Unifying Models and Registration: A Framework for Model-based Registration and Non-rigid Registration Assessment

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy in the
Faculty of Medical and Human Sciences
2006

Roy Samuel Schestowitz

Division of Imaging Science and Biomedical Engineering

Contents

List of Figures

Abstract

Statistical models of shape and appearance are widely used for the analysis of biomedical images. Two deficiencies of these models is that they require annotation across a large number of images in order to be built and, having built such models, it is then difficult to reason about their validity or assess their quality. Herein, a method is described which addresses both problems and establishes a unified solution. In order to construct the models rapidly, corresponding structures must be brought into the state of dense overlap. Image registration is the mechanism whereby a set of images can be analysed in a common frame of reference and models then derived from it. The thesis provides a solution to the recurring need to compare such models and extends the method as to provide an image registration assessment method, which does not require ground truth. The thesis also deals with a complementary case where images are registered by minimising the complexity of models. Overall, the proposed framework can be perceived as one which ucombines registration and modelling, taking advantage of the fact that they are innately the same. Registration provides correspondence across images and, given that correspondence, models of appearance can be built and registration

then assessed, without the need for ground-truth data.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

Dedication

I dedicate this thesis to my grandmother, who passed away amidst my studies.

Acknowledgements

I would like to express my thanks to:

- A Supervisor and role model, Chris Taylor, for investing plenty of time and effort to make my project a success.
- My Advisor, Steve Williams, for support, kind words of advice and perpetual encouragement.
- Katherine Smith, Carole Twining, and Stephen Marsland, who initiated much of the work which the thesis relates to, as well as builds upon.
- Tim Cootes, who helped me cope with C++ and VXL (among other technicalities). He has always done so while remaining patient and keeping pressure (or sense of demand) at a bare minimum.
- The EPSRC, for funding the project and connecting me with those who are involved in the MIAS-IRC. This 'umbrella' offered valuable feedback, advice, and a very motivational setting.

- My parents and siblings, as well as my grandfather, Avner Werner Max, for full financial support from my very first day at the University.

- My grandmother, who is no longer among us (passed away in 2005). She was a supportive and loving figure in a family which has been volatile for the past 2 years.

- Dr. David Baxter, whose advice led me to consider research at Manchester University and, more particularly, the division where I ended up enrolling for this degree.

- Prof. Tony Hegarty, who urged me to accept an academic route, rather than be distracted by the wrong career paths.

- David Robinson, who offered valuable advice and encouraging words towards the end of my Ph.D. He was one of those who urged me to put in extra time and rigour, gearing up towards completion.

- Other friends at the health club, who inspired and provided comfort whenever things went amiss or uncertainties loomed over.

- The Open Source community, which enabled this research to be carried out using versatile, distributable, Free software.

- Mathworks, for enabling me to share code in their Web site and thereby reach a very wide audience, putting to use some personal programming ambitions. Many of my endeavours were boosted by knowledge that the outcome would be shared among the MATLAB users community.

- Roland Selby, Mark O'Leary and several other people at Manchester Computing, who permitted me to concentrate on my Ph.D. studies while at work. They were surprisingly understanding, particularly when pressure was mounting.

- David Kennedy of the Center for Morphometric Analysis at MGH, for providing the fully-annotated brain images. Additional images from age-matched normals in a dementia study were generously provided by Prof. Alan Jackson, University of Manchester.

Publications

Portions of the work described in this thesis has also appeared in:

Conference papers

- Carole Twining, Tim Cootes, Stephen Marsland, Vladimir Petrovic, Roy Schestowitz, and Chris Taylor. A Unified Information-Theoretic Approach to Groupwise Non-Rigid Registration and Model Building. Presented in *Information Processing in Medical Images (IPMI)*, Lecture Notes in Computer Science, vol. 3565, pp. 1-14, 2005.

- Tim Cootes, Carole Twining, Vladimir Petrovic, Roy Schestowitz, and Chris Taylor. Groupwise Construction of Appearance Models using Piece-wise Affine Deformations. Presented in *British Machine Vision Conference (BMVC)*, vol. 2, pp. 879-888, 2005.

- Roy schestowitz, Carole Twining, Tim Cootes, Vladimir Petrovic, Chris Taylor, and Bill Crum. Assessing the Accuracy of Non-Rigid Registration With and Without Ground Truth. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2006.

- Carole Twining, Tim Cootes, Stephen Marsland, Vladimir Petrovic, Roy schestowitz, and Chris Taylor. Information-Theoretic Unification of Groupwise Non-Rigid Registration and Model Building. In *Medical Image Understanding and Analysis (MIUA),* vol. 2, pp. 226-230, 2006.

- Roy Schestowitz, Carole Twining, Tim Cootes, Vladimir Petrovic, and Chris Taylor. A Generic Method for Evaluating Appearance Models. Presented in *Proceedings of MIAS-IRC Plenary Meeting*, 2006.

- Roy schestowitz, Carole Twining, Tim Cootes, Vladimir Petrovic, Bill Crum, and Chris Taylor. Non-Rigid Registration Assessment Without Ground Truth. Presented in *Medical Image Understanding and Analysis (MIUA)*, vol. 2, pp. 151-155, 2006.

Peer-reviewed papers/symposia

- Roy Schestowitz, Carole Twining, Tim Cootes, and Chris Taylor. Image Registration by Model Criteria. Presented in *Proceedings of MIAS-IRC Plenary Meeting*, pp. 16-17, 2004.

- Roy Schestowitz, Bill Crum, Vladimir Petrovic, Carole Twining, Tim Cootes, and Chris Taylor. Assessing the Accuracy of Non-Rigid Registration. Presented in *Proceedings of MIAS-IRC Plenary Meeting*, pp. 25-26, 2005.

- Vladimir Petrovic, Tim Cootes, Carole Twining, Roy Schestowitz,

and Chris Taylor. Groupwise Construction of Appearance Models using Piece-wise Affine Deformations. Presented in *Proceedings of MIAS-IRC Plenary Meeting*, pp. 21-22, 2005.

Journals (submitted, under review)

- Roy schestowitz, Carole Twining, Vladimir Petrovic, Tim Cootes, Bill Crum, and Chris Taylor. Evaluating Non-Rigid Registration without Ground Truth. *IEEE Transactions on Medical Imaging*.

Code and material produced throughout the project was generalised and publicised in the form of utilities and tutorials. It received over 35,000 downloads at MATLAB Central and had the author ranked 1^{st} in the world at one point. This work is now internationally recognised and receives close to 2,000 downloads per month.

Accompanying CD-ROM

Videos demonstrating parts of the thesis have been put on a supplementary CD-ROM. This CD-ROM is bound to back cover of this thesis. It contains illustrative animations that are sometimes referenced in the text. The filenames on the CD-ROM are either arbitrary, chronological, or correspond to figure/section numbers in the text.

Peripheral files (GIF-formatted):

- p1.gif: combined model built from the MGH data. 7 modes of variation are normally (i.e. drawn from Gaussian distribution) varied simultaneously.
- p2.gif: combined model built from the MGH data. 10 modes of variation are normally varied simultaneously.
- p3.gif: the registration assessment framework illustrated schematically
- p4.gif: discrepancy image (checkerboard-type composition) of 2 brain images. The sequence shows the discrepancy image as an SSD-

based registration proceeds and, in response, the discrepancy image evolves.

- p5.gif: one-dimensional registration example. Rows in the matrix represent intensity vectors (1-D images) being registered to the remainder of the vector set in a pair-wise fashion.

Larger peripheral files (compressed or uncompressed AVI format, no proprietary codecs required):

- p6.avi: a set of 10 1-D vectors are being aligned using multi-edge clamped plate spline (CPS) warps. 10 iterations (passes) through the data are shown, unregistered images on the left and progressively re-registered images on the right.
- p7.avi: a set 10 1-D vectors, as visualised in 3-D space, are being aligned. 50 iterations through the data are shown in the sequence.
- p8.avi: the first and second modes of a combined (shape and intensity) model which is automatically built
- p9.avi: two 1-D images are being registered. Unregistered images are shown on the left and progressively registered – on the right.
- p10.avi: 1-D vectors being registered using a model-based objective function

- p11.avi: 10 simplified 1-D vectors, which are composed of 4 edges, are being registered. The objective function is based on mean-squared-differences.
- p12.avi: 10 1-D vectors with are being registered using an objective function that based on minimisation of the complexity of a point distribution function
- p13.avi: A large number of 1-D vectors (visualised as rows) being registered by considering just one vector at a time
- p14.avi: automatically-built combined model of a bump. The model is built automatically from the raw training set.
- p15.avi: A large-scale illustration of 1-D bump registration

Prologue

"The classical *synthesis problem* of computer graphics can be formulated as the problem of generating novel images corresponding to an appropriate set of parameters describing the camera viewpoint and aspects of the scene. The inverse *analysis problem* of estimating object labels as well as scene parameters from images is the classical problem of computer vision..."

– *David Beymer* [5].

Chapter 1

Introduction

"A mathematician is a device for turning coffee into theorems."

– *Paul Erods.*

THIS thesis outlines a novel approach to the evaluation of statistical modelbs of appearance [23], which can also be used to assess the quality of non-rigid registration (NRR) algorithms. Additionally, a method is presented for registering images, using model complexity as a criterion which provides a figure of merit [64]. The work is motivated by the observation that, given a set of registered (i.e. fully-aligned) images, appearance models can be built automatically and then be evaluated [63]. Another key observation is that generative models of any reasonable form can become the direct product of registered images, which need not necessarily be non-rigidly aligned. By reconstructing/generating images from the model, these registrations can be assessed. The ability to assess registration algorithm is important for benchmarks and comparative studies

that help improve or hand-tweak a given NRR algorithm.

This work contributes to the framework of modelling, as well as the popular and intricate study of non-rigid image registration [29]. The approach is broad and generic, but the thesis will focus on 2-D brain images, alluding to the existing extension and implementation in 3-D, as well as similar work on face images. At the 'proof-of-concept' stages, 1-D images are used as well. These are helpful in validation experiments that exploit synthetic images whose nature is well understood.

The overall aim of the work is to demonstrate that appearance models can be built automatically and registration be driven by the quality of appearance models. Concurrently, both models and registration are implicitly evaluated, owing to their innate bond. This reciprocal relationship is not only proven, but it is also shown to have promise in practical applications. This facilitates and manifests a variety of important experiments, including benchmarks that help discern one registration or model construction algorithm from inferior ones [65].

1.1 Image Registration

In medical imaging, one particularly important task to tackle is that which involves simplification of the vast amounts of information at hand. With advancements in technology, more data is gathered than a human is able to analyse. The level of redundancy and excess in the available data requires that various steps should make it more manageable. By

reducing the complexity or data and making it more cohesive, valuable information can be extracted from it, whereas unwanted residues are left aside.

An expert in a specialised field, for example, may wish to perform an analysis on a considerable number of images acquired from particular groups of subjects. Each group will often be characterised by various distinguishing features, but in order to study the group *as a whole*, images from each subject need be better assimilated, e.g. through transformation, to the remainder. As a result of applying sensible transformations to the image, understanding of structural change, whether pathological or not [73], becomes more trivial. There are two common cases where such studies have significant merits:

1. **An intra-subject study.** This involves analyzing a series of images taken from the same subject and comparing them. These images can be acquired over a period of time, as means of learning the progression and regression of atrophies, for instance. In other scenarios, the images might be a series of analogous slices, which are extracted from a three-dimensional volume, where there is a need to correct and compensate for motion.

2. **An inter-subject study.** Typically, such studies involve a comparison between two (or more) groups of subjects. In a medical context, this makes possible the discovery of symptoms for a certain groups of patients, observing how they deviate from 'normals'.

It is evident that the need to compare images is rather fundamental. It

forms the very basis for much of the above to become practical. Non-rigid registration (NRR) is the methodology used to address problems of this kind. NRR algorithms are intended to annul variations in pose, as well as in form, across a collection of images. Given a set of images, all of which contain something similar, one wishes to repeatedly transform them until they appear most identical [21]. There is no consensus which suggests that only one particular algorithm should be used. The problem is highly under-constrained and many different algorithms have been proposed to solving it. There is a wealth of algorithms that compete over performance, where a measure of performance is, in itself, subjective. There is a clear distinction, however, between two general approaches: groupwise and pairwise. Each case will be considered in turn.

NRR of both pairs or groups of images is used widely as a basis for medical image analysis and actual applications include structural analysis, atlas matching and change analysis [13]. The aim of NRR is to find, automatically, a meaningful dense correspondence between a pair (*pairwise* registration), or across a group of images *groupwise* registration). A typical algorithm consists of a representation of the deformation fields that encode the spatial variation between images, an objective function that quantifies the degree of misregistration, and a method of optimising the objective function with respect to the deformation fields. As different algorithms generally produce different results when applied to the same set of images [93], there is also a clear need for methods to evaluate the results of NRR. One interesting question to address, for instance, is whether a groupwise registration outperforms a similar pairwise ap-

proach.

Various methods of evaluation have been proposed [36, 60, 66]. One approach is to construct artificial test data, applying known deformations to real or synthetic images. This allows algorithms to be evaluated by attempting to recover the applied deformations, but does not allow the results of NRR to be assessed 'in-line' in real applications. An alternative approach is to provide anatomical ground truth for the images to be registered, then measure the degree of anatomical correspondence following NRR. One such method is described in this thesis as a 'gold standard', but the need for expert annotation of the images renders the approach too time-consuming and subjective for routine application. These problems motivate the search for a method of evaluation that can be used routinely in real applications, without the need for ground truth. There is potential in the use of statistical models – a potential that arises owing to numerous overlaps between NRR and modelling.

1.2 Models and Registration

The approach adopted for assessment is based on the observation that, given a set of non-rigidly registered images – however obtained – it is possible to construct a statistical model of appearance that takes account of both the shape and texture variation across the set. Models of this type have been used extensively as a basis for image interpretation by synthesis [12]. To build a model one can exploit the dense correspondence

across the set of images established by the NRR. The key idea that underpins the approach is that, if the correspondence is poor, the resulting appearance model will be unsatisfactory. When the correspondences are correct, the model will be simpler. The model will also faithfully reflect on and be able to reproduce the correspondent images. This observation transforms the problem of evaluating non-rigid registration into one of evaluating the model generated from the result of registration.

1.3 Exploiting the Tie Between Registration and Models

The main merit of the work is the introduction of a generic method for assessing the quality of non-rigid registration [63]. The method *does not* require ground truth, but rather depends solely on the registered images. Consider the case where NRR is applied to a *set* of images, providing a dense correspondence between images. Given this correspondence, it is possible to build a generative statistical model of appearance variation for the set. The quality of the resulting model will depend on the quality of the correspondence. Measures of model *specificity* and *generalisation* can be used to assess the quality of the model and, hence, the quality of the correspondence from which it is derived. The approach does not depend on the specifics of the registration algorithm or the form of the model.

Validation of this approach is performed by measuring the change in

model quality, as the correspondence of an initially registered set of MR images of the brain is progressively perturbed, and comparing the results with those obtained using a method based on the overlap of ground-truth anatomical labels. This demonstrates that, not only is the proposed approach capable of assessing NRR reliably without ground truth, but that it also provides a more sensitive measure of misregistration than the overlap-based approach. It is then possible apply the new method to compare the performance of different registration algorithms on a several sets of MR images of the brain, demonstrating that the method is able to discriminate between different methods of registration in a practical setting.

Since models are used throughout the entire process, evaluating the quality of models is possible and different methods of constructing appearance models can thus be compared [65]. Also of interest are methods that enable models to be built directly from the data whilst models serve as the similarity measure in the objective function. This is essentially a case or reversing the problem, attaining good registration by optimising the quality a model whose data is manipulated [64].

1.4 Contributions

The contribution of the work is two-fold. On the one hand, it is demonstrated that one is able to evaluate the quality of NRR, without needing ground-truth data. Thus, a comparison between numerous NRR algorithms can be made without cumbersome manual annotation. On the

other hand, one is also able to build models automatically, using registration algorithms, and then evaluate the resultant models. All in all, this provides a framework for automatic or semi-automatic analysis of arbitrarily large amount of data, assuming that enough images are made available for an appearance model to be constructed (the caveat).

1.5 Thesis Organisation

The structure of the thesis is as follows:

Chapters 2 and 3 provide an augmented description of the background to both the assessment of registration, and the construction of appearance models (respectively), explaining in more detail the link between the two.

Chapter 4 outlines previous work on MDL for shape model. This work is essential as it comprises some of the ground work, upon which this thesis is based.

Chapter 5 briefly outlines an algorithm that enables non-rigid registration algorithms to be driven by minimisation of model complexity and it also shows some corresponding results. The results are not only include a registered set of images, but also their appearance model.

Chapter 6 defines two quantitative measures of model quality as well as registration quality, and discusses their implementation.

Chapter 7 is intended to focus on method validation. The behavior of aforementioned measures is investigated by measuring the effect of deliberately perturbing the registration of an initially registered set of images. The results are compared to those obtained using a 'gold standard' method of assessment, based on measuring the overlap of manually-annotated ground truth. The results demonstrate that our new measures are closely correlated with those based on ground-truth, and that the proposed approach is actually *more* sensitive to misregistration

Chapter 8 presents practical applications. The measures developed are used to compare three NRR algorithms applied to the registration of sets of 2-D MR brain images, demonstrating the superiority of a fully groupwise registration algorithm over a repeated pairwise approach.

Chapter 9 described the extension of the method to 3-D, as well as limitations.

Chapter 10 lists several possible extensions and several ways forward.

Chapter 11 draws conclusions and contains a summary of the contributions of the work.

Chapter 2

Non-rigid Registration

"Having a set, popular formula does inhibit you."

– *George Shearing.*

THE aim of non-rigid registration is to identify an anatomically-meaningful, dense (i.e., pixel-to-pixel or voxel-to-voxel) correspondence across a set of images. This correspondence is typically encoded as a set of spatial deformation fields, one for each image, such that when the deformations are applied to the images, corresponding structures are brought into alignment.

A typical registration algorithm proceeds by optimising an objective function, which depends on the similarity of the images after alignment, with respect to the set of deformations [54]. As well as the objective function, it is necessary to define the representation used for the deformation fields and the method for finding the optimum of the objective function. Different choices lead to different registration results and competing methods

of NRR – hence the need for an objective, easily-applied method of assessment, as described in the remainder of this chapter.

While the thesis' primary point of focus is assessment and comparison of NRR algorithms [63, 9], the remainder of the chapter explains how constituent parts of the registration process interact with one another. It also surveys a variety of methods that are actively used to achieve a fully-functional NRR algorithm, alluding to the new model-based algorithm, which is described in Chapter 6.

2.1 Background

Image registration is an essential image processing step, which has entered several domains where reliable acquisition of fully-aligned images cannot be assured [29] or relationships between images turn out to be overly complex. The significance of this problem is made most apparent when alignment of large *groups* of images needs to be achieved [80]. In some cases, the images under consideration are rather different in terms of their nature, even though they contain exactly the same type of object. This leaves place for ambiguity – and consequently – misinterpretation.

Misalignment in images can result from movement of subjects or objects of interest, change in view-point, or changes to general conditions at the acquisition site. Misalignments can also be artifacts of morphological changes, or physical anomalies that are due to change in mass and elasticity of organs [30]. Changes in form can be observed over time some-

where within an object's constituent parts, e.g. the involuntary changes in the form of the subject's lungs. In some circumstances, as later discussed, misalignment incurs due to profound changes in the form of objects (typically *subject*s and their anatomy) being scanned. A state of near-perfect alignment, which is reached through NRR algorithms, is a key step that should often be completed before any analysis stage of a collection of images is safely embarked on. This facilitates and caters for better understanding of the contents of several images (and more cohesively so, as a group).

Given a collection of images, all of which depict the same object, one wishes to *transform* them in one way or another, so that they appear as similar as possible to one another. The solution to this problem is never unique as there will be infinitely many solutions, i.e. transformations, which lead to similar results. As such images may not contain precisely the same elements, there is rarely a ground-truth solution either, i.e. there is no definite one-to-one correspondence between imaged objects. Absence of or reappearance of finer elements, for example, implies that no point-to-point correspondence can always be determined, so good solutions need be *approximated* instead.

The field which is associated with this problem is uniformly referred to in the literature as "non-rigid registration". The work described in subsequent chapters underlines and extends a methodology that is devised to assess the quality of NRR, based on solutions reached by NRR algorithms. It is therefore important to elaborate on what is involved in any NRR algorithm, highlighting different approaches and parts of the pro-

cess. Principles of registration, particularly with respect to approaches which this thesis revolves around, will be dealt with in turn. The subject is by all means broad and for deeper understanding of alternative approaches, cited literature can be carefully read [29].

Approaches to NRR – those on which research around the world is based – are rather distinct, but are all built upon the same ideas. There are commonalities and so-called 'components', which NRR is logically based on. This chapter identifies these components, giving an overview with special emphasis on methods which are said to hold the very key to better NRR algorithms.

2.2 Transformation

2.2.1 Overview

Image registration ordinarily involves the manipulation of image pixels (or in a volumetric context – voxels). The product of this is dense mapping that describes where each pixel/voxel moves once the transformation (or warp) gets applied. This morphometric manipulation is done in accordance to a set of rules and with a common grand goal, which requires that transformation as a whole is sufficiently versatile. There are various image warping methods that achieve a full transformation which affects an entire image. The warps are subjected to conditions that make them valid, i.e. transformation is carried out under the imposition of strict constraints.

It is commonly desirable to attain a maximal cross-image similarity estimate [59] as images are being warped. In essence, a greater degree of overlap amongst a group of images is sought. This similarity can be reach by applying warps to the images and this should optimally be done with minimal extents of distortion because the integrity of the image should be preserved. Better similarity is achieved by applying *changes* to these images, which is where transformation fits in. Transformation is the means by which one approaches greater similarity among several images.

2.2.2 Transformation Types

One can perceive the different transformation types as though they pertain to different levels of 'interference' – that is – the interference to analysis and intervention with the integrity of data being manipulated. Some of the more permissive transformations violate and eliminate a state of reversibility. Once applied, there is ambiguity which prevents the transformation from being retracted (applied in reverse). A typical classification of transformation types is as follows (ordered by increasing interference or severity):

Rigid Transformation. Permits translation (relocation in space), rotation, and scaling (albeit only uniform size changes, i.e. shrinkage and enlargement)[1]. In hyperspace, normalised shape attributes are altogether preserved, so the process is usually concerned with

[1]More strictly, the inclusion of scaling makes this a Similarity transformation, rather than rigid.

a more fundamental alignment. Such alignment is ordinarily intended to position all data instances (images or volumes) upright and centred at the origin of a hyperspace, with a fixed size of 1 unit at most (fixed-sized hypersphere). All images are virtually confined to lie inside a bounding structure (a circle or sphere, in 2-D and 3-D respectively). In 3-D, for instance, there is a total of 6 degrees of freedom so a rigid transformation will be wholly characterised by a tuple of 6 parameters[2]. These parameters fully describe a rigid transformation.

Affine Transformation. Allows an image to *stretch and skew* along at least one axis (corresponding to a parametric dimension), yet not *necessarily* along several of them simultaneously. This ensures that a homogeneous scaling – that which affects all dimensions uniformly – will not be invalidated. Despite the fact that consistency is compromised, lines which were parallel before an affine transformation is applied will remain parallel after the transformation is applied[3]. Reconstruction is said to be possible so this transformation is invertible. Essentially, for a given affine transformations $T_a(x)$, where x is a vectorised representation of an image (or volume), and its inverse $T_a^{-1}(x)$, the expression $T(T_a^{-1}(x) = Id(x)$ must hold true. It retains a level of simplicity, which makes it easy to determine and resolve. This proves to be an important constraint when the practicability of

[2] 1 value for scaling, 3 for *x, y* and *z* coordinates and 2 for rotation, e.g. the *xy* and *yz* angles θ_1 and θ_2.

[3] Other transformations such as shear and taper, on the contrary, are not parallelism-preserving. The importance of this rigorous constraint is that the distance between any two points remains proportional to the transformation.

warps is further debated.

Non-rigid Transformation. All other valid transformations fall into this category [21]. In principle, no inviolable constraints are in place, but quite clearly, in a practical setting, a non-rigid transformation attempts to preserve some of the key structures[4] in the image while abstaining from excessive tearing and folding [62, 77]. This means that each pixel in the range must map to another and no pixel is left undefined. More on this is to be discussed later, in the context of diffeomorphism.

The images of an apple in Figure 2.1 illustrate the effect that each transformation type is permitted to have on the original image shown on the left.

As the figure suggests, the appearance of an object remains identical under rigid transformations. Images objects are allowed strictly to grow, shrink, move, and rotate. Affine transformation allows an object to lose its original form whereas non-rigid registration is far more permissive, so the object can be subjected to rather arbitrary deformations.

