Automatic Construction of Parts+Geometry Models for Initialising Groupwise Non-rigid Registration

A thesis submitted to

The University of Manchester

for the degree of

Doctor of Philosophy

in the Faculty of Medical and Human Sciences

Pei Zhang

CONTENTS

Final word count: 20447

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Groupwise non-rigid image registration is a powerful tool to automatically establish correspondences across sets of images. Such correspondences are widely used for constructing statistical models of shape and appearance. As existing techniques usually treat registration as an optimisation problem, a good initialisation is required. Although the standard initialisation—affine transformation—generally works well, it is often inadequate when registering images of complex structures. In this thesis we present a sophisticated system that uses the sparse matches of one or more parts+geometry models as the initialisation. We show that both the model/s and its/their matches can be automatically obtained, and that the matches are able to effectively initialise a groupwise non-rigid registration algorithm, leading to accurate dense correspondences. We also show that the dense mesh models constructed during the groupwise registration process can be used to accurately annotate new images. We demonstrate the efficacy of the proposed system on three datasets of increasing difficulty, and report on a detailed quantitative evaluation of its performance.

DECLARATION

Candidate Name: Pei Zhang

Faculty: Medical and Human Sciences

Thesis Title: Automatic Construction of Parts+Geometry Models for Initialising Groupwise Non-rigid Registration

Declaration to be completed by the candidate:

I declare that no portion of this work referred to in this 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.

Signed: Date: December 1, 2011

COPYRIGHT

The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright")[1] and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see `http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487`), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see `http://www.manchester.ac.uk/library/aboutus/regulations`) and in The University's policy on presentation of Theses.

---

[1] This excludes materials already printed in academic journals, for which the copyright belongs to the mentioned journal and publisher.

ACKNOWLEDGEMENTS

ABOUT THE AUTHOR

Pei Zhang received the BEng degree in aircraft power engineering and MEng degree in signal and information processing from Northwestern Polytechnical University, Xi'an, China, in 2004 and 2007 respectively. He joined ISBE in September, 2008, when he commenced his PhD study. He has published the following papers related to the work in this thesis:

- **P. Zhang** and T. F. Cootes. "Automatic Construction of Parts+Geometry Models for Initialising Groupwise Registration", IEEE Transactions on Medical Imaging, To appear;

- **P. Zhang** and T. F. Cootes. "Automatic Part Selection for Groupwise Registration", In *Proceedings of International Conference on Information Processing in Medical Imaging (IPMI)*, volume 6801, pages 636–647, 2011 (**Best Poster Award**);

- **P. Zhang**, S. A. Adeshina, and T. F. Cootes. "Automatic Learning Sparse Correspondences for Initialising Groupwise Registration", In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 6362, pages 635–642, 2010.

FREQUENTLY USED NOMENCLATURE

Acronyms

AAM   Active Appearance Model, 25

ASM   Active Shape Model, 25

CPS   Clamped-Plate Spline, 39

DP   Dynamic Programming, 45

EM   Expectation-Maximisation, 36

FBM   Forward-Backward Matching, 58

GA   Genetic Algorithm, 30

JS   Jensen-Shannon, 42

MDL   Minimum Description Length, 30

PDM   Point Distribution Model, 36

TPS   Thin-Plate Spline, 60

Variables & Functions

$\{\boldsymbol{p}_{i,0}\}$ All parts generated on the reference I_0, 58

ω A set of arcs, 50

G_{opt} Best parts+geometry model, 59

$\boldsymbol{p}_{i,k}$ Pose of the best match candidate of part i on image I_k, 56

$\mathcal{P}_S$ A set of selected parts, 57

$\boldsymbol{X}$ Positions of a dense set of points, 83

$\boldsymbol{S}_{ij}$ Covariance Matrix, 51

G A parts+geometry model, 50

$f_{ij}(\cdot,\cdot)$ Geometry cost, 50

I_k Image k, 54

K Number of used match candidates, 50

Q_i A measure of localisability of part i, 75

m Number of parts, 50

$\overline{\boldsymbol{X}}$ Mean dense points, 85

$\boldsymbol{d}_{ij}$ Mean separation of two parts, 51

$\bar{I}$ Mean reference image, 53

N_I Number of images, 54

N_{gen} Number of generations, 59

N_G Number of parts+geometry models, 84

N_{pop} Number of populations, 59

$f_i(\cdot)$ Part cost, 50

$M_{t,k}$ Best match of the parts+geometry model G_t on image I_k, 59

$\mathcal{G}$ A set of parts+geometry models, 59

$\boldsymbol{x}$ Centre position, 50

$\boldsymbol{p}$ Pose parameter, 55

R Region of interest, 54

I_0 Reference image, 55

r Lower bound of the number of times that each part is used, 74

C Total cost, 50

U Model utility or quality, 54

t_Q Threshold, 75

t_R Threshold, 75

$W(\cdot)$ Some transformation function, 54

ONE

INTRODUCTION

1.1   Motivation

Statistical models of shape and appearance are an important approach in the field of computer vision. They are a kind of generic method that uses a parameterised model to represent the shape and texture variations of an object of interest. They are widely used in many tasks, such as object recognition, object tracking and medical image analysis. The family of such models is large. Notable examples are the Active Shape Model (ASM) [18, 19], the Active Appearance Model (AAM) [15, 16], the 3D Morphable Model [8, 49, 92] and the Active Blob [80].

Amongst the most successful are the ASM and AAM, because they have become *de facto* standards in practical applications of statistical models of shape and appearance. The literature on the ASM and AAM is extensive. Most of the work aims to improve

- *modelling*—the ability of a model to deal with changes of viewpoint, illumination and occlusion;
- *fitting*—the efficiency and accuracy of matching a model to a set of images;
- *generalisation*—the ability of a model to represent unseen instances of the object, particularly those significantly different from the training set;

In contrast, less attention has been paid to the problem of automatic model construction. However, this is by no means a trivial problem.

To build ASMs and AAMs, we need a set of landmarks accurately placed across a group of training images. Here, a landmark refers to

> *a point of correspondence on each object that matches between and within*
> *populations* [33].

If the training set is small, it is possible to do manual annotation. However, if there are hundreds or thousands of training images (*e.g.* building models for different subjects, each containing many images), manual labelling will become extremely time-consuming and thus is not practical. What's worse, manually labelling 3D images is almost impossible due to the difficulties of visualisation. Furthermore, manual labelling will inevitably introduce bias to landmarks and thus to the resulting correspondence. According to [27], an inappropriate choice of correspondence may lead a model to (1) have poor generalisation; (2) generate implausible instances of the object; (3) use too many parameters to represent the variations of the object. Consequently, it is highly desirable to seek an approach which can *automatically* establish the *optimal* correspondences across large scale or 3D images.

Groupwise non-rigid image registration [20, 42, 43, 50, 83] is thus developed for this purpose. Such techniques generally treat registration as an optimisation problem which is solved with local minimisation strategies. Hence, they are likely to fail without good initialisation. For example, a common approach with groupwise registration is to use a simple affine transformation for initialisation, and then to perform non-rigid registration to an evolving mean to obtain more exact results. Although this may work for images of simple structures, such as faces [20], it can fail hopelessly when registering images of complex structures. An example is the radiographs of the human hand shown in Figure 1.1a, where there are considerable shape variations and multiple similar sub-structures (*e.g.* the joints), resulting in many local minima in the groupwise stage. The affine initialisation is thus insufficient. Figure 1.1b shows the average of a set of hand radiographs after affine alignment. This has failed to register the fingers adequately, leading to a poor mean. Further non-rigid registration onto this mean tends to diverge and make the result even worse.

A more sophisticated approach to initialisation is to deliberately find a sparse set of distinctive and corresponding points (Figure 1.2a), which define a roughly correct deformation field, where groupwise registration can be initialised using interpolation techniques. To automatically obtain those points, we use one or more feature detectors. There are a number of choices available, such as Shape Context [6] or SIFT [64]. However, there are two major problems when applying the feature

(a) Hand Examples   (b) Affine mean
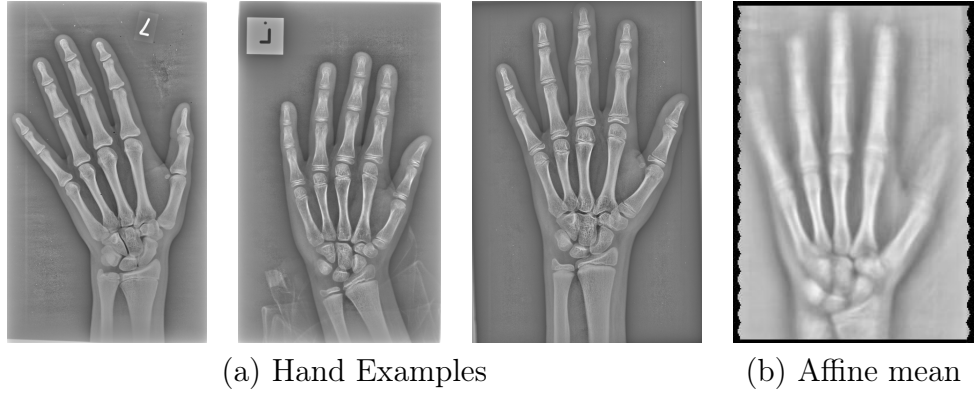
Figure 1.1: Examples of hand radiographs and the mean of a set after affine alignment (right).

(a)   (b)
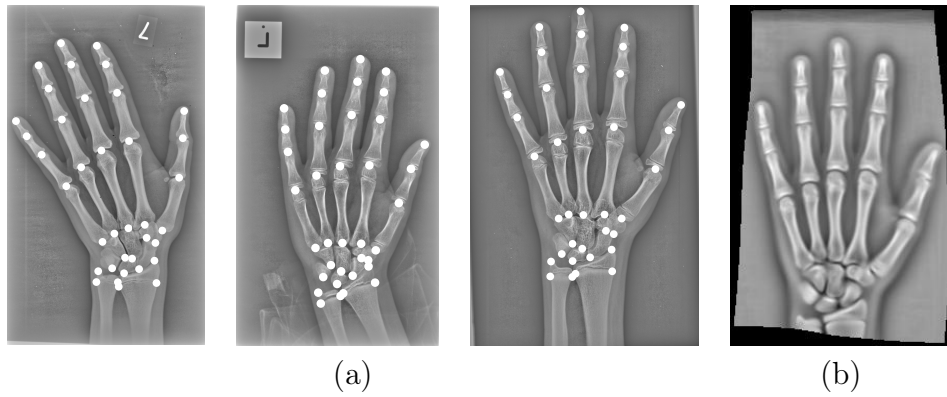
Figure 1.2: (a) An example of a set of distinctive and corresponding points; (b) The mean image by aligning the image set with the points.

detectors directly.

Since almost all the existing feature detectors can result in a large number of points, one problem is how to distinguish the useful points from the others. Once we can obtain a set of distinctive points on one image, we may be able to find their corresponding counterparts on the other images by point matching. This leads to another problem—how to correctly localise those corresponding points, because no detector can guarantee consistent and accurate localisation. In other words, a point on one image may be matched anywhere on another image, no matter if it is distinctive or not. A major reason is the inherent ambiguity in point matching. This may come from the cluttered background, or more commonly the object itself. For example, due to the self-similarity, searching for a point at a joint of a hand may result in many strong responses (Figure 1.3). All may be similar enough to the true joint to pass a threshold when using a suitable classifier. Hence, the correct match may not be identified. A set of inconsistently and incorrectly matched points will,

(a) A joint point   (b) Matches on one image
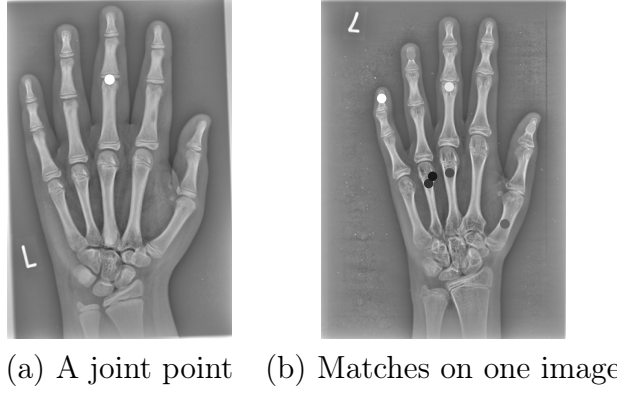
Figure 1.3: Best 10 matches of a joint point on one image using SIFT method. The brightest is the best match and the darker ones are strong competitors. Due to the distraction of similar joints, the method fails to pick out the correct match.

(a)   (b)   (c)
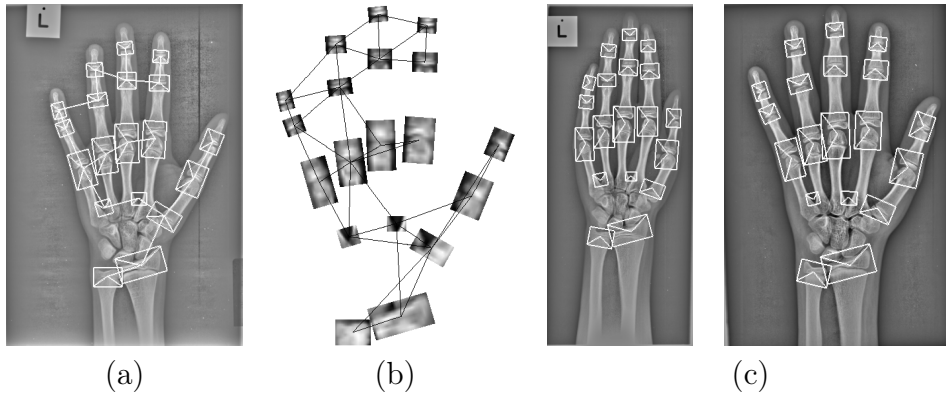
Figure 1.4: (a) A set of geometrically constrained parts; (b) The resulting parts and geometry model; (c) Examples of the best matches of the model.

like the affine transformation, lead to unsatisfactory initialisation.

A solution to the above problems is to introduce extra information to the point selection and matching. A natural choice is the spatial relationship between the points. The parts+geometry models (also called pictorial structures or part-based models) [11, 35, 37, 41], which represent an object using a set of geometrically constrained parts (Figure 1.4a-b), are well suited to our purpose.

To distinguish useful points, we can construct a parts+geometry model using a small set of points, search the image set with the model and examine how often the model gives good matches. A model built from a set of useful points is more likely to give good and consistent localisation than a model from a set of poor ones. Hence, the most useful points can be obtained by constructing a large number of parts+geometry models using different sets of points, comparing the quality of models and selecting the best. With the knowledge of pairwise geometric relationships between points, the parts+geometry model is able to disambiguate the multiple re-
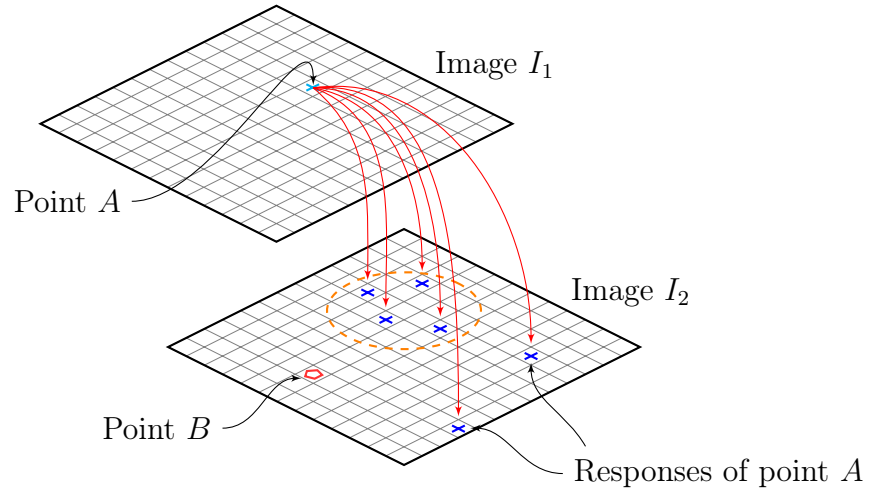
Figure 1.5: An illustration of the effect of geometric constraint between a pair of points. Suppose A and B are two neighbour points on some object of interest. By using some feature detector we can obtain a set of responses for point A on image I_2. Now suppose we are sure about the position of point B on image I_2. Due to the geometric constraint, point A may only appear within a region near to B (indicated by an orange ellipse), leading the responses outside the region to be ignored. This is why geometric relationships can be used to disambiguate multiple responses.

sponses of each point. This is because only the plausible responses of a point can appear within a small region defined by the other point and those outside will be heavily penalised (Figure 1.5). It is the above advantages of the parts+geometry model that motivate us to use it to seek the sparse set of corresponding points and thus good initialisation for groupwise registration.

1.2 Objective

We aim to provide groupwise non-rigid registration with a satisfactory initialisation by using parts+geometry models. The desired method should be fully automatic so as to maintain the unsupervised attribute of groupwise registration. In other words, the sparse set of points which can lead to good initialisation should be obtained without human intervention. Our specific aims are to explore

- the best approach to representing and matching parts;
- an effective way to propose a subset of useful parts from a large set;
- how to automatically construct parts+geometry models;
- how to use parts+geometry models to obtain a sparse set of points which can satisfactorily initialise groupwise non-rigid registration;

- how to initialise groupwise non-rigid registration using a set of points;

- how to evaluate the efficacy of the proposed system.

1.3 Overview of the Methodology

To obtain the desired set of points we perform the following three steps:

(1) constructing a large pool of candidate parts using one image from the given image set;

(2) searching for the match candidates for each part across the image set;

(3) selecting a subset of useful parts which can be used to best initialise groupwise registration.

The positions of the selected parts and their best matches on the other images form the points we are after.

Each part is defined by a pose (position, scale and orientation) and a signature (summarising the local image information). In this work we investigate two different choices for part modelling: intensity patches and SIFT descriptors.

To select the best subset of parts we use parts+geometry models. However, there are two challenges. As the number of the possible combinations of the parts is huge, one challenge is how to explore the numerous resulting parts+geometry models. The other challenge is how to define the term "best", that is, how to evaluate the quality of a model in localising a set of corresponding points.

In this work we use two different approaches to selecting the best parts. One is an *optimisation based* method, in which we use a variant of the Genetic Algorithm (GA) to select the optimal subset [103]. The other is a *voting based* method, where we use a voting strategy to propose a set of good parts and a greedy search to select the desired subset [104]. In both methods we use a minimum description length (MDL) criterion[1] to compute the quality of the parts+geometry models.

Both of the above methods use a single parts+geometry model for initialisation. Specifically, the selected parts are used to build a parts+geometry model, which is used to localise the best match for each part on each image. The resulting sparse points are then used to initialise a groupwise registration algorithm. However, this

---

[1] The MDL principle is essentially a model selection strategy. Suppose we have a set of model candidates, all of which represent a same dataset. The MDL criterion always chooses the simplest one as the best model, which is a formalisation of Occam's Razor.

single-model initialisation strategy might not be reliable, because the model may fail on some of the images, thus introducing inconsistency to the set of points. We therefore develop another strategy, a *multi-model initialisation* strategy, which uses many different parts+geometry models to do initialisation.

To evaluate the performance of the proposed system, we can simply compare its results with manually labelled landmarks. An alternative way is to use the registered data to annotate new images. This can be done by using the found correspondences to construct statistical models of shape and appearance, which are then matched to new data. If the correspondences are accurately established, the model should give good annotation results.

1.4 Contributions

In this thesis we describe a complete system that uses the sparse matches from one or more parts+geometry model(s) to initialise groupwise non-rigid registration. We show that both the model(s) and the matches can be automatically obtained. We also show that the matches are able to give satisfactory initialisation, leading to accurate dense correspondences. Specifically, our contributions include:

- a novel technique for unsupervised learning of parts+geometry models;
- two different ways to automatically choose the best parts+geometry model;
    - a variant of GA;
    - a voting scheme+a greedy search;
- two different strategies for initialisation;
    - single-model initialisation;
    - multi-model initialisation;
- application of the proposed system to large scale datasets, producing state-of-the-art registration results;

1.5 Experimental Setup

We demonstrate the efficacy of the proposed system on three different datasets of increasing difficulty:
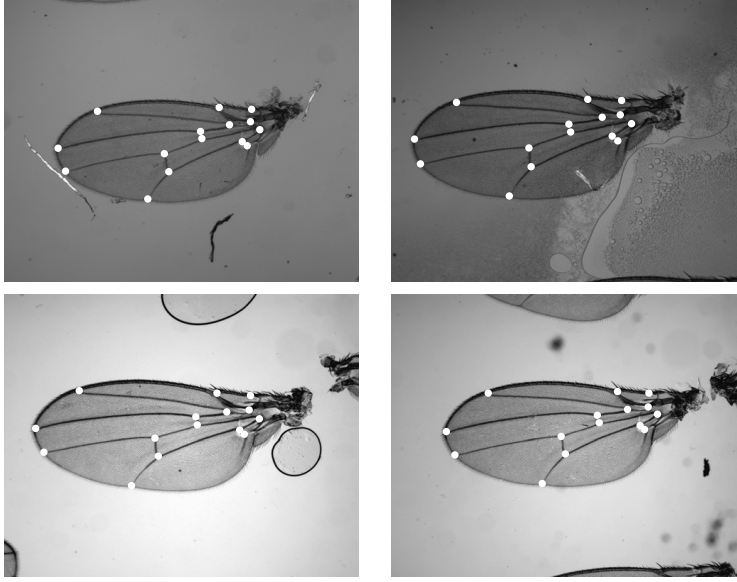
Figure 1.6: Examples of the fly wings and associated landmarks.
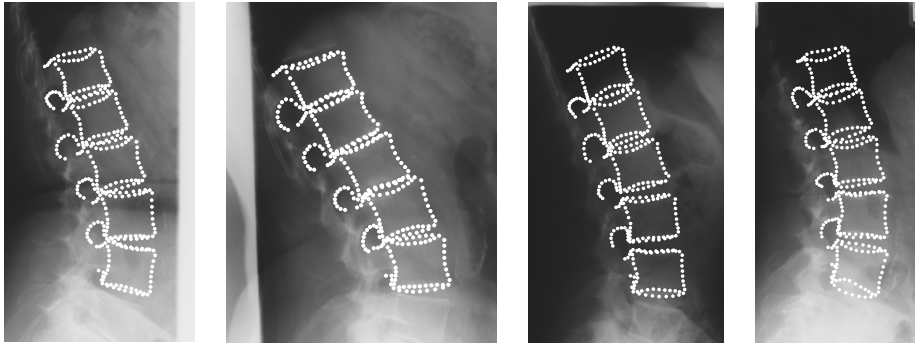
Figure 1.7: Examples of the spines and associated landmarks.

(1) 200 digital micrographs of female fly wings. Each image has a size of 1280×1022 pixels and is marked with 15 points by human expert. This dataset generally does not show too much shape variation;

(2) 200 radiographs of the hands of children (aged between 10-13), taken as part of study into bone ageing. The image size varies across the set, with a height ranging from 1000 to 1700 pixels. Each image has 37 manually labelled landmarks. The resolution of this set of images is 0.2 mm per pixel;

(3) 200 radiographs of the lumbar spine. The image height varies from 1500 to 1700 pixels. Each image has 337 manual landmarks placed around the outline of the vertebrae. This set of images has a resolution of 0.1694 mm per pixel. This dataset is very challenging as the shape of spines can change dramatically and includes a wide range of orientations.

Examples of these three datasets and their manual landmarks are given in Figures 1.6, 1.2 and 1.7. In all experiments the manual annotations are only used to evaluate the performance of the automatic system.

The whole system was developed using C++ except for the SIFT matching where Matlab was used. No parallel computing was implemented. We reported the running time of the system based on a laptop with an Intel T9300 CPU (2.50 GHz), where only one core was used.

1.6 Outline of the Thesis

In the following chapter we present a variety of methods related to (semi-)automatic construction of statistical models of shape and appearance. We explain the advantages of groupwise non-rigid registration compared to the other methods. We also describe various parts+geometry models and their applications in medical image analysis. Finally, we discuss the difference or the advantages of our work compared to other relevant work.

We introduce a complete system for initialisation in Chapter 3, where we take the optimisation based method as the example to describe the details of the system, including the construction, selection and use of parts+geometry models. We test our system on all three datasets, showing its superiority to the affine initialisation [20] and a supervised approach to constructing parts+geometry models [1].

In Chapter 4 we explain the voting based method. We make a detailed comparison with the optimisation based method, showing that the voting based method can give similar or better results. We also investigate some important parameters, indicating the possibility of using the voting based method to speed up the system.

The multi-model initialisation strategy is covered in Chapter 5, where we detail the motivations behind such strategy. We systematically study the influence of different reference images on the single-model and multi-model initialisation schemes, showing that the latter can give consistently good results regardless of the choice of reference images.

In Chapter 6 we describe the methods of using the dense correspondences resulting from groupwise registration to build statistical models and annotate new images. We show that accurate annotations can be achieved by a simple elastic mesh model initialised with the dense correspondences.

We conclude this thesis in Chapter 7, where we discuss the limitations of the proposed system and point out future research directions.

