformation model and the speed of the alignment.

Registration of both intra- and inter-patient images has been the subject of extensive study in medical imaging literatures. There are various registration algorithms which have been classified in many different ways, based on image dimensionality, registration basis, geometric or intensity transformation, degree of interaction, optimisation procedure, modalities, etc. Detailed descriptions of each category and corresponding theoretical background can be found in [34, 75, 152, 156]. Appendix A describes some of the algorithms available online [3, 80, 185].

2.2.1 Components of registration algorithms

Each registration algorithm has four components [57]:

- source and target datasets;
- a similarity measure of how well the images match;
- a transformation model that defines how one image can be deformed to match another;

- an optimisation process that delivers an optimal parameter set that maximises the matching criteria.

Source and target datasets might be represented by the raw intensities, curves and surfaces, landmarks, feature images, or a combination of the above.

Similarity measures

Principally, registration techniques can be divided into landmark and intensity based schemes. The former are based on identification of corresponding landmarks in two images or volumes, which could be anatomical features presented in both images, pins or markers fixed or inserted into the patient and visible in every image, or a set of distinctive and easy detectable objects, such as lines, curves, points, line intersections, boundaries, etc.

Firstly, landmark based algorithms attempt to extract the features from images and then to compute a transformation based on the correspondences between these features. The presence of invasive markers permits an accurate registration; however, the procedure might cause discomfort, and there is a small risk of infection or tissue damage.

Therefore, registration is often performed using a set of features suitable for image matching [74, 145, 220]. The features extracted from the source and reference images are individually compared, aiming to choose the best match [145, 216].

Landmark matching is relatively fast to compute and a large number of algorithms for various applications such as MRI and CT brain studies [231], vascular and spine images [10] has been developed. Variations in shape and intensity of objects in DCE-MRI data complicate extraction of well matched landmarks and so DCE-MRI registration with landmark algorithms is not popular.

When images are not rich in well distinguishable details, intensity or area based registration methods are applied [110, 161, 199]. These algorithms use all (or a large proportion of) the data in each image, define a measure of intensity similarity between the images, iteratively optimise the parameters describing the orientation of the data, generate a solution, align source to target accordingly, and evaluate the transformation using various similarity measures. The process is repeated until the accuracy threshold is reached or the similarity measure is maximised /minimised [47, 93, 113, 193].

The choice of the similarity measure depends on the scanning modality as well as the geometric and intensity differences between the images. Algorithms used for intermodality registration have to be insensitive to the tissue intensity differences introduced by different modalities. One of the first algorithms for PET-MRI data registration pro-

posed by Woods et al. [266] was based on the assumption that, at each intensity in the MR image, the range of the corresponding PET intensities is small [93]. CT-MRI alignment was first performed with an algorithm [70], where intensities of MR images were re-mapped or transformed in such a way that soft tissue in both the MR and CT images appeared bright. Later, more general measures based on the joint intensity histogram and mutual information were proposed [93, 205].

Registration of multiple images of the same patient acquired using the same modality is often performed using a correlation coefficient as a measure of similarity. One image is moved with respect to another until the strongest relationship between the intensities in one image and the intensities in the corresponding location in another image is found (in this thesis we aim to find a transformation that delivers the largest correlation coefficient).

Another measure, similar to the correlation coefficient, is the sum of squared intensity differences. Here, alignment is adjusted until the smallest sum of squared differences is found. Alternatively, we can define the ratio between two images – this alignment is the basis for a ratio image uniformity algorithm, where the variance of this ratio is calculated [93].

Performance of all these algorithms is similar, except when the underlying assumption about the linear relationship between the intensities in both images is violated due to changes in local image brightness, shading, etc. For this reason it is sometimes desirable to pre-segment parts of the image prior to registration or to exclude those which change intensity dramatically [75]. An alternative approach is to account for intensity and brightness variations explicitly [168].

Transformation model

The process of registration is determined by a transformation model, which characterises the type and number of possible deformations (degree of freedom, DOF) and defines how one image can be deformed to match another [103].

Normally, rigid, affine, piecewise affine, or non-rigid transformation models are considered. 3D rigid transformation is composed of three rotations and three translations; it is a linear transformation, normally used for within-subject registration.

An affine transformation can be thought of as a crude approximation to a fully non-rigid transformation. It is defined by 12 DOF (3 rotation, 3 translations, 3 stretches and 3 shears) and is used for within-subject registration when there is global gross-overall distortion (e.g. MR to CT registration). It is a linear transformation, that allows compensating for a combination rigid motion, scaling, and a skew about 2 or 3 axes.

Piecewise or local affine models are used when different parts of an image require

individual affine transforms. They are suitable for modelling of local tissue motion and can be thought of as a simple extension to fully non-rigid transformation.

Non-rigid or elastic registration is defined by a non-linear transformation with a larger DOF. It can describe various motions and is normally used for inter-subject registration and distortion correction. However, too much flexibility in the transformation can lead to undesirable results. Often the degree of non-rigidity need to be explicitly controlled by use of smoothness constraints (e.g. bending energy or strain energy) or a limited DOF (e.g. tensor splines).

The transformations could be global (applied to an entire image) or local (applied to image subsections). Some problems that are intrinsically locally rigid (such as registration of individual vertebrae from the images of a spinal column [171]) are often solved by splitting the image into sub-images meeting local rigid body constraints.

Several authors have shown the superiority of locally sensitive registration methods over global ones [76, 89]. Local transformations are normally combined with a global smoothness constraint [168, 186] imposed on transformation parameters. In this case continuity of the transformations is assumed for an entire image.

Pyramid approaches [2] for supporting scaled image analysis are often applied along with these transforms. An image pyramid consists of a sequence of copies of the original image in which both sample density and resolution are decreased in regular steps. These copies are obtained by convolving the original image with Gaussian or Laplacian kernels. This approach allows isolating critical components of the image so that they can be easily accessible to analysis [2], thereby recovering a larger range of distortions.

Optimisation techniques

Optimisation is a process of minimising or maximising a similarity measure calculated for source and target images. The majority of registration algorithms based on an intensity matching paradigm require an iterative approach to the optimisation problem. The registration starts from an initial position (that can be manually or automatically defined) and proceeds by calculating a series of approximate solutions aiming to increase the image similarity.

In each iteration, the current estimate of the transformation is used to re-calculate the similarity measure. The optimisation technique then makes another estimate of the transformation until the desired similarity threshold is reached. There are several popular approaches that are widely used for MR image registration [152, 194]:

Gradient descent is a straightforward fast optimisation method, however it might con-

verge to local minima.

Conjugate gradient descent is an iterative method, that is based on the assumption that second derivatives of the function to be minimised exist. The method chooses successive descent directions such that it is guaranteed to reach the minimum in a finite number of steps. In theory it is better than Gradient Descent. However, second derivatives can be difficult or impossible to estimate for real images.

Annealing algorithms such as simulated annealing [126] and deterministic annealing [208] are widely used for clustering, compression, classification, and regression, because they assist in avoiding local optima, are applicable to many different structures and architectures, and are relatively fast [126, 208].

Genetic algorithms maintain a pool of solutions rather than just one. New candidate solutions maybe be generated by 'mutation' or 'combination' of two solutions from the pool. Probabilistic criteria, similar to those used in simulated annealing, are used to select the candidates for mutation or combination, and for discarding excess solutions from the pool.

The Levenberg - Marquardt algorithm is an iterative technique that finds a local minimum of a function that is expressed as the sum of squares of nonlinear functions. It has become a standard technique for nonlinear least-squares problems and can be thought of as a combination of steepest descent and the Gauss-Newton method. When the current solution is far from the correct one, the algorithm behaves like a steepest descent method: slow, but guaranteed to converge. When the current solution is close to the correct solution, it becomes a Gauss-Newton method [138, 153].

The Newton method is a root-finding algorithm that uses the first few terms of the Taylor series of a function in the vicinity of a suspected root [67]. Newton-Raphson and Quasi-Newton methods are efficient iterative techniques used to optimise Newton method based minimisation.

2.2.2 Resampling and interpolation theory

To complete a registration process, we need to warp a source into a target. However, application of the estimated registration map is unlikely to result in a neat alignment of source and target images. Various interpolation methods are used to compute the exact intensity values in the transformed image [124]. Interpolation is a mathematical method

of creating missing data. For images presented as a regular discrete grid, the interpolation of point (x,y) is:

$$g(x,y) = \sum_{i=-N}^{N} \sum_{j=-N}^{N} f(x_i,y_j)h(x-x_i)h(y-y_j), \tag{2.1}$$

where $f(x,y)$ is the image intensity at the position (x,y) and h is the weighting function applied to N neighbouring current point (x,y) samples. Interpolation of 3D images is a straightforward extension of the 2D case, assuming that image pixels or voxels are equally spaced along each direction.

Interpolation methods differ in the choice of weighting function and size of the neighbourhood. The most popular techniques include nearest neighbour, linear, B-spline, cubic, truncated sinc, and windowed sinc.

The simplest is a nearest neighbour based interpolation, in which the intensity value closest to x is assigned to the transformed pixel. This method is computationally effective; however, it produces positional errors of at most half a pixel. The method has been found unsuitable for interpolation on images acquired by various scanning modalities [114].

Linear or bilinear interpolation assumes that the intensity function is linear in the neighbourhood local to a current point. This approach allows reducing the positioning error; however, the resulting image is blurred by averaging the neighbouring pixels.

Higher order interpolation methods, such as cubic interpolation, allow improving the quality of resampling [123, 201]. With cubic interpolation we overcome problems associated with the nearest neighbour and linear interpolation techniques [165], but it is still faster than other more complex interpolators. A number of authors confirm that cubic interpolation is sufficient for MRI data resampling [94, 108].

2.2.3 Registration of DCE-MRI data

In DCE-MRI studies acquired from RA patients, pixels corresponding to the abnormal tissue change their intensity values over time. It is assumed that aligned images/volumes within the slice/study are identical except for these local intensity changes [95]. This permits assessment of the tissue vascularity as the intensity change reflects disease progression or response to treatment.

Accurate location of the enhancing tissues is crucial for efficient diagnosis and disease treatment. However, the patient's motion might introduce various artefacts, causing erroneous enhancement of some tissues, and thereby preventing efficient data assessment.

Registration is used to compensate for enhancement attributable to the patient's mo-

tion. It is a challenging task [36] and the majority of available algorithms (Appendix A), are not designed to deal with local contrast and intensity changes, fail to perform on DCE-MRI data.

Early methods for DCE-MRI data registration [131, 279] described motion using rigid transformations. These algorithms calculate the ratio of one image to the other on a pixel-by-pixel basis, and then iteratively deform images aiming to minimise variance of this ratio. This approach is not efficient for motion correction in the soft tissues surrounding rigid structures.

Recently, motion in DCE-MRI data was modelled using optical flow and affine transformation [58, 60, 81, 207]. However, standard optical flow techniques rely on the assumption that the intensity of target and source images remains constant [15, 117], while the position of objects in the source image change. This assumption is inapplicable in dynamic contrast-enhanced MRI datasets.

Contrast enhancement introduces new information into images of the dynamic series, so registration cost functions that depend on information content are confounded by the appearance of changes both in contrast and shape of tissues, leading to erroneous results. Local intensity changes in areas affected by the disease do not allow us to apply standard registration schemes such as [80, 98]. Therefore, there is a need for a new way to deal with the contrast and brightness variations.

An algorithm presented in [186] was designed to deal with partly occluded or lost data by co-registering it to atlases. It is an intensity based algorithm, where the contrast and brightness variations are explicitly modelled using a technique suggested in [168]. Geometric transformations are described with the affine model; the minimum square error is used as a cost function. Both intensity and geometric parameters are estimated simultaneously for each pixel location, and a global smoothness constraint [18, 190] is imposed on the geometric, contrast and brightness parameters. The entire procedure is built upon a differential multiscale framework [169] that permits capturing both large- and small-scale transformations.

2.3 Medical image segmentation techniques

Image segmentation is an important pre-processing step that has been studied by many researchers [82, 115, 229]. The aim of a segmentation algorithm is to split an image into non-overlapping constituent regions, which collectively represent the entire image, or to extract regions of interest, each having consistent properties [96, 179].

In medical applications, image segmentation is used to classify different anatomy fea-

tures, such as bones, muscle, blood vessels, soft tissue, etc. from the background and from each other. It is also used for identification of the anatomical areas of interest or as a preprocessing step for data analysis.

In treatment and diagnosis of multiple sclerosis, cancer and RA studies, segmentation of regions of interest is used for tumours and lesion size estimation, calculation of thickness of the cartilage, and for visualisation for surgical planning and simulation. Image-guided surgery is an important application of segmentation. Recent advances in technology have made it possible to acquire images of a patient while surgery is taking place. The goal is then to segment relevant regions of interest and overlay them on an image of the patient to guide the surgeon.

Some registration algorithms, e.g. nonlinear warps, perform differently when operating on segmented / raw data. For instance, even within the same subject, the brain can move slightly within the skull over time; registering the skull may hinder registration of the brain, so segmentation is used as a pre-processing step.

DCE-MRI datasets are often processed on voxel-by-voxel basis. Some tissues (such as bone marrow and trabecular bone) do not provide valuable information in MR scans and therefore need to be excluded from the analysis.

Poor resolution, partial volume effects and intensity inhomogeneity typical of DCE-MRI data complicate their segmentation [189]. A signal measured for each voxel arises from the entire tissue in a small, three-dimensional cuboid, and therefore might represent a mixture of different tissue types. This makes the boundaries of the regions of interest blurred and ambiguous. Various artefacts which occur because of intensity inhomogeneity significantly degrade performance of segmentation algorithms, as they often assume that the intensities of a particular tissue are constant over the image [30, 226]. To address these problems, many segmentation algorithms employ 'soft' clustering [188, 258] or a probabilistic model for tissue segmentation [12, 238, 265], where rather than making a binary decision on whether or not a pixel belongs to an object, authors allow for uncertainty in the location of the boundaries [56, 265].

Snakes [122], Active Shape Models (ASMs), and Active Appearance Models (AAMs) [53] are often used to segment organs and tissues with a low degree of shape variation. In ASMs, a statistical model of object shape and shape variation is derived from a training set, where each sample is described by a shape vector containing the coordinates of landmark points that correspond between shapes. The shape model is then used to generate new shapes, similar to those found in the training set, which are fitted to the data. Fitting the model to an image from the dataset involves finding model parameters which optimise some matching criterion between an image and a synthesised model example and requires

a measure of probability that an image point belongs to the boundary [53]. The AAM is a generalisation of the ASM approach that contains a statistical model of the shape and grey level appearance of the object of interest.

It has been demonstrated that these approaches can be successfully applied to segmentation of objects with fairly consistent shape and grey level appearance [17, 206]. In some applications, however, the statistical shape model may be too restrictive if the training set is limited; the grey level appearance model often does not deal effectively with the highly varying background seen in DCE-MRI datasets [204].

Approaches for segmentation of the soft tissue, where the degree of shape variation is high, usually rely on evidence in the image such as grey value and gradient information. For instance, level set [172] or graph cut [29] approaches are often applied to medical image segmentation. The level set technique is formulated as infinite-dimensional optimisation on a spatially continuous image domain [28], and is based on the idea of front propagation [68]. Graph cut is defined as minimal cuts of a discrete graph representing the pixels of an image [28].

Some algorithms operate solely on intensity values [75, 87, 229, 236], others involve spatial information [228,268] or ground truth (GT) knowledge about the shape or intensity of regions of interest [19]. Often these approaches assume that the tissue intensity and/or location and shape of regions of interest across the data are constant. Such assumptions fail with dynamic data, complicated by the effect of contrast agent and high inter-patient variability.

Often techniques incorporate manual estimation of initial parameters or prior knowledge about the anatomy or image properties [182, 230]. This can significantly improve segmentation results; however manual tuning of the algorithm makes this solutions infeasible for analysing large amounts of data [6, 202]. There are surveys that attempt to classify segmentation algorithms based on the techniques they incorporate [75, 265].

Fully automated segmentation is a challenging task [51] and there is no general application independent solution which does not require supplementary knowledge about the image class, scanning modality or properties of regions of interest [82, 189, 230, 236, 277]. Here we provide a brief overview of two popular techniques, which are often used for DCE-MRI data segmentation.

2.3.1 Snakes

The active contour model or snake is defined as an energy-minimising spline, where the snake energy depends on its shape and location within an image [229]. This approach for

image segmentation was first introduced by Kass et al. [122] and is still widely used for various image analysis and understanding tasks.

It can be thought of as a technique for matching a deformable model to an image by means of energy minimisation. Local minima of this energy then correspond to desired image properties.

The active contour can be parametrically defined as $\mathbf{s}(p) = [x(p), y(p)]$, where $x(p)$ and $y(p)$ are x,y coordinates along the contour and $p \in [0,1]$. The energy function is a combination of internal and external forces [122]:

$$\begin{aligned} E^*_{snake} &= \int_{\mathbf{s}(p)} E_{snake}(\mathbf{s}(p))dp = \\ &= \int_{\mathbf{s}(p)} \Big\{ E_{int}(\mathbf{s}(p)) + E_{image}(\mathbf{s}(p)) + E_{contraints}(\mathbf{s}(p)) \Big\} dp \end{aligned} \tag{2.2}$$

In Equation 2.2 E_{int} defines the internal forces, which represents the internal energy of the spline due to bending:

$$E_{int} = \alpha(p)\left|\frac{d\mathbf{s}}{dp}\right|^2 + \beta(p)\left|\frac{d^2\mathbf{s}}{dp^2}\right|^2, \tag{2.3}$$

where $\alpha(p)$ and $\beta(p)$ specify the elasticity and stiffness of the snake. The external energy E_{image} defines the external forces that come from the image over which the snake lies; is minimsed at the features of interest, such as boundaries. The last term $E_{contraints}$ reflects external constraints imposed, for example, by a user. If the snake is near some desirable feature, the rest of the snake can be adjusted to account for it. If the snake converged to a local minimum that was defined as incorrect, this term can force the snake away to a different minimum.

From the calculus of variations, a snake that minimises E^*_{snake} must satisfy the Euler equation, where ∇ is the gradient operator:

$$\alpha\mathbf{s}''(p) - \beta\mathbf{s}''''(p) - \nabla E_{ext} = 0 \tag{2.4}$$

This can be viewed as the force balance equation:

$$F_{int} + F_{ext} = 0, \tag{2.5}$$

where internal force $F_{int} = \alpha\mathbf{s}''(p) - \beta\mathbf{s}''''(p)$ discourages stretching and bending and

external force $F_{ext} = -\nabla E_{ext}$ pushes the snake towards the desired edges.

There are various approaches available for minimisation of Equation 2.4 [55, 122]. Generally, the snake is made dynamic by treating $\mathbf{s}$ as function of time t as well as p, i.e. $\mathbf{s}(p,t)$. Then, the partial derivative of $\mathbf{s}$ is set with respect to t as in Equation 2.6:

$$\mathbf{s}_t(p,t) = \alpha\mathbf{s}''(p,t) - \beta\mathbf{s}''''(p,t) - \nabla E_{ext} \tag{2.6}$$

A numerical solution to Equation 2.6 is achieved by discretising the equation and solving the system iteratively. Most snake implementations use either a parameter which multiplies $\mathbf{s}_t$ in order to control the temporal step-size, or a parameter to multiply ∇E_{ext}, which permits separate control of the external force strength.

For our experiments we used a publicly available implementation provided by Xu and Prince [270], which is based on defining the external energy via a gradient vector flow. Even though this model allows for relatively flexible initialisation of the snake, it still requires the initial contour to be in the vicinity of a correct solution. Clearly, it is not trivial to find initialisation optimal for pre- and post-contrast DCE-MRI data with high inter-patient variability, presence of high ambiguous boundaries and intensity artefacts.

2.3.2 Region growing

Region growing is a technique that classifies pixels into regions based on pre-defined criteria [88] and might be useful for delineation of tumours and lesions [85, 191]. Region growing is often used in every day medical practice for segmentation of various tissues and semi-automatic identification of regions of interest [195].

It starts with definition of 'seed' points, from which the regions are grown by appending to each seed those neighbouring pixels that satisfy a predefined criterion. This could be the grey level or colour value of the pixel.

Location of the seed points and growth threshold (a similarity measure between the seed and neighbouring pixels) are based on the nature of the application and image data under consideration.

Descriptors such as colour, intensity, and texture are local and often do not account for the history of the 'growth'. Therefore, many authors utilise information about the size, position, and shape of the object of interest based on prior information.

Region growing tests the statistics inside the region; however it is sensitive to typical DCE-MRI data noise and contrast variations, which cause resulting regions to have irregular boundaries and small holes.

Figure 2.1 illustrates regions corresponding to the bone interiors taken from four dif-

ferent pre-contrast images. It is clearly visible that the contrast and intensity of the bone marrow and surrounding tissues varies significantly from study to study. The 4^{th} region in the figure exhibits oedema, which is causing a part of the bone to erode and appear black. Region growing with a seed inserted in the middle of the bone region would not deliver accurate results.

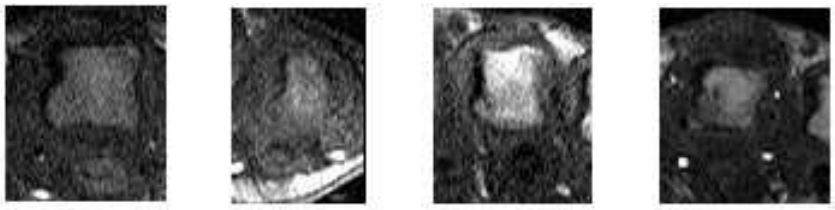


Figure 2.1: Regions corresponding to the bone interiors (brighter colours) and surrounding tissues from different DCE-MRI studies. Contrast and intensity of the bone marrow and surrounding tissues varies significantly from study to study.

Therefore, in the presence of artefacts, ambiguous boundaries and intensity changes, snakes and region growing often do not deliver an optimal solution. Besides, it might be time-consuming to find an initialisation that allows snakes or region growing to perform successfully on datasets acquired from various patients. Therefore, we seek an approach that would be robust to the contrast / intensity changes and user independent of the parameter selection problem.

2.4 Evaluation techniques

Algorithm evaluation is an important step towards establishing its adequacy for a particular application or its general efficiency [5, 78, 263]. Different segmentation evaluation metrics [45, 46, 66, 271, 276] have been proposed. They are used for quality assessment of segmentation results, as well as algorithm comparison. Evaluation of different algorithms on the same dataset allows choosing the most efficient one for a particular application. Evaluation of a single algorithm on different datasets provides information about its robustness, and ability to handle data acquired under different conditions and by different modalities.

Evaluation metrics can be divided into two groups: supervised and unsupervised. Supervised approaches [46, 66] are based on computing a dissimilarity measure between the results of machine segmentation and known correct information: the Ground Truth (GT). Unsupervised evaluation [45, 276] assesses the quality of segmentation by considering

different statistics derived from the properties of an image and segmentation, without knowledge of GT. Supervised evaluation is useful in method verification; unsupervised – in actual clinical practice, where volume of data is too high to supervise.

Existing evaluation metrics are usually demonstrated on synthetic datasets, rarely agree with each other, and put serious constraints on image properties. These constraints are not generally valid for DCE-MRI imagery, which is complicated by low contrast and intensity, local blur, patient movement artefacts, and the presence of ambiguous boundaries. Therefore, there is a need for new supervised and unsupervised approaches suitable for DCE-MRI data.

2.4.1 Supervised evaluation techniques

Earlier performance of the majority of segmentation algorithms [104, 105, 224] has been quantified based on opinions of human observers. This precludes analysis of large numbers of images, as well as meaningful comparison of slightly different results [115, 271]. Moreover, variation in evaluation results produced by different experts can be significant. If image quality is poor, the boundaries of the region of interest are ambiguous and observers' judgements on the expected segmentation might be uncertain.

An example of such a boundary is shown in Figure 2.2; the region is partly ambiguous and overlays outlined by 2 experts independently do not coincide. To deal with this problem, the most common approach is to involve the opinion of more than one human observer, and then to 'average' their judgements [46, 263].

Two procedures to evaluate an average curve, given two or more curves, are described in [46] and [26]. The former is based on establishing one-to-one correspondence between the points in the curves outlined by multiple observers using a modification of the methods for shape registration. The later is based on a shape-based interpolation method. Details can be found in the references; for our experiments the former procedure was used.

Alternatively, evaluation can be performed using statistical supervised metrics, which assess the quality of segmentation by using GT information about regions. GT is usually outlined manually by an expert before segmentation results are available; the performance of segmentation algorithms can then be assessed according to measured discrepancy. Such evaluation methods are also known as 'empirical goodness methods' [276].

However, the location of GT is usually dependent on expert opinion, and therefore subjective. Chalana and Kim [5, 46] attempted to solve this problem by computing a 'percentage statistic', which considers whether computer generated boundaries differ from GT outlined by multiple experts as much as GT boundaries differ from each other. The

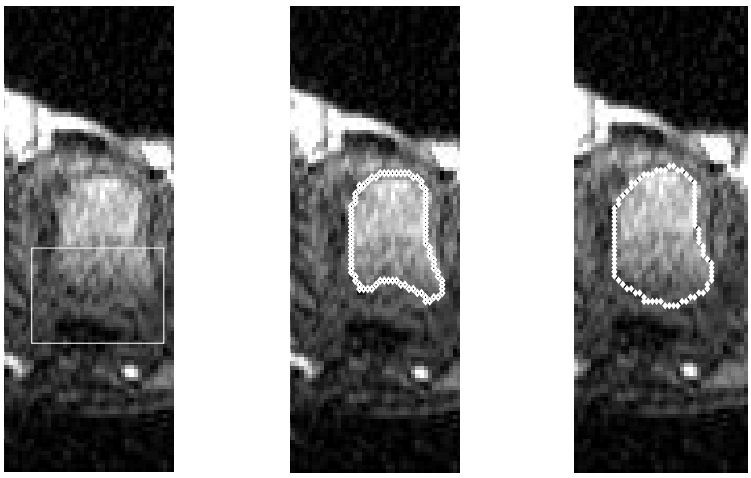


Figure 2.2: A region from a DCE-MRI study of the MCPJs with a partly ambiguous boundary (left), and two possible GT overlays, highlighted by different human experts (middle, right).

authors state that their method outperformed the technique suggested by Williams [263], where the opinion of an observer is compared with joint agreement of other observers. These results have been analytically and empirically analysed in [5] and found inconsistent.

The problem of GT inconsistency remains open, and is receiving attention in many areas of imaging and computer vision [64]. The need for an objective metric, which performs evaluation based on GT, but at the same time accounts for possible inconsistencies, has been identified by many researchers. A solution to this problem is to allow an error measure, yielding acceptable differences between the GT and the segmentation results. Such a solution can be adjusted for possible observer error or poor image quality.

All supervised evaluation techniques which allow acceptable segmentation error are based on one of two approaches [16, 276]: misclassified area [66] or a measure of boundary displacement [271]. Either can define an appropriate metric for assessing segmentation quality and efficiency.

2.4.1.1 The mutual overlap area based approach, and its limitations

The Mutual Overlap (MO) approach, also known as a Dice evaluation metric of a spatial overlap, is based on computing the area of overlap between a GT overlay and a segmented region [27, 66, 115]. This is illustrated in Figure 2.3.

The area is normalised to the total area of the two defining regions; if A_1 is the area of the segmented region, A_2 is the area of the GT region, and MO is the area of their mutual

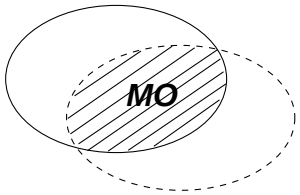


Figure 2.3: Mutual overlap: machine segmented region is shown in white, GT in red.

overlap, then the mutual overlap metric is defined as:

$$M_{MO} = \frac{2 \cdot MO}{A_1 + A_2}. \tag{2.7}$$

It is customary to measure acceptable quality with respect to this metric by setting a percentage threshold for M_{MO}, usually greater than 50% [27], but this will vary to reflect the strictness of the definition.

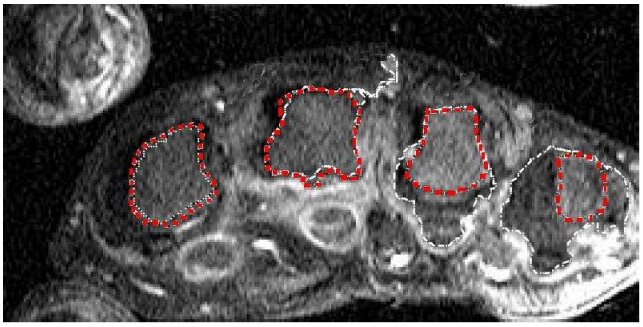


Figure 2.4: Four regions corresponding to bone interiors segmented in a sample DCE-MRI image. Machine segmented regions are shown in white, GT in red.

This approach is popular and seen to work well on, for example, binary, RGB or some satellite data, but its performance on DCE-MRI data is not always adequate:

- A DCE-MRI image segmented by a region growing algorithm [87] is shown in Figure 2.4; machine results are shown in white and corresponding GT in red. The M_{MO} measure of these 4 regions is 81%, 74%, 53%, and 11% respectively. With a threshold value of 70%, the first and the second regions are considered 'correct', whereas in fact the second regions does not reflect the bone's properties, which, from a clinician's viewpoint, makes this sort of segmentation of little value.

- Often images are corrupted by noise or complicated by patient movement – then boundaries become ambiguous and the location of GT is not obvious even for ex-

perts. Assessment of segmentation results should reflect acceptable error accounting for the extent of the ambiguous sections, or assigned by an expert. Current uses of this metric rarely if ever compensate for such factors [27].

- Experiments with DCE-MRI datasets show that boundaries of machine segmented regions may partly overlap GT, be located strictly inside or outside GT overlays, or include very local significant deviations (for example, the second region in Figure 2.4).

  M_{MO} considers the area of regions. Therefore, long thin 'tails' attached to the regions will not influence the metric performance. Therefore, M_{MO} is of most value only when the distances from GT to the segmentation output are unimodal with low variance, and so do not include very local significant deviations such as in the boundary of the 2nd region in Figure 2.4 (which will be referred to as 'tail'). If the boundary pixels form a tail the evaluation can be confusing – because the tail area is small, M_{MO}, which considers area inside the regions boundary, is still high.

- The metric cannot be applied to non-closed boundaries. Segmentation or boundary detection algorithms applied to regions of varying intensity and texture often result in partial boundaries, or they recognise interior areas of regions as boundaries. An example is shown in Figure 2.5. For such segmentations, GT may be expected to be a closed boundary, but segmentation may deliver open or closed curves. For such images, M_{MO} is of little use. Even if an open boundary has been closed by some approach, boundary confidence while passing through low contrast areas will be questionable. In this situation the metric will be affected by the quality of the image as much as by the quality of segmentation.

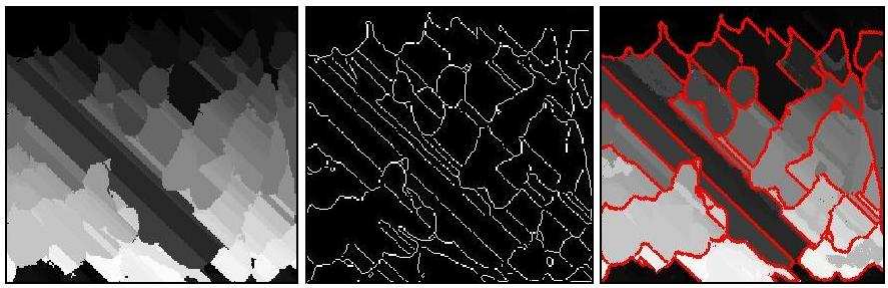


Figure 2.5: Left: An image [taken with permission from [254]] segmented by the Canny edge detector [39]. Middle: Results of a boundary detection algorithm. Right: GT (in red).

Despite these drawbacks, this metric is widely used for evaluation of segmentation algorithms executed on medical imagery [27, 38, 48, 192].

2.4.1.2 Hausdorff distance based approach

In 1977 an approach considering deviation of boundaries from GT was proposed [271]. In these experiments the M_{MO} metric [66] (known at that time) was shown to be inadequate since it often departs from the human judgement and does not reflect the spatial information inherent in pixel misclassifications. An alternative metric, which was based on computing the Euclidean distance between all segmented and GT pixels inside the regions' boundaries, was proposed. This is related to the Hausdorff measure between the sets [209].

The *Hausdorff distance* (HD) between two sets A and B is computed by finding the minimum distance from each element of one to some element of the other, and then finding the maximum such distance [209];

$$h(A,B) = \max_{a \in A} \left\{ \min_{b \in B} d(a,b) \right\} \tag{2.8}$$

where $d(a,b)$ is some suitable distance metric, commonly the Euclidean distance between a and b. The Hausdorff distance is oriented (asymmetric); usually $h(A,B) \neq h(B,A)$. A general definition of the Hausdorff distance between two sets is [209]:

$$H(A,B) = \max \left\{ h(A,B), h(B,A) \right\} \tag{2.9}$$

This defines a measure of the sets' mutual proximity, indicating how far (at most) two sets of points are from each other. $H(A,B)$ has been adapted as an evaluation measure for the quality of the segmentation results, where A is a GT region and B a machine segmented region, interpreting the sets A, B as the interiors of the two regions [46].

A debate persists on which evaluation metric should be preferred [43, 115, 217, 276]; few authors evaluate algorithms using both. A comparative study of the supervised evaluation metrics [43] favours M_{MO} and demonstrates that the *HD* based metric [271] does not provide adequate evaluation of segmentation in the case of under-segmented regions, favours over-segmented regions, and produces measurements that are highly dependent on confidence in GT. Conversely, it has been shown that evaluation provided by M_{MO} does not correspond to human observers' opinions, and a HD based metric is preferable [217].

2.4.2 Unsupervised evaluation techniques

Difficulties in acquiring GT such as imprecise definition, paucity of information, and time consumption have motivated research into unsupervised evaluation [41, 45, 276], but most of this work has been demonstrated on binary or synthetic images, with limited experiments on MRI, CT and imagery acquired by other scanning modalities. We are unaware of unsupervised evaluation metrics being successfully employed in domains with such characteristics.

Generally unsupervised evaluation metrics are based on the location, shape, size, contrast, or intensity of segmented regions [187]. Unsupervised metrics depend on either global image statistics, which can be derived from the set of all pixels in the image, or on regional statistics, which characterise the regions segmented.

Recent research on unsupervised evaluation methods [42] has shown that among the best known techniques [276] that require no human intervention, there are several which produce relatively consistent results from tests run on a range of imagery.

2.4.2.1 Existing unsupervised approaches

Intra-region uniformity criterion

A criterion proposed by Weszka and Rosenfeld [259] and Levine and Nazif [139] assumes that an adequate segmentation algorithm produces images with high intra-region uniformity. For untextured images, uniformity can be computed on the basis of intensity variance evaluated at every pixel of the foreground [139]. Let A_I denote the area of the segmented image I, A_i the area of the foreground region R_i, and f intensity of a pixel; then σ^2 represents the intensity variance of the whole image, and σ_i^2 that of the foreground [219]:

$$\sigma_i^2 = \sum_i \sum_{(x,y)\in R_i} \left[f(x,y) - \frac{1}{A_i} \sum_{(x,y)\in R_i} f(x,y) \right]^2 \tag{2.10}$$

$$\sigma^2 = \sum_{(x,y)\in I} \left[f(x,y) - \frac{1}{A_I} \sum_{(x,y)\in I} f(x,y) \right]^2 \tag{2.11}$$

In [139] the uniformity measure is given as:

$$\sigma_i^2 = \sum_i \sum_{(x,y)\in R_i} \left[f(x,y) - \frac{1}{A_i} \sum_{(x,y)\in R_i} f(x,y) \right]^2 \tag{2.12}$$

Sahoo [211] proposed normalising this measure as in Equation 2.13, where C_i is a normalising factor.

$$\overline{\sigma_i^2} = 1 - \sigma_i^2 / C_i \tag{2.13}$$

In [219] area of an image and area of a region are taken as normalising factors and the uniformity measure is given as in Equation 2.14:

$$M_1 = \frac{A_i}{A_I} \frac{\sigma_i^2}{\sigma^2} \tag{2.14}$$

Inter-region contrast based criterion

Later Levine and Nazif [139, 170, 176] proposed an evaluation measure known as the inter-region uniformity criterion. It is assumed that good segmentation produces regions of uniform intensity with high contrast along borders. If a grey scale image contains a region R_i with average intensity level f_0 and average intensity of the local background f_b, the measure is defined as the following:

$$M_2 = \frac{|f_0 - f_b|}{f_0 + f_b} \tag{2.15}$$

Zeboudj contrast

The Zeboudj criterion [42, 274] is based on the difference in the internal and external contrast of a region segmented. It is assumed that an adequate segmentation should split an image into regions of high contrast. This criterion considers separately pixels on the region border and within its interior; the contrast of the pixels on the border of a correctly segmented region is expected to be significantly different from the contrast of the pixels inside the region. The Zeboudj measure is based on the combined principles of maximum inter-region and minimal intra-region disparity measured in the pixel's neighbourhood.

The disparity between two pixels with grey levels $f(s)$ and $f(t)$, can be defined by Equation 2.16, where G is the maximum of the grey levels:

$$c(s,t) = |f(s) - f(t)| / (G-1) \tag{2.16}$$

The intra-region disparity of the region R_i with area A_i is given by Equation 2.17, where $W(s)$ is the neighbourhood of the region's pixels.

$$I(R_i) = \frac{1}{A_i} \sum_{s \in R_i} \max \left\{ c(s,t),\ t \in W(s) \cap R_i \right\} \tag{2.17}$$

External disparity of the region is given by Equation 2.18, where F_i is the region boundary of the length l_i.

$$E(R_i) = \frac{1}{l_i} \sum_{s \in F_i} \max \left\{ c(s,t),\ t \in W(s), t \notin R_i \right\} \tag{2.18}$$

Then, the disparity of the region R_i can be defined by Equation 2.19, where the measurement $C_{R_i} \in [0;1]$. A schematic diagram is illustrated in Figure 2.6.

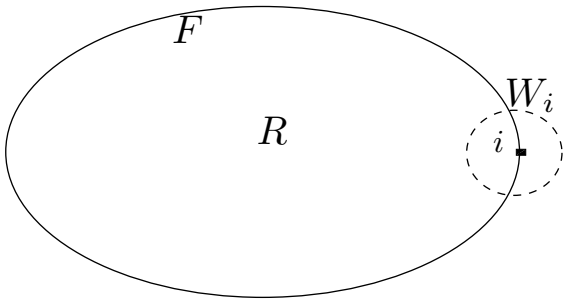


Figure 2.6: A region R with boundary F; a boundary pixel i and its neighbourhood W_i.

$$C_{R_i} = \begin{cases} 1 - I_i/E_i & \text{if } 0 < I(R_i) < E(R_i) \\ E_i & \text{if } I(R_i) = 0 \\ 0 & otherwise \end{cases} \tag{2.19}$$

Finally, the Zeboudj criterion is given by Equation 2.20, where A_I is the area of the image, and A_i is the area of the i^{th} region:

$$M_3 = \frac{1}{A_I} \sum_i A_i \times C_{R_i} \tag{2.20}$$

Classifier based metrics

Since different evaluation metrics make their judgements in different ways, they give diverse results on the same segmentation output. This provides a number of evaluation references, which may be combined to produce a single reference [45,275]; schemes such as support vector machines [45] and Bayesian classifiers [275] have been used.

An algorithm presented in [45] evaluates segmentation outputs by combining results of several stand-alone unsupervised metrics by means of support vector machines (SVM). A number of segmentations are evaluated by an unsupervised and a supervised (mutual overlap based) metric. This permits the assignment of a weight to each unsupervised metric associated with accuracy of its prediction; higher-weighted metrics are used to

train a SVM classifier and the metrics are combined into a subset that will be used for evaluation of a particular application. The classifier is trained to make a prediction about a new segmentation using predictions of individual unsupervised metrics as features. The authors performed a comparative study of the unsupervised metrics, and demonstrated that for a given application only a small subset of metrics produces an accurate prediction. Also, incorporating an inappropriate metric into a classifier worsens performance of the algorithm. It has been seen to perform well on synthetic images, but results are poor when it has been applied to MR and satellite imagery.

Zhang et al. [275] also suggest combining stand-alone metrics' outputs using various classifiers (simple classifier, weighted majority (WM), Bayesian, and SVM). Using various strategies the authors try to predict which of the given segmentations is better.

No learning stage is required for a Zhang's simple classifier; the decision is made alone with the predictions provided by the majority of the individual metrics. A weighted majority based approach uses an on-line learning approach and is based on combining the weighted opinions of individual metrics. Initially, each metric is given a weight of one; during the training stage, a number of segmentations are evaluated by human observers and opinions compared with the prediction given by the individual metrics. If a prediction is wrong, then the weight of that metric is decreased. Then in the evaluation stage, two segmentations are compared and the weighted majority algorithm predicts which one is better.

A further development was to combine primitive classifiers with a naïve Bayesian approach, assuming the scores of each metric are conditionally independent from each other. Results discussed in the literature [275] are the best yet seen.

These algorithms have been demonstrated on several images from the Berkeley dataset [154] (108 images were used for the training dataset and 91 for evaluation). The authors have shown that combining metrics without the learning stage produces poor results, but that classifiers trained using supervised learning techniques have produced relatively good results. However, the authors do not discuss the appropriateness of each metric to a particular application and how results would change with variation in the number of metrics, different imagery or a training dataset of different size. There were no experiments with MRI or DCE-MRI data.

Other metrics

There are metrics developed for evaluation of segmentation algorithms executed on colour images (Lui and Yang evaluation criterion [148], Borsotti criterion [24]). The Borsotti criterion has been modified to be applicable for grey scale image segmentation evaluation

[42], but performed poorly. An overview on this type of evaluation can be found in [5, 24, 42, 148].

Some other evaluation metrics (Cochran's homogeneity measurement based metric [52], Pal's intra-region uniformity based metric [178], Sahoo shape based measure [211]) require selection of a threshold, which is often done arbitrarily, thus limiting the applicability of the methods.

2.4.2.2 Issues with existing metrics

There are several problems affecting performance of the currently accepted unsupervised metrics:

- Noise and local blur may make object boundaries span several pixels [189]; thus segmentation algorithms will deliver objects with low contrast and partly ambiguous boundaries. Existing metrics [42, 139, 259, 274] usually assume that an image is segmented into regions of high contrast with well-defined boundaries.

- Interpretation and comparison of results delivered by unsupervised metrics are confusing. Some are based on contrast changes, others use pixel ratios; some are normalised to the image/region size, others not.

- Simple evaluation measures such as [139, 259, 274] operate without any user intervention, but are seen to be limited, while the more sophisticated metrics [45, 275] require a training stage where an algorithm learns from set of manually outlined GT segmentations.

  However, a simple combiner of votes does not produce adequate results [275] and some supervised learning should be incorporated. The size of the training dataset and the number of metrics included in the classifier influence performance of the algorithm. A study [45] shows that increasing / decreasing the size of a training dataset or incorporating an inappropriate metric into a classifier can worsen performance of the unsupervised approach. The choice of an optimal set of stand alone metrics, appropriate size of the training dataset, to obtain the best match image domain remain unclear.

2.5 Quantitative analysis of DCE-MRI datasets

DCE-MRI is used extensively in a wide range of applications involving different organs and pathologies [119, 132, 177, 197, 198]. It has evolved as an important method for eval-

uating various diseases of the musculoskeletal system [33, 72, 252]. This technique provides information about tissue vascularity, perfusion and capillary permeability [250] and therefore permits assessment of the degree of inflammation and post treatment progress evaluation.

After performing a dynamic study, a large number of images are available to be evaluated qualitatively and quantitatively. Evaluation of a series of images obtained with DCE-MRI can be performed in different ways.

A simple, but subjective, qualitative method is the 'naïve review method', in which an observer examines the contrast enhancement sequentially on all images of the dynamic sequence [202]. With this method detection of small areas of enhancement, or areas with discrete enhancement (especially in the wrist studies) can be difficult.

Early qualitative analysis methods were based on image subtraction, in which the first image (i.e. before contrast injection) is subtracted from all subsequent images of the dynamic study [72, 250]. The subtraction images are then viewed one by one. With such methods it is possible to detect the most highly enhanced tissues (for biopsy or injections). However, estimation of measures such as the magnitude of enhancement and time of onset of enhancement or recognition of the late enhancing tissues such as fat on the early subtracted images is difficult.

Quantitative analysis of DCE-MRI data can be performed using two fundamentally different groups of methods: pharmacokinetic [31, 134, 245] and heuristic [62, 107, 121, 130, 195]. Pharmacokinetic methods [31, 134, 245] provide a framework that can be used to link the physics of MRI signal acquisition and the underlying patho-physiology that governs contrast agent kinetics. Comparative analysis of these methods can be found in [196, 243].

Pharmacokinetic methods [31, 134, 245] rely on a common set of assumptions regarding the properties of the principal compartments and their interactions, but adopt different representation for temporal variations of the contrast agent concentration in the blood plasma. In [134] the contrast agent concentration in the blood plasma is measured in individual subjects, which makes implementation of this method in clinical settings difficult, especially when high spatial resolution and multi-slice coverage are required [196].

In [31, 245] the contrast agent concentration in the blood plasma is represented as a theoretical function in response to a chosen form of the input function (injection), which is often idealised as a delta function. Such a representation fits experimental data well when the temporal resolution of the DCE-MRI is low and acquisition time long. With higher resolution the characteristic shape of the contrast agent uptake in the tissue of interest resembles a sigmoid, which cannot be accurately described by these methods

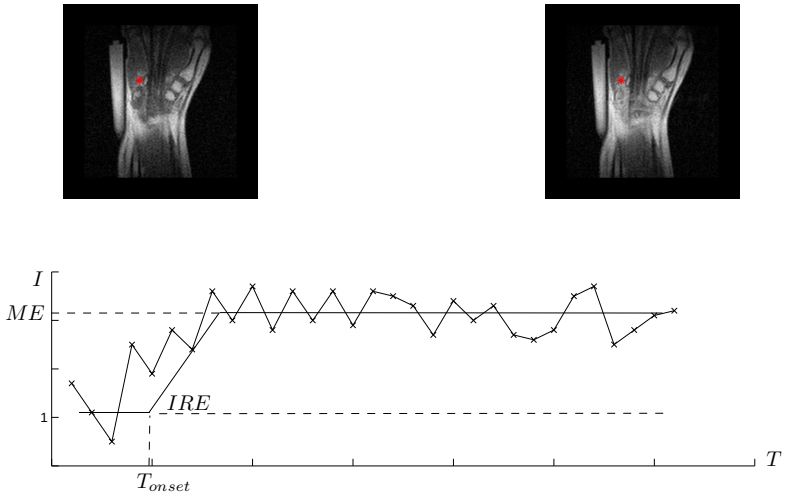


Figure 2.7: Top: Pre- and post-contrast images of the wrist. Bottom: Signal intensity (I) vs. time (T) curve approximated by a piecewise linear model. Parameters ME, IRE, and T_{onset} estimated for this curve.

[243]. Furthermore, long acquisition times incur more noise as a result of movement and provoke patient discomfort.

In clinical practice, it is impossible to assess the accuracy with which pharmacokinetic variables reflect the true underlying changes in concentration of the contrast agent [119]. The accuracy of the estimates will depend on the pharmacokinetic model used and the signal to noise ratio in any individual case. This is a particular problem with applications where noise is the dominant, or only, cause of variation of contrast agent concentration [119].

Alternatively, contrast enhancement can be quantified in terms of heuristic parameters such as maximum enhancement (ME), initial rate of enhancement(IRE), and time of onset of enhancement (T_{onset}). These heuristic parameters have been seen to correlate with pharmacokinetic measurements of inflammation [83, 128, 240]. In contrast to pharmacokinetic parameters, heuristic estimation is relatively straightforward (see Figure 2.7).

There is clear terminological confusion over the maximum rate of enhancement (MRE) and the initial rate of enhancement (IRE). In fact, what we call IRE in Figure 2.7 measures MRE, and will be equal to the initial rate of enhancement only when the wash-in phase (the intensity increase) is linear. However, due to the noise present in the data, a reliable estimate of IRE is often impossible. Therefore, we proceed with computation of MRE, which in all further experiments will be referred to as IRE in common with other literature.

Most such analysis hitherto has examined individual signal intensity curves derived from user defined regions of interest (ROI analysis) or on voxel-by-voxel basis.

The validity of ROI analysis relies on the position and size of ROI, as its misplacement might result in a $20-30\%$ difference in measurements [50, 162]. This implies poor reproducibility of the techniques which describe the shape of the enhancement curves [119].

The first attempt to perform objective voxel-by-voxel analysis was published in [195], where the enhancement curves were constructed for the all pixels in the dynamic slice. This freed the analysis from ROI placement, but the approach has its own drawbacks and limitations, listed in the next section.

2.5.1 Current analysis of DCE-MRI datasets

There are no standardised methods for heuristic analysis of the data acquired with low or high field scanners. Currently, for the metacarpophalangeal joints (MCPJs) and wrist studies dynamic curves are calculated from an approximately 2-3 mm^2 ROI positioned in the area of maximal visual enhancement [50, 195]. Measurements of *IRE* and *ME* contain both spatial and temporal information making the results vulnerable to the size and position of the ROI [162].

A semi-automated approach proposed for DCE-MRI data of the MCPJs analysis [195] was the first attempt to perform quantitative analysis objectively. It uses the commercially available software ANALYZE [202] for manual segmentation and identification of tissues of interest. Signal intensity vs. time curves (I) are normalised over a mean baseline computed from the first three values (b), and their geometrical properties such as height and slope are considered. The normalised signal is:

$$b = \sum_{t=1}^{3} \frac{I(t)}{3}, \;\; \hat{I}(t) = \frac{I(t)}{b}, \; t = 1 \ldots T \tag{2.21}$$

In Equation 2.21 T is the number of dynamic frames in the temporal slice; in the first three no enhancement is expected to be observed and therefore they are often taken as a baseline.

To assess the extent of RA various parameters, such as ME, IRE, and T_{onset} are computed from the enhancement curves. The parameters are estimated by passing an averaging window of length $n = 5$ (a number empirically found for the MCP data [195]) over the signal intensity vs. time curve and determining the gradient of the linear best fit in each window [195]. The maximal such gradient is taken as IRE, and the instant at which this occurs recorded as time of onset of enhancement T_{onset}. ME is found as a maximum of mean intensity values calculated in each window. Figure 2.8 illustrates estimation of *IRE*

with the moving-window algorithm[1].

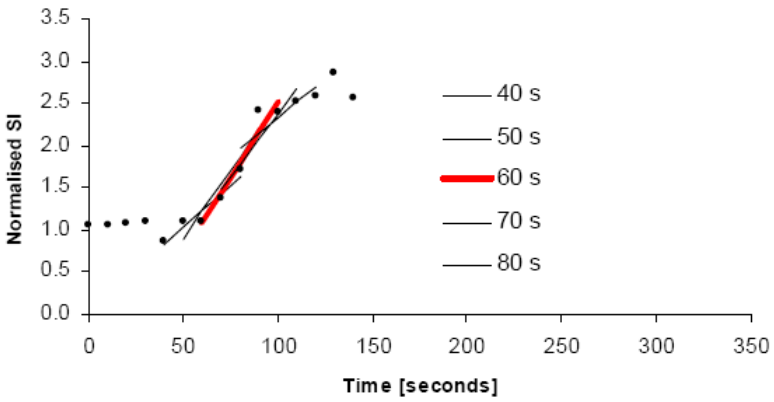


Figure 2.8: Estimation of ME, IRE, and T_{onset} with the moving-window approach.

At this stage, pixels in which signal intensity vs. time curves exhibit $T_{onset} > 60s$ or $ME < 1.2$ are regarded as unlikely to be of clinical interest since either the take-up has not been appreciable or the behaviour is outside the expected time interval [195]. This permits the measurement of N_{total} – the total number of enhancing pixels, which allows for assessment of a patient's condition.

Computation of the parameters with the moving-window will depend highly on the degree of noise present in the data, scanning equipment, and the time-course of the arrival of the contrast agent [111, 195]. The size of the moving window needs to be adjusted to process the data acquired by a particular scanner or at fixed acquisition settings (as different scanner generate data of different noise level). This implies manual intervention of the human operator, which makes the results subjective.

To visualise the extent of inflammation, parameters are presented in the form of parametric maps, which are 2D images depicting these parameters. Thus, a parametric map is a 2D representation of a chosen property of interest (e.g. ME,IRE) superimposed on the anatomy image.

The first mention of parametric maps can be dated to the late 1980s – early 1990s. They were used for DCE-MRI data analysis acquired in pre-clinical contrast agents trials. Later in the 1990s, clinical trials were performed on patients [31, 130, 134, 245]; this is when the modern definition of the maps originated [180, 235, 242, 244, 246].

The value of the parametric mapping technique lies in speed and ease of interpretation of DCE-MRI datasets and its simple display, suitable for clinical interpretation even by non-experts. The benefits of parametric mapping are obvious; however, the technique is

[1]Taken [with permission] from [195].

not widely adopted in medical practice. There are no established standards for assessment of quality of the parametric maps that might explain what degree of noise is acceptable and how well a parametric map reflects the activation events. Currently, evaluation of the parametric maps' quality is performed by experienced observers and clinical experts.

Therefore, current analysis of DCE-MRI data [50, 195] enables semi-automated computation of the parameters, but has some drawbacks and limitations. Firstly, as no preprocessing is employed, voxel-by-voxel analysis might be misleading as it is assumed that each voxel represents the same area of tissue throughout the acquisition, which is of course wrong due to the presence of patient's motion.

Secondly, more accurate estimation of the parameters is desirable. For example, estimated with the moving window method [195] T_{onset} corresponds to the time at which intensity is increasing fastest, and is clearly larger than the actual time of onset of enhancement. This estimate is, of course, trivial to improve given this time and the gradient IRE.

Furthermore, there is often a proportion of curves in which the maximal intensity has not been reached, indicating constant leakage into locally available extra-cellular space. Existing methods do not allow identification of tissues at which signal intensity did not peak during the acquisition of DCE-MRI data, which might lead to inaccurate estimation of the parameters. Such locations should be identified.

Lastly, all issues mentioned prevent accurate estimation of the total number of enhancing pixels (N_{total}), which indicates the extent of RA [180].

2.6 Segmentation of blood vessels

Due to the high vascularity that occurs in disease affected tissues, values of ME and IRE corresponding to the blood vessels and such tissues will be depicted in a very similar manner in the parametric maps. This complicates visual analysis of the data and does not permit an objective automated assessment of the inflammation. Therefore, blood vessels should be excluded.

Some algorithms attempt to determine the centre of the vessel paths and then employ various segmentation and tracking algorithms [149] or utilise prior knowledge about the segmentation task [35, 249] to reconstruct the vessels' tree structure. Other applications [146, 213] achieve sequential contour tracing by incorporating features such as central vessel point and search direction [146]. A semi-automatic graph representation approach for vessel tracking has been introduced in [136]. In [97], the authors present a coronary artery tracking system that incorporates information within subsections of an image for

stable tracking. Features such as vessel width and direction are required; also the starting point should be given by a user. Classification and comparison of various tracking based approaches can be found in [125].

Approaches discussed in [125, 159, 269] apply explicit models to extract the vasculature from DCE-MRI datasets. These include deformable models, parametric models, and template matching. Such methods generally require manual or semi-automatic initialisation based on prior information about the diameter and location of the vessels [127, 167, 210].

Several techniques have been proposed for automated initialisation of such algorithms [164,270,272] and their adaptation to a particular topology [143,160,221]. They generally require a pre-processing stage that allows generating an initial contour, which is further refined [63, 79, 218, 238].

Several authors use template matching approaches that attempt to recognise a structural model (template) in an image [20,255]. In arterial extraction applications the arterial tree is usually represented in the form of a series of nodes connected in segments. Formulation of the model often requires user interaction and prior information about the diameter and location of the vessels.

Even though template matching and tracking methods might produce good results, there is still a need for manual initialisation or selection of appropriate starting points. Alternative methods are based on tissue classification.

Commonly, tissue classification is done based on intensity values. The very first methods [242, 245] assumed that each voxel that enhances more than a certain threshold is vascular in origin. However, this approach can lead to exclusion of up to 50% of voxels from the image in enhancing tumours and other very vascular enhancing tissues [119].

In brain studies [49, 264] it is customary to assume that all tissues presented in an image belong to one of three groups. The first one, with the lowest intensity response is presented by cerebrospinal fluid, bone, and background air; the second, with middle intensity corresponds to brain tissues, and the third, with high intensity – to subcutaneous fat and arteries [100]. Furthermore, it is assumed that intensity of blood vessels is generally higher than intensities of all other tissues.

Using these assumptions, researchers employ various functions to model distribution of the intensities to automatically or semi-automatically classify the tissues [49,101,264]. Such methods do not require user interaction, and are based on the information that could be extracted from the data.