2.2.3 Diffeomorphism

There are important factors to consider when selecting a transformation method. Ideas which were introduced so far in this section confirm that

[4] A random uncontrollable transformation will disintegrate basic structures in the image and can make valid interpretation impossible.

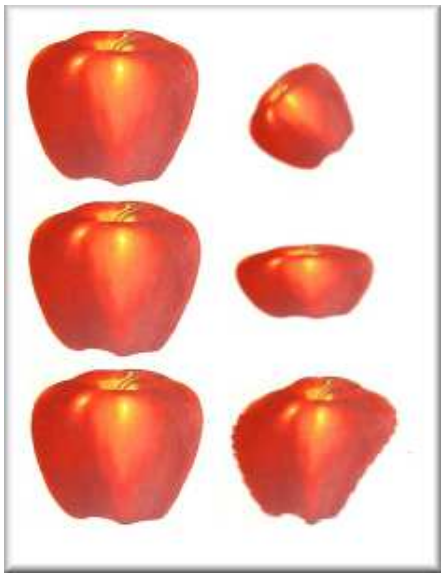

Figure 2.1: Registration examples. On the left column: the original, unwarped image; on the right column, from top to bottom: rigid, affine, and non-rigid transformations.

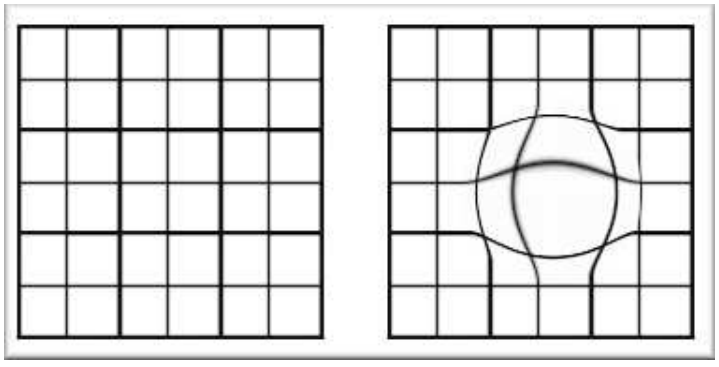

Figure 2.2: A pseudo-non-rigid warp example. The effect of the warp is shown on the right hand side.

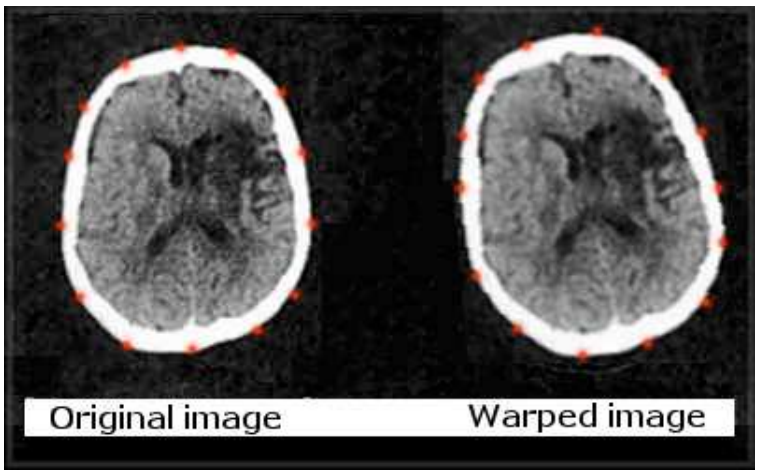

Figure 2.3: An example of image warping in medical contexts (the human brain). Red points that are overlaid on the skull depict knot-points for the splines that render a transformation, which is based on clamped plate splines.

there is an ever-increasing need for non-rigid registration algorithms that prevent the 'erosion' of image structures. Diffeomorphic [75] functions are *invertible, continuous* and *one-to-one* mappings, which can be applied to a given image[5]. These functions can be described by local geometric transformations that have an effect on groups of pixels, or the plane that pixels are embedded in.

Diffeomorphic transformations that are used in this work were initially devised by Twining and Marsland [76]. These benefit from having continuous derivatives at the boundaries, unlike for example, these proposed by Lötjönen and Mäkelä [45]. Diffeomorphism is a key property which is not a necessity. It is, however, a good warp attribute to have in real-world applications.

What Invertibility, continuity and one-to-one mappings mean, in simpler

[5]More generally, such functions are mappings defined over a matrix or a vector, which herein is analogous to an image.

terms, is that for each transformation:

1. The transformation has an calculable inverse transformation. This way, any transformation can be reversed, i.e. its effect retracted.

2. The transformation affects *all* data (e.g. image pixels) within its boundaries so it has a spatially-contained effect[6]. This means that every point must move as would be expected to give a continuous flow of intensities.

3. No two points should be mapped onto the same point as this would 'strip off' areas of the image, depleting them from data. These effects are also known as tearing and folding, both of which are notorious artifacts that need to be avoided.

Diffeomorphic warps are applied to the space in which images will be embedded. That newly-defined plane is intended to bring the collection of structures across the set of input images closer together. This ultimately brings a number of images to correspondence of better quality. The quality varies depending on a pre-defined objective function, as well as the warp representation and the similarity measures, as later explained.

Looking more closely at diffeomorphic functions, the spread of resampled points can be defined purely by a function and a reparameterisation which alters this function to find preferable matches. A monotonically-increasing function describes the distance of all points[7] from an arbi-

[6]A pixel of course can be mapped onto the exact same original position, but the idea is that a continuous flow should prevail.

[7]A continuous function is independent of the number of points. Therefore, the complexity can be increased progressively to obtain finer, more accurate results.

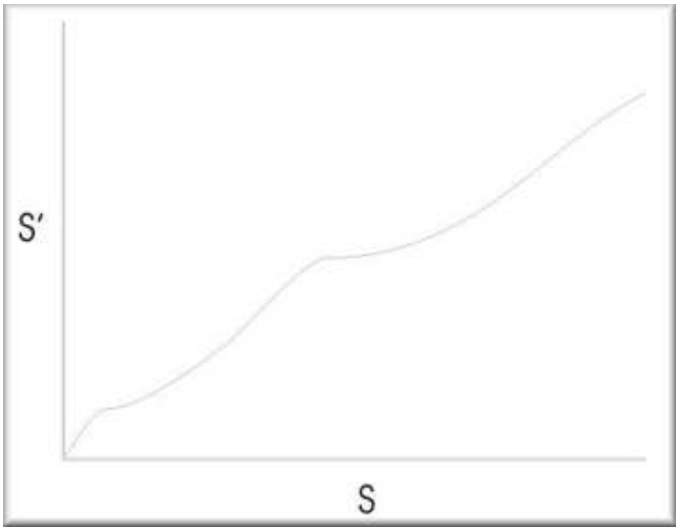

Figure 2.4: Monotonically-increasing function illustrated in a simplistic case. Each point is mapped from position S onto position S' along a one-dimensional curve.

trary point on the curve in such a way that will not violate their original sequential order.

Figure 2.4 shows what is meant by a monotonically-increasing function. The following expression is a more formal description and its exact inverse may hold instead (alluding to a monotonically-decreasing function). Consider the case

$$\forall(u \in S \wedge v \in S \wedge u < v) \rightarrow f_{mon}(u) < f_{mon}(v) \tag{2.1}$$

where f_{mon} is the monotonically-increasing function used and $f_{mon}(S) = S'$. More simply, the derivative at any point must be positive, i.e. $0 < \theta < 90$ so that $0 < tan(\theta) < 1$. In Figure 2.4, the distance or offset along the curve is guided by the value which was determined by the function above.

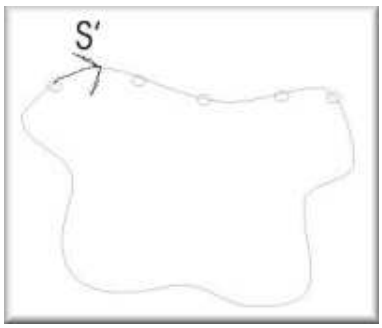

Figure 2.5: Reparameterisation example in 1-D. A point moves along the curve a distance S' from the origin. All other points will do so as well to make this a continuous reparameterisation, Each point is moved some distance away from the origin, but no point can override another which leads to a clash (and thus ambiguity in interpolation).

In this particular way, all points which lie on the curve can be moved *simultaneously,* without 'colliding' with one another and new descriptors of shape become available. Instead of describing the movement of each individual point, an arbitrary number of points can be shifted according to one modifiable function.

2.3 Similarity Measures

2.3.1 Introduction

The rest of this chapter examines and connects concepts, techniques, and ideas which are being employed to 'glue' together warps and similarity measures. Warps and similarity are the two main components of any non-rigid registration algorithm. More broadly, there are 3 separate 'compartments' to consider, namely:

1. Warps
2. Similarity
3. Objective function

Having covered the first, the latter two points will be explained in greater detail with reference to practical considerations. The approach taken is that an image needs to be gradually warped until it matches another. The match is estimated using similarity measures. The process of warping and similarity falls under one generic objective function, which is an 'umbrella' in this context. In that sense, an objective function serves as a bridge for warps and similarity, making the selection of warps improve the similarity. Objective functions are then handled by an general optimiser – that which selects warps that increase similarity. The remainder of this section deals with various methods of measuring image similarity.

2.3.2 Methods of Measuring Similarity

.

There are various ways of measuring a perceived similarity between two images. Sum of pixel- or voxel-wise differences (as well as sum-of-*squared*-differences or even the *mean* rather than a summation) emerges as the most intuitive method, which merely accumulates pixel- or voxel-wise differences between the images. This measure, however, is rather poor at gauging a meaningful extent of similarity, particularly if positions where

the images which lie in hyperspace[8] are a just short distance apart. This can be thought of a case of slight miscorrespondence, wherein the images are almost fully aligned. This fairly intuitive measure is a good one to use when convergence in NRR is foreseen.

Other image similarity measures are relatively immune to larger spatial displacements and mild variabilities in actual form. Histograms of intensity values in the images have seen a noticeable rise in practical use. Intensity values are accounted for globally, or even locally, e.g. inside regions whose impact on similarity estimation should be greater. These measures prove to be far better assessors of similarity under most circumstances. Mutual information and normalised mutual information, as described by Studholme [71], provide good histogram-based measures that see high usage in existing non-rigid registration algorithms. Each one will be dealt with in turn in the remainder of this section.

Another method for measuring similarity makes use of the correlation ratio. Due to the nature and scope of this thesis, it is less relevant and, in principle, goes back over half a century ago [37].

Mutual Information (MI)

Viola [81] developed a method[9] of measuring similarity between two images by repeatedly comparing histograms of pairs of images [54, 44].

[8] One can think of images as though they have been reduced to a vector of pixel values, which map onto a position in a high-dimensional space.

[9] The discovery of mutual information is also attributed to Maes, yet the work was sparked by Viola in the mid-nineties.

When measuring mutual information, one computes *informational overlap* across images. If two images are properly aligned, their joint histogram is indicative of where *sharp* grey-value peaks are located, as well as the sharpness value of these peaks. Under the complementary case, which is mis-registration, the joint histogram occupancy is expected to include peaks with low magnitude and new steep peaks can emerge. By defining a joint information (or entropy) to be $H(A, B)$ and a the information contained in a single histogram A to be $H(A, B)$, it is reasonable to argue that MI calculates $H(A) + H(B) - H(A, B)$. There are variants thereof [62], but the prime idea is that joint information is subtracted from the sum of information present in the two individual images.

Normalised Mutual Information (NMI)

Studholme [71] and Maes [46] suggested that normalisation should be applied to mutual information. Several steps are involved in this normalisation process[10]. The main difference is that the expression used for MI is significantly extended and divided by a normalisation term. The method is predominantly used in non-rigid registration as it is generic, adaptable to new data, and yields better results.

[10]There is an additional distinction between symmetric and asymmetric normalised mutual information, but rationale for this requires the full technical recipe. The dissertation at `http://www.lans.ece.utexas.edu/~strehl/diss/node107.html` summarises the way in which NMI is evaluated.

Sum-of-Squared-Differences

One of the most intuitive and least resource-intensive approaches is the sum of differences and its variants. Pixels are being compared in two images, one pixel at a time, and their (potentially squared) grey-level difference are calculated. A sum over all pixel-wise differences is accumulated or averaged over, which obtains a measure that is based on the sum-of-squared-differences (SSD). Mean-of-squared-differences (MSD) is merely the case where the differences are averaged over, rather than summed up. Other variants include the case where differences are not raised to the power of two (squared).This method is usually powerful if the two images compared are closely aligned and their intensity values are relatively continuous and low in contrast. In many cases, MSD/SSD will tolerate a low level of locally-situated difference, while contrariwise, MI and NMI properly handle sparse dispersion of pixels in some localised region.

Suggestions have been made over the years with regards to the issue of speeding up similarity measures. The above measures depend heavily, from an efficiency point-of-view, on the dimensions of an image. A multi-resolution approach, for example, can be used to speed up the entire process. Blurring or averaging, followed by re-sampling or sub-sampling, allows for images of smaller size to be manipulated and complexity to be quadratically lessened. As the similarity measures are proportional to the images size, far better performance can be achieved by a transition from coarse to finer resolution. Pluim *et al.* [55] studied the effects that this approach will have on the measurement of similarity.

2.4 Groupwise versus Pairwise Registration

A distinction is made between two approaches to tackling the NRR problem. Some take the approach wherein one image from a set is chosen as a reference (or template) and the remainder of the set is transformed to fit that reference. This means that, at the very end, all images will be assimilated to the particular reference, which was arbitrarily and possibly ill chosen. The other approach is based on the idea that images should be transformed to resemble the *entire* set, irrespective of an arbitrary choice of a reference image.

Debates over the validity (or lack thereof) – that which is inherent in the pairwise approach – are of great relevance to the work presented hereafter. The subjective choice of a reference image implies that the results are highly dependent upon this choice. This leaves room for ambiguity and bias, which motivates the need to evaluate the results of NRR algorithms, as well as become independent of any selections that make the problem non-deterministic.

2.5 Assessment of NRR

Two main approaches to assessing the accuracy of NRR algorithms have been described in the Introduction chapter – one based on the recovery of known deformation fields, the other based on measuring the overlap of ground-truth annotations after registration. Both approaches are valid, but neither is easy to apply routinely, and both are better suited to off-line

evaluation of algorithms, rather than *in-line* evaluation of the results of NRR in practical applications.

2.5.1 Recovery of Deformation Fields

One way to test the performance of a registration algorithm is to apply it to some *artificial* data where the correct correspondence is known. The correspondence is obtained by manual annotation, which can be refined by repeating the process, reducing or altogether annulling the effect of subjective errors. The STAPLE algorithm [87] from Warfield *et al*. addresses such problems. Estimation maximisation is used to account for a number of independent observations made by experts who annotate image labels (segmentation). No estimate can be considered to be an objective truth, so the distribution will be random. Making use of this observation was shown to lead to substantial improvement, as well as more pronounced and correct boundary edges. STAPLE exploits knowledge about uncertainty that is inherent in erroneous annotation where subjectivity prevails.

There is another approach that thrives on ground-truth data and artificial warps that get applied to it. Having obtained ground truth, degradation of this correspondence can be applied. Such test data is typically constructed by applying sets of known deformations (either spatial or textural) to real images. This artificially-deformed data is then registered, and evaluation is based on comparing the deformation fields recovered by the registration algorithm with those that were applied originally [60, 66].

This approach can be used to compare the performance of different NRR algorithms but, since it relies on the creation of artificial test data, cannot be applied in-line. Also, the validity of the approach depends on the ability to construct artificial deformations which mimic the variability found in real images of a given type, which is difficult to guarantee.

2.5.2 Overlap-Based NRR Assessment Methods

An alternative approach is based on measuring the alignment [36], or overlap [36, 60] of anatomical structures annotated by an expert, or obtained as a result of (semi-)automated segmentation. This has the disadvantage that manual annotation is expensive to obtain and prone to subjective error, whilst reliable automated or semi-automated segmentation is extremely difficult to achieve – indeed if it was available it would often obviate the need for NRR.

In later chapters, an overlap-based approach is used to provide a 'gold standard' method of assessment. The method requires manual annotation of each image – providing an anatomical/tissue label for each voxel – and measures the overlap of corresponding labels following registration, using a generalisation of Tanimoto's overlap coefficient [4]. Each label for a given image is represented using a binary image but, after warping and interpolation into a common reference frame, based on the results of NRR, a set of fuzzy label images is we obtained. These are combined in a generalised overlap score [9] which provides a single figure of merit aggregated over all labels and all images in the set:

$$\mathcal{O} = \frac{\sum_{\text{pairs},k} \sum_{\text{labels},l} \alpha_l \sum_{\text{voxels},i} MIN(A_{kli}, B_{kli})}{\sum_{\text{pairs},k} \sum_{\text{labels},l} \alpha_l \sum_{\text{voxels},i} MAX(A_{kli}, B_{kli})} \tag{2.2}$$

where i indexes voxels in the registered images, l indexes the labels and k indexes image pairs (all permutations are considered). A_{kli} and B_{kli} represent voxel label values for a pair of registered images and are in the range $[0, 1]$. The $MIN()$ and $MAX()$ operators are standard results for the intersection and union of fuzzy sets. This generalised overlap measures the consistency with which each set of labels partitions the image volume. The standard error in $\mathcal{O}$ can be estimated in the normal way from the standard deviation of the pairwise overlaps.

The parameter α_l affects the relative weighting of different labels. With $\alpha_l = 1$, label contributions are implicitly volume-weighted with respect to one another. This means that large structures contribute more to the overall measure. Later chapter also consider the case where α_l weights labels by the inverse of their volume (which makes the relative weighting of different labels equal), where α_l weights labels by the inverse of their volume squared (which gives regions of smaller volume higher weighting), and where α_l weights labels by their complexity, which is defined as the mean absolute voxel intensity gradient over the labelled region.

An overlap score based on a generalisation of the popular Dice Similarity Coefficient (DSC) would also be possible but, since DSC is related monotonically to the Tanimoto Coefficient (TC) by DSC = 2TC/(TC+1) [67] it was not considered further.

Chapter 3

Models

"If you optimize everything, you will always be unhappy."

– *Donald Knuth.*

THIS chapter contains a gentle introduction to models of shape and appearance – those which are being evaluated at the core of the work. The approach to ground-truth-free evaluation of NRR depends on the ability, given a set of registered images, to construct a generative statistical model of appearance. The approach of Cootes et al [12, 24], who introduced models that capture variation in both shape and texture (in the graphics sense), was adopted. These models have been used extensively in medical image analysis. They assist interpretation tasks in, for example, brain morphometry and cardiac time-series analysis [26, 61, 69].

Other approaches to appearance modelling could also be considered as, in this application, one relies only on the generative property of such mod-

els. Nevertheless, the focus in this chapter remains appearance models of shape and intesity, which are dealt with almost exclusively.

3.1 Statistical Models

3.1.1 The Top-down Approach

This section explains the motivation and nature of active appearance models as an image analysis method. Image analysis is a broad and generic problem that can be tackled in various ways. This analysis is fundamental and essential to many routine tasks such as industrial inspection, motion analysis [90], face recognition [20], and medical image understanding [70]. What makes this problem intrinsically laborious is one's inability to take into account single pixels independently and study the structures that they form together, cohesively. The goal of analysis and interpretation is not only to tackle such problems properly, but also to do so efficiently. This needs to be done in a manner that is not excessively affected by the size of the image, i.e. it ought to be scalable.

Analysis involves *measurements* of meaningful structures in an image, as well as explanations pertaining the *form* of these structures. In order to extract and study information from particular meaningful structures, image *segmentation* needs to precede. Segmentation is concerned with the identification of several regions of interest, which may be characterised as belonging to the same object. By sub-dividing the image into such

Figure 3.1: A target image **T** (greyscale bacground) is being overlaid with a high-level representation (the model **M**), which seeks a good fit in the target image by transforming itself.

regions, understanding of the nature of its constituent components can more intuitively be gained.

With models, one concentrates on abstraction and adopts a top-down approach to analysis of images. The approach relies on high-level knowledge about the visual attributes of one specific structure. Alternatively – and often more usefully – this abstraction can represent and embody a *collection* of structures that *together* form another aggregated structure. The reason why such an approach is referred to as a top-down approach is that it contains existing information which it attempts to *fit* to the problem posed[1]. It makes assumptions about the problem and is, in some sense, taking a preliminary, hypothesised overview on the structures in an image, as Figure 3.1 illustrates.

[1]A bottom-up approach considers low-level data and builds up towards knowledge of greater complexity which has a more substantial meaning. Top-down is the opposite approach where a model is aware of what it attempts to find, so it searches for a best match at lower levels.

3.1.2 Rationale

In many cases, image analysis tasks are better handled by a top-down approach. Not only is a top-down approach capable of gathering information about an image at hand, but in the case of model-based analysis, it is also capable of generating new images. The importance of this property lies in the fact that infinitely many unseen images can be derived from the model. These images can be used in various ways in a practical setting. One application, for example, involves the fitting to data, as explained at the end of this chapter. The model is deformed to resemble an image and, from this deformed model, images can be learned. Beyond this popular class of applications, worth mentioning is the fact that the thesis introduces a new application for models. On the one hand, models can be evaluated, whilst in practice, they can also be used to assess the quality of NRR.

3.2 Shape Models

Given a collection of images depicting an object which possesses common properties, it is possible to model the visual form (or *shape*) of that object. This model can be built in a way that makes it independent from subtle changes in view-point, object position, size etc. Such a model can be made robust to moderate levels of object deformation, too. The object which appears in the group of images need not be the exact same object; it can be an object belonging to one common *class*. Variation that is

typical for that class can be handled, and essentially be understood, as well. Learning of this variation can be done quite reliably with the help of elementary transformations. Such transformations were described in section 2.2 on page 35, but their functionality is limited and constrained. They are merely the means by which segmented images (or shapes) are forced into a state of alignment. The process of alignment says something about the variation in shape.

There are statistical methods that facilitate the encoding of the variability, which is *learned* during a so-called *training process*. That training process does not require more than an exhaustive pass through the set of images where objects appear, then probing the distribution of points. However, in order to interpret a large set of objects, several simplification steps are required. This results from the fact that most images where objects lie are relatively large-scale in practice. These are large enough to result in an exponential blow-up[2]. Reduction of the dimensionality of the data is thus needed.

3.2.1 Deformable Models

A method is sought which reduces the amount of information that is required to describe an object of interest. As well as describing that object, this method should consider the different forms this object may take – particular those which deem both suitable and valid. In practice, this is essentially achieved by selecting points of interest which lie within the

[2]Currently, model-based methods typically deal with only the order of tens of thousands of pixels. High-resolution medical images can contain millions of pixels/voxels.

image, preferably ones which are a representative sub-set of the whole image. Points need to be picked so that they jointly encapsulate knowledge about the object of interest. In most cases, edge detection is sufficient for capturing and selecting regions or points of greater significance in the image. Edges and corners tend to hold more information of value for subsequent analysis (e.g. segmentation). These points lead to better identification of the different objects residing in the image. Such points become what is referred to as *landmarks*.

Landmarks are positions in the image which can effectively distinguish one object from another object in a similar image. They are represented and encoded as a set of points in a set of images (see Figure 3.2 on the next page). They possess interesting spatial traits and they can be used to render curves (or contours), which together make up complete *shapes*. The concatenation of the coordinates of these landmarks describes an image (or rather the object being dealt with) using a more concise representation, which is lossy[3]. As an example, in 2-D, for n landmarks, a vector of size $2n$ can faithfully describe the shape of the object present in an image. This object can be encoded as follows.

$$(x_1, y_1, x_2, y_2, ..., x_n, y_n) \Rightarrow \mathbf{S} \tag{3.1}$$

where $\mathbf{S}$ is simply a discrete reconstruction of the shape in the image. This does *not* embody the actual image, but only key characteristics of it. This proves to be sufficient for good approximations to be made.

[3]The loss is a necessary evil when the complexity of the data is reduced. The greater the number of landmark points, the higher the fidelity (as well as complexity).

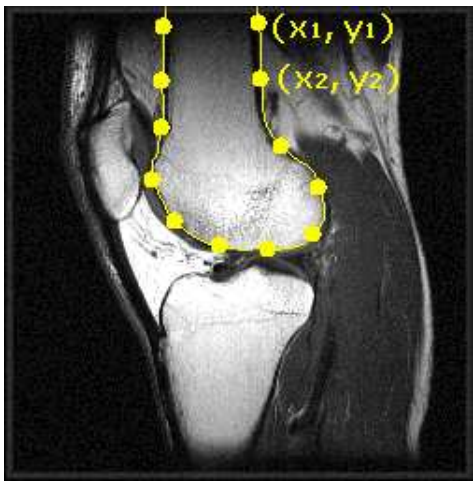

Figure 3.2: Landmark identification and mark-up in medical images

3.2.2 Correspondence

Chapter 2 alluded the the notion of image correspondence. The use of the term, in that context, referred to a dense, pixel-to-pixel correspondence. From here onward, when the notion of models and landmark points is discussed, the term "correspondence" deals with correspondence across landmark points, rather than that which involves all pixels contained in the images.