34

LITERATURE REVIEW

One of the motivations of this work is to use groupwise non-rigid registration to solve the problem of automatic construction of statistical models of shape and appearance. But, groupwise non-rigid registration is by no means the only solution. There are a number of other choices available. In this chapter we outline each method and highlight the strength of groupwise non-rigid registration.

Below we first use an example, the AAM [15, 16], to briefly introduce the statistical model of shape and appearance. We then focus on various ways of automatically constructing or learning statistical models in Sections 2.2-2.4. We cover the literature on parts+geometry models in Section 2.5. We review other relevant work in Section 2.6. We conclude this chapter in Section 2.7

2.1   Active Appearance Models

An AAM comprises a *shape* model and a *texture* model. Here, texture refers to

> *the pattern of intensities or colors across an image patch* [16].

To represent an object instance the AAM manipulates the two models to generate the shape s and the texture t of the instance. This can be done by either using the two models separately [68] or using them together as a combined model [16].

Usually the shape model and the texture model are defined as follows [15]:

$$s = \bar{s} + P_s b_s, \qquad (2.1)$$

$$t = \bar{t} + P_t b_t. \qquad (2.2)$$

$\bar{s}$ and $\bar{t}$ are the mean shape and the mean texture of a set of training object instances. $\boldsymbol{P}_s$ and $\boldsymbol{P}_t$ represent the major shape and texture variations across the set. $\boldsymbol{b}_s$ and $\boldsymbol{b}_t$ are the shape and texture coefficients. The shape model (2.1) is also called the Point Distribution Model (PDM) [19].

As we can see from the above equations, the essence of constructing an AAM is to learn a set of model parameters: $\bar{s}$, $\bar{t}$, $\boldsymbol{P}_s$ and $\boldsymbol{P}_t$. Consequently, it does not matter in which way these parameters are learned. This analysis is also applicable to other kinds of statistical models of shape and appearance, ASMs for instance.

Usually we use a set of manually labelled images, each containing an object instance, to learn the parameters of a statistical model. However, there is also considerable work attempting to minimise this human intervention. In general, there are three different kinds of approaches to (semi-)automatically learning model parameters: (1) direct parameter learning; (2) annotation propagation; (3) groupwise non-rigid registration. As the name suggests, the first kind of methods learns parameters directly from the training images. There is no need to explicitly establish correspondences across the training set beforehand. In contrast, the other two methods mimic the manual training process, where correspondences are first established and then used to learn the parameters. Below we review each kind of methods in more detail.

2.2 Direct Parameter Learning

In early work [4] the learning of model parameters is regarded as an image coding problem. The optimal parameters are achieved when the model encodes the entire image set the most efficiently. The learning strategy is to iteratively update one model parameter in turn, by assuming that all other parameters are known.

Kokkinos and Yuille [54] used an expectation-maximisation (EM) algorithm [30] to learn $\bar{t}$ and $\boldsymbol{P}_s$ by regarding $\boldsymbol{b}_s$ as latent variables. In the expectation step $\boldsymbol{b}_s$ is estimated by an AAM fitting algorithm [68] with the current estimates of $\bar{t}$ and $\boldsymbol{P}_s$, while the maximisation step is to update $\bar{t}$ and $\boldsymbol{P}_s$ using the previously computed $\boldsymbol{b}_s$. To initialise the EM algorithm $\bar{t}$ is computed by averaging all the training images in a reference frame, and $\boldsymbol{P}_s$ is obtained by clustering the ridges and edges in the frame onto sharp contours. A similar work can also be found in [53].

De la Torre Frade and Nguyen [29] repeated the following two steps to learn the

model parameters of a kernel based AAM: (1) matching the AAM to the images to estimate the geometric transformation; (2) aligning the images using the estimated transformation to update the model parameters. The initial AAM can be obtained by averaging all the images.

Ferrari *et al.* [40] developed an approach to building a PDM from a group of cluttered images. A set of local shape contours is extracted from each image. A set of model shape contours is then determined by choosing the local contours that have the highest occurrences across the image set. An initial $\bar{s}$ is created by composing the local contours that are most similar to the model contours. $\bar{s}$ is then refined by matching it back to each image and averaging the found matches in a common reference frame. Repeating the above process leads to a polished $\bar{s}$ as well as a set of matches, which can be further used to learn $\boldsymbol{P}_s$.

2.3 Annotation Propagation

The basic idea of this approach is to annotate an image set by first labelling a few images and then propagating the landmarks to the other images. Below we present various approaches to propagation in terms of the labelling workload.

Tong *et al.* [88] used a small set of manually labelled images to annotate the whole image set. The idea is to propagate the landmarks to the unlabelled images using a set of globally consistent pairwise transformations. Each transformation is defined between a common reference and an image in the set. The transformations for the labelled images are computed off-line, while those for the unlabelled are estimated by an iterative procedure. Each iteration involves using the current estimates of transformations to propagate the annotations, and using the propagation to calculate new transformations. The procedure converges when the sum of dissimilarity between each pair of images in the set is minimum. Later, Liu *et al.* [61] extended the above work by propagating a set of contours defined by the manual landmarks.

Langs *et al.* [56] annotated a group of images by only labelling one image. A number of interest points are extracted on a reference image using the Canny edge detector. A subset of points is initially selected using the Shape Context descriptor [6], and their matches on the other images are also localised by the same descriptor. The positions of the points on each image are then modified by a PDM built in a leave-one-out manner, and further refined using an MDL principle. Repeating this

procedure leads to a set of correspondences, which can be used to propagate the landmarks on one image to the other images. A coarse-to-fine strategy may be used to further improve the accuracy of the correspondences.

Donner *et al.* [32] also used a single annotation but labelled the whole image set in a different way. A set of parts on a reference image is carefully selected with the help of interest point detectors. A densely connected parts+geometry model is then built and used to search the other images for the best match on each. By ranking the set of best matches, the model is rebuilt with the top ranked match. The new model is used to search the images that are not used in the model construction, and then rebuilt as above. Once all the images have been considered, a final model is obtained and used to label the whole image set. A similar work can be found in [1].

Walker *et al.* [93, 95] attempted to automatically annotate an image sequence. The approach is to first localise a set of distinctive points on one frame and then detect them throughout the sequence. Each pixel within a user defined boundary is represented by a set of partial Gaussian derivatives, which form a feature vector. All of the feature vectors at each pixel result in a feature space, where the distinctive points can be localised at areas of lowest probability density. By assuming that the object does not undergo significant scaling changes across the sequence, the equivalent points on the rest of frames are obtained by a local search.

Later, the above work was extended to find correspondences across a group of spatio-temporally unrelated images [94]. The same technique is used to seek the salient points between each pair of images. A set of pairwise transformations is then estimated using the found points. To establish correspondences a rectangular mesh is placed on each image. For each mesh, the positions of its nodes are iteratively updated by computing a weighted mean of all other meshes, each aligned to the current mesh using the previously estimated pairwise transformations.

2.4 Groupwise Non-rigid Registration

The motivation behind groupwise registration is to construct an unbiased atlas (or mean) for a group of images/curves of the same object. The atlas can be used to study the structural variability of the object across the set. To compare the atlas with each image/curve we also need to estimate a set of deformation fields during groupwise registration, which can be used to warp each image to the atlas frame

for easy comparison. As the deformation field is usually defined by a set of control points, groupwise registration is equivalent to establishing correspondences across the image/curve set.

Below we focus on the literature on groupwise non-rigid image registration upon which this work is built. The basic idea of such method is to first estimate an initial atlas by registering each image to a common reference and then to refine the atlas based on some criterion. Common choices of the reference image include: (1) an arbitrarily selected image; (2) an evolving group mean; (3) an implicit image. We review various approaches according to the choice of the reference in Sections 2.4.1-2.4.3. We also cover some relevant work on groupwise registration of curves/surfaces in Section 2.4.4.

2.4.1 Registering to an Arbitrary Image

Guimond *et al.* [46, 47] described a method of computing an average model for a group of images. A demons algorithm [86] is used to warp every image to a fixed reference image. Averaging the aligned images leads to a group mean, which can be used to compute the desired model with the help of a mean deformation.

Rueckert *et al.* [76, 77] performed B-spline registration [78] between a chosen reference and each free image. The resulting dense correspondences are used to build a variant of PDM to study the anatomical structure variations across a population.

Cootes *et al.* [17] used groupwise registration to seek dense correspondences across an image set. An affine transformation together with a set of simple warps are used to ensure diffeomorphic deformation between each pair of images. By randomly choosing an image as the reference, pairwise non-rigid registration is performed to find the initial correspondences. A groupwise stage is then used to iteratively refine those correspondences by optimising a minimum message length [75] based objective function, leading to the final results.

Different from above, Marsland *et al.* [66] performed groupwise registration using a flexible reference. A set of control points together with clamped-plate splines (CPSs) [89] are used to represent the deformation fields between each image. The positions of the control points are iteratively updated by minimising an objective function, which measures the sum of dissimilarity between each pair of images. At each minimisation stage, the algorithm repeats the following two steps: (1) warping

each free image to the current reference to update the set of deformation fields; (2) selecting the image as the new reference which is the most similar to the others under the current estimates of deformation.

2.4.2 Registering to an Evolving Mean

Early work by Vetter *et al.* [91] computes pixelwise correspondences across a group of images. An arbitrary image is selected as the reference image, and correspondences between the reference and other images are initially estimated using an optical flow technique. An initial group mean is then computed using those correspondences. A linear model of shape and appearance is constructed using the current estimates of the correspondences, and fitted to each image in the set. Correspondences between each image and its model match are computed using optical flow and composed with those matches, leading to the new estimates of the correspondences across the set. The mean is recomputed and taken as the new reference. The final results are obtained when a stable mean is achieved.

Bhatia *et al.* [7] constructed an atlas using B-spline registration [78] while minimising a cost function based on normalised mutual information [85]. The optimisation is achieved by using a gradient projection method under the constraint that the overall displacements of the control points across the set are zero.

Twining *et al.* [90] regarded groupwise registration as a model selection problem, to which the MDL principle is well suited. An iterative algorithm is used to find the optimal correspondences. Each iteration involves (1) computing a group model, which consists of a mean image, a set of spatial warps and a set of residual images; (2) updating the deformation fields (represented by CPSs [89]) across the image set such that the description length computed from the current model is minimum.

Cootes *et al.* developed a similar method in [20, 21]. The differences are (1) piecewise affine transformation is used to define the deformation fields; (2) a simple statistical model of shape and appearance is built and used to calculate the description length.

Sidorov *et al.* [83] extended the approach in [21] by working in a low dimensional space. The idea is to use a small set of control points to iteratively estimate the increments of the deformation fields. The number of points is fixed during the whole process, but the points themselves are randomly generated at each iteration.

Joshi *et al.* [50, 63] addressed the problem of building an unbiased atlas for an image set which undergoes large diffeomorphic deformation, where the arithmetic mean of the set is no longer an appropriate estimate of the desired atlas. An objective function is used to compute the intensity and deformation differences between the current atlas and each free image. An iterative procedure—either a fluid flow algorithm [50] or a variational algorithm [63]—is used to minimise the cost function to get the final atlas. To initialise the procedure the arithmetic mean is used. However recently, Wu *et al.* [101] argued that a sharp mean should be used instead as the arithmetic mean is usually too blurred to guide the subsequent registration.

2.4.3 Registering to an Implicit Image

Learned-Miller [58] proposed to jointly align a set of images without explicitly using a reference image. This is achieved by minimising the sum of entropies of pixel stacks, each of which is a collection of intensity values sampled at an equivalent location across the image set. The underlying assumption is that if all images are well aligned to each other, low entropy can be expected amongst a majority of pixel stacks. The deformation fields between each image are iteratively estimated until convergence. Later, Huang *et al.* [48] extended this framework to register real world images (*i.e.* faces) instead of those well-controlled (*e.g.* handwritten digits in [58]). As it is no longer suitable to calculate entropy based on intensity values for real world images, a set of clustered SIFT descriptors [64] is used instead. Another extension could be found in [5], where the deformation fields are represented by B-splines [78] rather than the affine transformation as used in [58].

2.4.4 Groupwise Registration of Curves/Surfaces

Davies *et al.* [24–28] developed an information-theoretic approach to computing the optimal correspondences across a set of manual contours. The initial correspondences are established by parameterising each contour to a topological primitive, for instance, a circle for 2D [24–27] or a sphere for 3D [28]. A statistical shape model (*i.e.* a PDM) is then constructed using the correspondences, and its ability to represent the whole set of contours is computed as a description length [75]. The optimal set of correspondences is obtained by iteratively parameterising each contour until a minimum description length is achieved.

Frangi *et al.* [42, 43] described a method of computing 3D dense correspondences. An arbitrary shape is selected as the initial atlas, and all shapes are warped to the current atlas using a global quasi-affine transformation [84] as well as a local B-spline registration [78]. The current atlas is then updated by averaging the set of aligned shapes. Repeating this process leads to the final atlas, where a dense set of points can be generated by a marching cubes algorithm [62]. The correspondences are established by propagating the points onto each individual shape.

Wang *et al.* [96] considered using Jensen-Shannon (JS) divergence to register multiple sets of points. The idea is to minimise the difference (or divergence) between the probability density function of each point set (atlas-aligned) and that of the atlas. As the JS divergence is not well defined in continuous space, the difference is measured based on the cumulative distribution function of each point set, which is approximated by the Parzen window technique. The objective function is optimised by a gradient based strategy.

2.5 Parts+Geometry Models

The parts+geometry model is popular in the object recognition community, and is increasingly applied to medical imaging. Its power lies in the use of a flexible configuration of the object and a probabilistic inference network. The model is generative and can be learned in either a supervised or an unsupervised fashion. As the literature is vast, we confine ourselves to those most relevant. A general overview of the field can be found in [73]. Below we first describe a variety of parts+geometry models for object recognition, and then briefly introduce their applications in medical image analysis.

2.5.1 Object Recognition

In early work [11–13] a set of manually selected parts together with a shape constraint is used to localise/recognise objects of the same class (*e.g.* faces from different persons). The shape constraint is a joint probability density function of the spatial relationships between the parts. To localise/recognise an object, a set of hypotheses is first created to describe its possible positions. Each hypothesis comprises a set of candidate locations for each part, which are generated using some point detector.

The best hypothesis used for localisation/recognition can be selected by using the shape constraint to compute the likelihood of each hypothesis and ranking them all.

Weber *et al.* [97–99] extended the above work by automating the construction of the parts+geometry model. Useful parts are extracted by an interest point operator and selected using a standard k-means clustering algorithm. An initial model is then determined by an exhaustive search in a greedy style, where all possible models of a fixed number of parts (*i.e.* 3) are evaluated on a validation dataset and ranked by performance. In this process an EM algorithm is used to learn the model parameters. To improve the representative power of the initial model more parts may be added by repeating the above process.

Fergus *et al.* [37] further developed the work of Weber *et al.* by considering large scaling changes as well as the appearance variation (which Weber *et al.* omitted). The Kadir and Brady detector [51] is used to localise salient parts over scale. Each of the identified parts is then cropped from the image and bound to a small patch (*e.g.* 11×11 pixels). The appearance variation can be obtained by applying the Principal Component Analysis to a collection of those patches from the whole training set.

Later, Fergus *et al.* made further extensions to their work in [38] and [39]. A combination of feature detectors (*e.g.* curves, Kadir and Brady detector) rather than a single detector is used to localise the salient parts [38]. A simple star model instead of the fully connected constellation model is used to reduce the model complexity [39].

In [35] Felzenszwalb and Huttenlocher revisited the pictorial structures algorithm [41]. Their main contributions are (1) developing an efficient inference algorithm for the pictorial structures of an acyclic graph (*i.e.* a tree), and (2) reformulating the objective function of pictorial structures from a statistical point of view (*i.e.* maximum a posteriori probability).

As tree-structured models have limited representative ability, Crandall *et al.* [22] introduced to the above framework another kind of graphical structures—k-fans, which can be more powerful and yet computationally tractable. A k-fan model has k reference parts, which are fully connected. Every other part is only connected to the reference parts. By varying k one can manage the representative ability of the model.

To automatically learn the k-fan models Crandall and Huttenlocher proposed an EM based approach in [23]. A set of candidate parts is randomly generated and

then pruned according to their ability to distinguish positive and negative training examples. A number of model hypotheses are then created by first randomly generating a set of reference parts and then adding non-reference parts which are highly likely to be connected to the reference parts. The best model hypothesis is used to initialise an EM algorithm to learn the model parameters.

Zhu *et al.* developed various hierarchical parts+geometry models [105–109]. In [109] the parts+geometry model is constructed by recursively composing simple structures (*i.e.* edges) to larger and meaningful parts (*i.e.* eye contours). Each clique of simple structures (corresponding to a large part) at each level is independently represented by a model of the same form with the full parts+geometry model.

They attempted to introduce probabilistic grammars (*e.g.* AND, OR), which are widely used in natural language processing [65], to the parts+geometry model in [105, 106]. A full model consists of a sole OR node, a set of AND nodes and a set of triplets. The OR node represents a set of parts of the object. An AND node may describe either the background or a part which consists of several hierarchically compositional triplets. A triplet is a triangle of each vertex describing a simple structure, which is extracted by the Kadir and Brady detector [51] and represented by the SIFT descriptor [64]. To construct the full model, an iterative procedure is used to learn the model structure and an EM algorithm is used to learn the model parameters.

Zhu *et al.* gave another example of hierarchical models in [107, 108]. An abundance of low level image information (*e.g.* intensity, gradient, Gabor filters) and shape information (*i.e.* triplets [105, 106]) are used to compose high level parts (*i.e.* patches) and shape contours (*i.e.* collections of triplets), which are then composed in a similar way to form the full model. The model can be defined as a generative model [108] or a hybrid discriminative-generative model [107], and can be learned either in an unsupervised fashion [108] or a weakly supervised fashion (*i.e.* from one single image) [107].

None of the above methods deals with the problem of consistent localisation. They only focus on whether a particular part is present on the object or the background. They do not care whether the part occurs at corresponding positions of the object. In this work we explicitly address this problem. We show that the parts which are likely to be consistently located can be proposed by a voting scheme.

2.5.2 Medical Image Analysis

Toews and Arbel [87] proposed to use a variant of 1-fan model to study the structural variability across a group of brain images. A common structure is selected as the reference part and all other parts are modelled independently, with respect to the reference part. SIFT [64] is used to extract interesting parts across the image set. A variant of mean shift algorithm [14] is used to select a small set of distinctive parts, which are treated as the non-reference parts and used to learn the model parameters.

Schmidt *et al.* [79] used a fully connected parts+geometry model to localise and label the vertebrae of lumbar spine in 3D magnetic resonance images. The parts are manually selected and their match candidates are localised by classifiers [59]. An A^* search algorithm [69] together with dynamic programming (DP) is used to infer the model.

Donner *et al.* [31] described a similar approach to localising the anatomical structures in 2D images (*e.g.* hand X-rays). The parts are represented by the descriptors computed from local regions based on gradient vector flow [102], and their match candidates are localised around symmetry points. Given a set of parts, a Delaunay triangulation is used to define the graphical structure, where the relative distance and orientation of each pair of parts are modelled. The match of the model to an image is approximated by the algorithm given in [100].

Recently, Potesil [74] presented a method of using the parts+geometry model to locate different organs across a set of 3D whole-body computed tomography images.

Seghers *et al.* [81] used a densely connected parts+geometry model for image segmentation. They compared two different approaches, iterative DP and mean field annealing, to approximating model inference, and showed that the former is better. Babalola and Cootes [3] tackled a similar problem by using AAMs initialised with parts+geometry models.

2.6 Other Work

Langs *et al.* [55] described a method of constructing sparse shape models from unlabelled images, by finding multiple interest points and using an MDL approach to determine optimal correspondences, finding the model which minimises the description length of the feature points. Another related approach was developed by

Karlsson and Åström [52], who built patch models to minimise an MDL function, estimating the cost of explaining the whole of each image using the patches (by including a cost for the regions not covered by patches).

Both of the above approaches represent shape with a PDM. Such representation is useful for local optimisation, but cannot efficiently deal with multiple candidates. By instead learning a parts+geometry model, where the geometry is modelled with a sparse graph, we can take advantage of DP algorithms which can efficiently find the global optima where multiple candidates are present. Like Karlsson and Åström, our cost function is based on explaining the whole of an image region, but in our case this is done by constructing a model of the whole image using non-rigid deformation based on the centres of the parts (see Section 3.1.3 for details). In addition our goal is somewhat different—we seek a sparse set of points which can be used to initialise a local optimisation based groupwise non-rigid registration scheme.

A number of authors [2, 60, 72] also noticed the limitation of global affine transformation to register complex objects. They proposed to perform local registration (*e.g.* affine registration [2]) independently on a set of parts of the object. The resulting local deformation fields are used to compose a good, global deformation field. Usually the parts are manually selected based on some prior (*e.g.* anatomical structures [2]), but automatic methods also exist [60, 72]. Although it is possible for such methods to identify a set of corresponding parts between a pair of images, it is a challenge to localise such parts across a group of images without considering geometric information.

Work on articulated registration [67, 71] deals with certain objects, but requires carefully designed models, specialised to each application. In contrast, our system is generic and can work with a wide range of objects.

2.7 Summary

We have reviewed various approaches to (semi-)automatic construction of statistical models of shape and appearance. The idea of direct parameter learning is appealing, but can only lead to a set of implicitly defined correspondences. It is not clear in what sense these correspondences are optimal. Methods of annotation propagation cannot avoid human intervention. Although automatic methods [93, 94] exist, they can only deal with images of modest scaling and orientation changes. Also, the resulting

correspondences are not optimal. Groupwise non-rigid registration, however, avoids the above problems. It can automatically compute a set of explicit correspondences in some sense optimal (*e.g.* MDL [20], minimum entropy [58]). The correspondences not only can be used to construct statistical models but also can be applied to various tasks, such as image segmentation and data fusion. This is why this work is built upon groupwise non-rigid registration.

THREE

A COMPLETE SYSTEM FOR GROUPWISE REGISTRATION⋆

This chapter describes a complete system for groupwise registration which forms the basis for the following chapters. By "complete" we mean that the only input of the system is a set of unlabelled images and the output is a dense set of corresponding points. All other jobs, such as finding sparse points and using them to initialise groupwise registration, will be done automatically by the system given appropriate parameters.

We first introduce the definition, construction and evaluation of a parts+geometry model. We then describe how to represent and model a part of an object in Section 3.2, where we will present two different kinds of part models: a patch based model and a SIFT based model. To select the best parts+geometry model, or equivalently, the optimal subset of parts, we introduce an optimisation based approach which uses a variant of GA in Section 3.3. We then explain how to use the resulting sparse set of points to initialise groupwise registration in Section 3.4. We examine the performance of the proposed system in Section 3.5. We conclude this chapter with a discussion on the potential problems of the system in Section 3.6.

⋆ Parts of this chapter appeared as "Automatic Learning Sparse Correspondences for Initialising Groupwise Registration" in proceedings of MICCAI 2010 [103] and will appear in IEEE Transactions on Medical Imaging.

3.1 Parts+Geometry Models

3.1.1 Model Definition

A parts+geometry model G represents an object of interest using a set of parts together with a collection of pairwise geometric relationships between each part, which are defined by a set of arcs ω (Figure 3.1a-b).

Before applying G to an image, we require that the match candidates for each part have been localised on that image. This can be done by using each part to search the given image independently and retaining the best K match candidates (see Section 3.2 for detail). The position of a match candidate is determined by its centre position $\boldsymbol{x} = (x, y)^{\mathrm{T}}$. A match of G is thus essentially a combination of match candidates, each of which corresponds to one part (Figure 3.1c).

Let $f_i(\boldsymbol{x}_i)$ be the part cost, which measures the degree of dissimilarity between part i and its match candidate at the position $\boldsymbol{x}_i$. Let $f_{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ be the geometry cost, which calculates the degree of deformation mismatch between the arc formed by a pair of parts i and j and the arc formed by their match candidates at $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. To find the best match of G to the given image, we minimise the following objective function

$$C = \sum_{i=1}^{m} f_i(\boldsymbol{x}_i) + \alpha \sum_{(i,j)\in\omega} f_{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j), \tag{3.1}$$

where m is the number of parts and α affects the relative importance of the part and geometry cost. The best match of G defines a set of best matches for its parts.

(a) (b) (c)
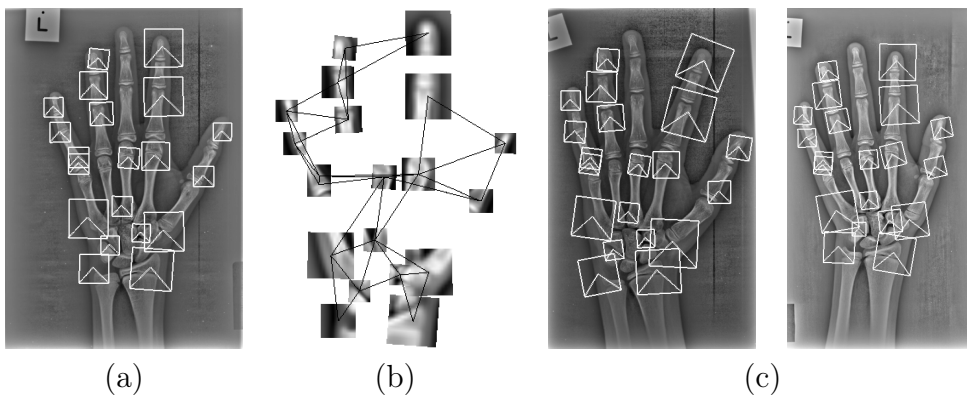
Figure 3.1: (a) A set of automatically generated parts; (b) The resulting parts+geometry model. We only show the model at its mean position, that is, the distance between part i and part j is $\boldsymbol{d}_{ij}$ (see below). The shape variations of the model are not given; (c) Examples of the best matches of the model.