Even though a lot of blood vessel segmentation techniques have been proposed [69, 125, 159, 230, 237], the problem of accurate vessel segmentation from DCE-MRI data

remains a challenging task. In DCE-MRI data, due to the in-flow effect, some vessels appear bright in the pre-and post-contrast images, while the intensity of others increases with contrast agent inflow, which would lead to low signal-to-noise ratio and an intensity inhomogeneity within the region. The conventional segmentation methods that are based on the image intensity fail when there is a significant signal intensity change within the vessel area. Furthermore, the intensity and contrast between background and vessels, or inside the vessel, may vary from region to region and patient to patient. Therefore, local absolute intensity statistics in the vessel and background regions may not be reliable, or the intensity gradient magnitude on the vessel boundary may not be large enough for reliable segmentation [237].

2.7 Conclusion

Approaches for analysis of DCE-MRI data assume that relationship between the contrast agent concentration and signal change are known and well-defined [244], and also that all intensity changes at each voxel can be attributed to the contrast leakage and that each voxel represents the same tissue type.

However, intensity changes in an DCE-MRI dataset will depend on the data acquisition parameters, dose of the contrast agent, and scanning equipment. Hardware instability, magnetic field inhomogeneity and subject motion during the imaging can introduce artefactual enhancement.

At the end of an examination, a radiologist receives a dataset of up to 300 2D images, which can be corrupted by noise and patient motion artefacts. A reader semi-automatically or manually segments images, in order to locate the tissues of interest and isolate markers and then makes a decision about the patient's condition by viewing the images one by one.

Assessment and comparison of the datasets acquired in follow-up examinations is performed in a similar manner. Such evaluation is obviously highly subjective and results are not easy to reproduce. There is no technique for comparison of the data acquired from the same patient on two different scanners. Also, the quality of the data can render the entire analysis invalid [9].

Therefore, we are dealing with several problems. Firstly, there is a need for efficient pre-processing techniques that can compensate for patient motion, locate tissue of interest, and thereby contribute to data fidelity. Secondly, efficient quantitative techniques that allow assessment and interpretation of the results of multiple examinations are required.

Some standard registration and segmentation techniques fail to perform under assump-

tions imposed by the dynamic data acquisition settings. Registration of the images located at the beginning of temporal slices can be successfully performed by, for instance, the iterative closest point algorithm based on edge features selected with Canny edge detector [39], however, when the local contrast and brightness change become prominent, the algorithm fails. Therefore, in this thesis we will introduce new segmentation and registration algorithms that are suitable for DCE-MRI data pre-processing.

We have also developed a new quantitative analysis technique adequate for DCE-MRI data evaluation and interpretation, including estimation of various inflammation-related parameters, assessment of the pattern of contrast agent uptake, and visualisation of the activation events.

In the following chapters we illustrate how analysis of DCE-MRI data involved in rheumatoid arthritis studies can be performed using optimised registration, new segmentation and robust quantitative evaluation techniques.

Chapter 3

Magnetic Resonance Imaging data

3.1 Image modalities for monitoring RA

In recent decades, medical imaging became a very powerful tool, offering new possibilities for diagnostic investigations. Development of imaging techniques such as positron emission tomography (PET), computer tomography (CT), ultrasonography (US), single photon emission computed tomography (SPECT), and magnetic resonance imaging (MRI) enables physicians to examine various disorders in a non-invasive fashion by means of high-resolution three-dimensional images.

Medical imaging modalities can be divided into two major categories: anatomical and functional. Functional modalities provide information on the metabolism of the underlying organs, and anatomical – illustrate the anatomy of body parts and organs. The later are commonly used for rheumatoid arthritis assessment.

CT permits good definition of bone change. However, a patient's exposure to ionising radiation might influence changes in RA soft tissues [4]. Evaluation of rheumatoid arthritis with X-ray is relatively inexpensive, widely available and has standardised methods for interpretation [106]. However, it also has a number of limitations, such as its inability to reliably determine structural change in less than 6-12 months and the need for experienced readers to interpret the images [241].

Musculoskeletal US is a rapidly emerging technique for detection of the soft tissue lesions in inflammatory rheumatoid diseases [91, 215, 251]. The main advantages of ul-

trasonography is the absence of radiation, low running costs, and good visualisation of the tendons and joint space [102]. However, the quality of the examination is highly dependent upon the skill of the operator and the use of optimal equipment. The former leads to problems with reproducibility based on inter-observer variability and the use of different scanners [241]. The latter implies that without a high-quality imaging system US examination might not be accurate. Conventional ultrasound equipment for musculoskeletal work is equipped with standard 7.5–10 MHz transducers; however systems with even 20 MHz transducers have a limited field of view, poor beam penetration, and do not allow the evaluation of structures deeper than 1.5 cm. Moreover, not much data on the prognostic value of US in RA is available [215]. Therefore, its long term diagnostic value is unknown.

MRI is a highly sensitive technique for disclosing early rheumatoid erosions and has been shown to be better than other modalities for the detection of inflammatory changes in RA joints [11, 215]. Various studies illustrate that MRI is sensitive to the follow up analysis of bone damage, detecting soft tissue lesions, synovitis, and early erosive changes.

Performance of MRI scanning has been enhanced by the introduction of a dynamic setup [248]. The acquisition of images in a dynamic manner permits recording of the tissue behaviour over time, which delivers more comprehensive information about the tissue.

3.1.1 Magnetic resonance imaging

In MRI, a patient is placed within a high intensity magnetic field; the strength of the field might vary from 0.2T[1] to 8T [1, 214] depending on the scanning equipment [32]. The induced magnetic field causes the magnetic moments of the hydrogen atoms, presented in the tissues within the patient, to align along the principal direction of the superconducting magnet. Low-level radio waves are then transmitted through the patient causing the magnetic moments of the hydrogen nuclei to resonate and re-emit microwaves after each pulse.

When the energy source is turned off, the protons return to equilibrium, emitting the absorbed energy as a time varying radio wave, which is recorded as the MRI signal. This is done by measuring a current which is induced during the relaxation process in a radio-frequency (RF) coil due to the time-varying magnetic flux caused by the relaxing nuclei. Interpretation of the current as a function yields a sine wave decaying over time [90].

Spatial locations of the scan can be determined by varying the magnetic field about

[1] The strength of the magnetic field is measured in Tesla [T].

the patient in different directions at different times. Magnetic gradients are generated by three orthogonal coils, oriented in the x, y and z directions of the scanner. Encoding of the spatial information into the signal is accomplished by superimposing three orthogonal magnetic field gradients, resulting in a spatially dependent resonance frequency and phase of hydrogen nuclei. The signal, which likewise is a function of time, is converted by the inverse Fourier transform into the spatial domain to produce an image [248].

$T1$ / $T2$ relaxation time

The MRI signal arises from the spin-lattice relaxation time ($T1$) and spin-spin relaxation time ($T2$) of hydrogen nuclei [22]. $T1$ and $T2$ are biological tissue dependent parameters and the basis of the tissue contrast in MR images.

The $T1$ effect in the relaxation process is due to the return of the high energy state protons to the low energy state via exchanging their 'extra' energy with the neighbouring protons. The value of $T1$ indicates how quickly the spinning nuclei will emit their absorbed RF into the surrounding tissue.

The $T2$ is a tissue-specific time constant for protons, which depends on the energy exchange of the proton and nearby nuclei. Due to the interaction, protons lose their phase coherence and, therefore, magnetisation. $T2$ measures the decay in magnetisation and allows tissue types to be distinguished.

A subtle variant of the $T2$ technique is called T_2^* imaging. Formation of T_2^* imaging allows for additional sensitivity to relaxation processes (however image resolution is sacrificed). Due to this property, T_2^*-weighted sequences are used for functional MRI (fMRI), evaluation of the baseline vascular perfusion, cerebral blood volume using injected agents; in these cases, there is an inherent trade-off between image quality and detection sensitivity [151].

The $T2$ relaxation is temperature dependent. At a lower temperature molecular motion is reduced and the decay times are reduced. Fat has a very efficient energy exchange and a relatively short $T2$. Water is less efficient than fat in the exchange of energy, and has a long $T2$ time. On a $T2$-weighted scan, fat-, water- and fluid-containing tissues are bright. Damaged tissue tends to develop oedema, which makes a $T2$-weighted sequence sensitive to pathology, and generally enables us to distinguish pathologic tissue from normal tissue.

The $T1$ time affects the tissue contrast. Due to the high mobility of the water molecules, the water nuclei do not give up their energy to the surrounding tissue as quickly as fat, and therefore, on a $T1$ contrast MR image, fat will appear bright and water dark [22]. This makes $T1$ sequences useful for the assessment of the condition of various tissues such as brain, joint, and spine.

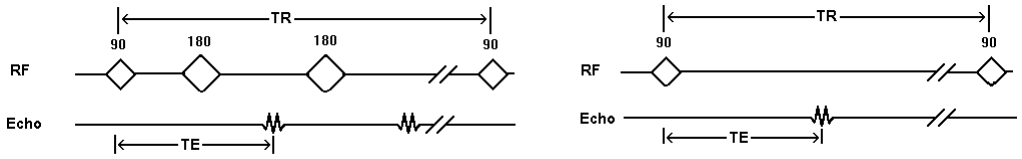


Figure 3.1: MR sequences. Left: Dual echo SE sequence. Right: Basic GRE sequence.

Repetition time / echo time

The pixel intensity $I(x,y)$ in an MR image can be described by the following equation [248]:

$$I(x,y) \propto \rho(x,y) \underbrace{\left(1-\exp\left[-\frac{TR}{T1}\right]\right)}_{T1-weighting} \underbrace{\exp\left[-\frac{TE}{T2}\right]}_{T2-weighting}, \tag{3.1}$$

where $\rho(x,y)$ is proton density, TE is echo time, and TR repetition time. TR is the time between the application of an RF pulse and the start of the next RF pulse, TE is the time between the application of the RF pulse and the peak of the echo detected [22]. An echo is the emission of energy in the form of an electromagnetic resonance signal of a nucleus after its excitation.

TE and TR influence the level of sensitivity of a particular tissue to the differences in $T1/T2$-relaxation process and therefore MR image characteristics. Adjustment of TR and TE causes the signal to become more sensitive to the relaxation process and allows for the enhancing of the contrast between specific types of soft tissue.

In practice the contrast of MR images can be altered by varying the imaging parameters TE and TR, which have to be adjusted by the operator to reasonable values in order to discriminate between various tissues [248].

MR sequences

The appearance of tissues in an MR image is influenced by proton density (i.e. number of hydrogen nuclei) and the temporal course of the $T1/T2$ relaxation process after applying certain RF-pulse sequences. A pulse sequence is a set of RF and gradient pulses repeated during a scan [248]. Figure 3.1 illustrates two types of MR pulse sequences: spin echo (SE) and gradient echo (GRE). The SE pulse sequence is the most commonly used pulse sequence. The pulse sequence timing can be adjusted to give $T1$-weighted, proton density, and $T2$-weighted images. The two variables of interest in SE sequences are TR and TE. All spin echo sequences include a slice selective 90^0 pulse followed by one or more 180^0

degree refocusing pulses.

The GRE sequences show a wide range of variations compared to the SE. The basic sequence is specified by TR, TE and an additional parameter – a flip angle. The flip angle of the spins is usually at or close to 90^0 for a spin echo sequence, but commonly varies over a range of about 10 to 80 degrees with a gradient echo sequence. The larger flip angle gives more $T1$ weighting to the image, and the smaller more $T2$.

MR sequences can be two-dimensional with a single 2D section acquired at a time, or three-dimensional with a volume of multiple sections obtained in a single acquisition.

Other imaging parameters

In MRI, anatomic coverage is determined by the size, location, and orientation of a user-defined solid rectangular volume. The possible orientations of the volume are axial, sagittal, coronal, or oblique.

Data is usually acquired in slices (a virtual slice through a 3D object). Typically, slices are parallel to one another; they may be contiguous or overlapping. A typical slice thickness varies from just under 1mm to about 5mm.

Spatial resolution or Field of View (FOV) describes the dimensions of a slice or a cross-section of a volume. Small FOV implies high resolution and small voxel size.

Contrast agents

Both $T1$- and $T2$-weighted images are acquired in medical examinations, however often they do not adequately show the anatomy or pathology. Therefore, the contrast between different types of healthy tissue as well as between healthy and pathologically altered tissue is further improved by the administration of a contrast agent, that affects the relaxation times of the tissues [50, 133, 174, 215, 253].

A contrast agent may be simply water, taken orally, for imaging the stomach. Superparamagnetic contrast agents (e.g. iron oxide nano-particles) became available in the early 90s [257]. These agents make tissues appear very dark on T_2^*-weighted images and may be used for liver imaging – normal liver tissue retains the agent.

The most common paramagnetic contrast agent is a gadolinium compound [133, 256]. Gadolinium diethylene triamine pentacetic acid (Gd-DTPA) is an extra cellular contrast agent that selectively alters MRI signal intensity throughout its distribution volume (blood plasma and extra cellular fluid).

Gd-enhanced tissues and fluids appear extremely bright on the $T1$-weighted images, which permits detection of the vascular tissues and assessment of brain perfusion or syn-

ovial tissue.

Initially, only structures immediately containing the Gd-DTPA contrast agent, like blood vessels, appear bright in the images. This is a dynamic or arterial phase of a DCE-MRI examination [230]. The timing of the arterial phase depends on the size of the contrast injection, location of the arterial bed and patient's cardiopulmonary status.

During the second phase, known as an early-delayed or venous phase [230], arteries usually remain opacified, although they may not appear as bright as during the arterial phase. After the venous phase comes a late-delayed or equilibrium phase. Vascular opacification during this phase depends on the type of contrast agent [230].

In DCE-MRI, a temporal variation of the MRI signal intensity occurs following intravenous administration of the contrast agent. The time course of signal changes corresponds to the underlying changes in local bulk tissue concentration of the agent, which in turn depends on the degree of inflammatory activity.

Therefore, DCE-MRI is an efficient technique for the assessment of the extent of inflammation and monitoring the treatment-induced changes in RA [83, 128, 240, 243].

3.2 DCE-MRI datasets involved in this research

Images involved in this research are acquired from the hand and wrist of patients involved in RA studies. Before the scanning, a patient is pre-injected with the contrast enhancing agent Gd-DTPA, then sequences of 3D volumes are acquired from the joints over a period of time during which some tissues exhibit reaction to the contrast enhancing agent.

Gd-DTPA induces selective enhancement of signal intensity in well perfused tissues and where capillary walls allow contrast penetration. As illustrated in Figure 3.2, inflamed synovium and blood vessels enhance significantly, the surrounding muscle exhibits a low degree of enhancement, and cortical bone and cartilage experience no enhancement.

Figure 3.2: Pre- and post-contrast images from a temporal slice acquired with the high-field scanner. In post-contrast image, inflamed synovium around the second bone region and blood vessels are enhanced significantly, the surrounding muscle exhibits a low degree of enhancement, and cortical bone and cartilage experience no enhancement.

Schematically the data from a single DCE-MRI experiment is shown in Figure 3.3. A 3D volume is a set of images acquired at S scans over time T. A temporal slice is a sequence of 2D images acquired from the same physical location at different time instances. A single 2D image in the DCE-MRI study is called a dynamic frame.

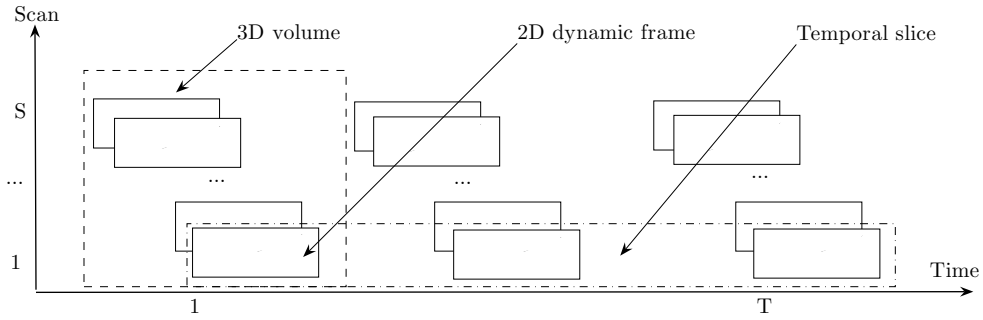


Figure 3.3: Structure of 4D DCE-MRI experiment: 3D volumes of images are composed of S scans and acquired over time T.

The acquisition parameters that define temporal scope of the imaging are specific for a given MRI scanner and are chosen by a radiologist to ensure maximum exposure of the disease affected tissues.

3.2.1 DCE-MRI data acquired by the high field scanner

Datasets from 10 patients with active RA were acquired with a 1.5T MRI scanner (Gyroscan ACS NT, Phillips Medical Systems, Best, The Netherlands). During the scanning patients were positioned prone, with their arm extended in front of the head and a linear circular 11cm diameter surface coil placed on the dorsum of the hand [195].

The positioning of the imaging volume is illustrated in Figure 3.4, where the lines indicate the positions of transverse slices [195], superimposed on a coronal and sagittal cross sections[2].

The imaging volume encompasses four (2^{nd} – 5^{th}) MCP joints and was positioned using a set of precise anatomical landmarks (such as MCP joints' capsules) to ensure adequate coverage of the joints at the baseline and follow-up scans [195].

A standard dose of 0.1 mmol per kilogram body weight of Gd-DTPA (Magnevist, Schering, Berlin, Germany) was used. The injection was administrated manually with a constant injection rate over a period of 7.1 seconds. Figure 3.2 illustrates baseline and post-contrast MR images of the MCP joints.

[2]Taken with permission from [195].

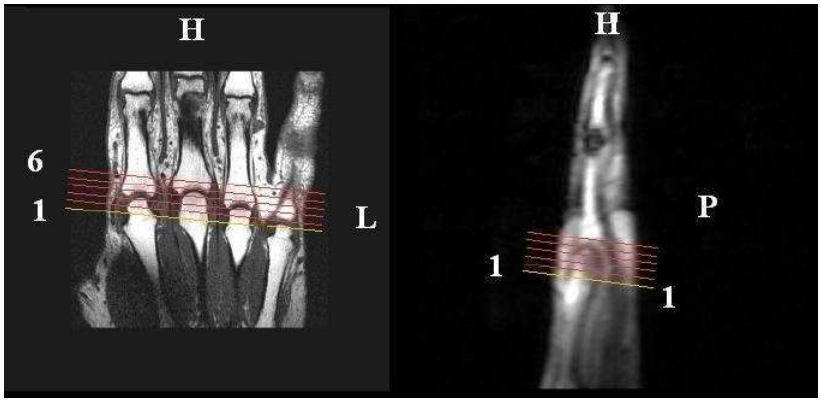


Figure 3.4: Positioning of the imaging volume in the high field hand RA study.

Imaging was performed using a 3D $T1$ weighted spoiled gradient echo sequence: TR / TE / Flip Angle = $14/3.8/40^0$; FOV = 100mm, 6 slices, 3mm slice thickness, 20 dynamic scans at 7.1 seconds intervals with a 128×256 image matrix. A single DCE-MRI dataset consists of 20 sets of $6 \times 128 \times 256$ volumes. The imaging time after the contrast injection was approximately 142 seconds (20 time instances).

3.2.2 DCE-MRI data acquired by the low field scanner

23 MRI examinations from patients with active RA, 4 healthy controls, and 1 patient with no RA, but suffering from occult wrist pain, were performed using a $0.2T$ musculoskeletal dedicated extremity scanner (E-scan, Esaote Biomedica, Genoa, Italy). The patients were examined in the supine position with the hand along the side of the body. For signal collection, a receiver-only cylindrical solenoid wrist coil was used.

The slice planes were either axial through the first carpal row or coronal through the middle part of the hand. Figure 3.5 illustrates pre- and post-contrast images of the wrist in the axial and coronal directions.

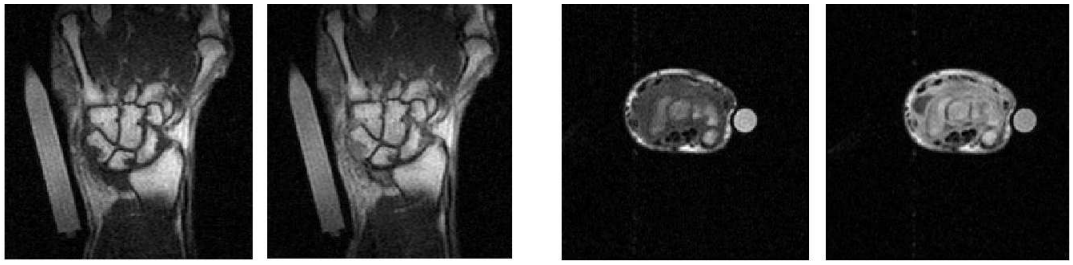


Figure 3.5: Left: Pre- and post-contrast images of the wrist scanned in the coronal direction with GRE sequence. Right: Pre- and post contrast images taken in in axial direction through the first carpal row, using SE sequence.

Figure 3.6: Pre- and post-contrast images with the blood vessels enhancing at different rate. Blood vessels in squares correspond to the arteries, and blood vessels in circles to veins.

The injection of Gd-DTPA (Magnevist, Schering AG, Berlin, Germany) was given at a dose of 0.2 mmol/kg of body weight over 30 seconds through a 21mm butterfly needle in the cubital vein.

Following the Gd-DTPA injection, the dynamic sequence was acquired making 22-30 consecutive fast spin-echo or fast gradient-echo images in three pre-positioned planes every 10-15 seconds. Slice thickness in the coronal direction was 4mm, and in the axial direction slice thickness is 5mm. Acquisition parameters for $T1$-weighted SE are TR/TE: 600/18 ms, FOV/ imaging matrix: 180×180mm / 192×192; for and the axial/coronal turbo 3D $T1$ gradient echo TR/TE: 38/16, FOV/matrix: $180 \times 180 \times 100$mm / $192 \times 160 \times 72$.

3.3 Data analysis

3.3.1 Veins and arteries

In DCE-MRI, the intra venous injection of the contrast agent is imaged on the first pass through the arterial system. The sequential scans essentially record the speed of the Gd-DTPA penetration, flowing through the veins and arteries.

It was noticed that the blood vessels corresponding to the arteries enhance even in the absence of Gd-DTPA, whereas blood vessels corresponding to the veins appear bright only in the images acquired at later time instants. Figure 3.6 illustrates pre- and post-contrast images from a temporal DCE-MRI slice with the blood vessels enhancing at different rates.

This observation can be explained by the fact that the blood vessels will enhance in response to the contrast agent and due to the inflow effect[3]. Ideally, penetration of the contrast agent into the volume of interest should occur within 10 seconds from the onset

[3]The inflow effect describes the fact that highly oxygenated blood flowing through the vessels generates signal even in the absence of the contrast agent.

of scanning. At this time, enhancement of the synovium and the blood vessels is expected to begin. However, even before the contrast agent penetrates the tissue of interest, the highly oxygenated blood flowing through the arteries provides a lot of signal due to the inflow effect. Therefore, blood vessels corresponding to arteries appear enhanced even in the first images in a temporal slice.

The inflow effect is not visible in the veins, which are wider than arteries, and therefore maintain slower blood flow. Thus, blood vessels corresponding to the veins appear bright only in the post-contrast images.

3.3.2 Bones and markers in the DCE-MRI datasets

The boundaries of the bone cross-sections are rigid, each bone is surrounded by cartilage, blood vessels and muscle. The bone cross-sections (which include both cortical and trabecular bone as well as bone marrow) will hereafter habitually be referred to as 'bones'.

The bones may not be visible in every image in the DCE-MRI study due to the poor inter slice resolution of the image, or the physical location of the MR scan. The phalanges of the joints are organised as shown in Figure 3.7, so a 2D temporal slice may lie between them, and some bones could be imaged partially or missed as in Figure 3.7 (right).

A temporal slice captures the same physical location of the joints. Therefore, it can be assumed [195] that, in the absence of significant patient movement, if a bone is visible in one image in the temporal slice, then it should be visible throughout the slice.

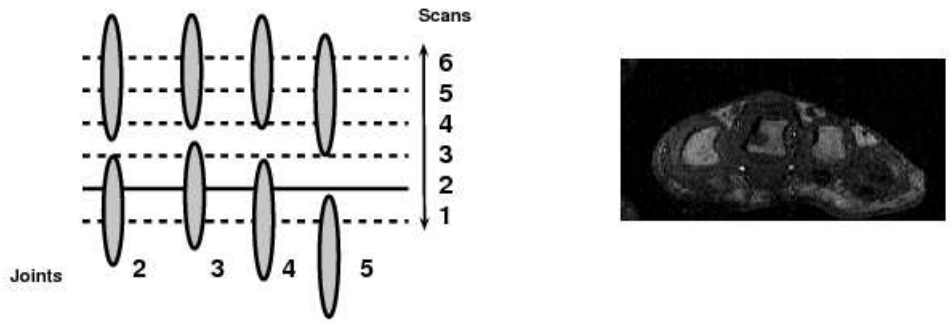


Figure 3.7: Left: A coronal plan of the phalanges. Six scans of four phalanges of the joints. Right: An axial MR section. With the second scan only three bones are imaged.

Some images acquired with the low field scanner contain a marker, which is located close to the wrist. The marker is used to normalise image intensities and in our case was a garlic pill attached to the patient's hand with a bandage.

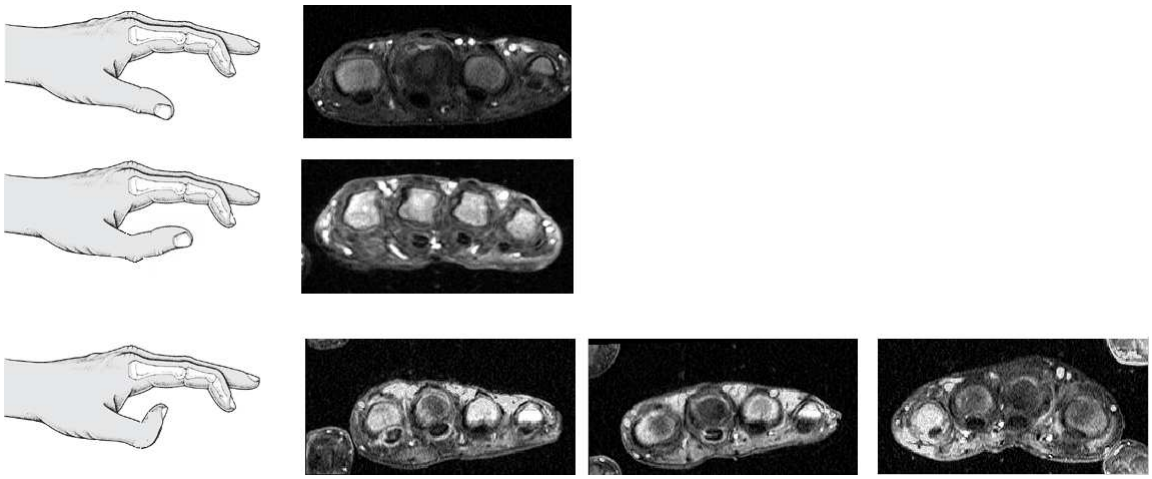


Figure 3.8: Location of thumb in the images acquired with the high field scanner. Picture of the hand is taken from [59]. Top row: The thumb is not depicted. Middle row: Only a small part of the thumb is visible in the MR image (left side). Bottom row: The thumb was bent during the examination. It is represented by two regions on the left or right side of the 2^{nd}–5^{th} MCP joints in the corresponding MR image.

Images acquired with the high field scanner might depict a thumb. The thumb is connected to the trapezium of the joint and is located on one of the sides, parallel to the arm. The thumb can be easily rotated by 90^0, on a perpendicular level compared to the palm (see Figure 3.8).

3.3.3 Normalised coordinate system

We manually outlined contours of the bone interiors, joints' envelope (a boundary that separates background, markers, and thumb, if present, from the interior of the joints, which includes the soft tissue, skin, bone, and muscle) and blood vessels in 20 DCE-MRI slices of the MCP joints acquired with the high field scanner. In order to analyse their location, we present our results in a normalised coordinate system.

Let $\{x_i, y_i\}_{i=1\ldots N}$ denote coordinates of the MCP joints' envelope in the coordinate axes X and Y, where N is the number of pixels in the boundary. Firstly, we transformed the data to be zero-mean by subtracting $(\frac{\sum_i x_i}{N}, \frac{\sum_i y_i}{N})$ from $\{x_i, y_i\}_{i=1\ldots N}$. Then, principal component analysis was performed on the zero-mean dataset and the coordinates $\{x_i, y_i\}_{i=1\ldots N}$ were rotated to the principal axes. Therefore, the transformation to the new coordinates $\{p_i, q_i\}_{i=1\ldots N}$ from $\{x_i, y_i\}_{i=1\ldots N}$ can be described by Equation 3.2, where spatial notation of the coordinates was dropped for the sake of clarity and θ is the angle between the original and principal axes.

$$\begin{bmatrix} p \\ q \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x - \frac{\sum x_i}{N} \\ y - \frac{\sum y_i}{N} \end{bmatrix} \tag{3.2}$$

Lastly, the coordinates $\{p_i, q_i\}_{i=1...N}$ are normalised by the standard deviation of p_i and q_i. Figure 3.9 illustrates the location of the joint envelopes, bone contours, and blood vessels outlined in randomly chosen 15 DCE-MRI slices of the MCP data acquired by the high-field scanner in the normalised coordinate system. All six slices from the DCE-MRI studies were included.

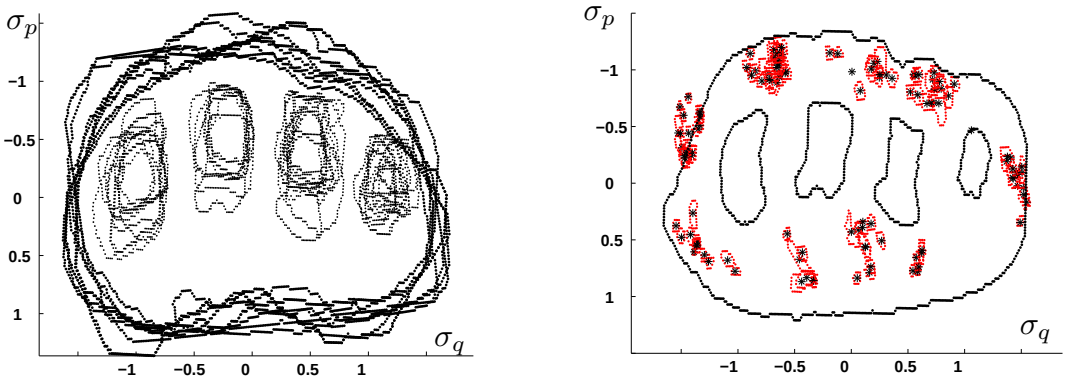


Figure 3.9: Left: The joints' envelopes and bone contours outlined in randomly chosen 15 DCE-MRI slices acquire by the high-field scanner and plotted in the normalised coordinate system. Slices 1 to 6 from various patients were included. Right: The centers (in black) and contours (in red) of the blood vessels outlined in 15 DCE-MRI slices of the MCP joints. σ_p and σ_q correspond to the standard deviation of p_i and q_i.

The bones appear centrally in the vertical / minor direction (and predictably in the horizontal / major direction)in respect to the joints' envelope boundary; the blood vessels tend to cluster close to the joints' boundary away from the bone interiors. However, some vessels might appear within the synovial tissue or near the bone interiors.

To evaluate the location of the vessels within the study, we manually outlined them in six dynamic frames acquired from the MCP joints of a patient. Figure 3.10 illustrates their appearance in the images received at the first and last scans. Due to the large distance between the temporal slices, we can observe dramatic differences in the location and size of the vessels. Figure 3.11 illustrates the vessels in six temporal slices (shown from different viewpoints).

This preliminary analysis of the data indicates that there are some similarities in appearance of the tissue of interest, such as bone interiors and joint envelopes in the inter-patient data. Even when all six slices were included in the analysis, we could observe

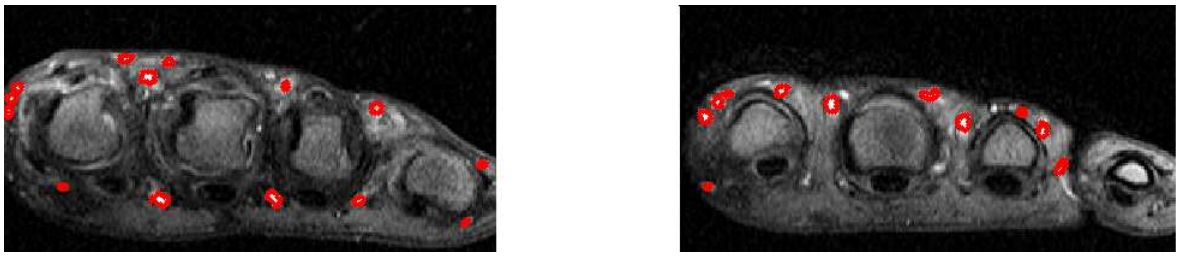


Figure 3.10: Blood vessels outlined in the post-contrast dynamic frames in the 1^{st} and 6^{th} temporal slices from the dataset acquired from a patient with the high-field scanner.

clear margins. Slice thickness, structure, and the small size of the vessels could be responsible for dissimilarities in their location. Often, blood vessels branch, which due to the large distance between slices, might appear as a blood vessel was divided into two or three vessels.

Figure 3.9 illustrates that bone interiors appear to cluster in location and blood vessels do not. Future analysis of the data will relate to this ground truth information, which provides the basis for some segmentation techniques described in the next chapters.

3.4 Conclusion

In recent years contrast-enhanced dynamic MR imaging has become a commonly used method for diagnosis and monitoring of inflammatory diseases. It permits acquiring information about the tissue vascularity and perfusion through measuring the speed of the contrast agent penetration into the target tissue. However, widespread use of DCE-MRI is limited by the need for further technical improvements and development of software algorithms for data analysis.

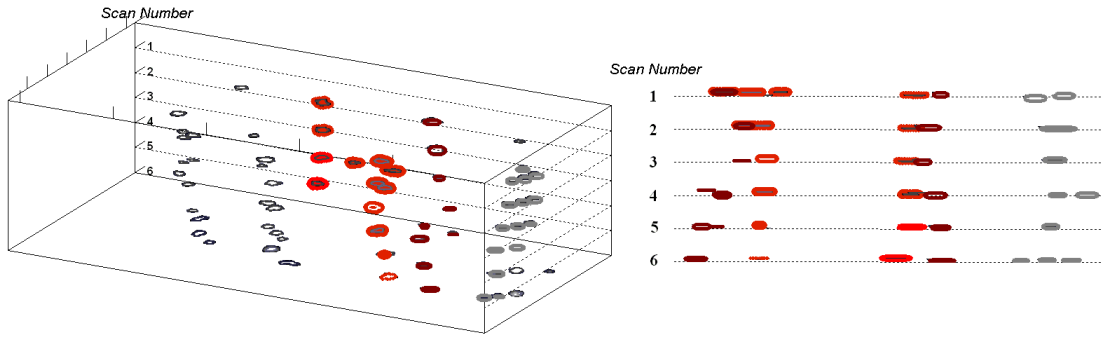


Figure 3.11: Position of the blood vessels in each slice from a sample DCE-MRI study; 3D view from different angles.

Acquisition parameters and properties of the datasets acquired in a dynamic manner by means of the high and low field MRI scanners in the presence of the Gd-DTPA contrast agent were discussed in this chapter. Ultimately, image quality is constrained by the limits imposed by a scanner, such as signal-to-noise ratio, imaging time, and contrast. In dynamic imaging, the speed of the data acquisition plays an important role. There is always a trade off between longer acquisition times, which permit higher image resolution, better contrast and larger field of view, and the degree of noise in images due to patient motion and hardware calibration.

Formation of an MRI image, which reflects clinically relevant anatomy and physiology, is achieved by manipulating the image contrast with the appropriate data acquisition parameters. The parameter choice will depend on the capability of a scanner, speed of the acquisition, desirable tissue contrast, and tissue characteristics.

Protons in environments corresponding to different materials have different longitudinal and transverse relaxation times, $T1$ and $T2$. The differences between these parameters and timing of the data acquisition are used to produce the contrast between the materials in an MR image. The flexibility of the MR contrast enhanced due to the presence of the contrast agents permits depiction of the anatomical and functional information of use in many clinical applications.

Optimisation of the parameters for a given task can be challenging [163]. Limited availability of the quantitative characterisation of MR tissue parameters and the variability of these characterisations restrict selection of the imaging parameters such as TR and TE. Efforts in this direction are progressing, however, for targeted applications [25, 140, 267].

Development of faster imaging sequences that produce images at higher resolution and efforts to reduce the system cost by reducing the magnetic field strength incur a signal-to-noise penalty. Currently, a great deal of effort has been invested in the image post-processing methods, which allow for data quality enhancement.

The post-processing techniques for data analysis normally take into account the nature of the data acquisition (MR data parameters) and anatomy of the region under investigation, which often restrict them to be application dependent. Normalisation of the position, shape and size of a tissue of interest in intra- and inter-patient data simplifies such analysis.

Ideally, image processing methods should be fully automated. However, there is a trade off between the amount of human input and the adaptability of the approach to a wide range of applications. Fully automated algorithms can normally process only a restricted range of data, but semi-automated set-up or manual initialisation allows for a wider spectrum of applications.

In the following chapters we present several algorithms for registration, segmentation and its evaluation, and quantitative analysis of the DCE-MRI data acquired from the hand and wrist of patients involved in RA studies.

Chapter 4

Image registration

Approaches for quantitative analysis of DCE-MRI data assume that signal intensity vs. time changes at each voxel can be attributed to the contrast leakage. However, patient movement can introduce artefactual enhancement with implications to the extracted measurements. Prior to the quantitative analysis images need to be aligned. In this chapter we demonstrate the value of registration.

The algorithm presented in Periaswamy et al. [184] explicitly models geometric, contrast, and brightness variations and allows for efficient alignment of synthetic, PET, MRI, CT, and X-ray images. It was designed to deal with occlusions and missing data via an expectation-maximisation (EM) step. However, the authors admit that this step is time-consuming and not robust, which causes the algorithm to fail on dynamic contrast-enhanced datasets where the effect of the contrast agent is prominent.

We have modified the algorithm [184] for efficient application on DCE-MRI data. Firstly, the EM step was eliminated to reduce computational time. Secondly, an incremental approach to the registration was introduced. This approach has been motivated by the fact that some tissue significantly changes its intensity and brightness in the post-contrast images and therefore, alignment of the post-contrast source to the pre-contrast target may not be accurate.

Alignment of the images/volumes from the contrast-enhanced studies in the incremental rather than pair-wise fashion allows us to take into account the contrast and brightness variations that occur locally in the soft tissue and, therefore, to reduce the registration error.

In this chapter a 3D extension of the algorithm [184] will be documented. The technique is going to be presented at three stages. At the beginning, it is assumed that there are no temporal contrast and brightness variations in the data and images/volumes can be brought into alignment using a purely geometric inform. Then, two extra parameters that explicitly describe the contrast and brightness variations are added in the registration model. Lastly, a global smoothness constraint is imposed on the geometric, contrast, and brightness parameters.

The modified 3D version of the algorithm is documented in Section 4.1[1]. An incremental approach to the registration problem is introduced in Section 4.2. A need for the smoothness constraint and performance of 2D/3D algorithms enhanced with various types of transformations on DCE-MRI data of the hand and wrist acquired by high and low field scanners is discussed Section 4.4.

4.1 3D registration algorithm

4.1.1 Local affine

Let $f(x,y,z,t)$ and $f(x,y,z,t-1)$ be source and target 3D images. It is assumed that there are no contrast/brightness variations between the volumes and transformations can be modelled using pure geometric parameters:

$$\begin{aligned} f(x,y,z,t-1) = \\ f(m_1x+m_2y+m_3z+m_{10}, m_4x+m_5y+m_6z+m_{11}, \\ m_7x+m_8y+m_9z+m_{12}, t) \end{aligned} \tag{4.1}$$

where $(m_1,...,m_9)$ represent the affine parameters, and $(m_{10},...,m_{12})$ – the translation parameters; $\vec{m} = (m_1,...,m_{12})$ are estimated locally for each small neighbourhood, but for the sake of clarity their spatial notation is dropped.

A least square measure has been employed to deduce parameters $\vec{m}$ via minimising the cost:

$$\begin{aligned} E(\vec{m}) \quad = \quad & \sum_{x,y,z\in\Omega} [f(x,y,z,t-1) - f(m_1x+m_2y+m_3z+m_{10}, \\ & m_4x+m_5y+m_6z+m_{11}, m_7x+m_8y+m_9z+m_{12}, t)]^2 \end{aligned} \tag{4.2}$$

where the sum is over the spatial support Ω of $f(\cdot)$ (Ω denotes a neighbourhood of the

[1] Terminology and notation used in Section 4.1 are heavily based on the description given in [184] [Taken with permission].

current pixel; the size of the neighbourhood is discussed in Section 4.1.3). The error function was approximated by the first order truncated Taylor series expansion [94]:

$$\begin{aligned} E(\vec{m}) \approx & \sum_{x,y,z\in\Omega} \Big(f(x,y,z,t-1) - [f(x,y,z,t-1) + \\ & + (m_1x + m_2y + m_3z + m_{10} - x) f_x(x,y,z,t-1) + \\ & + (m_4x + m_5y + m_6z + m_{11} - y) f_y(x,y,z,t-1) + \\ & + (m_7x + m_8y + m_9z + m_{12} - z) f_z(x,y,z,t-1) - f_t(x,y,z,t-1)] \Big)^2 \end{aligned} \quad (4.3)$$

$$\begin{aligned} E(\vec{m}) = & \sum_{x,y,z\in\Omega} [f_t(x,y,z,t-1) - (m_1x + m_2y + m_3z + m_{10} - x) f_x(x,y,z,t-1) - \\ & - (m_4x + m_5y + m_6z + m_{11} - y) f_y(x,y,z,t-1) - \\ & - (m_7x + m_8y + m_9z + m_{12} - z) f_z(x,y,z,t-1)]^2 \end{aligned} \quad (4.4)$$

In Equation 4.3 $f_x(\cdot)$, $f_y(\cdot)$, $f_z(\cdot)$, and $f_t(\cdot)$ are spatial and temporal derivatives of $f(\cdot)$. Further, the error function may be expressed in vector form:

$$E(\vec{m}) = \sum_{x,y,z\in\Omega} [k - \vec{c}^T \vec{m}]^2, \quad (4.5)$$

where scalar k and vector $\vec{c}$ are defined as:

$$\begin{aligned} k &= f_t + x f_x + y f_y + z f_z \\ \vec{c} &= (x f_x, y f_x, z f_x, x f_y, y f_y, z f_y, x f_z, y f_z, z f_z, f_x, f_y, f_z)^T \end{aligned} \quad (4.6)$$

The quadratic error function from Equation 4.2 has been transformed in a linear form using the Taylor expansion. Therefore, its minimisation can be performed analytically by differentiating the error function with respect to the unknowns:

$$\frac{dE(\vec{m})}{d\vec{m}} = \sum_{x,y,z\in\Omega} -2\,\vec{c}[k - \vec{c}^{\,T}\vec{m}] \quad (4.7)$$

Setting this result to zero, and solving for $\vec{m}$ yields:

$$\vec{m} = \Big[\sum_{x,y,z\in\Omega} \vec{c}\,\vec{c}^{\,T} \Big]^{-1} \Big[\sum_{x,y,z\in\Omega} \vec{c}\,k \Big] \quad (4.8)$$

With this approach a locally affine mapping can be found between the source and target images or volumes. However, the effect of the contrast agent has not been taken into account.

4.1.2 Intensity variations

The model described above assumes no contrast or brightness variations between the source and target images/volumes. This assumption fails when the volumes acquired in the presence of the contrast agent are aligned to the pre-contrast target. Therefore, to account for the contrast and brightness variations, two new spatially varying parameters m_{13} and m_{14} are introduced into the initial model. The transformation between the source and target volumes takes the following form:

$$m_{13} f(x,y,z,t-1) + m_{14} = \\ f(m_1 x + m_2 y + m_3 z + m_{10}, m_4 x + m_5 y + m_6 z + m_{11}, m_7 x + m_8 y + m_9 z + m_{12}, t)$$

It is here assumed that the contrast and brightness change locally in an affine manner. Obviously, the affine approximation does not describe accurately the real processes, however it reflects the fact that the contrast and brightness changes in soft tissues occur locally and allows us to perform differentiation. As earlier, the error function is approximated by a first order Taylor series expansion and differentiated to its unknowns $\vec{m}$ that now consists of 14 components. The result is set to zero, and the solution takes the following form:

$$\vec{m} = \Big[\sum_{x,y,z \in \Omega} \vec{c}\,\vec{c}^{\,T} \Big]^{-1} \Big[\sum_{x,y,z \in \Omega} \vec{c}\,k \Big], \tag{4.9}$$

where vector $\vec{c}$ and scalar k are:

$$\begin{aligned} k &= f_t - f + x f_x + y f_y + z f_z \\ \vec{c} &= (x f_x, y f_x, z f_x, x f_y, y f_y, y f_z, x f_z, y f_z, z f_z, f_x, f_y, f_z, -f, -1)^T \end{aligned} \tag{4.10}$$

These additional terms allow for efficient registration of the data complicated by the contrast and brightness variations. However, at this stage it is assumed that the geometric, brightness and contrast parameters are constant within a small neighbourhood Ω. To relax this assumption, the smoothness constraint is imposed on the volume space.

4.1.3 Smoothness constraint

The first term in Equation 4.9 is assumed to be invertible. This can be guaranteed by choosing a neighbourhood of large size [184]. However, the assumption that the local affine, brightness and contrast parameters are constant within a neighbourhood is only likely to be held when the area of the neighbourhood is small.

To avoid making a decision on the optimal size of the neighbourhood, the assumption

that the parameters $\vec{m}$ do not change within the neighbourhood is replaced with a smoothness assumption, which implies that physical properties in the neighbourhood of a space or within the time interval do not change abruptly [18, 141, 147].

We imposed the smoothness constraint on the model parameters $\vec{m}$ and now the error function consists of two terms: the smoothness constraint E_s and the geometric and intensity transformation constraint E_b:

$$E(\vec{m}) = E_b(\vec{m}) + E_s(\vec{m}) \tag{4.11}$$

The geometric and intensity transformation term is defined as:

$$E_b(\vec{m}) = [k - \vec{c}\,\vec{m}]^2, \tag{4.12}$$

where $\vec{c}$ and constant k are given by Equations 4.10.

The smoothness term $E_s(\vec{m})$ penalises solutions proportional to the local change in each parameter across a small spatial neighbourhood and is defined with the aid of positive constants $\{\lambda_i\}_{i=1\ldots14}$, that control the relative weight given to the smoothness constraint on a parameter m_p:

$$E_s(\vec{m}) \;=\; \sum_{p=1}^{14} \lambda_p \Big[\Big(\frac{\partial m_p}{\partial x}\Big)^2 + \Big(\frac{\partial m_p}{\partial y}\Big)^2 + \Big(\frac{\partial m_p}{\partial z}\Big)^2 \Big] \tag{4.13}$$

The value of λ, which is set to be the same for all parameters will be discussed in Section 4.4.1.

The error function, defined in such a way, allows for a locally smooth, but globally non-rigid transformation. Minimisation of the error function was done through differentiating, setting the result to zero and solving:

$$\frac{dE(\vec{m})}{d\vec{m}} = \frac{dE_b(\vec{m})}{d\vec{m}} + \frac{dE_s(\vec{m})}{d\vec{m}} \tag{4.14}$$

The derivative of $E_b(\vec{m})$ is:

$$\frac{dE_b(\vec{m})}{d\vec{m}} = -2\vec{c}[k - \vec{c}^T\vec{m}] \tag{4.15}$$

To compute the derivative of $E_s(\vec{m})$, the partial derivatives $\partial m_p/\partial x$, $\partial m_p/\partial y$, and $\partial m_p/\partial z$ are firstly estimated for each pixel location using their discrete approximations [65, 73, 116]. Then

$$\frac{dE_s(\vec{m})}{d\vec{m}} = 2L(\overline{\vec{m}} - \vec{m}), \tag{4.16}$$

where $\overline{\vec{m}}$ is the component-wise average of $\vec{m}$ over a small spatial neighbourhood, and L is a 14×14 diagonal matrix with λ_p on diagonal and zero off the diagonal. Setting $dE(\vec{m})/d\vec{m} = 0$, and solving for $\vec{m}$ at each pixel location would yield a solution. Convergence of optical flow registration methods was discussed in [166], where the problem was solved analytically.

4.2 Sequential registration

Let U^{AB} denote a transformation obtained with the algorithm described in Section 4.1 when registering a source B to a target A. For a given sequence of images/volumes I_1, ..., I_N, the following transformations can be derived with the algorithm: $U^{I_1 I_2}$, $U^{I_1 I_3}$, ..., $U^{I_1 I_N}$. This transformation is also known as Lagrangian approach. Let $\Im_k$ denote the output of this transformation, applied to an image I_k, $k \in [1, N]$, where N is the number of images/volumes in the slice:

$$\Im_k = U^{I_1 I_k}[I_k] \approx I_1 \tag{4.17}$$

Differences in the geometry between the images/volumes acquired at the first and last time instances might not be significant, however the contrast and brightness variations, especially in the datasets acquired from the patients severely affected by rheumatoid arthritis, are dramatic. Therefore, it might be argued that alignment of the post-contrast images/volumes to the pre-contrast target might not be accurate. This approach is also known as Euclidian.

To minimise the registration error the transformation could be performed in an incremental rather than pair-wise fashion: firstly, the transformation is estimated between the neighbouring pairs of images/volumes, and then the output of this sequential transformation is used as an initial solution for the basic registration algorithm. Such transformation corresponds to a hybrid Euclidian-Lagrangian approach.

This form of approach was first discussed in structure from motion estimation [99, 227], where for a given physical object and a set of views capturing this object from various viewpoints, levels of detail, and lighting conditions, the task is to reconstruct the structure of a scene from the motion of an observer. This is solved by analysing parts of the object in the photographs aligned in such a way that they show changes of the scene in an incremental manner. The approach confirms that it can be more efficient to deal with small rather than significant changes.

By analogy, in DCE-MRI datasets, it may be more efficient to determine a transformation between images/volumes, where the contrast/brightness variations are less signif-

icant.

Thus, firstly, neighbouring images/volumes are aligned. For example, in a temporal slice, an image acquired at the 2^{nd} time instant is registered to the image acquired at the 1^{st}, then an image acquired at the 3^{rd} time instant is aligned to the image acquired at the 2^{nd} time instant, and the output is registered to the 1^{st}, etc. Therefore, the final transformation is formed as a composition of the transformations:

$$V^{I_k} = U^{I_1 I_2} \otimes \ldots \otimes U^{I_{k-2} I_{k-1}} \otimes U^{I_{k-1} I_k} \tag{4.18}$$

This transformation applied to the k^{th} image yields the image J_k:

$$J_k = V^{I_k}[I_k] \approx I_1 \tag{4.19}$$

However, if alignment of the first few images/volumes was not perfect, registration error would propagate when registering images/volumes from later in the study, and alignment between the first and the last volumes would be erroneous.

To compensate for this possible error, a sequence of images/volumes registered in the sequential manner $\{J_k\}_{k=1\ldots N}$, was taken as an initial solution for the basic registration algorithm. Then, the final transform is defined as:

$$W^{I_k} = U^{I_1 J_k} \otimes V^{I_k} \tag{4.20}$$

and being applied to the k^{th} image yields the image L_k:

$$L_k = U^{I_1 J_k} \otimes (V^{I_k}[I_k]) = W^{I_k}[I_k] \tag{4.21}$$

With this approach a DCE-MRI study is considered as a whole, which permits reduction of the transformation error and allows compensation for contrast and brightness variations between the images.

Therefore, images / volumes from a DCE-MRI datasets can be registered using one of the suggested transformations:

- $\{\Im_k\}_{k=1\ldots N}$, obtained with the algorithm discussed in Section 4.1; *basic U-transformation*;
- $\{J_k\}_{k=1\ldots N}$, obtained with the sequential application of the algorithm; *sequential V-transformation*;
- $\{L_k\}_{k=1\ldots N}$, obtained with the incremental approach; *incremental W-transformation*.

Performance of 2D and 3D registration algorithms with U, V, and W transformations will be compared in Section 4.4.

4.3 System design

The following modifications in the system design were shown to be beneficial for algorithm performance. The effect of these modifications is discussed in [184].

- A 4 level Gaussian coarse-to-fine pyramid for the source and target images/volumes was built in a standard way [2]. Each level of the pyramid was obtained by convolving the previous level using a Gaussian kernel, followed by down-sampling by a factor of two[2].

  The source and target were registered at the coarsest scale to obtain an initial estimation of the registration map. This initial estimate was then used to warp the source image at the next scale. The warped source image was then registered with its corresponding target. This process was repeated at each level of the pyramid. A single registration map was maintained by accumulating successful registration maps estimated at each scale [14].

- Secondly, Taylor expansions used to approximate the error function were replaced with the Newton-Raphson style iterative scheme [147, 222] that provides a more accurate estimate of the actual error function [184]. After initial estimation of the registration parameters, the source is warped with the estimated parameters and registered again with the target. During each of these iterations, successful intermediate registration maps were accumulated to form a single registration map. These iterations were stopped when the average displacement of the estimated motion was less than 1 pixel. It was shown in [184] that five iterations improve the final estimate significantly.

- Lastly, partial derivatives $\partial m_p/\partial x$, $\partial m_p/\partial y$, and $\partial m_p/\partial z$ were estimated with the aid of the kernels developed by Farid et al. [73]. This set of derivative filters was specifically designed for multi-dimensional differentiation [73] and used in the original version of the algorithm [184].

  For 2D-2D registration, the images were first pre-filtered in time using the two-tap filter [0.5 0.5]. The derivative in the x direction is then estimated by first pre-filtering

[2] We also tried a Laplacian pyramid [37], which is computed as a difference between the original image and low-pass filtered image, but it provided slightly less accurate results.

the result in the y direction (using the 3-tap filter d_0=[0.2 0.5 0.2]), followed by differentiating in x with the filter d_1=[-0.4 0 0.4]. Similarly, the derivative in the y direction is estimated by first pre-filtering the result in x using d_0, followed by differentiating in y with the derivative filter d_1. The derivative in time is estimated by first pre-filtering in space (x and y) with d_0, followed by applying two-tap derivative filter. For 3D-3D examples, the following filters were used: d_0=[0.09 0.4 0.4 0.09] and d_1=[-0.2 -0.3 0.3 0.2] (see [73] for details).

4.4 Experiments and discussion

The described registration model was demonstrated to be highly efficient and robust for a wide range of synthetic, CT, PET, and static MRI data [184]. This section discusses the nature of motion in DCE-MRI studies and performance of the algorithm in application to dynamic contrast-enhanced data.

Evaluation of registration results was performed by measuring the residual difference between the corresponding anatomies over the image domain. There is no gold standard that can be used to validate the results of the registration algorithm on our data, therefore we proceed as follows:

- In images with a fiducial marker, residuals between the source and target within the area of the marker are assessed before and after registration using a variability measure.
- For studies without a marker, positions of the rigid bone interior before and after registration were analysed.

In Section 4.4.4 we provide 2D and 3D images before and after registration, so that the value of the algorithm can be assessed visually.

4.4.1 Parameter λ and analysis of the motion

Alignment of the data, acquired in follow-up examinations, performed over a period of time long enough for the anatomy to change, or the data captured in such a way that physical location of the scans in the source and target is slightly different, will require non-rigid methods.

Images from a DCE-MRI dataset acquired in a single examination do not exhibit dramatic anatomy change. Therefore, in the absence of the patient motion the images/volumes

from a dynamic dataset will be similar to each other except for local variations in brightness and contrast, which result from the response of the disease active tissues to the contrast agent. There might be a small change in the anatomy of the soft tissue due to the motion induced by the contrast agent or muscle relaxation. However, in general, the motion is expected to be locally rigid.

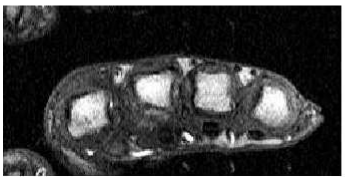
This was confirmed when considering the value of the parameter λ from Equation 4.13. In our experiments $\lambda = 10$, however the algorithm's performance does not change significantly with the higher values of λ (acceptable range for DCE-MRI data used in these experiments [1; 100]; in [184] $\lambda = 1 \times 10^{11}$). This indicates that the smoothness term E_s in the error function $E_b + \lambda E_s$ is very small, and therefore, the smoothness parameters imposed on the geometric transformations are largely irrelevant to the error. On average for 10 studies, $\partial m/\partial x$ is 0.007, $\partial m/\partial y$ – 0.01, and $\partial m/\partial z$ – 0.0015.

The intensity and brightness parameters vary from image to image in local non-rigid fashion. Abrupt changes are tolerable on the boundary of some structures, such as blood vessels. However, the smoothness constraint controls intensity and brightness changes in the synovial tissue. This allows generating a solution in which the mapping between images/volumes is described by nearly rigid geometric modulation with added constraint on non-rigid changes to the contrast and brightness.

4.4.2 2D and 3D registration examples

The target and source are 128×256 or 256×256 8-bit grey scale images with intensity values scaled into the range [0, 1]. The joints' envelopes have been segmented. In order to contend with border effects, each image has been padded with zeros to a size 270×270 before registration. The optimal size of padding was determined empirically: the maximum shift observed in images was 5 pixels. Figure 4.1 (right) illustrates a sample image used for registration.