3.2.3 Principal Component Analysis (PCA)

One can perceive and visualise the images dealt with as points in a high-dimensional space, as was earlier suggested. By placing all images in that space, it is expected that some cloud of points will be present at a specific, albeit somewhat confined, region. The breadth of this region (or the size of that cloud) will depend on the variation amongst the images (or more generally – data) which is being visualised.

PCA is a method which relies on Eigen analysis. In essence, it obtains the Eigen-vectors and Eigen-values of a cloud of points, decomposing it into a set of vectors with their magnitude. The highest Eigen-value corresponds to the most significant Eigen-vector (see the single-headed arrow in Figure 3.3). It symbolises the direction which best distinguishes the image data. As as arrow, it is expected to be the longest one too – that is – the one whose magnitude is the greatest[4]. This vector is considered to be the principal component which describes that data. it serves as the most effective discriminant.

In a recursive manner, and at each stage of the process, the current principal component is being studied (e.g. incorporated in a model's covariance matrix) and then set aside, by being reduced/annulled. The process is repeated until only negligible components (dimensions or vectors) remain present. As a whole, it is a progressive dimensionality reduction routine where data dimensionality us reduced at each stage, until only dimensions which contain noise are left. At each stage, this recursive function will therefore deal with simpler, denser, and more uniformly-distributed data. More and more principal components are set aside, leaving in tact data of lower dimensionality that occupies a relatively low volume in space.

A smaller number of components can then be used to express the variation up to a comparatively high level of accuracy. The process is lossy, much like any of the other stages in model construction, including the

[4]If one thinks of the cloud in n dimensional space as a placement of characteristics $(c_1, c_2 ... c_n)$, the principal component is one characteristic which best separates instances of the data. It picks the largest range of variation and uses it in decomposition.

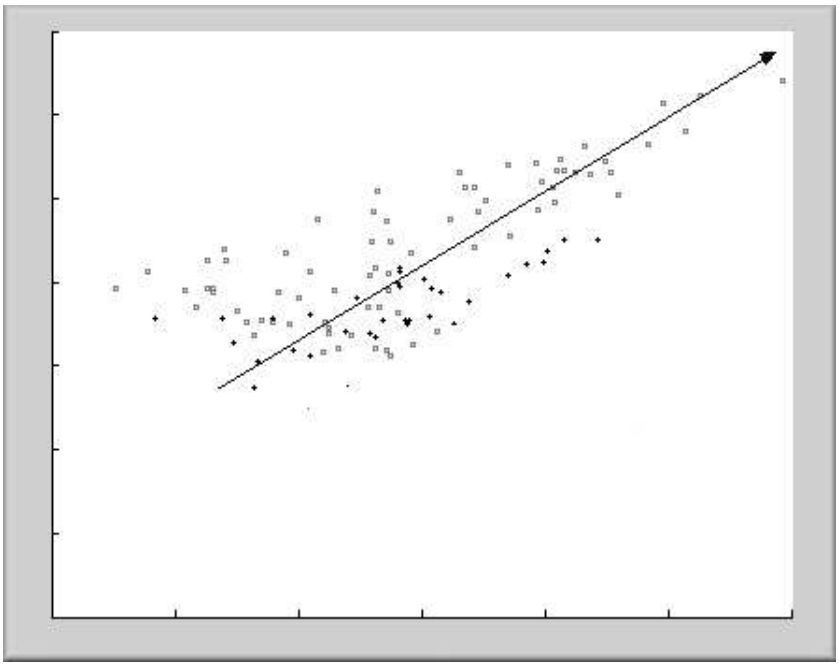

Figure 3.3: The principal component in a 2-D data scatter is indicated by the arrow

choice of a finite number of landmarks. That loss is being controlled in the sense that one can choose the minimal amount of variation which must be accounted for[5]. PCA is used to gain speed while retaining the best descriptors of variation or difference in shape and intensity. What this boils down to is the building of a model that is smaller in size and is easier to deal with. It is easier to deal with because: **(1)** it is smaller; **(2)** it is quicker, e.g. to reach convergence and **(3)** some of its key attributes have been decomposed, which may be useful in learning.

[5]A sensible choice might be, for example, 98% of the observed variation, which means that 2% of the variation is not accounted for. In practice, that 2% of the overall variation is usually the least informative and it is possibly made up from noise and error. Annulling this effect is, among other things, what PCA is intended to accomplish.

3.2.4 Model Construction

An integral part of any appearance model is its construction. This initial formation step defines what the model encapsulates. Construction also affects the validity and quality of the model.

The construction process can be broken down into various steps. The first step is concerned with the establishment of a model that is not only 'acquainted with' the *mean* form of some object (if not the image as a whole) in a set of images, but also the variation that can be applied to that mean in order to create new instances of that object. This model dictates which values several vectors can take. Each of these vectors can be translated into a visual form, i.e. image pixels which the vector affects.

More desirable models should never be excessively bendable. They should be generic, flexible and permissive, but remain strict and confined at the same time. These models should accept as valid more reasonable variations of the object under investigation. One property is referred to as Specificity as it forces the model to remain specific. Conversely, and in a complementary manner, the model should properly represent a large set of images and be general. We refer to this property as Generalisation and elaborate on it in Chapter 6.

There is a convenient mathematical method for expressing variation. It also happens to be linear, which has its pros and cons. The method involves assigning a parameter to each mode of variation. When change to these parameters is applied, the mean image is deformed accordingly and there will be a direct effect on its appearance. Rather usefully, each

valid image can be uniquely and fully described by the parameters which were used to generate it from the model. The synthetic appearance and its vector representations are equivalent and inter-changeable. What follows explains how models are being constructed, in technical terms.

3.2.5 Shape Models

Shape models are a simplified version of full appearance models. Chronologically, shape models precede appearance models as appearance is dependent on shape. Models of appearance essentially extend shape by adding dense intensity information.

To encode the shape of an object, landmarks need to be identified and statistical analysis applied to them. A vector of landmarks is formed which expresses spatial properties, namely a series of landmark coordinates. From a simple analysis, a mean shape is obtained and it can be denoted by $\mathbf{x}_{mean}$ or $\overline{\mathbf{x}}$. To arrive at this mean, the method most commonly used is Procrustes analysis. The generalised Procrustes procedure (or GPA for Generalised Procrustes Analysis) was developed by Gower in 1975 and has been adapted for shape analysis by Goodall in 1991. It studies each component of the vectors derived from the images and returns for each component a value that is said to be the mean. From here onwards, this vector which represents the mean of the data will be referred to as $\overline{\mathbf{x}}$. Each shape $\mathbf{x}$ can thus be formulated as indicated below

$$\mathbf{x} = \overline{\mathbf{x}} + \mathbf{P}_s \mathbf{b}_s. \tag{3.2}$$

The matrix $\mathbf{P}$ represents the Eigen-vectors of the covariance matrix (set of orthogonal modes of variation) and the parameters $\mathbf{b}_s$ control the variation of the shape by altering modes of variation. The parameters essentially describe the magnitude of the covariance of each element in the matrix. These parameters and the range within which they lie describe a level of freedom – that is – the freedom (or contrariwise – constraints) of the model. Eigen-analysis is used in the derivation of the expression above, as was discussed in this chapter.

It is important to note that landmark points could be chosen *arbitrarily.* However, this results in poor models. It lead to serious issues as images need to be correspondent, i.e. points should aligned to exhibit on their spatial commonality. Identification of objects is in most cases done by drawing lines or selecting surfaces which surround these objects, to form segments. Given continuous elements such as a lines or surfaces, it is not obvious how to suitably sample from them in the form of points. The choice of points affects the quality of reconstruction, as measured by the assigned errors.

With the concise landmark-based representation (described in Equation 3.1) assumed to be the convention and a collection of decent-sized vectors rather than a large collection of images and pixels/voxels, it should be possible to express (in a feasible way) the legal range[6] of each one of the vector components. This, in essence, establishes the *model*. It is an entity that can be manipulated to reconstruct all the shapes (or as later explained – images) it originates from, and far beyond that. This

[6]The legal range can be thought of as the values a parameters may take. In reality, a Gaussian distribution usually fits the observed range rather well.

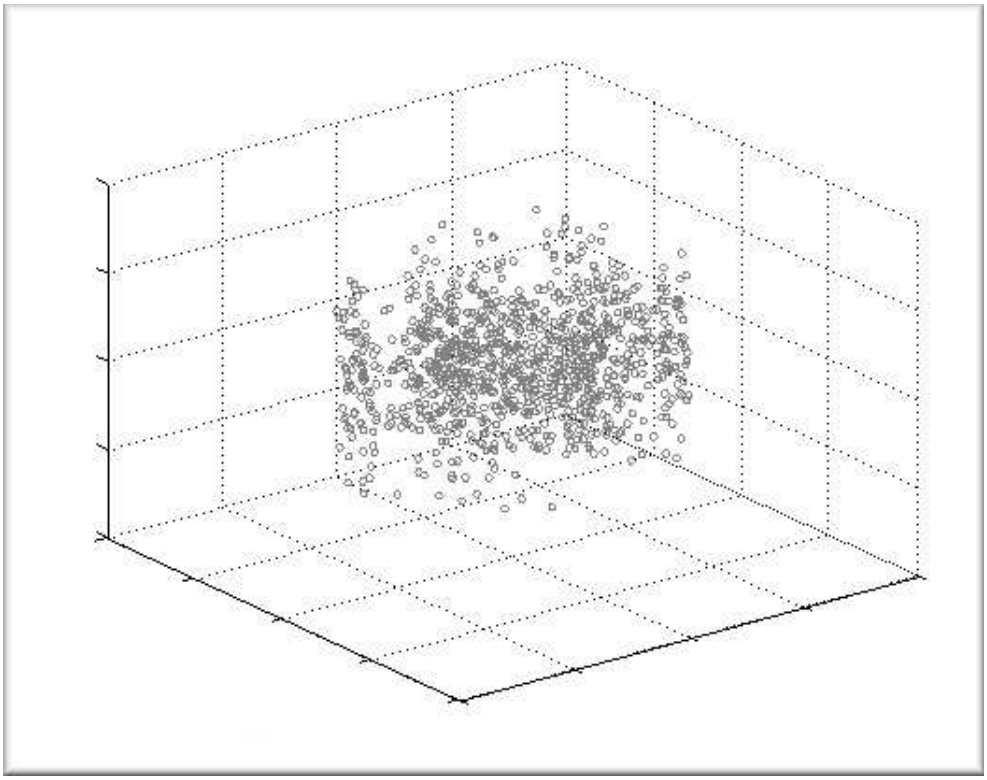

Figure 3.4: 3-D scatter of points, which illustrates data embedment in hyperspace

model encapsulates the variation which was learned from the data and it usually improves its performance as more legal examples are interpreted and 'digested' to support further training. Varying the parameters of the model can generate new (yet unseen) examples as long as that value variation is restricted by the legal range, as learned from the training examples. The vector representation mentioned beforehand can be also perceived as a description of a fixed location in space that comprises d dimensions (see illustrative scatter in Figure 3.4). This turns out to be a useful demonstrative idea as will be seen later when dimensionality reduction is applied.

To discuss caveats, shape models merely contain statistical information. They can be built from the images with overlaid landmark points identified and assembled. In order to make such a modeling approach possible, it is vital to seek consistency amongst the coordinates of all landmarks. This means that all points need to be projected onto a common space – a process whose purpose is to ease collective analysis. That process can also be thought of as an alignment step which somehow links to the next chapter. More issues that are concerned with normalisation, projection and the like are described in slightly more detail later in this thesis.

A human expert usually performs annotation or landmarking of the images with the aid of some computerised special-purpose tools. In recent years, alternatives which are automatic showed great promise [16] and they were also extended to 3-D [18]. The later chapter on MDL shape models is dedicated purely to that one strand of work which is so fundamental to the newly-proposed method.

3.3 Appearance Models

Appearance models were developed by Edwards *et al.* [23, 11]. Their greatest contribution, advantage, and essence lie within the fact that they contain grey-level data, rather than just shape-specific data. Incorporation of full colour was made possible as well (e.g. Stegmann *et al.* [69]) since colour can be simply thought of as an extension of the single grey-scale band being divided up into the most common separability: red,

green and blue components [7].

Appearance models contain intensity[8] information that is extracted from images by interpolating between landmark points. As a result, appearance models contain information about what an image *looks* like rather than just its *form,* as visualised by contours (or surfaces in 3-D). Grey-level values (also referred to as *intensity* or *texture*) could be systematically extracted from a normalised image and be stored in an intensity vector for subsequent steps of the algorithm. The normalisation process and representation of this intensity vector will be outlined later in this section.

What enables synthesis from appearance models to possess great resemblance to reality is the fact that, at the later stages of the construction process, a *combined* model is made available and it produces dense pixels rather than meshes or contours. The linear model incorporates *both* shape and intensity and it expresses the way in which a change in intensity affects shape, and visa versa (e.g. how an expansion results in darkening of an image region). The model studies notion of the *correlation* between the two – a notion that is dependent on the training data and principal component analysis algorithm chosen. Although appearance models are not as simple and fast to build as shape models, they contain all the information that is incorporated in shape models and, in that sense, are a superset of shape models.

Some techniques have been developed and employed to speed up the

[7]There are different possible colour schemes [53], but they need not have any affect on principles of sampling intensities.

[8]Patterns of intensities form texture.

matching of appearance models to image targets. Tasks such as the matching of an appearance model to some target image are described later in this chapter and are further illustrated in [10].

3.3.1 Intensity Models

The first stage in construction of an appearance model involves the sampling of texture. It assumed that a shape model was already obtained. Texture in this context is a patch of pixels intensities. In principle, having obtained a description of shape variation from a set of shapes, as well as their spatial correspondences, it is possible to identify homologous points in between these correspondences. This makes possible the approximations of the *denser* correspondence – that which involves larger, continuous parts of the image, rather than points only. The description below illustrates one possible way of sampling intensities. Construction of an intensity model is carried out in the exact same way as was done for shapes (Equation 3.2).

At this construction stage, each of the images, encoded as shape vectors, needs to be aligned to fit a bounding volume in space[9]. In practice, the properties of that space are implicitly defined by the mean shape[10]. Rigid (or Euclidean similarity) transformations, namely translation, scale and rotation, are rarely sufficient in warping all images/points into that common space. For example, in the case of human face recognition, different

[9] A normalisation step as such is similar to mapping onto a sphere, for instance.

[10] Oftentimes, the choice of the mean shape proves to be the least damaging choice, in terms of overall accuracy.

head sizes and facial expressions introduce great difficulties. Nonetheless, it is crucial that good fitting is obtained before the sampling of grey-levels.

Following these basic transformations which align all images, displaced control points on each image overlap and contain in between them shape-normalised patches. These patches are made available for construction of texture vectors. Barycentric arithmetics, renowned for their frequent utility in computer graphics and stereo vision, are used to identify the location of all corresponding points within a patch[11]. This location of point is directly affected by the warps applied to shift a given shape onto the space of the mean shape.

Triangle meshes are subsequently created by stretching lines between neighbouring control points and intensity values are captured one by one (along a chosen grid of points to be sampled) and are used to form a vector representative of texture. Each component in such a vector captures the intensity (or colour) of one single pixel, as was learned from the examples. Statistical analysis, which is not different from the former cases, results in a linear expression for texture

$$\mathbf{g} = \overline{\mathbf{g}} + \mathbf{P}_g \mathbf{b}_g. \tag{3.3}$$

g is the intensity and the other parameters are the same as for the shape.

[11]It is helpful to think of two different triangles and the relationship between points within these triangles. Centre of gravity (CoG) is used here to assign approximate correspondence.

The process is not different from dimensionality reduction in the case of shape.

The use of the algorithm above implies that, for small vectors (i.e. a low number of pixels sampled), coarse appearances will be easier to spot[12]. Objects will often appear to be nothing more than a collection of polygons that do not quite resemble realistic appearances[13]. To compensate for this, algorithms can be used for shading. In practical use, Geodesic interpolation is used and the results can be rather rich, considering the low dimensionality of the available data.

3.3.2 Combined Models

The models in Equations 3.2 and 3.3 take a linear form, so they are quite compact. This is a highly desirable property that makes the models flexible and manageable. The simplification is made possible owing to PCA, which reduces the size of vectors of shape and texture. As mentioned before, Eigen-analysis is involved in the process, but it has a few caveats. For example, in simpler cases, it assumes that none of the distributions is banana-shaped. This approach works more gracefully under the assumption that all distributions are normal. While this may be acceptable and plausible in practice, it adds a prerequisite to the method. There are other methods for decomposing data which resides in a high-dimensional space, but they will not be further explored.

[12]Analogically, in the case of shape, sharp-bended descriptors result from the low number of sample points.

[13]One of the main aims and great power of appearance models is full synthesis.

The two model components, $\mathbf{x}$ and $\mathbf{g}$ (the vectors above, which are a function in generative models), need to be merged in order to establish a new model that blends both types of variation. This expressive model accounts for both types of variability (shape and intensity) and holds within it the correlation between the two. It means that any variation in shape will affect intensity too, and vice versa.

The parameters $\mathbf{b}_s$ and $\mathbf{b}_g$ are aggregated to form a single column vector

$$\begin{Bmatrix} \mathbf{b}_s \\ \mathbf{b}_g \end{Bmatrix}. \tag{3.4}$$

The new vector is a simple concatenation of the two. However, since the values of intensity and shape can be quite different in terms of their nature and granularity, some weighting is needed to attain a state of equilibrium, under which both shape and intensity accrue and attain a sufficiently-noticeable effect and impact. It is a normalisation step. The danger is that if no weighing of any sort is applied, intensity values may supersede these of shape or vice versa. In less practical terms, if the extent of data values differs greatly, then spread of the points in space will be undesirably imbalanced. Thus, the components to be identified by PCA are not as beneficial as they otherwise would have been (very elongated distributions being one example). To use a 3-D analogy, if some values in the vectors are significantly greater than others, point vicinity takes a turn for the worse and the cloud might be flat instead of roughly spherical[14]. For a relatively spherical spread of data points (or those of almost

[14]As an example, intensity frequently takes values in the range $0..255$ whereas nor-

homogeneous variation), a greater number of large components will be available for selection by PCA. Consequently, the variation expressed by a fixed and constant number of principal components will be higher.

A weighing matrix that resolves the problem introduced above is by convention named $\mathbf{W}_s$[15]. The form in which coordinates get bound to $\mathbf{x}$ depends on the level of accuracy required, the image size and the number of dimensions, whereas for grey-level values, this form is dependent on the number of bits allocated per pixel[16]. With weighing in place, the aggregation takes the form

$$\begin{Bmatrix} \mathbf{W_s b}_s \\ \mathbf{b}_g \end{Bmatrix} \tag{3.5}$$

where $\mathbf{W}_s$ is chosen to minimise inconsistencies due to scale. Lastly, by applying another PCA stage to the aggregated data, the following combined model is obtained

$$\begin{aligned} \mathbf{x}_i &= \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{c}_i \\ \mathbf{g}_i &= \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c}_i \end{aligned} \quad . \tag{3.6}$$

The appearance (shape and brightness levels) is now purely controlled

malised shape coordinates lie between 0 and 1, so fractions such as $\frac{1}{255}$ can be used as coefficients. The two should then scale almost indifferently.

[15]The symbol s stands to *shape,* as by default this matrix scales the shape parameters only. It gives logically equivalent results to these of applying the factor $\mathbf{W}_g = \frac{1}{\mathbf{W}_s}$ to intensities.

[16]For colour it is common to use 24 bits and for grey-level just 8 bits. For more compact statistical appearance models, less than 8 bits (256 shades of grey) might suffice to achieve good results and, in medical imaging, 12 bits are nearly a standard in acquisition.

by the parameters $\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_n$ and there is no need to choose values for two 'families' of distinct parameters, as argued before. This combined model has the benefits of the dimensionality reduction performed, which is based on shape as well appearance. This means that it finally encompasses all the variation learned and the correlation between these two distinct components. Since PCA was applied, the number n of parameters $\mathbf{c}_i$ is expected to be smaller than (or in extremity – equal to) the number of parameters in $\mathbf{b}_s$ and $\mathbf{b}_g$ put together.

3.4 Active Models and Fitting

This section is concerned with search for model matches in images, for analysis purposes. This process is also known as model fitting. It is possible owing to the process characterised by active model *training*, which involves extending an appearance model. This aspect of the work explains and exemplifies how to *use* appearance models. They could be used in a variety of ways, but models can also be discarded once built if no special functionality is necessary, e.g. in the case of model-based registration (Chapter 4). Training will be dealt with first and fitting, which is a closely-related aspect, will be explained in the subsequent subsections.

3.4.1 Model Training

A statistical model is available for exploration at this stage, having built it from a set of training images. That model is a flexible deformable en-

tity [25] that can be used to describe the variation observed in the set of training images. It can also be used in a 'generative mode' to resemble instances of an object or a image[17] that resides in the range of the training set[18].

3.4.2 Model Fitting

The fitting of a model involves use of knowledge to analyse observations that have been made. To motivate model matching or fitting, one can argue that the previously-constructed model involved a learning (or training) process which should somehow be exploited. For it is now known what objects of some type look like, it is possible to recognise and capture new objects of the same type.

Models can be varied by changing values of their parameters. It is not obvious how one should deform the model to reach an appearance that is reasonably similar to a given image. It is a completely opposite and complementary problem that one who explores this model will be faced with: how can a model generate new image instances after similar existing images instances generated that one model? In some sense, an inverse operation is needed, so that the model can be used in the opposite direction to the means by which it was created. In practise, the task is

[17]The distinction between object or image is hard to make because the model can describe more than one valid independent object and usually represents only a partial region of the entire image. In a medical context, the term *atlas* fits well and it usually describes a single organ or anatomical structure.

[18]The word "range" is a terminological equivalent to the area which stretches in between the space of training set instances. It can be perceived as the space defined by a potentially-Gaussian distribution that makes up the training set.

not simple. The alteration of model values needs to be guided by minimisation that obtains the matching which is being sought. In an expectedly high-dimensional problems, the process is laborious, unless extra knowledge about this minimisation problem is provided and in advance. Such knowledg can then be utilised, e.g. level of tolerance for the optimiser to aim for.

3.4.3 Learning the Correlations

The way in which this problem can be solved involves learning how the parameters $\mathbf{c}_i$ affect the model[19] whilst compared to a standard target image – the image to which the model must be fit. Each parameter in $\mathbf{c}_i$ has an unequalled effect on different regions and aspects of the model, e.g. its size, intensities and so on. By changing the value of each such parameter and learning the change that is perceived in an image (using pixel-based comparison of some kind), a type of deformation index can be maintained. This index indicates which parameters should be changed and, if so, in what way and to what degree. The change is applied in order to approach good overlap between a model and some target image.

More formally, the algorithm works as follows. For the model parameters $\mathbf{c}_i$ where $1 < i < n$, a parameter change $\delta\mathbf{c}$ (where one parameter or more can be changed simultaneously) is applied to generate new shape and texture. $\delta\mathbf{c}$ expresses, in a vector-based representation, the offsets

[19]There are some more complex considerations as the model needs to be aligned properly (rigid transformation), as well as change its form.

that each of the original parameters $\mathbf{c}_i$ is subjected to. The exhaustive pixel-wise difference in intensity[20] is calculated in accordance with

$$\delta\mathbf{I} = \mathbf{I}_{model} - \mathbf{I}_{image} \tag{3.7}$$

to produce a new vector of intensities (the differences). This vector can also be converted to be made visual and demonstrate differences in a way that is interpretable to human observers. A simple measure of difference is used although this need not necessarily be the case. Sum-of-squares of the pixel differences is then used because larger quadratic differences will have a greater effect on the final measure and summation then only consists of positive values. For example, consider the values derived in Equation 3.9 and in 3.10. The former contains information about how the values of the vector in 3.8, and particularly their summed difference, get accentuated, whereas in the later case makes them almost negligible[21]. As an example to consider

$$\delta\mathbf{I} = sumofsquares(\{-1, 3, 5, 2, 6, -10, -1\}) \tag{3.8}$$

then becomes

$$\delta\mathbf{I} = sum(\{1, 9, 25, 4, 36, 100, 1\}) = 176 \tag{3.9}$$

[20]A simple raster scan that account for all pixels should clearly be fast under most contemporary computer architectures.