Currently the value of α is determined by preliminary experiments on a small subset of the data. We set $\alpha = 0.1$ for the fly wings and hands and $\alpha = 1.0$ for the spines in all experiments throughout this thesis.

Usually the part cost $f_i(\boldsymbol{x}_i)$ is calculated as the fit value of matching a part. Its computation depends on how the part is modelled. In Section 3.2 we describe two different kinds of part models. One computes $f_i(\boldsymbol{x}_i)$ as the intensity difference and the other computes it as the descriptor difference.

We compute the geometry cost $f_{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ as follows:

$$f_{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j) = ((\boldsymbol{x}_j - \boldsymbol{x}_i) - \boldsymbol{d}_{ij})^{\mathrm{T}} \boldsymbol{S}_{ij}^{-1} ((\boldsymbol{x}_j - \boldsymbol{x}_i) - \boldsymbol{d}_{ij}), \qquad (3.2)$$

where $\boldsymbol{d}_{ij}$ is the mean separation of two parts, and $\boldsymbol{S}_{ij}$ is an estimate of the covariance matrix. Both will be automatically estimated (see Section 3.1.2). The above function is equivalent to assuming that the relative position of the parts follows a Gaussian distribution $\mathcal{N}(\boldsymbol{d}_{ij}, \boldsymbol{S}_{ij})$. It only applies when the orientation and scale of the object are roughly equivalent in each image. If the object is likely to undergo significant scaling or orientation changes across the images, a more sophisticated function can be used.

Solving Equation (3.1) is a combinatoric problem. Arbitrary graphical structures will lead the problem to be NP-hard [10]. It thus becomes impractical to consider all possible matches of G if m and K are large. An approximate solution is usually adopted. However, for simpler structures there are fast, guaranteed optimal solutions. Below we describe how to build parts+geometry models of a tree-like structure for efficiency.

3.1.2 Model Construction

Suppose we have a set of parts and would like to use them to build a parts+geometry model. Before we can construct the model, we have to know its graphical structure and how to connect different parts to create arcs.

As complex graphical structures will incur significant computational difficulty, we only consider simple structures. For our purpose, there are two options, which are shown in Figure 3.2. If the graph is a tree (Figure 3.2a), then each part can be thought of as being connected to one parent, and any number of children. A variant of DP can be used to quickly find the optima in time of order $O(mK^2)$. A

(a) Tree (b) Tree of triangles
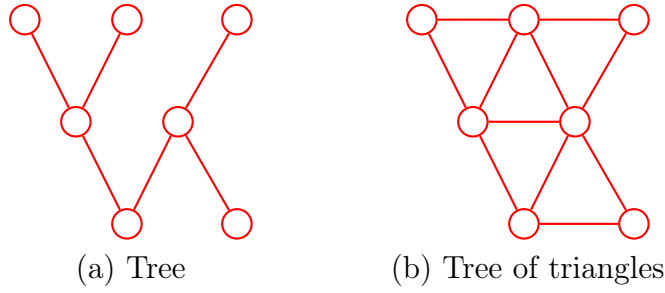
Figure 3.2: Example of two different graph connectivities.

more complex variant is similar to that used by [36], in which a network is created where each node can be thought of as having at most two parents (Figure 3.2b). The optimal solution for this can be obtained with a variant of DP, in $O(mK^3)$ time[1]. If K is modest, this is still fast. The additional links in the graph can lead to a more detailed representation. It is this advantage that motivates us to use the second graphical structure to build the parts+geometry models.

Now we describe how to create arcs for the given set of parts. Suppose we have used each part to localise its match candidates on a reference image (*i.e.* the one where the parts are generated). For each part, there will be a match candidate with minimum fit value. All such match candidates define an initial geometry, where a set of connecting arcs, ω, can be initialised based on the distances between the centres of the match candidates. Specifically, we use a variant of Prim's algorithm for the minimum spanning tree, where each part is connected to two "parents", rather than one. This involves creating the first arc from the two parts which are closest together. We then repeat the following steps until all parts are linked:

- compute the sum of the distances of each unlinked part to the closest two parts in the current linked set;
- select the part which has the minimum such distance, and link it to the two closest parts in the linked set.

The geometric relationships for each arc $(i, j) \in \omega$ (Equation (3.2)) are initialised with Gaussians, with a mean $\boldsymbol{d}_{ij}$ given by the separation in the reference image and standard deviation set to 25% of the length of the arc, that is, $\boldsymbol{S}_{ij} = \frac{1}{16}|\boldsymbol{d}_{ij}|^2\boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix. We then apply this initial parts+geometry model to each image to obtain the best match by minimising Equation (3.1). We group all of those matches and rank them by their fit value C. We use the best 50% of the

---

[1] We use the `mmn` library in the VXL software (`http://vxl.sourceforge.net`).

matches to re-estimate the geometric distributions of the arcs, leading to a robust parts+geometry model.

3.1.3 Model Evaluation

All the best matches of G define a sparse set of points and thus a correspondence across the image set. To choose the best one from the correspondences generated by different G, we have to evaluate the quality of the correspondences, or equivalently, the quality of the models.

We use an MDL principle to evaluate the quality of a model. To be more specific, we perform the following steps:

- Run G on every image to obtain the best match;
- Generate a set of points from each best match and augment it with a set of fixed border points (Figure 3.3b);
- Construct a statistical appearance model from the points and images;
- Compute the average cost of explaining each image using the appearance model.

We generate the set of points based on the centre of the best match of each part. The augmented points impose an artificial correspondence between the borders on each image, but allow piece-wise affine interpolation across the whole of the internal region.

Although detailed shape and intensity models could be used for the appearance model [57], in the following we use a simple mean reference image $\bar{I}$. This is

| (a) | (b) | (c) |
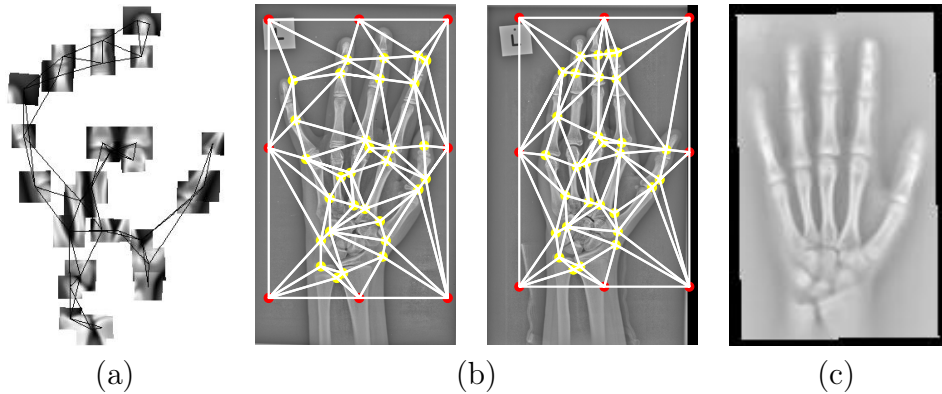
Figure 3.3: Control points and triangulation used for sparse correspondences of the hand: (a) a parts+geometry model; (b) augmented border points (red) and those generated from the best matches (yellow); (c) the resulting mean reference image.

computed as follows:

- Use Procrustes Analysis [19] to align the sets of points and compute the mean;
- Create a triangulation of the mean and use it to warp each image into a reference frame (Figure 3.3c);
- Compute the mean intensity in the reference frame.

To compute the cost of explaining the whole image set with this model, we warp $\bar{I}$ into the frame of each target image, and compute the sum of absolute differences over the region of interest

$$U = \sum_{k=1}^{N_I} \sum_{\boldsymbol{x} \in R} |I_k(\boldsymbol{x}) - \bar{I}(W^{-1}(\boldsymbol{x}))|, \tag{3.3}$$

where N_I is the number of images, $I_k(\boldsymbol{x})$ is the target image intensity at $\boldsymbol{x}$, R is the region of interest in the target frame and $W(\boldsymbol{y})$ is the transformation from $\bar{I}$ to I_k. We call U as the quality or utility of the parts+geometry model G.

According to Shannon codeword length [82], the description length of a value $\hat{x}$ drawn from some probability density function $p(\cdot)$ is given by $-\log(p(\hat{x}))$. By assuming that (1) each pixel is independently and identically distributed; (2) the difference between a pair of corresponding pixels in I_k and $\bar{I}$ follows an exponential distribution, $\exp\{-|I_k(\boldsymbol{x}) - \bar{I}(W^{-1}(\boldsymbol{x}))|/\sigma\}^1$, we can thus compute the description length of evaluating I_k with $\bar{I}$ using all the pixels in R as

$$\mathcal{L} = -\log\left(\prod_{\boldsymbol{x} \in R} \exp\{-|I_k(\boldsymbol{x}) - \bar{I}(W^{-1}(\boldsymbol{x}))|/\sigma\}\right) \propto \sum_{\boldsymbol{x} \in R} |I_k(\boldsymbol{x}) - \bar{I}(W^{-1}(\boldsymbol{x}))| \tag{3.4}$$

Hence, U is an approximation to the description length of evaluating the whole image set with $\bar{I}$.

Note that it is important to measure U in the target frame, and not in the mean reference frame. The latter will be different for different models, so a cost measured there cannot easily be used to compare models. Measuring in the target frame ensures that the same set of pixels is explained every time a different model is evaluated.

---

[1] This is found to be more robust than assuming a Gaussian distribution, which leads to a sum of squares measure.

3.2   Part Models

Now we address the construction of parts. Below we present two approaches to modelling the parts, leading to two different part models. One is a patch based model and the other is a SIFT based model.

3.2.1   Patch based Model

We define each part as a statistical model of the intensities over a fixed shaped region (typically a rectangle or an ellipse). We describe the centre position $\boldsymbol{x}$, scale s and orientation θ of such a region by a pose parameter $\boldsymbol{p} = (\boldsymbol{x}, s, \theta)$.

When applying a part i to an image I, a set of match candidates will be localised, each of which is defined within a local region on I with some pose parameter $\boldsymbol{p}$. Each match candidate has a vector of intensities $\boldsymbol{g}(I, \boldsymbol{p})$, which is sampled from its local region and normalised to have a zero mean and unit variance. The quality of fit when matching part i to a match candidate is computed as follows:

$$f_i(\boldsymbol{g}(I, \boldsymbol{p})) = \beta_i \sum_{j=1}^{n} |g_j - \bar{g}_{ij}|/\sigma_{ij}, \tag{3.5}$$

where $\bar{\boldsymbol{g}}_i$ is the vector of intensities of part i, estimated from a training set, $\boldsymbol{\sigma}_i$ is an estimate of the mean absolute difference from $\bar{\boldsymbol{g}}_i$ across that set, and the subscript j denotes the j-th element of $\boldsymbol{g}(I, \boldsymbol{p})$, $\bar{\boldsymbol{g}}_i$ and $\boldsymbol{\sigma}_i$, namely g_j, $\bar{g}_{ij}$ and σ_{ij}. β_i is a normalisation factor chosen so that the standard deviation of the fits across the training set is unity. We find this form (which assumes the data has an exponential distribution) gives more robust results than normalised correlation, which is essentially a sum of squares measure.

To estimate the parameters $\bar{\boldsymbol{g}}_i$, $\boldsymbol{\sigma}_i$ and β_i for part i, we first arbitrarily choose one image from the given image set as the reference image I_0. We then sample a local region on I_0 to give an initial estimate of $\bar{\boldsymbol{g}}_i$. By setting $\boldsymbol{\sigma}_i$ and β_i to some initial values (*e.g.* **1** and 1), we can search the best match candidate[1] for part i on each image which minimises Equation (3.5) (see below). To obtain a training set to re-estimate the above parameters, we collect the best match candidate from each image, rank the collected candidates by the quality of fit, and retain the best

---

[1]   Note that the difference between the best match candidate and the best match of a part. The former is determined by the part itself while the latter is selected by a parts+geometry model from a set of match candidates.

50% of these[1]. We use the match candidates in the training set to recompute $\bar{\boldsymbol{g}}_i$ by averaging their intensity vectors, and $\boldsymbol{\sigma}_i$ by averaging the absolute difference between $\bar{\boldsymbol{g}}_i$ and each vector. By fixing β_i to its initial value, we fit part i to the training set using the new $\bar{\boldsymbol{g}}_i$ and $\boldsymbol{\sigma}_i$ to recompute the fit values. The new β_i is set to the reciprocal of the standard deviation of those fit values.

Now we explain how to localise a set of match candidates for a part. We perform an exhaustive search at a range of positions, orientations and scales to find the match candidates which minimise Equation (3.5). For large parts this will be very computationally expensive, so we perform a multi-resolution search. We construct parts at a range of resolutions, with those at coarser resolution containing fewer pixels to cover the same image region[2]. We perform an exhaustive search for local optima at the coarsest scale to get a set of match candidates, and then refine each candidate at the finer resolutions. The search is done within a local region. Its centre is defined by the position of the part, and its width and height are identical and proportional to the image size, for instance, $0.4 \arg\min(w, h)$, where w and h are the width and height of the image. Typically, the range of scale is 1.1^i, $i = \{-1, 0, 1\}$, and the range of orientation is [-0.3, 0.3] with step of 0.1. During the search, we systematically vary the orientations and scales of a part to locate the candidates. Where two or more candidates overlap by more than f_o (typically 50%), we retain only the best. This approach allows us to quickly search large regions, usually resulting in a few tens of hypotheses.

To automatically construct a set of candidate parts, we use I_0 to generate a group of patches for a range of sizes, arranged in an overlapping grid pattern (Figure 3.4a). For each patch we use the above process to construct a part, and use it to search the images to find the match candidates used by the parts+geometry models.

As the number of candidate parts is usually huge, many poor ones will inevitably exist. We thus give a simple technique which can be used to select a subset of parts which may have good localisability. This can avoid distracting the algorithm described in Section 3.3.

Given part i, we use it to search each image to find its best match candidate. Let $\boldsymbol{p}_{i,k}$ be the pose of the best match candidate of part i on image I_k. We then define a patch whose size is much larger than that of part i. The intensity of the

---

[1] This may remove the poorly localised candidates and thus avoid their degrading the quality of part models.

[2] We use the `mfpf` library in the VXL software.

(a) (b) (c)
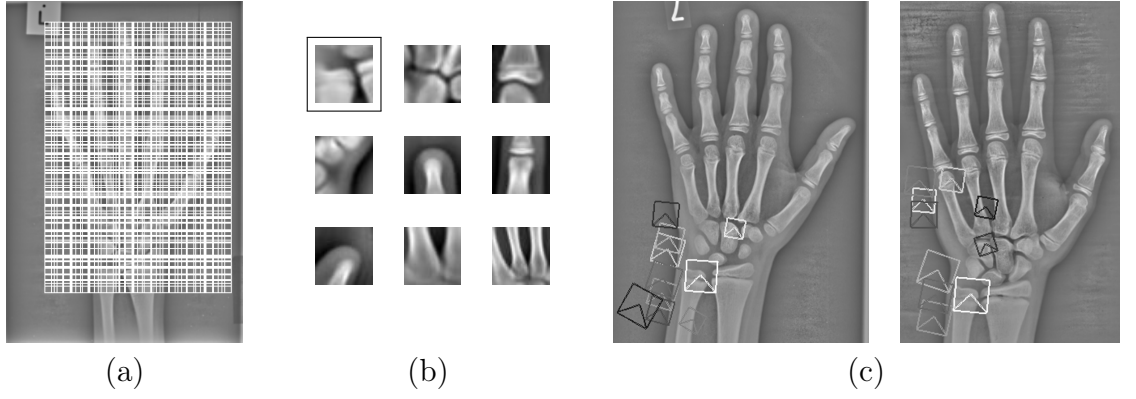
Figure 3.4: Illustration of patch based part models: (a) an overlapping grid; (b) typical part models; (c) best 10 match candidates of a part model (indicated by □) on two images. Intensity indicates rank (brightest=best match candidate).

patch is calculated by sampling each candidate at pose $\boldsymbol{p}_{i,k}$ and then averaging the samples. Let $\boldsymbol{u}_i$ be the intensity vector of a region which is centred at the patch and has the same size with part i. We measure the localisability of part i by

$$q_i = \min_{a \leq |\delta \boldsymbol{x}| \leq b} |\boldsymbol{u}_i(\delta \boldsymbol{x}) - \boldsymbol{u}_i|, \tag{3.6}$$

where $\boldsymbol{u}_i(\delta \boldsymbol{x})$ is the intensity vector of the region when displacing it by $\delta \boldsymbol{x}$ from the patch centre. We set $a = 1$ and $b = 4$ in all experiments throughout this thesis. If q_i is small then part i is likely to represent a flat region and thus is not good at localisation. If q_i is large, then part i may be well localised, suggesting a good part. We rank all the candidate parts by q and retain the best. We use $\mathcal{P}_S$ to denote the set of retained parts. Figure 3.4b shows some examples of the selected parts.

3.2.2 SIFT based Model

We can also represent each part using a SIFT descriptor[1] [64]. Similar to the patch based model, we also use a reference image I_0 to generate the candidate parts. We use two interest point detectors to propose the possible positions of the parts. One is an edge detector and the other is a variant of a local symmetry point detector [31], which returns local minima of a smoothed edge strength image. For each response point we generate parts at n_s different scales, where $s = s_0(\sqrt{2})^i$, $i = 1, 2, \ldots, n_s$ and s_0 defines the scale of the smallest region of interest to use. We also apply the same point detectors to the remaining images to extract interest points, computing

---

[1] `http://www.vlfeat.org`
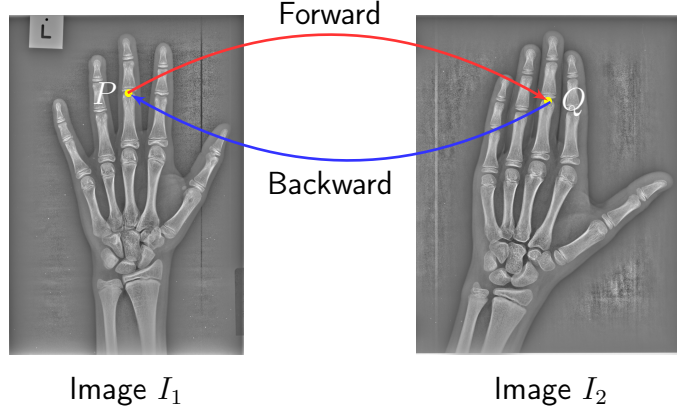
Figure 3.5: An illustration of the forward-backward matching algorithm.

their SIFT descriptors.

We use a variant of forward-backward matching (FBM) algorithm to select, a reduced set $\mathcal{P}_S$, from the whole set of candidate parts $\{\boldsymbol{p}_{i,0}\}$ (see Algorithm 1). The FBM was originally designed to find distinctive points between a pair of images [44]. Given a point P on image I_1 and its best match Q on image I_2, the key idea is that if P is a distinctive point then the best match of Q on image I_1 should be P (Figure 3.5). Here we extend the original FBM for a group of images. The underlying assumption is that a distinctive part is likely to appear in most of images and therefore has more chance to pass the forward-backward test. The fit of a SIFT based part model is the Euclidean distance between the SIFT descriptors of the model and the target point. For each part in $\mathcal{P}_S$, we normalise its fits to all the interest points on one image, rank the points by the fits and select the best K as the match candidates.

Algorithm 1: A variant of forward-backward matching algorithm

Input   : $\{\boldsymbol{p}_{i,0}\}$, I_0
Output : $\mathcal{P}_S$
Initialise: $\mathcal{P}_S \leftarrow \emptyset$, $N_m \leftarrow 0$

for all *candidate parts* **do**
    for all *training images* **do**
        $\boldsymbol{p}_{i,k} \leftarrow \texttt{getBestPointMatch}(\boldsymbol{p}_{i,0}, I_k)$
        $\boldsymbol{p}'_{i,0} \leftarrow \texttt{getBestPointMatch}(\boldsymbol{p}_{i,k}, I_0)$
        if $\boldsymbol{p}'_{i,0} = \boldsymbol{p}_{i,0}$ **then**
            $N_m \leftarrow N_m + 1$

    if $N_m/N_I > 50\%$ **then**
        $\mathcal{P}_S \leftarrow \mathcal{P}_S \cup \{\boldsymbol{p}_{i,0}\}$
    $N_m \leftarrow 0$

3.3 Selecting Useful Parts

As our goal is to find the best combination of the parts, we can regard it as an optimisation problem. We adopt an evolution based optimisation regime. Each subset of parts from $\mathcal{P}_S$ leads to a parts+geometry model, whose utility can be measured by Equation (3.3). We create an initial population by randomly sampling subsets from $\mathcal{P}_S$. For each set we generate a parts+geometry model G using the methods in Section 3.1.2 and then evaluate it to obtain a quality measure U. We then rank the members of the population (each a candidate set of parts) by U. We discard the worst 50%, and generate new candidates from pairs of candidates randomly selected from the best 50%. To generate a new subset we simply randomly sample from the union of parts from the two candidate parent sets. Repeating the

---

Algorithm 2: The optimisation based algorithm

Input : $\mathcal{P}_S$
Output : the best parts+geometry model G_{opt}
Initialise: m, N_{gen}, N_{pop}, $\mathcal{G} \leftarrow \emptyset$

// Randomly generate N_{pop} models of m parts from $\mathcal{P}_S$
$\mathcal{G} \leftarrow \texttt{RandomSample}(\mathcal{P}_S, m, N_{\text{pop}})$
$\{U_t\} \leftarrow \texttt{ComputeUtility}(\mathcal{G})$
$\mathcal{G} \leftarrow \texttt{sort}(\mathcal{G}, \{U_t\})$

for $i_{\text{gen}} \leftarrow 1$ **to** N_{gen} *generations* **do**
    // Retain the best 50% models
    $\mathcal{G} \leftarrow \texttt{RetainBest}(\mathcal{G}, N_{\text{pop}}/2)$

    // Generate another 50%
    for $i_{\text{pop}} \leftarrow 1$ **to** $N_{\text{pop}}/2$ **do**
        // Randomly select two models from $\mathcal{G}$
        $\{G_{t_1}, G_{t_2}\} \leftarrow \texttt{generateCandPairs}(\mathcal{G})$
        // Create a union of parts using the two models
        $\mathcal{P}_{i_{\text{pop}}} \leftarrow \texttt{CreatePartsUnion}(G_{t_1}, G_{t_2})$
        // Generate new model and add it to $\mathcal{G}$
        $\mathcal{G} \leftarrow \texttt{cat}(\mathcal{G}, \texttt{RandomSample}(\mathcal{P}_{i_{\text{pop}}}, m, 1))$

    $\{U_t\} \leftarrow \texttt{ComputeUtility}(\mathcal{G})$
    $\mathcal{G} \leftarrow \texttt{sort}(\mathcal{G}, \{U_t\})$

$G_{\text{opt}} \leftarrow \texttt{RetainBest}(\mathcal{G}, 1)$

Subroutine $\texttt{ComputeUtility}$
for all $G_t \in \mathcal{G}$ **do**
    for all *training images* **do**
        $M_{t,k} \leftarrow \texttt{getBestGraphMatch}(G_t, I_k)$
    $U_t \leftarrow \texttt{computeMDL}(\{M_{t,k}\})$ *// Using Equation (3.3)*

---

above process leads to the best G. We summarise this process in Algorithm 2.