A 4-level Gaussian pyramid is constructed for the source and target images. Each level of the pyramid is obtained by convolving the previous level with a low pass filter, followed by down-sampling. The transformation maps are estimated at each level as described above. The final estimate is then up-sampled and used as an initial estimate in the next level of the pyramid. Figure 4.2 (middle) illustrates resampled target and source volumes acquired in six scans.



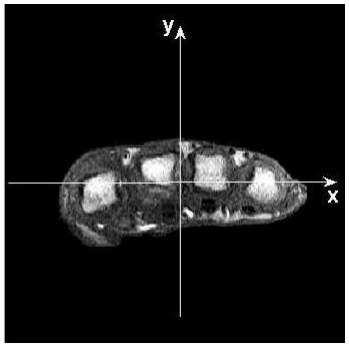


Figure 4.1: Left: Pre-contrast image of the MCP joints. Right: A sample image padded to 270×270; the joints' envelope was segmented using the algorithm described in Chapter 5.

4.4.3 Quantitative evaluation

We start by evaluating registration with the U-transformation, then compare its performance against the V- and W-transformations using 2D and 3D DCE-MRI data.

4.4.3.1 Range of values of registration parameters

Figures 4.3 and 4.4 illustrate a range of values of registration parameters (rotation and translation) obtained for in-slice and in-study motion correction with 2D and 3D registration algorithms with the U-transformation[3]. The x-axis translations are [-2, 0.5], the y-axis translation [-0.3, 5], and the rotations [-0.11, 0.12]. With 3D registration translations around the x, y, and z axis are in the range of [-5, 5], [-5.2, 1], and [-0.3, 0.2], respectively; rotations around the x, y, and z axis are in the range of [-0.2, 0.2], [-0.2, 0.2], and [-0.07, 0.1].

4.4.3.2 Mutual overlap based error

To perform evaluation with the mutual overlap metric[4] [66] the bone interiors in the DCE-MRI datasets acquired with the high field scanner and the markers in the datasets acquired with the low field scanner were segmented using algorithms described in Chapter 5. These regions are rigid and are not expected to enhance during the scanning procedure. Positions

[3]The box shows the lower quartile, median, and upper quartile values and the whiskers corresponds to the extent of the rest of the data. Maximum whisker length in units of inter-quartile range (standard is 1.5 $\times$ the inter-quartile range): Upper Inner Fence = 75th Percentile + (1.5 $\times$ Inter-quartile Range); Lower Inner Fence = 25th Percentile - (1.5 $\times$ Inter-quartile Range). Data points that lie outside of the fence, are considered to be outliers and marked as '+' [158].

[4]See Section 2.4.1.1 for metric description.

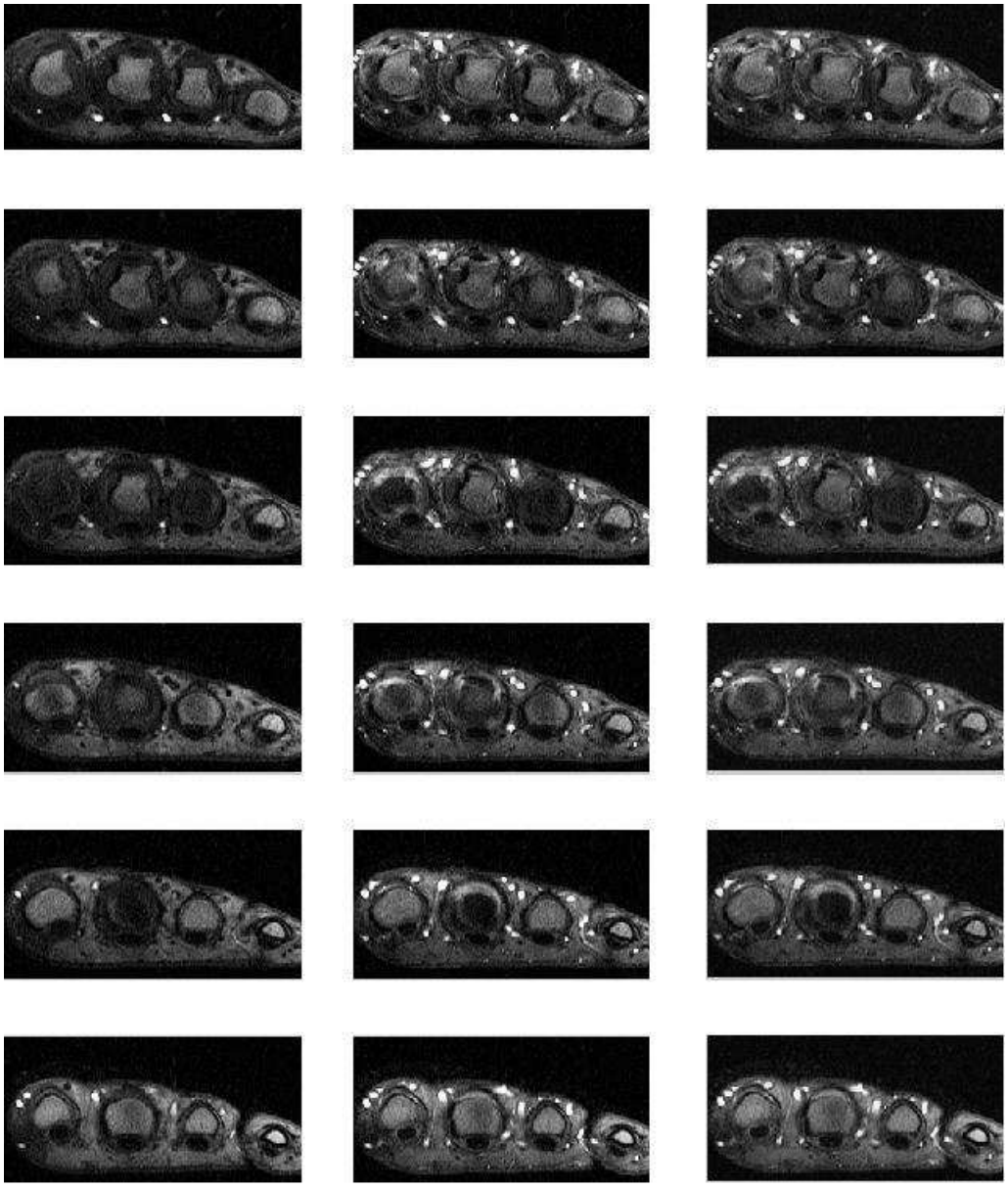


Figure 4.2: Resampled volumes of target (left), source (middle), and registered source (right) acquired with high field scanner from MCPJs. Six images in each column correspond to the six temporal slices. Registration was performed with the W-transformation.

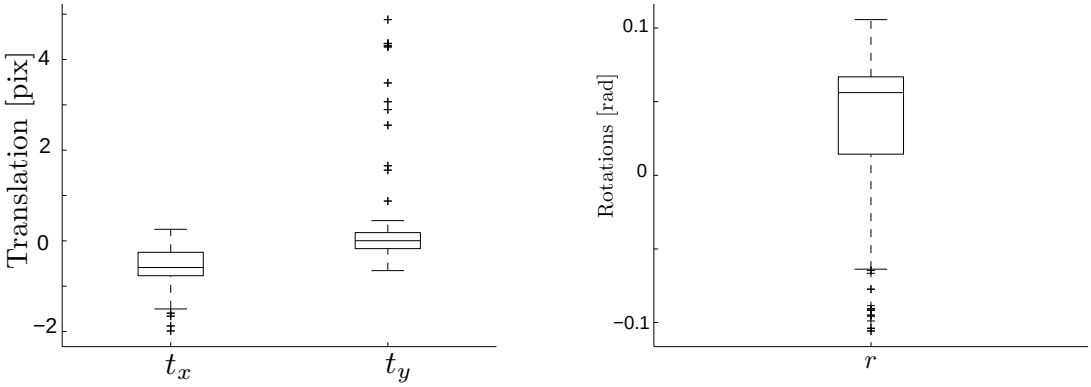


Figure 4.3: Translation (left) and rotation (right) parameters computed from randomly chosen 200 U-transformations estimated with the 2D algorithm applied to the data acquired with the high- and low-field scanners.

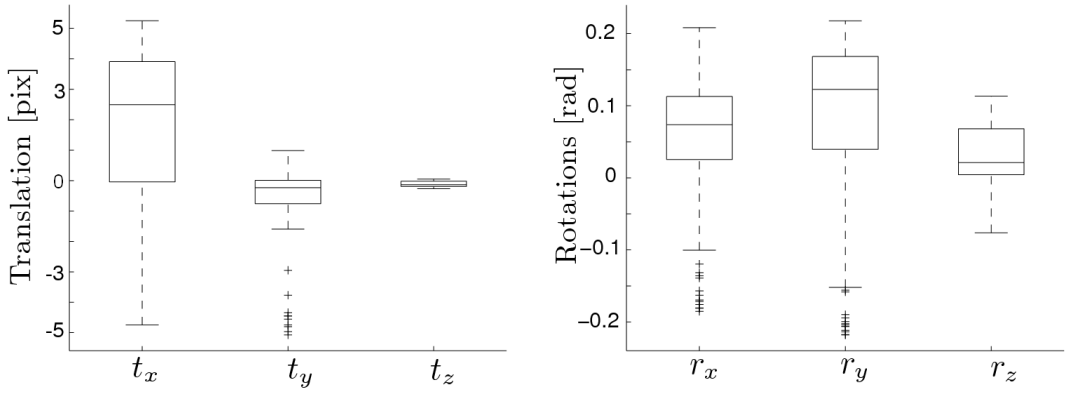


Figure 4.4: Translation (left) and rotation (right) parameters computed from randomly chosen 200 U-transformations estimated with the 3D algorithm applied to the data acquired with the high- and low-field scanners.

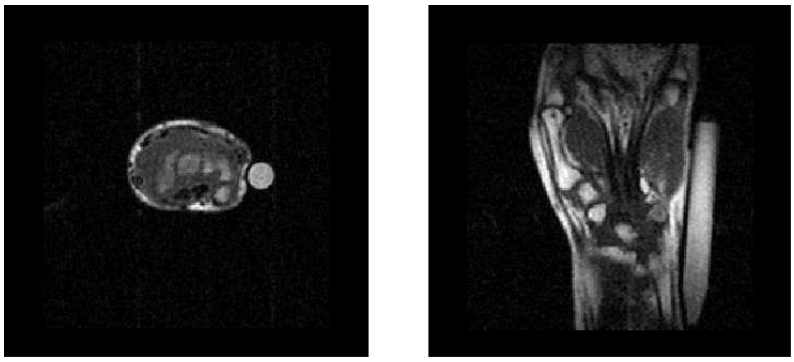


Figure 4.5: Images of joints acquired with the low field scanner in axial (left) and coronal (right) directions.

Figure 4.6: Left: A contour of a marker segmented in the source before (blue) and after (red) registration and superimposed on the target image. Right: Magnified marker.

of bone interiors and markers in the source and target images after registration is expected to be the same.

Images with the marker captured in the axial and coronal directions are shown in Figure 4.5. The change in the marker's position before and after the registration is illustrated in Figure 4.6.

A magnified target with superimposed contours of the joints' exterior and bone interiors segmented in the source image before (red) and after (blue) registration are shown in Figure 4.7. The difference in the location of the edges implies patient hand motion.

The application of the registration algorithm resulted in the accurate alignment of the contours of the joints' envelope and bone interiors in the source and target images. This has increased the mutual overlap from 0.8 to 0.96 in this example.

The same experiment was performed on 200 2D images (100 acquired with the low field, 100 with the high field scanner) and 100 3D volumes (50 acquired with the low field

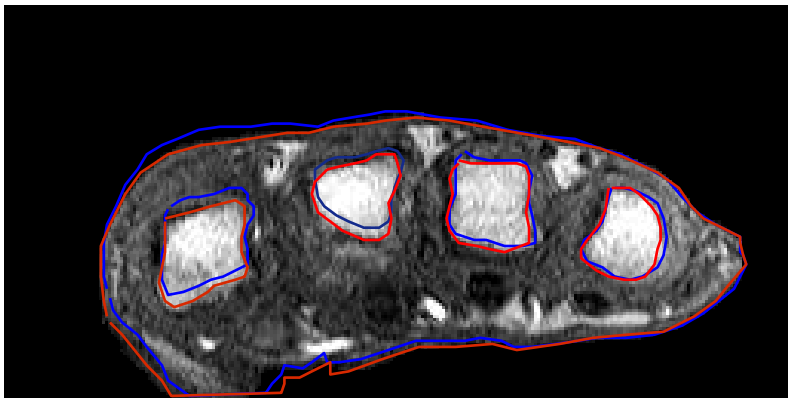


Figure 4.7: Contours of the joints' envelope and bone interiors outlined in the source before (blue) and after (red) registration and superimposed on the target image.

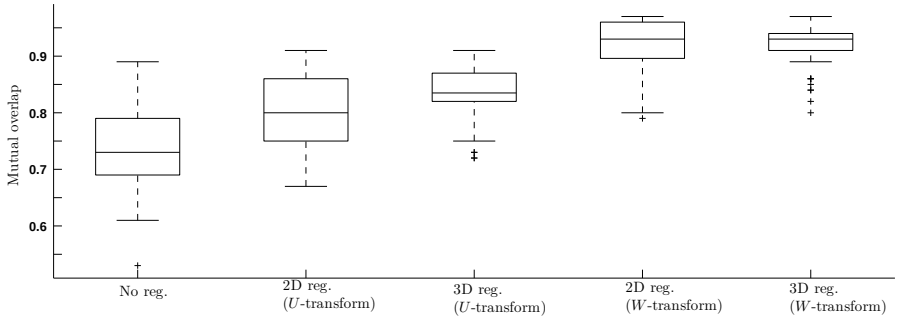


Figure 4.8: From the left: The mutual overlap between the source and target before the registration, after 2D and 3D registration with U-transformation, and after 2D and 3D registration with W-transformation. The same experiment was performed on randomly chosen 200 2D images (100 acquired with the low field, 100 with the high field scanner) and 100 randomly chosen 3D volumes (50 acquired with the low field and 50 with the high field scanner).

and 50 with the high field scanner). These images/volumes were registered with 2D/3D schemes with the U, V, and W-transformations. Figure 4.8 illustrates the results.

The mutual overlap between the markers/bones in the source and target images before the registration was on average 0.74 with the minimum at 0.53 and standard deviation 0.07. After 2D/3D registration with the U-transformation, it became 0.8/0.84 with standard deviations 0.06 and 0.05; after 2D/3D registration with the V-transformation – 0.81/0.86 with the standard deviations 0.04 and 0.04. Finally, when the W-transformation was applied, the mutual overlap became on average 0.91/0.92 with standard deviations 0.04 and 0.03.

This experiment illustrates that 2D and 3D registration schemes permit compensating for the patient motion. The algorithm enhanced with the W-transformation outperformed

the basic and sequential transformations and displayed a further increase in mutual overlap.

4.4.3.3 Comparison of 2D and 3D registration algorithms with various types of transformations

We have chosen a study with apparent in-slice motion and registered DCE-MRI images using 2D and 3D registration algorithms with various transformations. It is of interest to measure the differences in appearance of images after registration has been applied.

Figure 4.9 illustrates the target and source images before (2nd column) and after the 2D and 3D registration with the W-transformation. This study does not exhibit significant intensity change due to the injection of the contrast agent.

Using the 2D registration algorithm with the W-transformation we have estimated that translation in the x direction is on average 3-4 pixels, in the y direction 2-3 pixels, and the rotation 0.0762 radians. 3D registration delivers the following parameters: on average translations in the x, y, z-axis directions are 9, 4, and 2 pixels, the rotations around the x, y, z-axis are 0.06, 0.02, 0.05 radians.

Figure 4.10 illustrates the target images subtracted from the source before the registration (left), after 2D registration with the basic U-transformation (middle), and after 3D registration with the W-transformation (right). It is clear that the registered images contain less noise and the 3D registration of the volumes with the W-transformation allows for significant improvement in the alignment of the skin, inflamed synovial tissues and blood vessels.

To analyse the effect of the registration, the mean square errors (MSE) between the target and source before and after the registration were computed. Ideally, in the absence of the patient motion and contrast agent MSE between the registered images should be zero. However, due to the effect of the contrast agent, MSE between the pre- and post-contrast DCE-MRI images/volumes is always higher than zero, and measures the magnitude of the enhancement. Figures 4.11 illustrates MSE computed for images from the study illustrated in Figure 4.9 before and after the registration with various transformations. The results shown that MSE estimated with 3D registration algorithm enhanced with the W-transformation is the lowest.

To compare the effect of these different transformations, the same experiment was performed on 15 DCE-MRI studies acquired with the high and low field scanners. Only 15 slices (5 datasets) acquired by the low-field scanner had 22 images per slice, that is why this subset of data is used in this experiment. We have randomly chosen 15 slices from the data acquired with the high-field scanner.

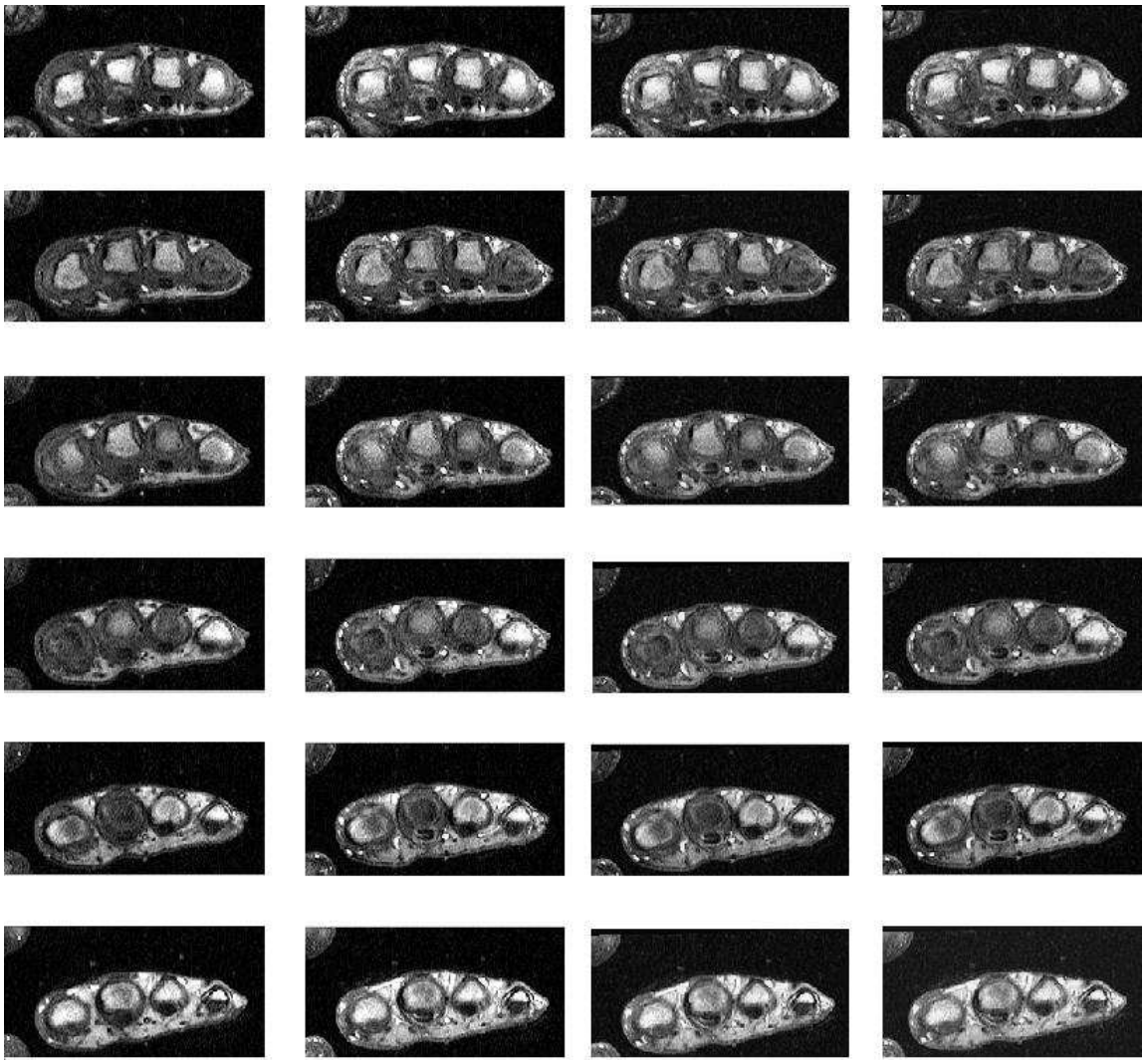


Figure 4.9: From the left: in each column 6 images representing resampled target volume, source volume, source after the 2D registration, source after the 3D registration with W-transformation. Six images in each column correspond to the six temporal slices. Volumes were acquired from the MCPJs with the high field scanner.

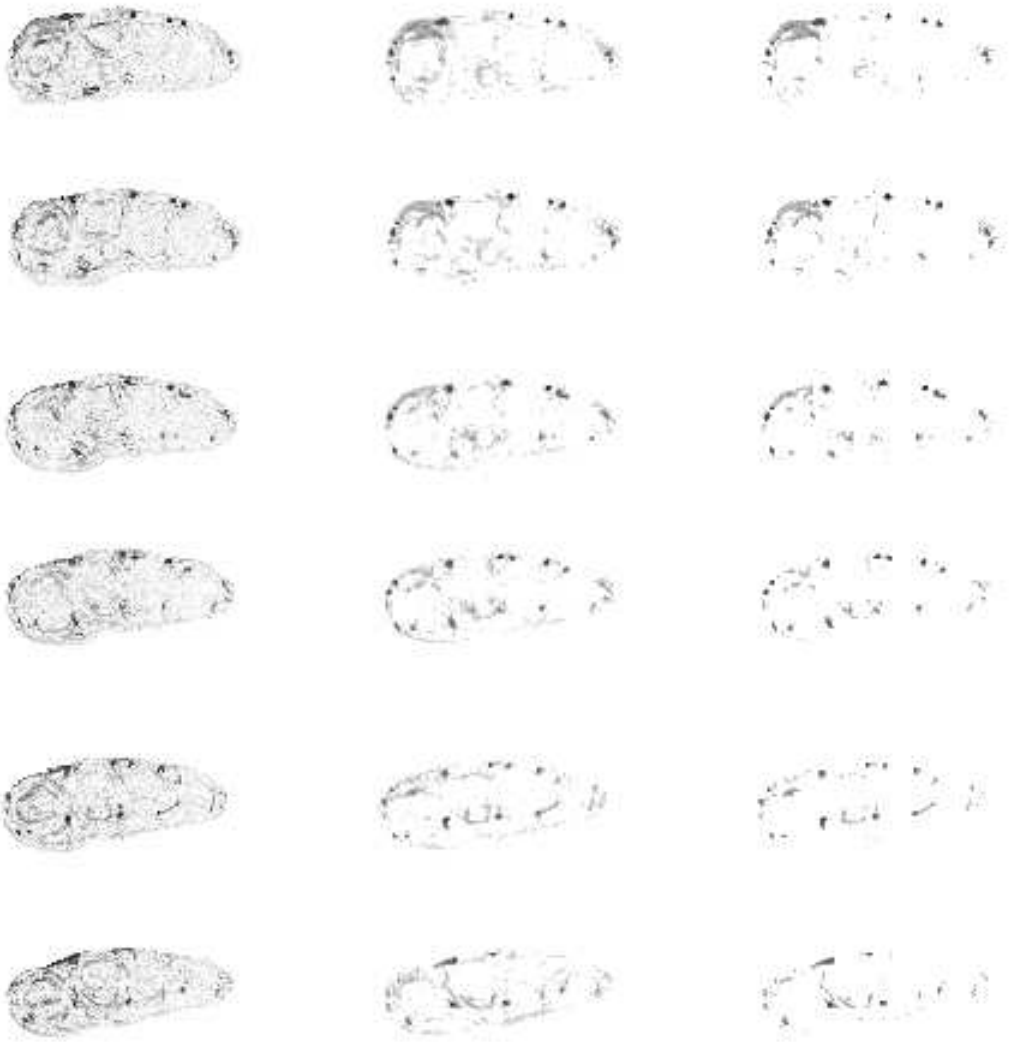


Figure 4.10: Resampled source volume subtracted from the target volume before registration (left); the source volume subtracted from the target volume after 2D registration with W-transformation (middle); the source volume subtracted from the target volume after 3D registration with W-transformation (right). Intensities are inverted. Six images in each column correspond to the six temporal slices. Volumes were acquired from the MCPJs with the high field scanner.

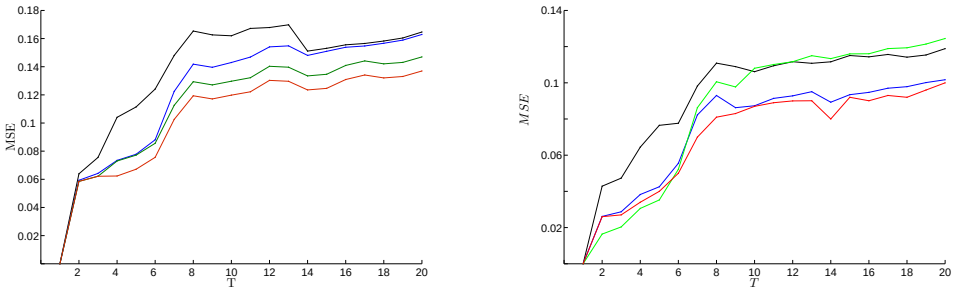


Figure 4.11: MSE computed between the source and target in a sample DCE-MRI slice (left) and a complete DCE-MRI study (right). MSE before the registration is shown in black, after registration with U-transformation in blue, with V-transformation in green, and with W-transformation in red.

The data acquired with the low-field scanner exhibit greater motion. Figure 4.12 illustrates MSE computed between the source and target after registration with the U and W-transformations. Note that in this experiment MSE can be attributed to both patient motion and contrast/intensity change. Let us consider the changes in MSE at each phase of the contrast enhancement:

Baseline: The first 3-5 images/volumes in a study do not exhibit significant contrast variations, and here MSE reflects the patient motion. The results in Figure 4.12 and 4.11 demonstrate that after the images/volumes were aligned MSE has been significantly reduced.

Wash-in, wash-out: The largest variation of the MSE was noticed at the 5^{th}–7^{th} time instants (wash-in) and after the 16^{th} for the high-field and after the 20^{th} for the low-field data (wash-out). At approximately 30s. after the injection of the contrast agent we observe the most significant intensity variation in the data and prominent patient motion. This results in an increase of artefactual enhancement in the data. When images/volumes were aligned and the artefacts due to the patient motion eliminated, MSE was significantly reduced.

Plateau: There is no significant intensity change at the plateau phase, which starts at approximately 12^{th} time instant. The intensity differences between the source and target should be close to a constant, therefore, errors can be attributed to patient motion. After images were aligned, MSE was significantly reduced.

Figure 4.11 demonstrates that MSE reflects behaviour of the enhancement; the shape of the MSE graphs before and after the registration corresponds to the expected change

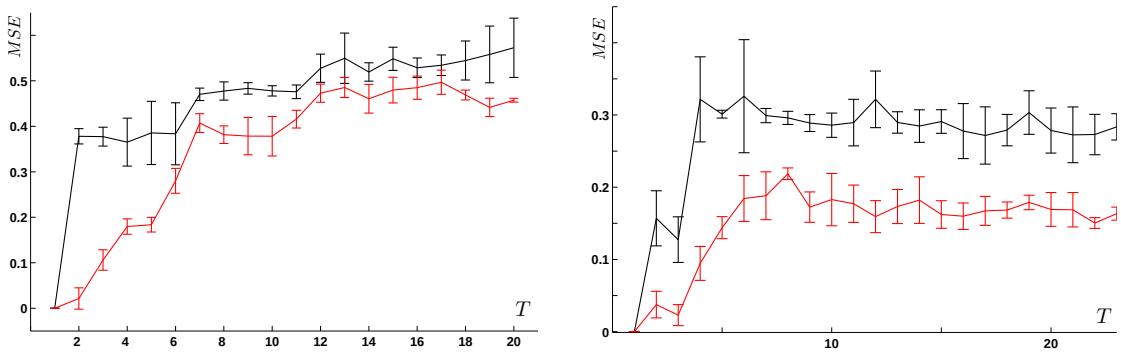


Figure 4.12: MSE computed between the target and source volumes aligned with registration with U-(black) and W-(red) transformations for 15 DCE-MRI studies acquired with the high (left) and low (right) field scanners. The length of the error bars is equal to two standard deviations. The baseline normally occurs between T_1 and T_5, wash-in between T_3-T_7, plateau between $T_7 - T_{16}$, and wash-out after T_{16}.

in intensity. The effect of the registration can be seen at the baseline and plateau phases, where no significant intensity changes are expected, and MSE curves extracted from the non-registered data reflect the artefactual enhancement.

Figure 4.13 illustrates MSE computed for a randomly chosen mix of 10 high and low-field DCE-MRI studies before and after registration with 2D/3D schemes with U and W-transformations. As expected 3D registration delivers more accurate results than 2D. The W-transformation outperformed the U-transformation – on average MSE has decreased from 0.18 to 0.1 when the 2D algorithm was applied with the W-transformation and from 0.12 to 0.08 with the W-transformation. We can conclude that the 3D registration algorithm with the W-transformation has delivered the most accurate results on the DCE-MRI datasets acquired with the high- and low-field scanners.

4.4.4 Visual inspection

To visualise in-slice patient hand movement during the scanning process, we fixed the location of the bone interiors, detected in the first image in a sample DCE-MRI temporal slice, and then plotted the temporal slice (20 images) with these bone interiors superimposed before and after registration.

Obviously, if a patient has moved, location of joints in the first image in the slice will not coincide with the location in other images in the slice. With no movement the location of the joints is the same throughout the slice.

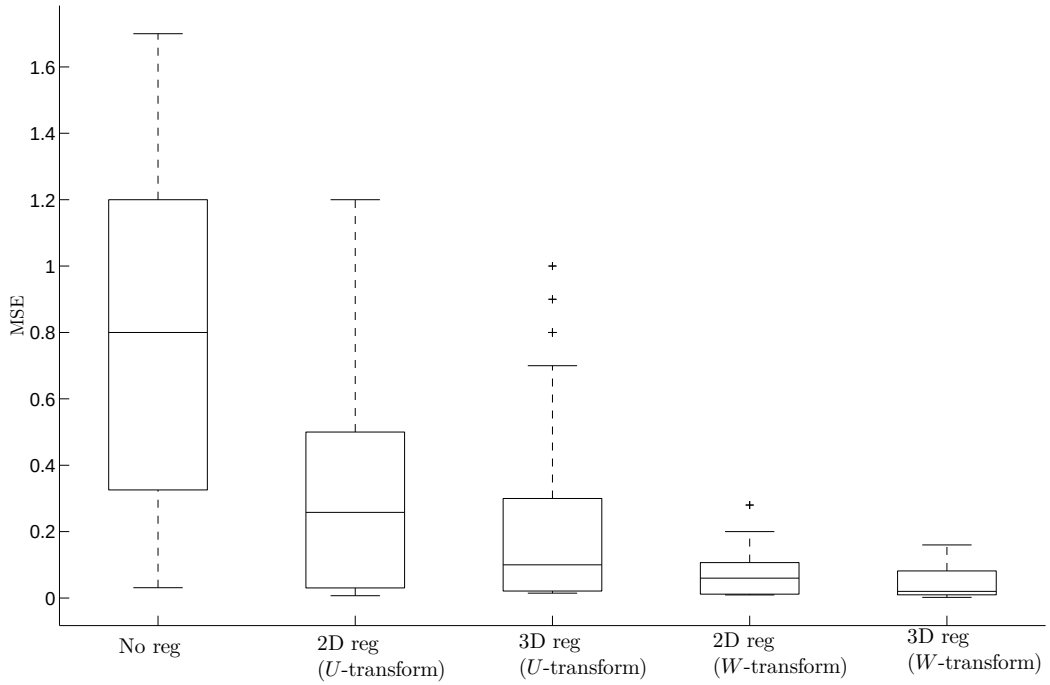


Figure 4.13: MSE computed for a randomly chosen mix of 10 high and low-field DCE-MRI studies before and after registration with 2D/3D schemes. From the left: MSE computed between the target and source images / volumes with 2D / 3D registration approaches with U- and W-transformations.

Magnified images from a DCE-MRI slice are shown in Figure 4.14 (top). The joints, detected in the target and superimposed on each image in the slice, are shown in white. The location of the joints detected in the target does not coincide with the location of the joints in images 3, 4, 5, etc.

Figure 4.14 (bottom) shows the same DCE-MRI temporal slice after 2D registration with the W-transformation. The location of the joints detected in the target has been fixed and superimposed on every image in the slice. The effect of registration is obvious – the bone interiors in the target and each registered source coincide accurately.

To illustrate in-volume motion, we automatically segmented joints' envelopes in the target and source images before and after registration, then subtracted the source volume from the target. Figure 4.15 illustrates the results.

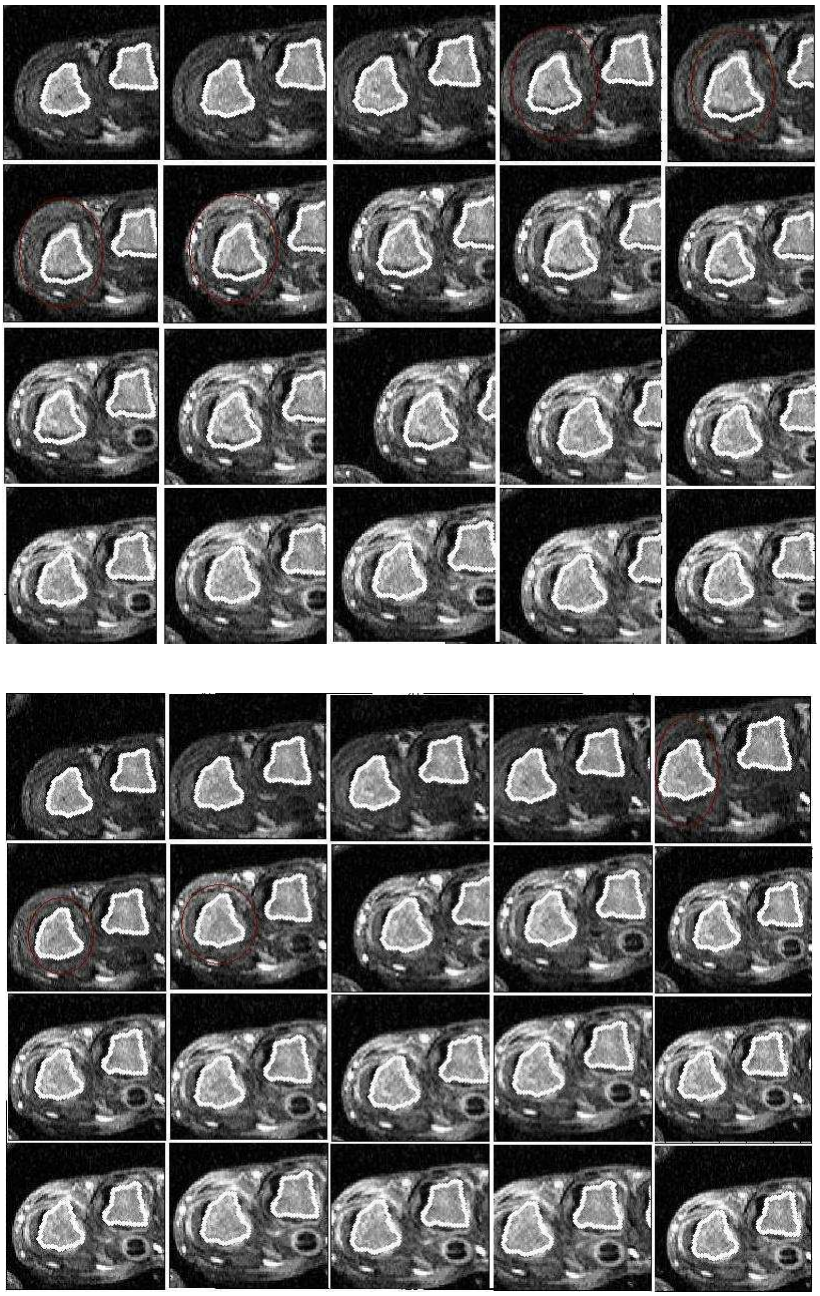


Figure 4.14: Magnified images from a DCE-MRI slice of the MCPJs acquired with the high field scanner before (top) and after (bottom) registration. The joints, detected in the target and superimposed on each image in the temporal slice, are shown in white. before registration, the location of the joints detected in the target does not coincide with the location of the joints in images 3, 4, 5, etc. After registration with W-transformation (bottom), the location of the joints detected in the target coincide with the location of the joints in images 3, 4, 5, etc.

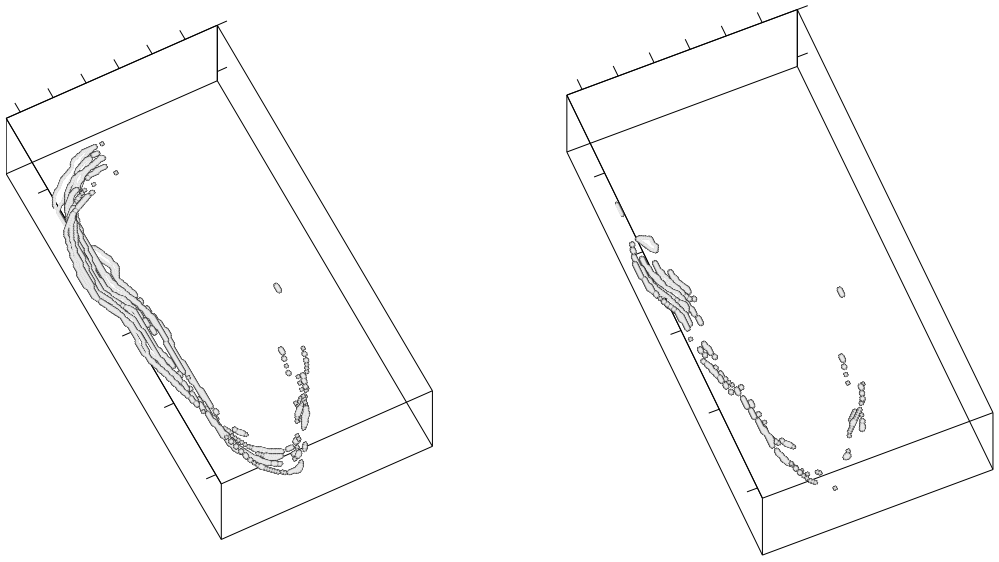


Figure 4.15: Automatically segmented joints' envelops in the target and source volumes of the MCPJs acquired with the high field scanner. The difference between the joint contours from the target and source volumes before (left) and after (right) registration.

4.5 Conclusion

In this chapter we discussed an intensity based registration algorithm, designed to align images and volumes acquired by means of MRI in the presence of the contrast agent. The algorithm is applicable to various DCE-MRI datasets and does not require user interaction. The approach presented here was originated by Periaswamy et al. [184, 186] and is based on optical flow techniques enhanced with an explicit modelling of the contrast and brightness variations.

The original registration algorithm was analysed in detail in [184]: simulations with various combination of the geometric, contrast, and brightness distortions demonstrate its robustness to a wide range of data. Here, we illustrate how to use the algorithm to perform adequately on DCE-MRI datasets acquired with the low and high field scanners.

Firstly, the EM step, which permits registration of partly occluded datasets, was eliminated to reduce computational time. Registration of two 256×256 images with the W-transformation can now be performed in less than 1 minute and two $256 \times 256 \times 6$ volumes in less than 6 minutes using software implemented in MatLab and run on a 2.79 GHz Windows machine with 1GB memory. In comparison, registration of two volumes with the original approach on the same machine requires 30 minutes.

Secondly, the algorithm has been augmented with an incremental approach that allows us to take into account significant variations in the contrast and brightness that occur in the post-contrast images/volumes from the dynamic datasets.

2D and 3D registration schemes with various transformations (U, V, and W) were applied to a large number of images/volumes. The results show that the 3D registration scheme permits better alignment than 2D as it compensates for in-study rather that in-slice motion. The incremental approach used for image/volume alignment (W-transformation) allows for further reduction of registration error.

Our results demonstrate that the smoothness constraint added to the error function does not influence significantly the algorithm's performance when recovering geometric distortions. However, it permits compensation for the error due to the local contrast and brightness variations in the soft tissues.

Registration of the images/volumes allows for significant improvement in the location of the blood vessels, bone interiors, skin, but most importantly synovial tissue. The reduced artefactual enhancement contributes to the data fidelity, which is crucial for further data analysis with quantitative approaches.

Chapter 5

Image segmentation

The inflamed synovium, which is a tissue of particular interest, is located inside the joints' envelope and outside the bone cross-sections. Therefore, in order to analyse a DCE-MRI dataset efficiently tissues within the bone interiors, which include both cortical and trabecular bone as well as bone marrow, markers, and thumbs need to be segmented out. In this chapter we introduce two algorithms: the first one for segmentation of the joints' envelopes in the hand and wrist datasets acquired with the low and high field scanners, and the second for segmentation of the bone interiors in DCE-MRI slices of the MCPJs acquired with the high field scanner.

Section 5.1 discusses an algorithm for segmentation of the joints' envelopes in DCE-MRI dynamic frames acquired from hand or wrist in the axial or coronal directions with the high and low field scanners. Section 5.2 introduces an algorithm for automatic segmentation of the rigid boundaries of the bone structures in DCE-MRI images of the MCPJs acquired with the high field scanner. The algorithm will be described as a two-step procedure. Firstly, the bone interiors are crudely located using a global thresholding technique and prior information about geometry of the MCP joints. Secondly, the boundaries of the detected regions are refined with an adaptive segmentation technique.

5.1 Segmentation of joint envelope

Some images acquired with the low field scanner contain a marker located close to the joints' envelope; images acquired with the high field scanner might depict a thumb (see Figures 3.5 and 3.8). Based on the prior knowledge about the anatomy of the joints and positioning of the patient during the data acquisition (see Section 3.2), the following assumptions about the marker and thumb size and location can be made:

- Empirically, it was estimated that the area of a thumb or marker in respect to the area of the joints' envelope is between 2% and 15%. Figure 5.1 illustrates the distribution of the size of thumb and marker manually outlined in 50 dynamic frames.
- The location of the joints' envelope within the imaging matrix is always approximately central.
- The location of a thumb in images acquired with the high field scanner in respect to the joints' envelope varies, but the thumb is always located in either the left or right corner of the imaging matrix. This is represented by one or two regions and might osculate with the joints' envelope. Figure 3.8 illustrates the physical positioning of the thumb and the corresponding MR images.
- The marker is located centrally, and is not attached to the joints' envelope.

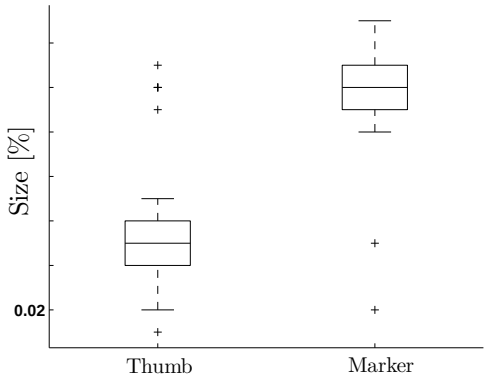


Figure 5.1: Size of the thumb and marker in proportion to the area of the joints' envelope.

Each dynamic frame from a DCE-MRI dataset is composed of a light foreground and a dark background in such a way that the regions of interest and background pixels have intensity levels grouped into two modes. An obvious way to extract the regions of interest

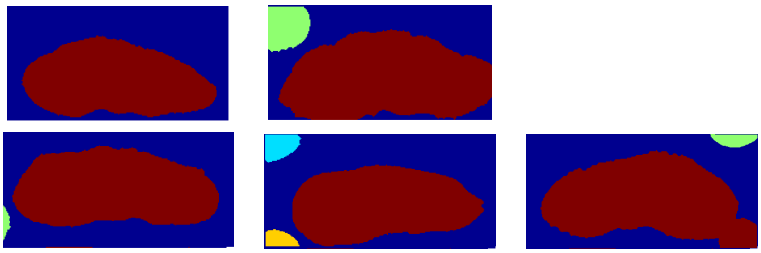


Figure 5.2: The result of the global thresholding on the images acquired with the high field scanner. The joints' envelope is shown in red.

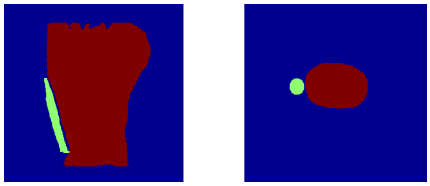


Figure 5.3: The result of the thresholding on the images acquired with the low field scanner. The joints' envelope is shown in red.

from the background is to select a threshold value that separates these modes. This has been done using a thresholding algorithm [200] with the threshold estimated in an iterative manner[1].

This allows the isolation of the joints' envelopes and markers / thumbs from the background. We remove the noise obtained as a result of this thresholding with morphological opening (a circle with the diameter of 2 is taken as a structural element). The result of this operation on several images is shown in Figures 5.2 and 5.3 (different colours indicate non-osculating (i.e. non-intersecting) regions, the joints' envelopes are shown in red).

After performing this, we are left with 1, 2 or 3 regions, one of which represents the joint envelope. The marker in the images acquired with the low field scanner never osculates with the joint envelope. Therefore, after the thresholding, we segment out the region with the smallest area.

Segmentation of a thumb, which might osculate with the joints' envelope, is not straightforward. To segment a thumb, an imaging matrix was firstly divided into four equal parts. Then the number and location of the region / regions in each quartile were analysed.

[1]Thresholding algorithm proceeds as follows. The histogram of pixel intensities from pre-contrast image is initially segmented into two parts using a starting threshold value such as a half the maximum intensity range. Then the sample mean of the grey values associated with the foreground pixels and the sample mean of the grey values associated with the background pixels are computed. A new threshold value is now computed as the average of these two sample means. The process is repeated, based upon the new threshold, until the threshold value does not change any more.

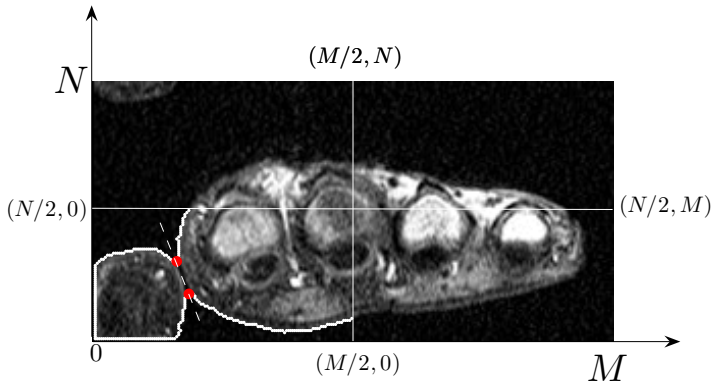


Figure 5.4: Segmentation of a thumb in the imaging matrix divided into four parts. Boundary of the joints' envelope with an osculating thumb superimposed on the intensity image. Points of maximum curvature are shown in red.

By considering the coordinates of the region / regions in each quartile, it was estimated whether or not the boundaries of the imaging matrix and region intersect. A quartile of the imaging matrix might contain:

1. Two non-osculating regions, boundaries of the imaging matrix and one of the regions intersect (upper left quartile of the image in Figure 5.4).

2. One region and its boundary intersects with the boundary of the imaging matrix (lower left quartile of the image in Figure 5.4).

3. One region and its boundary does not intersect with the boundary of the imaging matrix (upper right and low right quartiles of the image in Figure 5.4).

In the first and last cases, we take a region with the largest area as the part of the joint's envelope. In the second case, we assume that the thumb is osculating with the joint's envelope.

In the second case, to remove the thumb, firstly, two points of maximum curvature are found in the boundary of the region [144]. Such points are found as the two maxima; smoothing is used to ensure that multiple maxima are not detected at each pinch point. The points are shown in red in Figure 5.4. Note that we only need to consider a part of the region's boundary that does not intersect with the boundary of the imaging matrix. Secondly, these points are connected with a straight line, the region is split into two along the line, and the largest part of the region is selected as a part of the joints' envelope.

5.2 Segmentation of bone interiors

5.2.1 Preliminary segmentation

Within the joints' envelope we wish to distinguish bone, muscle, blood vessel, and synovial tissue. In pre-contrast images of the MCP joints, the intensity of the bone interiors and some blood vessels is high, whereas all other tissues appear dark. The bright and dark tissues within the joints' envelope can be separated with the iterative global thresholding technique discussed earlier. The results obtained for a sample DCE-MRI pre-contrast image are shown in Figure 5.5.

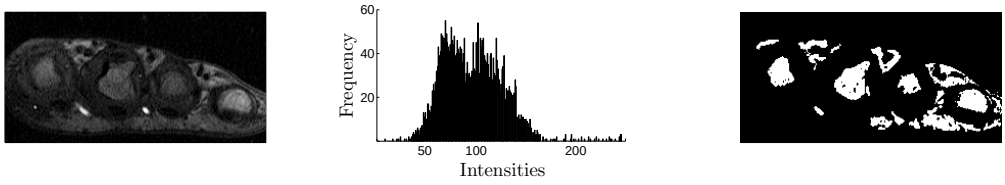


Figure 5.5: Left: Pre-contrast image of the MCP joints acquired with the high-field scanner. Middle: The histogram of the intensity values (assigned threshold value is 89). Right: The results of the global thresholding.

The imaging procedure and affect of the contrast agent cause bone interior intensity to change in images within a temporal slice, therefore, the thresholding might depict different parts of a bone in different images. In order to receive comprehensive information about a bone, it is beneficial to perform thresholding on several images within a temporal slice.

In the post-contrast dynamic frames, the synovial tissue surrounding bones enhances significantly. Therefore, the thresholding often classifies the synovium as a part of a bone, which prevents accurate segmentation of the bone interior. Figure 5.6 illustrates pre- and post-contrast dynamic frames with superimposed results of the thresholding.

Empirically, it was found that the effect of a contrast agent is not prominent in the first four images in a temporal slice. In our experiments these images were thresholded. Then, to distinguish between bone / non-bone regions obtained as the result of the thresholding, we used prior geometric information about the bones' position within the dynamic frames of the MCP joints.

For each MCP joint (2^{nd}–5^{th}) in the normalised coordinate system using prior information acquired earlier, we have defined a margin where it is presumed to appear. These margins are shown in the dashed line in Figure 5.7. Then the area of a rectangular was

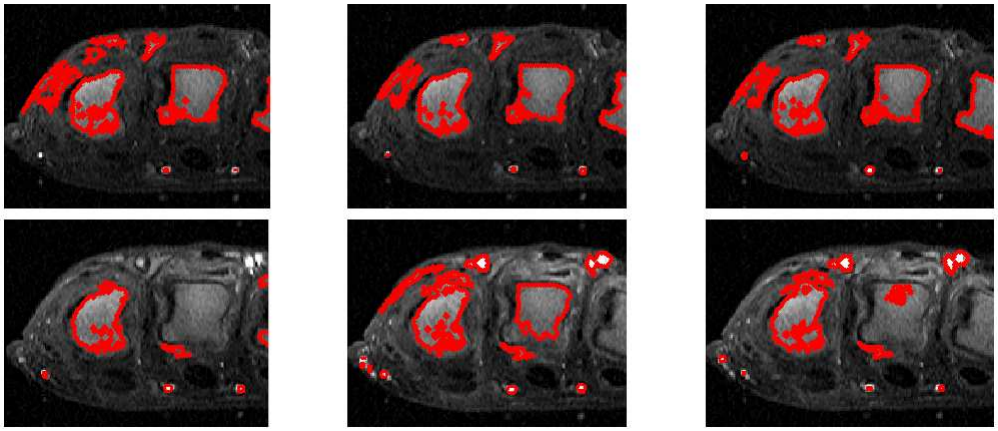


Figure 5.6: Pre- (top) and post- (bottom) contrast images with superimposed results of the thresholding. Boundaries of the detected regions are shown in red.

widened by adding 2 pixels to each side to allow some error. Final intervals in the normalised coordinate system for each bone are shown in solid line in Figure 5.7 (left). The result of the thresholding for a sample DCE-MRI dynamic frame is shown in Figure 5.7 (right).

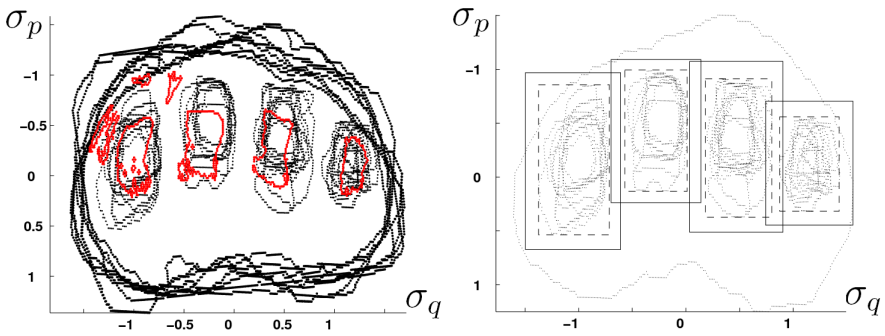


Figure 5.7: Left: Intervals where the MCP joints are presumed to appear (rectangular) in the normalised coordinate system. Right: Regions obtained as the result of the thresholding for a sample image in the normalised coordinate system. σ_p and σ_q are standard deviation of p_i and q_i defined in Section 3.3.3.

Regions that appear within the pre-defined interval are selected. To eliminate noise, the area of each selected region is required to be larger than 2% of the joints' envelope (it was empirically found that the size of a joint's interior is between 2% and 5% of the envelope). For each joint, all regions obtained as a result of this operation on the pre-contrast images are superimposed; a convex hull was drawn around them and taken as a 'segmentation mask'. Figure 5.8 illustrates the results obtained for several temporal slices.

It is, of course, unlikely that the detected boundaries will coincide precisely with the

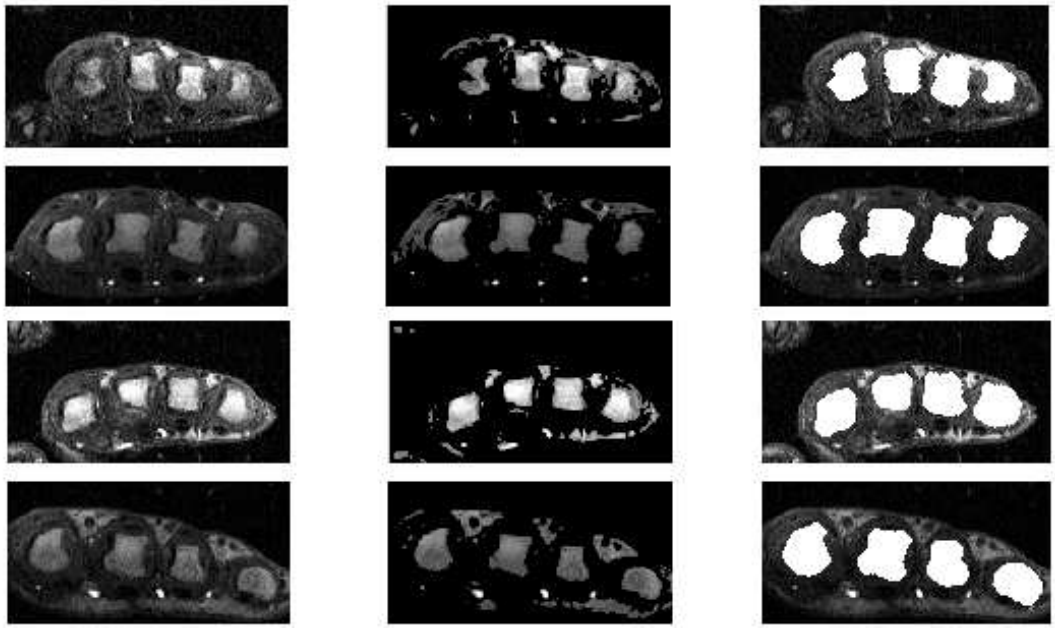


Figure 5.8: Left: Pre-contrast dynamic frames from different DCE-MRI studies acquired by the high-field scanner. Middle: The result of the global thresholding (shown as an intensity image). Right: The final segmentation mask (in white) superimposed on the pre-contrast image.

true boundary of the bones, although, due to the registration, the location of the final mask has been significantly improved. At this stage we will tolerate the results to be inaccurate.

5.2.2 Adaptive segmentation

The purpose of adaptive segmentation is to refine boundaries of the regions detected with the preliminary segmentation. The segmentation starts by locating the centroid of the segmentation mask, and determining the diagonal length of its bounding box. The normals equal to half this diagonal are drawn to a boundary pixel from inside and outside the region's boundary.

For each boundary pixel, along the normals we consider the image pixel intensity profile. Figure 5.9 (right) illustrates a profile obtained for a boundary pixel.

Two types of the intensity profiles have been noticed. If there are no artefacts and a boundary is one-pixel width, the profile looks like the one in Figure 5.10 (middle). If the bone is surrounded by the blood vessels or other artifacts the profile might look like the one in Figure 5.10 (right).