[21]This is similar to the need for a median measure, where average is sensitive to erratic values or salt-and-pepper noise.

as opposed to

$$\delta\mathbf{I} = sum(\{-1, 3, 5, 2, 6, -10, -1\}) = 4. \tag{3.10}$$

With this measure of intensity difference, relational information can be expressed between the parameter change and this difference as it appears in image space where a model is superimposed on some target. That information (merely a correlation) can be learned by using a pseudo-target image which is the model in its mean form. It can be used for basic comparison that says something about the model displacements and their corresponding effect[22].

This quantitative measure of difference obtained will indicate the approximate "goodness" of the parameter change (as perceived with the use of SSD or MSD) and not a more localised effect that the change has on the given image. This means that it will not necessarily be obvious what parts in the two entities (model and target) remain similar and which ones do not[23]. A type of a sequential data such as a vector is hence more useful as it retains the location of each computed difference value. Unsurprisingly, this also consumes far greater memory resources (and many vectors of this kind will in fact be necessary).

Under the premise that space is more expendable than time complexity,

[22]It is possible to learn the properties of rotation, as an exemple, by applying a rotation and looking at the difference between the resulting image and the original image. That is the main concept that this step is based upon, namely studying the relationship $Transformation \Longleftrightarrow Error$.

[23]The vector's distribution of values, i.e. positions with high absolute values, can address this question.

a vector of difference is calculated and the correlation can be formulated as follows.

$$\mathbf{c}_i \rightarrow \mathbf{c}_i + \delta\mathbf{c} \rightarrow \delta\mathbf{I} \tag{3.11}$$

This type of offset $\delta\mathbf{c}$, which was applied to the collection of parameters $\mathbf{c}_i$, is accompanied by a global change in intensity values across the image frame. This correlation can now be set aside and become accessible from an index as its size is proportional to the image size. Storage is dictated by the following relation:

$$\delta\mathbf{c} = \mathbf{A}\delta\mathbf{I} \tag{3.12}$$

where $\mathbf{A}$ is a matrix[24] that encapsulates the change in intensities due to the parameter/s change $\delta\mathbf{c}$. This is a matrix which is correspondent to an n-dimensional vector that expresses the change which was discovered off-line. It linearly defines (in a possibly high-dimensional space) the linear relation between change to the parameters and change to the intensities, or more precisely the *difference image*. It can be used to choose directions of change directly when performing a search and thereby avoid re-computation in a virtually recurring and almost identical problem.

The most fundamental (and perhaps even compact) algorithm will carry out the aforementioned steps for each of the modes of variation, as well as the linear geometrical transformations. This can be a very laborious

[24]The matrix $\mathbf{A}$ can be obtained using linear regression.

and cumbersome process, but it depends on the prescribed level of robustness . As subsequent stages illustrate, models that are not rich enough will fail to converge in difficult scenarios, a classic example of which is inappropriate initialisation.

The matrix **A** contains many numbers and the matrix forms a 'path-finding map' that guides exploration for good parameter changes; this property will be of great use when fitting the model to a target. In practice, such matrices are visualised by showing negative values as dark and positive one as increasingly brighter ones.

3.4.4 Target Matching

The final stage involves the use of the model above, as well as the correlations learned for that model, to do the fitting. It is possible to carry out a search which is driven by the calculated difference between the model and a given target image. In pragmatic terms, this means that fitting of the existing model will be improved until the model approximately overlaps the target[25]. The fitting is all done by changing the values of model parameters. The state of the model, having explored some space for a fit, holds in the form of parameter values some information about the target image. This information can be further analysed. One parameter in a model of faces, for example, could describe the vertical angle of given

[25]This process of fitting strives to converge to the global minimum (of difference measure). Realistically speaking, the model and the target never reach complete equivalence, namely the difference value of absolute 0. Even if the target was used to train the model, PCA would obscure the connection between the two, due to information loss.

faces. This is where the power of a statistical model lies – being able to describe something compound in a very compact form.

The search for model match is reliant on error (or conversely – similarity) measures which are repeatedly calculated after each attempted assignment in the model's parameter space. Having applied some change to the parameters, a new estimate of difference is obtained. Each such change in parameter values is primarily guided by the matrices described on page 77. These express the correlation between variation modes (the similarity transformations as well as modes of appearance change) and the intensity values which describe difference (or discrepancy in match).

The model, as shown in Figure 3.5 (or in Figure 3.1 on page 54), is initially placed somewhere inside the image frame, with reasonable proximity to its target. If the model is placed too far from its to-be target, there is a danger that it will be unable to converge to the target correctly. It will most likely get stuck in a local minimum (the global minimum being out of reach, as Subsection 4.3.3 explains). The reason why good initialisation is essential is that significantly large displacements are rarely learned off-line and the difference between the target and the model is quite meaningless unless there is at least some partial overlap or commonality.

The algorithm which is used to perform the search has the following general form:

- Place the appearance model $\mathbf{M}$ somewhere in the image, preferably at the centre where the target of interest (to be denoted by $\mathbf{I}$) is

likely to lie[26].

- With the appearance model in its current state and the static target, perform the following:

  - ✧ Calculate the differences between the model and the target. This can be done by synthesising **M** and calculating **M - I**.
  - ✧ Using the correlations learned off-line[27], set new values for the parameters c_i of **M**.
  - ✧ Compute the new difference measures between the model and the target (as previously).
    - ❑ Save the new state of the appearance model if the difference has been lowered, i.e. similarity is being approached.
    - ❑ If unsuccessful, re-adjust the value of model parameters, potentially with inclusion of a scaling coefficient $k = 1.5, 0.5, 0.25$ and so forth. This often achieves good results, although it is a heuristics-driven technique.

- Iterate while no stare of convergence has been reached and improvements are still observed at times.

More advanced methodologies and algorithms are used at present, but better clarity is achieved here by adhering to simplicity.

[26] Advanced knowledge about the problem is highly conductive at this stage. Otherwise, a bottom-up image analysis is a must.

[27] If these correlations are not available, guessing would be an alternative. It is important, however, to learn from the experience gained during this independent run of the program or else the optimisation would behave senselessly and lead to improvements being identified very slowly. General optimisers are assumed to make a good judgment as such.

Figure 3.5: Model and target fitting.

The technique of matching an appearance model to a target image can be depicted by a staged simulation or a large sequence of images resembling the one in Figure 3.5. Somewhat remarkably, only a few dozens of iterations are required in order to get good matching. This of course depends on the algorithm and the scale of the problem.

Chapter 4

MDL Shape Models

"All generalizations, with the possible exception of this one, are false."

– *Kurt Gödel.*

In the previous chapter, generative models of shape and intensity were described. Construction of such models is the factor that defines their quality as they are inherently based on statistics. Their dependence on landmark points – points that define correspondence – was also explained. This chapter explains how correspondences are exploited and improved in an iterative manner.

4.1 Shapes and Correspondence

Correspondences describe of how images in the set are related to one another. As they contribute information that is not present in the raw

images, they are essentially the 'glue' that assembles pertinent images and makes them an associative *set*. Poor models are built from poor point-to-point correspondence and, conversely, a good model is one which is built from data that is well aligned. A key observation to make is that models, by definition, are based on the existence of correspondences in images, or shapes. Manipulation of these correspondences affects the quality of the resultant model.

Different arguments can be made with regards to an arbitrary choice of landmark points in shape. These landmark points define a non-continuous, point-to-point correspondence. In the simpler case which is shapes, one needs to provide a precise and accurate set of landmarks that define a cross-shape relationship. Nevertheless, it is not sufficient only to determine how accurate the correspondence will be. It is also important to select a correspondence set that is representative of the shapes. The number of landmark points is, after all, finite. This means that sampling a set of corresponding point in one region while neglecting another leads to locally-optimised models which perform badly as a whole (globally). Herein we deal with two cases of discrepancy in correspondences. Firstly, the correspondence, if not dense, must be selected carefully. A selection of meaningless or arbitrary corresponding points, for example, leads to the construction of poor models. Secondly, there is the factor of accuracy in the identification of correspondences. This might not be an entirely objective task, so it needs to be assessed in one form or another.

This chapter concentrates on the case where increasingly-improved shape model are gradually being built. Correspondences across the set from

which they are derived get refined iteratively. The refinement is attained as better correspondences are chosen automatically, under a process called reparameterisation. Essentially, optimal models are built using an optimisation framework wherein better correspondences are sought. Correspondences are embedded in a high-parametric space that is explored by a general optimiser.

4.1.1 Landmark Selection

Past work by Kotcheff and Taylor [38] addressed the problem of building, progressively refining, and assessing shape models. The problem, which was tackled in part, involves the selection of good corresponding points on the curve of a shape. In practice, this is achieved by evaluating the choice of landmarks using a shape model – that which resulted from a choice of landmark points. There is a case is circularity here and an optimiser is intended to identify the point of balance, i.e. a state of optimality.

The determinant of the covariance matrix of the model is said to be a reasonable approximation of model quality. Poor models can be discerned from better ones since the product of mode variances should be small for assimilated shapes. That similarity is typically increased if the points on the curve of the shapes are correspondent. When PCA is applied to a poorly-correspondent shape data, incorrectness that prevails will increase the scale of the distribution. Consequently, the model built will be poorer or altogether invalid. It will contain more information than it needs to, due to errors in the correspondence.

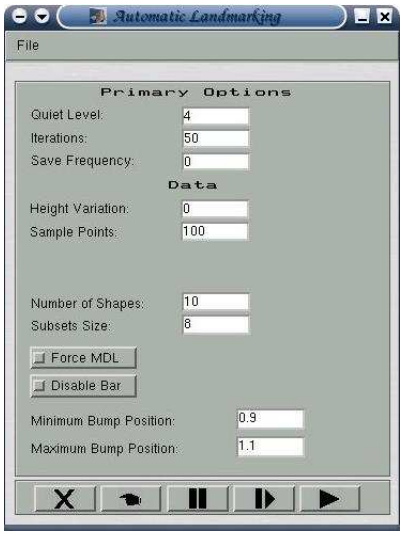

Figure 4.1: The graphical user interface for semi-automatic landmark selection.

4.1.2 Experimental Framework and Data

In most of the experiments that follow, a particular data type is used to make up a set of similar shapes. The data type is referred to as "brink and bump", owing to the elements from which it is composed (see Figure 4.2). A newly-constructed front end (shown in Figure 4.1) handled the process of landmark identification. Like most other applications that are used to carry out this work, the front end was constructed using Sun Microsystems' Java and Mathwork's MATLAB, under GNU/Linux. The new graphical user interface carries out the majority of the experiments described in the remainder of this chapter, as well as in subsequent chapters.

Figure 4.2 is intended to show not only the form of shape data which is being dealt with. It also depicts the results of using a given point-to-point

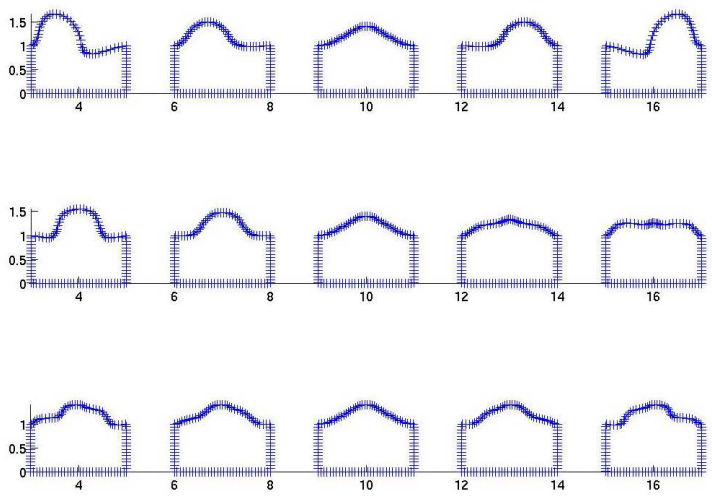

Figure 4.2: Unregistered bump-shaped synthetic data and its three principal modes of variation (± 2 standard deviations are shown). Each bump is composed of points, depicted as a plus sign, to make them distinguishable in the plots.

correspondence to construct a model of shape. Each point that is used to sample the shape (the dots are being connected using standard straight lines) is unique. Any point corresponds to another particular point in another shape. Since the figure shows shapes that were synthesised by the shape models, the original data will look very similar.

4.2 Learning Shapes

4.2.1 Principled Approach

A more recent approach, which is said to succeed the work of Kotcheff, made use of the minimum description length (MDL) principle [58] in shape modelling. An MDL-based criterion, rather than the determinant

of model variances, was used to evaluate the quality of shape models. Concepts from information theory (and particularly Shannon's entropy) were applied to the selection of preferable descriptors of shape [17].

The process of searching for better correspondences in shapes remains similar. A selection of points that describe a given shape is perpetually altered and their effect evaluated in order to find better shape models. In this case, a better model is said to be one that requires a more compact set of shapes to be passed as an encoded message[1]. The rationale is simple. Many shapes that are similar, based on the position of sample points, are more compressible. Many of them convey similar information, which is a trait that is perceived positively by a model, which becomes less complex.

4.2.2 Searching for Improved Model

Having defined a measure of model quality, there is an implicit measure of the quality of correspondences, too. The two are closely related, if not inherently the same. However, better correspondences ought to be found automatically.

In this particular context, a set of points serve as markers or descriptors of the outline of a shape. Each time points on the curve are selected, a different model is ultimately constructed. A good and compact statistical model is one whose variations are relatively small. The same may apply to the number of its control points. Such a model is found using a

[1] An alternative method involving B-fitting was proposed by Thacker *et al.* [72].

general optimiser, under which positions of corresponding points are altered, or the set of points reparameterised. Returning to NRR algorithms, MDL can be used as a similarity measure under an objective function that is iteratively evaluated for each reparameterisation of the points on the curve. The minimisation process is described in a reasonable level of details in Subsection 4.3.3 on optimisation. The more important part of this work is the use of an existing information-theoretic measure, namely MDL. It guides an autonomous search for good models. It is also possible to gain insight into the process by looking at its objective function.

4.3 Objective Function and Optimisation

4.3.1 Principles of Objective Functions

An objective function is an integral part of optimisation. It is responsible for solving the problem that it defines by stating a goal rather than the way it is achieved (which is where an optimiser fits in). The objective function returns a figure of merit for certain states/observations that are being probed, e.g. a given choice of landmark points that are embedded in a set of shapes. In the case of shape models assessment, the objective function will compute, for any given a choice of correspondences, how good the resultant model is. It is a function whose output value needs to be minimised[2], fundamentally by finding a set of values for its input

[2]Minimisation and maximisation are complementary and, in this case, a concise model has a low description length. In later chapters as well, model quality is consistently defined in a way which favours a model whose value is low.

parameters (e.g. sets of corresponding points). Such parameters are varied simultaneously in order for an optimal choice or an optimal solution to be picked from the many available choices. The greater the number of free (input) parameters, the more complex the function becomes and the longer it takes to solve it. This may be the general rule, albeit there are exceptions.

In the context of image registration, the objective function is most heavily based on similarity measures, as was briefly explained in the earlier chapter on NRR. However, there are more factors which can be taken into consideration. It is wise to enable this measure to be extended in some way. For example, it can be helpful to include the 'cost' of the warps that are used, so that the objective function is negatively affected by large warps. The reason why the cost of the warp is sometimes an integral part of the function is that complex warps are not as desirable as uncomplicated ones that perform the task equally well or even better. This cost is often considered a *regularisation term* which penalises sequences of warps [14] that form large trajectories in space. An optimiser, being a generic problem solver, will seek a solution that is simple rather than finding an odd trajectory in space that gives a similar solution. This may be case since the solutions are often not unique.

Objective functions are built to encapsulate in a concise and effective way everything that is repeatedly evaluated. They are therefore required to be a very efficient 'unit' (or black box) which will be invoked quite frequently. The *speed* of the registration will directly depend on the choice of an objective function that adds up results from warps, similarity cal-

culations and possibly more components. The *quality* of the registration will of course depend on this function, too.

To exemplify this with the use images, let two images $\mathbf{I}_m$ and $\mathbf{I}'_m$ be defined as the images before and after warping, respectively. Let a warping function $f_w(\mathbf{x})$ be defined as $f_w(\mathbf{I}_m, < parameters >) = \mathbf{I}'_m$, i.e. for a given set of parameters, the function will map the input image onto a new frame. For a similarity[3] function f_{sim}, the objective function then takes the form

$$f_{objective} = f_{sim}(f_w(\mathbf{I}_m, < params >), \mathbf{I}_r) + < reg - terms > . \tag{4.1}$$

where $f_{objective}$ is the objective function and $< reg - terms >$ is a regularisation term that takes account of the magnitude or severity of applied warps. The function's solver (an optimiser) then attempts to find a series of parameter values that will lead it to a globally-preferred solution. It does so by applying warps in the case of images or reparameterising along the curve in the case of shapes. The optimiser is exploring the space of the problem, as defined by the objective function at hand. More precisely, it attempts to find *assignments* for all parameters that describe the warps so that similarity is maximised (or difference minimised[4]).

This brief explanation about the objective function concludes and closes this section. It is intended to demonstrate the algorithmic approach that correspondence selection takes. Various algorithms (objective functions)

[3] It is assumed that for an objective function that needs to be minimised, the similarity measure will return small values for good similarity and vice versa.

[4] This function is said to minimise the sum of the difference between two images and another less significant term. The two images compared are the transformed image $\mathbf{I}'_m$ and the reference $\mathbf{I}_r$ in this case.

can be assessed by methods such as the one described by Warfield [86], or that which which is described in Chapter 7.

4.3.2 The MDL-based Objective Function

To recapitulate, objective functions define the means by which a solution is to be found. Efficiency remains a concern, so a sophisticated function that avoids constructing the model more frequently than necessary must be employed. The function used in this context needs to drive a search for shape correspondences using a suitable parameterisation (in the case of image registration – transformations which increase similarity across *all* images).

The nature of the problem and the methods of solving it convey the ulterior goal which is to minimise a description length of a model. The way this goal is achieved is different from the approach of most algorithms. Compared to the vast majority of methods to date, it takes a unique approach which is to use model encoding as a similarity measure. This relationship can be expressed using the formulation below.

In the case of image registration, consider a transformation function $W(\bullet, params)$. The construction of an appearance model can take the form $Model(\mathbf{x}_1, \mathbf{x}_2, .., \mathbf{x}_n)$ where $\mathbf{x}_i$ are the images used to train that model. One seeks a model that is more compact using the (simplified) function $F_{obj} = MDL(Model(\mathbf{x}_1 ..., \mathbf{x}_i .., \mathbf{x}_n)) - MDL(Model(\mathbf{x}_1 ..., W(\mathbf{x}_i, params) .., \mathbf{x}_n))$ where $params$ should be found to minimise this expression for each image vector $\mathbf{x}_i$. A succinct description of this algorithm follows.

- Repeat
  - ✧ For each image vector $\mathbf{x}_i$,
    - ❏ Optimise F_{obj} by altering the values of $params$.
- Until convergence.

In practice, to indirectly and quickly evaluate MDL, one can use an evaluate that is approximately related to MDL. What will then be calculated is $\sum_{i=1}^{n} log(\lambda_i)$ where $\lambda_{1<i<n}$ are the n Eigen-values of the covariance matrix whose magnitudes are the greatest. This is similar to the formulation of Kotcheff [38] where $\sum_{i=1}^{n} log(\lambda_i + \delta)$ is calculated to approximate

$$det(\mathbf{M} + \delta) \equiv \prod_{i=1}^{n} \lambda_i \propto \sum_{i=1}^{n} log(\lambda_i + \delta) \equiv log(det(\mathbf{M})) \tag{4.2}$$

where M is the covariance matrix under consideration.

4.3.3 Optimisation

4.3.4 Background

General optimisation is often used in the process of matching to or finding solutions. Optimisation complexity can be relatively high[5], so a cunning algorithm needs be devised. This process is, by convention, concerned with the minimisation[6] of the value of a function. That function most likely comprises more than a single variable, which makes it multi-dimensional.

A multitude of software packages that act as general optimisers exist and the way they operate and perform varies. Some even use a mixture of different algorithms depending on the stage of the optimisation and the changing granularity of the problem. Approximations and changes in granularity can lead to significant speed gains, as well as better search strategies.

Optimisation over a function that varies in many dimensions is a computationally-expensive process. Often this optimisation requires some *a priori* knowledge of the problem domain. Only by using some knowledge to devise *ad hoc* solutions can performance end up being satisfactory. In the case of image matching, advantages can be gained if the effect of variable alteration can be predicted in some way. An example of this was described in Section 3.4.2 on page 73 where pixel intensities have a dependency upon

[5]The behaviour of such a problem is not linear and it may cross over to the realms of quadratic programming (QP) where various parameters simultaneously control a function and minimisation is therefore by no means trivial.

[6]The complement is used to generalise it to become a maximisation problem.

a group of parameters. Given the difference between two or more images, or even some generic data which described *change* caused by value alternations in the objective function, it is possible to determine paths that lead to quick convergence.

For the problems at hand, common optimisation methods are gradient-descent and downhill simplex [56]. However, many other methods exist[7]. The advocated strategy would sometimes be a utilisation of mixtures of different methods with rational choice of the an algorithm at each stage. That may be needed because the different characteristics of the methods make them advantageous at different states throughout the optimisation process.

4.3.5 Problems

One of the flaws of existing optimisation methods is their inability to find a global minimum (or minima) reliably enough. In problems of very high-complexity, as in the case of model fitting in two or three dimensions, this can lead to shallow searches whose result is unacceptable. It is even more difficult to drive an optimisation without additional knowledge about the objective function and the problem. Assumptions about the behaviour of the curve along each of the axes[8] are otherwise made, based on observation. For example, one may assume that a face is located at the centre of the image and is upright.

[7] Among the popular methods: dynamic programming, genetic algorithms, Powell's, simulated annealing and steepest descent.

[8] Optimisation is a multi-dimensional problem that searches along hyper-spaces, some of which are orthogonal to the many existing axes.

The speed of the optimisation process can be improved at the expense of overall accuracy and error likelihood. If no exhaustive search[9] is carried out, there is then a danger of convergence at local minima. In most applications, any convergence at a local minimum would be highly undesirable although this may be better than a complete failure at identifying regionally-low points in the function. Local minima are a necessary evil for large, complex, and continuous functions.

In conclusion, there is a trade-off between speed and accuracy. However, accuracy can be achieved at a lower cost if more knowledge is acquired 'off-line', i.e. before the optimisation task actually begins. As expected, this also implies that many redundant computations will consume precious resources and time in order to train the optimiser.

4.4 Summary

This chapter has explained, but has not yet demonstrated, some of the advantages gained by using an MDL approach for choosing landmark points in a set of shapes. The notion of an objective function function was explained, as well the idea of using an information-theoretic objective function. Once this function is in place, there are various issues that are concerned with the optimisation regime. The way by which good solution are sought is rather crucial. Later chapters which review experiments elaborate further, using practical examples.

[9]Exhaustive search is impossible for continuous functions, but digital images take discrete values.

Chapter 5

Model-based Registration

"As order exponentially increases, time exponentially speeds up."

– *Ray Kurzweil.*

REGITRATION is the missing link which makes possible the construction of models without human intervention, i.e. without any interaction that involves annotation of data. This chapter outlines an approach for constructing appearance models using the images alone, without requiring any additional markup. The registration process is used to produce *dense* markup automatically and it serves the needed correspondences for the process of model building. This exemplifies the reciprocity and tight relationship between the task of registration – that which establishes dense correspondence – and models that utilise this correspondence.

5.1 Overview

An important component of the proposed framework, as described in subsequent chapters, is the construction of appearance models from a given set of correspondence *automatically*. The chapter begins by explaining a simplified registration framework that brings one dimensional data into a state of alignment. This is followed by more detailed explanation about the way alignment is used to construct models of appearance directly. Throughout this section, an approach is described for automatic construction of appearance models using a criterion of *complexity*, but one could in principle replace this with other criteria, as indicated in benchmarks that appear later. Technical details described herein frequently refer back to earlier work [16, 17] that backed the idea of building *shape* models, which are progressively refined, by assessing their complexity. Current work is distinct owing to the inclusion of intensity data in the model. Finally, a method for evaluating such models is described in greater depth.

5.2 Warps

As discussed in Chapter 2, a fundamental part of any NRR algorithm are means of transformation. As Chapter 2 adhered to a broader perspective – a perspective along the lines of general ideas – there needs to be a more elaborate explanation of the types of warps used at this stage, as well as the remaining stages.