3.4 Initialising Groupwise Registration

The final parts+geometry model is used to search the whole set of images for the best matches for each part. These sparse points are then used to initialise a groupwise non-rigid registration algorithm [20]. A hexagonal grid of control points is placed on one image (Figure 3.6a), and propagated to all the other images using a Thin-Plate Spline (TPS) [9, 34] interpolation based on the sparse points (Figure 3.6b). A mean reference image (Figure 3.6c) is constructed by warping each image into an average shape using the points, and computing the mean. The non-rigid registration
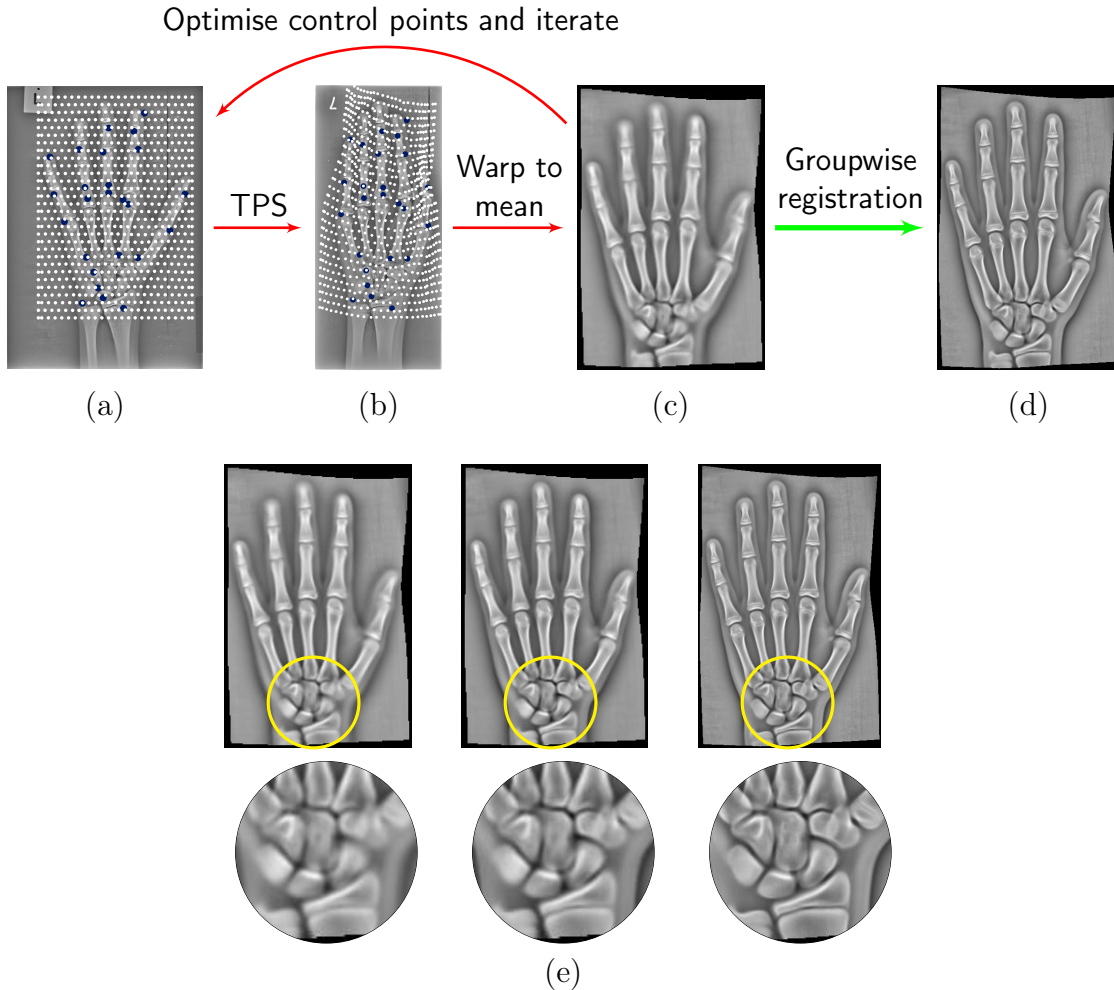
Figure 3.6: Illustration of the process of groupwise non-rigid registration initialised by a single parts+geometry model: (a) a reference image; (b) an example of propagated points; (c) initial mean; (d) final mean; (e) evolution of the mean reference image during groupwise registration. The dark blue points represent the sparse points while the white ones are the dense control points.

method is then used to move the control points on each image so as to optimise the match with the reference image. The reference image is recomputed from the refined control points and the process is repeated. A multi-resolution framework is used, in which the early stages work on coarse images, and later stages refine the result on higher resolution images. The result is a dense correspondence across the whole image set. For details the reader is referred to [20]. Figure 3.6d gives an example of the final mean after groupwise registration. Figure 3.6e illustrates the evolution of the mean reference image during groupwise registration. Closeups are provided for the area inside the yellow circle.

3.5 Experiments

We demonstrate the efficacy of the automatic system on the three datasets introduced in Chapter 1. We randomly selected 100 images from each dataset and used them for the test. Below we first show that affine initialisation is insufficient for groupwise non-rigid registration. We then report a detailed quantitative evaluation of the performance of the system in Section 3.5.2. We focus on the effects of two important parameters: the number of parts m and the choice of part modelling. In Section 3.5.3 we compare the system with a supervised method to further demonstrate its performance.

3.5.1 Insufficiency of Affine Initialisation

We used the fly wings and the hands to demonstrate the problem of affine initialisation. We computed the best affine transformation to register each image with a reference frame using a semi-exhaustive search followed by local optimisation of the transformation parameters. We used the resulting affine transformation to initialise the groupwise registration algorithm described above and ran it on both sets. In each case we applied 30 iterations, where one iteration involves registering every image with the current estimate of the mean, and then updating the mean. We repeated this process with another groupwise registration initialised by a 5-part parts+geometry model obtained using the optimisation based algorithm.

To evaluate the accuracy of the result we compare with a manual annotation. We used the resulting dense correspondence to warp each set of manual landmarks

(a) Micrographs of fly wings

(b) Radiographs of human hand
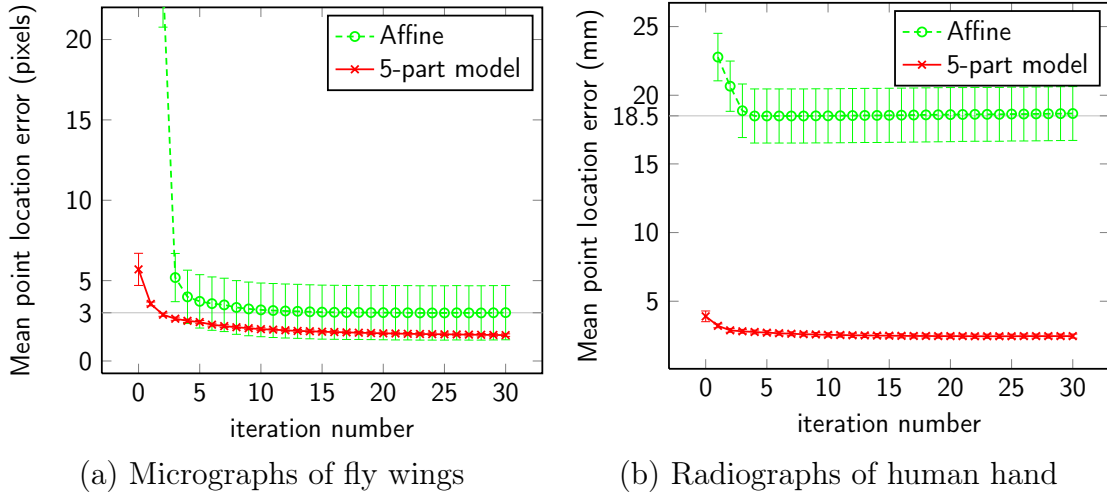
Figure 3.7: Point location errors of the dense correspondence resulting from two different initialisation methods. For the 5-part parts+geometry models, the standard errors in both cases are below 0.15 and therefore are almost invisible in the above figure. Note that how a good initialisation can significantly reduce both point location error and standard error of groupwise registration.

into a reference frame, computing their mean. We then projected this mean back to each individual image and calculated the mean absolute difference of the positions between each mean point and the original manual landmark. We repeated this process with the sparse set of points found by the parts+geometry model, leading to the point location error at iteration 0, which is essentially the error before groupwise registration.

The mean point location error against the iteration is plotted in Figure 3.7. It is not surprising that affine initialisation can work well on the fly wing set as they do not have too much morphological variation. However, the algorithm falls into a local minimum, and further registration cannot improve the result. For the hand set, the ability of affine initialisation is rather limited and further registration can even diverge. In contrast, the 5-part parts+geometry model can significantly improve the registration accuracy though it has a limited representative ability. This demonstrates that the initialisation is crucial to the performance of groupwise registration, and that a sophisticated initialisation is desirable even for those images where a simple one may appear to work.

Figure 3.7 also shows that groupwise registration can further improve the quality of the correspondence given by the parts+geometry model. For the fly wings, the point location error was 5.7 ± 1.0 pixels at iteration 0 and reduced to 1.6 ± 0.08 pixels

when groupwise registration converged. The percentage of improvement is around 72%. For the hand set, the initial error was 3.9±0.4 mm and the error at convergence was 2.5±0.2 mm, approximately 36% improvement.

As seen in Figure 3.7, groupwise registration of 7 iterations is enough to give reasonable results for the fly wings and hands. In addition, its running time is also relatively fast, at around 8 minutes. Below we thus used this 7-stage groupwise registration for all the experiments on the both sets. For the more challenging spine set, we ran one more iteration for luck. Note that these settings were also used in the other experiments throughout this thesis.

3.5.2 Performance of the System

An important parameter is m—the number of parts of a parts+geometry model. The choice of the method of modelling parts can also affect the performance of the system. In the following we first study the effects of these two parameters on the fly wings and hands, and then show the equivalent result on the spine set.

By systematically initialising patches in an overlapping grid on a reference image at a range of sizes, we automatically constructed over 1000 parts for the fly wings and over 1900 for the hands. We then used each part to search the whole set of images. The whole searching process took about 4.3 hours for the fly wings and 8.5 hours for the hands. We ranked all the parts by their localisability using Equation (3.6) and selected the best 250 for the fly wings and the best 500 for the hands.

We also extracted interest points on all images using the point detectors described in 3.2.2. We then computed the SIFT descriptors for each point at different scales. Where points had multiple descriptors at one scale (due to multiple orientations), we used only one of them. For each point at each scale, we constructed a SIFT based part model. We used Algorithm 1 to select the best parts, which are shown in Figure 3.8. In the experiment we set $s_0 = 8$ and $n_s = 3$ for both sets. The whole process took about 10 minutes and 30 minutes for the fly wings and the hands respectively.

We generated parts+geometry models using both kinds of part models, to allow us to compare their merits. We set $K = 50$ in all experiments throughout this thesis. We used the optimisation based method to select the best parts+geometry models, for a range of different numbers of parts m. The resulting sparse points were used

(a) Fly wings   (b) Hands
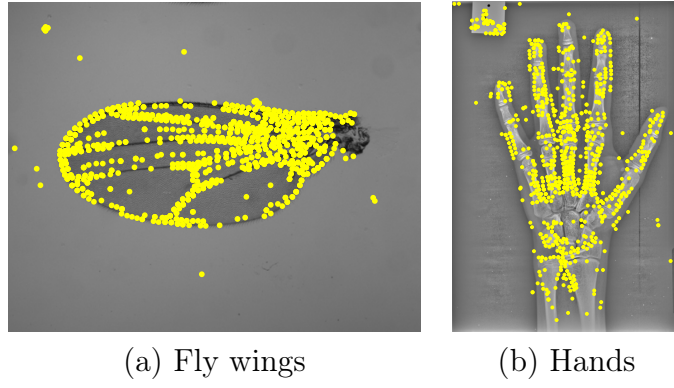
Figure 3.8: Examples of the SIFT based part models selected by the FBM algorithm.

(a) P+G Model   (b) Initial Mean   (c) Final Mean
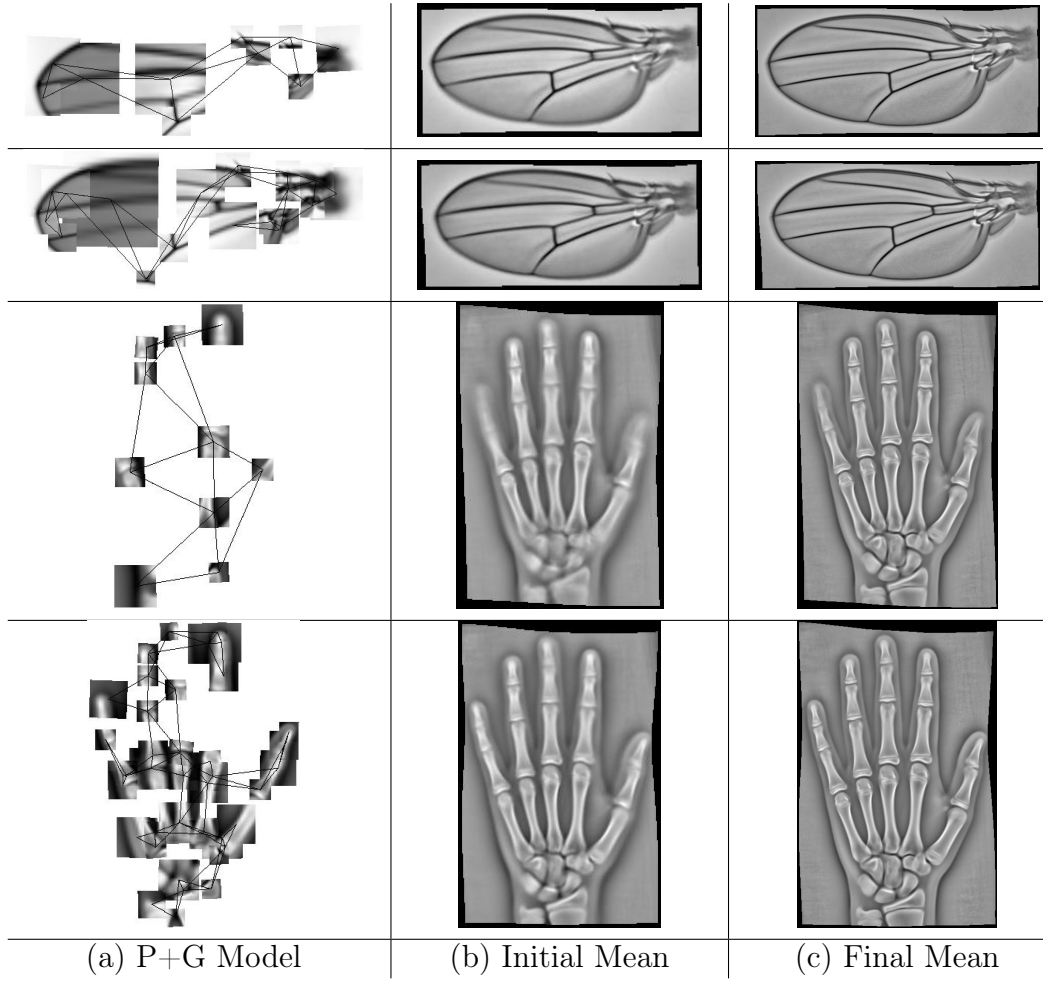
Figure 3.9: Examples of the best parts+geometry models (constructed from patch based parts) and resulting registration with different numbers of parts. Top: 10 and 20 parts for the fly wings. Bottom: 10 and 40 parts for the hand set.

to initialise groupwise registration. Examples of the resulting models for the patch based parts are shown in Figure 3.9a. The initial mean and final mean during the groupwise registration are shown in Figure 3.9b-c. Equivalent results for the SIFT based parts are shown in Figure 3.10a-c.

| (a) P+G Model | (b) Initial Mean | (c) Final Mean |
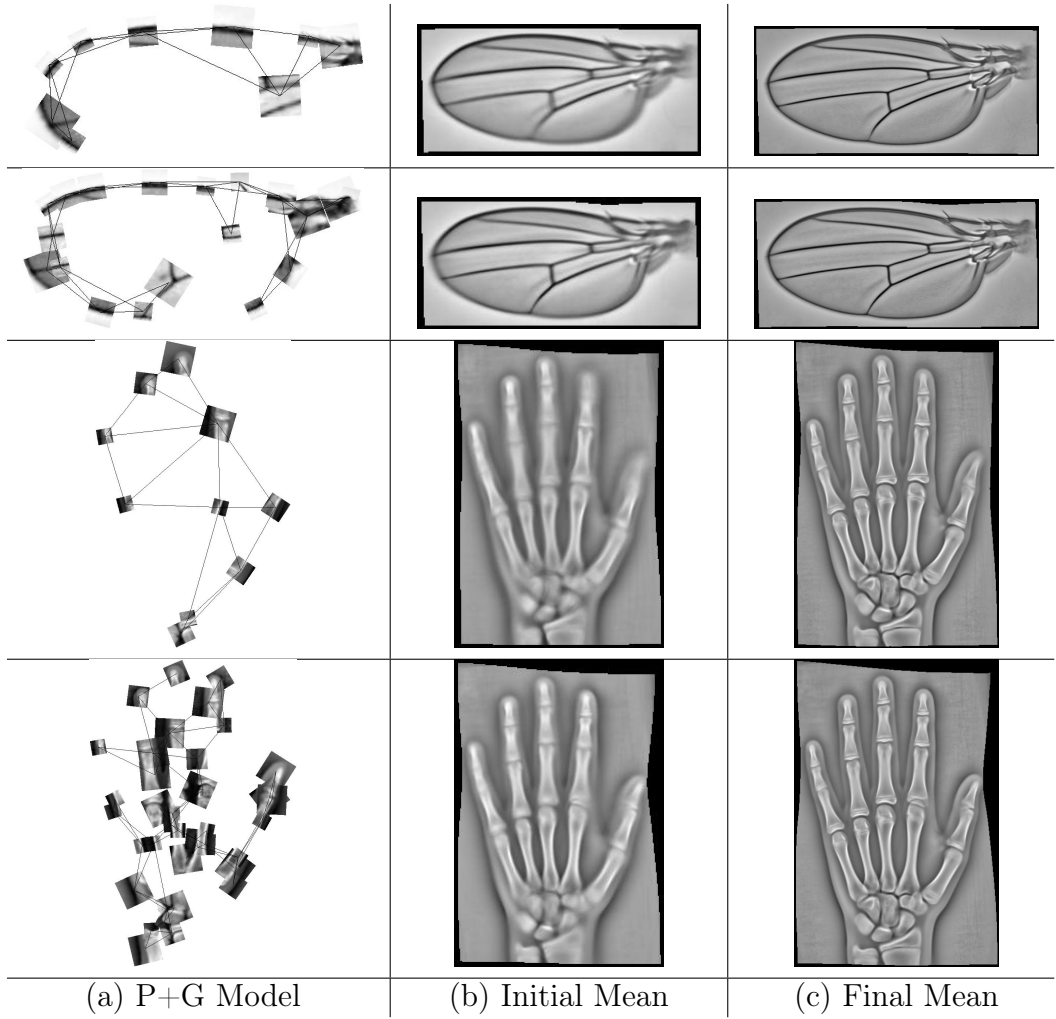
Figure 3.10: Examples of the best parts+geometry models (constructed from SIFT based parts) and resulting registration with different numbers of parts. Top: 10 and 20 parts for the fly wings. Bottom: 10 and 40 parts for the hand set.

The same protocol (point-to-point location error) was used to compute the registration error for both sets. Results for different numbers of parts are given in Table 3.1 and Table 3.2. We find that the patch based part models generally lead to better results than their SIFT based counterparts, though locating the former is much more computationally expensive. Additionally, there is not a clear relationship between the number of parts and performance—once sufficient parts are available to cover the main components of the object, adding further elements is unlikely to improve performance and may lead to a decline.

We also give the typical running time (hours) of the optimisation based approach in Table 3.1 and Table 3.2. The timing was based on the patch based part models. As we used the same parameters to select both kinds of part models for each dataset, the running time of the optimisation based method was similar for both models.

3. A COMPLETE SYSTEM FOR GROUPWISE REGISTRATION

Table 3.1: Point location errors (pixels) of the dense correspondence for the fly wings.

	Patch based			SIFT based			
m	Mean±s.e.	Median	90%	Mean±s.e.	Median	90%	Time
5	2.2±0.09	1.8	3.4	3.5±0.7	1.9	5.0	0.4
10	1.8±0.09	1.5	2.7	2.3±0.1	2.0	3.8	1.2
15	**1.6±0.1**	**1.4**	**2.2**	2.0±0.1	1.7	3.1	2.0
20	1.8±0.09	1.5	2.8	1.9±0.1	1.4	3.3	3.1
25	1.9±0.08	1.7	3.1	2.4±0.3	1.7	3.7	3.7

Table 3.2: Point location errors (mm) of the dense correspondence for the hand set.

	Patch based			SIFT based			
m	Mean±s.e.	Median	90%	Mean±s.e.	Median	90%	Time
10	1.4±0.08	1.2	2.6	1.7±0.2	1.1	2.7	3.3
20	1.2±0.08	0.9	2.0	1.4±0.1	1.0	2.1	5.2
30	1.2±0.09	0.9	1.9	1.4±0.1	1.0	1.9	8.0
40	**1.0±0.05**	**0.8**	**1.5**	1.7±0.1	1.1	3.4	10.3
50	1.0±0.06	0.8	1.6	1.5±0.1	1.2	2.7	13.0

Note that little time difference may exist because some of the parts may have fewer match candidates than what was considered in the experiment.

For the spine set, we only constructed patch based part models. Searching with all parts (over 900) took 5.2 hours. We retained the best 100 parts and repeated the above experiment. We evaluated the accuracy by computing the mean distance between each point and the nearest point on the equivalent curve on the manual annotation—this allows for "sliding" along curves in the manual annotation. This is found to be a more useful measure of the performance than direct point-to-point distance[1]. An example of the resulting parts+geometry models for the spine set is shown in Figure 3.11a. The initial mean and the final mean during the groupwise registration are shown in Figure 3.11b-c.

Table 3.3 shows the statistics of the resulting point-to-curve errors for different numbers of parts. The figures are confounded by considerable number of failures on difficult images, as well as the inherent ambiguity in model position—on some images the model can be shifted up by one vertebra. Despite this, good results were obtained. Table 3.3 also lists the typical running time (hours) of the optimisation based method on this dataset.

---

[1] Most of landmarks placed on the hands and fly wings are at positions that can be well localised (*e.g.* at corners), so a point-to-point measure is appropriate. The majority of the points on the spines are annotated along edges, and thus there is some ambiguity about exactly where on the curve of the edge they should be placed. A point-to-curve measure is thus more appropriate for them.

| (a) P+G Model | (b) Initial Mean | (c) Final Mean |
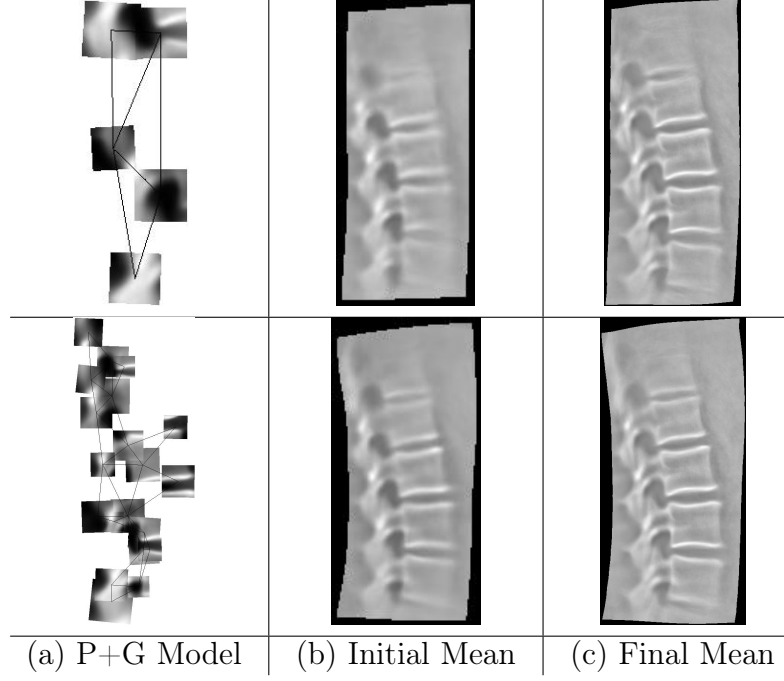
Figure 3.11: Examples of the best parts+geometry models and resulting registration with different numbers of parts. We only show the cases when $m = 5, 20$.

Table 3.3: Point-to-curve location errors (mm) of the dense correspondence for the spine set.

m	Mean±s.e.	Median	90%	Time
5	6.7±0.9	2.4	25.0	0.2
10	**3.2±0.5**	**1.9**	**4.1**	0.5
15	4.1±0.5	2.6	5.1	0.6
20	4.1±0.5	2.6	5.2	0.7

To visualise the performance of the system, we give examples of the projection of the mean annotation onto individual images in Figures 3.12 and 3.13. In general, our system can deal well with the challenging hand set, though mistakes may be made on some images. For the spine set, we get a good overall fit in many images, though there are commonly problems at the very top and bottom of the model (see the blurred regions in Figure 3.11). Our system makes a brave attempt at difficult examples on this challenging dataset (see the bottom row of Figure 3.13). On the fly wing set since our system works well, with no visual failures.

3.5.3 Comparison with Supervised Method

We compared our automatic system with a supervised method [1]. For each dataset, we manually chose a set of parts on the same reference image and constructed

Good results

Failure modes
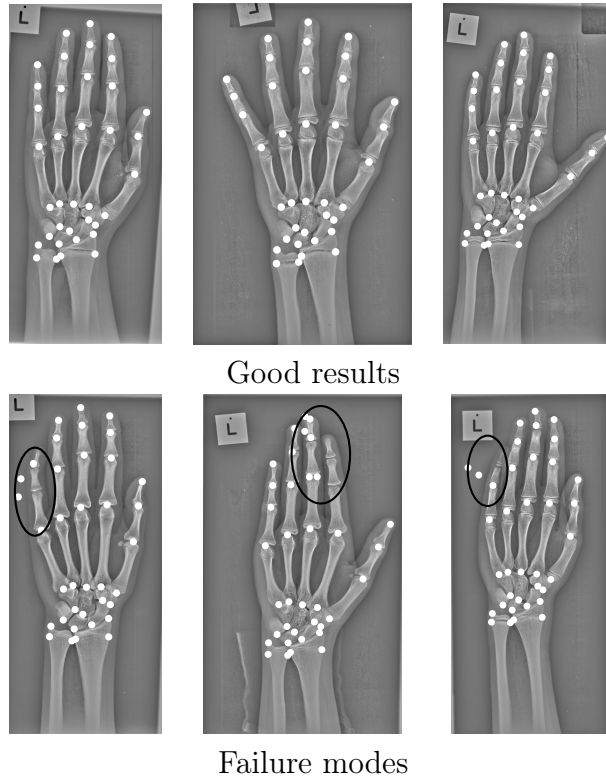
Figure 3.12: Projection of average points onto different hand radiographs (using 40 parts model). Failures are indicated by black ellipses.