The underlying model here assumes that the boundary pixel should separate a bright

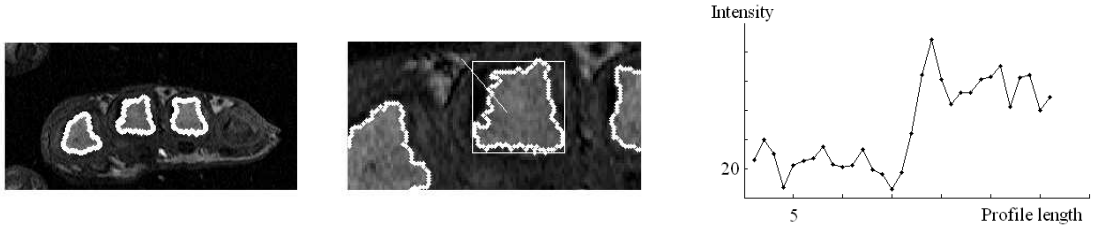


Figure 5.9: Left: A pre-contrast image with the segmented bone interiors. Middle: Magnified bone interior, its bounding box and two normals. Right: Corresponding intensity profile.

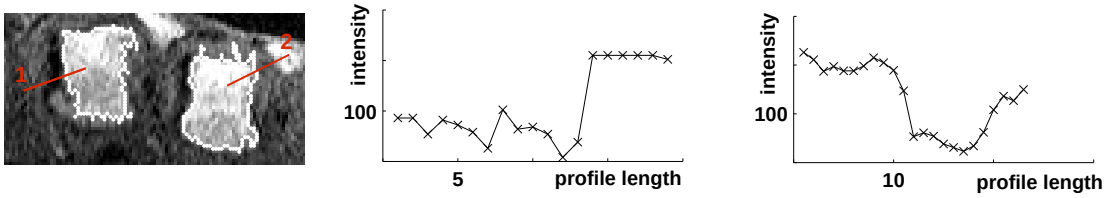


Figure 5.10: Left: Magnified results of the preliminary segmentations superimposed on the MR image with the normals drawn to different boundary pixels. Middle: Intensity profile corresponding to 1. Right: Intensity profile corresponding to 2.

inner area (bone) from a darker outer area. The length of an intensity profile is assumed to be long enough to cross the actual boundary of a bone.

By finding an optimal segmentation of the profile, we shrink or extend the mask obtained as the result of the preliminary segmentation along the normals with the aim of approaching the true boundary of the region. The problem of the boundary refinement can therefore be solved by using an efficient 1D signal segmentation technique discussed in the next section [2]

5.2.3 Signal segmentation

The procedure starts by modelling each intensity response as a sequence of segments formed by pixels of equal intensity. A sharp change between bright and dark areas indicates the 'optimum' boundary pixel. Here we assume the uniformity of the intensities in the vicinity of the region's boundary. If a current boundary pixel does not coincide

[2]This work was done in collaboration with Mike Pyatnizkiy, Biophysics, Russian State University, Moscow, Russia.

with the optimum boundary pixel, the boundary along the normal is moved towards the optimum boundary pixel.

A minimisation with the minimum least squares (MLS) algorithm [86], which minimises the sum of the squared deviations of the signal segments, has been chosen to analyse the signal.

Let θ denote a signal, which consists of N components $\{x_1, x_2, ..., x_N\}$. We use the notation $\xi = \theta(b, e)$ to define a segment of the signal θ, where $b \leq e$; b denotes the beginning and e denotes the end of the segment; and $b, e \in \{x_1, x_2, ..., x_N\}$. If $\xi_1 = \theta(b_1, e_1)$ and $\xi_2 = \theta(b_2, e_2)$, where $b_2 = e_{1+1}$, are two segments, then $\xi_1\xi_2 = \theta(b_1, e_2)$ denotes their concatenation.

A 1D segmentation S of θ into k segments is a sequence $\{\xi_1\xi_2...\xi_k\}$ of k segments such that $\xi_1\xi_2...\xi_k = \theta$ and each ξ_i is non-empty. We are interested in obtaining the segmentations of θ, where the segments are internally homogeneous.

In order to formalise this goal, we associate a cost function F with the internal heterogeneity of individual segments, and aim to minimise the overall cost of the segmentation.

Two assumptions have been made on the overall cost. Firstly, the cost of a single segment $F(\xi_i)$ is a function of data points $C(F(\xi_i))$. Secondly, the cost of 1D segmentation $C(F(\xi_1, \xi_2, ..., \xi_k))$ is a sum of the cost of its segments $\xi_1,\ \xi_2, ...,\ \xi_k$:

$$C(F(\xi_1, \xi_2, ..., \xi_k)) = \sum_{i=1}^{k} C(F(\xi_i)) \tag{5.1}$$

An optimal 1D segmentation of a signal using cost function F is such that the cost $C(F(\xi_1, \xi_2, ..., \xi_k))$ among all possible 1D segmentations larger than 1 segment per pixel is minimal. The cost of any segment can be defined as follows:

$$C(F(\xi_i)) = \frac{1}{N} \sum_{j=b_i}^{e_i} \left| x_j - \mu_i \right|^2, \tag{5.2}$$

where b_i is the beginning of the segment e_i is the end of the segment; μ_i is mean value of samples in the segment $\xi_i = \theta(b_i; e_i)$. More formally, the mean of the segment is defined as follows:

$$\mu_i = \frac{1}{e_i - b_i + 1} \sum_{z=b_i}^{e_i} x_z \tag{5.3}$$

In order to minimise the overall cost of the segmentation we have to find such a set of k segments $\xi_1, \xi_2, ..., \xi_k$ that provides the minimum sum of corresponding costs. Therefore, we seek $\min_{e_1, e_2, ..., e_{k-1}} C(F(\xi_1, \xi_2, ..., \xi_k))$.

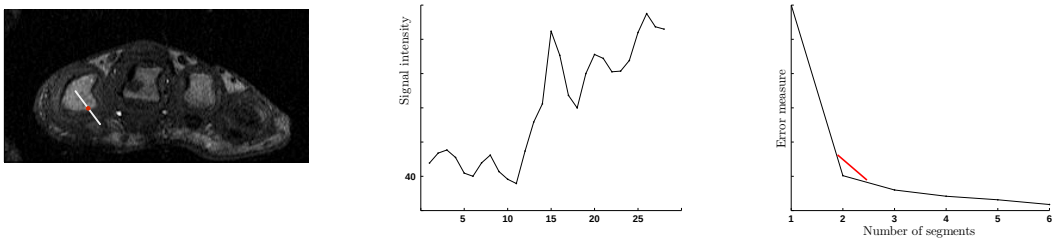


Figure 5.11: Left: DCE-MRI dynamic frame with normals drawn towards a boundary pixel. The point of the segmentation of the profile chosen by the algorithm is marked red. Middle: Corresponding intensity profile. Right: Corresponding graph. The minimal angle is shown in red.

The algorithm describes the best approximation of an intensity profile with 1, 2, 3, etc. straight lines. Now we need to determine the point on the intensity profile that corresponds to the true boundary pixel.

By splitting and merging segments we construct a graph, where the x-axis is the number of segments and the y-axis is the value of the cost function measuring the error of performing a segmentation at k segments. At some point, splitting data into a larger number of segments does not significantly change the value of the error function. This point corresponds to the optimal number of segments.

There is a number of approaches to estimate this point. For our experiments, we have adapted the one proposed by Salvador et al. [212]. Two straight lines are fitted in the data from the left and right sides; each line should include at least two pixels and together lines should cover all pixels. The fit where the angle between the lines is minimal is taken as a desired point. The intensity profile is then segmented at the point of the best fit of the two segments.

Figure 5.11 illustrates a profile corresponding to one of the boundary pixels and its segmentation into two segments. Signal profiles along the normals drawn to each boundary pixel were analysed and segment break points located, allowing a suitable shift in the estimate for the true boundary pixel.

5.2.4 Final steps

The boundary of a region might span over several pixels or be surrounded by a relatively thick cartilage, which appears dark in the MR images. In this case, a profile crossing cartilage, artefact or an ambiguous part of the boundary, will be segmented into more than two segments and therefore more than one potential boundary point will be obtained.

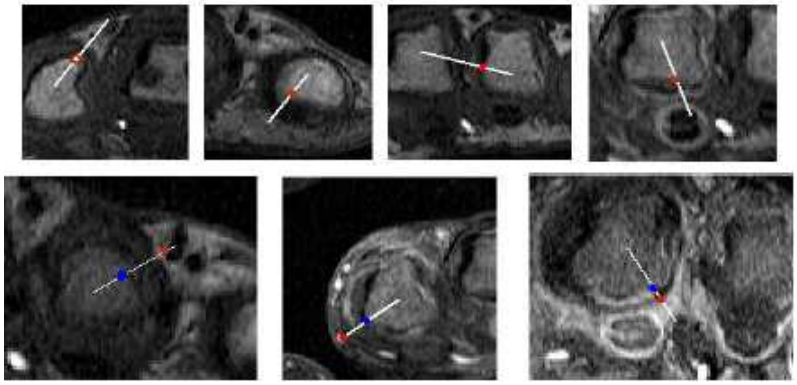


Figure 5.12: Results of the adaptive segmentation, when a profile crosses cartilage, artefact or an ambiguous part of the boundary and is segmented into more than two segments and therefore more than one potential boundary point will be obtained. Break points superimposed on the profiles are shown in red and blue.

Figure 5.12 illustrates several profiles, where segmentation delivers one (top row) and several (bottom row) points.

In order to restore the boundary of a bone, we employed the following strategy. Firstly, all profiles segmented into more than two segments were discarded. Empirically, it was noticed that location of these profiles is sparse, and the number of accepted pixels significantly exceeds the number of rejected pixels. Figure 5.13 illustrates two regions with accepted (in yellow) and rejected (in blue) boundary pixels.

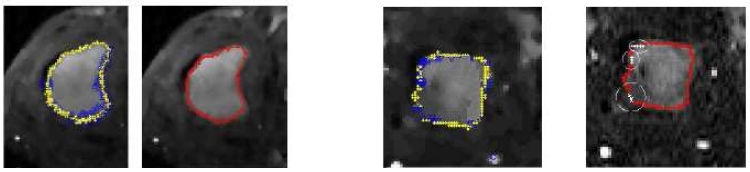


Figure 5.13: Boundaries with accepted (yellow) and rejected (blue) results of the adaptive segmentation. Results of the linear interpolation are shown in red, and the gaps closed with shortest path in white.

After exclusion of the pixels, a boundary might contain small (1 missing pixel) and big (order of 10 pixels) gaps. The location of the isolated missing pixels was restored with an interpolation technique applied to the nearest neighbours, and the gaps were eliminated by connecting pixels with the shortest path (straight line).

Sometimes, the shortest path does not yield the optimal solution. A bone in Figure 5.14 contains an erosion (dark line in the bottom left corner). A gap in the bone's boundary was closed with the shortest path, which cut off a part of the bone.

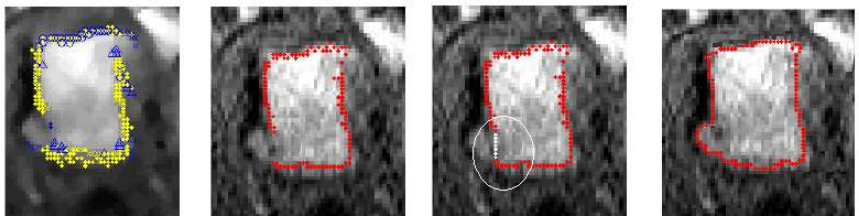


Figure 5.14: From the left: results of the adaptive segmentation (yellow – accepted, blue – rejected); the boundary after the interpolation; the boundary with a gap closed with the shortest path; the final boundary after the in-study interpolation.

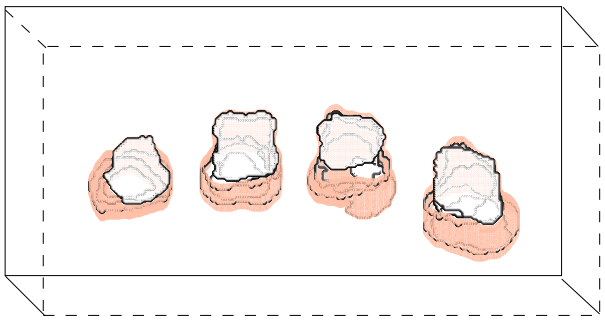


Figure 5.15: Depiction of the 3D bones in a sample DCE-MRI study. The morphology-based three-dimensional interpolation scheme [137] was applied on the six slices to produce this result.

The shape of such a region can be restored with an interpolation technique applied on 3D study. The morphology-based three-dimensional interpolation scheme [137] was applied on the six slices to produce accurate and smooth intermediate slices and volumetric data between the neighboring slices. Figure 5.15 illustrates the results for a sample study. A restored boundary of the bone interior is shown in Figure 5.14 (right).

The output of the algorithm on several regions is shown in Figure 5.16, where the results of the preliminary segmentation are shown in white, and the adaptive segmentation in red. Limitations and possible extension of this approach are discussed in Chapter 9.

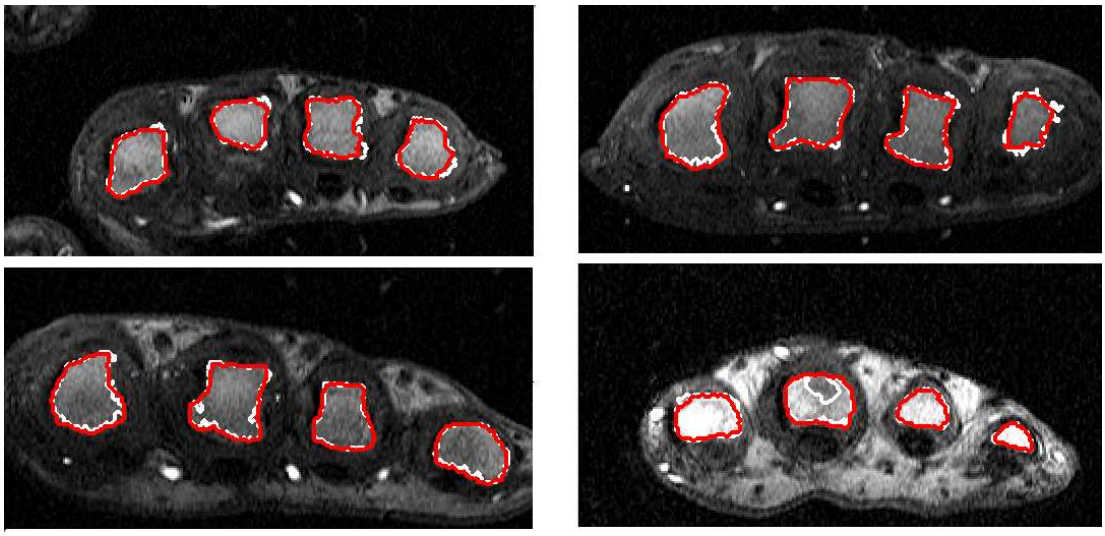


Figure 5.16: The bone interiors segmented with the preliminary segmentation (in white), and the adaptive segmentation (in red) in images drawn from four DCE-MRI studies acquired with the high-field scanner.

5.3 Experiments

5.3.1 Segmentation of joint envelope

In DCE-MRI datasets acquired with the high field scanner, 12 joints' envelopes osculate with the thumb, in 26 slices the thumb is depicted, but does not touch the joints' envelope, and 22 slices do not contain the thumb. 18 DCE-MRI datasets (54 temporal slices) acquired with the low field scanner contain a marker located in axial direction in 6 slices, and in coronal - in 48 slices; 10 datasets have no marker.

In dynamic frames depicting a thumb osculating with the joints' envelope, an expert[3] manually outlined the location of the envelope excluding the thumb using software ANALYZE [202]. Mutual overlap between the automatically segmented joints and ground truth (GT) overlay is on average 0.94, with the minimum 0.93, the maximum 0.97.

The mutual overlap between the manually and automatically segmented joints' envelopes in images with no thumb / marker is on average 0.96, with the minimum at 0.94 and the maximum at 0.97. The error of course might be attributed to the variability of the human observer's opinion on the precise location of the joints' boundary.

Figures 5.17 and 5.18 illustrate joints' envelopes segmented automatically (in white) and manually (in red). We asked three independent observers to outline the GT boundary

[3]This evaluation was performed by an experienced observer from the Department of Medical Physics, University of Leeds, Leeds General Infirmary, Leeds, UK.

for these regions and took an average of these subjectively defined boundaries as the final GT boundary [46, 263].

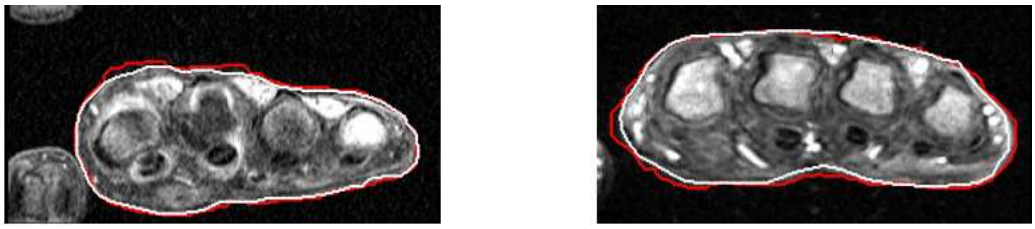


Figure 5.17: Post-contrast images from different DCE-MRI studies acquired with the high-field scanner, with the joints' envelopes segmented automatically (white) and manually (red).

Figure 5.18: Post-contrast images from different DCE-MRI studies acquired with the low-field scanner, with the joints' envelopes segmented automatically (white) and manually (red).

5.3.2 Segmentation of bone interiors

Recalling that in some images not all four bones will be observable, there are two separate evaluations to be performed. Firstly, it should be determined with what reliability we can judge whether or not a bone is present. An experienced observer has provided the ground truth which is in most cases a 'yes/no' judgement (that is, each of the joints 2–5 is or is not observable). In a small number of cases, the expert was unsure. We are able to compare our results with this clinical judgement.

Figure 5.19 (bottom row) illustrates the regions where the expert was unsure whether a bone interior is present or not. It is a challenging task to distinguish between the cartilage

and fat and water inside the actual bone, since in some images due to the data acquisition procedure the difference in intensities is subtle.

The algorithm has been tested on 10 DCE-MRI datasets; and bones in each of the 60 slices (6 per dataset) have been detected. 92% of the existing bones have been found. There were no false positive bones. If due to the objective reasons mentioned above a bone is missing, the algorithm is not expected to detect anything in this location. Table B.1 in Appendix B summarises the results.

In just a few cases the algorithm fails to distinguish bone from cartilage. Examples are shown in Figure 5.19. The bones, which have not been detected are of a small size. Some of them were removed at the thresholding stage, others – by the classifier, which was seeking the regions with a larger area. Table B.1 in Appendix B illustrates that in the slices spatially neighbouring a slice with the missing bones all joints were detected. This allows reconstruction of the under-segmented regions.

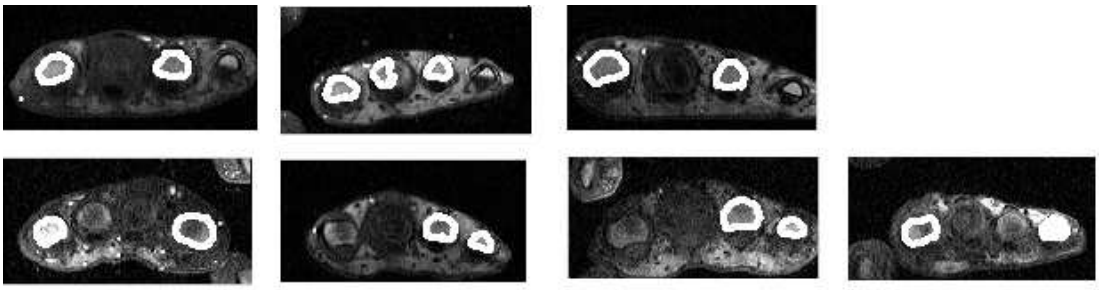


Figure 5.19: Sample images from different DCE-MRI studies acquired with the high-field scanner. Some bones in these regions were not detected by the algorithm. Contours of the detected bones are shown in white.

Secondly, the quality of the segmented boundaries should be evaluated. The best possible boundaries in more than a hundred images drawn randomly from the datasets were manually outlined, and the results of the segmentation were compared to what we will assume is the ‘correct’ solution.

The mutual overlap between the results delivered by the algorithm and the manually outlined contours exceeds 0.90. Detailed quantitative evaluation of the quality of segmentation outputs located with the adaptive segmentation technique will be discussed in the next chapter.

5.4 Conclusion

Two segmentation algorithms were presented in this chapter. The first one permits accurate segmentation of the joints' envelopes in the dynamic frames acquired with the low and high field scanners, and the second one – the bone interiors in the MCPJs images acquired with the high field scanner. The second algorithm has not been tried on data acquired by the low-field scanner.

Segmentation of the joints' envelopes was complicated by the presence of the thumb / marker whose location in respect to the joints' envelope varied. In cases when the thumb and joints' envelop osculated, the thumb was removed based on the assumption that the curvature in the points of osculation is maximal. The mutual overlap between the results delivered by the algorithm and the manually outlined contours exceeds 0.93.

Segmentation of the regions of interest such as bones in DCE-MRI data is challenging mainly because the borders of the bones do not have a continuous high gradient value. Instead, areas with very high gradient values can appear within the joints or the borders might have gaps with the low gradient values. Moreover, the shape and size of bones vary between the studies.

The algorithm discussed here detects the bone interiors and then refines their boundaries using the adaptive segmentation approach. 92% of existing bones were correctly identified with the preliminary segmentation.

The adaptive segmentation does not require a boundary to be smooth or of a constant width. It considers each boundary pixel independently and refines its position using an intensity profile that reflects the intensity change in the vicinity of this boundary pixel. No constraints were imposed on the shape or location of the region obtained as a result of the preliminary segmentation. This is beneficial for segmentation of the bones, whose shape was corrupted by partial erosions and oedema. The interior of such bones usually contains sharp edges which are of clinical interest. Chapter 6 discusses performance of the algorithm in more details and compares it with region growing and snakes.

Chapter 6

Evaluation techniques

This chapter pursues two purposes. Firstly, we discuss supervised and unsupervised evaluation metrics: show how to enhance an existing supervised approach, introduce a new unsupervised metric, and compare new and existing metrics' performance. Secondly, we use these new and old metrics to evaluate the adaptive segmentation discussed in Chapter 5, and to compare its performance with snakes and region growing [87, 122].

6.1 An enhanced supervised metric

In Chapter 2 we illustrated that a widely used and accepted supervised metric, based on mutual overlap (MO) between ground truth and segmented regions, is not fully appropriate for DCE-MRI data. The performance of recently developed algorithms is often evaluated with a Hausdorff distance (HD) based metric [271]. However, this metric does not afford an application-adaptable threshold for a degree of tolerance in segmentation error, and therefore cannot deal with local blur, partial volume effects and ambiguity of region boundaries, or inconsistency of human expert judgments.

The absence of a notion of 'tolerance' is problematic: should two algorithms be evaluated as similar by a metric it is possible that their actual performance is significantly different (in some ways), and this can be revealed by considering them with respect to, for example, a tolerance parameter. Often, edge detectors may deliver partial boundaries whose value, or lack of value, needs measuring. Using region interiors as the basis of

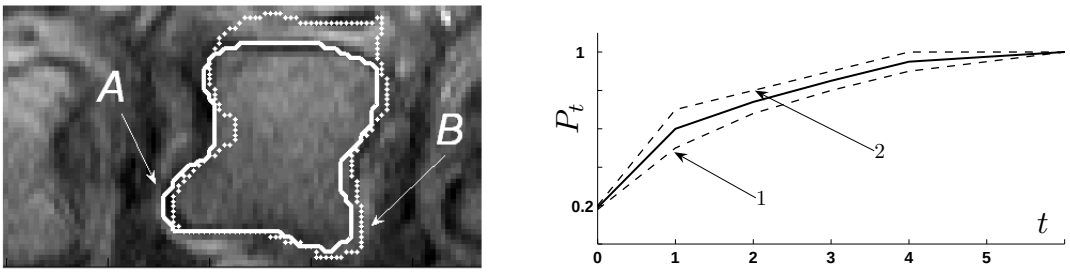


Figure 6.1: A segmented DCE-MRI image with GT shown in solid (A) and segmentation result – in dotted (B). Plots 1 and 2 show the proportion of pixels within threshold t from A to B, and B to A (P_t). The intermediate black line illustrates the percentage of pixels from both boundaries at which HD is no greater than t.

closeness, the HD based metric clearly precludes this.

We introduce a tolerance threshold t to describe how separated boundaries may acceptably become. For instance, due to the high noise level and ambiguity of regions' boundaries in DCE-MRI data, a difference between GT and a segmentation of 2-3 pixels is not usually a problem.

In Figure 6.1 a region segmented by a region growing algorithm [87], and corresponding GT are shown. The graph illustrates the effect of relaxing t: we measure the percentage of pixels within the given threshold distance. Clearly, this plot will be monotonically increasing and converges to 1. The underlying metric is Euclidean distance to nearest point on other perimeter.

For any two boundaries, we can compute the number of pixels of one that are within a threshold t of the other. This number, normalised by the total number of pixels in both boundaries, provides a metric $H_t(A,B)$ (the H_t metric), which gives the percentage of pixel-wise 'closeness' between two boundaries. Let N_A be the number of pixels in boundary A; then let $A_t(B)$ be the pixels of A within a distance t of a pixel of B. If N_{A_t} is the cardinality of $A_t(B)$, and N_B and N_{B_t} are defined similarly, we will write:

$$H_t(A,B) = \frac{N_{A_t} + N_{B_t}}{N_A + N_B} \tag{6.1}$$

For a given boundary, this metric will increase monotonically with t, and converge to 1. As it measures the distance between boundaries of the regions instead of the regions themselves, it permits evaluation of open boundaries.

The parameter t is an interval of tolerance, within which pixels from one boundary are considered as being in the vicinity of the other: this reflects the acceptable error of segmentation. It can be adjusted for the desired segmentation quality; for example, the width of ambiguous boundary sections, or the opinion of experts. The tolerance can be

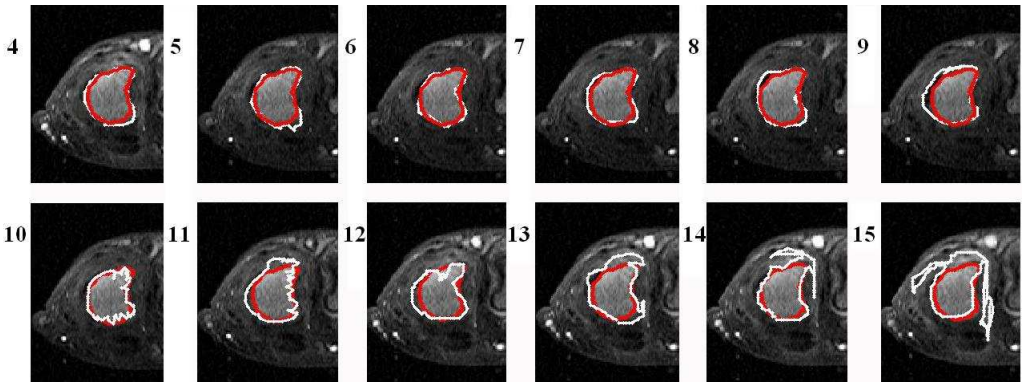


Figure 6.2: A bone interiors segmented by the region growing algorithm in the images from a sample DCE-MRI temporal slice acquired by the high-field scanner; images were acquired between T_4 and T_{15} time instants. Intensity change in the tissues surrounding the bones is noticeable. Machine segmentations are shown in white, GT – in red.

defined using domain knowledge, and therefore reflects an application-dependent acceptable segmentation error.

This metric produces easy to interpret and comprehensible results. It permits comparison of different algorithms on various datasets or tuning a method's parameters. Using it, we can detect the number of pixels which coincide precisely with the GT overlay, or assess the width of ambiguity.

Clearly, the combination of Equation 6.1 might have been defined in many ways, in particular as a simple mean $\frac{1}{2}(\frac{N_{A_t}}{N_A} + \frac{N_{B_t}}{N_B})$. We find the qualitative behaviour of these two definitions to be the same, while in the case t=0, Equation 6.1 has a correspondence with the MO definition (see Section 2.4.1) as the ratio of intersection to total perimeter.

6.1.1 Discussion: supervised evaluation

In registered DCE-MRI datasets the location of bone interiors is assumed to be constant through the temporal slice. A region growing was applied to these data; initial parameters, such as seed point and growing criteria, have been manually defined. In the data acquired with the high field scanner, the most noticeable intensity change is expected to appear in the sequence of images acquired between the 28 and 112 seconds (4^{th} - 15^{th} time instants. The quality of the segmentation results is expected to decrease during this interval because contrast enhancement during these times makes edges less clear. Figure 6.2 illustrates the results. It is noticeable that the segmentation quality degrades in the later images.

Figure 6.3 illustrates the segmentation quality measured by M_{MO} and H_3, where GT has been defined by experienced human input, which suggests that $t = 3$ represents inter-

observer difference.

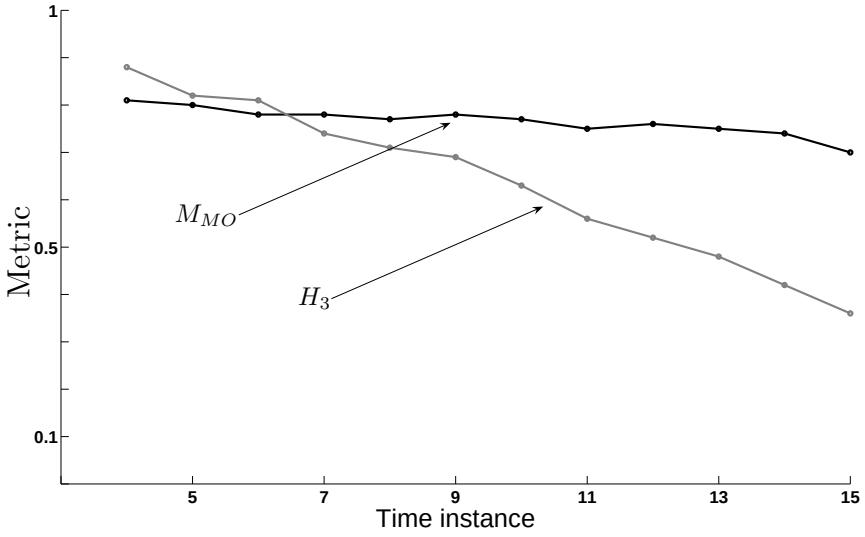


Figure 6.3: Evaluation of the segmentation results shown in Figure 6.2 by two supervised metrics M_{MO} and H_3.

We observe that both metrics detect a decline in segmentation quality over the 12 images, but M_{MO} provides less then adequate results in evaluation of segmentation of several images.

- M_{MO} suggests that the segmentation quality of the 11^{th} and 14^{th} regions is the same; both have $M_{MO} = 0.76$. The 14th region clearly has a long tail, but its shape and size do not impact on M_{MO} significantly.

- The 'tolerance' aspect of H_t has been useful: in the 11^{th} image the pixels of the segmentation are mostly within 6-7 of GT, but some pixels from the 14^{th} region are more than 10 pixels away. Segmentation quality obtained for these two regions has been analysed with H_t at different t (see Figure 6.4), clearly illustrating the different quality of result.

We can conclude that performance of the supervised metrics is in agreement on regions segmented so that their boundaries are complete, and do not contain tails. When, as in this application, tails are in evidence, H_t might be preferred. H_t also provides information on the extent of these tails, which allows more comprehensive segmentation quality evaluation.

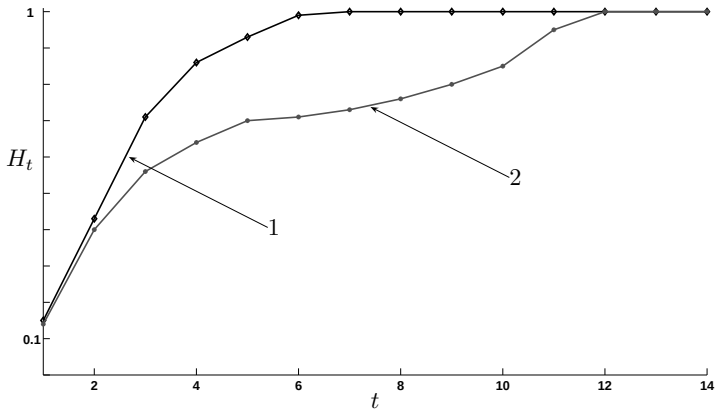


Figure 6.4: Graphs correspond to the analysis of region 11 and 14 from Figure 6.2 by the supervised metric H_t at different t. Graph (1) corresponds to the region 11 and graph (2) to the region 14.

6.1.1.1 Evaluation of segmentation algorithms with the supervised metric

In this section we will compare performance of three segmentation approaches – snakes [122], region growing [87], and the algorithm discussed in Chapter 5 using two supervised metrics M_{MO} and H_t. Examples of regions segmented by these algorithms are illustrated in Figure 6.5.

Region growing

Region growing techniques generally perform better on regions of homogeneous structure. In our application, intensities within regions are higher than outside, but are not homogeneous; regions detected with a low growing criterion are often located within GT and do not intersect with the actual boundary (see the 1^{st} region, 2^{nd} row in Figure 6.5).

When the criterion is high, the segmented region overlaps GT. Considering the first few images in the slice, we have manually selected initial parameters, and then these have been used to segment bones throughout the slice. The medical procedure causes intensity variations to be more pronounced in the last images in the slice, and therefore growing parameters for these are not optimal.

Evaluation with H_t reflects this (see Figure 6.6): very little of the segmented boundary coincides with GT (the mean of H_0 is 0.23 and maximum 0.3; the mean of H_1 is 0.3 and maximum 0.41). With a larger t, the quality of the segmentation results will be evaluated as higher. Evaluation at different intervals of tolerance shows that about 60% of boundary pixels segmented by region growing are located within a margin acceptable for our application (the mean of H_3 is 0.61 and maximum 0.7), but only a small percentage

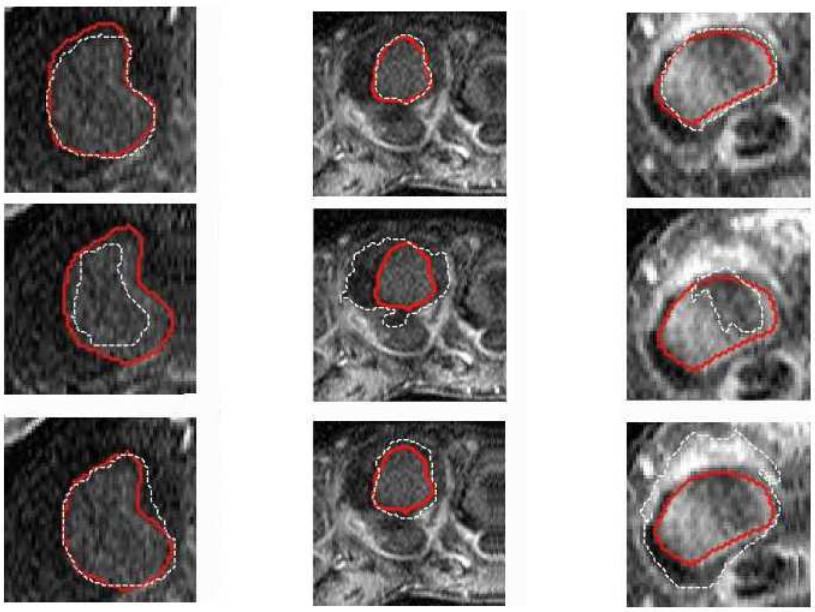


Figure 6.5: The regions (row 1,2,3) in images acquired by the high-field scanner from different DCE-MRI studies. In the top row: regions segmented by the adaptive segmentation algorithm; middle row: regions segmented by the region growing; bottom row: regions segmented by snakes. Machine segmentations are shown in white, GT in red.

are within 1 pixel from GT. Figure 6.6 illustrates the results obtained for 45 regions.

Boundaries of regions segmented by region growing very often contain tails, which can partly overlap with the actual contour of the boundary, but partly (normally very locally) extend toward the neighbouring bones. Results produced by the new metric show that the length of the tails might be 5-10 pixels, and these cannot be well assessed by the MO-based metric since the area of the tail does not adequately correspond to the 'error' it represents: it does not correspond to human opinion as demonstrated in Figure 2.4.

Snakes

The final regions delivered by snakes are close to actual boundaries, but often do not preserve original contours. Evaluation with H_0 and H_1 (Figure 6.7) shows that in many cases a high percentage of boundary pixels intersect with GT.

The initial parameters for snakes have been selected so that boundaries of segmented regions do not contain tails; this means that evaluation of quality can be performed adequately by either of the supervised metrics. Results are illustrated in Figure 6.7.

Evaluation results produced by both supervised metrics on regions segmented by snakes are comparable, as Figure 6.7 illustrates. We can observe that about 50% of de-

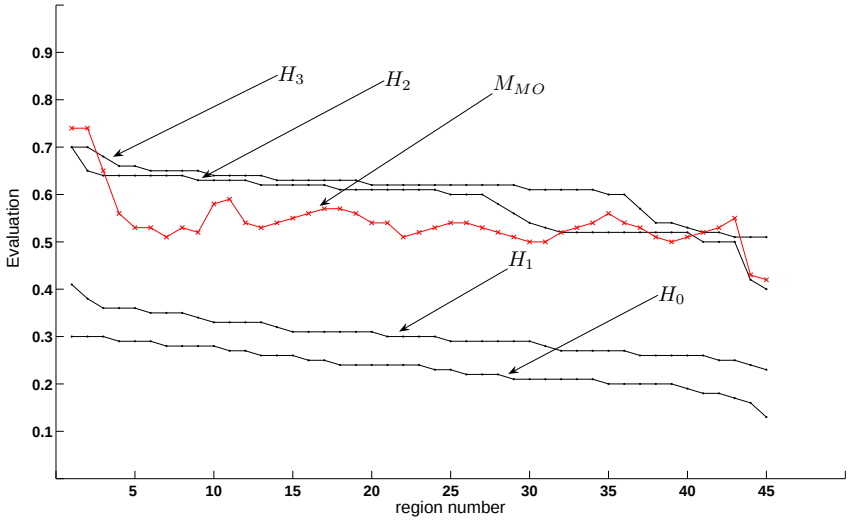


Figure 6.6: Evaluation results obtained with different metrics on 45 regions segmented by region growing. Results are sorted with respect to the behaviour of H_0.

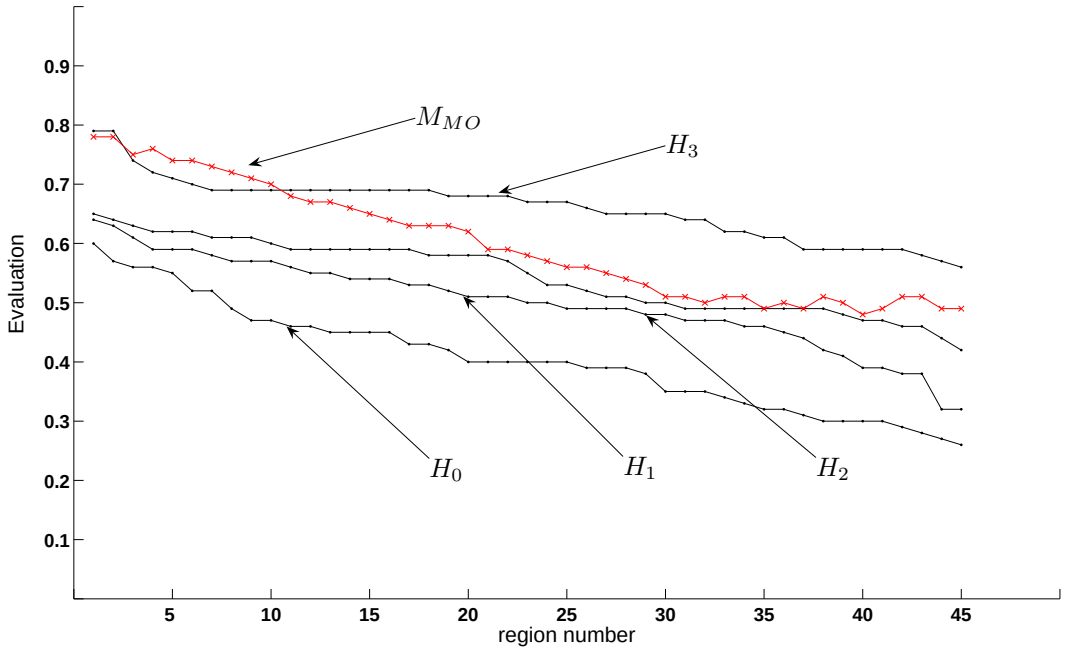


Figure 6.7: Evaluation results obtained with different metrics on 45 regions segmented by snakes. Regions are been sorted with respect to behaviour of H_0.

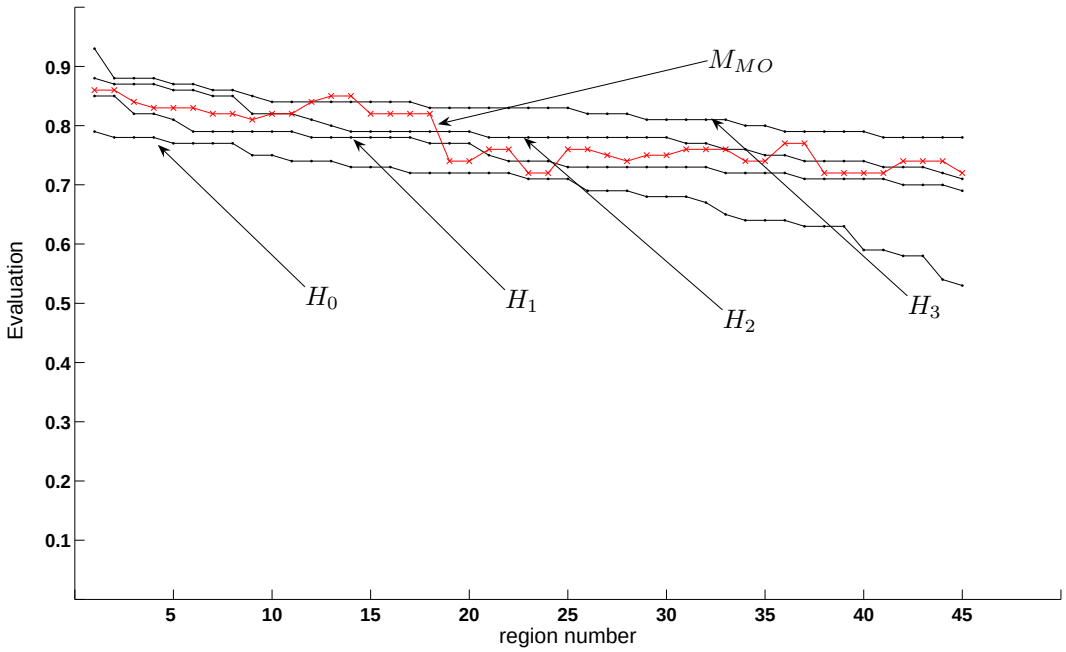


Figure 6.8: Evaluation results obtained with different metrics on 45 regions segmented by adaptive segmentation (Chapter 5).

tected pixels are placed correctly on a boundary (as in regions 7 and 8 in Figure 6.5), but more than 30% of all pixels detected are located further than 3 away from GT.

Adaptive segmentation

The adaptive segmentation based algorithm discussed in Chapter 5 has been specifically developed for this application, and is very successful at locating bone boundaries. In the majority of cases (Figure 6.8) results of high segmentation quality are produced: H_1 has mean 0.75, maximum 0.85; H_3 has mean 0.83, maximum 0.93. Divergence from GT may be due solely to subjective opinions of human observers. The algorithm performance is of interest when precise segmentation is required.

In images at the end of the slice, especially when the region boundaries are partially ambiguous, the algorithm might produce regions with tails (of extent 2-3 pixels). For this experiment we have chosen images from the beginning and the end of the slices taken from 10 DCE-MRI studies: the first 9 measurements in Figure 6.8 correspond to the regions from the images acquired at T_1. We can see that performance of both metrics on these images is in agreement. The remaining images have been randomly chosen, some

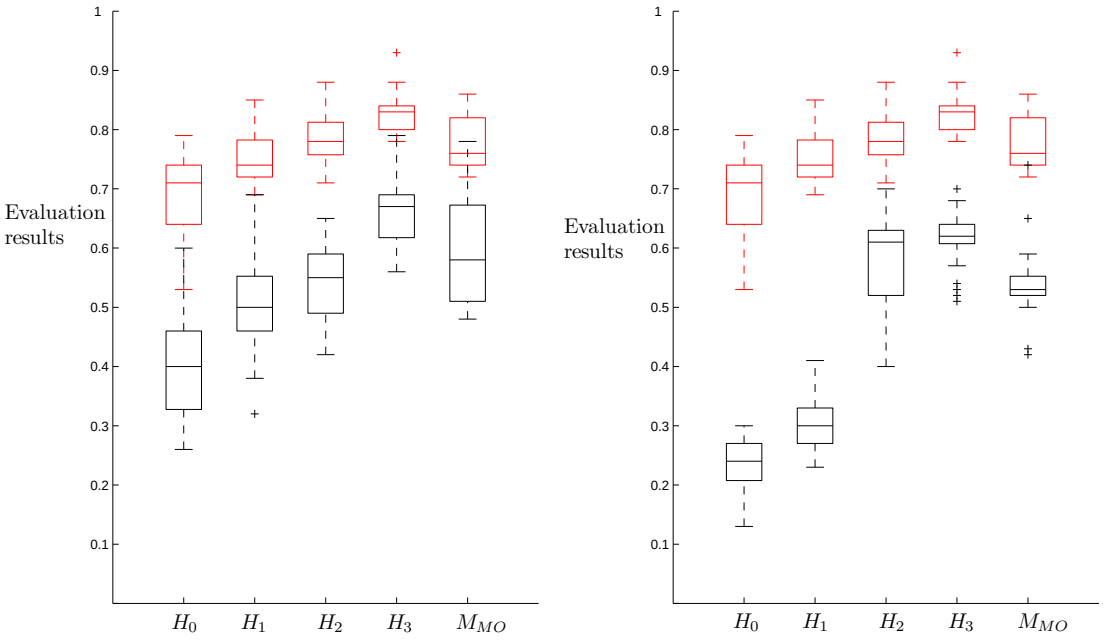


Figure 6.9: Left: Evaluation results obtained with different metrics (H_0-H_3, M_{MO}) for snakes (black) and adaptive segmentation (red). Right: Evaluation results for region growing (black) and adaptive segmentation (red).

containing tails or partly ambiguous boundaries; performance of the metrics on these regions is different, because the area of a tail does not influence significantly M_{MO}.

We learn that the H_t metric provides more comprehensive information on algorithm performance, assigning a threshold for segmentation error tolerance, allowing assessment of the width of ambiguous sections, and choosing an appropriate algorithm for an application. Much less information on boundary quality can be gained using M_{MO}; for some regions, M_{MO} provides an inadequate evaluation.

6.1.1.2 Evaluation of algorithm performance using supervised metrics

We compare three segmentation algorithms with respect to the quality of their segmentation results. Figure 6.9 shows box and whisker plots [158] of evaluation results for snakes, region growing, and adaptive segmentation algorithms.

While both supervised metrics suggest that the adaptive segmentation based algorithm outperforms the others, with M_{MO} we cannot assess which of snakes or region growing performs better on our data. Figure 6.9 illustrates that according to M_{MO} the median of

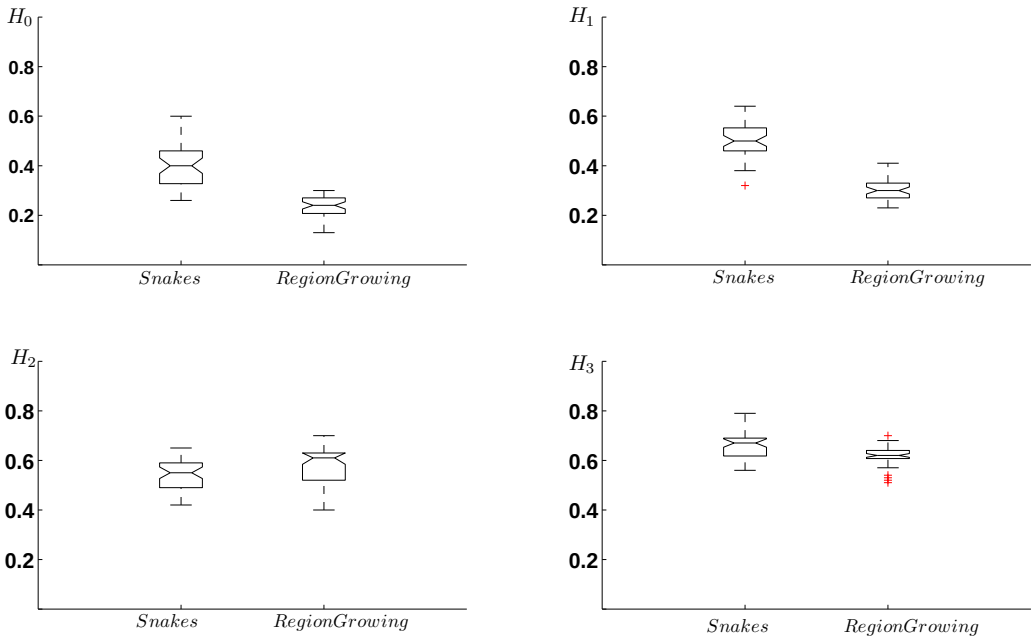


Figure 6.10: Evaluation of the results produced by snakes and region growing measured with H_t at different intervals of tolerance.

the evaluation coefficients obtained by snakes is higher than for the outputs obtained with region growing, but a large percentage of regions has been segmented with very similar quality. Thus, according to M_{MO} the performance of these two algorithms is indistinguishable (based on the similarity of medians).

Evaluation with H_t permits better analysis; the quality of the segmentation outputs produced by region growing and snakes at different intervals of tolerance are shown in Figure 6.10. By looking at the notches for the medians of the results produced by region growing and snakes in Figure 6.10, we can say that the two medians are significantly different at the 0.05 confidence level. Thus, the metric suggests that the performance of snakes is better.

H_0 and H_1 suggest that snakes outperform region growing (about 50% of the boundary pixels of the regions segmented by snakes and only about 25% of the pixels segmented by region growing are located in the vicinity of GT). At a larger interval of tolerance, such as 2-3 pixels, H_2 and H_3 suggest both algorithms perform similarly.

Based on H_t, we can conclude that the number of boundary pixels detected by snakes in the vicinity of GT is higher than the number detected by region growing, while the algorithm discussed in Chapter 5 outperforms both of them.

6.1.1.3 Evaluation of segmentation quality by human observers

The purpose of segmentation is often to locate regions which are of interest to a particular group of observers. The quality of algorithms is often judged by observers subjectively regardless of evaluation provided by metrics. Our previous experiments demonstrated that $H_0 - H_3$ are in agreement, and for certain regions M_{MO} behaves differently. Therefore, it is of interest to learn how judgments provided by the metrics correspond to the opinion of human observes.

We asked 30 experienced and 50 naïve observers to evaluate segmentation outputs obtained by the algorithms. Four pairs of images with segmentation outputs and GT were given to the observers (see Figure C.1 in Appendix C). Observers were asked to choose the 'better', in their opinion, segmentation output in each pair. No formal definition of 'better' was given. Table 6.1 shows observers' and metric preferences for each pair of regions.

	*Pair*1			*Pair*2			*Pair*3			*Pair*4		
	1	2	*ind*	1	2	*ind*	1	2	*ind*	1	2	*ind*
Exper.	3	23	4	2	27	1	13	2	15	14	4	12
Naïve	7	41	2	3	42	5	16	3	31	20	3	27
Total	10	64	6	5	69	6	29	5	46	34	7	39
H_t		$\surd$			$\surd$		$\surd$			$\surd$		
M_{MO}	$\surd$			$\surd$				$\surd$			$\surd$	

Table 6.1: Experienced (Exper), naïve (Naïve), and total number (Total) of observers who preferred region 1, region 2, or decided that regions are indistinguishable in quality (ind); H_t and M_{MO} preferences marked with $\surd$.

We received evaluation results for 80 subjects on 4 pairs of images. The results obtained for the first two pairs of outputs are consistent; evaluation of the last two pairs is less conclusive.

Image pair 1: The boundary of the first region contains long tails; the boundary of the second region is less fragmented and has no tails. The second region was preferred by H_t and 80% of observers.

Image pair 2: The first region is over-segmented and its boundary has long tails. The boundary of the second region has short tails. The second region was preferred by H_t and more than 86% of observers.

Image pair 3: Both regions are over-segmented, but the boundary of the second region contains shorter tails than the boundary of the first region. H_t prefers the first, M_{MO}

the second region. Most observers evaluated these regions as indistinguishable in quality (46 observers). However, the first region has been preferred by 29; the second region by only 5.

Image pair 4: Both segmentation outputs are over-segmented, the boundary of the second region contains longer tails than the boundary of the first region. H_t prefers the first, M_{MO} the second region. 33% of the observers evaluated these regions as indistinguishable in quality, the first region however was preferred by 34 observers, where the second one by 7.

Based on the judgments we can conclude that regions with tails or over-segmented regions do not provide accurate contextual information about the region of interest. Regions located within the ideal segmentation and regions with a larger number of boundary pixels, which coincide or in the vicinity of the GT overlay, are preferred.

This experiment demonstrates that observers base their judgments on a distance between machine segmentation output and GT, shape and smoothness of the boundary of segmentation, length of the tails, and amount of contextual information provided by the segmented countour.

Results obtained with this experiment intersect best with H_t – when supervised metrics disagree, evaluation with H_t was preferred by the majority of observers.

This experiment has of course its limitations. Using non-expert observers is imperfect because their interpretations are not based on medical knowledge. The observers were not explicitly told what segmentation is better, therefore a definition of a better boundary was decided by the observers. Obviously, opinion of the naïve observers is much less reliable than opinion of the experienced ones.

6.2 Unsupervised evaluation metrics

Metrics M_1–M_3 discussed in Chapter 2 have not been utilised extensively on MRI images, or images complicated by local blur or presence of partly ambiguous boundaries. The assumptions required by the metrics (i.e. sharp boundary) do not allow adequate evaluation of the results from such images. Here we propose a new unsupervised approach which can be used to assess the quality of segmentation results obtained from such datasets.

Boundary pixels may be located in areas of high and low contrast, in ambiguous sections of boundary, and might not represent GT at all. If we can confidently distinguish between pixels that represent a ‘reliable’ boundary of the object and pixels that do not, we

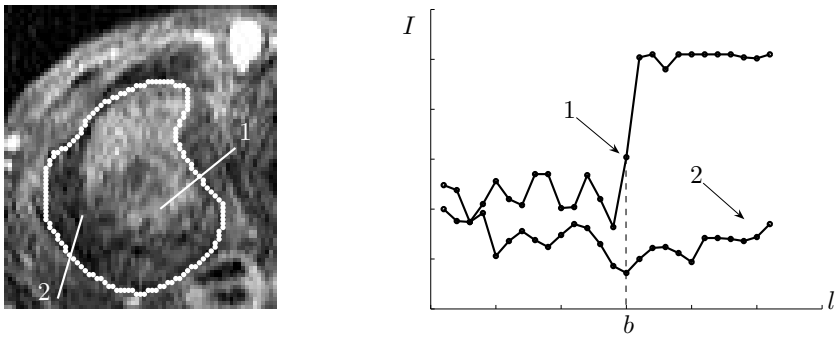


Figure 6.11: Left: Sample DCE-MRI region from an image of the MCPJs acquired with the high field scanner. A bone interior is in white. Two profiles (1 and 2) are plotted perpendicular to the boundary pixels. Right: Intensity values plotted along these profiles (1- corresponds to the profile 1 and 2 - to the profile 2); b – location of a boundary pixel.

can automatically compare quality of segmentation outputs, or performance of segmentation algorithms based on their ability to detect long reliable boundaries.

6.2.1 A new unsupervised evaluation metric

Consider a normal to a boundary pixel and an intensity profile drawn along it toward and from the boundary pixel. Let $\mathbf{I_1}$ denotes the intensity profile within the region and $\mathbf{I_2}$ the profile without. Let the length of each profile be $l \in [1;L]$, where L is the length of the longest profile (at this stage restricted only by image size). The choice of L will be considered later. Such profiles may or may not correspond to that expected at an 'edge' – see Figure 6.11.