Clamped-plated splines [48] are invertible, diffeomorphic warps that are

used here almost exclusively. The notion of diffeomorphism (see Subsection 2.2.3) was introduced to describe functions that map a group of pixels onto new positions without collision or ambiguity (notably the effect of tearing of folding). Diffeomorphism is by all means an important attribute for any type of warp to possess. Moreover, due to practical considerations, the warps used should also be computationally inexpensive.

The warps currently in use affect a rounded confined region (this extends to elliptic-spherical region in 3-D) and they can be wholly characterised by their location, radius, and magnitude. These warps are parameterised by their horizontal and vertical location, while magnitude and radius are simple pertinent values. To transform images, many such warps are placed and applied at different locations and scales to one image at a time. Their position is chosen randomly, by drawing number from a Gaussian distribution. Good results are carried on to subsequent iterations while bad ones get discarded. Towards the later stages of the algorithm, only small local warps, much as in the case of reparameterisation in shape, will entail constructive results, i.e. an increased inter-image similarity.

5.3 Using Models as a Similarity Measure

An objective function describes the framework which is used to register images. The objective function described in this section comprises warps based on biharmonic clamped-plate splines, as well as an implicit similarity measure, which is an approximation of the quality of a model. Any

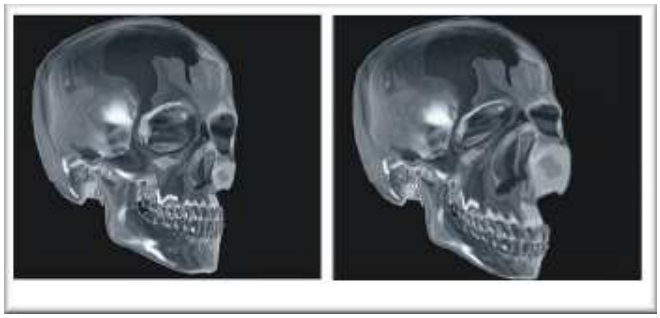

Figure 5.1: An arbitrary warp applied to image. On the left: image before warp is applied; On the right: image after warping.

collection of unregistered images can be used to establish such a model, but only a properly-registered set of images (that which results in high groupwise similarity) builds a good model. This observation can be exploited to create a similarity measure that not only deals with *pairs* of images, but can also deal with large *sets*.

5.3.1 The Registration Algorithm

This section presents the model-based objective function in the form of pseudo-code. Initially, the algorithm is demonstrated using simplistic one-dimensional data. The algorithm can be conceptually divided into two parts as follows:

Initialisation

- Generate images or retrieve them from a file.
- Optionally, apply image smoothing.

- Choose image reference. By default, the image closest[1] to the mean of all images gets selected.

Main Loop

- For a predefined number of iterations in the registration :
  - ✧ Set the level of precision for the optimiser to reach. Ideally it should increased (or tolerance lowered) when advancements toward the goal made are small, i.e. when registration is approached.
  - ✧ Repeat for all images:
    - ❑ If the current image is not a reference[2]:
      - ▷ Set up the positions of knot-points (random placements drawn from a Gaussian distribution are typically chosen).
      - ▷ Given the knot-points positions, apply warps to the current image and seek the warp parameters which minimise the cost $f(x)$, where x is the complexity of the model built from the entire set of data.
    - ❑ end if
  - ✧ end repeat
- end for

[1]Proximity is calculated based on the Euclidean distance, which identified the references that is, on average, nearest to all other images.

[2]The reference remain static in this case.

- Statistics and registration logging take place.

At the core of this objective function lies an evaluator of the complexity of the model. How this value gets computed is the core idea which is discussed further in this section.

5.3.2 Algorithm Visualised

Figure 5.2 shows the process that is outlined above. The framework is demonstrated in a simplified form in both cases (schematically and algorithmically). A reference image, as seen at the top of the figure, remains unaffected while all other images are manipulated in the way which is described in Figure 5.3. These are used to construct a model from which a certain complexity measures can be derived, e.g. description length. Based on that measures of complexity, subsequent warps are applied to the group of images.

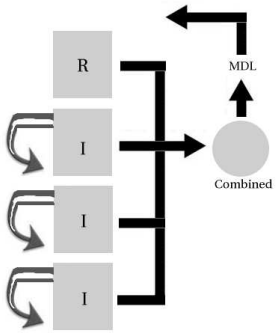

Figure 5.2: Schematic of the registration algorithm. A reference image (R) and the remainder of the warped set of images (I) form a combined model (circle) which is evaluated in an MDL-like manner to refine the subsequent warps.

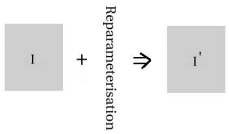

Figure 5.3: Current algorithm at a lower level. The idea of a reparameterisation is shown by emphasising that images are formed by aggregation of the previous image with some parameterisation.

5.3.3 The Data

The 1-D data which is dealt with hereafter is a simple elliptic bump. It varies in 3 distinct yet related ways, as shown in Figure 5.4. The data is being perturbed by varying the positions of the points that have it sampled (y-values), as shown in Figure 5.5.

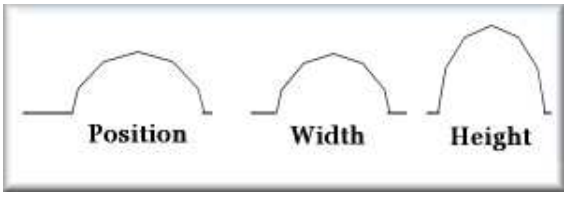

Figure 5.4: Illustration of the three variation modes.

Figure 5.6 shows 6 different examples of what the data appears like in its most simplistic form, i.e. when only 2 points (the edges) are used to sample it. In practice, however, one can deal with the bump as though it is a vector of image intensities. Dealing with several such bumps (1-D images) in turn, we can visualise them as shown in Figures 5.7 and 5.8.

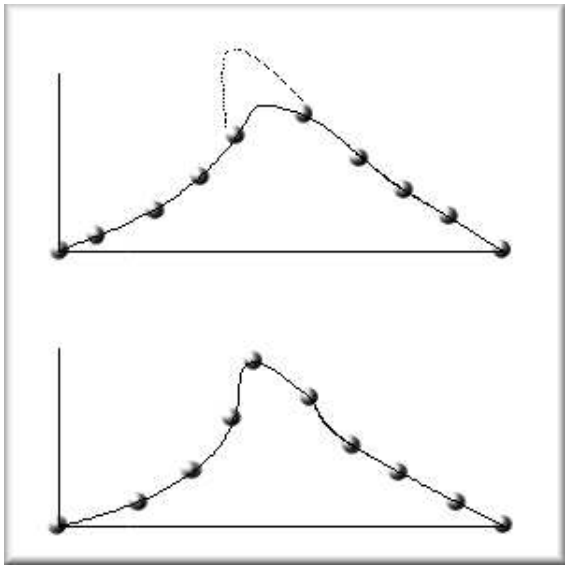

Figure 5.5: Movement of sample points and resampling of the curve that connects the points.

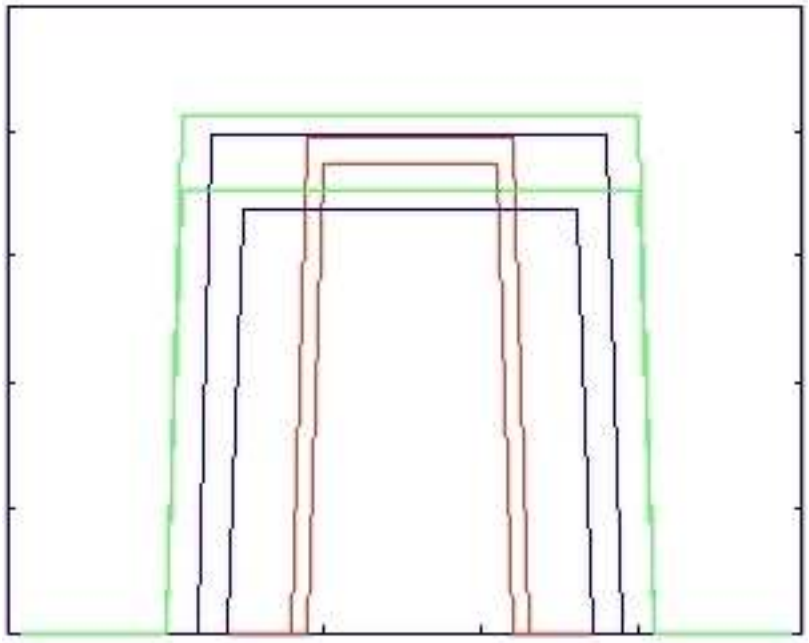

Figure 5.6: An simplified set of bump data. Different instances are indicated by distinct colours (or shades).

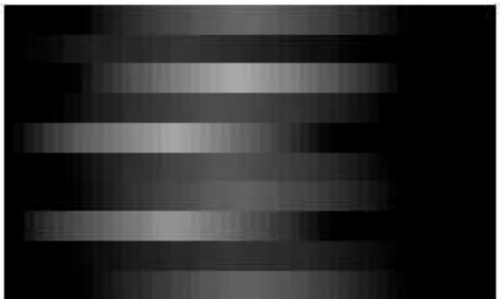

Figure 5.7: Data being registered. The registration process is visualised by an image composed of data vectors. The columns are 1-D vectors interpreted as grey-scale pixels.

Figure 5.8: A larger example of pixel representation for 1-D bump data.

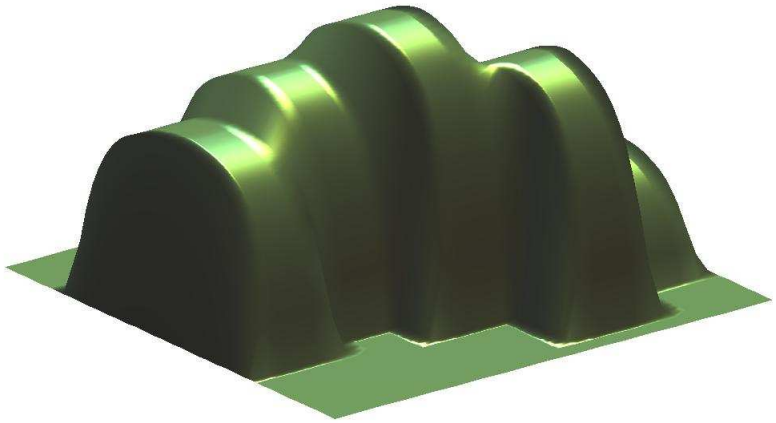

Figure 5.9: Original dataset depicted in 3-D. A set of size 5 is shown before application of any warps, which are intended to align the data.

The first step taken by the registration program is the generation of some random 'bumps'. These bumps varied in their height and width; the step size of the bump (the steep ends of the flat pinnacle) was fixed, i.e. the bump was initially flat at the top.

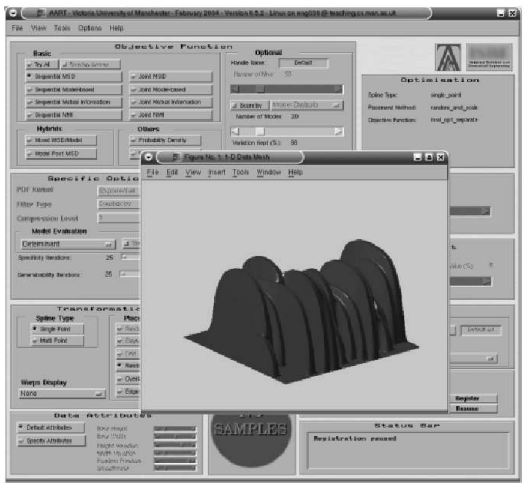

Figure 5.10: Autonomous Appearance-based Registration (AART): the program built to handle registration and model building.

Although the property of height was not intended to be ignored during registration (its signal is the intensity, which cannot be discarded), it was expected that it would remain unchanged due to CPS-based warps being perfectly diffeomorphic[3]. The bumps were all symmetric and the height took one value from the set {*hi, low*} where $lo = 0$ and $0.7 < hi < 1$. The data was therefore far simpler than any 1-D data which is not constrained in any way. The height of the bump and the position at which the bump goes high could conjointly define that bump so two real numbers (a tuple) at the minimum would suffice to reconstruct each bump.

[3]While the bump may have its form tweaked and manipulated, its highest peak should be preserved although it may move leftward or rightward. This assumes that boundaries for the warp intensity are honoured.

5.3.4 Early Experiments

Early work on the model-based objective function was fruitful. Images of the bump were registered using a variety of registration methods, including the model-based one. The existing algorithm was well-behaved. The problem, however, was not thoroughly understood as warps were permitted to degrade parts of the data. What follows are some exemplary experiments which, together with correspondent videos (see accompanying CD-ROM) demonstrate the process of registration visually, with progression. Below are perils and solutions to problems encountered on the path a working algorithm.

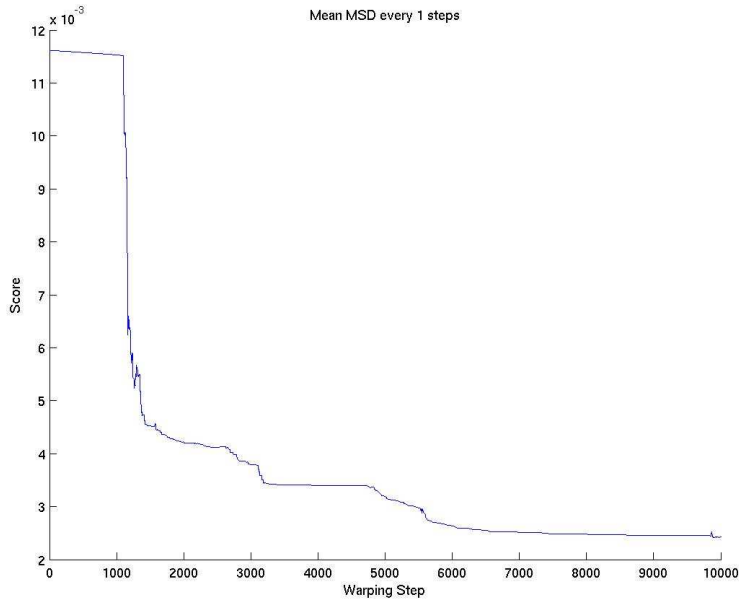

Figure 5.11: Mean MSD measures at each point during the model-based registration of 10 data instances.

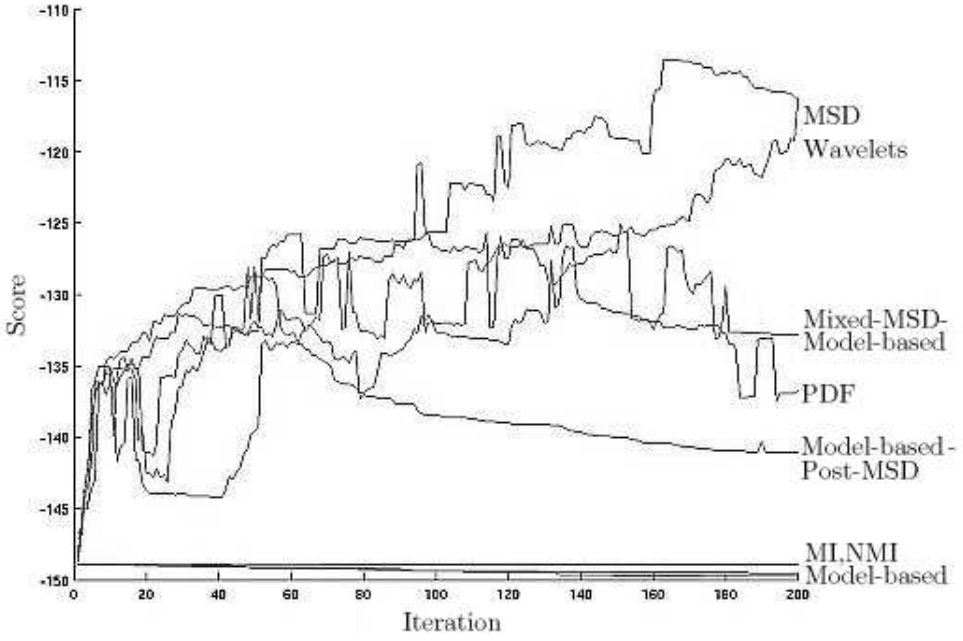

Figure 5.12: A comparative analysis of different objective functions. It illustrates that the model complexity decreases only for the newly-proposed objective functions. The Y-Axis value is an indicator of model compactness.

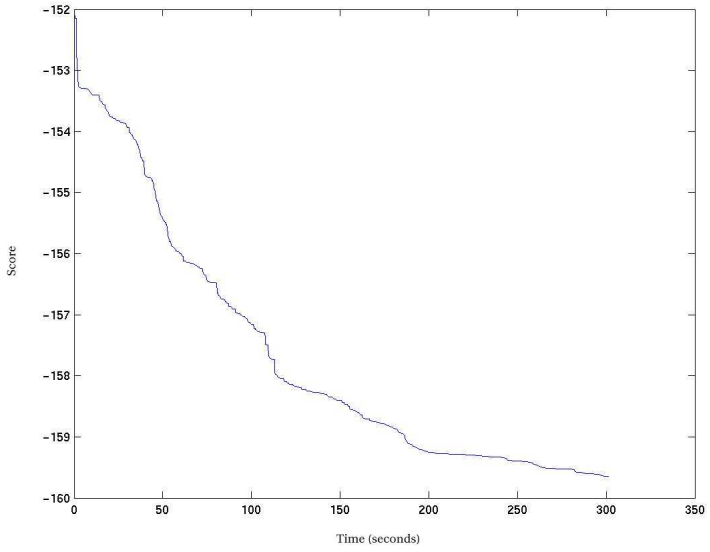

Figure 5.13: Images being registered according to the description length of the entire set of size 10. The X-axis indicates run-time time in seconds.

MDL and Models

It is realised that model residuals must to be included in some form or another (e.g. description length) in the objective function. As one example of the need for this issue to be resolved, see Figure 5.14. When not accounted for properly, the quality of the model can be perceived as though it surpasses the point of optimality. The incomplete term for description length is described in [16].

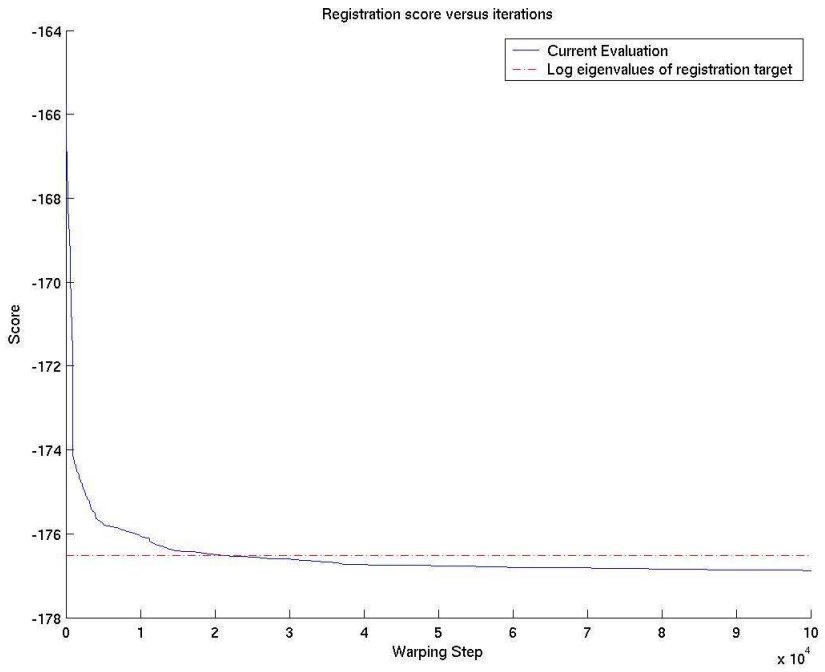

Figure 5.14: A long optimisation with the successful model-based algorithm shows that it surpasses what is questionably the correct solution (indicated by the red dotted line).

Automatic Landmark Selection

More considerable work was focused on shapes and the selection of landmarks which define point distribution models (PDM's). These are somewhat analogous to shape models. Work on shapes can be sub-categorised as follows.

Subsets

The idea of this approach is to speed up the algorithm by essentially pyramiding the whole set (see Figure 5.15) and building up towards a much quicker convergence.

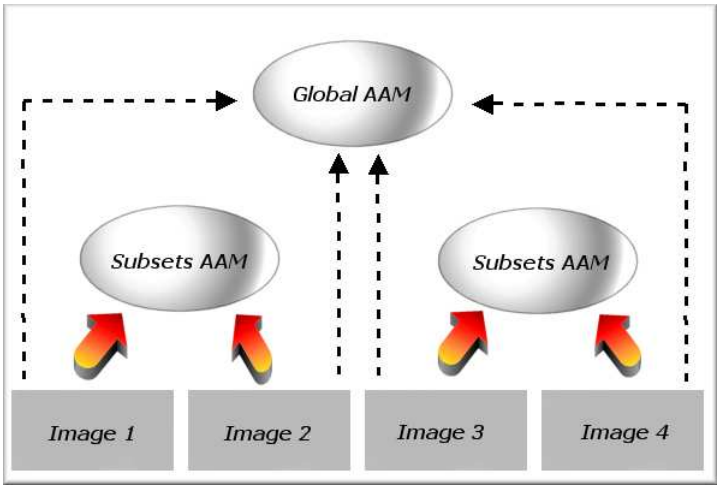

Figure 5.15: Illustration of the approach taken in registration using subsets.

This hierarchy permits larger sets to be dealt with, going as high as hundreds – something which has thus far been impractical. The figure shows how subsets are chosen in the context of image registration to create

smaller AAM's. In practice the choice is stochastic. By registering subsets, a near globally optimal AAM can be constructed. Similar principles can be shown for shapes.

Instead of treating large sets and optimising over these, smaller sets can be handled, thereby lightening the burdens of large Eigen analyses. Figures 5.16 and 5.13 illustrate that subsets appear to result in better and quicker descent[4]. The time required to optimise over subsets is significantly reduced.

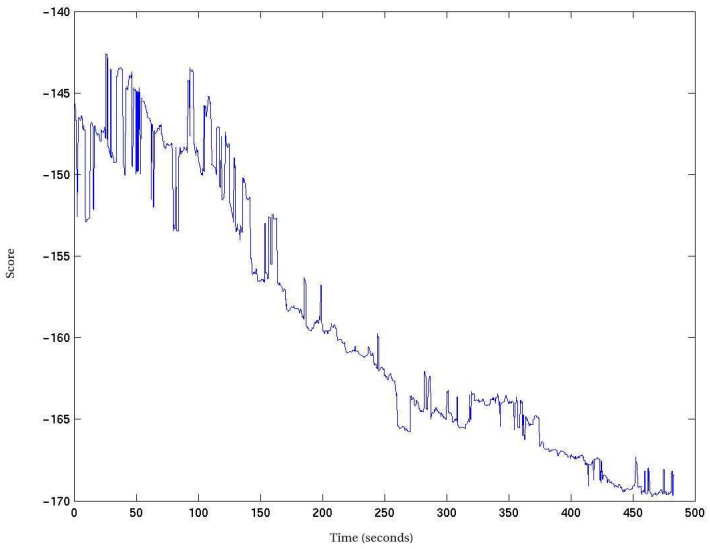

Figure 5.16: Images being registered according to the description length of random subsets comprising 4 images each. A choice of subset changes every 10 iterations. It is evident that the score goes lower, but the time required is then greater.

Figure 5.16 shows that a subset-driven approach is slower though it is able to bring about some great improvements after an initial instability

[4]This excludes the start when subsets require time to stabilise by preliminary warps.

at the start. That slow start can be explained by pointing out that an insufficient number of different subset choices was cycled through. As a result, a rather localised optimisation is performed while the overall set benefits very little.

Varying Optimiser Tolerance

As part of speed-up through modification, an adaptive precision approach was taken. The rate of convergence is changed as the process goes on and so is the speed of the algorithm. There is more to be investigated to ensure the approach invariantly results in gains. It is also worthwhile to see if the choice of tolerance can be made more preferable, based on empirical evidence. Experiments with varying values for tolerance showed that there was promise in an approach that seeks coarse solutions at the start and refines them further as it went along.

Taboo Search (TS)

To improve the performance of the search for good reparameterisations, a better optimisation method was sought. Among the methods explored was taboo search. Taboo search is a technique that attracted some interest in the 1990's [27]. It makes use of knowledge about the search space while optimising. Thus, it can look up previous decisions and reach good solutions rather rapidly. It is similar to Simulated Annealing from a theoretic point-of-view.

Inference for Images

Further improvements to the registration algorithms were inspired by experience acquired in past work. Previously, *ad hoc* improvements were made to the landmark selection algorithm. One such example is detection and rollback of any parameterisations which cause the assessor to report degradation in value. The optimiser, by its very nature, may return declining values once bad warps/parameterisations have been picked. Conventionally, this needs to be fixed manually, by intervening with decisions outside the optimisation routine. A mild adjustment, on the other hand, can automate this.