Table 3.4: Point-to-point location errors (pixels) of the dense correspondence for the fly wings.

Method	m	Mean±s.e.	Median	90%
Affine	–	3.5±1.7	1.8	2.5
Manual	9	2.1±0.2	1.2	4.0
Automatic	15	**1.6±0.1**	**1.4**	**2.2**

Table 3.5: Point-to-point location errors (mm) of the dense correspondence for the hand set.

Method	m	Mean±s.e.	Median	90%
Affine	–	18.5±2.0	9.9	41.8
Manual	20	**1.0±0.05**	**0.8**	**1.3**
Automatic	40	1.0±0.05	0.8	1.5

a parts+geometry model using those parts (Figure 3.14). Following [1], we used the model to search all images to localise the sparse correspondence, which was used to initialise the same groupwise registration algorithm. Figure 3.15 gives the examples of the sparse matches as well as the resulting mean images. The final registration results are given in Tables 3.4–3.6. For comparison purpose, we also include the results from the affine initialisation and the very best results from the

Generally good matches

Heroic Failures

Hopeless Failures
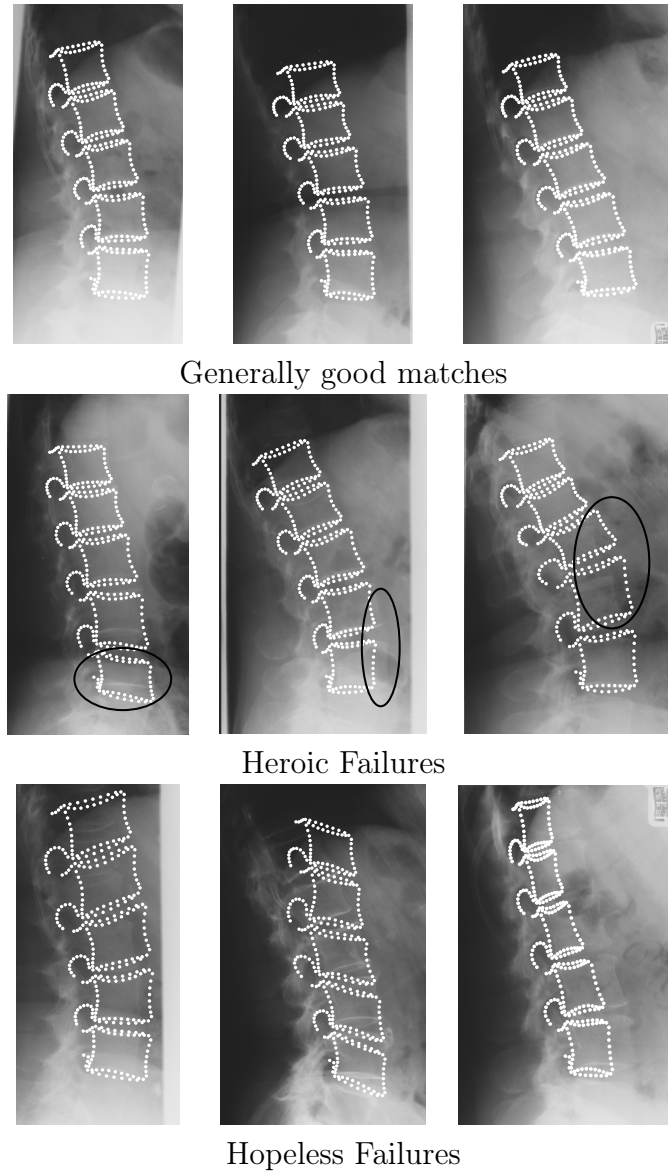
Figure 3.13: Projection of average points onto different spine radiographs (using 10 parts model). Some of the failures are indicated by black ellipses.

Table 3.6: Point-to-curve location errors (mm) of the dense correspondence for the spine set.

Method	m	Mean±s.e.	Median	90%
Affine	–	37.6±2.9	28.8	74.8
Manual	13	10.9±0.6	8.3	20.9
Automatic	10	**3.2±0.5**	**1.9**	**4.1**

proposed system. We find that our system can give similar or better results than a manual choice, demonstrating that our system does help to select useful parts for initialisation. Although the manual selection of parts only needs a few minutes, it is clearly worth spending much more time for better registration accuracy. Results from the spine set suggest that a set of manually selected parts may seem sensible,

(a) Fly wing   (b) Hand   (c) Spine
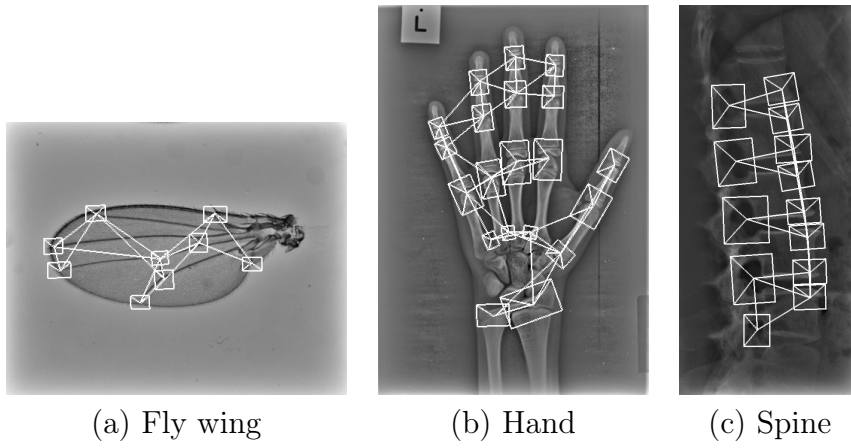
Figure 3.14: Manually constructed parts+geometry models for three datasets.

(a) Good Match   (b) Failure   (c) Final Mean
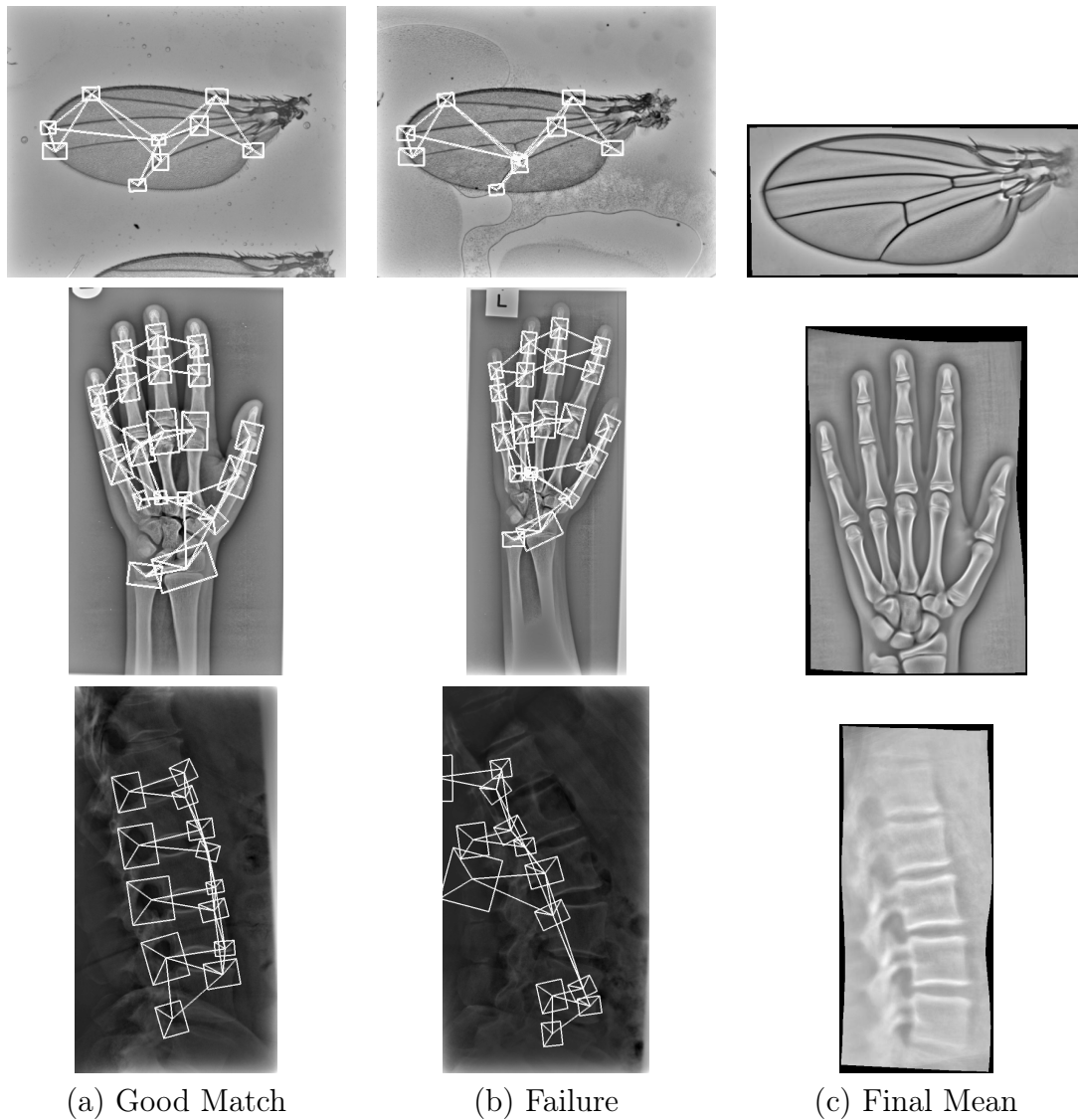
Figure 3.15: Examples of the sparse matches found by the manually constructed parts+geometry models as well as the mean images resulting from groupwise registration.

but may not actually give good initialisation, perhaps because some parts are not reliably found in each image. In this case the automatically chosen parts significantly

outperform a "natural" set chosen by a human (Figure 3.14c)

3.6 Conclusions and Discussion

We have described an automatic system that is able to initialise groupwise registration. This is achieved by using the sparse matches of a parts+geometry model. As we do not use any domain knowledge of the object of interest, the technique potentially can be used for a wide range of datasets.

Experimental results show that our algorithm can achieve good results most of the time on three different datasets, two of them being particularly challenging. We show that our method is superior to the classic affine initialisation in every case. Not only can it further improve the registration result given by the affine initialisation, but also it can give satisfactory results wherever the latter fails utterly. We also show that our automatic system is comparable or better than a supervised approach.

However, the system still suffers from the following problems[1]:

Low efficiency   This includes both searching with part models and selecting the best subset of parts. The computational burden of the former mainly arises from the large number of candidate parts, because searching with one part is reasonably fast. Clearly, if we could reduce the number of candidate parts from thousands to hundreds or even tens, without loss of useful parts, we can significantly reduce the running time of the system. Of course, a further speedup of the system may be achieved by using more efficient part modelling and matching techniques.

The complexity of the optimisation based method is approximately given by $O(N_{\text{gen}}N_{\text{pop}}N_I m K^3)$. Although the computational efficiency of the system can be improved by reducing m and K, we may risk the loss of accuracy. Also, we cannot easiliy manipulate N_{gen} and N_{pop} to speed up the system, because we have to set them reasonably large to ensure the local minima are reached. Hence, if we would like to have a faster system, we have to find some other way to select the useful parts. In chapter 4 we describe a voting scheme which can better initialise groupwise registration as well as use fewer parts+geometry models or even fewer images.

---

[1]   We henceforth discuss the system based on the patch based parts, which have been shown to be more powerful than those SIFT based.

Difficult choice of m Exploring a good value for m requires repeating experiments. Also, m may vary with respect to the object of interest. Hence, it is desirable to develop a strategy that is able to automatically determine the optimal value of m. The voting scheme can be used to solve this problem.

Poor definition of localisability   We use a simple technique to remove the parts with poor localisability, which does not always work well. Parts like and , which are difficult to accurately locate, can be easily included. In chapter 4 we introduce a more effective definition of localisability from a statistical point of view.

Limitation of a single model   The performance of the system only depends on the quality of the final model. However, there are many factors that may degrade that model: (1) the localisability of the parts, *i.e.* resulting in too many poor matches and missing the correct match; (2) the cost function (3.1), *i.e.* Gaussian distribution not sufficient to capture the large shape variations of the object; (3) the GA, *i.e.* leading to suboptimal model due to either many local minima or the poor fitness function—the MDL principle. As a result, it is very unlikely for a single model to deal well with the whole set of images. Quite often, a single model can only tackle a subset of images well and fails on the difficult ones. However, we may overcome this problem by using multiple parts+geometry models, which will be described in Chapter 5.

FOUR

A VOTING STRATEGY FOR PART SELECTION*

In Chapter 3 we showed how to select a subset of candidate parts, which are likely to accurately locate points. This was done by examining the shape of the fit responses of the part models. However, this pre-selection may fail to include useful parts, and does not guarantee that the chosen parts can give consistent and unambiguous matches. In this chapter we present the use of a voting strategy to explicitly explore the localisability of the whole set of candidate parts.

To use the system introduced in Chapter 3, we have to specify the number of parts to be used. This is a common problem of the methods of automatic construction of parts+geometry models [23, 37, 99]. As this parameter, m, can significantly affect the accuracy of the resulting correspondence, it is desirable to automatically determine its proper value. In this chapter, we describe a greedy search algorithm to automatically choose m.

4.1 Method

4.1.1 Voting

The key idea of the voting scheme is that if a match candidate for a part is selected by many different parts+geometry models, it is likely to be the best match for the part (Figure 4.1). Furthermore, if there are significantly more votes for the best match compared to alternatives, then it suggests that the part is a good one. On

---

* Parts of this chapter appeared as "Automatic Part Selection for Groupwise Registration" in the proceeding of IPMI 2011 [104] and will appear in IEEE Transactions on Medical Imaging.

(a) Part models   (b) Part+geometry models

(c) Match candidates   (d) Good localisability   (e) Poor localisability
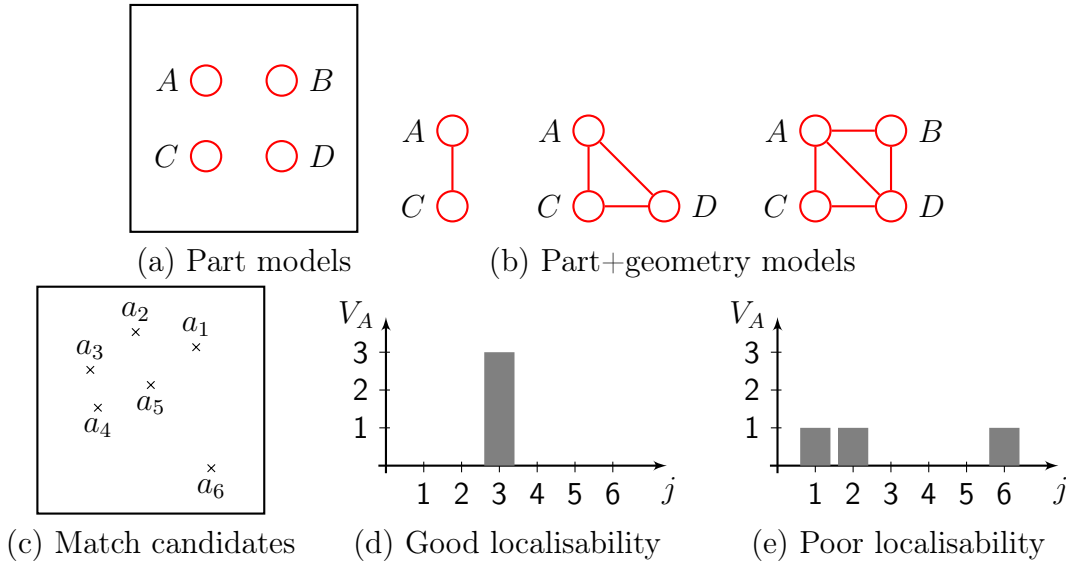
Figure 4.1: (a) Consider four different parts A, B, C and D in an image. (b) We can use part A to construct three different parts+geometry of random configuration. (c) We use these models to search another image and each of them will localise a match candidate a_k for part A. (d) If all three models choose the same match candidate, say a_3, then it is likely to be the best match of A and there will be a peak of votes at a_3. If such a peak is observed on most of the images, it suggests that part A is good at localisation. (e) If the models do not agree with each other, the votes may be approximately uniformly distributed amongst the match candidates. It is thus hard to tell which one is the best match. If this occurs quite often, we may conclude that part A cannot be reliably located.

the other hand, if there are several candidate positions with similar numbers of votes, it suggests that the part is ambiguous. A part which consistently selects "good" matches on many images is likely to have good localisability; otherwise, it may not be suitable for localisation. This suggests that we can propose good parts by analysing the votes of their best matches. We formalise this as follows.

Given a set of parts, we construct at least r different parts+geometry models for each part. Each such model contains a random number, m, of parts. We use each model to identify the best match on each image for each part, and record votes for each match candidate. Let $V_{i,k}^j$ be the number of votes for candidate j of part i on image k—the number of times that the candidate j is selected as the best match of part i. Let $T_{i,k} = \sum_j V_{i,k}^j$ be the total votes cast by all parts+geometry models for the position of part i on image k. Then $L_{i,k}^j = V_{i,k}^j/T_{i,k}$ is an estimate of the probability that candidate j is the correct match. If $L_{i,k}^{\hat{j}}$ is near 1, where $\hat{j} = \arg\max_j L_{i,k}^j$, then it suggests that many different models voted for a same

location, and thus part i has good localisability on I_k. If it is near zero, it suggests that the models voted for a range of different candidates, with no one standing out—part i cannot be reliably located on I_k. The localisability of part i can therefore be measured using $Q_i = \frac{1}{N_I} \sum_k L_{i,k}^{\hat{j}}$, the average probability of the best match over the set of images. We can then rank the parts by Q_i—those with the highest values are likely to be the most reliable.

4.1.2 Greedy Search

We use a greedy search algorithm to select an initial sparse subset, by first choosing the highest ranked part, then selecting the next parts which are sufficiently well separated from the current set and whose probability pass a threshold t_Q. We define two parts to be overlapping if their mean separation is less than a threshold, usually set to the average radii of the two parts. To ensure only reliable parts are included in the initial set, the threshold t_Q is set high, and thus some relevant areas of the object may be missed. We therefore extend the set by adding a part if it improves the utility of the resulting model (see Section 3.1.3). The result of this process is a set of parts and associated sparse matches on every image. The number of parts is automatically chosen by the algorithm. We summarise the above procedure in Algorithm 3.

4.2 Experiments

We compare the voting based method with the optimisation based method described in Section 3.3. We used the same sets of images and candidate parts as used in Section 3.5.2. Note that here we consider the whole set of candidate parts, not just those pre-selected. We also used the same protocol to calculate the registration accuracy.

4.2.1 Performance of the Approach

We generated a large number of parts+geometry models with random m, repeating until each part has been considered in at least $r = 100$ different models. We computed the average probability measure, Q, for each part, and retained those with $Q > t_R$ as well as in a roughly defined boundary (Figure 4.2a). We chose an initial

---

Algorithm 3: The voting based algorithm

Input : $\{\boldsymbol{p}_{i,0}\}$
Output : G_{opt}
Initialise: r, $\mathcal{G} \leftarrow \emptyset$, $\mathcal{P}_S \leftarrow \emptyset$, $\mathcal{P}_{\text{opt}} \leftarrow \emptyset$, $V_{i,k}^j \leftarrow 0$

// *Repeat using each part to build at least r different*
// *models, where m is randomly chosen from a range*
$\mathcal{G} \leftarrow \texttt{RandomSample}(\{\boldsymbol{p}_{i,0}\}, r)$

// *Cast the votes*
for all $G_t \in \mathcal{G}$ **do**
    for all *training images* **do**
        $M_{t,k} \leftarrow \texttt{getBestGraphMatch}(G_t, I_k)$
        // *Update the votes for candidates of each part*
        $V_{i,k}^j \leftarrow \texttt{CountVotes}(G_t, M_{t,k}) + V_{i,k}^j$

// *Analyse the distribution of votes for each part*
for all *candidate parts* **do**
    for all *training images* **do**
        $T_{i,k} \leftarrow \sum_j V_{i,k}^j$
        $\hat{j} \leftarrow \arg\max_j V_{i,k}^j / T_{i,k}$
    $Q_i \leftarrow \frac{1}{N_I} \sum_k L_{i,k}^{\hat{j}}$
$\{\boldsymbol{p}_{i,0}\} \leftarrow \texttt{sort}(\{\boldsymbol{p}_{i,0}\}, \{Q_i\})$

// *Retain the parts with high Q*
$\mathcal{P}_S \leftarrow \texttt{RetainBest}(\{\boldsymbol{p}_{i,0}\})$

// *Select an initial subset of reliable parts*
for all $\boldsymbol{p}_{i,0} \in \mathcal{P}_S$ **do**
    if $Q_i > t_Q$ `&&!`$\texttt{isOverlapped}(\boldsymbol{p}_{i,0}, \mathcal{P}_{\text{opt}})$ **then**
        $\mathcal{P}_{\text{opt}} \leftarrow \mathcal{P}_{\text{opt}} \cup \{\boldsymbol{p}_{i,0}\}$

// *Extend the subset based on Equation* (3.3)
$G \leftarrow \texttt{BuildGraphModel}(\mathcal{P}_{\text{opt}})$
$U_{\text{new}} \leftarrow \texttt{ComputeUtility}(G)$

while 1 **do**
    $U_{\text{old}} \leftarrow U_{\text{new}}$
    $U_{\text{new}} \leftarrow \infty$
    for all $\boldsymbol{p}_{i,0} \in \mathcal{P}_S$ **do**
        if $!\texttt{isOverlapped}(\boldsymbol{p}_{i,0}, \mathcal{P}_{\text{opt}})$ **then**
            $G \leftarrow \texttt{BuildGraphModel}(\mathcal{P}_{\text{opt}} \cup \{\boldsymbol{p}_{i,0}\})$
            $U \leftarrow \texttt{ComputeUtility}(G)$
            if $U < U_{\text{new}}$ **then**
                $\boldsymbol{p}_{\text{min}} \leftarrow \boldsymbol{p}_{i,0}$
                $U_{\text{new}} \leftarrow U$

    if $U_{\text{new}} < U_{\text{old}}$ **then**
        $\mathcal{P}_{\text{opt}} \leftarrow \mathcal{P}_{\text{opt}} \cup \{\boldsymbol{p}_{\text{min}}\}$
    else
        `break`

$G_{\text{opt}} \leftarrow \texttt{BuildGraphModel}(\mathcal{P}_{\text{opt}})$

---

a) $Q > t_R$ b) Initial parts   c) Final parts   d) Final model
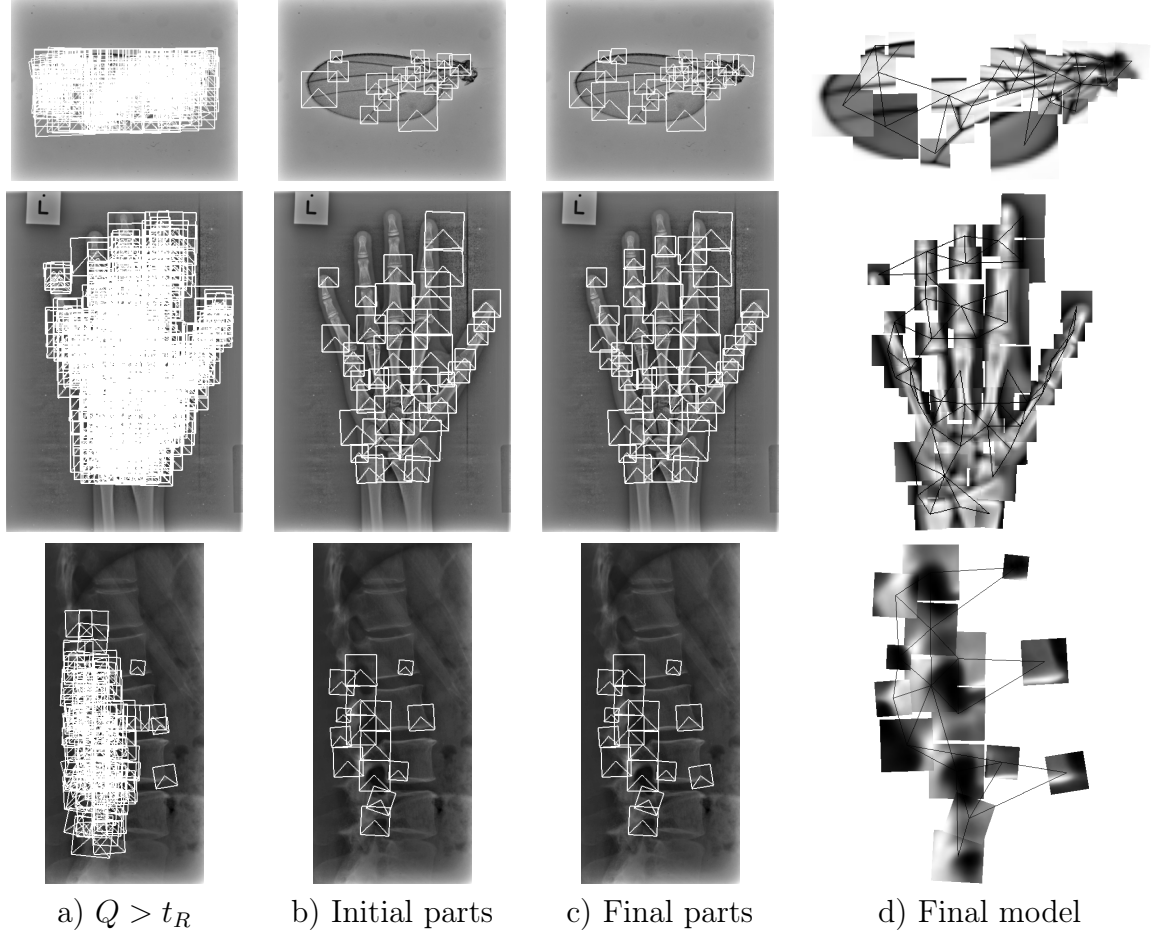
Figure 4.2: Results of the voting based method on the three datasets. From top to bottom, fly wing, hand and spine.

subset of non-overlapping parts with $Q > t_Q$ (Figure 4.2b), and extended the set with further parts which may better encode the entire training set in terms of an MDL principle (Figure 4.2c). In the experiment we set $t_R = 0.9$ and $t_Q = 0.95$ for the hands and fly wings, $t_R = 0.5$ and $t_Q = 0.55$ for the spines (lower thresholds were required for this challenging dataset as few parts gave consistently high values of Q). To evaluate the merit of the added parts, we constructed two parts+geometry models: one was built from the initial set and the other was from the extended set (Figure 4.2d). We used the two models to search each dataset respectively and initialised groupwise registration with the resulting sparse correspondences.