To estimate the strength of a boundary pixel, we measure the intensity change along the profiles. We select weights as in Equation 6.2 and build a weighted sum on each of the intensity profiles by assigning heavier weights to intensities closer to the putative boundary pixel. This is illustrated in Figure 6.12.

$$\alpha_1 > \alpha_2 > ... > \alpha_l > 0, \ \sum_i \alpha_i = 1 \tag{6.2}$$

The weights α_i may be selected in many ways: for simplicity we have chosen them to be

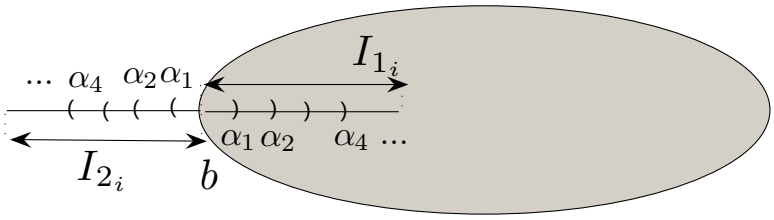


Figure 6.12: Diagram illustrating the intensity profiles ($\mathbf{I}_{1_i}$ inside the region and $\mathbf{I}_{2_i}$ outside the region) and weights α drawn for a boundary pixel b.

linearly decreasing to 0. The strength S can be defined as:

$$S(l) = \left| \sum_{i=1}^{l} \alpha_i \mathbf{I}_{1_i} \; - \; \sum_{i=1}^{l} \alpha_i \mathbf{I}_{2_i} \right| \tag{6.3}$$

Acknowledging that the evidence of a boundary pixel is likely to extend over more than a step change, there are many models we might use to fit the data: the sigmoid is popular. In an environment of probable noise, we use instead for convenience a piecewise linear function as shown in Figure 6.13 (left) – the sacrifice of accuracy in many applications will be dominated by noise effects. This can be described by Equation 6.4 in which there are two parameters: M gives the contrast across the edge and c captures how sharp (or fuzzy) it is. The weight function is given by Equation 6.5 (and is illustrated in Figure 6.13 (right)).

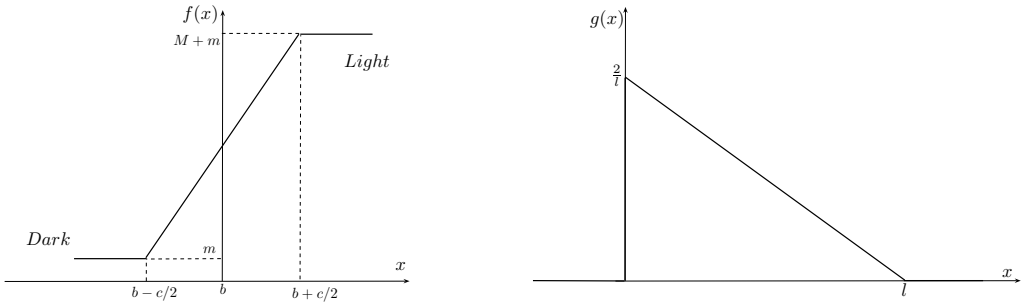


Figure 6.13: Functions f (left) and g (right) defined by Equation 6.4 and 6.5.

$$f(x) = \begin{cases} m & x \leq b - c/2 \\ Mx/c + M/2 + m & x \in [b - c/2; b + c/2] \\ m + M & x \geq b + c/2 \end{cases} \tag{6.4}$$

$$g(x) = \begin{cases} \frac{2}{l} - \frac{2}{l^2}x & x \in [0; l] \\ 0 & elsewhere \end{cases} \tag{6.5}$$

Let I_1 and I_2 denote weighted intensity profiles for a boundary pixel b. These are defined by the equations below, where $\mathbf{f}$ relates to the image intensity profile:

$$I_1 = \int_b^{l+b} \mathbf{f}(x) g(x-b) dx \tag{6.6}$$

$$I_1(l) = \begin{cases} \frac{-5Mc^2}{24l^2} + \frac{3Mc}{4l} - \frac{c^2 m}{4l^2} + \frac{mc}{l} & l \leq c/2 \\ \\ \frac{Mc^2}{24l^2} - \frac{Mc}{4l} + M + m & l \geq c/2 \end{cases} \tag{6.7}$$

$$I_2 = \int_{b-l}^{b} \mathbf{f}(x) g(b-x) dx \tag{6.8}$$

$$I_2(l) = \begin{cases} \frac{-Mc^2}{24l^2} + \frac{Mc}{4l} - \frac{c^2 m}{4l^2} + \frac{mc}{l} & l \leq c/2 \\ \\ \frac{-Mc^2}{24l^2} + \frac{Mc}{4l} + m & l \geq c/2 \end{cases} \tag{6.9}$$

Therefore, the strength of a boundary pixel S is given by Equation 6.10. An $S(l)$ curve corresponding to a 'perfect' profile is shown in Figure 6.14.

$$\begin{aligned} S(l) &= \left| I_1 - I_2 \right| = \\ &= \begin{cases} \frac{Mc}{2l} - \frac{Mc^2}{6l^2} & l \leq c/2 \\ \\ M - \frac{Mc}{2l}\left(1 - \frac{c}{6l}\right) & l \geq c/2 \end{cases} \end{aligned} \tag{6.10}$$

Given p and q, $1 \leq p < q \leq L$, we can estimate M and c from this equation. These estimates for different p, q will vary depending on noise effects, boundary misplacement and proximity of p and q to the asymptote. In order to estimate (M, c) reliably, for a natural number δ we will take several p, q pairs from an interval $[l_\delta, l_\delta + \delta] \subset [1; L]$ that makes this set of estimates most robust. We make this judgement by seeking the set of estimates that is most tightly clustered as a result of enforcing the model of Equation 6.10. For an interval $[l, l + \delta]$ we can generate n_δ estimates for $[M, c]$:

$$n_\delta = \frac{\delta(\delta + 1)}{2} \tag{6.11}$$

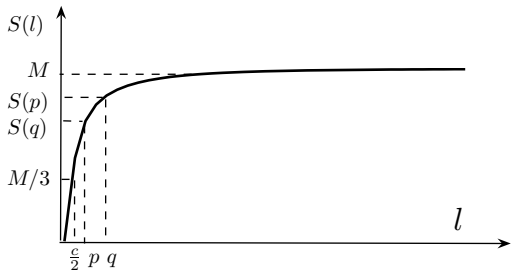


Figure 6.14: The strength of a boundary pixel $S(l)$ for the ideal profile.

These form a set $\left\{(M_k, c_k)\right\}_{k=1..n_\delta}$ with a centroid at (μ_M, μ_c). We will normally expect $\delta \geq 3$ in order to provide a non-trivial number of estimates.

For each interval $[l, l+\delta]$ with estimates $\{(M_k, c_k)\}_{k=1...n_\delta}$, we envelop the estimates with an ellipse centered at the centroid, and axes orientated along the principal axes of the data, with aspect ratio determined by the ratio of the data variances. For a given δ, we use the ratio between the cardinality of the subset and the ellipse perimeter as a set compactness measure, and select the maximal such.

Performing this for each acceptable δ, we now aggregate all estimates into one set $\{(M_k, c_k)\}_{k=1...n_\delta, \delta=3...\delta_{max}}$, and determine the means μ_M, μ_c and variances σ_M^2, σ_c^2. There is no statistical justification in theory for this aggregation but at the same time, we have no evidence that using this larger set will reinforce any bias. The lower values of l will give better results, being less likely to be corrupted by other information, but we have not used this observation. An issue remains, however, over the quality of any estimate of c when that parameter is low since most values of $S(l)$, particularly for higher l, will be near the asymptote and estimates of c may be ill conditioned.

In fact, this presents no serious problem; if c is 'low' then we observe a step edge and can get a (probably) reliable estimate of M from low values of l. Empirically, we discover that for $c > 4$, the approach we outline is reasonable, while smaller values represent a near step-change in intensity. In the application we come to consider, which is characteristic of challenging domains, c is commonly in the range $[5, 8]$.

Now for a given boundary pixel $\mathbf{x}_i$ we will have

$$f_i = f(\mathbf{x}_i) = (\mu_M^i, \sigma_M^i, \mu_c^i, \sigma_c^i) \tag{6.12}$$

These statistics will form the basis of future decisions about the validity of the pixel; in most circumstances, μ_M is good if high, and μ_c is good if low; σ_M and σ_c can provide a relative confidence in the estimates. These parameters are of course only relative in value,

and have no absolute interpretation (unless the image domain provides it – for instance, when there is a clear evidence that pixels with intensities above a certain threshold are correctly segmented).

We will define a confidence measure $\xi(\mathbf{x}_i) = \xi(f_i)$, $0 \leq \xi \leq 1$. Then, the metric of segmentation quality can be defined as:

$$Q = \frac{1}{N} \sum_{i=1\ldots N} \xi(\mathbf{x}_i) \tag{6.13}$$

The choice of $\xi(f_i)$ can be determined by the application – in Section 6.2.2 we illustrate some choices and compare their results with some of the established metrics.

6.2.2 Experiments and discussion

6.2.2.1 Evaluation of segmentation quality

Synthetic imagery

Human observers usually cannot gauge the quality of segmentation in any absolute sense and will judge one to be 'better' or 'worse' than another [276]. 'Goodness' is a relative property. Similarly, segmentation metrics in practical use will normally address themselves to a set of measurements made on an image or images, using a number of algorithms or algorithm parameter settings.

In order to compare quality of a number of segmentation outputs we set

$$M_{\max} = \max_{i=1\ldots N_k, k=1\ldots B} \mu_M^i \tag{6.14}$$

$$M_{\min} = \min_{i=1\ldots N_k, k=1\ldots B} \mu_M^i, \tag{6.15}$$

where B is the number of segmentations and N_k is the number of boundary pixels in the k^{th} segmentation. $c_{\max}$, $c_{\min}$ are defined similarly. A simple approach for estimation of the strength coefficients is to set

$$\xi_1(f_i) = 1 - \frac{M_{\max} - \mu_M^i}{M_{\max} - M_{\min}} \tag{6.16}$$

The quality of the k^{th} segmentation output is defined as

$$Q_1(k) = \frac{1}{N_k} \sum_{i=1}^{N_k} \xi_1(f_i) \tag{6.17}$$

Here, we assume that the pixel with $\mu_M = M_{\max}$ and the pixel with $\mu_M = M_{\min}$ are actually the best (and correctly detected) and the worst (and mis-segmented) ones in all segmentation outputs and therefore it is reasonable to grade the quality of the other pixels by comparing them to these cases. With the metric Q_1 only the magnitude of the intensity change defines the strength of the pixel. This metric is of value when all pixels after evaluation have very similar values of c_i.

Synthetic regions used in [45] for metric evaluation and comparison are a good example of such a boundary. Figure 6.15 illustrates two regions segmented by snakes. Here, boundary pixels are coloured according to ξ_1: pixels with maximal $\xi_1^i = 1$ are plotted in red, and $\xi_1^i = 0$ in yellow, with intermediate colours accordingly[1].

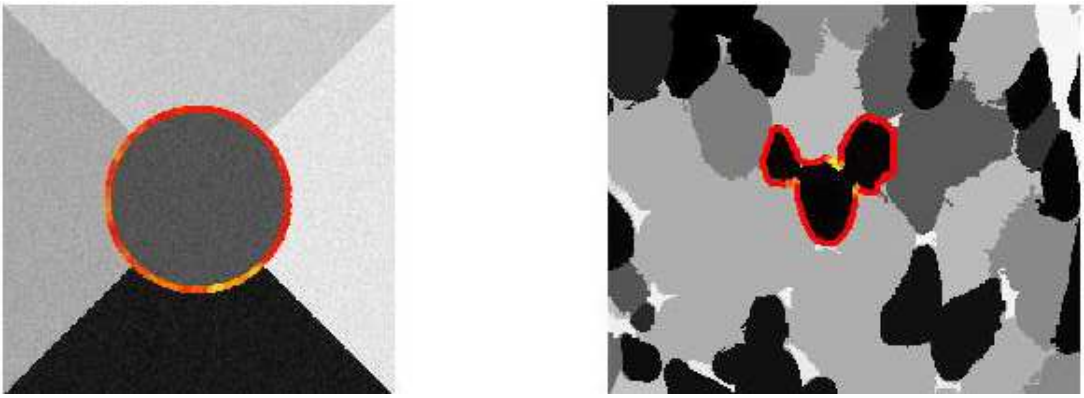


Figure 6.15: Synthetic regions taken from the database provided by [45] segmented by snakes; colours of the pixels from the machine segmented boundary correspond to ξ_1.

Consider a region segmented by region growing with two different threshold values, and a seed point located in the centre of the circle. Figure 6.16 (middle) illustrates the sensitivity of parameter selection: in an approximate sense the segmentation is good, but it is poor in many issues of detail. Previous authors have used artificially poor segmentation to provide comparisons: Figure 6.16 (right) is an example. It is obvious to a human observer that the quality of the segmentations shown in Figure 6.16 decreases from left to right.

Our metric illustrates that the first segmentation in Figure 6.16 is preferable and the hand segmented boundary has the lowest quality. The metric's judgments correspond to the opinion of a human observer. Thus, we can conclude that the metric here permits automatic evaluation and comparison of segmentation outputs obtained with different algorithms on synthetic images.

[1] We use this colour scheme throughout this section.

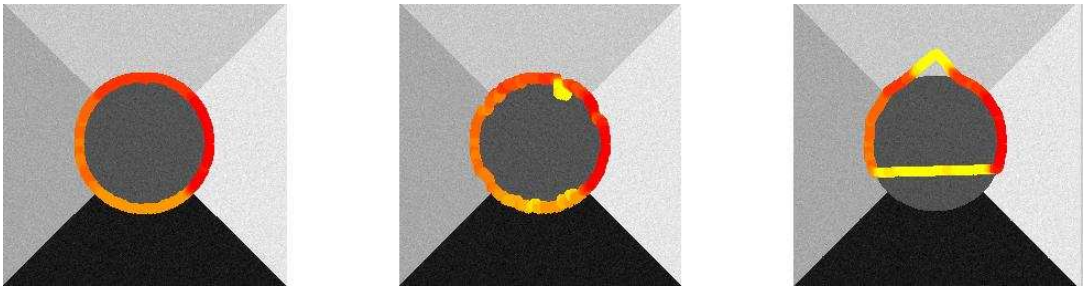


Figure 6.16: Left: Synthetic region from the database provided by [45] segmented by region growing $Q_1 = 0.96$. Middle: Synthetic region from the database provided by [45] segmented by region growing with different threshold value $Q_1 = 0.72$. Right: Hand segmented boundary, deliberately poor $Q_1 = 0.54$. Colours of the boundary pixels correspond to ξ_1.

DCE-MRI data of the MCPJs

Similarly to Equation 6.16 we can derive an approach using the estimate c. For some applications, mis-segmented pixels are those that separate areas of low contrast and/or located in the fuzzy/ambiguous parts of the boundary. Such pixels may have a relatively low estimate of M and high estimate of c. One way to incorporate the influence of c is:

$$\xi_2(f_i) = (1-\lambda)\left(1 - \frac{M_{\max} - \mu_M^i}{M_{\max} - M_{\min}}\right) + \lambda\left(\frac{c_{\max} - \mu_c^i}{c_{\max} - c_{\min}}\right) \tag{6.18}$$

and define Q_2 in the same manner as Q_1. λ determines the relative influence that M and c have on the final estimate of the pixel's strength.

For our application we are able to assume that:

- Pixels that separate areas of high contrast (high M), and are not in the ambiguous/fuzzy parts of the boundary (low c) probably represent the actual boundary of a region;
- Pixels that separate areas of low contrast (low M) and the intensity change along their normal is not sharp (large c) probably do not represent good boundary;
- No assumptions can be made about pixels with large M and low c or low c and low M. These pixels can represent a weak edge or an ambiguous section of a boundary.

If we can assume that estimates of M and c are equally significant, then $\lambda = 0.5$. Later it will be illustrated that for our application the optimal in correlation with human opinion λ is in the range $[0.4; 0.6]$.

Figure 6.17 illustrates segmentation results obtained with snakes and adaptive segmentation based algorithm, and Q_2. A larger percentage of the pixels on the right has been assigned lower strength coefficients, and according to Q_2 the quality of the left output is higher.

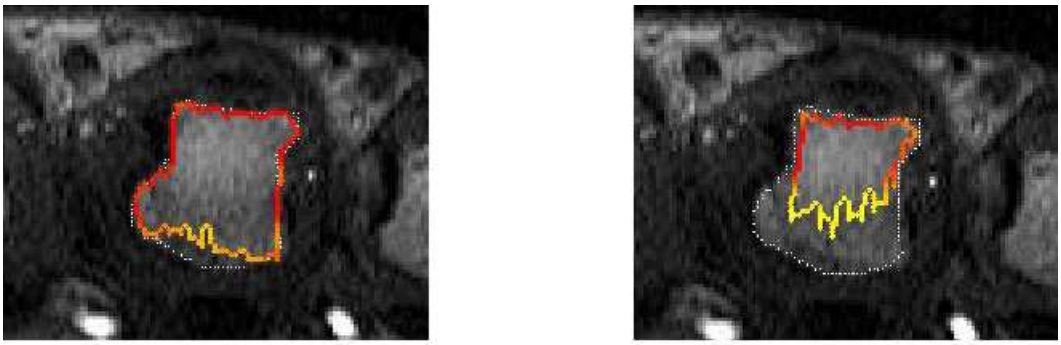


Figure 6.17: Segmentation results of the algorithm discussed in Chapter 5 (left) $Q_2 = 0.87$ and snakes (right) $Q_2 = 0.41$ obtained on a sample image from the database provided by [45]. GT is shown in white. Colours of machine segmented boundaries correspond to ξ_2.

Adjusting λ emphasises the influence of M or c. Figure 6.18 illustrates behaviour when $\lambda = 0.1$ (middle) and $\lambda = 0.5$ (right). The image on the left illustrates a boundary with blue lines indicating direction normals, their length corresponding to c.

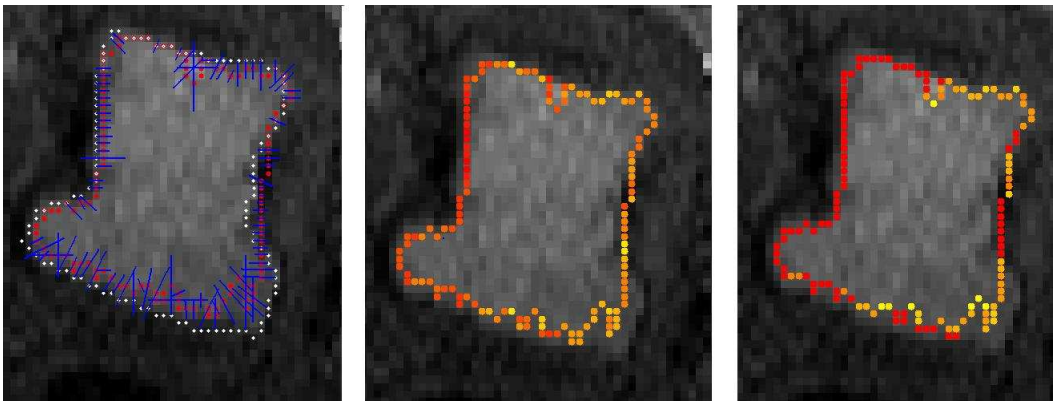


Figure 6.18: Left: Machine segmentation is shown in red, GT in white; blue lines are plotted in the direction of normals, their length corresponds to c. Middle: Colours of the boundary pixels correspond to ξ_2 ($\lambda = 0.1$), $Q_2 = 0.81$. Right: ξ_2 ($\lambda = 0.5$), $Q_2 = 0.87$.

With low λ the influence of M is more significant than c. This might be acceptable for synthetic images used in [45], where machine segmented boundaries are assumed to be constant width.

Segmentation outputs obtained on the MCP data are usually more than 1 pixel width; also, correctly detected boundaries might be strong (high M and low c) or weak (low M and low c). When λ is low, the strength of the weak and mis-detected parts of the boundary will be similar. Therefore, weak, but still correct parts of the boundary will be thought of as mis-segmented.

Figure 6.18 (middle) illustrates a segmentation evaluated with $\lambda = 0.1$. Some pixels from a weak part of the boundary have been assigned relatively low coefficients. When $\lambda = 0.5$ (Figure 6.18 right), and the estimate of c is considered, the strength of the pixels is defined much more accurately: pixels from a weak part of the boundary have been assigned higher strength.

This experiment illustrates that estimates of both M and c should be taken into account when comparing quality of segmentation outputs obtained on data such as these. Q_2 is suitable for quality assessment when information about contrast between background and foreground, and the fuzziness of the region boundary is relevant.

Incorporating confidence estimates in the evaluation metric

Use of the estimates M and c can be enhanced by incorporating confidence measures computed along with M and c. The standard deviations σ_M and σ_c are most unlikely to carry any physical meaning, but Equation 6.19 offers a simple approach that allows an assumption that some estimates are definitely wrong ($\beta = 0$);

$$\beta_M^i = \frac{\max_i(\sigma_M^i) - \sigma_M^i}{\max_i(\sigma_M^i) - \min_i \sigma_M^i} \quad \beta_c^i = \frac{\max_i \sigma_c^i - \sigma_c^i}{\max_i \sigma_c^i - \min_i \sigma_c^i} \tag{6.19}$$

Another simple alternative may be;

$$\beta_M^i = \frac{1}{\sigma_M^i} \quad \beta_c^i = \frac{1}{\sigma_c^i} \tag{6.20}$$

There are clearly many ways we might incorporate such confidence measures; one is to attempt to improve Q_1 and Q_2. We might use these confidences as the basis of a

weighted mean filter;

$$\nu_M^i = \frac{\sum_{j=i-h}^{i+h} \beta_M^j \cdot \mu_M^j}{\sum_{j=1-h}^{j=i+h} \beta_M^i} \tag{6.21}$$

$$\nu_c^i = \frac{\sum_{j=i-h}^{i+h} \beta_c^j \cdot \mu_c^j}{\sum_{j=1-h}^{j=i+h} \beta_c^i} \tag{6.22}$$

where h is some suitably chosen boundary window width. Similarly to Equations 6.16 and 6.18 the weighted coefficients of the pixels' strength are:

$$\xi_3^i = 1 - \frac{M_{\max} - \nu_M^i}{M_{\max} - M_{\min}} \tag{6.23}$$

$$\xi_4^i = (1-\lambda)\left(1 - \frac{M_{\max} - \nu_M^i}{M_{\max} - M_{\min}}\right) + \lambda \frac{c_{\max} - \nu_c^i}{c_{\max} - c_{\min}} \tag{6.24}$$

with Q_3 and Q_4 defined correspondingly.

We applied Q_3 to evaluate the quality of segmentation results obtained on a number of real life images [254]. One of the images is shown in Figure 6.19; segmentation results are obtained with region growing with two different threshold values and a seed point located in the centre of the region. Boundary pixels are coloured according to their strength given by Equation 6.23.

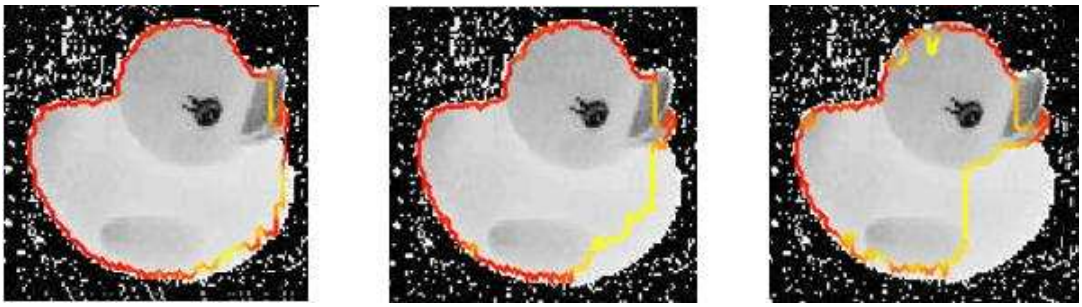


Figure 6.19: Segmentation results of region growing with different threshold value on a real world image [254]: From the left $Q_3 = 0.81, 0.70, 0.49$. Colours of the boundary pixels correspond to ξ_3.

It is obvious to a human observer that the quality of the segmentation results in Fig-

ure 6.19 (left) is the best and the quality of the region on the right is the worst. The metric Q_3 generates corresponding results.

We applied Q_4 to evaluate the results of snakes with different initial parameters on a sample DCE-MRI image of the MCP joints. Figure 6.20 illustrates the segmentation outputs, with boundaries coloured according to ξ_4: Q_4 prefers the result at the left of Figure 6.20.

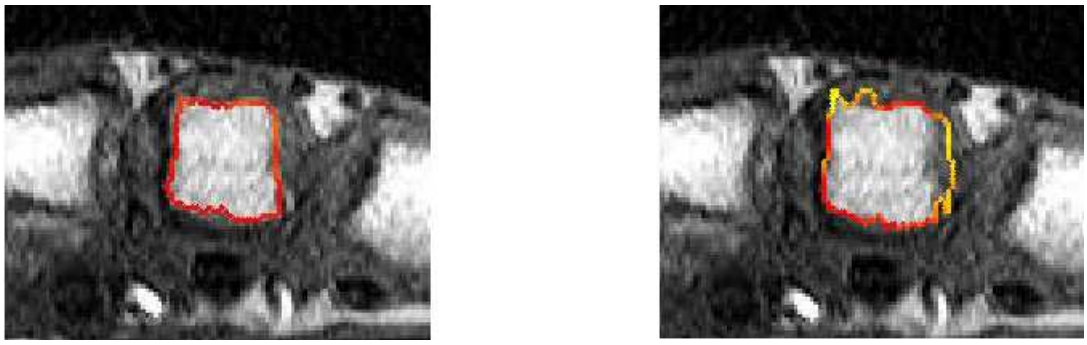


Figure 6.20: Segmentation results of snakes on DCE-MRI images of the MCPJs. GT is shown in white. Colours correspond to ξ_4; $Q_4 = 0.98$ (left), $Q_4 = 0.72$ (right).

Figure 6.21 illustrates a boundary segmented by region growing and evaluated by Q_1, Q_2 ($\lambda = 0.5$), Q_3, and Q_4 ($\lambda = 0.5$). Boundary pixels are coloured according to $\xi_1 - \xi_4$. The upper left part of the boundary clearly represents the actual boundary of the region, while the lower part is mis-segmented, and there are some pixels located in the weak part of the boundary.

ξ_1 has classified pixels from the weak part of the boundary as mis-detected; the colouring of the boundary pixels is inhomogeneous. When c is incorporated and strength is estimated using ξ_2, the majority of the correctly segmented pixels from the weak parts of the boundary have been assigned higher strength (Figure 6.21, region 2).

When the confidence based filter has been applied (ξ_3 and ξ_4), the boundary evaluations look much smoother. However, when the pixel strengths have been estimated with ξ_3 the colours in the boundary are inhomogeneous (Figure 6.21, region 3). The best evaluation is obtained with ξ_4: there are no outliers in the strong and mis-detected parts of the boundary, pixels located in the weak part have higher strength coefficients than mis-detected pixels, and the colours in the corresponding image (4^{th} region in Figure 6.21) correspond well to the quality of the segmentation. Our experiments show that for our application evaluation with ξ_4 generates the best information on segmentation quality (see Table 6.3).

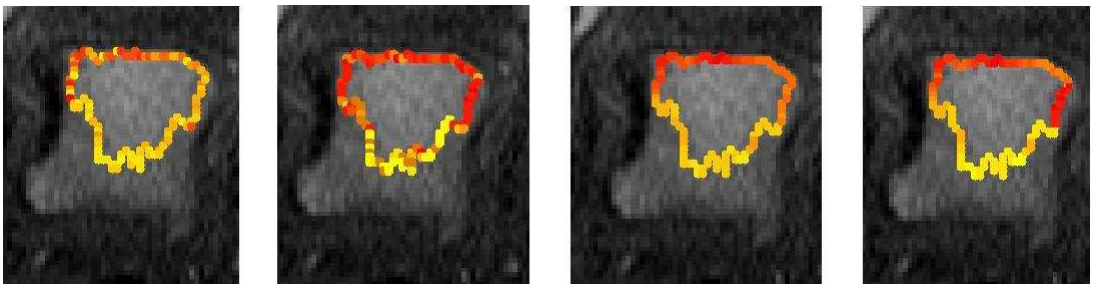


Figure 6.21: A region has been segmented by the region growing algorithm and the quality of the segmented boundary was evaluated by the metrics Q_1-Q_4. The colour in the images correspond to the coefficients ξ_1-ξ_4 (left to right).

6.2.2.2 Comparison of metric performance on synthetic and real world imagery

In a recent survey [45], several metrics [139, 259, 274] have been evaluated on a dataset of synthetic regions using M_{MO} as an objective measure. The results of the experiment conducted in the survey [45] are summarised in Table 6.2, where Spearman's rank correlation coefficient [112] was used to evaluate the strength of the relationship between the variables.

For this experiment the authors used a database of 400 synthetic images segmented by various algorithms; sample images and segmentation outputs are shown in Figure 6.22. The results show that the metrics do not agree with each other and have demonstrated very poor performance on the test imagery.

Metric	M_1	M_2	M_3	M_{MO}
M_1	1	0.02	0.06	0.20
M_2		1	0.18	0.16
M_3			1	0.56

Table 6.2: Correlation between various unsupervised approaches and M_{MO}: the data are taken from [45] and [44].

We have conducted experiments on 100 synthetic images (only 100 out of 400 were made available) from [43, 45], segmented by various algorithms (snakes, region growing, and adaptive segmentation based approach). Segmentation outputs have been evaluated by M_{MO}, H_t and Q_1, Q_3 and by the established unsupervised metrics $M_1 - M_3$. The evaluation results have been normalised to be in the range from 0 to 1 and then sorted according to the quality of the results obtained with H_1. Figure 6.23 illustrates the results.

Evaluation results produced by the supervised metrics correlate highly with results produced by the new metrics. The behaviour of established unsupervised and supervised

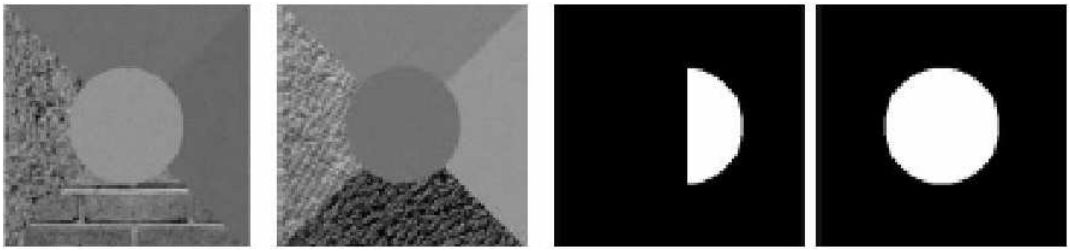


Figure 6.22: The first two images are sample images from the dataset used for experiments in [45], the last two images – segmentation results used in the experiments published in [45].

metrics does not agree. The new metrics show that there is obvious decrease in the quality of the segmentation results, whereas the established metrics failed to capture this trend. Table 6.3 shows the correlation coefficients between metrics.

Metric	M_1	M_2	M_3	Q_1	Q_2	Q_3	Q_4	M_{MO}	H_0	H_1
M_1	1	0.22	0.12	0.23	0.21	0.23	0.20	0.18	0.09	0.15
M_2		1	0.15	0.19	0.18	0.15	0.16	0.05	0.03	0.09
M_3			1	0.15	0.1	0.15	0.11	0.33	0.37	0.32
Q_1				1	0.80	0.96	0.88	0.83	0.88	0.86
Q_2					1	0.89	0.97	0.81	0.80	0.81
Q_3						1	0.92	0.88	0.89	0.87
Q_4							1	0.85	0.88	0.84

Table 6.3: Correlation coefficients between unsupervised (M_1-M_3, Q_1-Q_3) and supervised approaches (M_{MO}, H_0, and H_1) approaches (synthetic data).

The test imagery does not satisfy the established metrics' requirements, such as homogeneity of background and foreground intensity, and high contrast between region segmented and the background. The data in Tables 6.2 and 6.3 show that the established unsupervised metrics' performance on synthetic images is poor; they do not agree with each other and behaviour does not correspond to that of the supervised approaches.

Performance of the metrics Q_1 and Q_3 is similar to the performance of the supervised approach. The correlation between the results is high, which indicates that metrics generate reliable evaluation results.

6.2.2.3 Performance of metric on DCE-MRI data of the MCPJs

Experimental results published in [42] and [45] show that the established metrics' performance is also poor on real world satellite imagery and computed tomography images of the brain. To our knowledge none of the established unsupervised metrics has been used

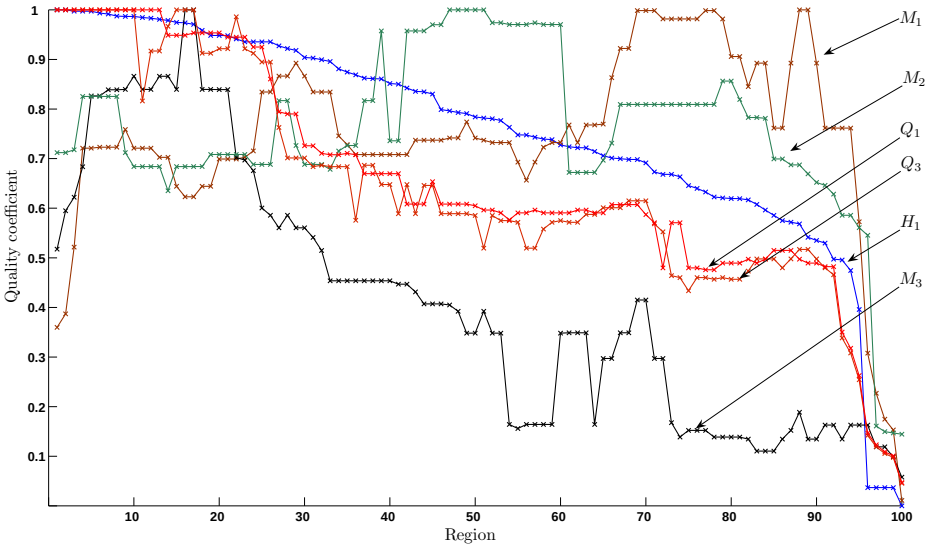


Figure 6.23: Evaluation results obtained with different metrics on 100 synthetic regions segmented by various algorithms. Segmentation outputs have been evaluated by M_{MO}, H_t and Q_1, Q_3 and by the established unsupervised metrics $M_1 - M_3$. The evaluation results have been normalised to be in the range from 0 to 1 and then sorted according to the quality of the results obtained with H_1.

for evaluation of segmentation quality obtained on DCE-MRI or MRI data. It is of interest to assess the metrics' performance on a large dataset with the rich variety of problems presented by this domain.

The dataset is composed of 140 regions, segmented by various algorithms (region growing, snakes, and adaptive segmentation based approach). Results have been evaluated by the supervised (H_t and M_{MO}) and unsupervised ($M_1 - M_3$, Q_2 and Q_4) metrics. Three independent observers were asked to outline GT for the most challenging regions. The per-pixel difference between these individual judgments averages 3 pixels, which suggests an appropriate value for t in evaluating H_t.

We have calculated the correlation coefficients between supervised and unsupervised approaches; these are shown in Table 6.4. Figure 6.24 illustrates evaluation results produced by 5 unsupervised and 2 supervised metrics on 140 regions.

Those metrics which use intensity or contrast uniformity of the segmented regions as an evaluation criterion [139, 259] perform poorly on DCE-MRI data. They show slightly better performance on images at the beginning of the temporal slices, where contrast between the correctly segmented region and the background is higher. Results on images from the end of temporal slices often do not agree with evaluation performed by super-

Metric	M_1	M_2	M_3	Q_1	Q_2	Q_3	Q_4	M_{MO}	H_3
M_1	1	0.14	0.12	0.19	0.23	0.21	0.14	0.33	0.21
M_2		1	0.11	0.17	0.19	0.16	0.11	0.29	0.26
M_3			1	0.15	0.20	0.17	0.14	0.37	0.33
Q_1				1	0.79	0.91	0.81	0.76	0.78
Q_2					1	0.79	0.96	0.82	0.86
Q_3						1	0.82	0.83	0.81
Q_4							1	0.84	0.87

Table 6.4: Correlation coefficients between unsupervised (M_1-M_3, Q_1-Q_3) and supervised approaches (M_{MO}, H_0, and H_1) applied to DCE-MRI data.

vised metrics.

M_3 (Zeboudj) provides evaluations which correspond better to the supervised metrics evaluation. This metric operates on the contrast changes in the neighbourhoods of pixels along the boundaries; it shows good performance when regions have been segmented with the adaptive segmentation based approach. These are usually very close to GT, which leads to sharper contrast changes between inner and outer areas.

The behaviour of Q_2 and Q_4 is similar on the MCP data. Results produced by the new unsupervised metrics show the best performance in correlating with H_3 and M_{MO}. This experiment shows that the new metrics generate reliable results in evaluation of DCE-MRI data of the MCPJs and can be used when automated comparison of segmentation outputs is required.

6.2.3 Limitations of the approach and application dependant parameters

There are obvious limitations of the proposed approach. As discussed in Section 2.4.2 all unsupervised approaches that provide results of useful quality require human intervention at some stage. Our method does not require a training stage, however there are two parameters (L, and for some metrics λ) that have to be manually chosen before the metric can be applied.

Optimal length of the profile L

Two factors should be considered when choosing L: the likely width of boundaries under evaluation and the proximity of local artefacts. By the width of the boundary we understand the image distance over which the edge exists. The length of the profile L should

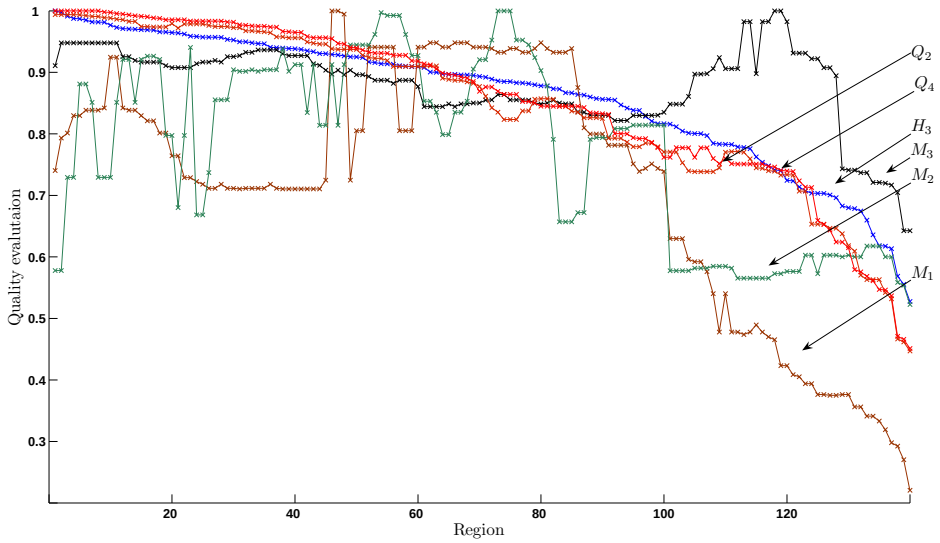


Figure 6.24: Evaluation results produced by the metrics on 140 randomly chosen regions from the DCE-MRI images of the MCPJs. Regions have been sorted according to the quality of results obtained with H_3.

be large enough to capture the intensity change across boundaries, but small enough to prevent corruption of profiles by the artefacts.

For many applications, it is difficult to define a unique position for a boundary. Much research has been carried out to evaluate true locations of boundaries [46, 263], but the problem remains unsolved.

Thus, in order to determine the width of a boundary, several experts (at least two) should be asked to outline outer and inner contours of several boundaries. Figure 6.25 (left) illustrates boundaries outlined by two experts. Suppose t_1 is the maximum distance between all outer boundaries, t_2 that between all inner boundaries, and D the maximal separation between closest inner and outer boundaries. Then in order for a profile to capture the true intensity change, L should be longer than $D+t_1+t_2$.

Recognising that particularly large values of l are probably unnecessary and undesirable, we are not interested in L being higher than it needs to be. The proximity of artefacts can be also estimated based on expert opinion. In our application, local artefacts are blood vessels and other bone regions, and experiments with our datasets have shown that $L \in [7;15]$ is reasonable, and provides sufficient estimates of M and c. Figure 6.26 demonstrates different choices of L. $S(l)$ captures the ambiguity of the boundary (high estimate of c implies more ambiguous boundary) and the intensity change. It is crucial to choose L large enough for the intensity change $S(l)$ to reach its maximum.

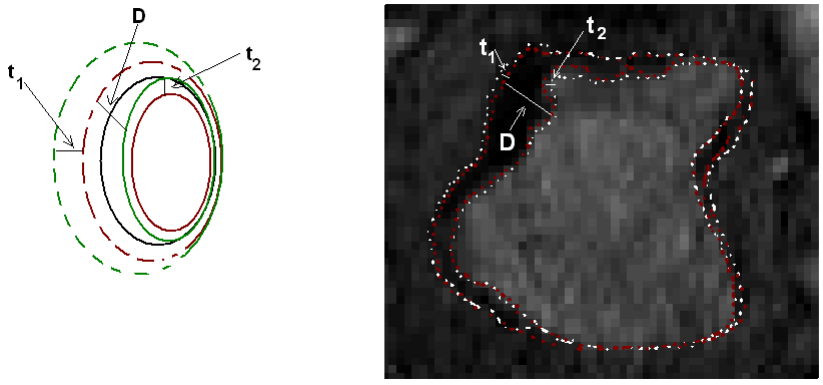


Figure 6.25: Synthetic and MRI regions with inner and outer boundaries outlined by two experts. Inner boundaries in solid, outer in dashed line; the same colour indicates results obtained from the same observer. t_1 is the maximum distance between all outer boundaries, t_2 that between all inner boundaries, and D the maximal separation between closest inner and outer boundaries.

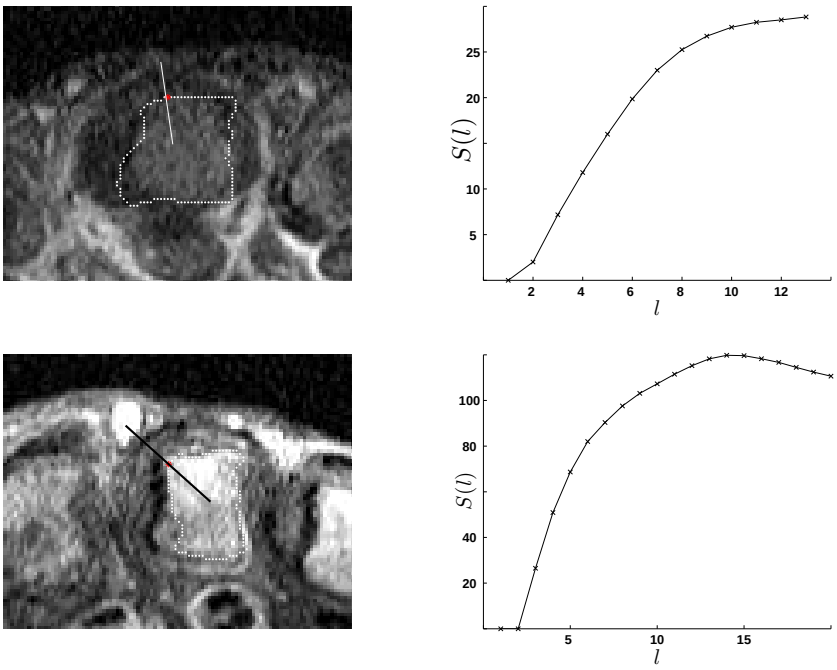


Figure 6.26: Left: Regions in sample images of the MCP joints acquire by the high field scanner with a segmented boundary shown in white and a profile drawn through the boundary pixels shown in red. Right: $S(l)$ graphs corresponding to the profiles.

Optimal choice of λ

Choice of λ and sensitivity of the metrics to it will depend on a particular application. For example, when the boundary under evaluation is supposed to be of a constant width, but has estimates of c indicating that it is not, then it can be assumed that segmentation delivered erroneous results. Here more weight should be placed on the term c. Another example, if after evaluation all estimates of c are similar and correspond to the expected width of a boundary, then more weight should be put on evaluation provided by parameters M.

To select an optimal λ for a new application a user can evaluate several regions with one of the supervised metrics, and then choose λ in such a way that the unsupervised evaluation corresponds to the results provided by the supervised metrics. In the absence of any ground truth information, equal weight on the M and c terms can be placed. For our experiments we found $\lambda = 0.5$ satisfactory.

6.3 Conclusion

Evaluation of segmentation algorithms is an intrinsic part of image processing. In this chapter we discussed several issues associated with supervised and unsupervised evaluation and proposed solutions to the evaluation problem.

We have shown that currently accepted supervised and unsupervised metrics are not always adequate in application to MR imagery, which is complicated by local blur, partial volume effects, intensity variations, subtle contrast, and patient movement artefacts.

The mutual overlap based metric does not deal well with objects whose boundaries contain tails, and is not applicable to open and ambiguous boundaries. It does not permit assigning an acceptable segmentation error and cannot assess inconsistencies in GT information.

Existing stand alone unsupervised approaches usually rely on image characteristics they measure and require constraints on image properties, which often cannot be satisfied in medical or real life imagery applications. Other unsupervised approaches aggregate stand alone metrics using classifiers and employ various learning strategies in an attempt to improve accuracy. Performance of such aggregate metrics improves with increasing size of training set, but the issue of choice of primitive metrics remains open.

We proposed a new supervised metric as an enhanced derivation of a Hausdorff distance based metric. It allows estimating the acceptable error of segmentation, which can be adjusted for desired quality; for example, the width of ambiguous boundary sections, or

the opinion of experts. This metric produces easy-to-interpret and comprehensible results. It permits comparison of different algorithms on various datasets or tuning a method's parameters. Using it, we can detect the number of pixels which coincide precisely with GT, or assess the width of ambiguity.

We also have considered several unsupervised approaches and demonstrated their limitations. We conclude that while it is probably impossible to develop an application independent unsupervised evaluation metric that would not require any human interference, we can propose a family of metrics valid for various types of images, including synthetic, real world, and DCE-MRI data. In such a way we can deal with the unfavourable properties of MR imagery and permit involvement of prior knowledge about the data.

Experiments have been performed using a large number of regions (synthetic and real) of various shapes, intensities, and contrast level, segmented by 3 algorithms of different behaviour and underlying criteria. We conducted experiments to compare evaluation results produced by the old and new supervised metrics against human observers opinion and learned that in most cases observers agree with the proposed metric. We have also objectively compared performance of the proposed unsupervised metrics against results obtained with supervised approaches and demonstrated that the correlation is high. We consider the new metrics to be an improvement on those prevailing.

Moreover, in this chapter performance of the adaptive segmentation part of the algorithm discussed in Chapter 5 was evaluated with MO and HD based approaches. Both supervised metrics suggest that the algorithm outperforms snakes and region growing, which generated disappointing results in most cases.

Chapter 7

Analysis of DCE-MRI data

In this chapter we present an alternative approach to voxel-by-voxel analysis, which overcomes problems associated with heuristic methods currently used for DCE-MRI data assessment and discussed in Section 2.5. We aim to allow fully automated analysis of signal intensity vs. time curves and objective estimation of the heuristics such as ME, IRE, and T_{onset}. Such analysis should permit quantitative assessment of the degree of inflammation and allow for time-efficient and objective evaluation of the patient's condition.

The results will be presented on the datasets acquired from active RA patients and healthy controls by high and low field scanners. The approach will be compared to the moving-window technique [195] discussed in Chapter 2

7.1 Classification of tissue behaviour

Behaviour of the signal intensity vs. time curves may be explained by the underlying phases of the data acquisition. Starting from a baseline, the perfused tissues absorb the contrast agent, and their intensity climbs up (wash-in phase); it usually increases up to a certain point and then exhibits a plateau (of variable width) followed by a wash-out phase (gradual signal intensity decrease).

Signal intensity vs. time curves are normalised over a baseline as discussed in Section 2.5.1. A sample generic normalised signal intensity vs. time curve ($\hat{I}$) and the durations of different phases of contrast enhancement are displayed in Figure 7.1, where δ_1

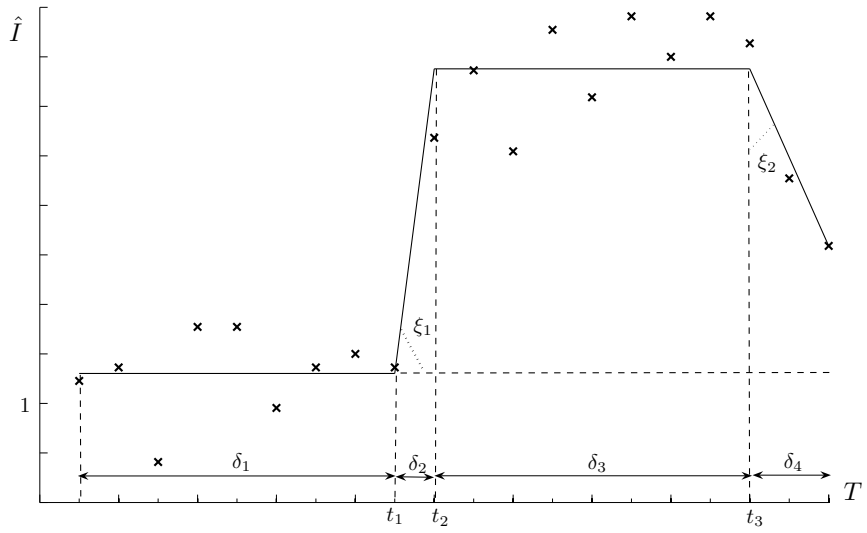


Figure 7.1: Signal intensity vs. time curve normalised over a baseline. T is the number of dynamic frames in a temporal slice; t_1, t_2, t_3 indicate the beginning of the wash-in, plateau, and wash-out phases.

is a baseline, δ_2 increase or wash-in, δ_3 plateau, and δ_4 wash-out periods; angles ξ_1 and ξ_2 represent the rates of the wash-in and wash-out. Some restrictions on the relationship between these parameters of the signal intensity vs. time curves may be derived:

$$\begin{aligned} &\delta_1 > 0,\ \delta_2 > 0,\ \delta_3 \geq 0,\ \delta_4 \geq 0, \\ &\xi_1 \in (0, \pi/2),\ \ \xi_2 \in (-\pi/2, \pi/2) \end{aligned} \tag{7.1}$$

We are implying a piecewise linear approximation[1] of the $\hat{I}$ and curves satisfying these restrictions will be approximated by one of the shapes shown in Figure 7.2.

S_1 – negligible enhancement: pixels at which no enhancement of note can be detected (these will include bone interiors).

S_2, $\overline{S_4}$, $\overline{S_5}$ – baseline/wash-in: signal intensity vs. time curves which clearly enhance but do not reach a plateau within T recorded instants.

S_3 – baseline/wash-in/plateau: pixels at which signal intensity vs. time curves reach the maximum and an intensity plateau develops.

S_4, S_5 – baseline/wash-in/{plateau}/wash-out: pixels at which the Gd-DTPA has dissipated and the intensity has detectably started to drop.

[1] Alternatively, curves can be approximated by more sophisticated models (e.g., a sigmoid), but we have not performed such experiments.

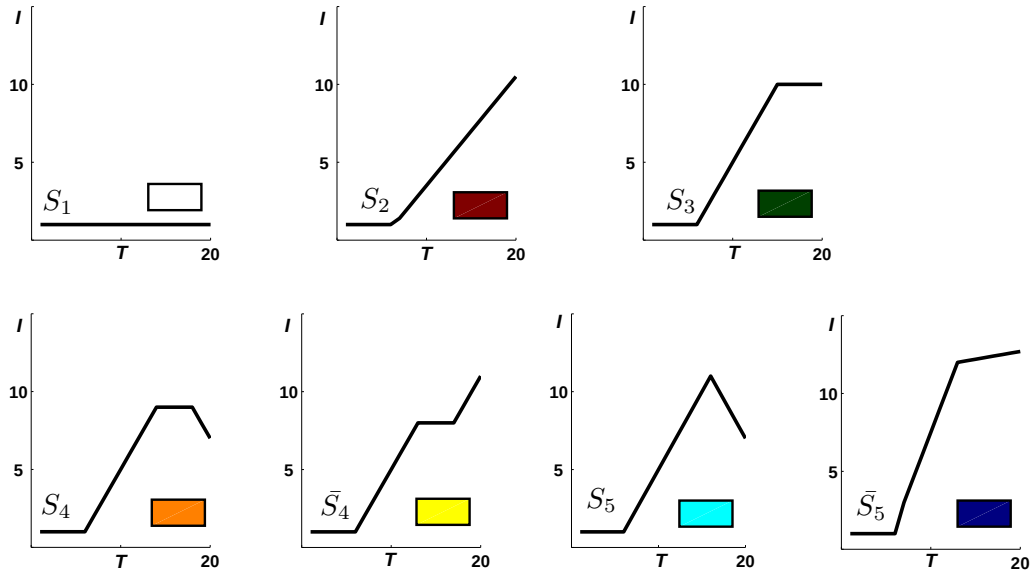


Figure 7.2: Possible shapes of the $\hat{I}$ curves: S_1, S_2, S_3, S_4, $\overline{S_4}$, S_5, $\overline{S_5}$.

S_1 corresponds to tissues which do not absorb the contrast agent, such as fat and water within the bone interiors, and display a relatively constant intensity within a temporal slice. S_3 describes the situation in which the tissue goes through the baseline, wash-in, and plateau phases. S_4 and S_5 show the presence of the relaxation phase of the tissue after the plateau. S_2, $\overline{S_4}$ and $\overline{S_5}$ illustrate the situation in which the take up time of the tissue is longer than the time of data acquisition, and the plateau has not been achieved.

The case $\overline{S_4}$ is included here only for completeness, and represents something we would not normally expect to observe. Here, while a change in the rate of increase is plausible, a significant plateau during the increase stage is not. Empirically we found that any observation of this model has a very short plateau (at most 3 time instants) equally well attributed to noise. We shall henceforward assume that all signal intensity vs. time curves can be modelled by one of the shapes proposed in Figure 7.2 excluding this special case.

We propose to use the knowledge of the underlying temporal pattern of the Gd-DTPA take-up to classify the $\hat{I}$ signals as an aid to noise reduction. This should permit more robust estimation of the heuristics, which will be extracted from the parameters of the fitted model rather than from the raw signals.

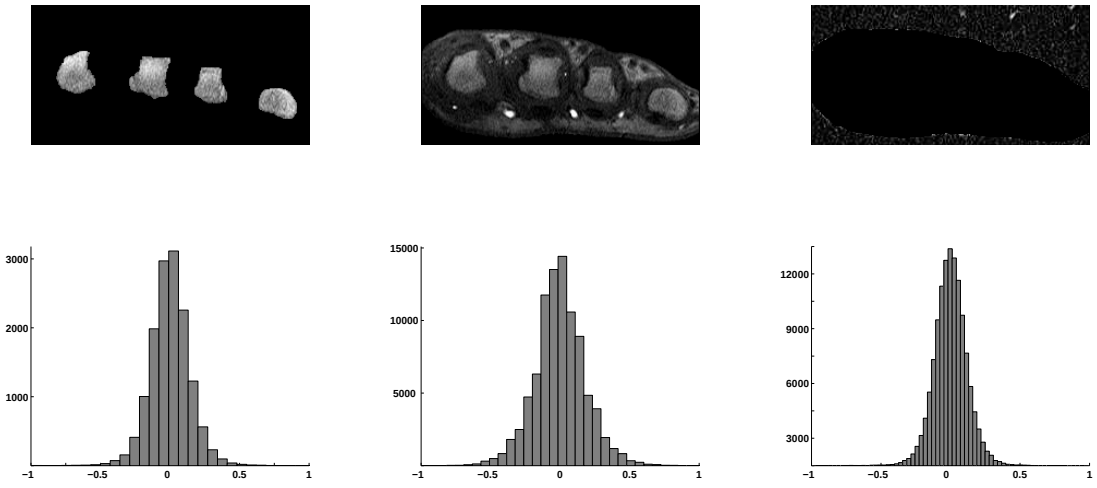


Figure 7.3: Top: The bone interiors, tissues within the joints' envelope, and the background segmented in a sample DCE-MRI temporal slice. Bottom: corresponding normalised noise distributions derived from the bone interiors, $T_1 - T_3$, and from the background pixels.

7.2 Noise model estimation

We seek an estimate of the distribution of the noise through which models will be fitted to data. It is assumed that there is no Gd-DTPA take-up in the tissues identified within the bone interiors and markers or in pixels outside the joints' envelope. Therefore, signals corresponding to these pixels may be approximated by a constant (the local signal mean), with variations being explicit noise measurements. It is also assumed that images acquired at $T_1 - T_3$ time instants are unaffected by any enhancement, and therefore these values may also be approximated by a constant, permitting a different measurement of noise. Note that it is not claimed at this point that these three noise estimates will necessarily come from or describe the same distribution.

Normalised noise distributions derived from three different sources are shown in Figure 7.3. There is clear similarity between these distributions which we can evidence using the Kolmogorov-Smirnov (KS) test [155]. This is a parameter independent test of goodness of fit, which has the advantage of making no assumption about the distribution of data. The KS test is based on a comparison between the empirical cumulative distribution functions (ECDF) of the datasets.