Comparison of Optimisation Regimes

Different optimisation regimes were investigated, e.g. by changing optimiser parameters for the standard optimisation function, as well as investigating approaches that are altogether different. Example of this is presented in the next section on automatic model-building in 2-D.

5.4 Model Building

5.4.1 Automatically Building Appearance Models

The previous section covered a variety of methods for registration of data. The output of such registration are warp fields that define one-to-one correspondences of points in the data. Having got these correspondences,

models can be built directly as described in § 3 on page 52. As the correspondence is a dense one, there is no limit in principle on the number of sample points from which the appearance model is constructed.

This section explores the extension of the image registration method to 2-D, as well as derivation of models from the registered data. What a registration algorithms establishes, regardless of the objective function used, is a dense correspondence in the images. Given this correspondence, however obtained, it is possible to learn the variation in shape and intensity. By applying the same algorithm to a set of one- or two-dimensional data, an appearance models can be built. The videos labeled `8.avi` and `14.avi` (see the accompanying CD-ROM) show the result of 1-D registration of the bump data using a model-based objective function. Combined models are shown, as well as shape and intensity models. Taking a similar approach to 2-D datasets, the same type of models can be built from brain data.

5.4.2 The Objective Function

In order to obtain a set of corresponding points, transformation of data need be involved. Two separate families of registration algorithms have been repeatedly used. The first method uses sum-of-squared-differences to measure similarity, whereas the second uses the minimum description length (MDL) framework. In both cases, transformation was handled solely using bi-harmonic clamped-plate splines. Although mutual information is a ubiquitous similarity measure, it did not prove to be as

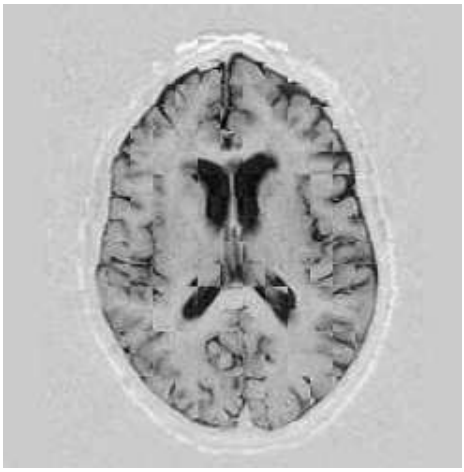

Figure 5.17: Discrepancy image showing the difference between two registered images. the objective function used was sum-of-squared-differences.

valuable in the experiments described in the remainder of this thesis.

Figures 5.18 and 5.17 show the registration process, by overlaying images and get a blocky blend that is the discrepancy image.

Lastly, as part of the experiments that investigate optimisation through minimisation of the objective function, a survey is shown which compared 3 methods. Figure 5.19 visually, as opposed to quantitatively, compares the level of refinement obtained through the different optimisation methods.

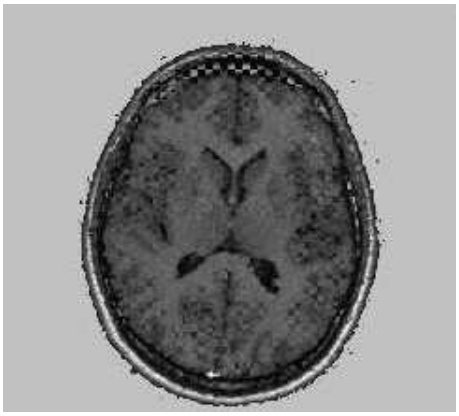

Figure 5.18: Discrepancy image showing the difference between two registered images (different from the image set shown in the previous figure). the objective function used was mutual information.

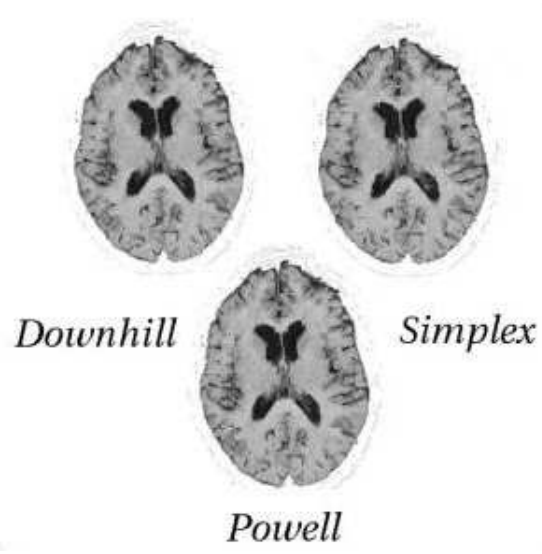

Figure 5.19: A survey of different registration optimisation methods. The figure shows the results in the form of discrepancy images, each for the results of a different optimisation method.

Chapter 6

Assessment of Models and Non-Rigid Registration

"Nothing in life is to be feared. It is only to be understood."

– Marie Currie.

THE most valuable contribution of the thesis is the introduction of a framework wherein NRR can be assessed without any use of ground truth. The framework is valuable not only because NRR applications are routinely used. This framework is considered useful because it offers a solution to a problem where cumbersome annotation is otherwise needed [9]. Novelty, on the one hand, should be attributed to seminal work by Davies *et al.* [16] while, on the other hand, the use of models in registration is unprecedented. Models were derived from a registration in the past [61], but they were not actively used to assist or drive registration and its assessment.

This chapter returns to discussing how models are built and proceeds to explaining an *ad hoc* method that is used to evaluate these models. Lastly, this evaluation method is applied to the problem of assessing the quality of NRR, merely by reversing a problem that is double-edged.

6.1 Building Appearance Models from Correspondences

As explained in Chapter 3, the key requirement in building an appearance model from a set of images, is the existence of a dense correspondence across the set. This is often defined by interpolating between the correspondences of a limited number of user-defined landmarks. Shape variation is then represented in terms of the motions of these sets of landmark points. Using the notation of Cootes et al [12], the shape (configuration of landmark points) of a single example can be represented as a vector x formed by concatenating the coordinates of the positions of all the landmark points for that example. The texture is represented by a vector g, formed by concatenating image values (texture) sampled over a regular grid on the *registered* image. This means that the a given element in g is sampled from an equivalent point in each image, assuming the registration is correct.

In the simplest case, the variation of shape and texture is modeled in terms of multivariate gaussian distributions, using Principal Component

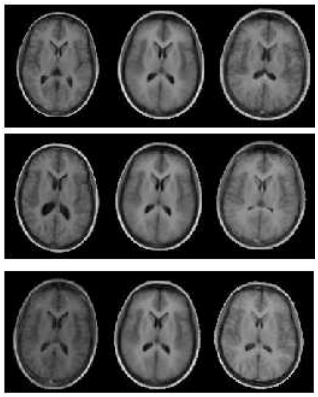

Figure 6.1: The effect of varying the first (top row), second, and third parameter of a brain appearance model by ± 2.5 standard deviations

Analysis (PCA) [34] to obtain linear statistical models of the form:

$$\begin{aligned} \mathbf{x} &= \overline{\mathbf{x}} + \mathbf{P}_s \mathbf{b}_s \\ \mathbf{g} &= \overline{\mathbf{g}} + \mathbf{P}_g \mathbf{b}_g \end{aligned} \tag{6.1}$$

where $\mathbf{b}_s$ are shape parameters, $\mathbf{b}_g$ are texture parameters, $\overline{\mathbf{x}}$ and $\overline{\mathbf{g}}$ are the mean shape and texture, and $\mathbf{P}_s$ and $\mathbf{P}_g$ are the principal modes of shape and texture variation respectively.

In generative mode, the input shape ($\mathbf{b}_s$) and texture ($\mathbf{b}_g$) parameters can be varied continuously, allowing the generation of sets of images whose statistical distribution matches that of the training set.

In many cases, the variations of shape and texture are correlated. If this correlation is taken into account, a combined statistical model is obtained. It has the more general form:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{c}$$

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c} \tag{6.2}$$

where the model parameters $\mathbf{c}$ control both shape and texture, and $\mathbf{Q}_s$, $\mathbf{Q}_g$ are matrices describing the general modes of variation derived from the training set. The effect of varying different elements of $\mathbf{c}$ for a model built from a set of 2-D MR brain images is shown in Figure 6.1. The number of modes (columns) in $\mathbf{Q}_s$ and $\mathbf{Q}_g$ is one less than the number of images. In practice, it is often possible to approximate images well, using fewer modes m.

Generally, we wish to distinguish between the meaningful shape variation of the objects under consideration, and the apparent variation in shape that is due to the positioning of the object within the image (the pose of the imaged object). In this case, the appearance model is generated from an (affinely) aligned set of images. Point positions $\mathbf{x}_{im}$ in the original image frame are then obtained by applying the relevant pose transformation $T_{\mathbf{t}}(\cdot)$:

$$\mathbf{x}_{im} = T_{\mathbf{t}}(\mathbf{x}_{model}) \tag{6.3}$$

where $\mathbf{x}_{model}$ are the points in the model frame, and $\mathbf{t}$ are the pose parameters. For example, in 2-D, $T_{\mathbf{t}}$ could be a similarity transform with four parameters describing the translation, rotation and scale of the object.

In an analogous manner, the image can also be normalised with respect to the mean image intensities and image variance,

$$\mathbf{g}_{im} = T_{\mathbf{u}}(\mathbf{g}_{model}), \tag{6.4}$$

where T_{u} consists of a shift and scaling of the image intensities. For further implementation details see [12, 24].

As noted above, a meaningful, dense, groupwise correspondence is required before an appearance model can be built. NRR provides a natural method of obtaining such a correspondence, as noted by Frangi and Rueckert [26, 61]. It is this link that forms the basis of our new approach to NRR evaluation.

The link between registration and modelling is further exploited in the Minimum Description Length (MDL) [80] approach to groupwise NRR, where modelling becomes an integral part of the registration process. This is one of the registration strategies evaluated here.

6.2 Generalisation and Specificity

A previous chapter, as well as Section 5.3, described how the results of NRR can be used to build a generative statistical model of image appearance. In this section, the method for quantitatively assessing the quality of the model is presented. The model is built from the registered data and, hence, the quality of the NRR from which the model was derived. Several variants of the approach are introduced, with the aim of finding one which is both robust and sensitive to small misregistrations.

A good model of a set of training data should possess several properties. Firstly, the model should be able to extrapolate and interpolate effectively

from the training data, to produce a range of images from the same general class as those seen in the training set. This property will be called *generalisation ability* from this point onwards. Conversely, the model should not produce images which cannot be considered as valid examples of the class of image modelled. That is, a model built from brain images should only generate images which could be considered as valid images of possible brains. This will be called the *specificity* of the model. In previous work, quantitative measures of *specificity* and *generalisation* were used to evaluate shape models [17]. The extension of these ideas to images (as opposed to shapes) is presented here. Figure 6.2 provides an overview of the approach.

Consider first the training data for the model, that is, the set of images which were the input to NRR. Without loss of generality, each training image can be considered as a single point in an n-dimensional image space. A statistical model is then a probability density function (pdf) $p(\mathbf{z})$ defined on this space.

To be specific, let $\{\mathbf{I}_i : i = 1, \ldots \mathcal{N}\}$ denote the $\mathcal{N}$ images of the training set when considered as points in image space. Let $p(\mathbf{z})$ be the probability density function of the model. A quantitative measure of the *specificity* S of the model is defined, with respect to the training set $\mathcal{I} = \{\mathbf{I}_i\}$ as follows:

$$S_\lambda(\mathcal{I}; p) \doteq \int p(\mathbf{z}) \mathbf{min}_i \left(|\mathbf{z} - \mathbf{I}_i| \right)^\lambda \, d\mathbf{z}, \tag{6.5}$$

where $|\cdot|$ is a distance on image space, raised to some positive power λ (for the remainder of this chapter only the case $\lambda = 1$ will be considered). That

is, for each point $\mathbf{z}$ on image space, the nearest-neighbour to this point in the training set is found. Then, the powers of the nearest-neighbour distances it is summed up, weighted by the pdf $p(\mathbf{z})$. Greater specificity is indicated by *smaller* values of S, and vice versa. In Figure 6.3, diagrammatic examples of models with differing specificity is given.

The integral in equation 6.5 can be approximated using a Monte-Carlo method. A large random set of images $\{\mathbf{I}_\mu : \mu = 1, \ldots \mathcal{M}\}$ is generated, having the same distribution as the model pdf $p(\mathbf{z})$. The estimate of the specificity (6.5) is:

$$S_\lambda(\mathcal{I}; p) \approx \frac{1}{\mathcal{M}} \sum_{\mu=1}^{\mathcal{M}} \mathbf{min}_i \left(|\mathbf{I}_i - \mathbf{I}_\mu|\right)^\lambda, \tag{6.6}$$

with standard error:

$$\sigma_S = \frac{SD_\mu \left\{\mathbf{min}_i \{|\mathbf{I}_i - \mathbf{I}_\mu|^\lambda\}\right\}}{\sqrt{\mathcal{M} - 1}}, \tag{6.7}$$

where SD_μ is the standard deviation of the set of μ measurements. Note that this definition of S does not require that the space of images is constructed. Instead, one simply needs to be able to define distances between images. This is discussed in Section 6.3 below.

A measure of generalisation similarly is defined, simply reversing the direction of the nearest-neighbour distance measure:

$$G_\lambda(\mathcal{I}; p) \doteq \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \mathbf{min}_\mu \left(|\mathbf{I}_i - \mathbf{I}_\mu|\right)^\lambda, \tag{6.8}$$

with standard error:

$$\sigma_G = \frac{SD_i\left\{\mathbf{min}_\mu\{|\mathbf{I}_i - \mathbf{I}_\mu|^\lambda\}\right\}}{\sqrt{\mathcal{N}-1}}. \tag{6.9}$$

That is, for each member of the training set $\mathbf{I}_i$, the distance to the nearest-neighbour in the sample set $\{\mathbf{I}_\mu\}$ is computed. Large values of G correspond to model distributions which do not cover the training set and have poor generalisation ability, whereas small values of G indicate models with better generalisation ability.

Note here that both measures can be further extended, by considering the sum of distances to k-nearest-neighbours, rather than just to the single nearest-neighbour. However, the choice of k would require careful consideration and in what follows, expeiments shown are restricted to the single nearest-neighbour case.

6.3 Image distance measures

The definitions provided for specificity and generalisation require a measure of separation in image space. The most straightforward way to measure the distance between images is to treat each image as a vector formed by concatenating the pixel/voxel intensity values, then take the Euclidean distance. This means that each pixel/voxel in one image is compared against its spatially corresponding pixel/voxel in another image. Although this has the merit of simplicity, it does not provide a very

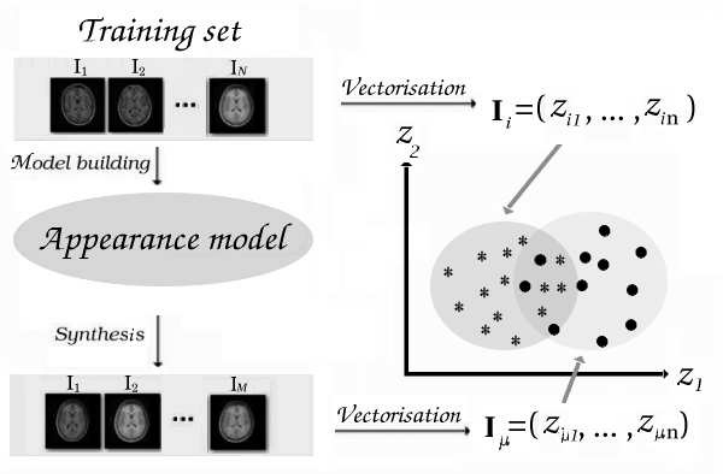

Figure 6.2: The model evaluation framework: A model is constructed from the training set and used to generate synthetic images. The training set and the set generated by the model can be viewed as clouds of points in image space ($\mathbf{I}_i$ represented by stars, and $\mathbf{I}_\mu$ represented by dots).

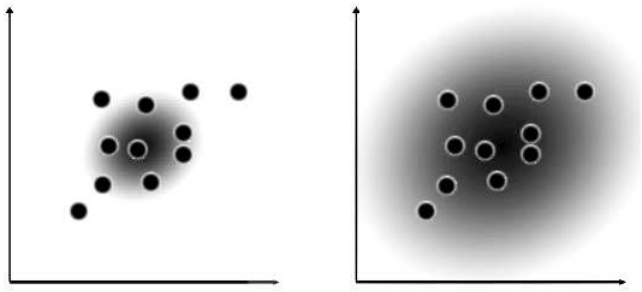

Figure 6.3: Training set (points) and model pdf (shading) in image space. **Left:** A model which is specific, but not general. **Right:** A model which is general, but not specific.

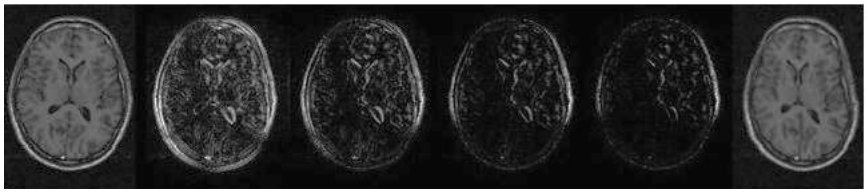

Figure 6.4: A comparison between shuffle difference images evaluated using various size neighbourhoods (radius r). **Left:** original image, **right:** warped image, **centre, from the left:** shuffle distance with $r = 1$(Euclidean), $1.5, 2.9$ and 3.7 pixels.

well-behaved distance measure since it increases rapidly for quite small image misalignments [84]. Another possibility is the use of Image Euclideadn Distance (IMED), as proposed in a recent paper. While this approach was implemented and tested, it was not considered further due to efficiency and constraints.

This observation led to consideration of an alternative distance measure, based on the 'shuffle difference', inspired by the 'shuffle transform' [39]. Given two images $\mathbf{I}_1(\mathbf{x})$ and $\mathbf{I}_2(\mathbf{x})$, the shuffle distance between them is defined as

$$D_s(\mathbf{I}_1, \mathbf{I}_2) = \frac{1}{n} \sum_{\mathbf{x}} \mathbf{min}_i \|\mathbf{I}_1(\mathbf{x}) - \mathbf{I}_2(\mathbf{N}_i(\mathbf{x}))\| \tag{6.10}$$

where $\|\cdot\|$ is the absolute difference, there are n pixels (or voxels) indexed by $\mathbf{x}$, and $\{\mathbf{N}_i(\mathbf{x})\}$ is the set of pixels in a neighbourhood of radius r around $\mathbf{x}$.

The idea is illustrated in Figure 6.5. Instead of taking the sum-of-squared-differences between corresponding pixels, the minimum absolute difference between each pixel in one image and the values in a neighbourhood around the corresponding pixel is used. This is less sensitive to small misalignments, and provides a better-behaved distance measure. The tolerance for misalignment is dependent on the size of the neighbourhood (r), as is illustrated in Figure 6.4.

It should be noted that the shuffle distance as defined above depends on the direction in which it is measured (see Figure 7.1), hence is not a true distance. It is trivial to construct a symmetric shuffle distance, by averaging the distance calculated in both directions between a pair of

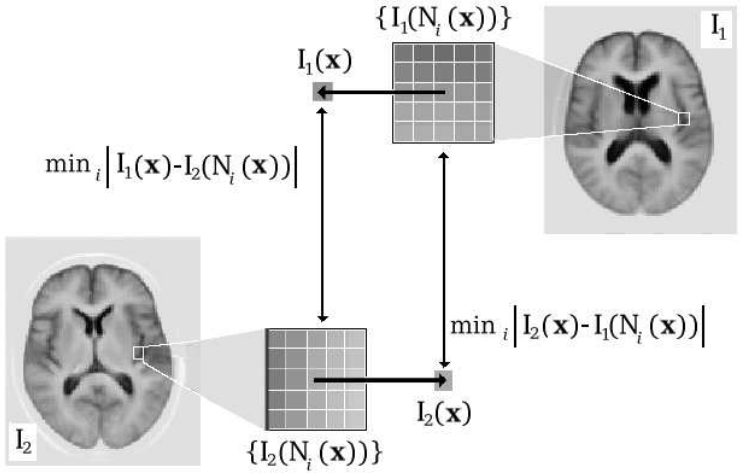

Figure 6.5: The calculation of a shuffle difference image

images. It was found, however, that the improvement obtained was not significant, and did not justify the increased computation time. In what follows, the asymmetric shuffle distance is used exclusively.

Chapter 7

Validation Methodology and Experiments

"We know nothing in reality; for truth lies in an abyss."

– *Democritus.*

In order for the assessment method to deem credible, one need to test it against data where the correct solution is known. Then, in attempts to arrive at a particular true answer, one can confirm correctness and validity of the method. In the next couple of chapters, the most comprehensive set of experiments is described in detail. The experiments represent a culmination of the work since they address and solve a problem which is commonly explored. They embody just one aspect of the work, which successfully unifies a broader whole. While building of models using NRR and evaluation of any type of model is possible, the chapter only concentrates on in-depth experiments where the quality of NRR of two brain

sets is studied.

Two sets of experiments were performed, one designed to validate the model-based approach for evaluating NRR, the other to demonstrate its use in a practical application. The latter experiment is described in the next Chapter.

In the first set of experiments the aim was to show that Specificity and Generalisation are valid measures of the degree of misregistration of a group of images. It is expected that, as registration is degraded, Specificity and Generalisation should respond accordingly. If the measures are indicative of the quality of NRR, then they can be employed in benchmark-type NRR studies.

A set of registered images, for which ground-truth labels were available, was used and a series of deformations was then applied. The labels were binary images, each of which corresponds to an anatomical compartment of the brain. As deformations were applied to the images and their accompanying labels, any binary image was interoplated to become fuzzy. Images and their ground-truth labels were deformed in tandem so, for every pixel in the image, its corresponding anatomical classification (derived from the corresponding label) moved along with it. The deformation was repeated with varying degress of magnitude, which were carefully controlled and studied from the warp fields. This introduced progressively-increasing misregistration. This made it possible to investigate how measures of Specificity and Generalisation varied, as a function of the known misregistration. The newly-created test set was composed of progressively-degraded pseudo-registrations that substitute real

NRR results where the correct solution is unknown, or subjected to bias. Generalised overlap was also measured for each of the deformed image sets, using the ground-truth labels, to provide a comparison with the existing method that is based on models.

In the second set of experiments the aim was to demonstrate that one could usefully discriminate between different NRR algorithms, by comparing results for the same dataset. Several algorithms were to perform a full registration on the same datasets and their results evaluated using the same three measures (metric) that have been validated.

7.0.1 Image Data

To conduct the experiments two different sets of MR images of the brain were used. The first, which will be referred to as the 'MGH Dataset' (see Acknowledgements), was a set of 2-D transaxial mid-brain slices, extracted at an equivalent level from each of a set of affinely aligned T1-weighted 3-D MR scans of $\mathcal{N} = 36$ normal subjects. As well as the images themselves, there is access to ground-truth data, in the form of dense (pixel by pixel) anatomical label maps for the gray and white matter, the caudate nucleus, and the lateral ventricles. These labels were further divided into left and right hemispheres. The anatomical labels were obtained by manual annotation under conditions of rigorous quality control. An example image and the corresponding label maps are shown in Figure 7.2.

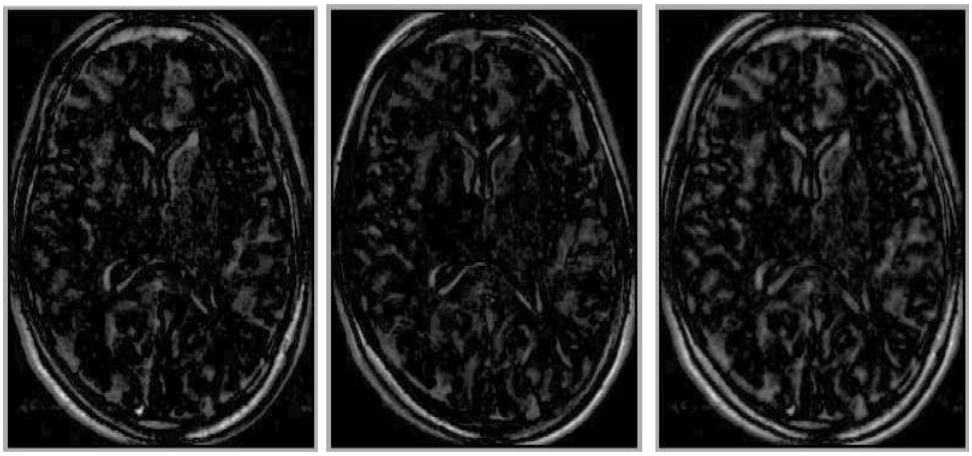

Figure 7.1: Examples of the shuffle difference image: from first to second (left), from second to first (centre), and the symmetrical shuffle difference image (right)

The set of images was non-rigidly registered using a Minimum Description Length (MDL) NRR algorithm [80], and this registration was used as the starting point for a systematic evaluation of the effects of misregistration. Although the initial registration is subjected and inclined to the MDL objective function, this provides a sufficiently-good starting point, assuming that each deformation degrades the registartion rather than improve it.