Tables 4.1–4.3 show the correspondence accuracy for the three datasets after dense registration. The best results from the optimisation based method are also included for comparison. In each table the last two rows show the registration results from the initial and final parts+geometry models.

We find that the voting based method can give similar or better results than

Table 4.1: Point-to-point location errors (pixels) of the dense correspondence for the fly wings.

Method	m	Mean±s.e.	Median	90%
Optimisation	15	1.6±0.1	1.4	2.2
Voting	19	1.2±0.03	1.1	1.6
	25	1.2±0.03	1.2	1.5

Table 4.2: Point-to-point location errors (mm) of the dense correspondence for the hand set.

Method	m	Mean±s.e.	Median	90%
Optimisation	40	1.0±0.05	0.8	1.5
Voting	37	1.1±0.06	0.9	1.5
	44	**0.9±0.03**	**0.8**	**1.2**

Table 4.3: Point-to-curve location errors (mm) of the dense correspondence for the spine set.

Method	m	Mean±s.e.	Median	90%
Optimisation	10	3.2±0.5	1.9	4.1
Voting	15	3.4±0.4	2.2	4.6
	16	3.0±0.4	2.1	3.8

the optimisation based one. We also find that a reasonable choice of t_Q is good enough to select useful parts. The resulting parts+geometry model can capture the morphological features of the object well. Adding more parts using the MDL criterion does not impair the localisability of the model and can help remove outliers, which may lead to a further improvement on registration accuracy.

4.2.2 Influence of Important Parameters

To investigate the effect of the two parameters in the voting based method—the lower bound of the number of models per part, r, and the number of images used to construct the models N_I, we repeated the above experiment with the same sets of images on two datasets—the fly wings and the hands. We first used all images in the training set and ran the algorithm for a range of different r. We then fixed $r = 100$ and ran the algorithm on sized subsets of the training set. The resulting parts+geometry models were used to find correspondence on the rest of images.

Figure 4.3 shows the impact of the two parameters on the accuracy of the resulting dense correspondence. For the hand set, even with r as low as 10 the voting based method can still give good results. The running time of the voting stage is

(a) Influence of the number of repeating times

(b) Influence of the number of training images
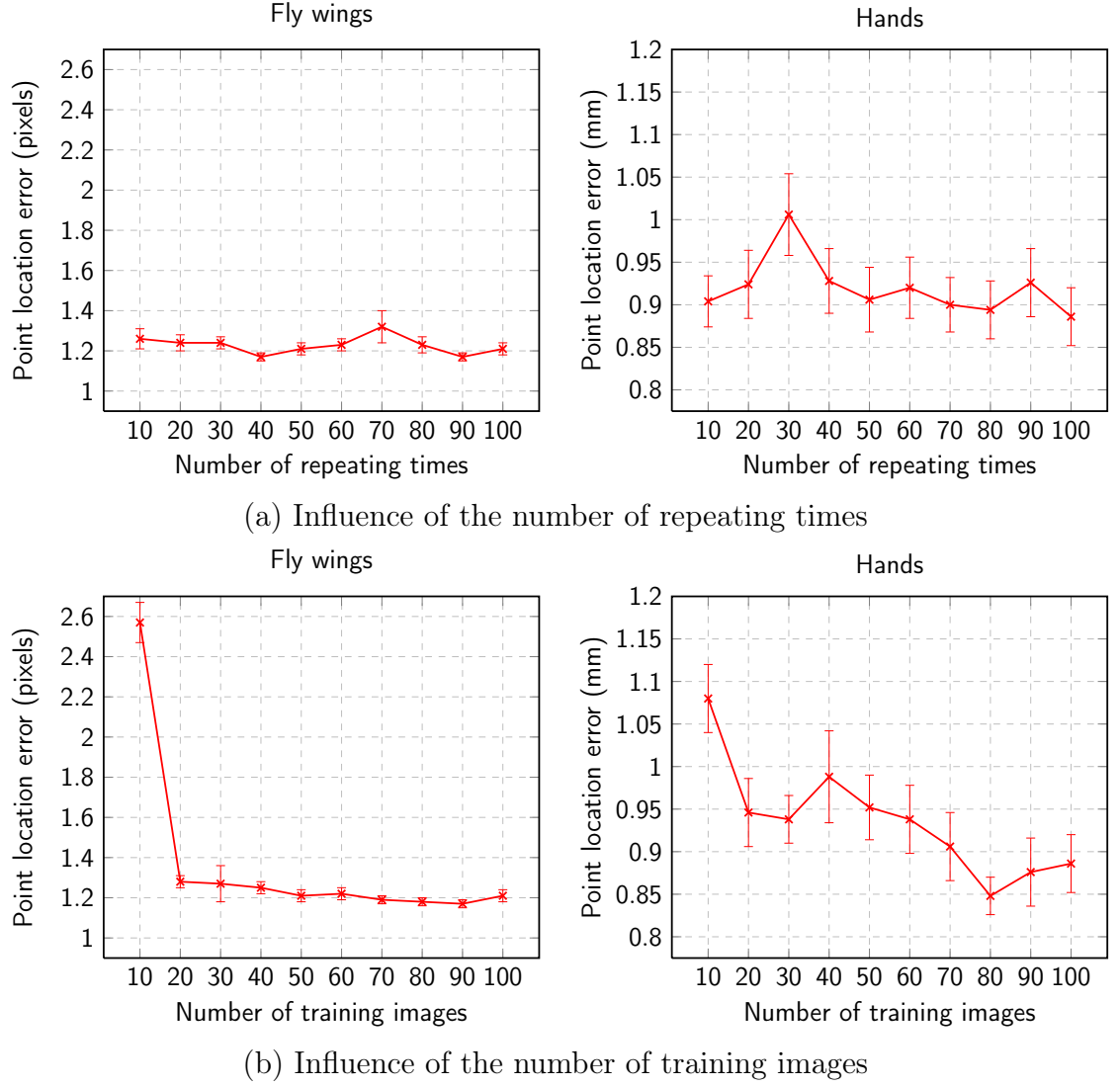
Figure 4.3: The influence of the two parameters on the accuracy of groupwise registration initialised by the voting based algorithm.

linear in r. Compared with the best result obtained using the optimisation based algorithm ($m = 40$), the voting scheme only took 2.7 hours to run for $r = 10$ while the optimisation one took 10.3 hours. As there is a linear relationship between the algorithm and the number of the training images, using fewer images can also speed up the algorithm. A similar pattern can also be observed on the fly wing dataset, where the voting scheme took 1.0 hours to run for $r = 10$.

4.3 Conclusions and Discussion

We have described a method that can effectively initialise groupwise non-rigid registration on datasets of simple or complex structures. This is achieved by the following three steps: (1) selecting a subset of parts with good localisability using a voting

strategy; (2) using a greedy search to add more parts to the set; (3) construct-ing a parts+geometry model with the chosen parts to do initialisation. Compared with the optimisation based method, this voting based algorithm can give similar or better initialisation, and can run much faster without loss of accuracy.

To choose the optimal subset of parts we use two thresholds. Although this simple technique appears to work well for all datasets reported, a more sophisticated approach is still desirable to avoid repeating trials for proper values (*e.g.* the spines). As will be shown in the next chapter, this technique is also sensitive to the choice of reference images on some datasets. Besides, the voting based method is still a single-model initialisation scheme. As a result, it suffers from the same problem as the optimisation based method, that is, it is unlikely to obtain a one-size-fits-all model. This can be clearly seen from the results on the spine set, where both methods have similar registration errors, suggesting that the model found by the voting based method still fails quite often on this difficult dataset. In the following chapter we will describe how to use multiple parts+geometry models to boost the performance of the system.

FIVE

MULTI-MODEL INITIALISATION

We have described two different approaches to initialising groupwise registration. Although methodology varies, they all attempt to seek a single parts+geometry model for good initialisation. However, this single-model initialisation strategy is problematic. There are two reasons.

One is that we use an MDL criterion in both methods to help select the optimal model. There is no guarantee that the model favoured by such criterion will give the best initialisation and thus registration of the whole set. Quite often, other highly ranked models may give a better choice. Figure 5.1 shows the accuracy of groupwise registration initialised by each of the top 10 models selected using the MDL criterion. Clearly, the best model (model 1) does not result in the best overall registration.

The other reason is that due to the limit of the image set and imperfection of the algorithm, it is unlikely for a single model to capture every possible variation of the object. As a result, the learned model may fail on some of the images, particularly when dealing with objects of complex structures. For an example, see Figure 5.2.

Figure 5.2 also suggests that it is always possible to learn a model which can deal well with a subset of images. Different models may work with different subsets, which are likely to be complementary to each other. By combining the best results from a set of models, we may achieve better initialisation than that from a single model.

In this chapter we explore using multiple parts+geometry models to initialise groupwise registration—a multi-model initialisation scheme. We build this scheme
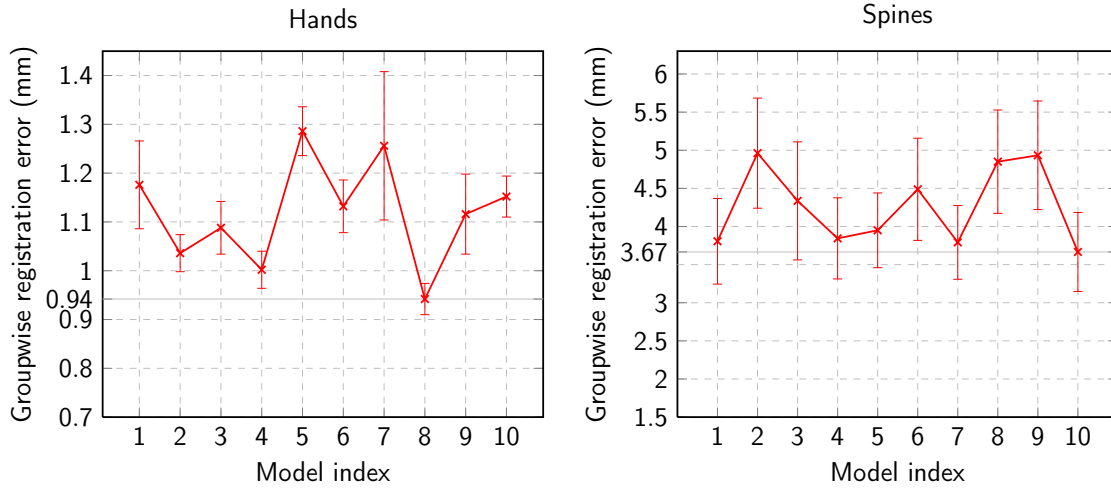
Figure 5.1: An illustration that the best model in terms of an MDL principle does not guarantee the best registration result. Models 1 to 10 (the smaller the better) are those most favoured by the MDL principle. We use each of them to initialise groupwise registration and calculate the resulting registration error. We repeat this process on the hand and spine datasets used in Section 3.5.2, and plot the registration error against the model index. For each model, we use $m = 30$ for the hands and $m = 10$ for the spines.

| (a) Parts+geometry model A | (b) consistent matches | (c) Failures |
| (d) Parts+geometry model B | (e) Failures | (f) consistent matches |
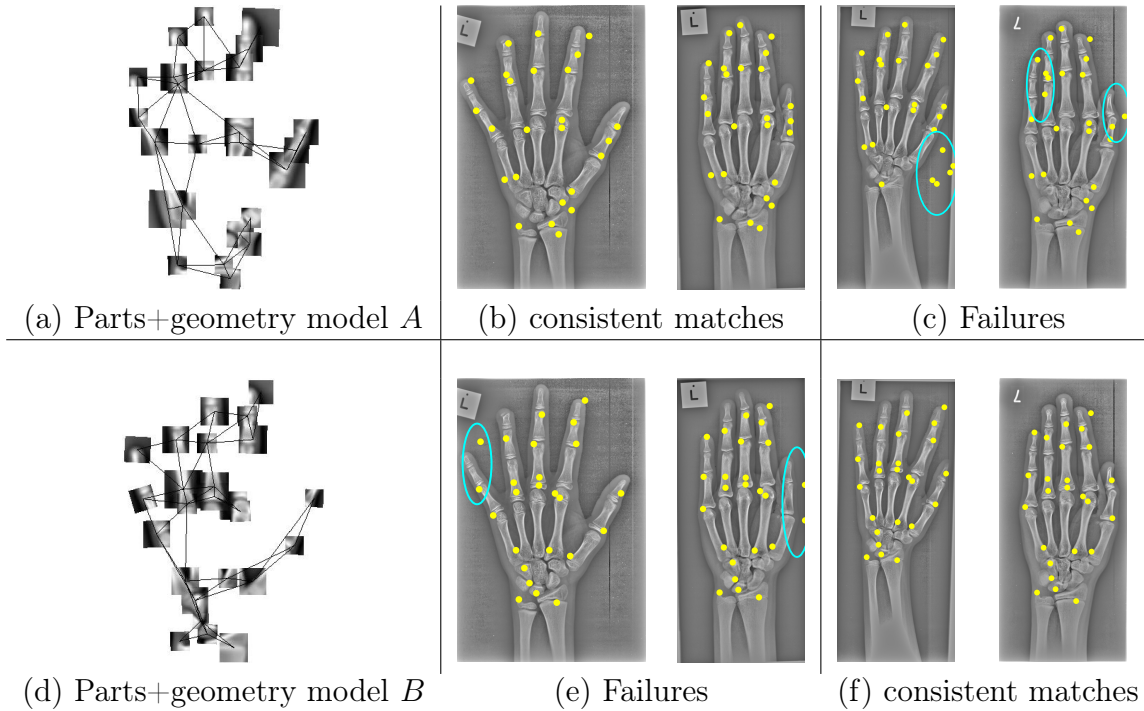
Figure 5.2: Left column: two parts+geometry models. Right four columns: matches of the models. Top row: model A can find consistent matches on two images but fails on the other two. Bottom row: model B can give consistent matches on those where model A fails but fails where model A works well. Failures are indicated by cyan ellipses.

upon the optimisation based method, as it naturally generates a population of roughly good models.

5.1 Method

Let the term *pattern* refer to the sparse set of points found on an image by a particular model, which thus defines the correspondence for that image to a reference frame for that model. When using multiple models we obtain multiple patterns on each image. As shown in Figure 5.2, some patterns are apparently better than the others. By replacing those poor patterns with good ones, we can modify and improve the correspondence defined by the single model.

An overview of the multi-model initialisation scheme is given in Figure 5.3. Given a set of parts+geometry models, we use each of them to search the image set to obtain a sparse correspondence, where a dense correspondence can be generated using a TPS interpolation. Poor patterns will lead to poor dense points (indicated by the red box in Figure 5.3). The quality of the pattern can be evaluated by warping the target image to a mean image using the dense points, and comparing the similarity between the warped image and the mean. To choose the best pattern for an image, we use each of the associated sets of dense points to warp the image to the mean and compute the similarity. A new correspondence can thus be established by grouping the sets of dense points related to the best patterns across the images. We give the details of each step below.

5.1.1 Dense Points

As we cannot easily compare the quality of different patterns for an image, a solution is to transfer the correspondence information encoded in each pattern to a common set of points. A simple method is to generate a dense set of points on a reference image and propagate the points to the other images using a TPS, as illustrated in Figure 3.6. We use $\boldsymbol{X} = (x_1, y_1, x_2, y_2, \ldots, x_n, y_n)^{\mathrm{T}}$ to denote the positions of the dense points.

5.1.2 Quality of the Pattern

An observation from groupwise registration is that if the correspondences across the image set are well established, we should be able to obtain a crisp mean image. Furthermore, if a pattern is good, its related image should be similar to the mean when comparing the two in the same frame, and *vice versa*. Hence, we can use a
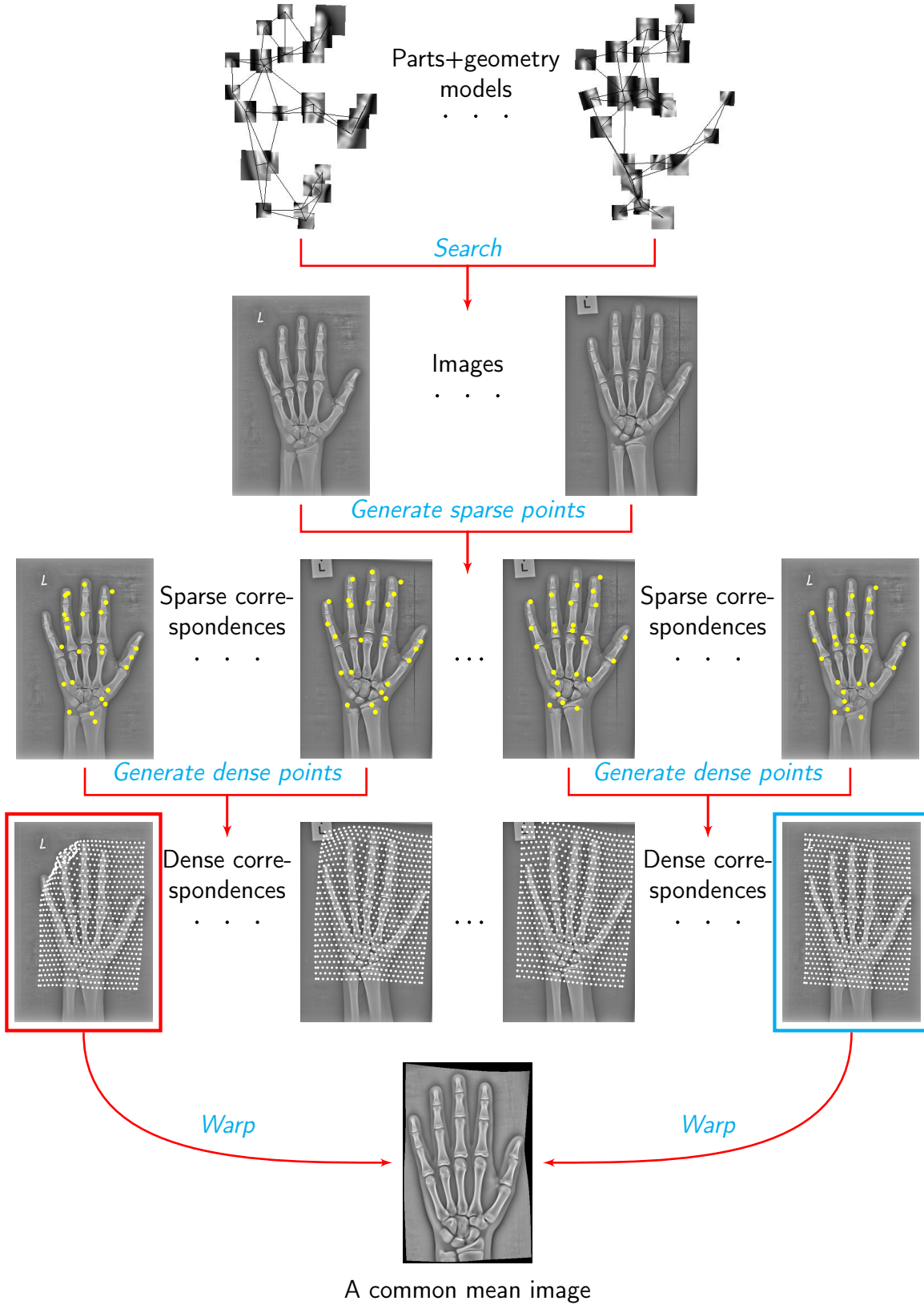
Figure 5.3: An overview of the multi-model initialisation strategy.

mean image to evaluate the quality of a pattern.

Suppose we have a set of images $\{I_k | k = 1, \ldots, N_I\}$ and N_G parts+geometry models. We use $\{\boldsymbol{X}_k^l | l = 1, \ldots, N_G\}$ to denote the sets of dense points associated

with image I_k. Let $\bar{I}$ be the mean image and $\overline{X}$ be the dense points in the mean (see below). Given an image I_k and one of its dense point sets X_k^l, a warp from $\bar{I}$ to I_k is uniquely defined by $\overline{X}$ and X_k^l. We write this warp as $z' = W(z : \overline{X}; X_k^l)$, where z is a point in $\bar{I}$ and z' is the corresponding point in I_k. To evaluate the quality of the pattern related to X_k^l, we warp I_k onto $\bar{I}$ so as to compare them in a same frame, and use the following function

$$D_k^l = \sum_{z \in R} \left| I_k(W(z : \overline{X}; X_k^l)) - \bar{I}(z) \right|, \qquad (5.1)$$

where R is a region of interest in the mean frame. This function computes the absolute intensity difference over the region of interest between the warped image $I_k(W(z : \overline{X}; X_k^l))$ and the mean. Note that all the images have been preprocessed to standardise their intensity ranges.

5.1.3 Mean Image

Given $\{X_k^l | k = 1, \dots, N_I\}$, we can perform the following steps to compute the mean image $\bar{I}^l$:

(1) Use Procrustes Analysis [19] to align each X_k^l to a reference frame[1], computing a mean shape $\overline{X^l}$;

(2) Create a triangulation of the mean shape using the Delaunay algorithm;

(3) Warp each image I_k to the reference frame, computing $I_k(W(z : \overline{X^l}; X_k^l))$,[2] $z \in R$;

(4) Compute the mean image using $\bar{I}^l(z) = \frac{1}{N_I} \sum_{k=1}^{N_I} I_k(W(z : \overline{X^l}; X_k^l))$, $z \in R$.

5.1.4 A Common Mean

Different sets of points $\{X_k^l | k = 1, \dots, N_I\}$ will lead to a different mean image $\bar{I}^l$. Hence, the quality D computed using different means cannot be compared directly. This can be solved by using a single, common mean. We take the following steps to compute this common mean image:

(1) For $l \leftarrow 1$ to N_G

---

[1] The choice of the reference frame is free. Any image in the set can be used for this purpose.

[2] $I_k(W(z : \overline{X^l}; X_k^l))$ is computed by piece-wise linear interpolation between corresponding triangles in $\overline{X^l}$ and X_k^l.

    (a) Compute the mean image $\bar{I}^l$ using the sets of points $\{\boldsymbol{X}_k^l | k = 1, \ldots, N_I\}$;

    (b) Warp every image to the mean $\bar{I}^l$, computing D_k^l using Equation (5.1);

    (c) Rank all images by D_k^l and select the top 50% of the images as the set $\mathcal{S}^l$;

(2) Find a common set of images $\mathcal{S} = \bigcap_{l=1}^{N_G} \mathcal{S}^l$—selecting the images in all sets $\{\mathcal{S}^l\}$;

(3) For each image in $\mathcal{S}$, average associated sets of dense points, $\widehat{\boldsymbol{X}}_k = \frac{1}{N_G} \sum_{l=1}^{N_G} \boldsymbol{X}_k^l$;

(4) Compute a mean image using $\mathcal{S}$ and the sets of points $\{\widehat{\boldsymbol{X}}_k\}$. Take this as the common mean.

5.1.5   Pattern Selection

Once we have the common mean image, we can take the following steps to select the best pattern for each image:

(1) For $l \leftarrow 1$ to N_G

    For $k \leftarrow 1$ to N_I

        Warp I_k to the common mean and compute D_k^l;

(2) For $k \leftarrow 1$ to N_I

    Select the best $\boldsymbol{X}_k^{\hat{l}}$ from $\{\boldsymbol{X}_k^l | l = 1, \ldots, N_G\}$ such that $D_k^{\hat{l}}$ is minimum.