The KS test statistic is the greatest discrepancy between the cumulative frequencies of two distributions. If we have two experimental cumulative frequency distributions: $F_N(x)$ containing N events, and $F_M(x)$ containing M events, the test statistic is defined by

Equation 7.2 [155]:

$$D_{MN} = \max_x |F_N(x) - F_M(x)| \tag{7.2}$$

$$KS = D_{MN} \times \sqrt{[MN/(M+N)]} \tag{7.3}$$

Appropriate tables allow suitable conclusions of confidence to be drawn from the observation KS. In our experiments significance level of 95% was used. It may be argued that the noise level is expected to be the lower in the first few images, where any patient movement artefacts and the Gd-DTPA influence are insignificant. This would suggest deviations from the $T_1 - T_3$ baseline as the best approach. Section 7.4.1 illustrates that we are secure in accepting these three distributions to be indistinguishable. Accordingly, we model noise as an aggregate of that detected from these three different sources.

7.3 Determining best model fit

We proceed by attempting to fit each of the models illustrated in Figure 7.2, and considering as likely candidates any for which the implied noise matches in some sense that which we expect to see. For each model the piecewise linear best fit is determined in a least-squares sense (minimising also with respect to t_1, t_2, t_3); each such 'fit' then implies T noise measurements. We then compute KS for each model, rejecting those in which we would have low confidence. Note that we are interested in matching noise distribution and not minimising noise observation; the latter would always preclude the simpler models such as S_2, S_3 in favour of S_4.

In the event of more than one model being acceptable, the asymptotic statistic p derived from the test permits discrimination of which is 'best'. This statistic is known as Kuiper's coefficient of associations and measures maximum absolute difference between the cumulative functions of the distributions; the error distribution with lower absolute difference is preferred. The statistic p ranges from 0 to 1 and is defined as:

$$p = \frac{1}{2}\Big(\max_x |F_N(x) - F_M(x)| + \max_x |F_M(x) - F_N(x)| \Big) \tag{7.4}$$

Experimenting with this approach, most pixels are fitted best by models S_3, S_4 and S_5, as expected. Figure 7.4 illustrates an example in which each pixel in the slice has been coloured in accordance with its best fit (the colours correspond to those of Figure 7.2).

Clearly, some pixels potentially fit more than one model, and it is of interest to determine how often ambiguity arises. Figure 7.5 plots this with respect to the data from

Figure 7.4; the histogram at the bottom of the figure illustrates the frequency and depth of ambiguities at 95% significance level.

We see that the modal observation is 2 models, but many pixels also satisfy 3 models, while some of the signal intensity vs. time curves can satisfactorily be approximated by all models. This raises questions about the likely quality of the second (or third) best fit, and whether the chosen model best fits the actual data behaviour. To this end, we adapt the approach.

7.3.1 Simplified approximation scheme

Our observations suggest that many ambiguities seen within the S_i models usually do not reflect differing underlying physical behaviour, but are issues of detail probably decided by local noise. Therefore, we will classify the tissue under examination into one of four broad behaviours, which we will label and define as:

M_0 – negligible enhancement. Referring to the linear fits, this is S_1. Some tissue located within cortical and trabecular bone, inactive joints, skin and disease unaffected areas do not absorb Gd-DTPA and are not expected to show intensity enhancement in the later frames of temporal slices. Where recognised, we will not colour such pixels in future representations.

M_1 – baseline/wash-in. There is often a proportion of curves in which by the end of the scanning procedure the maximal intensity has not been reached, indicating constant leakage into locally available extra-cellular space. The Gd-DTPA absorption and signal intensity vs. time curves enhancement continue after the scanning has been completed. These are models S_2 and $\overline{S_5}$, which we will colour red.

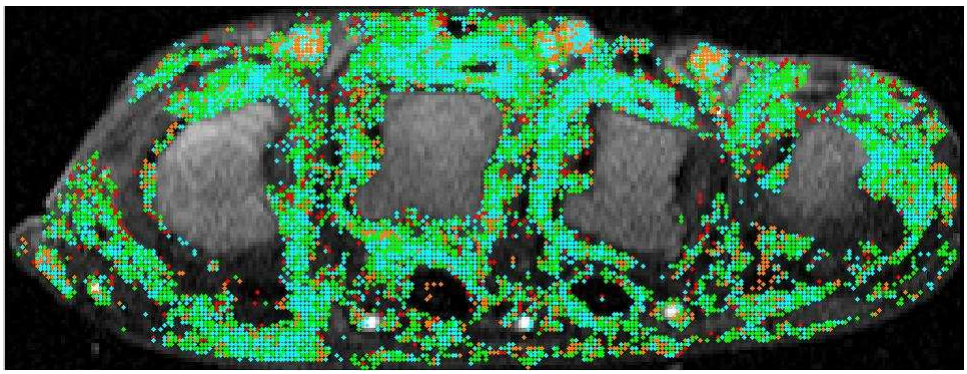


Figure 7.4: An illustration of the model ‘best fit’: pixels, where signal intensity vs. time curves assumed S_1 are plotted in white, S_2 in red, S_3 in green, S_4 in orange, $\overline{S_4}$ in yellow, S_5 in cyan, and $\overline{S_5}$ in blue.

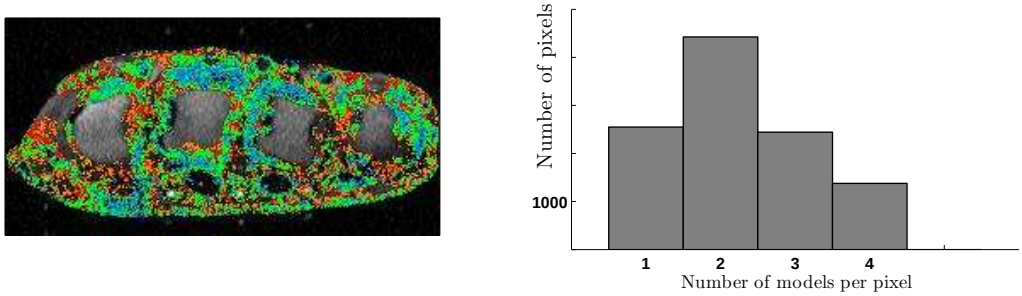


Figure 7.5: Left: Pixels plotted in colours corresponding to the number of approximation models (S_1, S_2, S_3, S_4, S_5,and $\overline{S_5}$) that fit: 1 model – cyan; 2 models – green; 3 models – red; 4 models – orange. Right: The histogram summarises the frequencies: number of models fit per pixel vs. number of pixels.

M_2 – baseline/climb/plateau. Full absorption of the Gd-DTPA by the tissues. This is S_3, which we shall colour green.

M_3 – baseline/climb/plateau/decline. The wash-out phase is observed at the end of the scanning procedure. These are models S_4 and S_5 which we shall colour blue.

A recolouring of Figure 7.4 with this scheme is given in Figure 7.6. It is still possible for more than one model to fit at a given pixel, but this is now distinctly unusual. Figure 7.8 illustrates this. Thus, simplification of the classification significantly reduces the number of ambiguities seen within the S_i models.

7.3.2 Spatial filtering

The behaviour of $\hat{I}$ need not be considered in isolation. It is reasonable to suppose that neighbouring tissue will behave in a similar way; but inspection of Figure 7.6 suggests significant local variability. Arguing that the models M_1, M_2, M_3 represent 'ordered' behaviour, we have passed a selective 3×3 median filter, which is a form of a spatial filter, over the labels represented by Figure 7.6 (only pixels labelled 1, 2 or 3 are computed in the median).

The size of the filter will depend on the application and resolution of the images. Empirically, we found that for our data results obtained with median filter sizes from 3×3 to 7×7 correlate well with human judgements. With the larger size, areas with the most active synovitis are oversmoothed.

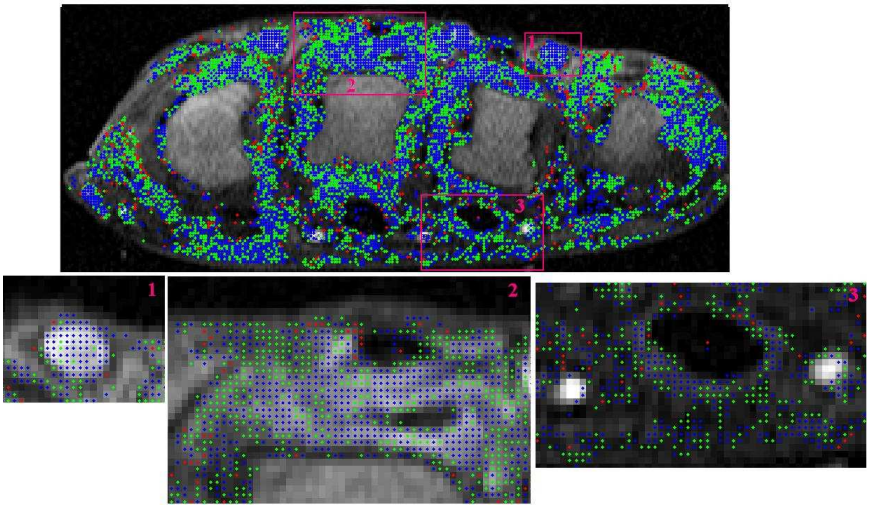


Figure 7.6: Resulting map of the contrast agent uptake obtained for a sample DCE-MRI temporal slice of the MCP joints acquired with the high field scanner. The colours in the image correspond to the models each normalised signal intensity vs. time curve assumed: M_1 -red, M_2-green, and M_3 -blue. Below are three magnified regions: (1) blood vessel, (2) inflamed tissue, and (3) skin region.

This results in the relabelling of some pixels – of the order of 20% – 30% of those within the perimeter of the joints' envelope excluding bone interiors. We now adopt for these pixels the best fit of the imposed model.

Since these are no longer in our sense 'best fits' we will have incurred different error residuals – we find that in the majority of cases the model is changed to the second best as indicated by the p statistic, suggesting that the noise distribution will not be perturbed radically.

We have considered the aggregate distribution of such and compared it to the adopted error model; the KS statistic allows us to deduce it is indistinguishable from that of the model to which we are working. Figures 7.6 and 7.7 show an example of this procedure, and Table 7.1 summarises the effect.

Label	Before	After
M_1	705 (6%)	940 (8%)
M_2	4822 (41%)	5762 (49%)
M_3	6233 (53%)	5058 (43%)

Table 7.1: Populations of the different labels of Figures 7.6 and 7.7.

This example illustrates that after the filtering the colours show distinguishable clusters. There is a clinical plausibility to these patterns: blue – presence of the wash-out

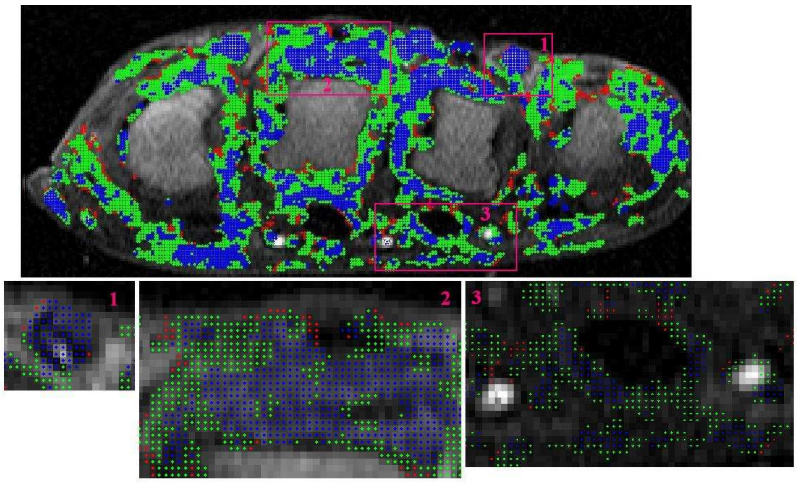


Figure 7.7: A median filtering of the preceding image.

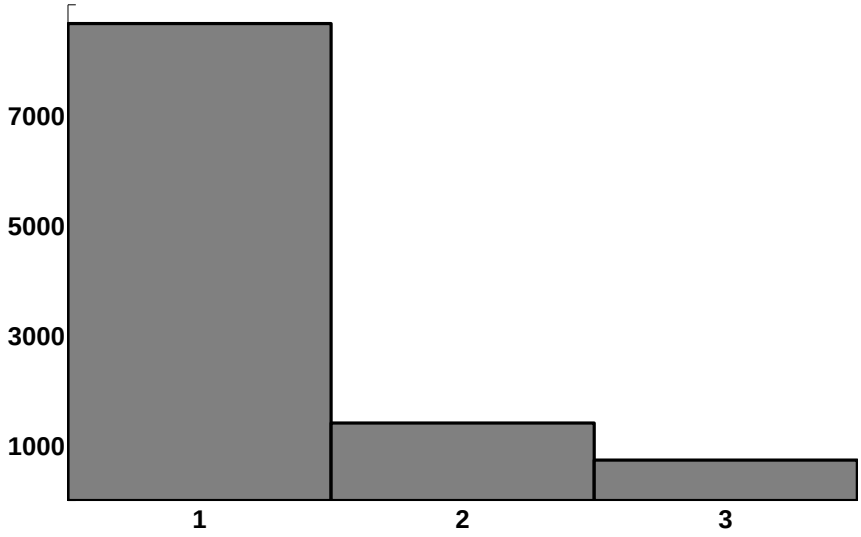


Figure 7.8: Referring to Figure 7.6, the number of approximation models ($M_0 - M_3$) a normalised signal intensity vs. time curve satisfies.

phase, green – plateau, red – continuing wash-in. In particular, it is possible to observe that the blood vessels and tissues with the most active inflammation are blue, an expected behaviour.

Some magnifications are also shown in Figures 7.6 and 7.7. They depict blood vessel, disease affected area and skin. Note that the few green pixels visible inside the blood vessel in the 1^{st} region in Figure 7.6 have been coloured blue; this identifies the vessel. An affected area has been split into several clusters of blue and green. This experiment shows that the enhancement is clearly visible, with patterns revealed by the filtering.

Figure 7.9 shows the results obtained for a slice from a different study. Here, significant areas become visible in which the intensity has not reached its maximum. This implies significant incompleteness in any conclusions drawn from the data; the converse would be true in the absence of any red pixels.

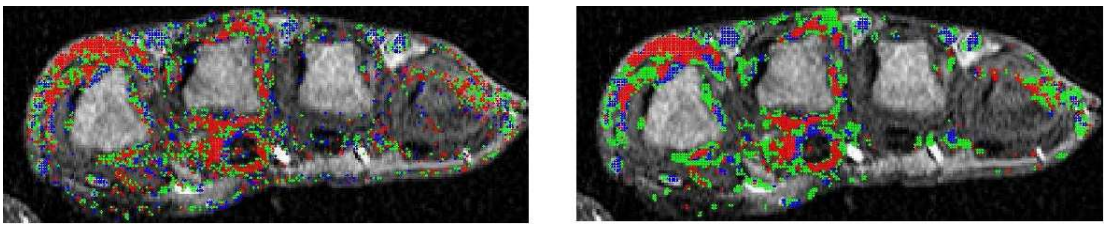


Figure 7.9: Left: An image coloured according to the estimated approximation models $M_1 - M_3$. Right: The same image after median filtering.

The median filer in our case is used as a voting mechanism. It is acceptable if there are only two labels in the neighborhood, but slightly nonsensical if there are more than two labels. For example, it would not produce an adequate results being applied to following labelling [1112333]. However, empirically it was found that majority of the pixels within the tissue of interest are labelled M_2. That is why, the results of the filtering are plausible. For more complex data, a more sensitive filters such as Markov Random Field or a rank filter, would be more suitable.

7.3.3 Computing ME, IRE, and T_{onset}

Given a presumed model, it is straightforward to extract heuristics ME, IRE and T_{onset} from the linear approximation rather than from the raw data; Figure 7.10 illustrates this.

Note that t_1 of Figure 7.10 is interpreted as actual time of onset of enhancement, and there will be differences in these measurements when compared with the moving window technique. These are discussed in Section 7.4.2.

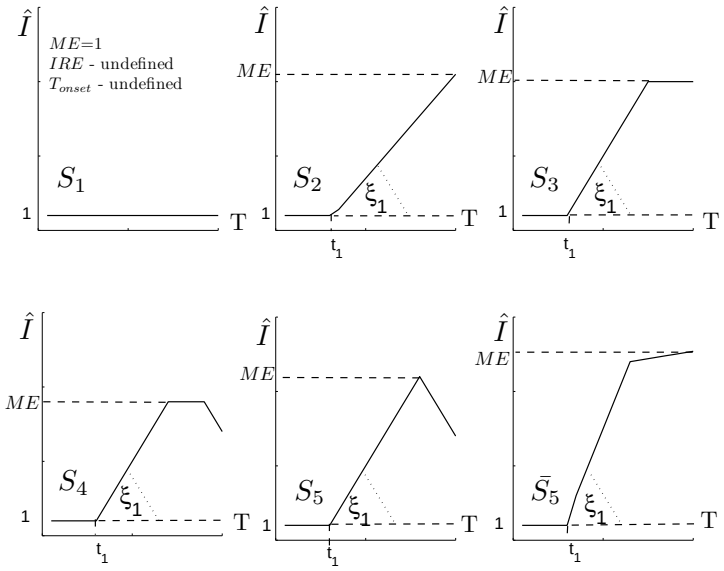


Figure 7.10: Estimation of the parameters ME, IRE, and T_{onset} for each approximation model. ME has not been reached for S_2 and $\overline{S_5}$.

7.4 Discussion

7.4.1 Consideration of the noise model

In Section 7.2 we referred to different sources of noise that we can measure explicitly: N_1 derived from the first 3 time instants, N_2 from the bone interiors over all time instants, and N_3 from the background pixels. We argued that distributions of these different sources of noise for our purposes are indistinguishable, and that we can aggregate them.

If we experiment with the models separately, we find N_1 (the 'deviation from the baseline' model) to be slightly stricter – more signal intensity curves were approximated by M_0 and therefore heuristics such as IRE and T_{onset} were not defined for these $\hat{I}$. This is illustrated in Figure 7.11 where the labellings derived from the noise models (N_1, N_2, and N_3), are superimposed on the intensity post contrast image. Table 7.2 summarises the data from these figures.

It is not immediately clear that noise characteristics will remain unchanged during the procedure. We have considered the noise distribution across all T_i; the KS test suggests that the distribution of these errors is not distinguishable from the model we have constructed.

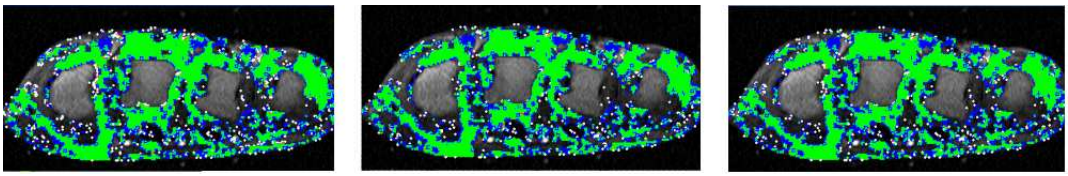


Figure 7.11: Pixels are labelled according to the best fit from Figure 7.1. Noise is estimated from N_1 model (left), N_2 model (middle), and N_3 model (right). Colour labellings were derived from the noise models (N_1, N_2, and N_3) and superimposed on the intensity post contrast image of the MCPJs acquired with the high field scanner. White colour corresponds to pixels, whose normalised SI curves were approximated by M_0, red - by M_1, green - by M_2, and blue – by M_3.

Model / Noise model	N_1	N_2	N_3	$N_1 \cup N_2 \cup N_3$
M_0	1194 (12%)	1152 (12%)	1176 (12%)	1186 (13%)
M_1	51 (2%)	42 (1%)	43 (2%)	46 (1%)
M_2	5588 (60%)	5584 (60%)	5500 (57%)	5589 (60%)
M_3	2459 (26%)	2514 (27%)	2490 (29%)	2515 (26%)

Table 7.2: The relative populations of the different labels in Figures 7.11. The values mean the number of pixels.

7.4.2 Visualisation of the heuristics *ME* and *IRE*

Parametric maps are generally used to characterise functional anatomy and disease-related changes [50,180,195]. When the heuristics *ME* and *IRE* are computed, a parametric map is built to reflect the general behaviour of the tissues. Figure 7.12 illustrates a parametric map of *ME*. The colour coding here considers the value of a parameter and plots lower values in red, moving to yellow then white as the values increase. In the parametric maps, values of the heuristics are normalised by their maximum and colour coding is performed on the data from the range [0,1]. The maximum values of *ME* and *IRE* are normally associated with voxels located within blood vessels, therefore the degree of inflammation of other tissues will be colour coded in respect to the blood vessel enhancement. This, of course, prevents objective comparison of the data acquired on different scanners.

Figure 7.13 illustrates a parametric map of *ME* constructed with different approaches. The study chosen exhibits significant patient movement, and the effect of the registration to compensate for this should be clear. The differences are further illustrated in the magnified images shown in Figures 7.14 and 7.15.

According to the opinion of experienced observes and clinical experts, the maps constructed with the model-based technique provide sharper shapes of the bone contours, blood vessels, and reduced skin enhancement. These visual results are explained in Sec-

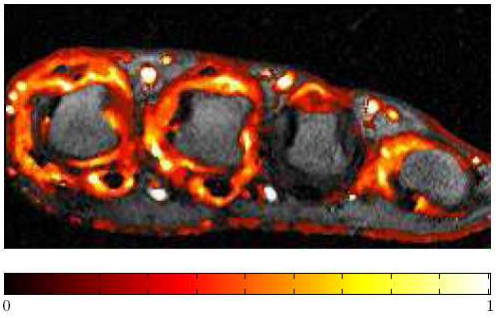


Figure 7.12: Parametric map of *ME* and a colourbar. The lower values of the parameters are plotted in red, then yellow and white as the values increase. The maximum value of the parameter corresponds to the blood vessels' enhancement.

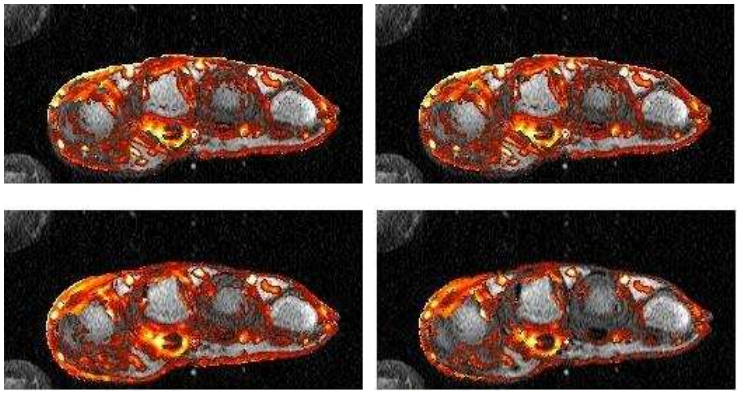


Figure 7.13: A parametric map of *ME* obtained with the moving-window approach (left top), with the benefit of a spatial median filter (right top), with the further benefit of the registration (left bottom), and a parametric map obtained using the model-based method (right bottom).

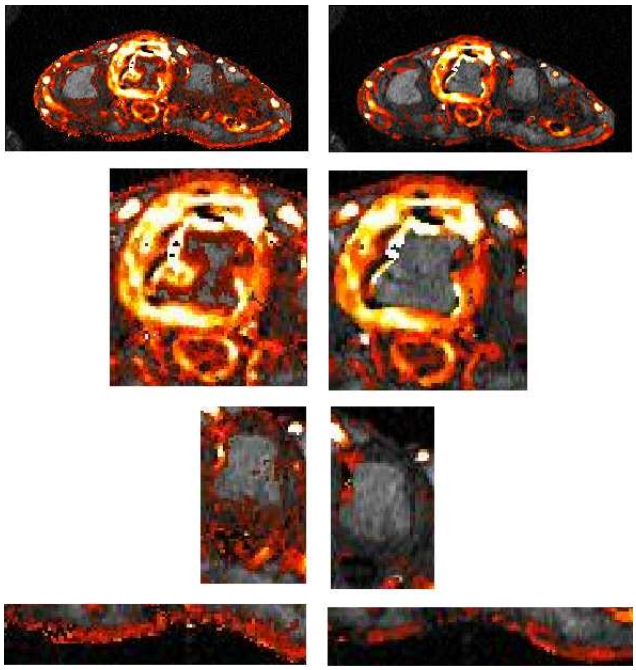


Figure 7.14: Parametric maps of *ME* obtained with the moving-window (left) and model-based (right) approaches. The blowups show reduced skin enhancement, changes in behaviour inside the blood vessels and bone interiors.

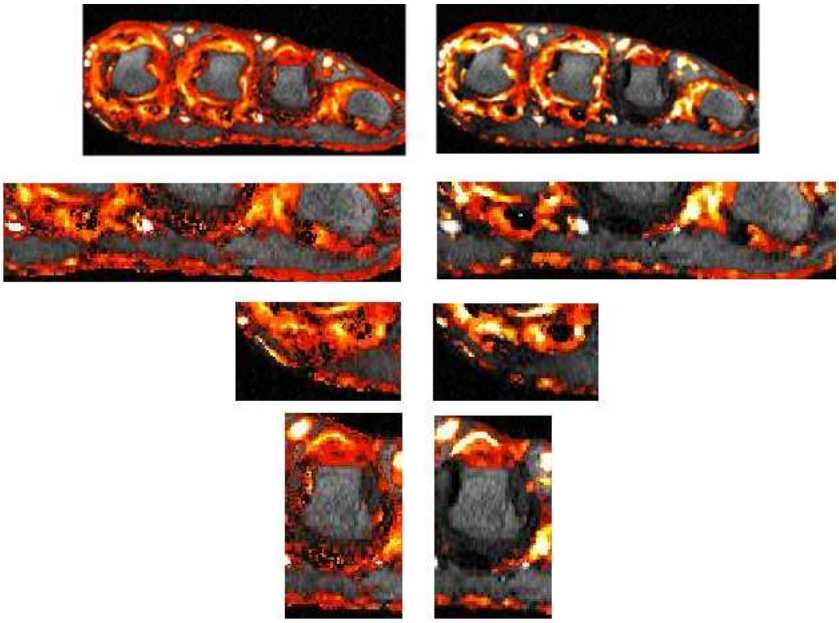


Figure 7.15: Parametric maps of *IRE* with the moving-window (left) and model-based (right) approaches.

tion 7.4.5

Besides heuristics *ME* and *IRE*, clinicians often use a range of parameters derived from $\hat{I}$ signals; examples include 'time to reach 90% of *ME*' [119]. It is clear that with the approach presented the estimation of such parameters is straightforward.

In some studies the difference in the parametric maps obtained by different algorithms is not dramatic, in others it is more significant. Even if we observe a difference in the number of pixels contributing to the parametric maps, the values of *IRE* and *ME* for the accepted pixels are comparable. However, estimation of the heuristics from the approximations fitted through noise allows minimising the noise influence. Therefore, the proposed method provides greater confidence in the measurements.

7.4.3 T_{onset}

Another important measure used to characterise tissue behaviour is the time of onset of enhancement [119]. This time estimated using the moving-window algorithm (T_{IRE} in Figure 7.16) corresponds to the instant where the maximum of the gradient occurs. This estimated can be adjusted to the actual time of onset, given an estimate of *IRE*, its instant of occurrence, and an assumption of a baseline of 1. A simple approach is to determine the straight line intersection illustrated in Figure 7.16. Let us define this time as T'_{onset}.

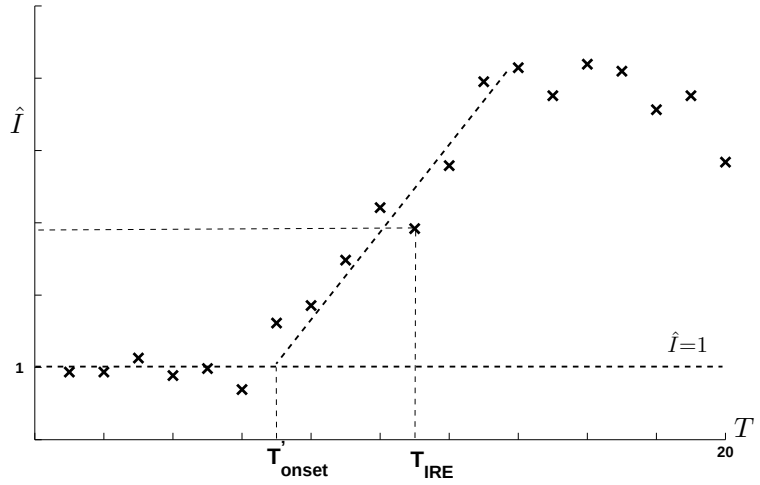


Figure 7.16: Determining T_{onset} given T_{IRE} and *IRE*. Intersection of the maximum gradient line and a baseline defines T'_{onset}. The point T_{IRE} is estimated as an average within a moving window of size 5.

The model-based method provides an alternative approach (t_1 of Figure 7.1) and we

have compared these.

Two histograms of the statistic T_{onset} estimated for one of the DCE-MRI slices with the model-based and improved moving-window approaches are shown in Figure 7.17; the statistic has only been computed at locations deemed 'interesting' in both approaches. Figure 7.18 shows the distribution of per-pixel differences in these estimates.

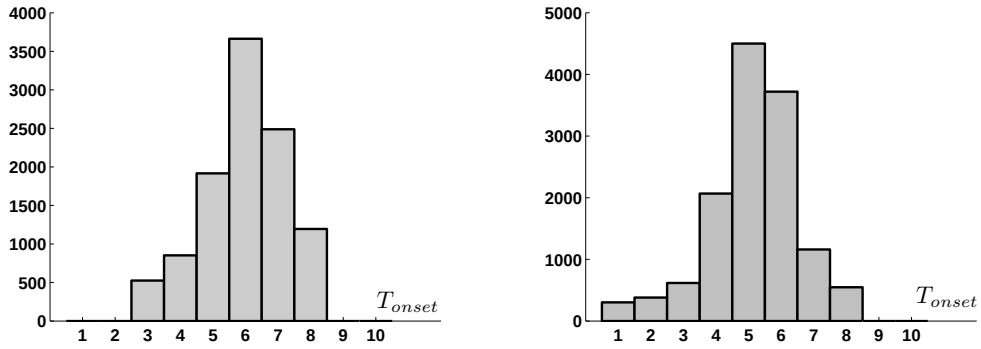


Figure 7.17: Histograms of T_{onset} estimated with the model-based (left) and modified moving-window (right) approaches.

The histogram in Figure 7.18 exhibits bias, obtained as a result of per-pixel subtraction of t_1 from T'_{onset}. Recalling that the size of the window is 5, the histogram demonstrates that if the duration of the wash-in is less than 5 time instants then we might expect $t_1 > T'_{onset}$, and if the duration is more than five then $t_1 < T'_{onset}$. Figure 7.18 suggests that the latter case dominates. Examples are shown in Figure 7.19.

In the great majority of cases, when estimates are different, these estimates differ in magnitude by at most 1 and so there is arguably no difference, but there are still many cases in which the difference is appreciable (around 20% of locations of interest). On inspection, it is hard to say which approach generates more clinically plausible interpretation of $\hat{I}$.

An example of this is shown in Figure 7.19 (left); onset of enhancement estimated with the improved original approach $T'_{onset} = 7$ and with the model-based approach $t_1 = 6$. This experiment has been performed on 100 randomly selected curves.

In some cases the estimates give the same results, however, in about 20% of cases the estimate of T_{onset} provided by the model-based approach is preferable to that of the moving-window. One such example is shown in Figure 7.19 (right).

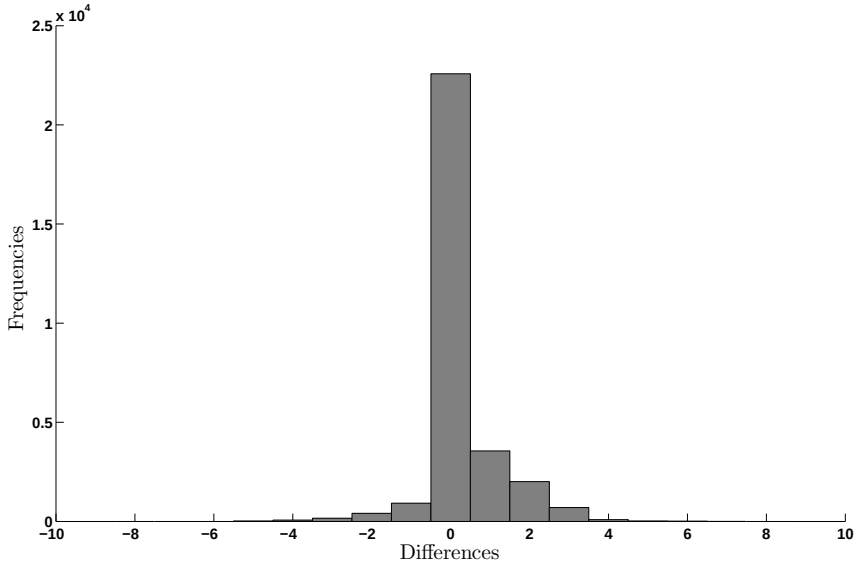


Figure 7.18: Distribution of the differences between T_{onset} estimates.

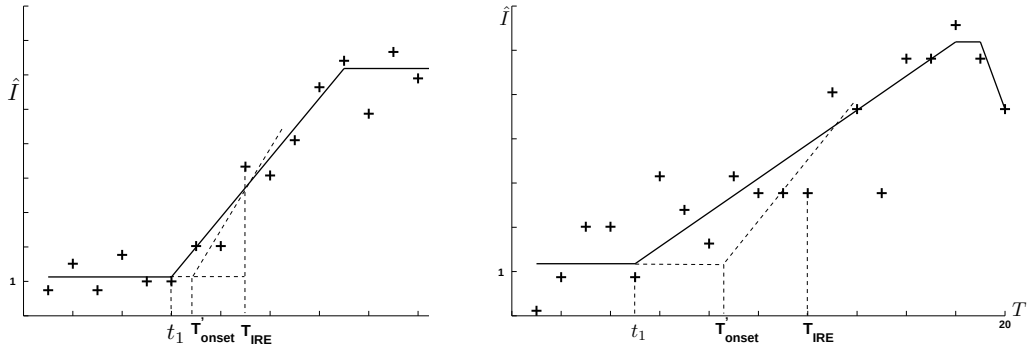


Figure 7.19: Different estimates of T_{onset}. Left: The model-based approach generates a more reliable result. Right: The choice of which estimate should be preferred is unclear.

7.4.4 Number of enhancing voxels N_{total}

The correct estimation of a parameter such as the percentage of enhancing voxels N_{total} is useful for the evaluation of a patient condition (assessment of the extent of RA and tracking disease progression) more accurately, and potentially estimating the degree of RA based on quantitative rather than the pain scoring or visual assessment based [174] measures.

Subtraction of pre- and post-contrast images is often used to compute a number of enhancing voxels [21, 175, 233, 278]. Areas with synovitis, erosions, and bone oedema are manually or semi-automatically pre-segmented, and the number of enhancing voxels located within these areas is computed to score the disease progression or response to

treatment. With the moving-window approach this judgement has been made via criteria such as requiring $ME > 1.2$ and $T_{onset} < 60s$ [195].

Our method permits an improved measure: we label as 'non-enhancing' any voxel which assumes the label M_0. Further, we are able to identify tissues that did not absorb enough Gd-DTPA to exhibit maximum enhancement.

Table 7.3 shows the number of enhancing pixels (N_{total}) normalised to the total number of pixels processed in the temporal slice (that is, pixels within the joints' interior) for a random selection of DCE-MRI studies acquired with the high and low field scanners. Due to the minimised artefactual enhancement, there is a clear qualitative difference in these numbers. Figure 7.24 illustrates the results obtained on several slices with these approaches.

Study	Model-based approach	Moving-window approach
High-field data		
1	0.58 (22%)	0.68
2	0.59 (6%)	0.63
3	0.42 (10%)	0.51
4	0.39 (13%)	0.57
5	0.39 (2%)	0.57
Low-field data		
1	0.68 (3%)	0.97
2	0.57 (6%)	0.86
3	0.32 (1%)	0.89
4	0.42 (4%)	0.97
5 (healthy control)	0.1 (10%)	0.98

Table 7.3: The number of enhancing voxels estimated with the moving-window and model-based approaches. In the second column, we include the percentage of pixels corresponding to locations where we observe continuous Gd-DTPA absorption.

7.4.5 Codings of the Gd-DTPA take-up and parametric maps of *ME* and *IRE*

We are further able to map the Gd-DTPA take-up by plotting the adoption of models $M_1 - M_3$. Blood vessels usually assume M_3, indicating (as expected) presence of the wash-out phase. Most signal intensity vs. time curves corresponding to the disease-affected tissues normally assign models M_2 or M_3, indicating a plateau of intensity and full absorption of the Gd-DTPA. However, some require the model M_1, suggesting a wash-in phase continues after the procedure has been completed.

Figure 7.20 illustrates this for different studies; a predominance of red indicates the procedure is incomplete. Information of this nature may clearly be of use in tuning the procedure.

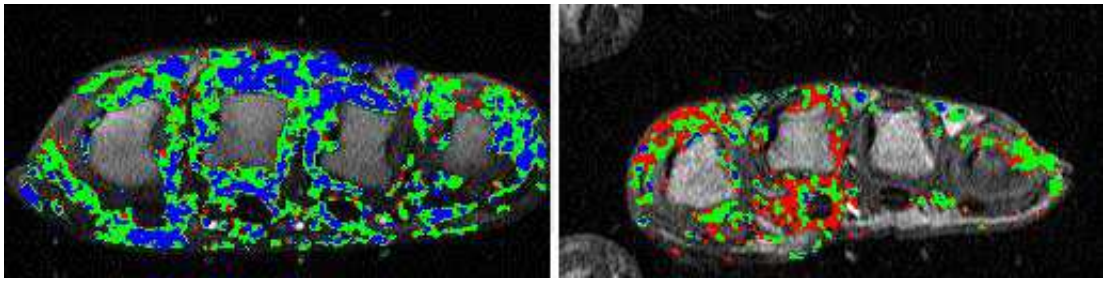


Figure 7.20: Gd-DTPA take-up maps: highly perfused tissues and blood vessels are usually modelled by M_3 (blue), inflamed tissues by M_2 (green), tissues where procedure is incomplete by M_1 (red).

Figures 7.21 and 7.22 show parametric maps for *ME*, *IRE* and Gd-DTPA uptake built for a DCE-MRI study of the MCPJs using the model-based and moving-window approaches.

We note that:

- The model-based technique permits the accurate separation of non-enhancing bone marrow and fat, muscle, which assumes low-intermediate enhancement, and synovitis, which normally assumes high-intermediate for active RA patients.

- Before images were aligned, voxels located in the skin area were composed of a mixture of the tissues (dermis and epidermis), and exhibited intermediate enhancement. When the skin layers were aligned with the registration approach, only pixels within the epidermis remained enhanced. This has reduced the overall width of the skin enhancement.

- With the moving-window algorithm 'interesting' voxels are those, where $ME > 1.2$ and $T_{onset} < 8.5$. Often, heuristics estimated for signal intensity vs. time curves located within the blood vessels did not satisfy this approach, and the corresponding pixels were eliminated. With the model-based technique only pixels whose $\hat{I}$ curves assumed M_0 were excluded. These pixels never appear within the blood vessels. Therefore, blood vessels partially ignored in the maps obtained with the moving-window method are in evidence with the model-based technique.

- Parametric maps obtained with the proposed method are less corrupted by noise

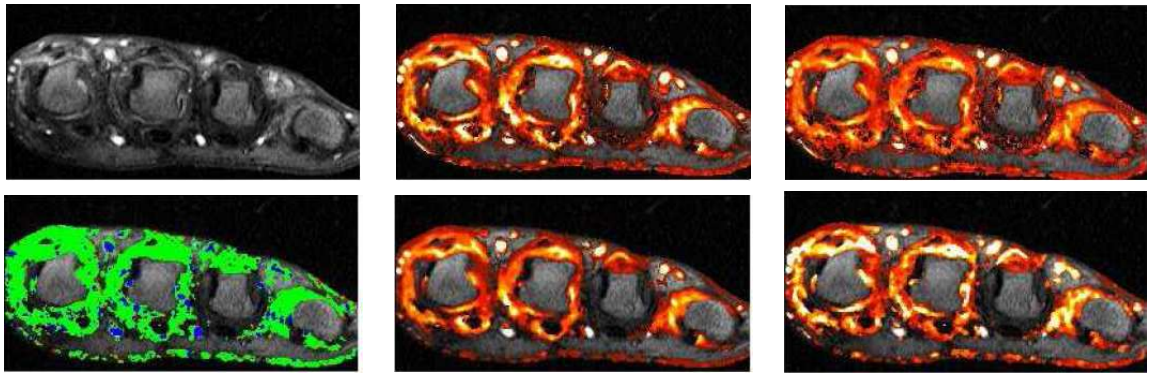


Figure 7.21: A dynamic frame from a temporal slice with a small amount of motion analysed with the moving-window and model-based approaches. Top row: Post-contrast image, parametric maps of *ME* and *IRE* obtained with the moving-window approach. Bottom row: Gd-DTPA take-up map, parametric maps of *ME* and *IRE* obtained with the model-based approach.

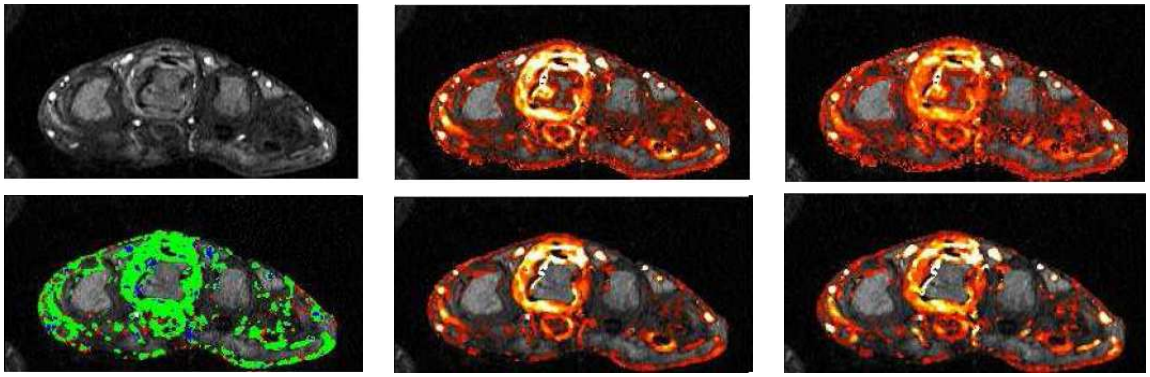


Figure 7.22: A dynamic frame from a temporal slice with a significant motion analysed with the moving-window and model-based approaches. Top row: Post-contrast image, parametric maps of *ME* and *IRE* obtained with the moving-window approach; Bottom row: Gd-DTPA take-up map, parametric maps of *ME* and *IRE* obtained with the model-based approach.

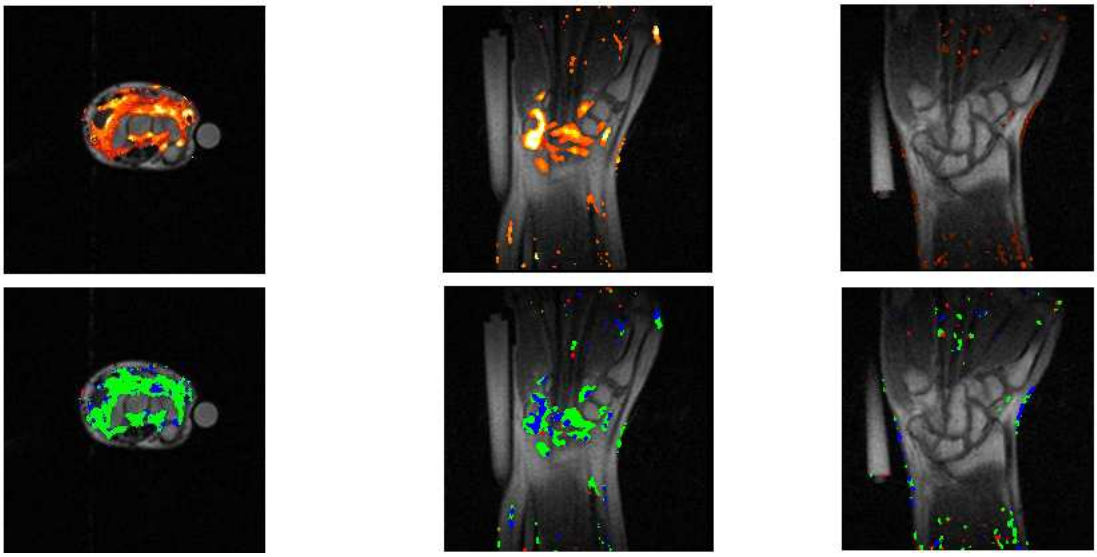


Figure 7.23: Parametric maps of *ME* (top) and Gd-DTPA uptake (bottom) for datasets acquired with the low-field scanner. Left: Patient with active RA, SE sequence. Middle: Patient with active RA, GRE sequence. Right: Healthy control, GRE sequence.

and provide a clearer visualisation of the bone contours, blood vessels, and disease-affected areas.

- We are able to detect areas of subtle and sparse enhancement (Figure 7.23) which are very difficult to locate by viewing images one by one or even with the subtraction method. These areas are of particular interest for the assessment of RA in wrist studies. Figure 7.23 illustrates maps of *ME* and Gd-DTPA take-up constructed with the model-based method for DCE-MRI data of the wrist joint acquired by the low field scanner.

The data obtained from the low-field scanner was also processed with the moving-window technique. We experimented with windows of different size – a larger size window (> 5) smoothes out the details of the signal intensity vs. time curves and does not allow for the efficient differentiation of tissues. Figure 7.24 illustrates the parametric maps constructed with the moving window (top) and model-based (bottom) techniques. The joints were segmented automatically before the algorithms have been applied. Figure 7.25 shows parametric maps of *ME* when no segmentation was applied.

In the parametric map constructed with the moving-window technique pixels located within the marker and bone interiors were colour-coded, implying response to the Gd-DTPA. In the corresponding map obtained with the model-based approach no pixels within the bones or marker were coloured as they assumed model M_0.

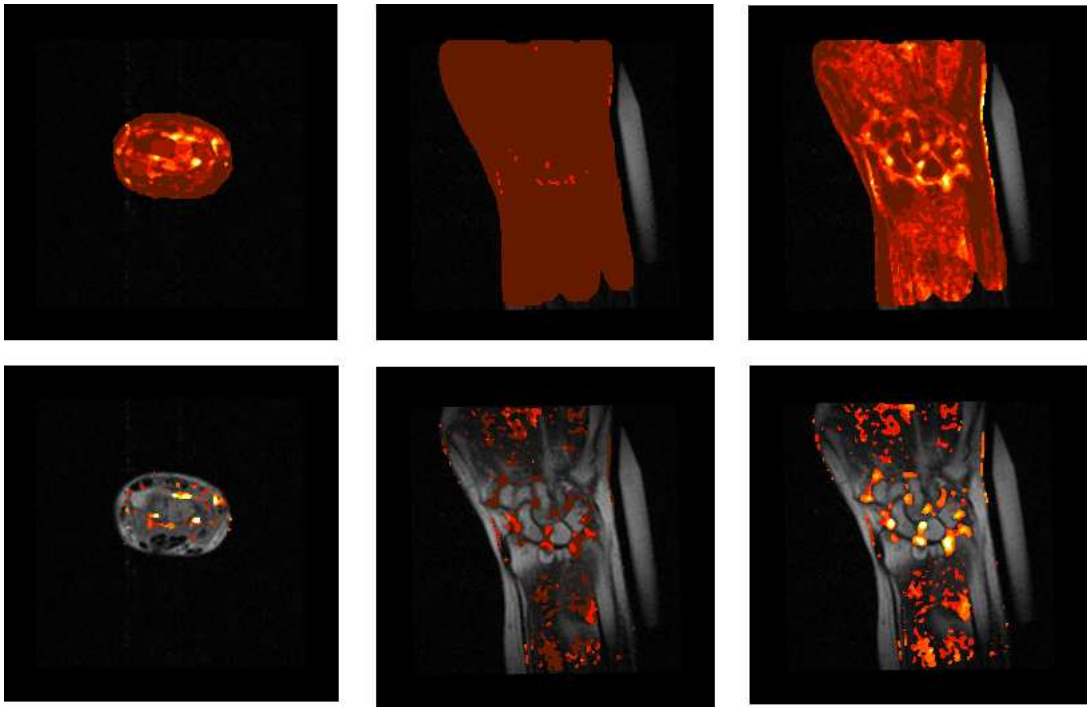


Figure 7.24: Top: Parametric maps of *ME* and *IRE* obtained with the moving-window technique (window size 5×5). Bottom: Corresponding results obtained with the model-based technique.

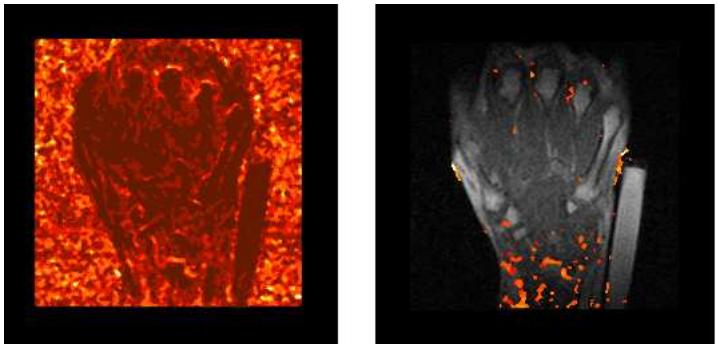


Figure 7.25: Parametric maps of *ME* obtained with the moving-window technique (left) and the model-based (right) techniques. The joints' contour was not segmented.

7.5 Results

This section presents the experimental results that confirm the reliability of the model-based approach for RA assessment, and illustrate the benefits of such analysis for evaluation of the disease progression[2].

Healthy controls. One of the datasets was acquired from a subject with no RA, but suffered from occult wrist pain – possibly due to a ganglion in the wrist joint, that was not found in the post contrast sequences [50]. Pre-, post-contrast images and parametric map of *ME* and Gd-DTPA take-up constructed for this subject are shown in Figure 7.26.

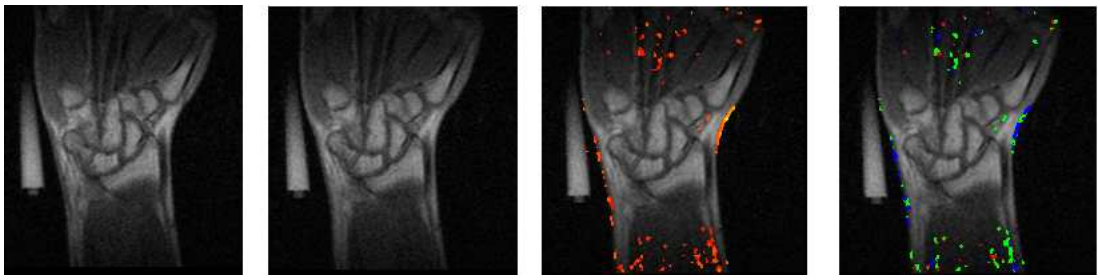


Figure 7.26: Pre-, post-contrast images and parametric maps of *ME* and Gd-DTPA uptake for patient with no diagnosis of RA.

Disease assessment in the healthy controls is problematic. It is challenging to differentiate RA patients from those who suffer from wrist pain but are not affected by the disease.

Currently, this analysis is performed with a ROI based method. Usually the position of the ROI is chosen at the patient's suggestion or presumptively based on the anatomical landmarks [50]. The ROI covers the area where the patient feels pain the most. Such evaluation might result in erroneous diagnosis if the ROI is misplaced.

The proposed technique eliminated the problem with ROI placement and allows objective assessment of such patients. Medical experts from Frederiksberg hospital performed a visual inspection of the results illustrated in Figure 7.26. They concluded that the maps clearly show that the patient does not have inflammatory arthritis, however exhibits some tissue reaction on the contrast agent. Quantitatively for this patient we found that $N_{total} = 0.06$, *ME* is on average 1.3 with the maximum at 2.4.

[2]The observers viewed images with prior knowledge of which method was used to generate them (i.e. not a blind study).

The condition of this patient was also evaluated using ultrasound with a Doppler technique; experiments were performed by specialists from the Parker Institute, Denmark. The results obtained with the model-based technique correspond to the US evaluation, which showed mild colour Doppler activity in the wrist leading to the conclusion that the patient suffered from a mild unspecific irritation of the wrist.

Parameters for RA patients and healthy controls. Figure 7.27 illustrates *ME* and N_{total} estimated for 33 subjects with active RA (129 temporal slices) and 4 controls (12 temporal slices). The maximum enhancement in healthy controls is always below 20%, with the total number of enhancing pixels N_{total} less than 5%. The enhancing pixels are located sparsely in the skin and blood vessels. On average in slices corresponding to the patients with active RA *ME* is between 2.5 and 4 with the maximum reaching 14, and N_{total} is between 0.3 and 0.4, depending on the degree of inflammation. This could provide a good metric for identifying healthy controls and measuring disease progress of RA patients.

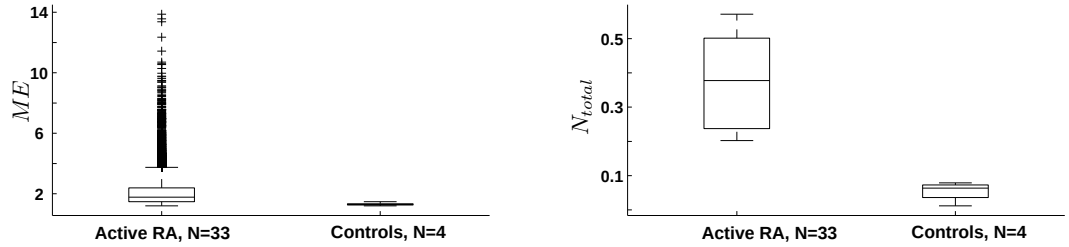


Figure 7.27: Box-and-whisker plot of *ME* (left) and N_{total} (right) for patients with active RA and healthy controls; N – is a number of patients.

Assessment of data from follow-up examinations. We have analysed several datasets acquired in follow-up examinations. The first patient was scanned 2 times in the axial plane after the injection of the intra-articular glucocorticoid. Figure 7.28 illustrates parametric maps.

Parametric maps in Figure 7.28 suggest a diminished perfusion in the visible pannus[3] and clear improvement of the patient's condition. This is an expected treatment effect. The experts also confirmed that this information is not available with the conventional images after Gd-DTPA contrast, where the patient had high and unchanged synovitis at the follow-up.

[3]Pannus is a medical term for a hanging flap of tissue. Pannus consists of skin and fat. In people suffering from rheumatoid arthritis, pannus tissue eventually forms in the joint affected by the disease, causing loss of bone and cartilage. [260]

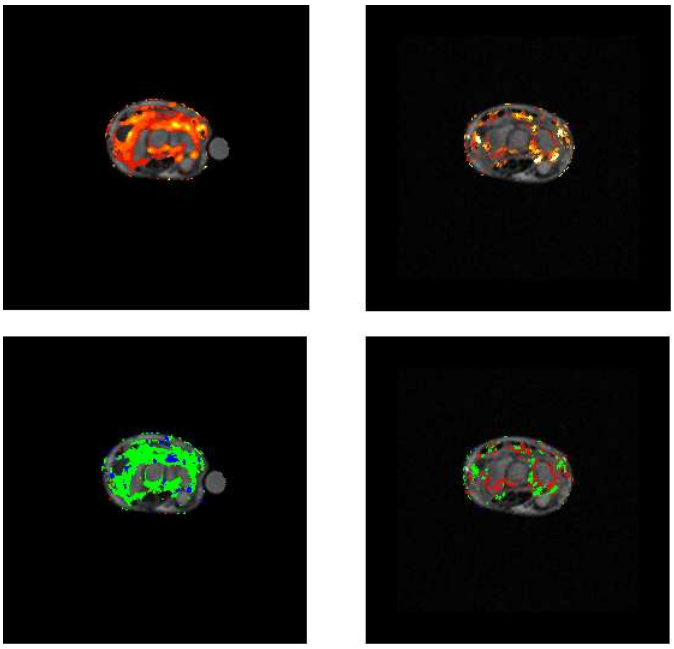


Figure 7.28: Parametric map of *ME* and Gd-uptake constructed for RA patient in the first (left) and last (right) examinations. The patient has shown improvement.

Another patient was re-scanned in the coronal direction three times after injection of a steroid. He had had short clinical relief, but got worse a few days after each injection with pain and discomfort in the wrist. This information is reflected by the parametric maps shown in Figure 7.29. The first image was acquired at the first examination, second – after the injection of a steroid and illustrates minor improvements. After a short period, this patient got worse, and the image on the right illustrates this.

7.6 Conclusion

Quantitative analysis of contrast enhanced dynamic MRI datasets involved in monitoring and assessment of RA has been discussed. We presented a technique that allows extraction of various parameters such as ME, IRE, and T_{onset}, which are essential for the assessment of the data acquired from patients with rheumatoid arthritis.