The second set of images, which will be referred to as the 'Dementia Dataset', consisted of a set of 2-D transaxial mid-brain slices, extracted at an equivalent level from each of a set of affinely-aligned T1-weighted 3-D MR scans of $\mathcal{N} = 104$ subjects entered into a clinical study of dementia. These images varies in terms of intensity, size, and shape from the former set, which means that the method cannot be fit and customised to work with just a particular set of data. This makes the arguments as regards validation ever more compelling. While other datasets such as face images were used in validation experiments where models are shown to

degrade, in terms of Generalisation and Specificity, as a function of deformation, their scope and contribution is assumed to be unnecessary for the method's validity to be defended.

7.0.2 Perturbing the Initial Registration

In order to perform a systematic evaluation of the effects of misregistration, multiple image sets were created, based on the MGH Dataset, but with controlled degrees of misregistration. To create a misregistered set, the original image set was taken and had applied to it a set of smooth pseudo-random spatial warps, based on biharmonic Clamped Plate Splines [79]. The warp for each image was controlled by 25 randomly placed knot-points, each displaced in a random direction by a distance drawn from a Gaussian distribution whose mean controlled the degree of misregistration introduced. This provided a very general family of warps. The direction and magnitude of these warps were carefully studied to ensure that all parts of the image were subjected to homogenous deformations. The degree of misregistration was analysed by measuring d, the average magnitude of pixel displacement over the whole image. This was done by tracing the per-pixel shift in the warp fields. The standard deviation of these warp was verify the validity of this Gaussian distribution. A total of 70 misregistered image sets were generated– 10 warp-set instantiations for each of 7 different values of d (0.0643, 0.249, 0.685, 1.36, 2.21, 2.76, and 4.15 pixels). Examples of warped images are shown in Figure 7.3.

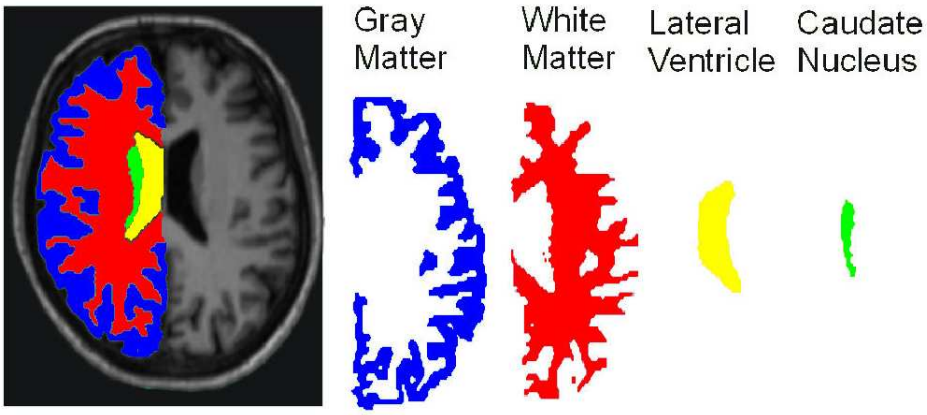

Figure 7.2: An example affinely-aligned brain image and its accompanying anatomical labels, both overlaid and expanded, for gray matter, white matter, the lateral ventricles, and the caudate nucleus. The labels are also divided into left and right.

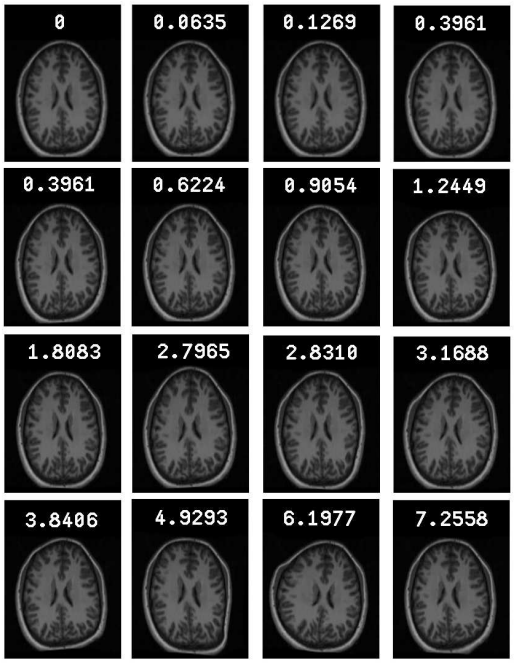

Figure 7.3: An original image from the MGH Dataset (top left) and examples of warped versions of the same image obtained using different values of d, the mean pixel displacement (shown on each image).

7.0.3 Validation using Warped Images

Given the 70 image sets described above, each with known average misregistration, d, the relationship between d and Specificity, Generalisation, and Generalised Overlap was investigated, by calculating the mean and standard error for each measure over the 10 warp instances for each value of d. In total, there were 71 image sets to study. One is the original registered set and the other 70 image sets comprise 10 instantiatuion for each value of d.

For each misregistered image set, Specificity and Generalisation were calculated, as described in Section 6.2, using m = 15 modes of variation for the model and $\mathcal{M}$ = 1000 synthetic images drawn from a Gaussian distribution, as described in Section **??**. This was repeated for values of shuffle radius, r, of 1 (Euclidean distance), 1.5, 2.1 and 3.7, as defined in Section 6.3, corresponding to circular neighbourhoods contained within 1x1, 3x3, 5x5 and 7x7 pixel patches respectively. These experiments were repeated with 2.5%, 5.0% and 10% Gaussian intensity noise added to the misregistered images, in order to investigate the sensitivity of the model-based measures to image noise. This makes possible to argue in favour of the robuestness of these measures to noise, as well as knowing its caveats.

Similarly, Generalised Overlap with volume, equal, inverse volume and complexity weightings were calculated, as defined in Chapter 2. The mean and standard error for each measure over the 10 warp instances for each value of d was also calculated.

7.0.4 Sensitivity

The size of perturbation that can be detected in the validation experiments will depend both on the change in the values of the measures as a function of misregistration and the standard error of those values. To quantify this, the sensitivity of a measure was defined as follows.

$$D(m;d) = \frac{1}{\sigma_m}\left(\frac{m(d) - m(0)}{d}\right), \tag{7.1}$$

where $m(d)$ is the value of the measure for some degree of deformation d, σ_m is the standare error of the estimate of $m(d)$. $D(m;d) = 1$ is the change in d required for $m(d)$ to change by one noise standard error, which indicates the lower limit of change in misregistration d which can be detected by the measure. D is a function of d; to simplify comparison between different methods of evaluation, we also use the mean sensitivity over a range of values of d.

In order to compare the sensitivities of different methods of evaluation, the expected error in D also needed to be estimated. Since the validation experiments provided repeated estimates of $m(d)$, one can obtain empirical estimates of the errors in $m(d)$, $m(0)$, and σ_m. These can be combined, using error propagation, to estimate the uncertainty in the estimate of sensitivity.

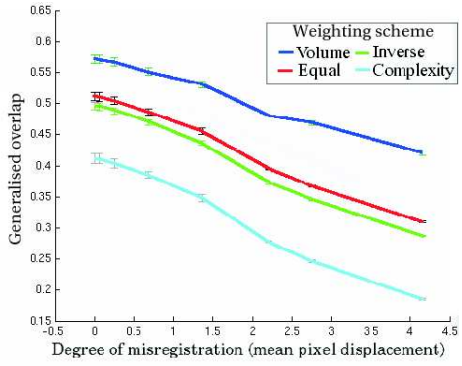

Figure 7.4: Overlap measures (with corresponding $\pm$ one standard error errorbars) for the MGH dataset as a function of the degree of degradation of registration correspondence, d. The various graphs correspond to the various tissue weightings as defined in Chapter 2.

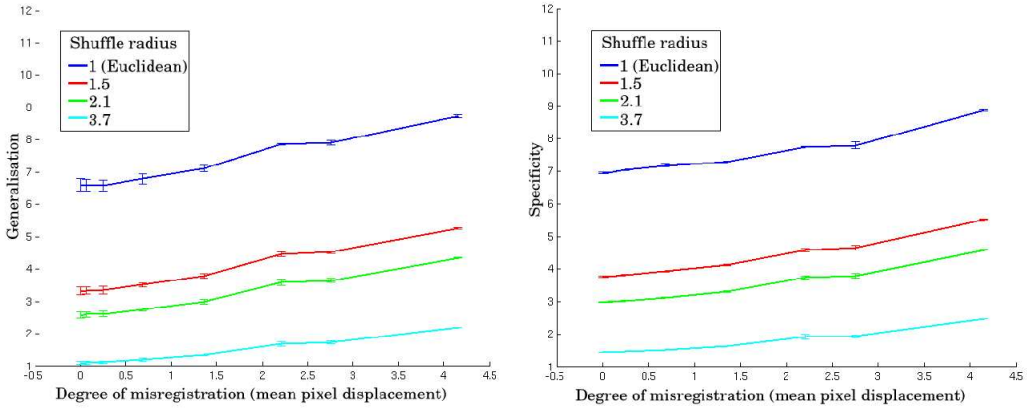

Figure 7.5: Generalisation & Specificity for various definitions of image distance (varying shuffle radius) with corresponding $\pm$ one standard error errorbars as a function of the degree of degradation of the registration correspondence d for the MGH dataset

7.1 Results

Figure 7.4 plots each of the four variants of the generalised overlap measure, as a function of d, the degree of misregistration. As expected, the value decreases monotonically with increasing misregistration, in each case. This shows that the two gold-standard measures of misregistration (mean pixel displacement and ground-truth overlap) are in agreement, which validates the experimental framework.

Similarly, Figure 7.5 plots Generalisation and Specificity as functions of d, for different values of shuffle radius r. The results are qualitatively similar to those obtained for generalised overlap, except that both measures *increase* monotonically with increasing misregistration, as expected (see Section 6.2. These results show that, over the range of misregistrations investigated, the model-based measures are good surrogates for d, the mean pixel misregistration. Since the warps used to introduce controlled misregistration were of very general form, there is no reason to suppose that this result is dependent on the pattern of misregistration.

7.1.1 Sensitivity

Figure 7.6 shows the results of applying sensitivity analysis to the validation study. These demonstrate that Specificity is more sensitive (is able to detect smaller misregistrations) than the overlap-based approach, which is in turn more sensitive than Generalisation. Note from the error bars that these differences are statistically significant. Maximum sensi-

tivity is achieved with a shuffle radius of 1.5 or 2.1. The most sensitive generalised overlap measure is obtained using label-complexity weighting.

7.1.2 Effect of Noise

The validation experiments were repeated and sensitivity analysis reported above with added image noise. Although the absolute values of the model-based measures were shifted upwards, as would be expected, there were no changes in the relative values, nor any systematic or statistically significant changes in sensitivity, even for 10% added noise.

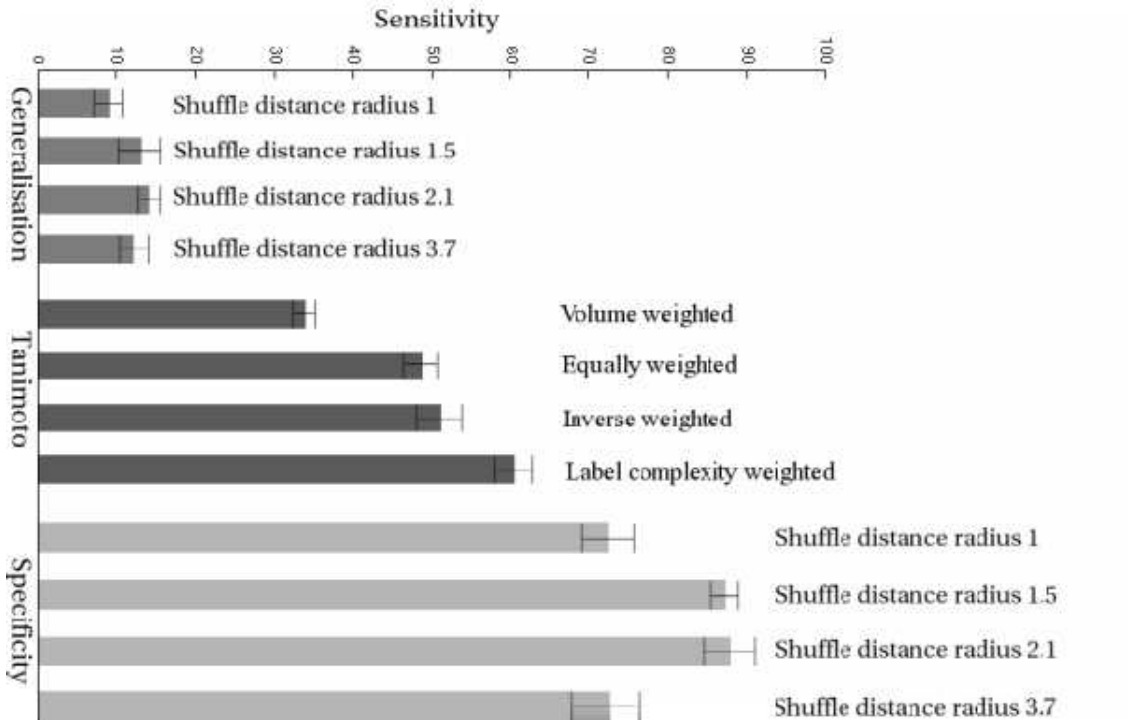

Figure 7.6: Mean sensitivity of different NRR assessment methods over the full range of deformations d, shown with $\pm$one standard error errorbars

Figure 7.7: The first mode (± 2.5 standard deviations) of an appearance model built automatically by group-wise registration.

Chapter 8

Application to Non-Rigid Registration Evaluation

"Civilisations can only be understood by those who are civilised."

– Alfred North Whitehead.

THIS chapter presents one among the practical applications of the word presented thus far. Apart from assessment of appearance model of the brain and human face, one is able to assess the quality of registration algorithm. The following benchmark is the most extensive set of experiments performed and it demonstrates that an important routine task such as NRR can be indeed simplified, by obviating the need for ground truth data.

8.1 Comparing Registration Algorithms

To illustrate the application of model-based evaluation in practice, the NRR results obtained using three different methods for registering a group of images were compared, as described in more detail below. The intent was to establish whether it was possible, in a practical setting, to detect significant differences in performance between different NRR algorithms. All three registration methods used the same piecewise affine representation of image warps [15] and the same multi-resolution optimisation framework. The same number of iterations (function evaluations) were used in each case.

The three registration algorithms were applied to two datasets. The MGH Dataset was used because it allowed the evaluation results obtained using Specificity and Generalisation to be compared with an evaluation based on the Generalised Overlap measure (using ground truth). For these experiments $\mathcal{M}$ = 500 synthetic images were used to estimate Specificity and Generalisation. The Dementia Dataset was used because it was more representative of a typical clinical study, and it is important to demonstrate that the results were not dataset-specific. For these experiments $\mathcal{M}$ = 1000 synthetic images were used.

The three registration methods used were as follows.

8.1.1 Pairwise Registration to a Reference

A commonly used approach to registering a group of images is to register each image in the group in turn to a reference image chosen from the group, using a pairwise objective function (e.g., [61]). We used this approach as a baseline, with a sum of absolute intensity differences objective function (which gave slightly better results than sum of squared differences or mutual information).

Pairwise approaches to registration can produce reasonable correspondences, but suffer from the problem that the results obtained depend on the choice of reference. Refinements of the basic approach are possible, where the reference is initialised and updated so as to be representative of the group of images as a whole. It is important to note, however, that even in this case the correspondence for a given image is determined solely by the information in the image and the reference. More recently, there has been considerable interest in *groupwise* methods which aim to make more systematic use of the information in the complete set of images when establishing correspondence. The remaining two methods we tested fall into this category.

8.1.2 Groupwise Congealing Algorithm

Learned-Miller et. al. [49] originally introduced their 'congealing' algorithm for registering a set of hand-written digits. The aim was to avoid the arbitrary selection of a co-ordinate frame, by repeatedly registering

each image with an evolving "average" model. Given the current set of transformed images (initially the original images), for each pixel position, i, the probability density function of intensities, v, at that position across the set of images, $p_i(v)$ is estimated. The objective function is then the sum of entropies of these distributions across the whole image, $F = \sum_i \int p_i(v) \log p_i(v) dv$. A set of image deformations were optimised to minimise this. In later work on registering sets of 3-D medical images [92], the objective function was approximated by $\sum_j \sum_i \log p_i(v_{ij})$, where v_{ij} is the value of pixel i in deformed image j. During optimisation, each image was warped so as to bring pixels with similar intensities into correspondence across the set. We implemented this later approach.

8.1.3 Groupwise MDL Algorithm

A groupwise method which uses a Minimum Description Length (MDL) formulation [80] has previously been described. The main idea is that the complete set of images can be encrypted as a coded message, and the description length [58] in bits used as an objective function. Rather than encoding the raw images, the encoding uses an appearance model, built using the estimated correspondences, to approximate the data; the encoding needs also to include details of the model itself and of the discrepancy between each image and its model approximation. As the registration proceeds, the correspondences, and hence the appearance model, are continually updated so as to minimise the description length.

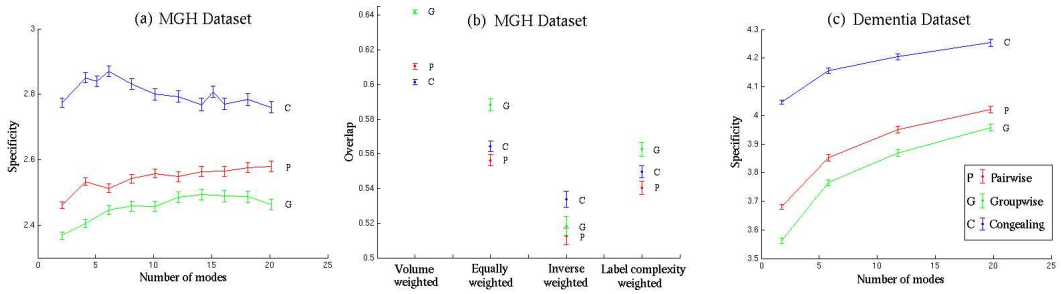

Figure 8.1: **Left and right:** Generalisation and Specificity of the three registration methods as a function of the number of modes included in the appearance

8.2 Results of Comparison

Figure 8.1 compares the performance of the three registration algorithms outlined in Section 8.1. All the measures tested in the previous section were computed, but we show results for only the most sensitive model-based method. Figures 8.1(a) and (c) show Specificity calculated using a shuffle radius of 2.1, for different values of m, the number of modes used to build the generative model. Figure 8.1(b) shows generalised overlap using different weightings. The results shown in Figure 8.1(a) suggest that the MDL groupwise approach gives the best registration result for the MGH Dataset, followed by Pairwise and Congealing in order of decreasing performance – irrespective of the value of m. Inspection of the error bars shows that these differences are statistically significant. The results for Generalised Overlap, shown in Figure 8.1(b), are more complicated, with the performance of the different NRR algorithms ordered differently for different weightings, though inspection of the error bars shows that many of the differences are not significant. Overall, the same

general pattern emerges as for Specificity, with the Groupwise method generally best (statistically significantly in two cases), but with no significant difference between Pairwise and Congealing in most cases. The results for inverse volume weighting generally lack significance, but are inconsistent with those obtained using the other weighting schemes. Volume weighting gives the best separation between the different variants, and places the three methods in the same order as Specificity. Overall, this supports the interpretation that Specificity give results that are generally equivalent to those obtained using Generalised Overlap, but with higher sensitivity. Finally, the Specificity results shown in Figure 8.1(c) for the Dementia Dataset, place the three methods in the same order.

Chapter 9

Extensions to 3-D

"If you're not part of the solution, you're part of the precipitate."

– *Henry J. Tillman.*

THE model and NRR assessment methods were extended to operate on three-dimensional data. Rather than handling images, the methods then deal with volumes and, in accordance, shuffle distance neighbourhoods become a box or sphere of voxels, rather than a square or a disk.

The chapter alludes to work which is performed at present. Thus, in this particular chapter, along with the subsequent chapter on future paths, planning and strategies are described rather than complete work.

Figure 9.1: A multi-resolution approach illustrated in 2-D. Coarser representations are shown at the top levels and the original image lies at the bottom.

9.1 Speed Limitations

There are several tricks-of-the-trade, which can be used to speed up the process of registration assessment and model evaluation in 3-D. One strategy, for example, involves rescaling the data, then dealing with smaller versions of the whole, as shown in Figure 9.1.

The basic concept is that, if the scale of the problem is reduced, it can be handled at a coarse level and then iteratively handled at finer level, until the original unscaled data is reached.

Progressive refinement of the numerical results is a strategy which can also be used in 2-D. However, only when the scale of the problem rises to include a third dimension, *ad hoc* solutions become a necessity. Clearly, results of an assessment experiment that operates on a reduced-size set will miss information that exists at the finer levels of granularity. Thus, only an approximation can ever be obtained.

9.2 Progressively-improved Estimates

A strategy worth employing is one which offers a figure of merit that is gradually made more accurate. In cases where a benchmark is performed, conclusions can be reached early on. In the case where synthetic images are considered, this strategy is a possibility even without scaling. The size of the synthetic set can be simply limited. However, an assessment method that is quicker to deliver an estimate is one which makes use of the entire set (whether coarse or not) rather than use just a subset that is subjected to bias.

There different ways of reducing the complexity of the problem at hand. In practice, this means that a box of voxel may be sliced into 8 (2^3) equal-sized boxes which are then used in the analysis, or even more usefully, the box should be rescaled to become 8 times smaller, in terms of volume.

9.3 Selective Assessment of Slices

In general, there is not much which distinguishes the method's use in 2-D and in 3-D, other than efficiency factors. However, several possibilities emerge owing to the fact that 3-D data can be interpreted differently once its dimensionality is reduced. For example, one can choose one representative slice from a larger volume and refine the evaluation by considering more slices, one at a time. This suffers from the fact that voxels whose position varies in the third dimension, i.e. it moves between the slices

due to warping, will not be treated appropriately. All these issues, along with other pitfall, will be addressed in the future.

Chapter 10

Future Exploration

"We know nothing in reality; for truth lies in an abyss."

– *Democritus.*

HAVING introduced a small family of methods which successfully solve the problems at hand, it would be most desirable to look ahead and make proposals. Several deficiencies of the methods are yet to be addressed and various extensions implemented. While the methods accomplish their goals, there is place for further refinement and improvements. Generalisations, customisations, and further simplifications can be envisioned and they are all motivated by known drawbacks.

The chapter lists possible paths that take the existing frameworks and enable them to perform better and even incorporate additional functionality.

10.1 Pitfalls

Computational loads are an important factor that has become a barrier, particularly in 3-D. There are particular steps in the algorithm whose computational cost is far greater than the remainder. Firstly, one must consider the long time that is required to synthesise many images from appearance models and subsequently use them in an evaluation. The greater the number of synthetic images, the more accurate the results. This relationship means that there is no clear point of balance. Sufficiency in the evaluation can never be attained.

Secondly, the more time-consuming process involved the computation of inter-image distances. With the added complexity of a third dimension, as well as a shuffle distance with large neighbourhood sizes, there is a considerable cost which is proportional to the number of voxels at hand.

Another problem one can identify lies in the fact that computation of Sensitivity and Generalisation is not principled. This can be corrected by calculating the self-normalising pseudo-entropy of graphs [52]. This graph represents the distances between images (edges) where vertexes are individual images. Entropic graphs is an area that was explored in great depth, yet it turned out to be rather complex due to the need to estimate many parameters, using a Monte-Carlo simulation. Although efforts to adopt the method have been conceded, there is place to propose another paradigm for dealing with this issue. The *ad hoc* nature of Specificity and Generalisation leaves plenty of room for new measures to evolve. Whether alternative method would be equally cheap to compute

remains an unknown. As in many such large-scale problem, simplicity has its merits, too.

The issues are dealt with in depth in the next section.

10.2 Extending the Scheme

There are a few proposed improvements that can further improve the validity and accuracy of the metrics.