By grouping $\boldsymbol{X}_k^{\hat{l}}$ across the image set we can obtain a new correspondence $\{\boldsymbol{X}_k^{\hat{l}} | k = 1, \ldots, N_I\}$. Although we can use it to directly initialise groupwise registration, in the following experiment we use it to generate a sparse correspondence. This is more efficient, and allows fairer comparison with the single-model scheme which only outputs sparse sets of points.

    To create the sparse correspondence we (1) generate a sparse set of points on an image I_k using the best parts+geometry model (in terms of model utility); (2) project the sparse points to the other images using the piece-wise affine transformation between different $\boldsymbol{X}_k^{\hat{l}}$.

5.2 Experiments

5.2.1 Performance of Multi-model Initialisation

We compare this multi-model initialisation strategy with the single-model one. We used the same sets of images as used in Section 3.5.2. For each dataset, we randomly selected 10 images as the reference images, constructed the patch based parts for each reference, and ran the optimisation based method and the voting based one respectively using each set of parts. For both methods, all parameters were set the same as used in Section 3.5.2 and Section 4.2.1, except that we used $r = 10$ for the voting here. The initialisation with both methods was done exactly the same as before.

We used the parts+geometry models output by the optimisation based method

Table 5.1: Point-to-point location errors (pixels) of the dense correspondence for the fly wings using different reference images.

(a)

Method	m	Mean±s.e.	Med.	90%
Opt.	20	1.4±0.05	1.3	1.9
Voting	21	1.5±0.05	1.4	2.1
Multi	10	1.3±0.03	1.3	1.8

(b)

Method	m	Mean±s.e.	Med.	90%
Opt.	15	1.5±0.06	1.3	2.3
Voting	27	1.6±0.1	1.2	2.1
Multi	15	1.4±0.03	1.3	1.8

(c)

Method	m	Mean±s.e.	Med.	90%
Opt.	5	1.5±0.06	1.3	2.1
Voting	26	1.3±0.04	1.2	1.8
Multi	5	1.3±0.03	1.2	1.6

(d)

Method	m	Mean±s.e.	Med.	90%
Opt.	25	1.6±0.06	1.4	2.4
Voting	22	1.6±0.08	1.4	2.5
Multi	5	**1.4±0.05**	**1.3**	**1.9**

(e)

Method	m	Mean±s.e.	Med.	90%
Opt.	20	2.1±0.1	1.7	3.8
Voting	24	1.6±0.08	1.3	2.9
Multi	15	1.6±0.06	1.4	2.3

(f)

Method	m	Mean±s.e.	Med.	90%
Opt.	5	1.4±0.04	1.3	1.9
Voting	24	**1.1±0.02**	**1.1**	**1.4**
Multi	5	1.2±0.04	1.2	1.6

(g)

Method	m	Mean±s.e.	Med.	90%
Opt.	10	2.0±0.07	1.9	2.7
Voting	18	**1.2±0.03**	**1.1**	**1.6**
Multi	5	1.4±0.04	1.3	2.0

(h)

Method	m	Mean±s.e.	Med.	90%
Opt.	5	1.7±0.3	1.4	1.9
Voting	25	1.3±0.05	1.2	1.8
Multi	5	1.3±0.04	1.3	1.7

(i)

Method	m	Mean±s.e.	Med.	90%
Opt.	15	1.4±0.04	1.3	1.9
Voting	27	1.8±0.1	1.4	3.2
Multi	5	1.3±0.03	1.3	1.7

(j)

Method	m	Mean±s.e.	Med.	90%
Opt.	10	1.2±0.03	1.2	1.5
Voting	22	1.9±0.1	1.6	3.1
Multi	10	1.2±0.03	1.2	1.5

Table 5.2: Point-to-point location errors (mm) of the dense correspondence for the hands using different reference images.

(a)

Method	m	Mean±s.e.	Med.	90%
Opt.	20	1.0±0.04	0.9	1.5
Voting	44	1.0±0.06	0.8	1.5
Multi	20	1.1±0.05	0.9	1.7

(b)

Method	m	Mean±s.e.	Med.	90%
Opt.	40	1.1±0.05	1.0	1.6
Voting	45	1.3±0.09	1.0	2.2
Multi	30	1.1±0.04	1.0	1.6

(c)

Method	m	Mean±s.e.	Med.	90%
Opt.	50	1.3±0.1	0.9	1.7
Voting	36	1.1±0.06	1.0	1.4
Multi	40	1.2±0.07	1.0	1.9

(d)

Method	m	Mean±s.e.	Med.	90%
Opt.	30	1.0±0.05	0.9	1.5
Voting	40	1.0±0.04	0.9	1.4
Multi	50	0.9±0.04	0.9	1.3

(e)

Method	m	Mean±s.e.	Med.	90%
Opt.	20	1.0±0.09	0.8	1.2
Voting	43	1.4±0.08	1.1	2.9
Multi	40	0.9±0.03	0.8	1.4

(f)

Method	m	Mean±s.e.	Med.	90%
Opt.	20	1.1±0.04	0.9	1.6
Voting	46	1.0±0.05	0.9	1.4
Multi	40	1.0±0.03	0.9	1.4

(g)

Method	m	Mean±s.e.	Med.	90%
Opt.	40	1.1±0.07	0.9	1.6
Voting	37	1.2±0.07	1.0	2.1
Multi	30	1.0±0.04	0.9	1.5

(h)

Method	m	Mean±s.e.	Med.	90%
Opt.	50	1.1±0.08	0.9	1.6
Voting	41	1.0±0.07	0.9	1.4
Multi	50	0.9±0.03	0.8	1.3

(i)

Method	m	Mean±s.e.	Med.	90%
Opt.	40	1.1±0.05	0.9	2.0
Voting	44	1.0±0.05	0.9	1.4
Multi	50	1.0±0.04	0.9	1.6

(j)

Method	m	Mean±s.e.	Med.	90%
Opt.	30	1.2±0.09	1.0	1.7
Voting	29	1.2±0.1	0.9	1.6
Multi	50	1.1±0.06	0.9	1.7

to do multi-model initialisation. Specifically, we retained the top 10 models for each number of parts, fed them into the approach described in Section 5.1, and used the resulting sparse correspondences to initialise groupwise registration.

We used the same protocol to calculate registration errors. We show part of the results in Tables 5.1–5.3. For full results, see Appendix A. Each sub-table (a-j) represents an experiment done for a reference image. Only the very best results for each method are given here. We used Welch's (two-tailed) t-test ($p < 0.05$) to compare the result from each method so as to find out the one which is significantly better than the others (emphasised in bold font). Only the means and standard deviations ($\propto$ the standard errors reported) were used for such test. We can see that the multi-model initialisation strategy performs similarly or better than the single-model one in most of cases. The power of multiple models can be clearly seen from

Table 5.3: Point-to-curve location errors (mm) of the dense correspondence for the spines using different reference images.

(a)

Method	m	Mean±s.e.	Med.	90%
Opt.	15	2.5±0.3	1.9	3.6
Voting	26	6.4±0.5	4.5	13.2
Multi	20	2.1±0.1	1.7	3.1

(b)

Method	m	Mean±s.e.	Med.	90%
Opt.	15	2.6±0.3	1.8	3.4
Voting	20	9.8±0.5	8.0	19.0
Multi	5	2.1±0.3	1.6	2.8

(c)

Method	m	Mean±s.e.	Med.	90%
Opt.	10	4.4±0.6	2.7	6.0
Voting	25	11.0±0.5	9.7	18.3
Multi	15	**2.7±0.3**	**1.9**	**4.0**

(d)

Method	m	Mean±s.e.	Med.	90%
Opt.	20	4.0±0.5	2.5	7.6
Voting	21	3.6±0.4	2.3	6.5
Multi	5	**2.1±0.1**	**1.8**	**3.1**

(e)

Method	m	Mean±s.e.	Med.	90%
Opt.	10	3.6±0.5	1.9	5.3
Voting	31	6.5±0.4	5.3	10.8
Multi	15	**2.1±0.1**	**1.7**	**3.2**

(f)

Method	m	Mean±s.e.	Med.	90%
Opt.	20	3.3±0.5	1.6	4.1
Voting	27	6.7±0.5	4.7	15.9
Multi	20	**1.9±0.1**	**1.6**	**3.1**

(g)

Method	m	Mean±s.e.	Med.	90%
Opt.	15	3.5±0.5	2.1	5.9
Voting	16	4.8±0.4	3.6	9.2
Multi	10	**2.0±0.3**	**1.6**	**2.6**

(h)

Method	m	Mean±s.e.	Med.	90%
Opt.	10	4.4±0.5	2.6	8.5
Voting	15	9.3±0.4	8.2	14.2
Multi	10	3.6±0.4	2.5	5.1

(i)

Method	m	Mean±s.e.	Med.	90%
Opt.	10	3.6±0.5	2.2	3.9
Voting	14	13.6±0.6	12.6	20.9
Multi	10	2.6±0.2	2.0	3.9

(j)

Method	m	Mean±s.e.	Med.	90%
Opt.	15	2.7±0.4	1.8	4.3
Voting	37	4.4±0.4	3.4	5.7
Multi	10	**1.6±0.1**	**1.5**	**2.5**

the results on the spines, where the improvements are the most significant compared with those achieved on the other two datasets. Despite of the slight improvements on the fly wings and hands, the multi-model initialisation scheme still effectively corrects some poor matches and enhances the consistency of the correspondences (*e.g.* reduced standard error or outliers).

To visualise the improvements achieved by the multi-model initialisation strategy on the spines, we show the final mean images in Figure 5.4. These images are from the experiment summarised in Table 5.3j. We can see that the unlikely deformation of the bottom vertebra (indicated by yellow circles), which occurs in both single-model initialisation approaches, has been largely corrected by different parts+geometry models.

Note that our method can work with parts+geometry models of different num-

(a) Optimisation   (b) Voting   (c) Multi-model
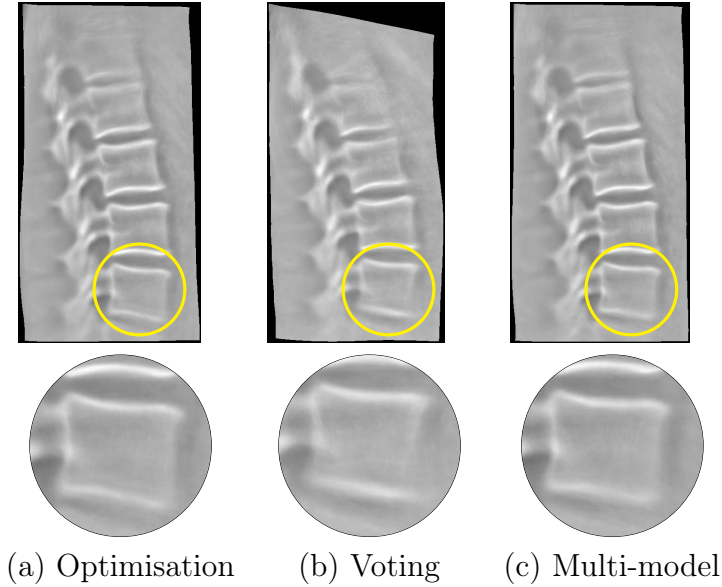
Figure 5.4: An example of the final mean images from both the single-model and multi-model initialisation schemes.

Table 5.4: Dense registration errors resulting from models of different numbers of parts.

Dataset	Mean±s.e.	Median	90%
Fly wings (pixels)	1.4±0.05	1.3	2.0
Hands (mm)	1.2±0.07	0.9	2.4
Spines (mm)	1.7±0.08	1.4	2.6

bers of parts. Table 5.4 shows the groupwise registration results using the models from the last two rows in Tables A.1j–A.3j respectively. We selected the best 5 models for each number of parts. Hence, 10 models in total were used to initialise groupwise registration for each dataset. The registration errors are similar to those given in Tables A.1j–A.3j.

5.2.2 Influence of Reference Images

To show the effects of different choice of the reference image, we computed the median $\hat{c}$ of the 10 medians c_i and the mean absolute difference $\text{MAD} = \sum_{i=1}^{10} |\hat{c} - c_i|/10$ for each method on each dataset. We compared $\hat{c}$ with the best median $c_{\min}$ of the 10 cases and MAD with the corresponding standard deviation.

We summarise the results in Table 5.5. Note that these results were computed from the original data not from those given in Tables 5.1–5.3, where approximations have been applied. We find that the choice of reference images only has a small effect on the results of the multi-model initialisation strategy for all three datasets.

Table 5.5: The influence of choice of reference images on the single-model and multi-model initialisation strategies.

(a) Fly wings (pixels)

Method	$\hat{c}$	MAD	c_{min}	s.d.
Opt.	1.3	0.1	1.2	0.3
Voting	1.3	0.1	1.1	0.2
Multi	1.3	0.06	1.2	0.3

(b) Hands (mm)

Method	$\hat{c}$	MAD	c_{min}	s.d.
Opt.	0.9	0.04	0.8	0.9
Voting	0.9	0.06	0.8	0.6
Multi	0.9	0.04	0.8	0.3

(c) Spines (mm)

Method	$\hat{c}$	MAD	c_{min}	s.d.
Opt.	2.0	0.3	1.6	5.0
Voting	5.0	2.5	2.3	4.0
Multi	1.7	0.2	1.5	1.0

Although a similar pattern can be observed for both single-model initialisation approaches on the fly wings and hands, the performance of the two methods varies more on the spines. For example, the performance of the voting based method varies dramatically from one reference image to another one.

5.3 Conclusions and Discussion

We have described a strategy that can effectively initialise groupwise non-rigid registration. This is achieved by using a set of parts+geometry models. Experiments show that this scheme is able to achieve very good results and can significantly outperform earlier approaches which only use a single model. By using such scheme, we have achieved the best result on the spines (shown in Table 5.3j) so far, much better than any observed by using the single-model initialisation approaches. We also compared this multi-model initialisation strategy with the single-model one in terms of the influence of reference images. We find that the multi-model scheme is the least sensitive to the choice of reference images, suggesting that a robust system can be expected.

Current work indiscriminately uses the top models output by the optimisation based method. If some models result in too many poor matches, the performance of the multi-model scheme will be inevitably degraded. Moreover, if different models share too many common parts, redundancy will arise. This will dilute the advantage of using multiple models to do initialisation, since different models may fail on the same images so that there is no chance to rectify those faults. Hence, it is desirable

to explore how to effectively choose a good set of parts+geometry models.

92

SIX

LABELLING NEW IMAGES[⋆]

Once we have successfully registered a set of images, we can use the resulting dense correspondences to build a statistical model to annotate new images. There are three motivations.

Accurate annotations achieved on the new images can be regarded as a further demonstration of the performance of the proposed system. The underlying assumption is that if the correspondences are well established across an image set A, a statistical model constructed using those correspondences should give good fitting results on a new image set B that is similar to A. Furthermore, suppose we would like to establish correspondences across a large set of images, say hundreds. If we feed all the images to the system directly, the running time may be long, as shown in the previous chapters. Hence, a more efficient way is to first apply the system to a small set of images and then use the model learned from this subset to establish correspondences on the rest of images. Finally, as annotating new images is a reasonable application of the system, it will be useful to explore which kind of statistical models can work with the system best.

In this chapter we describe three different methods of using the registered data to label the new images: (1) a simple TPS interpolation; (2) a combination of a PDM and an AAM; (3) an elastic mesh model. All three methods have the same first step, that is, a parts+geometry model learned on a training set is used to search the new images for the sparse points. The difference is that the first method does not use statistical models while the other methods do. Specifically, the former uses a

---

⋆   Parts of this chapter appeared as "Automatic Part Selection for Groupwise Registration" in the proceeding of IPMI 2011 [104] and will appear in IEEE Transactions on Medical Imaging.

TPS interpolation based on the sparse points to obtain the automatic dense points, while the latter achieve this by using the sparse points to initialise statistical models. Below we assume that the sparse points have been localised on the new images.

6.1 Methods

6.1.1 TPS Interpolation

The TPS interpolation is possible because we use the same parts+geometry model to search both the training images and new images. By randomly selecting an image from the training set as the reference image, we can easily project the dense points on the reference to the new images using a TPS interpolation, as shown in Figure 3.6a-b. As this may introduce bias to the projected points, we instead project a mean, which is constructed by warping the set of dense points on each training image to a reference frame and averaging the sets of aligned points.

6.1.2 PDM+AAM

We use the dense points on the training images to construct an AAM (*e.g.* the white points in Figure 3.6a-b). In addition, for each training image we concatenate together the sparse points and the dense points (*e.g.* both the blue and white points in Figure 3.6a-b), and build a PDM from the joint sets of points. This captures the relationship between the sparse and the dense points.

Given a new image, we use the learned parts+geometry model to localise the sparse points, which are used to initialise the PDM to predict the positions of the dense points (assumed unknown and assigned zero weights in the matching). The AAM is then initialised by the predicted dense points and used to refine the positions of those points, leading to the final annotation.

6.1.3 Elastic Mesh Model

We now consider using an elastic mesh model to label the new images. The model is defined in the mean image generated during the registration phase. Since we create a sequence of mean images of increasing resolution as the registration progresses (Figure 3.6e), there is a sequence of models.

To initiate the labelling process we randomly select one image from the training set as the reference image, and project the dense points on the reference to each new image using a TPS interpolation. The projected dense points are used to initialise the mesh model at the lowest resolution, estimating the new positions of the dense points (see below), which are then used to initialise the model at higher resolution. The process repeats until all mesh models have been considered.

As before, we use $\overline{\boldsymbol{X}}$ to denote the dense points in the mean image $\bar{I}$ and $\boldsymbol{X}_k$ for the dense points on a new image I_k. The piece-wise linear warp from $\bar{I}$ to I_k is given by $\boldsymbol{z}' = W(\boldsymbol{z} : \overline{\boldsymbol{X}}; \boldsymbol{X}_k)$, where $\boldsymbol{z}$ and $\boldsymbol{z}'$ represent any pair of corresponding points in $\bar{I}$ and I_k. We define a quality of fit measure as

$$F(\boldsymbol{X}_k) = \sum_{\boldsymbol{z} \in R} \left| I_k(W(\boldsymbol{z} : \overline{\boldsymbol{X}}; \boldsymbol{X}_k)) - \bar{I}(\boldsymbol{z}) \right| / \sigma_r + \gamma F_s(\boldsymbol{X}_k) \tag{6.1}$$

where R is a region of interest in $\bar{I}$, σ_r is an estimate of the noise of the residuals, $F_s(\boldsymbol{X}_k)$ is a shape regularisation term and γ is a weighting constant (set to 0.1 in all experiments). In the following we use a locally elastic form,

$$F_s(\boldsymbol{X}_k) = \sum_i \left| \boldsymbol{x}_{ki} - \hat{\boldsymbol{c}}_{ki} \right|^2 / \sigma_z^2 \tag{6.2}$$

where $\boldsymbol{x}_{ki}$ is the position of the i-th point of $\boldsymbol{X}_k$, σ_z^2 is the variance of $|\boldsymbol{x}_{ki} - \hat{\boldsymbol{c}}_{ki}|$, and $\hat{\boldsymbol{c}}_{ki}$ is the position predicted for $\boldsymbol{x}_{ki}$ given the position of its neighbours. For an internal point in a regular mesh, this is the centre of gravity of its neighbours, elsewhere it is a suitably weighted sum of its neighbours. A gradient descent approach is used to optimise the positions of the dense points. The derivatives of $F(\boldsymbol{X}_k)$ can be computed efficiently by displacing each point in turn, as moving one point only affects the nearby triangles.

6.2   Experiments

For each dataset, we chose 100 images as the training set and used the rest of images as the test set. The training sets were the same with those used in Section 4.2.1. We used the parts+geometry models learned in Section 4.2.1 (from the extended sets) and the resulting registration to annotate the test sets.

The AAMs were built to retain sufficient modes to explain 95% of the shape and

95

Mode 1 (±2.5s.d.)

Mode 2 (±2.5s.d.)

Mode 3 (±2.5s.d.)

Mode 4 (±2.5s.d.)
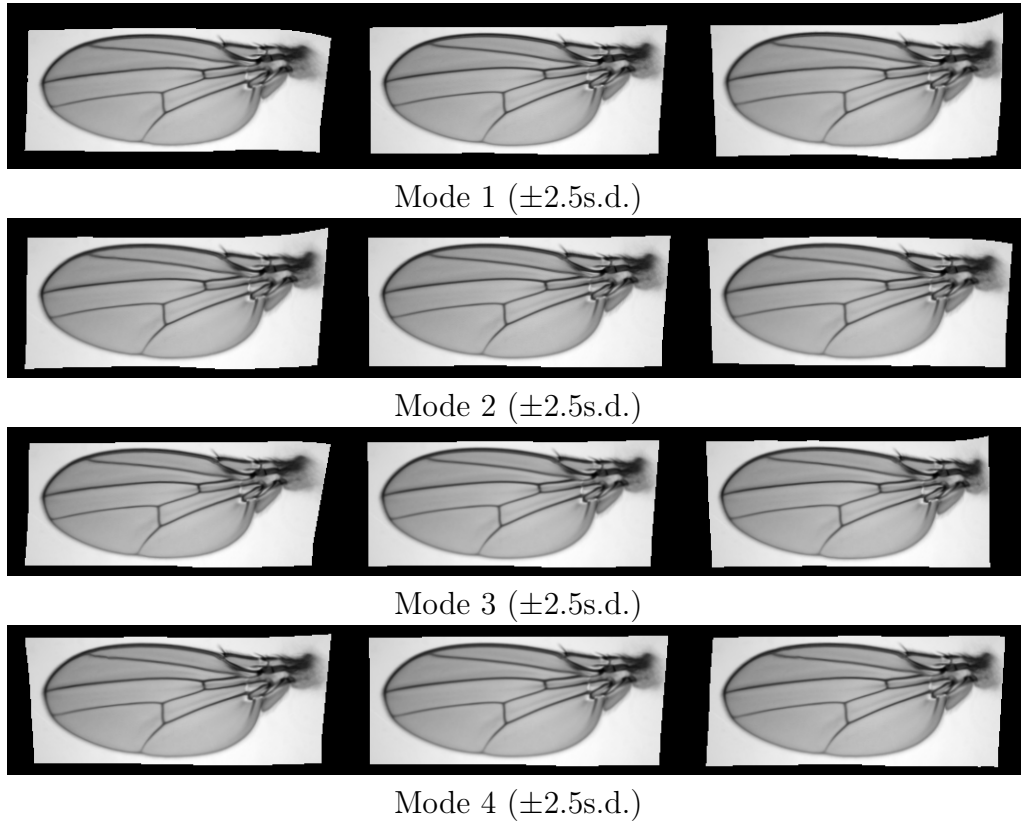
Figure 6.1: First four modes of shape variation of an AAM built from 100 automatically registered wing micrographs.

Mode 1 (±2.5s.d.)

Mode 2 (±2.5s.d.)

Mode 3 (±2.5s.d.)

Mode 4 (±2.5s.d.)
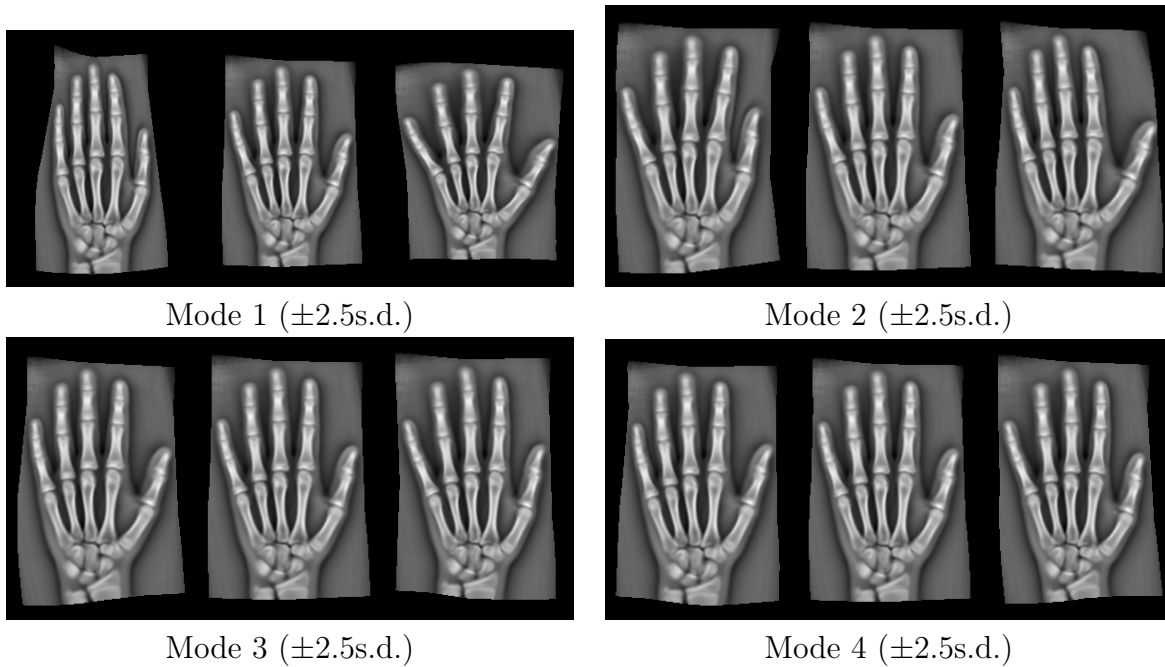
Figure 6.2: First four modes of shape variation of an AAM built from 100 automatically registered hand radiographs.

texture variance. Figures 6.1–6.3 show the first four modes of shape variation of the models built from all three datasets.