These heuristics are derived from linear approximations rather than from raw signal intensity curves, making their estimation robust to the subjective opinion of the operator and noise effects. The choice of the 'best' model for each curve allows for accurate tissue classification. Voxel-by-voxel analysis eliminated a need for ROI placement and a choice of an optimal moving-window size, which makes the results fully automated,

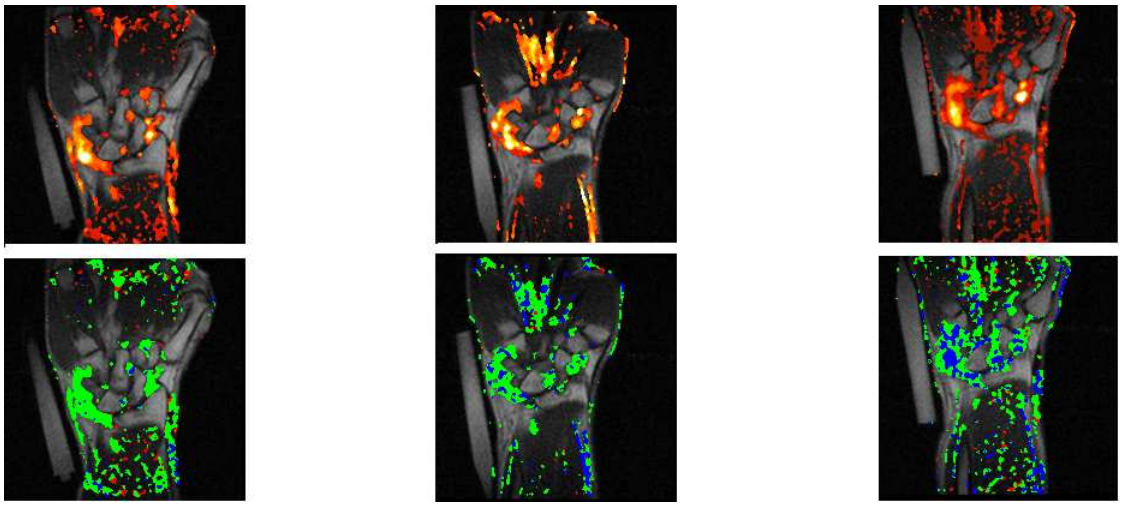


Figure 7.29: Parametric maps constructed for a patient re-scanned 3 times after the injection of a steroid. The first image was acquired at the first examination, second – after the injection of a steroid and illustrates minor improvements. After a short period, this patient got worse, and the image on the right illustrates this.

easily reproducible and objective.

This approach to the extraction of the heuristics and parametric maps permits easier visual assessment of the degree of inflammation in RA patients, which allow for a more accurate analysis of the extent of the disease and differentiation of various tissues. Moreover, indications of Gd-DTPA take-up, hitherto unavailable, provide improved identification of tissue behaviour according to its temporal pattern of the contrast agent uptake.

Among the randomly chosen patient cohort in this pilot study, parametric maps of *ME* and *IRE* constructed for healthy and RA patients are noticeably different, corresponding to our expectations and clinical judgments provided by the experienced observers and radiologists. The results demonstrate that the model-based technique is sensitive and may be useful in the diagnosis and follow-up examinations of the patients who are receiving disease-controlling treatment.

The method provides a numeric evaluation upon which clinical and research decisions can confidently be made. The possibility of the evaluation of the data acquired from the low-field MRI scanners further extends the usability of the method as such scanners are more patient friendly [214] and cost efficient compared to the high-field machines.

Chapter 8

Blood vessel segmentation

The assessment of disease progression or patient response to treatment can be done via constructing parametric maps and computing the number of enhancing pixels within a certain threshold of values of the parameters ME and IRE. In terms of these parameters behaviour of the blood vessels corresponding to the veins is very similar to the behaviour of the inflamed synovium: normalised signal intensity vs. time curves exhibit significant intensity change, assume model M_2 or M_3 and high values of ME and IRE. This might complicate objective and visual assessment of the data. In this chapter we will discuss an algorithm for the detection of the blood vessels in DCE-MRI data of the MCP joints acquired with the high field scanner.

8.1 Segmentation algorithm

The tissues within the joints' envelope will be classified into vessel / non-vessel with a three-stage, coarse-to-fine approach.

- Firstly, the shape of the normalised signal intensity vs. time curves is analysed by principal components analysis (PCA) applied to temporal slices. The behaviour of the normalised signal intensity curves extracted from pixels located within the blood vessels and synovium is similar at the baseline, wash-in, and plateau phases. At the wash-out phase, their behaviour starts to diverge: the vessels are expected to leak out the contrast agent earlier. However, due to the short duration of this phase,

not all pixels within the blood vessels exhibit a wash-out. Therefore, classification of the tissues based on the signal intensity vs. time curve behaviour during the wash-out is imperfect.

- Subsequently, these results were refined using more specific parameters such as *ME* and *IRE*.
- In the last stage, the spatial relationship between the pixels located within the vessels were considered through probabilistic modelling.

8.1.1 Principal component analysis

Firstly, to assess a pattern of behaviour in the signal intensity vs. time curves, we applied PCA to 20 temporal slices randomly chosen from 10 DCE-MRI datasets of the MCP joints. Pixels at which the normalised signal intensity vs. time curves assumed M_0 and M_1 were excluded. Figure 8.1 illustrates the mean and principal components from this data sample.

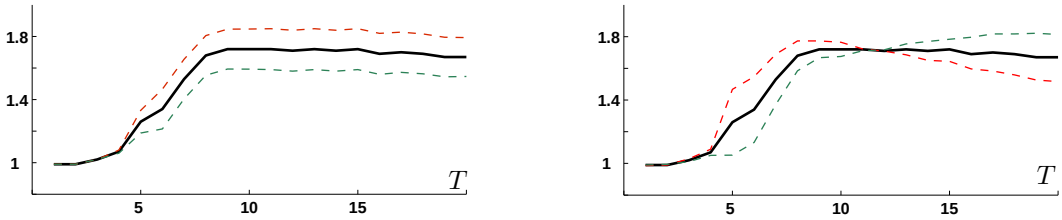


Figure 8.1: The mean (in bold), the mean $\pm$ 2 standard deviations of the first principal components (left), and the mean $\pm$2 standard deviations of the second principal component (right). The mean + component is shown in red; the mean - component – in green.

The first two components capture 97% of the information present in the data. The shape of the temporal course of the first component corresponds to the behaviour of the inflamed tissues, and the second to the blood vessels.

Secondly, we manually outlined synovial tissue and blood vessels in these 20 DCE-MRI slices. Figure 8.2 illustrates the mean and the mean $\pm$ two first principal components, estimated for pixels located within the perfused tissues (left) and the blood vessels (right).

Figures 8.2 illustrates that the mean $\pm$ 2 principal components(standard deviations) corresponding to the blood vessels exhibits a clear wash-out, and the mean $\pm$ 2 principal components corresponding to the synovial tissues might exhibit an intensity plateau and slight intensity increase. This behaviour is consistent and reflects presence of the wash-out phase in the blood vessels.

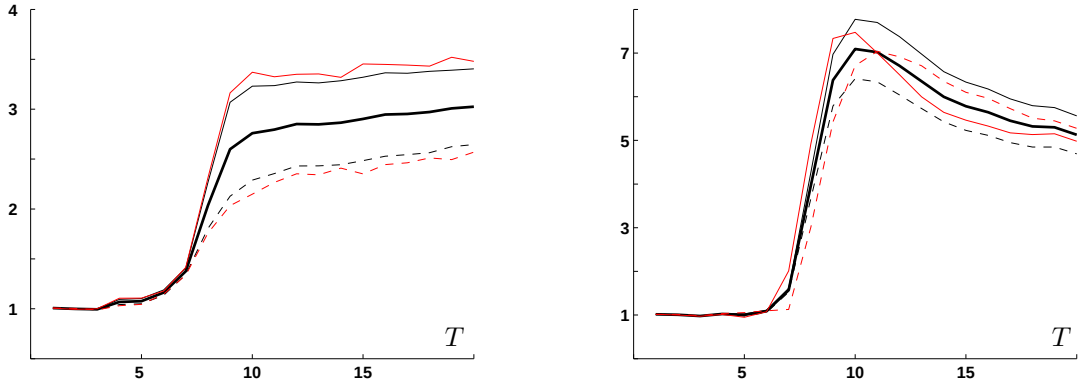


Figure 8.2: The mean $\pm$ 2 standard deviation of the first (in red) and second (in black) principal components estimated for the inflamed synovial tissue (left) and blood vessels (right). The mean + component is shown in solid line.

Using this result we attempt to discriminate between the vessels and synovial tissue. Let μ and v_1 denote the mean and the first principal component estimated for the blood vessels, and η and w_1 the mean and the first principal component estimated for the synovial tissue outlined in a 10 slice data sample. Thus, the following two models can be tried:

$$\begin{aligned} \phi &= \mu + \alpha v_1 \\ \varphi &= \eta + \beta w_1 \end{aligned}$$

where α and β are multipliers for the principal components; ϕ describes behaviour typical for the blood vessels and φ for the synovial tissue.

All signal intensity vs. time curves in a sample temporal slice were projected on ϕ and φ to calculate α and β. This reflects the correlation between the model and a curve. Figure 8.3 illustrates a sample $\hat{I}$ curve (in black) derived from a blood vessel (left) and a synovial tissue (right) projected on ϕ (red) and φ (blue).

Using these PCA descriptors each $\hat{I}$ curve can be assigned a provisional vessel / non-vessel label: if for a $\hat{I}$ curve $\alpha < \beta$, then the corresponding pixel is assigned a label 'vessel', otherwise, when $\beta > \alpha$, 'non-vessel'. Figure 8.4 illustrates the results obtained for a sample DCE-MRI slice. Euclidian distance based comparison is valid as eigenvalues are similar (i.e. two sets have approximately the same variance).

At this stage, approximately 50% of pixels were classified correctly. The algorithm did not deliver false negative results; all errors were false positive. This result suggests

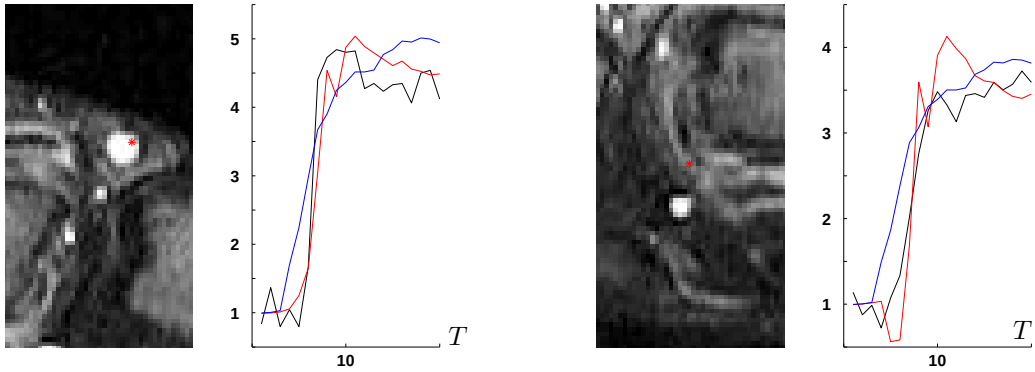


Figure 8.3: Sample $\hat{I}$ curves (in black) derived from a blood vessel (left) and synovial tissue (right) projected on ϕ (red) and φ (blue).

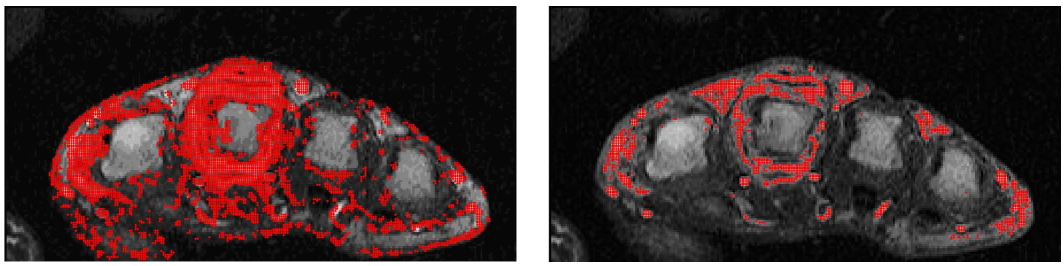


Figure 8.4: Left: A post-contrast image from a sample DCE-MRI slice with enhancing pixels shown in red. Right: The same image after classification with PCA: pixels classified as vessels are shown in red.

that PCA can only crudely classify the tissues; it fails to separate synovial tissue from the blood vessels, however isolates tissues with a high vascularity from the rest of the tissues present in the joints.

We will proceed by analysing the behaviour of the signal intensity curves corresponding to the pixels labelled as vessels at this stage.

8.1.2 Spatial relationship

Empirically, it was found that distributions of the heuristics *ME* and *IRE* on the pixels classified with PCA as vessels are and Gaussian[1]. Figure 8.5 illustrates distribution of *ME* and *IRE* derived for a sample DCE-MRI slice.

The measure T_{onset} was less informative. The wash-in phase in the vessels that enhance as the result of the inflow effect starts early (less than 4 time instants); the wash-in

[1]Kolmogorov-Smirnov statistical test of goodness of fit confirmed this observation. The GMM was first fit, then KS done on each Gaussian

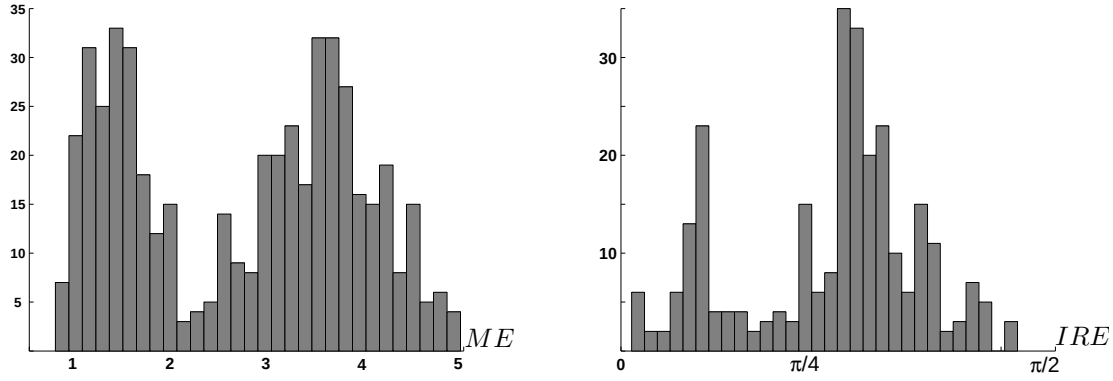


Figure 8.5: Histograms of *ME* and *IRE* for a sample DCE-MRI temporal slice of the MCP joints acquired with the high-field scanner.

of the vessels enhancing in response to the contrast agent starts at about 3^{rd}-5^{th} time instant. There are no strict constraints on when the wash-in can be expected in the synovial tissues. Therefore, we used only three parameters $< ME, IRE, M >$ to describe behaviour of signal intensity vs. time curves.

We fit a mixture model of two 3 dimensional Gaussian functions to the data with the expectation maximisation algorithm (EM) and label the two components as vessel / non-vessel based on the assumption that the mixture, in which all components of the mean are higher represents a vessel. Unsupervised EM algorithm was chosen over a supervised approach due to the data variability.

We define the set of parameters for each pixel of a dynamic frame in a temporal slice as $X = \{\mathbf{x_i}, i = 1...N\}$ where N is the number of pixels. A corresponding field of labels $L = \{l_i, i = 1...N\}$ is defined, where $l_i \in \{\lambda_1, \lambda_2\}$ denotes a pixel as vessel/non-vessel. The task is to find the assignment of labels to each pixel which maximises the probability of the observed parameters. Assuming independence between pixels, we can write this probability as:

$$p(X|L) = \prod_i p(\mathbf{x_i}|l_i, \theta)t(l_i), \tag{8.1}$$

where t represents a prior on the proportion of each class within the data, and $p(\mathbf{x_i}|l_i, \theta)$ the conditional probability distribution of the parameters $< ME, IRE, M >$ given the label (and model).

Given the independence assumption made, the assignment of labels maximizing the

probability can be found separately for each pixel:

$$l_i^*(x_i) = \arg\max_{l_i} p(\mathbf{x_i}|l_i,\theta)t(l_i) \tag{8.2}$$

The noise present in the data and imprecision in the model we used leave some error (evaluated in Figure 8.8). Therefore, we exploit the fact that the blood vessels have non-negligible spatial support by assuming that neighbouring voxels are likely to have the same label in the absence of significant differences in the grey level between them.

8.1.3 Final steps

Markov random field (MRF) [84] based filtering allows refinement of our initial assignment of labels to L. We define a set of cliques C which represents the connections between pixels in the image (considered as nodes on a graph). We have adopted an 8-connected neighbourhood such that the cliques contain all pairs of pixels which are neighbouring on the 8 compass points (north, north-east, etc.). Note that the cliques are unordered, for example north/south relationships are considered identical.

The energy function, which is a mechanism for modelling contextual information, is described by Equation 8.3, where we omitted dependence on the fixed parameters θ for the sake of clarity.

$$E = -\sum_{i=1}^{N} \log p(\mathbf{x}_i|l_i) + \alpha \sum_{<i,j>\in C} \Psi(f_i, f_j, l_i, l_j) \tag{8.3}$$

$\alpha \geq 0$ controls the relative importance of the terms. Based on the empirical observations, α was set to 1, giving equal weight to both terms, Ψ is defined as the following, where f is a pixel intensity in the post-contrast image, $|C|$ is the number of cliques:

$$\Psi = \begin{cases} 0 & l_i = l_j \\ 1 - e^{-\frac{|f_i - f_j|^2}{2\sigma^2}},\ \sigma = \sqrt{\frac{1}{|C|}|f_i - f_j|^2} & l_i \neq l_j \end{cases} \tag{8.4}$$

For an input set of pixels P and a set of labels L, the goal is to find a labelling or a mapping from P to L which minimises the energy function. This assignment of the labels to the pixels was found with the mincut-maxflow algorithm, which is known to give the global minimum [29, 239].

We can refine region boundaries using the adaptive segmentation based algorithm discussed in Chapter 5. An example of a blood vessel's boundary is shown in Figure 8.6,

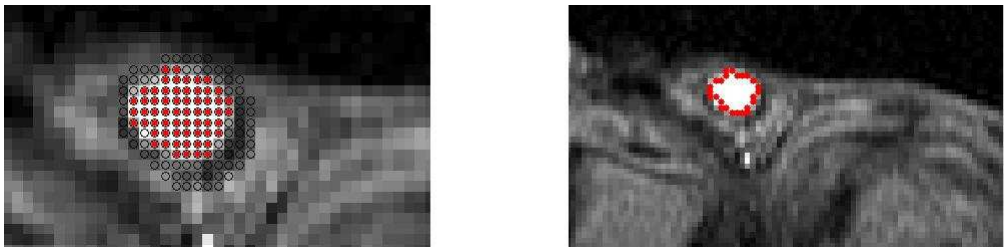


Figure 8.6: Left: A boundary of a blood vessel refined with an adaptive segmentation technique; the position of the initial boundary is shown in black and the final boundary in red. Right: Contour of the red boundary shown on the left.

where the initial boundary is shown in black and the final boundary in red. The method works, but is not necessary given the MRF output.

Figure 8.7 illustrates results of the algorithm on several slices, where blood vessels are outlined in the parametric map of *ME*.

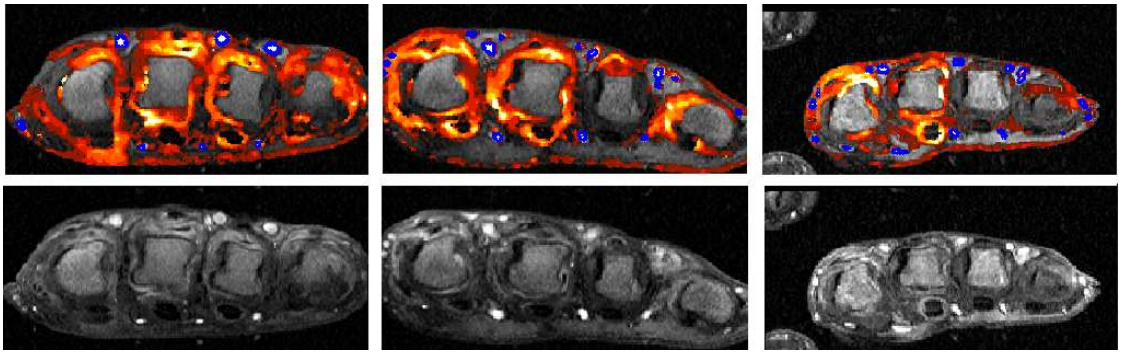


Figure 8.7: Top: Parametric maps of *ME* corresponding to DCE-MRI slices with segmented blood vessels (the contour is shown in blue). Bottom: Corresponding post-contrast images.

8.2 Experiments and discussion

PCA classification As the first step of the algorithm, we attempted to analyse the pattern of behaviour in the blood vessels and perfused tissue using PCA. Figure 8.4 illustrates the results of PCA based classification on a sample DCE-MRI slice; pixels classified as vessels are shown in red.

Our experiments demonstrate that PCA fails to segment out the blood vessels, however, it separates severely affected tissue and blood vessels from the rest of the tissue in the joints.

We performed PCA-based classification on 60 DCE-MRI slices of the MCP joints. The results show that approximately 30% of the total number of pixels within the joints' interior without bones are identified as vessels, $50-60\%$ of these pixels classified correctly.

Numerical evaluation of the pixel reduction is given in Table D in Appendix D. Figure 8.12 illustrates the percentage of the pixels that correspond to the vessels in respect to the total number of pixels.

EM based classification We applied the EM algorithm to the pixels classified as vessels with PCA. Each such pixel was labelled l_1/l_2 based on the value of the heuristics $\{ME, IRE\}$ and a model number M. We evaluate performance of the EM algorithm with a different number and combination of the initial parameters using Receiving Operating Curves (ROC) space[2].

Five DCE-MRI studies (30 slices) with manually segmented blood vessels were used in this experiment. The true positive rate was computed as the ratio between the true positive pixels and a total number of pixels within the blood vessels, and the false positive rate as the ratio between the false positive pixels and a total number of true non blood vessel pixels.

The heuristics ME and IRE and model number M computed for a DCE-MRI slice were normalised to be zero-mean and unit variance. Figure 8.8 illustrates the performance of the best classifiers, where the true positive rate is the sensitivity and the false positive rate is equivalent to one minus specificity.

Clearly, the number and combination of the initial parameters influence performance of the algorithm. In order to choose the best classifier the area under the ROC curve [261] was computed. Before we compute the area under the curve, the convex hull was formed. Using the convex hull approach might have changed the results. However, the difference in the classifiers' performance is significant, and therefore influence of the convex hall was not dramatic. The classifier with the $\{ME, IRE, M\}$ set of parameters delivered the best results. Figure 8.9 illustrates performance of this classifier on 6 post-contrast images.

Pixels identified as false positive normally appear in the disease affected areas and their heuristics are very similar to those of blood vessels. Pixels identified as false

[2] In a ROC curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (100-Specificity) for different cut-off points. Each point on the ROC plot represents a sensitivity/specificity pair. A test with perfect discrimination (no overlap in the two distributions) has a ROC plot that passes through the upper left corner (100% sensitivity, 100% specificity). Therefore the closer the ROC plot is to the upper left corner, the higher the overall accuracy of the test.

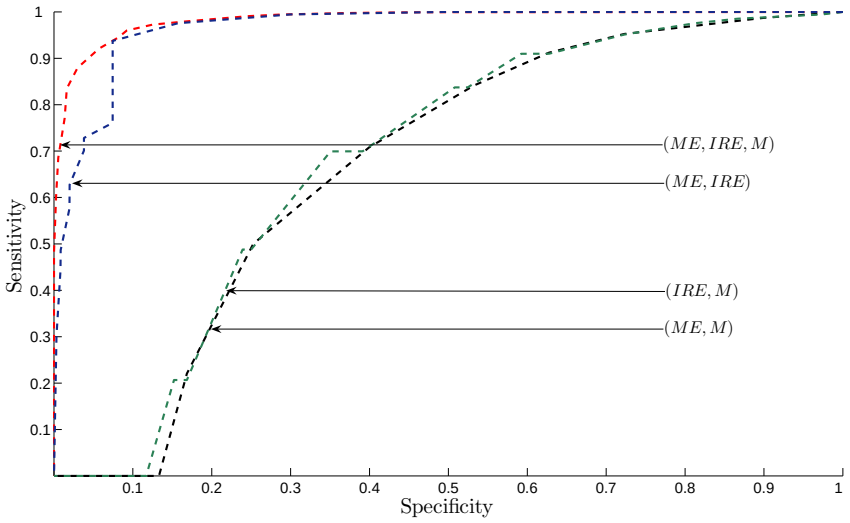


Figure 8.8: ROC curves for EM classifiers with different combination of the measures.

negative normally surround blood vessels and their exclusion / inclusion might be due to observer's mis-detection. Figure 8.10 illustrates false positive and false negative pixels for one of the DCE-MRI slices.

8.2.1 Blood vessel detection

In this section we evaluate performance of the segmentation algorithm on 60 DCE-MRI slices of the MCP joints. Firstly, the ability of the algorithm to detect vessels in temporal slices will be assessed, then we will compare automatically segmented and manually outlined blood vessels using the mutual overlap based metric.

The number of the blood vessels per slice varies from 8 to 17. Table 8.1 illustrates a number of automatically vs. manually detected blood vessels.

Table 8.1: Detection of the blood vessels in temporal slices [Number of blood vessels delivered by the algorithm / Total number of vessel in a temporal slice]; S – scan number; P – patient number.

P/ S	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
S_1	9/9	14/14	16/16	14/14	9/9	12/12	12/12	12/12	10/10	17/17
S_2	9/9	11/11	14/15	12/13	11/13	9/10	12/13	10/10	9/9	14/16
S_3	9/9	11/12	17/17	14/14	12/12	11/11	11/12	8/9	8/10	14/15
S_4	8/9	13/13	16/17	12/14	9/10	12/13	10/12	6/8	12/12	15/15
S_5	8/8	10/11	17/17	15/15	11/11	12/12	13/13	10/10	12/13	15/15
S_6	9/9	12/12	16/17	16/16	11/11	10/10	13/13	12/12	12/12	16/16

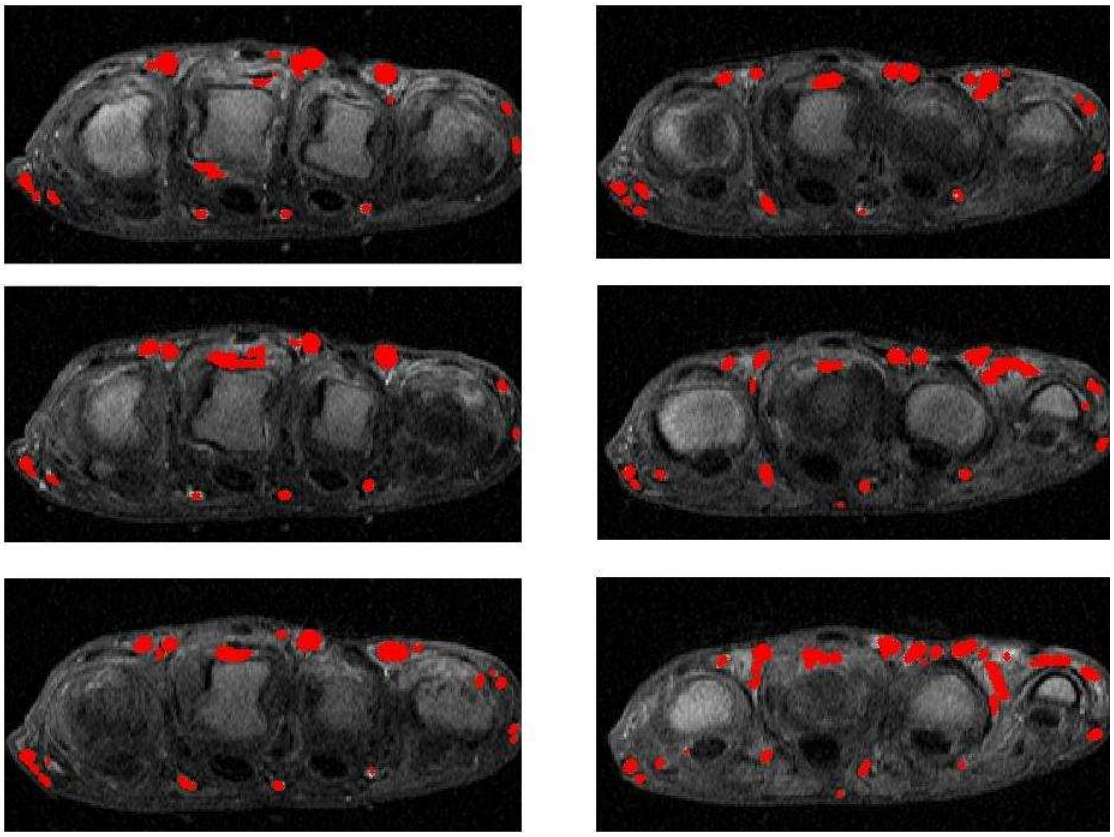


Figure 8.9: Six post-contrast images from different DCE-MRI slices; pixels classified as vessels are shown in red.

In our experiments, the algorithm did not deliver false positive results; thus, no post-processing to remove over-segmented regions is needed. However, some blood vessels of small size (area less than 5 pixels) were not detected. The location of these under-segmented vessels can be recovered when a 3D vessel tree is reconstructed. To illustrate the approximate location of the vessels we applied an interpolation technique [137] on the vessels' location. Figure 8.11 illustrates the result.

The quality of the automatically segmented vessels was compared with manually outlined regions on each step of the algorithm (PCA, PCA+EM, PCA+EM+MRF); for the 60 temporal slices we computed a number of pixels within the vessels in respect to the overall number of detected pixels. Figure 8.12 illustrates the results.

There is an incremental increase in segmentation quality: after PCA based classification about 50% of pixels were classified correctly; EM refines this result to 70% of the detected pixels representing vessels. MRF filtering increases this to 90%.

This experiment illustrates that the proposed strategy generates promising results. On average the algorithm detects 92% of vessels in dynamic MRI slices of the MCPJs, with

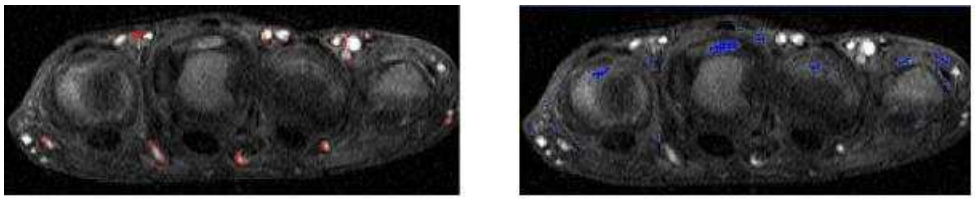


Figure 8.10: False negative in red (left) and false positive in blue (right) pixels detected with the EM classifier.

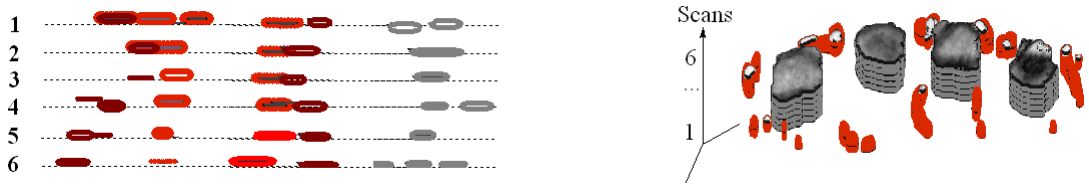


Figure 8.11: Left: Location of the blood vessels in a sample DCE-MRI study. Right: Depiction of the vessel tree and bone interiors (3D view).

mutual overlap between GT and obtained segmentations exceeding 90%.

8.3 Conclusion

The algorithm presented in this chapter allows for accurate segmentation of the blood vessels from DCE-MRI datasets of the hand joints acquired with high-field scanner. The algorithm has not been applied to the data acquired with the low-field scanner, as the blood vessels are not visible in these images.

We demonstrated how a combination of approaches motivated by the physiological properties of the individual tissues, such as speed of absorption and concentration of a contrast agent, can be employed to describe the behaviour of the vessels and synovitis.

MRF-based filtering incorporating prior information about the smoothness and the data term, describing intensities of the pixels in post-contrast images, was used to exploit the fact that the pixels located within the blood vessels are unlikely to have significant differences in the grey level. It allowed us to remove ambiguities due to low contrast and partial volume effects, which significantly improved the segmentation quality.

In our application the number and shape of vessels vary significantly within the study, and the vessel segmentation problem was solved on a slice-by-slice basis, rather than in 3D space. However, for an application where the slice thickness is smaller, the extension of the algorithm would be straightforward. Future work should focus on testing of the method on datasets acquired from different joints, extending the approach to perform on 3D images, and investigation of the algorithm's performance when other heuristics are

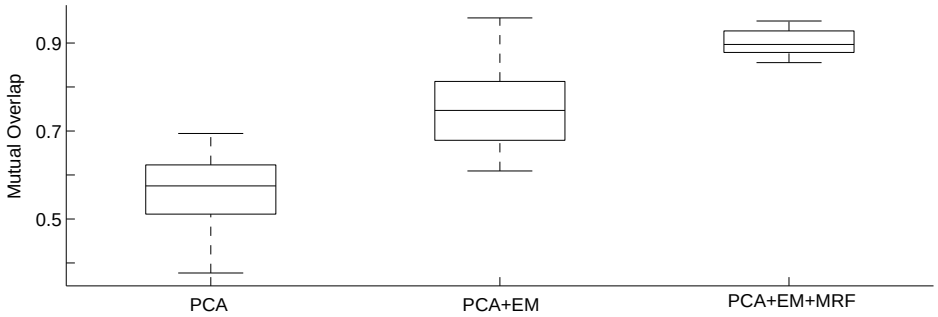


Figure 8.12: A number of pixels within the vessels in respect to the total number of detected pixels at different steps of the algorithm: PCA / PCA+EM / PCA+EM+MRF.

included.

Chapter 9

Conclusions

9.1 Summary of work

Magnetic resonance was first applied in regard to biological tissue assessment in the late 1970s, however its recognition as a clinical imaging modality occurred only in the late 1980s. It is still a developing technology, moving from the manual qualitative to fully automated quantitative analysis of tissue conditions.

With constantly evolving hardware and the increasing use of MRI in clinical applications, there is a high demand for efficient and robust image reconstruction and analysis algorithms. The use of quantitative measurement techniques that can represent information about the tissue in numerical rather than subjective terms increases [181].

The current trend is towards developing fully automated strategies that will be independent of a particular machine, magnetic field strength, pulse sequence or operator. The results obtained with such methods are potentially more reproducible because effects related to particular MR machines and parameters have been removed.

MR already plays an important role in diagnostic imaging such as monitoring disease progression, assessment of the patient's response to treatment, and treatment selection. DCE-MRI has recently become a promising modality for RA diagnosis, with the main thrust being the detection of the disease at an early stage, when disease-modifying drugs can be used. Early diagnosis of rheumatoid arthritis and early, aggressive treatment can help prevent joint damage and deformity, which ultimately means for a patient the difference between a relatively normal life style and a disability. Data acquired in the presence of the contrast agent, especially by low field MRI scanners, is often corrupted by noise due to patient motion or hardware instability. The poor quality of the data slows down its reading, moreover, might influence diagnostic decisions.

Existing DCE-MRI data analysis approaches, as outlined in Chapter 2, include no pre-

processing techniques that can contribute to data fidelity and may produce highly subjective results. For example, the nave manual viewing method, where the interpretation of the data is performed by an experienced radiologist who analyses the MR intensity images, obtained as a result of a baseline image subtraction from the post-contrast data [72, 250], is time-consuming and highly subjective. The inter-observer errors associated with this method can reach 20% [50].

Results produced by semi-automated methods such as the region-of-interest approach [50] depend highly on the position and size of ROI. Inaccurate ROI placement may result in a 20-30% error [162].

Recently, voxel-by-voxel analysis of DCE-MRI data gained attention from both researchers and medical doctors. Pharmacokinetic methods [31, 134, 245] and heuristic approaches [195], which consider the statistics related to the intensity change in DCE-MRI datasets, have been demonstrated to provide sufficient information for discerning different types of tissue. These methods focus on evaluating the temporal component of DCE-MRI data as given by intensity vs. time course changes of the individual voxels, enabling physicians to analyse tissue condition based on its response to the contrast agent.

Currently, in clinical practise it is impossible to assess the accuracy with which pharmacokinetic variables reflect the true underlying changes in concentration of the contrast agent [119]. The accuracy of the estimates will depend on the pharmacokinetic model used and the signal to noise ratio in any individual case. This is a particular problem with applications where noise is the dominant, or only, cause of variation of contrast agent concentration [119].

Furthermore, both pharmacokinetic and heuristic approaches assume that the intensity change at each voxel can be attributed to the contrast leakage. However, patient movement can introduce artefactual enhancement with implications on the extracted measurements. Nevertheless, voxel-by-voxel analysis based methods represent a more reliable alternative to the contemporary naïve data reading, and might become widely accepted in the future.

The aim of this thesis was to develop automatic techniques for the analysis of dynamic contrast-enhanced magnetic resonance imaging data involved in rheumatoid arthritis studies of the hand and wrist joints. The framework of the algorithms presented here permits the enhancement of data quality as well as its objective analysis. The essential motivation behind this work has been its acceptance in everyday clinical environments.

Unlike previous work, the approach for DCE-MRI data analysis presented here employs efficient segmentation and registration algorithms that compensate for patient movement and contribute to data fidelity, and a modelling technique that permits the objective and robust computation of heuristics describing the shape of the signal intensity curves.

In this work we did not seek to duplicate or improve the pharmacokinetic methods, but instead presented an efficient approach that allows the enhancement or substitution of widely used methods such as naïve viewing and ROI based method. A comparison of the results produced by our approach and the established pharmacokinetic methods would be beneficial, and future work should venture in that direction.

Our approach to data analysis permits the automated assessment of the disease progression and patient response to the treatment. Such analysis is largely user independent and time efficient. The methods are designed to deliver objective and reproducible results. Each algorithm introduced in this work supports radiologists at certain steps in the DCE-MRI data analysis.

Registration compensates for the problems associated with subject motion during the imaging and, therefore, validates the assumption that each voxel within DCE-MRI slices represents a particular tissue type. In the presence of contrast/brightness changes, the artefacts associated with the motion have been successfully eliminated with the 3D registration algorithm enhanced with the W-transformation.

Registration algorithm has been applied to datasets acquired with low- and high-field scanners and demonstrated promising results.

Segmentation algorithms were designed to locate the tissues of interest within the joints' envelope and to exclude the bone interiors and blood vessels from analysis of DCE-MRI data of the MCPJs. This permits more efficient data processing and objective evaluation and interpretation of the parametric maps.

Segmentation of joints envelopes was performed on the datasets acquired with low- and high-field scanners. Bone interiors and blood vessels were segmented in 2D images acquired by the high-field scanner; these algorithms have not been tried on the data acquired with the low-field scanner.

Quantitative analysis has the capability to objectively evaluate a patient's condition and to track the disease progression. Furthermore, we proposed a robust scheme for tissue behaviour classification, which is based on the tissues' temporal pattern of contrast agent uptake. Such analysis might be useful for guiding data acquisition and the contrast agent dose estimation.

This approach has been demonstrated on datasets acquired by low- and high-field scanners.

The outcome of each algorithm was carefully evaluated by quantitative techniques, developed as part of this thesis, as well as detailed discussion of its visual presentation with

experienced observers and radiologists[1]. This evaluation has shown to be of particular importance, because it is often difficult or even impossible to acquire definite reference labels which perfectly reflect the biological truth.

According to clinical experts the automated analysis allows the depiction of different disease activity in separate compartments of the joints; it delivers more differentiated and comprehensive information to the reader regarding the areas with the most active perfusion, and might have a positive impact on RA studies' timelines, cost, and success.

9.2 Discussion and algorithm improvements

The work described in this thesis could be extended in a number of ways. There are obvious enhancements that could improve the computational efficiency of the algorithms, such as re-implementing the methods using more efficient programming languages and optimisation techniques. This section discusses assumptions and limitations of the presented algorithms, and possible improvements to them.

Image registration As discussed in Chapter 4 the main component of the motion in DCE-MRI data acquired from RA patients is physical tremor associated with the instability of a patient's hand. The anatomy of the tissues does not change significantly in the images acquired within a short time period. Our experiments with the parameter λ, responsible for the regularisation of the geometric / intensity and smoothness terms in the registration algorithm presented in Chapter 4, indicate that the influence of the smoothness term is not significant.

The experiments presented here were only performed with the data acquired from the same patient, therefore the registration could have been accomplished with a rigid approach accompanied by a method for the intensity change equalisation in pre- and post-contrast images.

The registration algorithm used as the basis for the approach discussed in Chapter 4 has been shown to perform on inter-patient data [184]. Therefore, one of the possible extensions of the registration approach is its testing and possible adjustment to for performance on datasets acquired in the follow-up examinations, where the anatomy of the soft tissue such as synovium changes dramatically.

[1]Results were viewed by medical experts from the Parker Institute, Copenhagen, Denmark and Clinica Reumatologica, Genoa, Italy, and experienced observes from the Academic Unit of Medical Physics, University of Leeds, Leeds General Infirmary, Leeds, UK.

Image segmentation The algorithms for background, bone, and blood vessels segmentation were specifically designed for this application. Each of the algorithms was developed independently. Currently, each image within a temporal slice is segmented in 2D space, then a 3D structure of the anatomy is formed using the interpolation method [137].

The current trend in medical image segmentation is towards the generalisation of segmentation approaches. Potentially, the MCPJs tissue segmentation task can be reformulated in terms of the texture modelling [40] or level sets [172].

These methods do not depend on the geometry or position of the tissue of interest, and allow simultaneous segmentation of the tissues within the joints, which should shorten the processing time and permit detecting the edge of the anatomy more precisely.

Alternatively, given a larger number of samples, the problem of bone and blood vessel segmentation could have been solved using active shape models [53]. Recent publications demonstrate the applicability of the ASM to recover various topologies [142] and discuss how initialisation and optimisation problems can be overcome [54].

Evaluation In this thesis we often used both supervised metrics: the new one and mutual-overlap based. We have compared behaviour of the metrics and found that the proposed approach generates more reliable results. However, currently, mutual overlap metric is widely accepted and therefore, majority of the authors report their results using this metric. Thus, in order to algorithms developed against the ones published in the literature, mutual-overlap based approach was used.

It was demonstrated that unsupervised approaches that incorporate a number of stand alone measures via SVM [45] and Bayesian classifiers [275] have shown very promising performance. However, the metrics [45] and [275] have never been tested on MR data; on the other hand, the family of unsupervised metrics proposed here performed well on such datasets. Therefore, unification of these approaches could be the next step in unsupervised evaluation.

Quantitative analysis Here we have presented results obtained on a limited pilot study of images. The data have been acquired with both high and low field MRI scanners, using different sequences, and acquisition parameters. This demonstrates the adaptability of the approach with this domain.

The nature of inflammatory diseases such as cancer and brain tumours leads us to

believe that the algorithms will be useful in these applications. The applicability of the method to DCE-MRI datasets acquired from other organs / body parts and by different scanners, of course, needs to be examined.

In order to describe the tissue behaviour, we have employed linear approximations. This approach may be further enhanced by incorporating more sophisticated models. However, complexity of the model might incur a longer computational time.

Currently, we perform the data analysis using the segmentation and registration steps as a pre-process to the contrast modelling. This is a time-consuming approach, as each $\hat{I}$ curve needs to be approximated by four models. Incorporating heuristic modelling into the classifier of the registration method might improve the efficiency of the approach.

Visualisation Parametric maps of the heuristics *ME* and *IRE* are colour coded in such a way that the highest values of the heuristics and consequently the brightest colours correspond to the enhancement of the blood vessels. Therefore, the magnitude of the synovial tissue enchantment is coloured relative to the blood vessels' enhancement.

This allows the evaluation of intra-patient inflammation; however, in the absence of the vessels or when comparing datasets acquired from different patients, this assumption needs to be replaced.

With the appropriate clinical expertise, it would be possible to generate intervals in the heuristics' values and assign a colour to each interval, so that it reflects the tissue condition. Such colour coding would then allow objective inter-patient and inter-sequence comparison.

Currently, parametric maps illustrate the magnitude of inflammation in each temporal slice. Visualisation of the synovial tissue in 3D space would permit an even better visual assessment of a patient's condition and allows for a more accurate diagnosis.

9.3 Subsequent steps in the analysis of DCE-MRI data

The algorithms presented in this thesis support radiologists throughout the analysis of DCE-MRI data. The registration and segmentation contribute to the data fidelity; the quantitative analysis and visualisation of the heuristics give a visual feedback that enables

radiologists to localise inflamed synovial tissue and to analyse the magnitude and spread of the disease. However, the final diagnosis still requires the radiologist's expertise.

The next step in the assessment of RA by means of DCE-MRI is the automatic diagnosis of a patient's condition. For this purpose it would be essential to derive a new scoring procedure that would be based on the objective heuristic parameters rather than the subjective opinion of an operator [174].

During clinical diagnosis, a tissue is commonly investigated from a variety of viewpoints [247]. Diagnostic techniques such as DCE-MRI, X-ray, and CT provide a broad range of complementary information exposing tissue features. As an extension of this work, it would be interesting to understand how methods proposed in this thesis can be extended to processing the input derived from other scanning modalities. It will be challenging to align, analyse, and correlate the information provided by data of different dimensionalities and resolutions.

To develop a new scoring system, it would be essential to compare information provided by the parametric maps and gold standard techniques such as ultrasound and static MRI. Moreover, correlation between the current scoring methods, such as ROI based and OMERACT-RAMRIS methods, and parametric maps needs to be established. This work would not be accomplished without sufficient clinical expertise and a large number of datasets.

Another direction in which this work can be extended is towards a more comprehensive understanding of the differences and similarities in the behaviour of the synovium and bone marrow oedema, which is essentially an inflammation inside the joint. It has recently been discovered that bone marrow oedema and erosive changes provide additional information about RA activity and may be used as very reliable markers [157, 232].

Using the proposed segmentation algorithm, we outlined contours of the bone interiors in the datasets acquired by the high-field MRI scanner and processed them with the model-based approach. The results illustrated in Figure 9.1 confirm the clinical opinion that this patient has visible inflammation within the joints.

The automatic classification of inflamed synovial tissue, bone marrow oedema, and erosions, and automatic measurement of the oedema volumes will require further development and optimisation of both segmentation and quantitative analysis approaches.

It is of course crucial to proceed with any further development of this work in collaboration with medical experts, who ultimately will ratify the medical benefit of any such work.

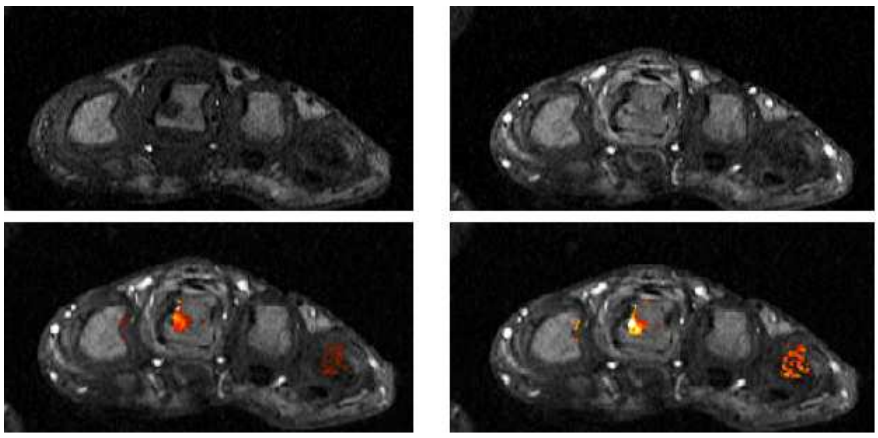


Figure 9.1: Top: Pre-and post-contrast images. Bottom: *ME* (left) and *IRE* (right) computed for pixels within the joints.

Bibliography

[1] A. M. Abduljalil, A. Kangarlu, X. Zhang, R. E. Burgess, and P.-M. L. Robitaille. Acquisition of human multislice MR image at 8T. *Journal of Computer Assisted Tomography*, 23(3):335–340, 1999.

[2] E. Adelson, C. Anderson, J. Bergen, P. Burt, and J. Ogden. Pyramid methods in image processing. *RCA Engineer*, 29(6):33–41, 1984.

[3] Ainapse Systems Ltd. Dynamic contrast-enhanced MRI (DCE-MRI) analysis, 2006. `http://www.xinapse.com/Manual/dce_mri.html` (last access on 03.05.07).

[4] E. Alasaarela, I. Suramo, O. Tervonen, S. Lahde, R. Takalo, and M. Hakala. Evaluation of humeral head erosions in rheumatoid arthritis: a comparison of ultrasonography, magnetic resonance imaging, computed tomography and plain radiography. *British Journal of Rheumatology*, 37(11):1152–1156, 1998.

[5] C. Alberola-Lopez, M. Martin-Fernandez, and J. Ruiz-Alzola. Comments on: A methodology for evaluation of boundary detection algorithms on medical images. *IEEE Transactions on Medical Imaging*, 23(5):658–660, 2004.

[6] American College of Rheumatology of Osteoarthritis Guideline. Recommendations for the medical management of osteoarthritis of the hip and knee. *Arthritis and Rheumatism*, 43(9):1905–1915, 2000.

[7] Arthritis Research Campaign. About arthritis. Online publication, 2004. `http://www.arc.org.uk/about_arth/default.htm` (last access on 03.05.07).

[8] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3D point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):698–700, 1987.

[9] E. A. Ashton. Quantitative imaging in clinical trials. Online publication, 2006. `http://www.actmagazine.com/appliedclinicaltrials/article/articleDetail.jsp?id=378344` (last access on 03.07.07).

[10] S. R. Aylward, J. Jomier, S. Weeks, and E. Bullitt. Registration and analysis of vascular images. *International Journal of Computer Vision*, 55(2-3):123–138, 2003.

[11] M. Backhaus, T. Kamradt, D. Sandrock, D. Loreck, J. Fritz, K. J. Wolf, H. Raber, B. Hamm, G. R. Burmester, and M. Bollow. A comprehensive approach comparing conventional radiography, scintigraphy, ultrasound and contrast-enhanced magnetic resonance imaging. *Arthritis and Rheumatism*, 42(6):1232–1245, 1999.

[12] J. Bailleul, S. Ruan, and D. Bloyet. Automatic atlas-based building of point distribution model for segmentation of anatomical structures from brain MRI. In *Proceedings of International Symposium on Signal Processing and Its Applications*, volume 2, pages 629–630, 2003.

[13] R. Bajcsy and S. Kovacic. Multiresolution elastic matching. *Journal of Computer Vision, Graphics, and Image Procession*, 46(1):1–21, 1989.

[14] J. Barron and M. Khurana. Determining optical flow for large motions using parametric models in a hierarchical framework. In *Proceedings of International Conference on Vision Interface*, pages 47–56, 1994.

[15] J. L. Barron, D. J. Fleet, S. S. Beauchemin, and T. A. Burkitt. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.

[16] M. Beauchemin and K. P. B. Thomson. The evaluation of segmentation results and the overlapping area matrix. *International Journal of Remote Sensing*, 18(18):3895–3899, 1997.

[17] G. Behiels, F. Maes, D. Vandermeulen, and P. Suetens. Evaluation of image features and search strategies for segmentation of bone structures in radiographs using active shape models. *Medical Image Analysis*, 6(1):47–62, 2002.

[18] M. Bertero, T. A. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889, 1988.

[19] J. C. Bezdek, L. O. Hall, and L. P. Clarke. Review of MR image segmentation techniques using pattern recognition. *Journal of Medical Physics*, 20(4):1033–1048, 1993.

[20] A. H. Bhalerao and P. E. Summers. Derivation of pressure gradients from magnetic resonance angiography using multi-resolution segmentation. In *Proceedings of IEE Conference on Image Processing and its Applications*, volume 1, pages 404–408, 1995.

[21] 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*, 48(3):614–624, 2003.

[22] R. Bitar, G. Leung, R. Perng, S. Tadros, A. R. Moody, J. Sarrazin, C. McGregor, M. Christakis, S. Symons, A. Nelson, and T. P. Roberts. MR pulse sequences: What every radiologist wants to know but is afraid to ask. *Journal of Radiological Society of North America of Radiographics*, 26(2):513–537, 2006.

[23] C. R. Boggis and S. M. Astley. Computer-assisted mammographic imaging. *Breast Cancer Research*, 2(6):392–395, 2000.

[24] M. Borsotti, P. Campadelli, and R. Schettini. Quantitative evaluation of colour image segmentation results. *Pattern Recognition Letters*, 19(8):741–747, 1998.

[25] P. Bottomley, C. Hardy, R. Argersinger, and G. Allen-Moore. A review of 1H nuclear magnetic resonance relaxation in pathology: Are T1 and T2 diagnostic? *Journal of Medical Physics*, 14(1):1–17, 1987.

[26] C. Bouma, W. Niessen, K. Zuiderveld, E. Gussenhoven, and M. Viergerver. Evaluation of segmentation algorithms for intravascular ultrasound images. In *Proceedings of the Conference on Visualization in Biomedical Computing*, pages 203–212, 1996.

[27] K. W. Bowyer. *Validation of medical image analysis techniques*, in: Handbook of Medical Imaging, Eds. J. Beutel, H. Kundel and R. van Metter, SPIE: Bellingham Washington, 2000.

[28] Y. Boykov, D. Cremers, and V. Kolmogorov. Graph-cuts versus level-sets. Tutorial in European Conference on Computer Vision, May 2006. `http://www.csd.uwo.ca/~yuri/Abstracts/eccv06-tutorial.html` (last access on 09.09.07).

[29] Y. Boykov, V. S. Lee, H. Rusinek, and R. Bansal. Segmentation of dynamic N-D data sets via graph cuts using Markov models. In *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 1, pages 1058–1066, 2001.

[30] B. H. Brinkmann, A. Manduca, and R. A. Robb. Optimized homomorphic unsharp masking for MR gray scale inhomogeneity correction. *IEEE Transactions on Medical Imaging*, 17(2):161–171, 1998.

[31] G. Brix, W. Semmler, R. Port, L. R. Schad, G. Layer, and W. J. Lorenz. Pharmacokinetic parameters in CNS Gd-DTPA enhanced MR imaging. *Journal of Computed Assisted Tomography*, 15(4):621–628, 1991.

[32] G. Brown. A home made collection of web base resources on MRI issues. Royal Adelaide Hospital, The Adelaide MRI Website, 2003. Online Report, `http://www.users.on.net/~vision` (last access on 09.09.07).

[33] J. Brown, D. Buckley, A. Coulthard, A. K. Dixon, J. M. Dixon, D. F. Easton, R. A. Eeles, D. G. Evans, F. G. Gilbert, M. Graves, C. Hayes, J. P. Jenkins, A. P. Jones, S. F. Keevil, M. O. L. G. P. Liney, S. M. Moss, A. R. Padhani, G. J. Parker, L. J. Pointon, B. A. Ponder, T. W. Redpath, J. P. Sloane, L. W. Turnbull, L. G. Walker, and R. M. Warren. Magnetic resonance imaging screening in women at genetic risk of breast cancer: imaging and analysis protocol for the UK multicentre study. *Journal of Magnetic Resonance Imaging*, 18(7):765–776, 2000.

[34] L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, 1993.

[35] A. Bulpitt, E. Berry, R. Boyle, J. Scott, and D. A. Kessel. A deformable model, incorporating expected structure information, for automatic 3D segmentation of complex anatomical structures. In *Proceedings of International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS)*, volume 1, pages 572–577, 2000.

[36] G. Buonaccorsi, C. Roberts, S. W. Cheung, Y. Watson, J. P. B. O'Connor, K. E. Davies, A. Jackson, G. Jayson, and G. J. M. Parker. Comparison of the performance of tracer kinetic model-driven registration for dynamic contrast enhanced MRI (DCE-MRI) using different models of contrast enhancement. *Journal of Academic Radiology*, 13(9):1112–1123, 2006.

[37] P. J. Burt and E. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532–540, 1983.

[38] P. Campadelli and E. Casirahgi. Lung field segmentation in digital posterior-anterior chest radiographs. In *Proceedings of International Conference on Advances in Pattern Recognition*, volume 1, pages 736–745, 2005.

[39] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[40] E. Cesmeli and D. Wang. Texture segmentation using Gaussian-Markov random fields and neural oscillator networks. In *Proceedings of IEEE International Conference on Neural Networks*, volume 3, pages 1529–1534, 1997.

[41] S. Chabrier, B. Emile, H. Laurent, C. Rosenberger, and P. Marché. Unsupervised evaluation of image segmentation application to multi-spectral images. In *Proceedings of International Conference on Advances in Pattern Recognition*, volume 1, pages 576–579, 2004.

[42] S. Chabrier, B. Emile, C. Rosenberger, and H. Laurent. Unsupervised performance evaluation of image segmentation. *EURASIP Journal on Applied Signal Processing*, 2006:1–12, 2006.