10.2.1 Normalisation

At present, values returned for various measures are dependent upon the size of the sets, the dimensionality and a few other free parameters. This makes it difficult to argue about and distinguish between results from different experiments, unless all conditions (free parameters) were identical. For example, an experiment performed with large images and small sets cannot trivially be compared against other experiments involving small images and very large sets. In order for all results to be numerically comparable, there need to be a normalisation stage, which accounts for the many free parameters simultaneously.

10.2.2 Investigating Robustness

One aspect which must never be neglected are the boundaries and edge cases. It is valuable to know where the methods cease to be valid as noise

levels supersede the signal. Having found the limitations of the method, their robustness can be improved. For instance, in the case of model assessment, one can improve the range of displacements where results can be differentiated by increasing the size of the shuffle neighbourhood. Prior experiments showed that the performance is degraded beyond a certain shuffle neighbourhood size, but there are other parameters that can be varied and their effect on the shuffle neighbourhood is not mutually exclusive.

10.2.3 Further Improvement of Sensitivity

Of particular interest is the notion of sensitivity as it enables one assessment method to be compared against another. Although sensitivities were shown to culminate at a particular value of the shuffle distance, other approaches that had been investigated could perhaps entail superior sensitivity, at the expense of computing power. It is worth exploring if a more complex approach, e.g. one which considers an average or median in a neighbourhood of pixels/voxels, outperforms shuffle distance.

Chapter 11

Summary and Conclusions

"If you're not part of the solution, you're part of the precipitate."

– *Henry J. Tillman.*

THE work covered in this thesis can be summarised as follows. Firstly, a novel framework was described which non-rigidly registers images using a model-based similarity measure. This framework is able to deal with any type of images and. while it requires a number of images in order to become practical (i.e. in order for a sensible model to be built), its performance not depend on the type of variation that is contained in the set of images. As a result of registering the images using a model-based approach, one also obtains an appearance model, which is progressively refined and whose quality is dependent on the quality of the registration algorithm. This establishes a framework for automatic construction of models that requires nothing but a registration algorithm.

The second part of the work is concerned with assessment. Two things are being assessed: the quality of any appearance model (or any generative model) and the quality of a groupwise registration. This opens up the possibilities of comparing NRR algorithms and hand-tweaking them for better performance.

11.1 Discussion

The results of the validation experiment reported in Section 7 are the most important outcome of the work presented here. They demonstrate a causal relationship between our Specificity and Generalisation measures, and a known (up to an additive constant) mean pixel displacement, d. A strong correlation between these model-based measures and a Generalised Overlap measure, based on ground truth, adds further weight to this interpretation. The fact that the relationship with d held good over many different instantiations of a very general class of perturbing warps, makes it unlikely (though not impossible) that there is any significant pattern dependence.

The results obtained with added noise are also encouraging, since it is a reasonable concern that the use of an intensity-based distance measure might make the model-based measures sensitive to noise. In the event, the approach seems robust to quite significant levels of noise. The fact that the absolute values of specificity and generalisation change when noise is added, mean that they would not be useful for comparing registration results for different image sets. Their ability to compare the

performance of different registration algorithms applied to the same set of images, the main intended use, is, however, unaffected.

Our results comparing the performance of different registration algorithms demonstrate that the model-based measures, and Specificity in particular, are sufficiently sensitive to misregistration to provide useful discrimination in a practical setting. There is, however, a potential concern that it is important to address. It might be argued that using a model-based approach to assessing registration favours methods which use a model-based objective function for registration (as in the experiments reported here). In practice, we do not believe that this is a problem.

First, as we have argued above, our validation results show that there is a causal relationship between the mean pixel displacement, d, and Specificity/Generalisation. It is thus irrelevant how a registration (or misregistration) has been obtained. Second, the MDL objective function we optimise in our model-based registration method measures a quite different property of the model to those we use in evaluation, so there is no element of 'self-fulfilling prophecy. In an ideal world it would, of course, be preferable to avoid even the possibility of bias, though it seems unlikely that one could devise a strategy for evaluation that had no relevance to achieving a good registration in the first place. We hope that, in due course, other ground-truth-free methods of evaluation will be developed, allowing a multi-perspective assessment of performance.

One obvious limitation of our approach to evaluation is that it can *only* be applied to groups of images. This could be considered an important

restriction, since many practical applications involve registration of pairs or very short temporal sequences of images. We would argue that, in fact, this is a necessary restriction, because it is only possible to arrive at a meaningful assessment of registration in the context of a population of images.

The experiments we have reported were performed in 2-D to limit the computational cost of running the large-scale evaluation for a range of parameter values and with repeated measurements. The extension to 3-D is, however, trivial, though the calculation of shuffle distance for 3-D images increases the computational cost significantly. We have implemented the method in 3-D and the time taken to evaluate the registration of 100 190x190x50 images using a shuffle radius of 2.1 and $\mathcal{M} = 1000$ is around 62.5 hours on a modern PC, which is short compared to most registration algorithms.

There are a number of issues that merit further investigation. We have studied a particular method of measuring image separation, but others, such as local correlation, would be worth exploring. Another interesting issue is whether it is possible within this framework to localise registration errors. We have performed some initial experiments, summing the shuffle difference maps between all pairs of images in the registered set. This gives some interesting results, highlighting areas of common misregistration, but it is not clear what quantitative interpretation could be placed on such maps. Finally, it is clear that our current measures of Specificity and Generalisation are not normalised – their values depend on the size of the set of registered images, the number of synthetic images

generated and so on. We are currently exploring the possibility of measuring more fundamental properties of the relationship between the real and synthetic image distributions, with a view to achieving a 'natural' normalisation.

11.2 Conclusions

We have described a method for registering images in a groupwise fashion including the quality of their appearance model in the objective fucntion. Not only does this enable us to reach good results from a groupwise perspective, but it also results in the automated construnction of appearance model, whose quality is assessed.

We have also described a model-based approach to evaluating the results of NRR of a group of images. The most important advantage of the new method is that it does not require any ground truth, but depends solely on the registered images themselves.

We have validated the approach by studying the effect of perturbing, progressively, the registration of an initially registered set of images, comparing the results with those obtained using a 'gold standard' measure based on the overlap of ground-truth anatomical labels. We have shown that our new method provides measures of registration accuracy that are monotonic functions of the known misregistration, and that one, *Specificity*, provides a more sensitive measure of misregistration than the approach based on ground truth.

The model-based approach requires a distance measure in image space, and we have also demonstrated that the use of shuffle distance, rather than Euclidean distance, improves the sensitivity of the approach.

We have further validated the approach, and illustrated its application, by performing a comparative evaluation of the results obtained using three different NRR algorithms, demonstrating the superiority of a fully-groupwise algorithm over a repeated pairwise approach.

It is important to emphasise that our approach is not restricted to evaluating model-based NRR algorithms, though we presented results for one such method; the model-based measures of registration accuracy can be applied to any set of non-rigidly registered images, however they were obtained. We have discussed the possibility of a bias in favour of model-based methods of registration and conclude that there is no major problem, though it would be desirable to compare results obtained using a range of ground-truth-free methods of evaluation.

Bibliography

[1] A. Abkar and H. Hedenmalm, "A Riesz Representation Formula for Super-Biharmonic Functions," Annales Academie Scientiarum Fennice Mathematica, vol. 26, pp. 305–324, 2001.

[2] J. Ahlberg and R. Forchheimer, "Face tracking for model-based coding and face animation," International Journal of Imaging Systems and Technology, vol. 13, pp. 8–22, 2003.

[3] C. J. Beeston and C. J. Taylor, "Automatic landmarking of cortical sulci," in Medical Image Computing and Computer-Assisted Intervention 2000, vol. 1935, pp. 125-133.

[4] 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.

[5] D. Beymer and T. Poggio. "Image Representations for Visual Learning," In Science, vol. 272, issue 5270, pp. 1905 -1909.

[6] K. K. Bhatia, J. V. Hajnal, B. K. Puri, A. D. Edwards, and D. Rueckert, "Consistent groupwise non-rigid registration for atlas construction," presented at ISBI, Arlington, VA, USA, 2004.

[7] F. L. Bookstein, "Principal Warps: Thin-Plate Spines and the Decomposition of Deformations," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, pp. 567–585, 1989.

[8] A. D. Brett and C. J. Taylor, "A method of automated landmark generation for automated 3D PDM construction," Image and Vision Computing, vol. 18, pp. 739–748, 2000.

[9] W. R. Crum, O. Camara, D. Rueckert, K. Bhatia, M. Jenkinson, and D. L. G. Hill, "Generalised overlap measures for assessment of pairwise and groupwise image registration and segmentation," in Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI), Lecture Notes in Computer Science, vol. 3749. Springer, 2005, pp. 99–106.

[10] T. F. Cootes, C. Beeston, G. J. Edwards, and C. J. Taylor, "A unified framework for atlas matching using Active Appearance Models," in Information Processing in Medical Imaging, Proceedings, vol. 1613, Lecture Notes in Computer Science, 1999, pp. 322–333.

[11] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, pp. 681–685, 2001.

[12] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," in Proceedings of the European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science, vol. 1407. Springer, 1998, pp. 484–498.

[13] W. R. Crum, T Hartkens, and D. L. G. Hill, "Non-rigid image registration: theory and practice," British Journal of Radiology, vol. 77, pp. 140–153, 2004.

[14] T. F. Cootes, S. Marsland, C.J. Twining, K. Smith, and C.J. Taylor, "Groupwise diffeomorphic non-rigid registration for automatic model building," In Proceedings of the European Conference of Computer Vision 2004, pp. 316–32.

[15] T. F. Cootes, C. J. Twining, V. Petrovic, R. Schestowitz, and C. J. Taylor, "Groupwise Construction of Appearance Models using Piece-wise Affine Deformations." in Proceedings of the British Machine Vision Conference (BMVC'04), Kingston UK, 2004.

[16] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor, "3D statistical shape models using direct optimisation of description length," In Proceedings of the European Conference on Computer Vision, Lecture Notes in Computer Science, vol. 2352, pp. 3–20, 2002,.

[17] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor, "minimum description length approach to statistical shape modeling," IEEE Transactions on Medical Imaging, vol. 21, pp. 525–537, 2002.

[18] R. H. Davies, C. J. Twining, P. D. Allen, T. F. Cootes, and C. J. Taylor, "Shape discrimination in the hippocampus using an MDL model," in Information Processing in Medical Imaging Proceedings, Lecture Notes in Computer Science, vol. 2732, pp. 38–50, 2003.

[19] T. F. Cootes, G. V. Wheeler, K. N. Walker, and C. J. Taylor, "View-based active appearance models," Image and Vision Computing, vol. 20, pp. 657–664, 2002.

[20] T. F. Cootes, G. V. Wheeler, K. N. Walker, and C. J. Taylor, "Coupled-view active appearance models," British Machine Vision Conference, vol. 1, pp. 52–61, 2000.

[21] B. M. Dawant, "Non-Rigid Registration of Medical Images: Purpose and Methods, A Short Survey," In Proceedings of the First IEEE International Symposium on Biomedical Imaging, 2002.

[22] S. Duchesne, J. C. Pruessner, and D. L. Collins, "Appearance-based segmentation of medial temporal lobe structures," NeuroImage, vol. 17, pp. 515–531, 2002.

[23] G. J. Edwards, C. J. T. Taylor, and T. F. Cootes, "Interpreting face images using active appearance models," In IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998.

[24] G. J. Edwards, T. F. Cootes, and C. J. Taylor, "Face recognition using active appearance models," In Proceedings of European Conference on Computer Vision, Lecture Notes in Computer Science, vol. 2. pp. 581–595, 1998.

[25] P. F. Felzenszwalb, "Representation and detection of deformable shapes," Computer Vision and Pattern Recognition, vol. 1, pp. 102–108, 2003.

[26] A. F. Frangi, D. Rueckert, J. A. Schnabel, and W. J. Niessen, "Automatic construction of multiple-object three-dimensional statistical shape models: application to cardiac modelling," IEEE Transations in Medical Imaging, vol. 21, pp. 1151–1166, 2002.

[27] F. Glover, E. D. Taillard, and D. de Werra, "A user's guide to taboo search," Annals of Operations Search, pp. 3–28, 1993.

[28] A. Guimond, J. Meunier, and J. Thirion, "Automatic computation of average brain models," presented at Medical Image Computing and Computer Assisted Intervention 1998, pp. 631–640, Cambridge, MA, USA.

[29] J. V. Hajnal, D. L. G. Hill, and D. J. Hawkes, "Medical image registration," Boca Raton, Fla. ; London: CRC Press, 2001.

[30] S. Haker, A. Tannenbaum, and R. Kikinis, "Mass Preserving Mappings and Image Registration," presented at Fourth International Conference on Medical Image Computing and Computer Assisted Intervention, 2001.

[31] D. W. Hansen, M. Nielsen, J. P. Hansen, A. S. Johansen and M. B. Stegmann, "Tracking eyes using shape and appearance," IAPR Workshop on Machine Vision Applications, pp. 201–204, 2002.

[32] A. Hill, C. J. Taylor, and A. D. Brett, "A framework for automatic landmark identification using a new method of nonrigid correspondence," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp. 241–251, 2000.

[33] T. Jebara, "Images as Bags of Pixels," presented at International Conference of Computer Vision, Proceedings, Nice, France, 2003.

[34] I.T. Joliffe, "Principal Component Analysis," Springer Series in Statistics, Springer, New York, 1986.

[35] S. C. Joshi and M. L. Miller, "Landmark Matching via Large Deformation Diffeomorphisms," IEEE Transaction on Image Processing, vol. 9, pp. 1357–1370, 2000.

[36] P. Hellier, C. Barillot, I. Corouge, B. Giraud, G. L. Goualher, L. Collins, A. Evans, G. Malandain, and N. Ayache, "Retrospective evaluation of inter-subject brain registration," in Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI), Lecture Notes in Computer Science, vol. 2208. Springer, 2001, pp. 258–265.

[37] J. F. Kenney and E. S. Keeping, “Mathematics of Statistics,” part 2, 2nd edition, Princeton, New Jersey: Van Nostrand, 1951.

[38] A. C. W. Kotcheff and C. J. Taylor, “Automatic construction of eigenshape models by genetic algorithm,” in Information Processing in Medical Imaging, vol. 1230, pp. 1–14, Lecture Notes in Computer Science, 1997. 2005.

[39] K. N. Kutulakos, “Approximate n-view stereo,” in Proceedings of the European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science, vol. 1842. Springer, 2000, pp. 67–83.

[40] A. Lanitis, “PROSOPO - A face image synthesis system,” Advances in Informatics, Lecture Notes in Computer Science, vol. 2563, pp. 297–315, 2003.

[41] A. Lanitis, C. J. Taylor, and T. F. Cootes, “Automatic Face Identification System Using Flexible Appearance Models,” Image and Vision Computing, vol. 13, pp. 393–401, 1995.

[42] A. Lanitis, C. J. Taylor, and T. F. Cootes, “Toward automatic simulation of aging effects on face images,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 442–455, 2002.

[43] L. LeBriquer and J. Gee, “Design of a statistical model of brain shape,” presented at Proceedings of IPMI, 1997.

[44] B. Likar and F. Pernus, “A Hierarchical Approach to Elastic Registration Based on Mutual Information,” Image and Vision Computing, vol. 19, pp.

[45] J. Lötjönen and T. Mäkelä, “Elastic Matching Using a Deformation Sphere,” presented at Fourth International Conference on Medical Image Computing and Computer Assisted Intervention, 2001.

[46] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, “Multimodality image registration by maximization of mutual information,” IEEE Transactions on Medical Imaging, vol. 16, issue 2, pp. 187–198, 1997.

[47] D. Marr, “Understanding complex information processing systems,” Vision, pp. 19–24, 1982.

[48] S. Marsland, C. J. Twining, and C. J. Taylor, “Groupwise non-rigid registration using polyharmonic clamped-plate splines,” presented at Medical Image Computing and Computer-Assisted Intervention, Montreal, Canada, 2003.

[49] E. G. Miller, N. E. Matsakis, and P. A. Viola, “Learning from one example through shared densities on transforms,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Volume 1, 2000, pp. 464–471.

[50] T. Mitchell, “Machine Learning,” New York ; London: McGraw-Hill, 1997.

[51] A. Najmi, R. A. Olshen, and R. M. Gray, “A criterion for model selection using minimum description length,” in Proceedings Compression and Complexity of Sequences, pp. 204–214, Salerno, Italy, 1997.

[52] H. Neemuchwala, A. O. Hero, and P. Carson, “Image registration using entropy measures and entropic graphs,” European Journal of Signal Processing, 2003.

[53] M. Petrou and P. Bosdogianni, “Image processing: the fundamentals,” ISBN 0471 99883 4, 1999.

[54] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, “Image Registration by Maximisation of Combined Mutual Information and Gradient Information,” IEEE Transaction on Medical Imaging, vol. 19, pp. 809–814, 2000.

[55] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, “Mutual information matching in multiresolution contexts,” in Image and Vision Computing, vol. 19, issues 1–2, pp. 45–52, January 2001.

[56] W. H. Press, “Numerical recipes in C/C++ : the art of scientific computing. ” Cambridge: Cambridge University Press, 2002.

[57] M. Rabbani and R. Joshi, “An overview of the JPEG 2000 still image compression standard,” Signal Processing: Image Communication, vol. 17, pp. 3–48.

[58] J. R. Rissanen, “Stochastic complexity in statistical inquiry,” presented at World Scientific Series in Computer Science, Singapore, 1989.

[59] P. Rogelj and S. Kovacic, “Similarity Measures for Non-Rigid Registration,” presented at Medical Imaging 2001: Image Processing.

[60] P. Rogelj, S. Kovacic, and J. C. Gee, “Validation of a nonrigid registration algorithm for multimodal data,” in Proceedings of Medical Imaging 2002, Image Processing, SPIE Proceedings, vol. 4684, 2002, pp. 299–307.

[61] D. Rueckert, A. F. Frangi, and J. A. Schnabel, “Automatic construction of 3-D statistical deformation,” IEEE Transactions on Medical Imaging, vol. 22, issue 8, pp. 1014–1025, 2003.

[62] D. Rueckert, A. F. Frangi and J. A. Schnabel, “Automatic construction of 3D statistical deformation models using non-rigid registration,” presented at Medical Image Computing and Computer-Assisted Intervention 2001.

[63] R. S. schestowitz, C. J. Twining, T. F. Cootes, V. S. Petrovic, C. J. Taylor, and B. Crum, “Assessing the Accuracy of Non-Rigid Registration With and Without Ground Truth,” IEEE International Symposium on Biomedical Imaging (ISBI), 2006.

[64] R. S. schestowitz, C. J. Twining, T. F. Cootes, and C. J. Taylor, “Image Registration by Model Criteria,” Presented in Proceedings of MIAS-IRC Plenary Meeting, pp. 16–17, 2004.

[65] Roy Schestowitz, Carole Twining, Tim Cootes, Vladimir Petrovic, and Chris Taylor, “A Generic Method for Evaluating Appearance Models,” Presented in Proceedings of MIAS-IRC Plenary Meeting, 2006.

[66] J. A. Schnabel, C. Tanner, A. Castellano-Smith, A. Degenhard, M. O. Leach, D. R. Hose, D. L. G. Hill, and D. J. Hawkes, "Validation of non-rigid image registration using finite element methods: application to breast MR images," IEEE Transactions on Medical Imaging, vol. 22.

[67] D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg, and R. M. Leahy, "Magnetic resonance image tissue classification using a partial volume model," NeuroImage, vol. 13, pp. 856–876, 2001.

[68] M. Sonka, V. Hlavac, and R. Boyle, "Image processing, analysis and machine vision," Pacific Grove, Calif. ; London: PWS Publishing, 1999.

[69] M. B. Stegmann, B. K. Ersboll, and R. Larsen, "FAME - A flexible appearance modeling environment," IEEE Transactions on Medical Imaging, vol. 22, pp. 1319–1331, 2003.

[70] M. B. Stegmann, "Analysis of 4d cardiac magnetic resonance images," Journal of The Danish Optical Society, vol. 4, pp. 38–39, 2001.

[71] C. Studholme, D.L.G. Hill, and D.J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," Pattern Recognition, vol. 32, issue 1, pp. 71–86, 1999.

[72] N. A. Thacker, D. Prendergast, and P. I. Rockett., "B-fitting: an estimation technique with automatic parameter selection," presented at British Machine Vision Conference 1996.

[73] P. M. Thompson, M. S. Mega, C. Vidal, J. L. Rapoport, and A. W. Toga, "Detecting Disease-Specific Patterns of Brain Structure using Cortical Pattern Matching and a Population-Based Probabilistic Brain Atlas," presented at Proceedings of the 17th International Conference on Information Processing in Medical Imaging, 2001.

[74] A. Toga, "Brain Warping," San Diego, CA, USA: Academic Press, 1999.

[75] A. Trouve, "Diffeomorphisms Groups and Pattern Matching in Image Analysis," International Journal of Computer Vision, vol. 28, pp. 213–221, 1998.

[76] C. J. Twining and S. Marsland, "Constructing diffeomorphic representations of non-rigid registrations of medical images," presented at Information Processing in Medical Imaging 2003.

[77] C. J. Twining, S. Marsland, and C. J. Taylor, "Groupwise non-rigid registration: the minimum description length approach," British Machine Vision Conference 2004.

[78] C. J. Twining and C. J. Taylor, "The use of kernel principal component analysis to model data distributions," Pattern Recognition, vol. 36, pp. 217–227, 2003.

[79] C. J. Twining, S. Marsland, and C. J. Taylor, "Measuring geodesic distances on the space of bounded diffeomorphims," in Proceedings of the British Machine Vision Conference (BMVC'02), 2002.

[80] C. J. Twining, T. F. Cootes, S. Marsland, V. Petrovic, R. Schestowitz, and C. J. Taylor, "A unified information-theoretic approach to groupwise non-rigid registration and model building." in Proceedings of Information Processing in Medical Imaging (IPMI), Lecture Notes in Computer Science, vol. 3565. Springer, 2005, pp. 1–14.

[81] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," International Journal of Computer Vision, vol. 24, pp. 137–154, 1997.

[82] K. N. Walker, T. F. Cootes, and C. J. Taylor, "Determining correspondences for statistical models of appearance," in Computer Vision - ECCV 2000, Pt I, Proceedings, vol. 1842, Lecture Notes in Computer Science, 2000, pp. 829–843.

[83] Y. Wang and L. H. Staib, "Elastic model based non-rigid registration incorporating statistical shape information," presented at Medical Image Computing and Computer-Assisted Intervention, 1998.

[84] L. Wang, Y. Zhang, and J. Feng, "On the euclidean distance of images," IEEE Trans. Pattern Anal. Machine Intell., vol. 27, pp. 1334–1339,

[85] Y. Wang and L. H. Staib, "Integrated approaches to non-rigid registration in medical images," presented at Workshop on Applications of Computer Vision, 1998.

[86] S. K. Warfield, J. Rexilius, P. S. Huppi, T. E. Inder, E. G. Miller, W. M. Wells, III, G. P. Zientara, F. A. Jolesz, and R. Kikinis, "An Entropy Measure to Assess Nonrigid Registration Algorithms for Statistical Atlas Construction," presented at Medical Image Computing and Computer-Assisted Intervention, 2001.

[87] Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging. 2004 Jul;23(7):903–21.

[88] J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, J. Maurer, C. R., R. M. Kessler, R. J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P. A. van den Elsen, S. Napel, T. S. Sumanaweera, B. Harkness, P. F. Hemler, D. L. G. Hill, D. J. Hawkes, C. Studholme, J. B. A. Maintz, M. A. Viergever, G. Malandain, X. Pennec, M. E. Noz, J. Maguire, G. Q., M. Pollack, C. A. Pelizzari, R. A. Robb, D. Hanson, and R. P. Woods, "Comparison and evaluation of retrospective intermodality brain image registration techniques," Journal of Computer Assisted Tomography, vol. 21, pp. 554–566, 1997.

[89] M. Xu and W. L. Nowinski, "Talairach-Tournoux Brain Atlas Registration Using a Metalforming Principle-Based Finite Element Method," Medical Image Analysis, vol. 5, pp. 271–279, 2001.

[90] C. Yam, M. S. Nixon, and J. N. Carter, "Automated person recognition by walking and running via model-based approaches," Pattern Recognition, vol. 37, 2004.

[91] L. Younes, "Deformations, Warping and Object Comparison: A Tutorial," 2000.

[92] L. Zollei, E. Learned-Miller, E. Grimson, and W. Wells, "Efficient population registration of 3D data," in Workshop on Computer Vision for Biomedical Image Applications: International Conference of Computer Vision (ICCV05), 2005.

[93] B. Zitova and J. Flusser, “Image registration methods: A survey,” Image and Vision Computing, vol.~21, pp. 977–1000, 2003.