[43] S. Chabrier, H. Laurent, B. Emile, C. Rosenberger, and P. Marché. A comparative study of supervised evaluation criteria for image segmentation. In *Proceedings of European Signal Processing Conference*, pages 1143–1146, 2004.

[44] S. Chabrier, C. Rosenberger, H. Laurent, B. Emile, and P. Marché. Evaluating the segmentation results of a gray-level image. In *Proceedings of European Signal Processing Conference*, pages 953–956, 2004.

[45] S. Chabrier, C. Rosenberger, H. Laurent, and A. Rakotomamonjy. Segmentation evaluation using a support vector machine. In *Proceedings of International Conference on Advances in Pattern Recognition*, volume 2, pages 889–896, 2005.

[46] V. Chalana and Y. Kim. A methodology for evaluation of boundary detection algorithms on medical images. *IEEE Transactions on Medical Imaging*, 16(5):642–652, 1997.

[47] H. M. Chen and P. K. Varshney. Mutual information-based CT-MR brain image registration using generalized partial volume joint histogram estimation. *IEEE Transactions on Medical Imaging*, 22(9):1111–1119, 2003.

[48] R. Chrastek, M. Wolf, K. Donath, H. Niemann, D. Paulus, T. Hothorn, B. Lausen, R. Lammer, C. Y. Mardin, and G. Michelson. Automated segmentation of the optic nerve head for diagnosis of glaucoma. *Medical Image Analysis*, 9(4):297–314, 2005.

[49] A. C. S. Chung, J. A. Noble, and P. Summers. Vascular segmentation of phase contrast magnetic resonance angiograms based on statistical mixture modeling and local phase coherence. *IEEE Transactions on Medical Imaging*, 23(12):1490–1507, 2004.

[50] M. A. Cimmino, S. Innocenti, F. Livrone, F. Magnaguagno, E. Silvesti, and G. Garlaschi. Dynamic gadolinium-enhanced MRI of the wrist in patients with rheumatoid arthritis. *Arthritis and Rheumatism*, 48(5):674–680, 2003.

[51] L. P. Clarke, R. P. Velthuizen, M. A. Camacho, J. J. Heine, M. Vaidyanathan, L. O. Hall, R. W. Thatcher, and M. L. Silbiger. MRI segmentation: methods and applications. *Journal of Magnetic Resonance Imaging*, 13(3):343–368, 1995.

[52] W. Cochran. Some methods of strengthening the common χ^2 tests. *Biometrics*, 10(4):417–451, 1954.

[53] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models – their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[54] T. F. Cootes and C. J. Taylor. An algorithm for tuning an active appearance model to new data. In *Proceedings of British Machine Vision Conference*, volume 3, pages 919–928, 2006.

[55] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Training models of shape from sets of examples. In *Proceedings of British Machine Vision Conference*, volume 1, pages 9–18, 1992.

[56] J. J. Corso, E. Sharon, and A. L. Yuille. Multilevel segmentation and integrated Bayesian model classification with an application to brain tumor segmentation. In *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 2, pages 790–798, 2006.

[57] W. R. Crum, T. Hartkens, and D. L. G. Hill. Non-rigid image registration: theory and practice. *British Journal of Radiology*, 77(2):140–153, 2004.

[58] W. R. Crum, C. Tanner, and D. J. Hawkes. Anisotropic multi-scale fluid registration: evaluation in magnetic resonance breast imaging. *Electronic Journal on Physics in Medicine and Biology*, 50(21):5153–5174, 2005.

[59] Current Medicine Group LLC. The online encyclopedia of medical images. `http://www.images.md/users/explore_results.asp` (last access on 19.07.07).

[60] K. M. H. S. T. N. D. Bystrov, V. Pekar. Motion compensation and plane tracking for kinematic MR-imaging. In *Proceedings of First International Workshop on Computer Vision for Biomedical Image Applications: Current Techniques and Future Trends*, pages 551–560, 2005.

[61] E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens. A unified framework for atlas based brain image segmentation and registration. In *Proceedings of International Workshop on Biomedical Image Registration*, pages 136–143, 2006.

[62] J. A. D'Arcy, D. J. Collins, I. J. Rowland, A. R. Padhani, and M. O. Leach. Applications of sliding window reconstruction with Cartesian sampling for dynamic contrast enhanced MRI. *Journal of Nuclear Magnetic Resonance in Biomedicine*, 15(2):174–183, 2002.

[63] E. Debreuve, M. Barlaud, G. Aubert, I. Laurette, and J. Darcourt. Space-time segmentation using level set active contours applied to myocardial gated SPECT. *IEEE Transactions on Medical Imaging*, 20(7):643–659, 2001.

[64] H. M. Dee and S. A. Velastin. How close are we to solving the problem of automated visual surveillance? A review of real-world surveillance, scientific progress and evaluative mechanisms. *Machine Vision and Applications: Special Issue on Video Surveillance Research in Industry and Academic*, 2007. Published online, `http://www.springerlink.com/content/0828801018545294/` (last access on 10.09.07).

[65] J. E. Dennis. Algorithms for nonlinear problems which use discrete approximations to derivatives. In *Proceedings of the 1971 26th Annual Conference*, pages 446–456, 1971.

[66] L. R. Dice. Measures of the amount of ecologic association between species. *Journal of Ecology*, 26(3):297–302, 1945.

[67] R. M. Dickau. Compilation of iterative and list operations. *Mathematica Journal*, 7(1):14–15, 1997.

[68] M. Droske, M. Meyer, M. Rumpf, and C. Schaller. An adaptive level set method for interactive segmentation of intracranial tumors. *Journal of Neurosurgical Research*, 27(4):363–370, 2005.

[69] J. C. Duncan and N. Ayache. Medical image analysis: Progress over two decades and the challenges ahead. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):85–106, 2000.

[70] P. A. Elsen, E.-J. D. Pol, T. S. Sumanaweera, P. F. Hemler, S. Napel, and J. R. Adler. Grey value correlation techniques used for automatic matching of CT and MR brain and spine images. In *Proceedings of SPIE Conference on Visualization in Biomedical Computing*, pages 227–237, 1994.

[71] P. A. Elsen, E. J. D. Pol, and M. A. Viergever. Medical image matching: a review with classification. *IEEE Engineering in Medicine and Biology Magazine*, 12(3):26–39, 1993.

[72] R. Erlemann, M. F. Reiser, P. E. Peters, P. Vasallo, B. Nommensen, C. R. Kusnierz-Glaz, J. Ritter, and A. Roessner. Musculoskeletal neoplasms: static and dynamic Gd-DTPA-enhanced MR imaging. *Journal of Radiology*, 171(3):767–773, 1989.

[73] H. Farid and E. P. Simoncelli. Optimally rotation-equivariant directional derivative kernels. In *Proceedings of International Conference on Computer Analysis of Images and Patterns*, pages 207–214, 1997.

[74] R. Fergus, P. Perona, and A. Zisserman. A sparse object category model for efficient learning and exhaustive recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 380–387, 2005.

[75] J. M. Fitzpatrick, D. L. G. Hill, and C. R. Maurer. *Image registration*, in: Handbook of Medical Imaging: Medical Image Processing and Analysis, Eds. J. M. Fitzpatrick and M. Sonka, SPIE: Bellingham, WA, 2000.

[76] J. Flusser. An adaptive method for image registration. *Pattern Recognition Letters*, 25(1):45–54, 1992.

[77] D. Foley-Nolan, J. P. Stack, M. Ryan, U. Redmond, C. Barry, J. Ennis, and R. J. Coughlan. MRI in the assessment of rheumatoid arthritis–a comparison with plain film radiographs. *British Journal of Rheumatology*, 30(2):101–106, 1991.

[78] L. A. Forbes and B. A. Draper. Inconsistencies in edge detector evaluation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2398–2404, 2000.

[79] A. F. Frangi, W. J. Niessen, and M. A. Viergever. Three-dimensional modeling for functional analysis of cardiac images: A review. *IEEE Transactions on Medical Imaging*, 20(1):2–25, 2001.

[80] K. Friston. SPM: Statistical Parametric Mapping, 1994. `http://www.fil.ion.ucl.ac.uk/spm/` (last access on 03.05.07).

[81] M. S. Froh, D. C. Barber, K. K. Brock, D. B. Plewes, and A. L. Martel. Piecewise-quadrilateral registration by optical flow – applications in contrast-enhanced MR imaging of the breast. In *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 2, pages 686–693, 2006.

[82] K. S. Fu and J. K. Mui. A survey on image segmentation. *Pattern Recognition Letters*, 13(1):3–16, 1981.

[83] K. Gaffney, J. Cookson, D. Blake, A. Coumbe, and S. Blades. Quantification of rheumatoid synovitis by magnetic resonance imaging. *Arthritis and Rheumatism*, 38(11):1610–1617, 1995.

[84] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[85] P. Gibbs, D. L. Buckley, S. J. Blackband, and A. Horsman. Tumor volume determination from MR images by morphological segmentation. *Electronic Journal of Physics in Medicine and Biology*, 41(11):2437–2446, 1996.

[86] G. H. Golub and C. F. van Loan. An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis*, 17(6):883–893, 1980.

[87] R. C. Gonzalez and R. E. Woods. *Image Segmentation*, in: Digital Image Processing, 2nd edition, Ed. McDonnell, Prentice Hall: New York, USA, 2002.

[88] R. C. Gonzalez and R. E. Woods. *Digital image processing*. Prentice Hall, New York, USA, 2nd edition, 2004.

[89] A. Goshtasby. Registration of images with geometric distortions. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1):60–64, 1988.

[90] T. A. Gould. How MRI works. Online publication, 2001. `http://health.howstuffworks.commri12.htm` (last access 24.08.07).

[91] W. Grassi, E. Filippucci, A. Farina, F. Salaffi, and C. Cervini. Ultrasonography in the evaluation of bone erosions. *Annals of the Rheumatic Diseases*, 60(2):98–103, 2001.

[92] H. Gutte, D. Jakobsson, F. Olofsson, M. Ohlsson, S. Valind, A. Loft, and L. E. A. Kjaer. Automated interpretation of PET/CT images in patients with lung cancer. *Nuclear Medicine Communications*, 28(2):79–84, 2007.

[93] J. V. Hajnal, D. L. G. Hill, and D. J. Hawkes, editors. *Medical Image Registration*. CRC Press, 2001.

[94] J. V. Hajnal, N. Saeed, E. J. Soar, A. Oatridge, I. R. Young, and G. M. Bydder. A registration and interpolation procedure for subvoxel matching of serially acquired MR images. *Journal of Computer Assisted Tomography*, 19(2):289–296, 1995.

[95] L. Hallpike and D. J. Hawkes. Medical image registration: an overview. *Journal of British Institute of Radiology on Imaging*, 14(6):455–463, 2002.

[96] R. H. Haralick and L. G. Shapiro. Image segmentation techniques. *Journal of Computer Vision, Graphics, and Image Processing*, 29(1):100–132, 1985.

[97] M. Hart and L. Holley. A method of automated coronary artery tracking in unsubtracted angiograms. In *Proceedings of IEEE Conference on Computers in Cardiology*, volume 1, pages 93–96, 1997.

[98] T. Hartkens, D. Rueckert, J. A. Schnabel, D. J. Hawkes, and D. L. G. Hill. VTK CISG registration toolkit: An open source software package for affine and non-rigid registration of single- and multimodal 3D images. In *Proceedings of Bildverarbeitung in der Medizin*, pages 409–412, 2002.

[99] R. Hartly and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 3rd edition, 2000.

[100] M. S. Hassouna and A. A. Farag. Stochastic segmentation of blood vessels from time-of-flight. In *Proceedings of IEEE International Conference on Image Processing*, volume 1, pages 29–32, 2005.

[101] M. S. Hassouna, A. A. Farag, S. Hushek, and T. Moriarty. Cerebrovascular segmentation from TOF using stochastic models. *Medical Image Analysis*, 10(1):2–18, 2006.

[102] M. Hau, H. Schultz, H. P. Tony, M. Keberle, R. Jahns, R. Haerten, and M. Jenett. Evaluation of pannus and vascularization of the metacarpophalangeal and proximal interphalangeal joints in rheumatoid arthritis by high-resolution ultrasound (multidimensional linear array). *Arthritis and Rheumatism*, 42(11):2303–2308, 2001.

[103] D. J. Hawkes. *Registration Methodology*, in: Medical Image Registration, Ed. M. R. Neuman, CRC Press: London, UK , 2001.

[104] M. Heath, S. Sarkar, T. Sanocki, and K. W. Bowyer. A robust visual method for assessing the relative performance of edge detection algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(12):1338–1359, 1997.

[105] M. Heath, S. Sarkar, T. Sanocki, and K. W. Bowyer. Comparison of edge detectors: a methodology and initial study. *Computer Vision and Image Understanding*, 69(1):38–54, 1998.

[106] D. M. Heijde. Plain X-rays in rheumatoid arthritis: overview of scoring methods, their reliability and applicability. *Journal of Baillieres Clinical Rheumatology*, 10(3):435–453, 1996.

[107] S. H. Heywang, A. Wolf, E. Pruss, T. Hilbertz, W. Eiermann, and W. Permanetter. MR imaging of the breast with Gd-DTPA: use and limitations. *Journal of European Radiology*, 171(1):95–103, 1989.

[108] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Electronic Journal on Physics in Medicine and Biology*, 46(3):R1–R45, 2001.

[109] D. L. G. Hill, D. J. Hawkes, J. E. Crossman, M. J. Gleeson, T. C. Cox, E. E. Bracey, A. J. Strong, and P. Graves. Registration of MR and CT images for skull base surgery using point-like anatomical features. *British Journal of Radiology*, 64(767):1030–1035, 1991.

[110] J. H. Hipwell, G. P. Penney, R. A. McLaughlin, K. Rhode, P. Summers, T. C. Cox, J. V. Byrne, J. A. Noble, and D. J. Hawkes. Intensity-based 2D-3D registration of cerebral angiograms. *IEEE Transactions on Medical Imaging*, 22(11):1417–1426, 2003.

[111] R. J. Hodgson, S. Connoly, T. Barnes, B. Eyes, R. S. D. Cambell, and R. Moots. Pharmacokinetic modelling of DCE-MRI of the hand and wrist in rheumatoid arthritis and the response to anti-TNF alpha therapy. *Journal of Magnetic Resonance in Medicine*, 58(3):482–489, 2007.

[112] R. V. Hogg and A. T. Craig. *Introduction to Mathematical Statistics*. Macmillan, New York, US, 1995.

[113] M. Holden, D. L. G. Hill, E. R. E. Denton, J. M. Jarosz, T. C. S. Cox, and D. J. Hawkes. Voxel similarity measures for 3D serial MR brain image registration. In *Proceedings of 16th International Conference on Information Processing in Medical Imaging*, pages 472–477, 1999.

[114] M. Holden, D. L. G. Hill, E. R. E. Denton, J. M. Jarosz, T. C. S. Cox, T. Rohlfing, J. Goodey, and D. J. Hawkes. Voxel similarity measures for 3D serial MR brain image registration. *IEEE Transactions on Medical Imaging*, 19(2):94–102, 2000.

[115] A. Hoover, G. Jean-Baptiste, X. Jiang, P. J. Flynn, H. Bunke, D. B. Goldof, K. Bowyer, D. W. Eggert, A. Fitzgibbon, and R. B. Fisher. An experimental comparison of range segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7):673–689, 1996.

[116] B. Horn. *Robot Vision*. MIT Press, Cambridge, MA, 1986.

[117] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Journal of Artificial Intelligence*, 17(1-3):185–203, 1981.

[118] I. Isgum, A. Rutten, M. Prokop, and B. van Ginneken. Detection of coronary calcifications from computed tomography scans for automated risk assessment of coronary artery disease. *Medical Physics*, 34(4):1450–1461, 2007.

[119] A. Jackson. Analysis of dynamic contrast enhanced MRI. *British Journal of Radiology*, 77(2):154–166, 2004.

[120] M. Jenkinson and S. Smith. A global optimisation method for robust affine registration of brain images. *Journal of Medical Image Analysis*, 5(2):143–156, 2001.

[121] W. A. Kaiser and E. Zeitler. MR imaging of the breast: Fast imaging sequences with and without Gd-DTPA. Preliminary observations. *Journal of European Radiology*, 170(3):681–686, 1989.

[122] M. Kass, A. Witkin, and D. Terzopoulos. Snakes, active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1987.

[123] R. G. Keys. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, Signal Processing*, 29(6):1153–1160, 1981.

[124] D. Kidner, M. Dorey, and D. Smith. What's the point? Interpolation and extrapolation with a regular grid digital elevation model. In *Proceedings of International Conference on GeoComputation*, 1999. Online publication `http://www.geovista.psu.edu/sites/geocomp99/Gc99/082/gc_082.htm` (last access 10.04.07).

[125] C. Kirbas and F. Quek. A review of vessel extraction techniques and algorithms. *ACM Computing Surveys*, 36(2):81–121, 2004.

[126] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[127] A. K. Klein, F. Lee, and A. Amini. Automatic quantitative coronary angiography with deformable spline models. *IEEE Transactions on Medical Imaging*, 16(5):468–482, 1997.

[128] H. Konig, J. Sieper, and K. J. Wolf. Rheumatoid arthritis: evaluation of hypervascular and fibrous pannus with dynamic MR imaging enhanced with Gd-DTPA. *Journal of Radiology*, 16(7):473–477, 1990.

[129] W. J. Koopman, D. W. Boulware, and G. Heudebert. *Clinical Primer of Rheumatology*. Lippincott Williams and Wilkens, 2003.

[130] C. K. Kuhl, P. Mielcareck, S. Klaschik, C. Leutner, E. Wardelmann, J. Gieseke, and H. H. Schild. Dynamic breast MR imaging: Are signal intensity time course data useful for differential diagnosis of enhancing lesions? *Journal of European Radiology*, 211(1):101–110, 1999.

[131] R. Kumar, K. Hanna, J. C. Asmuth, J. R. Bergen, C. Hulka, D. B. Kopans, R. Weisskoff, and R. Moore. Detecting lesions in magnetic resonance breast scans. In *Proceedings of SPIE 24th AIPR Workshop on Tools and Techniques for Modeling and Simulation*, pages 181–190, 1996.

[132] H. W. M. Laarhoven, M. Rijpkema, C. J. A. Punt, T. J. Ruers, J. C. M. Hendriks, J. O. Barentsz, and A. Heerschap. Method for quantitation of dynamic MRI contrast agent uptake in colorectal liver metastases. *Journal of Magnetic Resonance Imaging*, 18(3):315–320, 2006.

[133] M. Laniado, H. J. Weinmann, W. Schörner, R. Felix, and U. Speck. First use of Gd-DTPA/dimeglumine in man. *Journal on Physiological Chemistry and Physics and Medical NMR*, 16(2):157–165, 1984.

[134] H. B. Larsson, M. Stubgaard, J. L. Frederiksen, M. Jensen, O. Henriksen, and O. B. Paulson. Quantitation of blood-brain barrier defect by magnetic resonance imaging and Gadolinium-DTPA in patients with multiple sclerosis and brain tumors. *Journal of Magnetic Resonance in Medicine*, 16(1):117–131, 1990.

[135] W. C. Lavely, C. Scarfone, H. Cevikalp, R. Li, D. W. Byrne, A. J. Cmelak, B. Dawant, R. R. Price, D. E. Hallahan, and J. M. Fitzpatrick. Phantom validation of coregistration of PET and CT for image-guided radiotherapy. *Journal of Medical Physics*, 31(5):1083–1092, 2004.

[136] L. Lecornu, C. Roux, and J. J. Jacq. Extraction of vessel contours in angiograms by simultaneous tracking of the two edges. In *Proceedings of IEEE Conference on Engineering in Medicine and Biology*, volume 1, pages 678–679, 1994.

[137] T.-Y. Lee and W.-H. Wang. Morphology-based three-dimensional interpolation. *IEEE Transactions on Medical Imaging*, 19(7):711–721, 2000.

[138] K. Levenberg. A method for the solution of certain problems in least squares. *Quarterly of Applied Mathematics*, 2:164–168, 1944.

[139] M. D. Levine and A. M. Nazif. Dynamic measurement of computer generated image segmentations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):155–164, 1985.

[140] K. Li, G. Wright, L. Pelc, R. Dalman, J. Brittain, H. Wegmueller, D. Lin, and C. Song. Oxygen saturation of blood in the superior mesenteric vein: In vivo verification of MR imaging measurements in a canine model. *Journal of Radiology*, 194(2):321–326, 1995.

[141] S. Z. Li. Markov random field models in computer vision. In *Proceedings of European Conference on Computer Vision*, pages 361–370, 1994.

[142] J. Liu and J. K. Udupa. Oriented active shape models for 3D segmentation in medical images. In *SPIE Proceedings on Medical Imaging 2007: Image Processing*, pages 30–40, 2007.

[143] L. M. Lorigo, O. D. Faugeras, W. E. L. Grimson, R. Keriven, R. Kikinis, A. Nabavi, and C. F. Westin. CURVES: Curve evolution for vessel segmentation. *Medical Image Analysis*, 5(3):195–206, 2001.

[144] D. G. Lowe. Organization of smooth image curves at multiple scales. *International Journal of Computer Vision*, 3(2):119–130, 1989.

[145] D. G. Lowe. Distinctive image features from scale-invariant key points. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[146] S. Lu and S. Eiho. Automatic detection of the coronary arterial contours with sub-branches from an X-ray angiogram. In *Proceedings of IEEE Conference on Computers in Cardiology*, volume 1, pages 575–578, 1993.

[147] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.

[148] J. Lui and Y. H. Yang. Multiresolution color image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(7):689–700, 1994.

[149] D. Magee, A. Bulpitt, and E. Berry. 3D automated segmentation and structural analysis of vascular trees using deformable models. In *Proceedings of IEEE Workshop on Variational and Level Set Methods in Computer Vision*, volume 1, pages 119–126, 2001.

[150] D. Magee, A. Tanner, M. Waller, D. McGonagle, and A. Jeavons. Registration of PET and MR hand volumes using Bayesian networks. In *Proceedings of ICCV Workshop on Computer Vision for Biomedical Image Applications*, volume 1, pages 437–448, 2005.

[151] Magnetic Resonance Technology Information Portal. Website, 2006. `http://www.mr-tip.com` (last access on 03.05.07).

[152] J. Maintz and M. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

[153] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of SIAM*, 11(2):431–441, 1963.

[154] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of International Conference on Computer Vision*, volume 2, pages 416–423, 2001.

[155] F. J. Massey. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

[156] C. R. Maurer and J. M. Fitzpatrick. *A review of medical image registration*, in: Interactive Image Guided Neurosurgery, American Association of Neurological Surgeons: Park Ridge, IL, 1993.

[157] M. E. Mayerhoefer, J. Kramer, M. J. Breitenseher, C. Norden, A. Vakil-Adli, S. Hofmann, R. Meizer, H. Siedentop, F. Landsiedl, and N. Aigner. Short-term outcome of painful bone marrow oedema of the knee following oral treatment with iloprost or tramadol: results of an exploratory phase II study of 41 patients. *Rheumatology*, 46(9):1460–1465, 2007.

[158] R. McGill, J. W. Tukey, and W. A. Larsen. Variations of box plots. *The American Statistician*, 32(1):12–16, 1978.

[159] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 1(2):91–108, 1996.

[160] T. McInerney and D. Terzopoulos. T-snakes: Topology adaptive snakes. *IEEE Transactions on Medical Image Analysis*, 4(2):73–91, 2000.

[161] R. A. McLaughlin, J. H. Graeme, P. Penney, K. Rhode, A. Chung, J. A. Noble, and D. J. Hawkes. Intensity-based registration versus feature-based registration for neurointerventions. *IEEE Transaction on Medical Imaging*, 24(8):1058–1066, 2005.

[162] F. McQueen. Comments on the article by Cimmino et al.: Dynamic gadolinium-enhanced MRI of the wrist in patients with rheumatoid arthritis. *Arthritis and Rheumatism*, 50(2):674–680, 2004.

[163] E. McVeigh, M. Bronskill, and M. Henkelman. Optimization of MR protocols: A statistical decision analysis approach. *Journal of Magnetic Resonance in Medicine*, 6(3):314–333, 1988.

[164] V. B. Medina, R. C. Valdes, O. Yanez-Suarez, M. Garza-Jinich, and J.-F. Lerallut. Automatic initialization for a snakes-based cardiac contour extraction. In *Proceedings of Conference of the IEEE*, volume 3, pages 1625–1628, 2000.

[165] E. Meijering and M. Unser. A note on cubic convolution interpolation. *IEEE Transactions on Image processing*, 12(4):477–479, 2003.

[166] A. Mitiche and A. R. Mansouri. On convergence of the Horn and Schunck optical flow estimation method. *IEEE Transactions on Image Processing*, 13(6):848–852, 2004.

[167] C. Molina, G. P. Prause, P. Radeva, and M. Sonka. 3-D catheter path reconstruction from biplane angiography using 3D snakes. In *SPIE Proceedings on Medical Imaging*, pages 504–512, 1998.

[168] S. Negahdaripour and C. H. Yu. A generalized brightness change model for computing optical flow. In *Proceedings of International Conference of Computer Vision*, volume 2, pages 2–11, 1993.

[169] O. Nestares and D. J. Heeger. Robust multiresolution alignment of MRI brain volumes. *Journal of Magnetic Resonance in Medicine*, 43(5):705–715, 2000.

[170] W. S. Ng and C. K. Lee. Comment on: Using the uniformity measure for performance measure in image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):933–934, 1996.

[171] B. H. Okker, C. H. Yan, J. Zhang, S. H. Ong, and S. H. Teoh. Accurate and fully automatic 3D registration of spinal images using normalized mutual information. In *Proceedings of IEEE International Workshop on Biomedical Circuits and Systems*, volume 1, pages 1–8, 2004.

[172] S. Osher and J. A. Sethian. Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.

[173] M. Østergaard, A. Duer, U. Møller, and B. Ejbjerg. Magnetic resonance imaging of peripheral joints in rheumatic diseases. *Best Practice and Research. Clinical Rheumatology*, 18(6):861–879, 2004.

[174] M. Østergaard, C. Peterfy, P. Conaghan, F. McQueen, P. Bird, B. Ejbjerg, R. Shnier, P. O'Connor, M. Klarlund, P. Emery, H. Genant, M. Lassere, and J. Edmonds.

OMERACT rheumatoid arthritis magnetic resonance imaging studies. Core set of MRI acquisitions, joint pathology definitions, and the OMERACT RA-MRI scoring system. *Journal of Rheumatology*, 30(6):1385–1386, 2003.

[175] M. Østergaard, M. Stoltenberg, P. Gideon, K. Sorensen, O. Henriksen, and I. Lorenzen. Changes in synovial membrane and joint effusion volumes after intraarticular methylprednisolone. Quantitative assessment of inflammatory and destructive changes in arthritis by MRI. *Journal of Rheumatology*, 23(7):1151–1161, 1996.

[176] N. Otsu. A threshold selection method from gray-level histogram. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[177] A. R. Padhani. Dynamic contrast-enhanced MRI in clinical oncology: Current status and future directions. *Journal of Magnetic Resonance Imaging*, 16(4):407–422, 2002.

[178] N. R. Pal and S. K. Pal. Entropic thresholding. *An International Journal on Signal Processing*, 16(7):97–108, 1989.

[179] N. R. Pal and S. K. Pal. A review on image segmentation techniques. *Pattern Recognition Letters*, 26(9):1277–1294, 1993.

[180] J. R. Panting, P. D. Gatehouse, Z. G. Yang, M. Jerosch-Herold, N. Wilke, D. N. Firmin, and D. J. Pennell. Echo-planar magnetic resonance myocardial perfusion imaging: Parametric map analysis and comparison with thallium SPECT. *Journal of Magnetic Resonance Imaging*, 13(4):192–200, 2001.

[181] G. J. M. Parker. *Dynamic Gd imaging to measure permeability and extracellular space*, in: Physics and Engineering in Medicine in the New Millennium, Eds. P. F. Sharpe and A. C. Perkins, Institute of Physics and Engineering in Medicine: York, UK, 2000.

[182] T. Pavlidis. *Structural pattern recognition*. Springer Verlag, New York, 2nd edition, 1977.

[183] C. A. Pelizzari, G. T. Y. Chen, D. R. Spelbring, R. P. Weichselbaum, and C. T. Chen. Accurate 3-dimensional registration of CT, PET, and/or MR images of the brain. *Journal of Computer Assisted Tomography*, 13(1):20–26, 1989.

[184] S. Periaswamy. *General-purpose medical image registration*. PhD thesis, Department of Computer Science, Dartmouth College, Hanover, NH, USA, 2003.

[185] S. Periaswamy. Registration software. Website, 2005. `http://www.cs.dartmouth.edu/~sp/research.htm` (last access on 03.05.07).

[186] S. Periaswamy and H. Farid. Medical image registration with partial data. *Medical Image Analysis*, 10(3):452–464, 2006.

[187] R. A. Peters II and R. N. Strickland. A review of image complexity metrics for automatic target recognizers. In *Proceedings of 8th Meeting of Optical Engineering*, 1993. Invited paper `http://www.vuse.vanderbilt.edu/~rap2/papers/atrstc.pdf` (last access on 16.09.07).

[188] D. L. Pham and J. L. Prince. An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity inhomogeneities. *Pattern Recognition Letters*, 20(1):57–68, 1999.

[189] L. Pham, C. Y. Xu, and J. L. Prince. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2:315–338, 2000.

[190] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317(26):314–319, 1985.

[191] R. Pohle and K. D. Toennies. Segmentation of medical images using adaptive region growing. In *Proceedings of SPIE on Medical Imaging: Image Processing*, pages 1337–1346, 2001.

[192] M. Prastawa, J. H. Gilmore, W. Lin, and G. Gerig. Automatic segmentation of MR images of the developing newborn brain. *Medical Image Analysis*, 5(9):457–466, 2005.

[193] W. K. Pratt. Correlation techniques of image registration. *IEEE Transactions on Aerospace and Electronic Systems*, 10(3):353–358, 1974.

[194] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: the art of scientific computing*. Cambridge University Press, Cambridge, UK, 2nd edition, 1992.

[195] A. Radjenovic. *Measurement of physiological variables by dynamic Gd-DTPA enhanced MRI*. PhD thesis, School of Medicine, University of Leeds, Leeds, UK, 2003.

[196] A. Radjenovic, J. P. Ridgway, and M. A. Smith. A method for pharmacokinetic modelling of dynamic contrast enhanced MRI studies of rapidly enhancing lesions acquired in a clinical setting. *Journal of Physics in Medicine and Biology*, 51(9):187–197, 2006.

[197] W. E. Reddick, J. S. Taylor, and B. D. Fletcher. Dynamic MR imaging (DEMRI) of microcirculation in bone sarcoma. *Journal of Magnetic Resonance Imaging*, 10(3):277–285, 1999.

[198] R. J. Reece, M. C. Kraan, A. Radjenovic, J. P. Ridgway, J. P. Farnell, P. J. O'Connor, W. W. Gibbon, D. J. Veale, F. C. Breedveld, and P. Emery. Dynamic gadolinium enhanced MR monitoring of inflammatory changes in rheumatoid arthritis with the new DMARD leflunomide versus methotrexate. *63rd Annual Scientific Meeting of the ACR (Arthritis and Rheumatism)*, 42(9):364–365, 1999.

[199] B. Rezaie and M. D. Srinath. Algorithms for fast image registration. *IEEE Transactions on Aerospace and Electronic Systems*, 20(6):716–728, 1984.

[200] T. W. Ridler and S. Calvard. Picture thresholding using an iterative selection method. *IEEE Transactions on Systems, Man and Cybernetics*, 8(8):630–632, 1978.

[201] S. S. Rifman. Digital rectification of ERTS multispectral imagery. In *Proceedings of Symposium on Significant Results Obtained From the Earth Resources Technology Satellite-1*, pages 1131–1142, 1973.

[202] A. Robb, D. P. Hanson, R. A. Karwoski, A. G. Larson, E. L. Workman, and M. C. Stacy. Analyze: a comprehensive, operator-interactive software package for multidimensional medical image display and analysis. *International Journal on Computerized Medical Imaging and Graphics*, 13(6):433–454, 1989.

[203] R. A. Robb. *Biomedical Imaging, Visualization, and Analysis*. Wiley-Liss, A John Wiley and Sons, Inc.:Somerset, NJ, 2000.

[204] M. G. Roberts, T. F. Cootes, and J. E. Adams. Automatic segmentation of lumbar vertebrae on digitised radiographs using linked active appearance models. In *Proceedings of Medical Image Understanding and Analysis*, volume 2, pages 120–124, 2006.

[205] A. Roche, G. Malandain, N. Ayache, and S. Prima. Towards a better comprehension of similarity measures used in medical image registration. In *Proceedings*

of International Conference on Medical Image Computing and Computer-Assisted Intervention, volume 1, pages 555–566, 1999.

[206] M. Rogers and J. Graham. Robust active shape model search. In *Proceedings of the European Conference on Computer Vision*, pages 517–530, 2002.

[207] T. Rohlfing and C. R. Maurer. Shape-based averaging for combination of multiple segmentations. In *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 2, pages 838–845, 2005.

[208] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of IEEE*, 86(11):2210–2239, 1998.

[209] G. Rote. Computing the minimum Hausdorff distance between two point sets on a line under translation. *Information Processing Letters*, 38(3):123–127, 1991.

[210] D. Rueckert, P. Burger, S. M. Forbat, R. D. Mohiaddin, and G. Z. Yang. Automatic tracking of the aorta in cardiovascular MR images using deformable models. *IEEE Transactions on Medical Imaging*, 16(5):581–590, 1997.

[211] P. K. Sahoo, S. Soltani, A. K. C. Wong, and Y. C. Chen. A survey of thresholding techniques. *Computer Vision Graphics and Image Processing*, 41(2):233–260, 1988.

[212] S. Salvador and P. Chan. Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In *Proceedings of the Annual IEEE International Conference on Tools with Artificial Intelligence*, pages 576–584, 2004.

[213] L. Sarry and J. Y. Boire. Three-dimensional tracking of coronary arteries from biplane angiographic sequences using parametrically deformable models. *IEEE Transactions on Medical Imaging*, 20(12):385–392, 2001.

[214] A. Savnik, H. Malmskov, H. B. Thomsen, T. Bretlau, L. B. Graff, H. Nielsen, B. Danneskiold-Samsøe, J. Boesen, and H. Bliddal. MRI of the arthritic small joints: comparison of extremity MRI (0.2 T) vs. high-field MRI (1.5 T). *Journal of Arthritis Research and Therapy*, 11(6):1030–1038, 2001.

[215] A. K. Scheel, K.-G. A. Hermann, S. Ohrndorf, C. Werner, C. Schirmer, J. Detert, M. Bollow, B. Hamm, G. A. Müller, G. R. Burmester, and M. Backhaus. Prospective 7 year follow-up imaging study comparing radiography, ultrasonography, and

magnetic resonance imaging in rheumatoid arthritis finger joints. *Annals of the Rheumatic Diseases*, 65(5):595–600, 2006.

[216] C. Schmid and R. Mohr. Local grey value invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–535, 1997.

[217] B. Sendov. Hausdorff distance and image processing. *Russian Mathematical Survey*, 59(2):319–328, 2004.

[218] J. A. Sethian. Evolution, implementation, and application of level set and fast marching methods for advancing fronts. *Journal of Computational Physics*, 169(2):503–555, 2001.

[219] M. Sezgin and T. Marmara. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–165, 2004.

[220] L. Shao, T. Kadir, and M. Brady. Geometric and photometric invariant distinctive regions detection. *International Journal on Information Sciences*, 177(4):1088–1122, 2007.

[221] D. G. Shen, E. H. Herskovits, and C. Davatzikos. An adaptive-focus statistical shape model for segmentation and shape modeling of 3-D brain structures. *IEEE Transactions on Medical Imaging*, 20(4):257–270, 2001.

[222] J. Shi and C. Tomasi. Good features to track. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, 1994.

[223] W. C. Shiel. Rheumatoid arthritis. *WebMD, MedicineNet Incorporation*, March 2007. Online publication `http://www.medicinenet.com/rheumatoid_arthritis/article.htm` (last access om 09.09.07).

[224] M. C. Shin, D. B. Goldgof, and K. W. Bowyer. Comparison of edge detector performance through use in an object recognition task. *Computer Vision and Image Understanding*, 84(1):160–178, 2001.

[225] A. J. Silman and J. E. Pearson. Epidemiology and genetics of rheumatoid arthritis. *Journal of Arthritis Research and Therapy*, 4(3):S265–S272, 2002.

[226] J. G. Sled, A. P. Zijdenbos, and A. C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, 17(1):87–97, 1998.

[227] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3D. *ACM Transactions on Graphics (SIGGRAPH Proceedings)*, 25(3):835–846, 2006.

[228] Z. Song, N. J. Tustison, B. B. Avants, and J. C. Gee. Integrated graph cuts for brain MRI segmentation. In *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 2, pages 831–838, 2006.

[229] M. Sonka, V. Hlavac, and R. D. Boyle. *Image Processing, Analysis, and Machine Vision*. Thomson Learning, 3rd edition, 2008.

[230] M. Sonka, W. Liang, R. M. Stefancik, and A. Stolpen. *Vascular Imaging and Analysis*, in: Handbook of Medical Imaging - Medical Image Processing and Analysis, Eds. M. Sonka and J. M. Fitzpatrick, SPIE: Washington, USA, 2000.

[231] G. Subsol, N. Roberts, M. Doran, J. P. Thirion, and G. H. Whitehouse. Automatic analysis of cerebral atrophy. *Journal of Magnetic Resonance Imaging*, 15(8):917–927, 1997.

[232] H. Sugimoto, A. Takeda, and K. Hyodoh. Early-stage rheumatoid arthritis: prospective study of the effectiveness of MR imaging for diagnosis. *Journal of Radiology*, 216(2):569–575, 2000.

[233] H. Sugimoto, A. Takeda, and S. Kano. Assessment of disease activity in rheumatoid arthritis using magnetic resonance imaging: quantification of pannus volume in the hands. *Journal of Rheumatology*, 37(8):854–861, 1998.

[234] H. Sugimoto, A. Takeda, J. Masuyama, and M. Furuse. Early-stage rheumatoid arthritis: diagnostic accuracy of MR imaging. *Journal of Radiology*, 198(1):185–192, 1996.

[235] K. Sùilleabhàin. Dynamic contrast-enhanced MRI predicts response to antiangiogenesis agents. *Online Journal OncoLog*, 49(6), 2004. `http://www2.mdanderson.org/depts/oncolog/articles/04/6-jun/6-04-2.html` (last access on 03.05.07).

[236] J. S. Suri. *Segmentation Models*, in: Handbook of Biomedical Image Analysis, Eds. J. S. Suri and D. L. Wilson and S. Laximinarayan, Springer: New York, USA, 2005.

[237] J. S. Suri, K. Liu, L. Reden, and S. Laxminarayan. A review on MR vascular image processing algorithms: acquisition and prefiltering: Part I. *IEEE Transactions on Information Technology in Biomedicine*, 6(4):324–337, 2002.

[238] J. S. Suri, K. C. Liu, S. Singh, S. N. Laxminarayan, X. L. Zeng, and L. Reden. Shape recovery algorithms using level sets in 2-D/3-D medical imagery: A state-of-the-art review. *IEEE Transactions on Information Technology in Biomedicine*, 6(1):8–28, 2002.

[239] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. Comparative study of energy minimization methods for Markov random fields. In *Proceedings of European Conference on Computer Vision*, volume 2, pages 16–29, 2006.

[240] K. Tamai, M. Yamato, T. Yamaguchi, and W. Ohno. Dynamic magnetic resonance imaging for the evaluation of synovitis in patients with rheumatoid arthritis. *Arthritis and Rheumatism*, 37(8):1151–1157, 1994.

[241] P. C. Taylor. The value of sensitive imaging modalities in rheumatoid arthritis. *Journal on Arthritis Research and Therapy*, 5(5):210–213, 2003.

[242] P. S. Tofts. Modelling tracer kinetics in dynamic Gd-DTPA MR imaging. *Journal of Magnetic Resonance Imaging*, 7(1):91–101, 1997.

[243] P. S. Tofts, G. Brix, D. L. Buckley, J. L. Evelhoch, E. Henderson, M. V. Knopp, H. B. Larsson, T. Y. Lee, N. A. Mayr, G. J. Parker, R. E. Port, J. Taylor, and R. M. Weisskoff. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. *Journal of Magnetic Resonance Imaging*, 10(3):223–232, 1999.

[244] P. S. Tofts and E. P. du Boulay. Towards quantitative measurements of relaxation times and other parameters in the brain. *Journal of Neuroradiology*, 32(5):407–415, 1990.

[245] P. S. Tofts and A. G. Kermode. Measurement of the blood-brain barrier permeability and leakage space using DEMRI. *Journal of Magnetic Resonance in Medicine*, 17(2):357–367, 1991.

[246] P. S. Tofts, B. A. Rerkowitz, and M. Schnall. Quantitative analysis of dynamic Gd-DTPA enhancement in breast tumors using a permeability model. *Journal of Magnetic Resonance in Medicine*, 33(4):564–568, 1995.

[247] T. Twellmann. *Data-Driven Analysis of Dynamic Contrast-Enhanced Magnetic Resonance Imaging Data in Breast Cancer Diagnosis*. PhD thesis, Faculty of Technology, University of Bielefeld, Bielefeld, Germany, 2005. `http://www.techfak.uni-bielefeld.de/ags/ni/publications/media/Twellmann2005.pdf` (last access on 12.09.07).

[248] T. Twellmann, A. Saalbach, O. Gerstung, M. O. Leach, and T. W. Nattkemper. Image fusion for dynamic contrast-enhanced magnetic resonance imaging. *BMC BioMedical Engineering OnLine*, 3(10), 2004. `http://www.biomedical-engineering-online.com/content/3/1/35`.

[249] K. A. Vermeer, F. M. Vos, H. G. Lemij, and A. M. Vossepoel. A model based method for retinal blood vessel detection. *Computers in Biology and Medicine*, 34(3):209–219, 2004.

[250] K. L. Verstraete, Y. D. Deene, H. Roels, A. Dierick, D. Uyttendaele, and M. Kunnen. Benign and malignant musculoskeletal lesions: dynamic contrast-enhanced MR imaging–parametric 'first-pass' images depict tissue vascularization and perfusion. *Journal of Radiology*, 192(3):835–843, 1994.

[251] R. J. Wakefield, W. W. Gibbon, P. G. Conaghan, P. O'Connor, D. McGonagle, C. Pease, M. J. Green, D. J. Veale, J. D. Isaacs, and P. Emery. The value of sonography in the detection of bone erosions in patients with rheumatoid arthritis. *Arthritis and Rheumatism*, 43(12):2762–2770, 2000.

[252] E. Warner, D. B. Plewes, K. A. Hill, P. A. Causer, J. T. Zubovits, R. A. Jong, M. R. Cutrara, G. DeBoer, M. J. Yaffe, S. J. Messner, W. S. Meschino, C. A. Piron, and S. A. Narod. Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. *Journal of the American Medical Association*, 292(11):1317–1325, 2004.

[253] A. G. Webb. *Introduction to biomedical imaging*. IEEE Press Series on Biomedical Engineering, 2003.

[254] Website of the free real-world imagery. `http://www.freeimages.co.uk` (last access on 10.09.07).

[255] R. Weide, C. J. G. Bakker, and M. A. Viergever. Localization of intravascular devices with paramagnetic markers in MR images. *IEEE Transactions on Medical Imaging*, 20(10):1061–1071, 2001.

[256] H. J. Weinmann, R. C. Brasch, W. R. Press, and G. E. Wesbey. Characteristics of gadolinium-DTPA complex: a potential NMR contrast agent. *American Journal of Roentgenology*, 142(3):619–624, 1984.

[257] R. Weissleder, G. Elizondo, J. Wittenberg, C. A. Rabito, H. H. Bengele, and L. Josephson. Ultrasmall super paramagnetic iron oxide: Characterization of a new class of contrast agents for MR imaging. *Journal of Radiology*, 175(2):489–493, 1990.

[258] W. M. Wells III, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz. Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 15(4):429–442, 1996.

[259] J. S. Weszka and A. Rosenfeld. Threshold evaluation techniques. *IEEE Transactions on Systems, Man and Cybernetics*, 8(8):622–629, 1978.

[260] Wikipedia. The free encyclopaedia. `http://en.wikipedia.org/wiki/Medical_imaging` (last access on 03.05.07).

[261] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics*, 1(6):80–83, 1945.

[262] P. Willendrup, L. H. Pinborg, S. G. Hasselbalch, K. H. Adams, K. Stahr, and G. M. Knudsen. Assessment of the precision in co-registration of structural MR images and PET images with localized binding. In *Proceedings of the International Workshop on Quantification in Biomedical Imaging with PET and MRI*, pages 275–280, 2004.

[263] G. W. Williams. Comparing the joint agreement of several raters with another rater. *Biometrics*, 32(3):619–627, 1976.

[264] D. L. Wilson and J. A. Noble. An adaptive segmentation algorithm for time-of-flight MRA data. *IEEE Transactions on Medical Imaging*, 18(10):938–945, 1999.

[265] A. Wismuller, A. Meyer-Base, O. Lange, T. Schlossbauer, M. Kallergi, M. F. Reiser, and G. Leinsinger. Segmentation and classification of dynamic breast magnetic resonance image data. *Journal of Electronic Imaging*, 15(1):3020–3033, 2005.

[266] R. P. Woods, J. C. Mazziotta, and S. R. Cherry. MRI-PET registration with automated algorithm. *Journal of Computer Assisted Tomography*, 17(4):536–546, 1993.

[267] G. Wright, D. Nishimura, and A. Macovski. Flow-independent magnetic resonance projection angiography. *Journal of Magnetic Resonance in Medicine*, 17(1):126–140, 1991.

[268] Q. Wu, M. Salganicoff, A. Krishnan, D. S. Fussell, and M. K. Markey. Interactive lesion segmentation on dynamic contrast enhanced breast MRI using a Markov model. In *SPIE Proceedings on Medical Imaging 2006: Image Processing*, pages 1487–1494, 2006.

[269] C. Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3):359–369, 1998.

[270] C. Y. Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3):359–369, 1998.

[271] W. A. Yasnoff, J. K. Mui, and J. W. Bacus. Error measures for scene segmentation. *Pattern Recognition Letters*, 9(4):217–231, 1977.

[272] A. Yezzi, S. Kichenassamy, A. Kumar, P. Olver, and A. Tannenbaum. A geometric snake model for segmentation of medical imagery. *IEEE Transactions on Medical Imaging*, 16(2):199–209, 1997.

[273] T. S. Yoo, editor. *Insight into images: principals and practice for segmentation, registration and image analysis*. A. K. Peters Ltd., 2nd edition, May 2004.

[274] R. Zeboudj. *Filtrage, seuillage automatique, contraste et contours: du prétraitement á l'analyse d'images*. PhD thesis, Saint Etienne University, France, 1988.

[275] H. Zhang, J. E. Fritts, and S. A. Goldman. A co-evaluation framework for improving segmentation evaluation. In *Proceedings of the SPIE on Signal Processing, Sensor Fusion, and Target Recognition XIV*, pages 420–430, 2005.

[276] Y. J. Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition Letters*, 29(8):1335–1346, 1996.

[277] S. C. Zhu and A. L. Yuille. Region competition: Unifying snakes, region growing, and Bayes / MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, 1996.

[278] A. K. Zikou, M. I. Argyropoulou, P. V. Voulgari, V. G. Xydis, S. N. Nikas, S. C. Efremidis, and A. A. Drosos. Magnetic resonance imaging quantification of hand synovitis in patients with rheumatoid arthritis treated with adalimumab. *Journal of Rheumatology*, 33(2):219–223, 2006.

[279] C. Zou, A. Jiang, B. L. Buff, T. G. Mahon, and T. Z. Wong. Automatic motion correction for breasts MR imaging. *Radiology*, 198(3):903–906, 1996.

Appendix A

Registration algorithms

Automated Image Registration (AIR) [266] applicable for intra-subject image registration using a rigid-body model (PET, MRI), inter-modality (PET-MRI) registration, and inter-subject registration;

Automatic Mutual Information-based Registration (AMIR) [135] is a method for assessing the accuracy of CT/PET image registration;

FLIRT: FMRIB's Linear Image Registration Tool [120] performs linear (affine) intra- and inter-modal PET and MRI brain image registration;

Interactive Point Selection (IPS) [262] is a semi-automated landmark-based method with least-squares optimisation, applied for neuroreceptor PET and MRI studies;

Medical Image Processing, Analysis, and Visualisation (MIPAV) [8] performs landmark-based registration on PET, MRI, CT, or microscopy data;

Statistical Parametric Mapping (SPM) [80] incorporates a number of different cost functions related to the mutual information for registration of brain fMRI, PET, and SPECT images;

The Visualisation ToolKit (VTK) [98] is an image processing and visualisation tool. VTK implements affine, grid and thin-plate spline transformations with nearest neighboor, trilinear or tricubic interpolation on meshes, regular sampled, structure and unstructured grids.

Appendix B

Segmentation evaluation

Table B.1: Evaluation of the preliminary segmentation part of the algorithm. *P* – patient number; Slice – slice number. In each cell: [A number of bones detected by the algorithm / a number of bones detected by the expert].

Patient 1	Slice 1	Slice 2	Slice 3	Slice 4	Slice 5	Slice 6
MCPJ2	1/1	1/1	0/0	0/0	1/1	1/1
MCPJ3	1/1	1/1	1/1	1/1	0/0	0/0
MCPJ4	1/1	1/1	1/1	0/0	1/1	1/1
MCPJ5	1/1	0/0	1/1	1/1	1/1	0/1
Patient 2	Slice 1	Slice 2	Slice 3	Slice 4	Slice 5	Slice 6
MCPJ2	1/1	1/1	1/1	0/0	1/1	1/1
MCPJ3	1/1	1/1	1/1	1/1	0/0	0/0
MCPJ4	1/1	1/1	1/1	0/0	1/1	1/1
MCPJ5	1/1	0/0	1/1	1/1	1/1	0/1
Patient 3	Slice 1	Slice 2	Slice 3	Slice 4	Slice 5	Slice 6
MCPJ2	1/1	1/1	1/1	0/0	1/1	1/1
MCPJ3	1/1	1/1	0/0	0/0	0/0	1/1
MCPJ4	1/1	1/1	1/1	1/1	1/1	1/1
MCPJ5	1/1	0/0	1/1	1/1	1/1	1/1
Patient 4	Slice 1	Slice 2	Slice 3	Slice 4	Slice 5	Slice 6
MCPJ2	1/1	1/1	0/0	0/0	1/1	1/1
MCPJ3	1/1	1/1	1/1	0/0	0/0	1/1
MCPJ4	1/1	1/1	0/0	0/1	1/1	1/1
MCPJ5	1/1	1/1	1/1	1/1	0/1	1/1
Patient 5	Slice 1	Slice 2	Slice 3	Slice 4	Slice 5	Slice 6
MCPJ2	1/1	0/1	0/0	0/0	1/1	1/1
MCPJ3	1/1	1/1	1/1	1/1	0/0	1/1
MCPJ4	1/1	1/1	1/1	1/1	0/0	1/1
MCPJ5	1/1	1/1	1/1	1/1	0/1	1/1
Patient 6	Slice 1	Slice 2	Slice 3	Slice 4	Slice 5	Slice 6
MCPJ2	1/1	1/1	0/0	0/0	1/1	1/1
MCPJ3	1/1	1/1	1/1	1/1	0/0	1/1
MCPJ4	1/1	1/1	1/1	0/0	1/1	1/1
MCPJ5	1/1	0/0	1/1	1/1	1/1	1/1
Patient 7	Slice 1	Slice 2	Slice 3	Slice 4	Slice 5	Slice 6
MCPJ2	1/1	1/1	1/1	0/0	0/0	1/1
MCPJ3	1/1	1/1	1/1	1/1	0/0	1/1
MCPJ4	1/1	1/1	0/0	0/0	1/1	1/1
MCPJ5	1/1	0/0	1/1	1/1	1/1	1/1
Patient 8	Slice 1	Slice 2	Slice 3	Slice 4	Slice 5	Slice 6
MCPJ2	0/0	0/0	0/0	0/0	0/1	1/1
MCPJ3	1/1	1/1	1/1	1/1	0/0	0/0
MCPJ4	1/1	1/1	0/0	0/0	1/1	1/1
MCPJ5	0/0	0/0	1/1	1/1	1/1	1/1
Patient 9	Slice 1	Slice 2	Slice 3	Slice 4	Slice 5	Slice 6
MCPJ2	1/1	1/1	0/0	1/1	1/1	1/1
MCPJ3	1/1	1/1	1/1	0/0	0/1	1/1
MCPJ4	1/1	1/1	1/1	0/0	1/1	1/1
MCPJ5	1/1	0/0	1/1	1/1	1/1	1/1
Patient 10	Slice 1	Slice 2	Slice 3	Slice 4	Slice 5	Slice 6
MCPJ2	1/1	1/1	1/1	0/0	1/1	1/1
MCPJ3	1/1	1/1	1/1	1/1	0/0	1/1
MCPJ4	1/1	1/1	1/1	0/0	1/1	1/1
MCPJ5	1/1	0/0	1/1	1/1	1/1	1/1

Appendix C

Evaluation by human observers

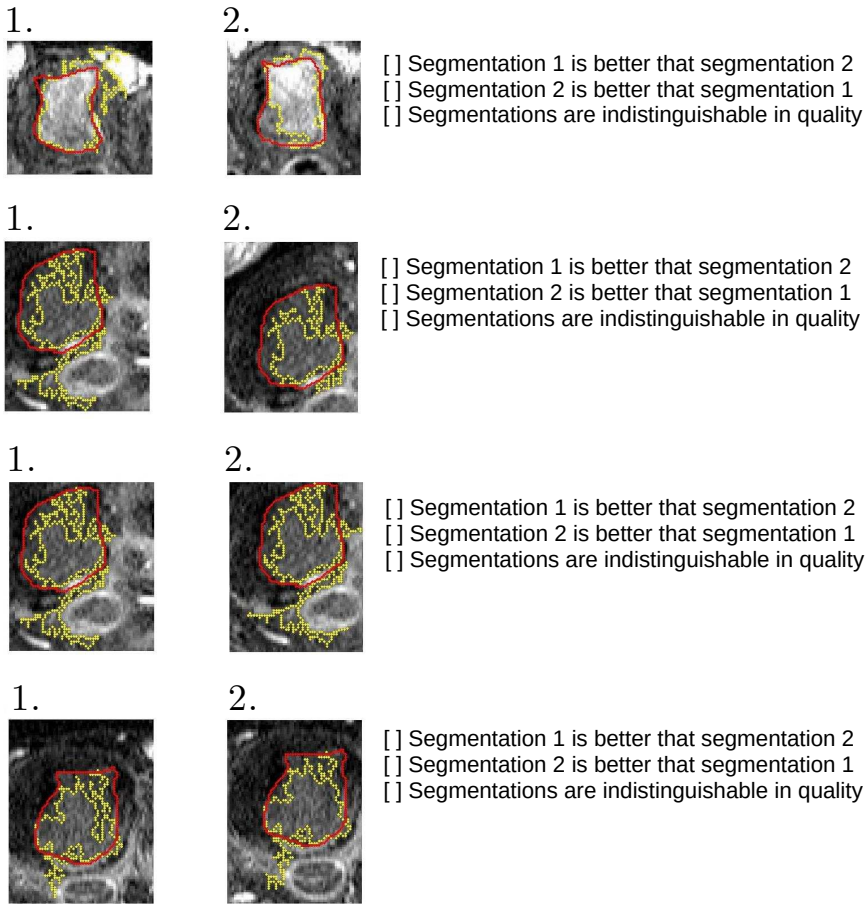


Figure C.1: Four pairs of segmentation outputs given to the human observers. GT overlay is shown in red, machine segmentation in yellow.

Appendix D

Number of pixels under consideration

[Number of pixels within joint interior excluding bone interiors / Number of pixels that exhibit wash-out phase]; *P* – patient number; Slice *N* – temporal slice number.

	Slice 1	Slice 2	Slice 3	Slice 4	Slice 5	Slice 6
P_1	17053/6807	18038/6515	17883/6694	18425/6490	16471/5504	16949/5908
P_2	11557/4845	11631/2382	11359/3801	11995/3532	10577/2265	10149/1921
P_3	14809/7455	15697/5473	15507/5033	13828/4271	12681/3761	11558/2743
P_4	17256/5241	18288/5187	18827/5354	19269/4642	16618/4140	12727/3034
P_5	14882/5174	15397/4241	15893/5118	15599/4522	16100/4500	12308/4150
P_6	12341/3756	13067/3807	12942/3582	12383/3216	11203/2778	10452/2706
P_7	11617/4515	11600/3513	11385/3368	12076/3224	10586/2659	9963/2305
P_8	19360/5726	23394/4916	18568/4885	18368/4840	17679/4300	16812/3521
P_9	13994/4003	13535/3623	13972/3423	14382/818	13516/3590	12004/2552
P_{10}	14008/4254	13512/3578	13656/3529	14400/3658	12766/3926	11935/2538