Mode 1 (±2.5s.d.) Mode 2 (±2.5s.d.)

Mode 3 (±2.5s.d.) Mode 4 (±2.5s.d.)
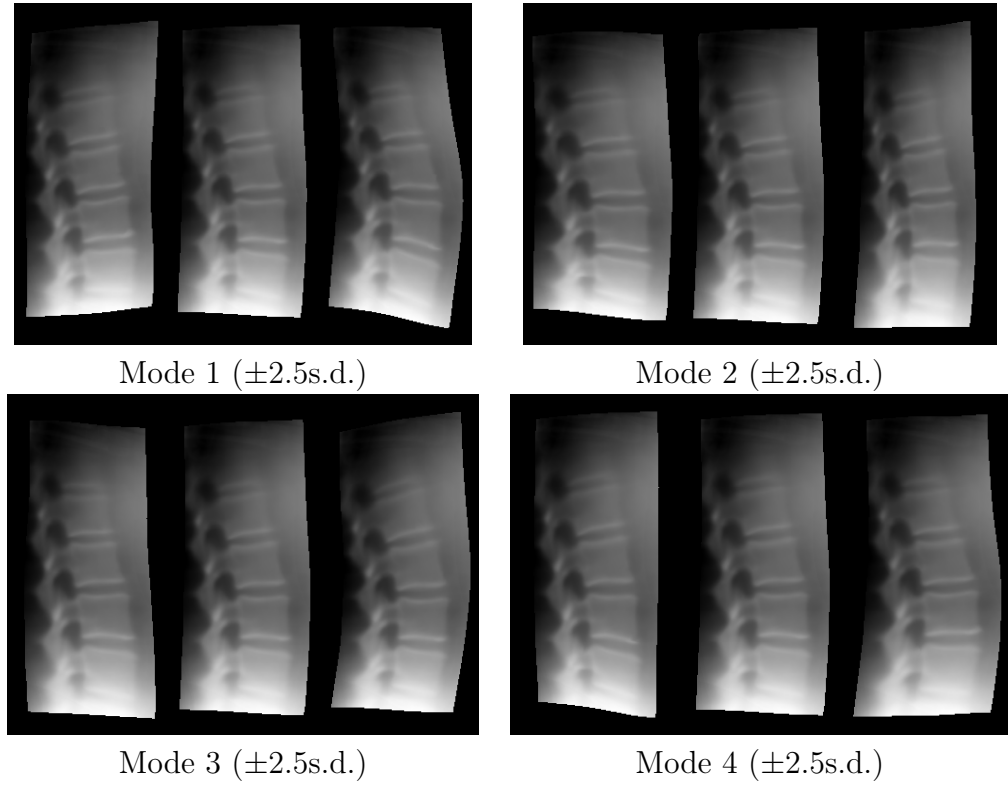
Figure 6.3: First four modes of shape variation of an AAM built from 100 automatically registered spine radiographs.

To compare the performance of the three methods, we warped the manual annotation of each training image to a reference frame, computing the mean. We then projected the mean onto each test image using the deformation field defined between the reference frame and the automatically labelled dense points. This is to avoid the bias that may be introduced by projecting the manual annotation on an arbitrarily selected training image. As before, we calculated the point-to-point location error for the fly wing and hand data, and point-to-curve error for the spines.

Tables 6.1–6.3 show the results on the three datasets. We can see that simply using TPS warping with the sparse points gives good results, but that these can be further improved by matching the elastic mesh model. Searching with the AAM may sometimes degrade results, though it works best for the spines. This is probably because the global shape model in the AAM is not sufficiently flexible to deal with local deformations, unlike the less constrained elastic mesh model. The mesh model is also the slowest method, requiring extensive non-linear optimisation. Note that the figures are comparable to those shown in Tables 4.1–4.3. This demonstrates that our system has successfully registered the given datasets.

Table 6.1: Point-to-point location errors (pixels) of the automatic annotations for the fly wings.

Methods	Mean±s.e.	Median	90%
TPS	2.6±0.4	2.0	4.7
PDM+AAM	3.1±0.2	3.2	4.2
Mesh Model	**1.5±0.2**	**1.2**	**2.7**

Table 6.2: Point-to-point location errors (mm) of the automatic annotations for the hands.

Methods	Mean±s.e.	Median	90%
TPS	1.7±0.1	1.6	2.8
PDM+AAM	1.8±0.08	1.8	2.5
Mesh Model	**1.4±0.1**	**1.4**	**2.3**

Table 6.3: Point-to-curve location errors (mm) of the automatic annotations for the spines.

Methods	Mean±s.e.	Median	90%
TPS	4.1±0.6	2.4	6.2
PDM+AAM	**3.5±0.5**	**2.1**	**4.5**
Mesh Model	3.9±0.6	2.2	6.0

6.3 Conclusions

We have described three different methods of using the correspondence resulting from the proposed system to label new images. One of them is to use a simple TPS interpolation, while the other two methods achieve the goal by constructing statistical models. Results show that robust, accurate annotations can be obtained by using a simple elastic mesh model when well initialised. The annotation accuracy on new data is comparable to the registration accuracy achieved on the training set. Unlike the PDM+AAM, the mesh model uses a looser shape constraint, which lends it more power to handle large shape variations and thus leads to more accurate annotations.

SEVEN

CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

We have described a complete system that is able to initialise groupwise non-rigid image registration effectively. This is achieved by automatically constructing one or more parts+geometry models and using its/their matches to do initialisation. The system is quite generic. It can work with any interest point descriptor and can cope with many complex objects, not restricted to those discussed in this thesis.

Extensive experiments show that our system can achieve very good results on three different datasets, two of them being particularly challenging. For example, if we compare the very best results obtained by the multi-model initialisation scheme with those from groupwise registration initialised with manual annotations (Table 7.1), we find that the accuracy of the dense correspondences established by our system is close to that of manual annotations.

We show that our system dramatically outperforms the standard affine initialisation on all three datasets. We also show that the parts automatically selected by our system can give similar or much better initialisation than a set of manually

Table 7.1: Comparison between the dense registration errors resulting from our automatic system and the manual annotations.

Dataset	Multi-model Initialisation			Manual Landmarks		
	Mean±s.e.	Median	90%	Mean±s.e.	Median	90%
Fly wings (pixels)	1.2±0.03	1.2	1.5	1.1±0.02	1.1	1.3
Hands (mm)	0.9±0.03	0.8	1.3	0.7±0.01	0.7	0.8
Spines (mm)	1.6±0.1	1.5	2.5	1.2±0.03	1.2	1.6

selected parts.

We have explored two different approaches to modelling the parts. One is a patch based method and the other is a SIFT based one. We show that the former is generally better than the latter, though less computationally efficient.

We have examined two different schemes to use the parts+geometry model/s for initialisation. One is a single-model initialisation scheme and the other is a multi-model one. We show that the latter outperforms the former in most of cases and is less sensitive to the choice of the reference image. We also show that one of the single-model initialisation schemes—the voting based method—can run quite fast without loss of accuracy.

To further demonstrate the performance of the system we use the resulting dense correspondences to build statistical models to label new images. We have compared three different approaches, showing that robust, accurate annotations can be obtained by using an elastic mesh model.

7.2 Future Work

We have mentioned some directions for future work in previous chapters. Now we would like to summarise them and point out new directions.

Effective Part Pre-selection & Efficient Part Matching A major problem of the current system is low efficiency. This is largely because of constructing and searching with huge number of parts, which can take up to approximately 91% of the total running time[1]. A straightforward approach to speeding up the system is thus to reduce the number of candidate parts. This, ideally, should propose a minimal superset of all useful parts shared by most of images. Moreover, the algorithm must be simple and fast enough, without greatly increasing the computational cost of the system. We can also speed up the system by using more efficient part matching techniques. We have developed an approach in light of the above ideas. We use β-stable features [45] to propose a small set of useful parts and use Random Ferns [70] to do fast matching. Preliminary experiments show that this approach can reduce the time of part searching to tens of minutes. However, more work is still needed for robust and accurate localisation.

---

[1] This is calculated based on Table 3.3.

Complex Graphical Structures Currently we use a tree topology for graphical models to allow efficient global optimisation. Using more sophisticated graphical structures will increase the representative ability of the models and may improve the performance of the system, as the added links may reduce the occurrence of the false graph matches. On the other hand, the complex topology will inevitably increase the inference difficulty and thus lead to a decline in efficiency. Hence, it is desirable to explore the influence of graphical structures on the performance of the system, finding out a trade-off between performance and efficiency.

Automatic Parameter Learning There are a number of parameters in the system. Most of them can be given roughly and will not affect the performance of the system too much. However, there are some key parameters, which currently have to be determined based on preliminary experiments. An example is α, which determines the localisability of parts+geometry models and may vary according to the object of interest. It is thus interesting to find effective ways to automatically determine such parameters of the system.

Dealing with Occlusions The current system can tolerate partial occlusions, which are present on some of the datasets. It cannot tackle large occlusions or missing data. Although this problem can be alleviated by using multiple parts+geometry models, it is still necessary to further develop the system to make it robust to the above outliers.

Selecting Good Parts+Geometry Models We show that the performance of the system can be boosted by using a set of roughly good parts+geometry models. We also notice that this multi-model initialisation strategy may, occasionally, perform slightly worse than the single-model one. We believe that this is not the problem of the multi-model strategy itself but the problem of the models used. If most of models have poor performance or share too many common parts, it is difficult for the system to benefit from the use of multiple models. Hence, it is interesting to explore how to choose a set of good models which have similar performance as well as minimum redundancy.

Automatic Segmentation of Mean Image After groupwise registration, we obtain a set of dense correspondences and a mean of the training images. We have

shown that a statistical model can be constructed using the dense correspondences and used to annotate new images. However, the new data may contain occlusions or missing parts, which may mislead the search of the statistical model. A way to circumvent this problem is to build a set of statistical models, each representing a local structure of the object. This can be achieved by breaking down the mean image into several pieces, projecting each piece onto the training images and constructing a model for each piece. A set of local models may lead to better annotations than a single, global model.

Extension to 3D Images   We only tested the system on 2D images. However, the key ideas extend naturally to 3D images, though computational efficiency issues may have to be addressed.

A

EXPERIMENTAL RESULTS

Table A.1: Point-to-point location errors (pixels) of the dense correspondence for the fly wings.

(a)

Method	m	Mean±s.e.	Med.	90%
Opt.	5	1.4±0.04	1.4	1.9
	10	1.8±0.1	1.5	2.7
	15	1.5±0.1	1.3	2.0
	20	**1.4±0.05**	**1.3**	**1.9**
	25	1.6±0.05	1.5	2.4
Voting	21	1.5±0.05	1.4	2.1
Mulit.	5	1.4±0.04	1.3	2.0
	10	**1.3±0.03**	**1.3**	**1.8**
	15	1.4±0.04	1.3	2.0
	20	1.4±0.04	1.3	1.8
	25	1.6±0.06	1.4	2.4

(b)

Method	m	Mean±s.e.	Med.	90%
Opt.	5	1.8±0.06	1.6	2.7
	10	3.7±0.2	3.4	6.3
	15	**1.5±0.06**	**1.3**	**2.3**
	20	2.0±0.08	1.8	3.0
	25	1.8±0.07	1.6	2.9
Voting	27	1.6±0.1	1.2	2.1
Mulit.	5	1.5±0.05	1.3	2.0
	10	1.7±0.05	1.6	2.5
	15	**1.4±0.03**	**1.3**	**1.8**
	20	1.4±0.05	1.3	1.9
	25	1.7±0.06	1.6	2.5

(c)

Method	m	Mean±s.e.	Med.	90%
Opt.	5	**1.5±0.06**	**1.3**	**2.1**
	10	2.3±0.1	1.9	4.1
	15	2.5±0.1	2.3	3.9
	20	2.0±0.07	1.8	3.0
	25	1.7±0.05	1.7	2.3
Voting	26	1.3±0.04	1.2	1.8
Mulit.	5	**1.3±0.03**	**1.2**	**1.6**
	10	1.4±0.03	1.4	1.8
	15	1.6±0.06	1.5	2.5
	20	1.8±0.08	1.6	2.9
	25	1.7±0.05	1.6	2.4

(d)

Method	m	Mean±s.e.	Med.	90%
Opt.	5	2.0±0.1	1.5	3.3
	10	2.9±0.1	2.5	5.0
	15	2.5±0.1	2.4	4.0
	20	2.1±0.08	1.9	3.4
	25	**1.6±0.06**	**1.4**	**2.4**
Voting	22	1.6±0.08	1.4	2.5
Mulit.	5	**1.4±0.05**	**1.3**	**1.9**
	10	1.6±0.05	1.4	2.3
	15	1.8±0.07	1.6	2.9
	20	1.7±0.06	1.6	2.7
	25	1.6±0.07	1.4	2.4

(e)

Method	m	Mean±s.e.	Med.	90%
	5	2.2±0.2	1.7	3.7
	10	2.6±0.1	2.0	4.6
Opt.	15	2.2±0.1	1.9	3.6
	20	**2.1±0.1**	**1.7**	**3.8**
	25	2.2±0.1	1.9	3.6
Voting	24	1.6±0.08	1.3	2.9
	5	1.8±0.09	1.5	2.7
	10	1.7±0.06	1.5	2.5
Mulit.	15	**1.6±0.06**	**1.4**	**2.3**
	20	1.8±0.07	1.7	2.8
	25	1.9±0.08	1.6	3.1

(f)

Method	m	Mean±s.e.	Med.	90%
	5	**1.4±0.04**	**1.3**	**1.9**
	10	1.8±0.07	1.6	2.8
Opt.	15	1.7±0.08	1.5	2.9
	20	1.9±0.08	1.7	3.0
	25	2.2±0.08	2.0	3.6
Voting	24	1.1±0.02	1.1	1.4
	5	**1.2±0.04**	**1.2**	**1.6**
	10	1.6±0.05	1.4	2.4
Mulit.	15	1.5±0.05	1.4	2.1
	20	1.5±0.07	1.3	2.3
	25	1.5±0.05	1.4	2.1

(g)

Method	m	Mean±s.e.	Med.	90%
	5	2.5±0.1	2.1	3.9
	10	**2.0±0.07**	**1.9**	**2.7**
Opt.	15	2.5±0.1	2.1	3.8
	20	2.4±0.1	2.2	3.9
	25	2.8±0.1	2.4	4.6
Voting	18	1.2±0.03	1.1	1.6
	5	**1.4±0.04**	**1.3**	**2.0**
	10	1.9±0.08	1.7	2.9
Mulit.	15	2.0±0.08	1.8	2.9
	20	1.9±0.08	1.7	3.1
	25	1.8±0.08	1.5	3.0

(h)

Method	m	Mean±s.e.	Med.	90%
	5	**1.7±0.3**	**1.4**	**1.9**
	10	2.1±0.1	1.9	3.6
Opt.	15	3.2±0.1	3.0	4.8
	20	2.0±0.08	1.8	3.0
	25	2.6±0.1	2.5	3.7
Voting	25	1.3±0.05	1.2	1.8
	5	**1.3±0.04**	**1.3**	**1.7**
	10	2.1±0.1	1.8	3.4
Mulit.	15	2.3±0.08	2.1	3.4
	20	1.7±0.06	1.6	2.5
	25	2.0±0.08	1.8	3.0

(i)

Method	m	Mean±s.e.	Med.	90%
	5	1.7±0.1	1.4	2.4
	10	1.5±0.06	1.4	2.2
Opt.	15	**1.4±0.04**	**1.3**	**1.9**
	20	2.3±0.09	2.2	3.4
	25	2.2±0.09	2.2	3.6
Voting	27	1.8±0.1	1.4	3.2
	5	**1.3±0.03**	**1.3**	**1.7**
	10	1.3±0.04	1.3	1.9
Mulit.	15	1.5±0.06	1.4	2.1
	20	1.5±0.05	1.4	2.2
	25	1.6±0.05	1.4	2.4

(j)

Method	m	Mean±s.e.	Med.	90%
	5	1.5±0.07	1.3	1.9
	10	**1.2±0.03**	**1.2**	**1.5**
Opt.	15	1.4±0.04	1.3	1.8
	20	1.9±0.06	1.7	2.7
	25	1.7±0.07	1.6	2.4
Voting	22	1.9±0.1	1.6	3.1
	5	1.4±0.04	1.3	1.9
	10	**1.2±0.03**	**1.2**	**1.5**
Mulit.	15	1.2±0.03	1.2	1.5
	20	1.3±0.04	1.2	1.7
	25	1.5±0.04	1.3	2.1

Table A.2: Point-to-point location errors (mm) of the dense correspondence for the hands.

(a)

Method	m	Mean±s.e.	Med.	90%
Opt.	10	2.0±0.2	1.1	4.8
	20	**1.0±0.04**	**0.9**	**1.5**
	30	1.4±0.1	1.0	3.1
	40	1.2±0.1	0.9	2.3
	50	1.2±0.09	0.9	1.7
Voting	44	1.0±0.06	0.8	1.5
Mulit.	10	1.8±0.1	1.2	3.8
	20	**1.1±0.05**	**0.9**	**1.7**
	30	1.2±0.06	1.0	2.2
	40	1.2±0.07	1.0	2.0
	50	1.2±0.08	1.0	1.7

(b)

Method	m	Mean±s.e.	Med.	90%
Opt.	10	1.5±0.07	1.3	2.3
	20	1.2±0.04	1.1	1.9
	30	1.2±0.06	1.1	1.7
	40	**1.1±0.05**	**1.0**	**1.6**
	50	1.5±0.07	1.3	2.1
Voting	45	1.3±0.09	1.0	2.2
Mulit.	10	1.6±0.06	1.4	2.2
	20	1.2±0.04	1.2	1.7
	30	**1.1±0.04**	**1.0**	**1.6**
	40	1.2±0.04	1.1	1.7
	50	1.2±0.05	1.0	1.8

(c)

Method	m	Mean±s.e.	Med.	90%
Opt.	10	2.0±0.2	1.4	3.2
	20	1.3±0.09	1.1	1.9
	30	1.4±0.2	0.9	2.2
	40	1.3±0.1	0.9	2.0
	50	**1.3±0.1**	**0.9**	**1.7**
Voting	36	1.1±0.06	1.0	1.4
Mulit.	10	1.5±0.08	1.3	2.4
	20	1.4±0.05	1.2	1.9
	30	1.2±0.08	1.0	2.1
	40	**1.2±0.07**	**1.0**	**1.9**
	50	1.3±0.08	1.0	1.9

(d)

Method	m	Mean±s.e.	Med.	90%
Opt.	10	2.0±0.1	1.7	3.0
	20	1.1±0.09	0.9	1.4
	30	**1.0±0.05**	**0.9**	**1.5**
	40	1.4±0.1	1.1	2.3
	50	1.2±0.08	0.9	2.0
Voting	40	1.0±0.04	0.9	1.4
Mulit.	10	1.7±0.08	1.5	2.5
	20	1.0±0.04	0.9	1.5
	30	1.2±0.06	1.0	2.0
	40	1.0±0.04	0.9	1.4
	50	**0.9±0.04**	**0.9**	**1.3**

(e)

Method	m	Mean±s.e.	Med.	90%
Opt.	10	1.4±0.08	1.2	2.1
	20	**1.0±0.09**	**0.8**	**1.2**
	30	1.1±0.05	1.0	1.7
	40	1.2±0.07	1.0	2.0
	50	1.1±0.09	0.9	2.0
Voting	43	1.4±0.08	1.1	2.9
Mulit.	10	1.4±0.06	1.2	2.0
	20	1.0±0.05	0.9	1.4
	30	1.0±0.05	0.9	1.5
	40	**0.9±0.03**	**0.8**	**1.4**
	50	0.9±0.04	0.9	1.3

(f)

Method	m	Mean±s.e.	Med.	90%
Opt.	10	1.6±0.1	1.3	2.2
	20	**1.1±0.04**	**0.9**	**1.6**
	30	1.2±0.05	1.1	1.6
	40	1.4±0.09	1.0	2.3
	50	1.2±0.04	1.1	1.7
Voting	46	1.0±0.05	0.9	1.4
Mulit.	10	1.7±0.1	1.4	2.5
	20	1.1±0.04	0.9	1.6
	30	1.0±0.04	0.9	1.5
	40	**1.0±0.03**	**0.9**	**1.4**
	50	1.1±0.04	1.0	1.5

(g)

Method	m	Mean±s.e.	Med.	90%
Opt.	10	1.7±0.1	1.4	3.0
	20	1.1±0.07	0.8	1.7
	30	1.6±0.1	1.1	2.8
	40	**1.1±0.07**	**0.9**	**1.6**
	50	1.1±0.06	0.9	1.9
Voting	37	1.2±0.07	1.0	2.1
Mulit.	10	1.9±0.1	1.5	3.3
	20	1.2±0.07	0.9	2.0
	30	**1.0±0.04**	**0.9**	**1.5**
	40	1.0±0.05	0.9	1.5
	50	1.1±0.05	0.9	1.9

(h)

Method	m	Mean±s.e.	Med.	90%
Opt.	10	1.3±0.1	1.1	1.8
	20	1.5±0.07	1.2	2.1
	30	1.2±0.09	0.8	2.0
	40	1.2±0.08	0.8	2.1
	50	**1.1±0.08**	**0.9**	**1.6**
Voting	41	1.0±0.07	0.9	1.4
Mulit.	10	1.4±0.08	1.1	2.3
	20	1.1±0.05	1.0	1.7
	30	1.0±0.05	0.9	1.4
	40	0.9±0.04	0.8	1.3
	50	**0.9±0.03**	**0.8**	**1.3**

(i)

Method	m	Mean±s.e.	Med.	90%
Opt.	10	1.7±0.1	1.5	2.9
	20	1.5±0.07	1.1	2.8
	30	1.2±0.07	1.0	1.8
	40	**1.1±0.05**	**0.9**	**2.0**
	50	1.1±0.06	0.9	2.3
Voting	44	1.0±0.05	0.9	1.4
Mulit.	10	1.4±0.06	1.2	2.1
	20	1.4±0.07	1.2	2.1
	30	1.2±0.05	1.1	1.7
	40	1.1±0.04	1.0	1.6
	50	**1.0±0.04**	**0.9**	**1.6**

(j)

Method	m	Mean±s.e.	Med.	90%
Opt.	10	2.1±0.1	1.6	4.7
	20	1.2±0.08	1.0	1.8
	30	**1.2±0.09**	**1.0**	**1.7**
	40	2.4±0.1	2.2	3.9
	50	1.3±0.07	0.9	2.1
Voting	29	1.2±0.1	0.9	1.6
Mulit.	10	1.7±0.07	1.5	2.6
	20	1.2±0.05	1.1	1.9
	30	1.2±0.06	0.9	1.7
	40	1.3±0.07	1.0	2.4
	50	**1.1±0.06**	**0.9**	**1.7**

Table A.3: Point-to-curve location errors (mm) of the dense correspondence for the spines.

(a)

Method	m	Mean±s.e.	Med.	90%
	5	7.8±0.9	2.8	25.0
	10	3.9±0.5	2.2	6.1
Opt.	15	**2.5±0.3**	**1.9**	**3.6**
	20	3.8±0.4	2.6	5.9
Voting	26	6.4±0.5	4.5	13.2
	5	2.3±0.3	1.6	3.2
	10	2.7±0.4	1.7	4.4
Mulit.	15	2.1±0.3	1.6	3.1
	20	**2.1±0.1**	**1.7**	**3.1**

(b)

Method	m	Mean±s.e.	Med.	90%
	5	4.8±0.6	3.0	7.0
	10	3.3±0.5	2.1	4.3
Opt.	15	**2.6±0.3**	**1.8**	**3.4**
	20	3.4±0.4	2.0	6.4
Voting	20	9.8±0.5	8.0	19.0
	5	**2.1±0.3**	**1.6**	**2.8**
	10	2.2±0.3	1.6	2.6
Mulit.	15	2.7±0.3	2.0	3.1
	20	2.7±0.3	2.3	3.2

(c)

Method	m	Mean±s.e.	Med.	90%
	5	5.7±0.7	3.4	8.8
	10	**4.4±0.6**	**2.7**	**6.0**
Opt.	15	6.9±0.6	4.7	20.6
	20	6.8±0.6	4.4	20.5
Voting	25	11.0±0.5	9.7	18.3
	5	3.7±0.3	2.7	6.6
	10	3.4±0.4	2.1	4.5
Mulit.	15	**2.7±0.3**	**1.9**	**4.0**
	20	3.6±0.4	2.4	6.0

(d)

Method	m	Mean±s.e.	Med.	90%
	5	7.9±0.7	5.0	21.3
	10	5.5±0.7	2.9	17.8
Opt.	15	4.7±0.5	2.9	7.0
	20	**4.0±0.5**	**2.5**	**7.6**
Voting	21	3.6±0.4	2.3	6.5
	5	**2.1±0.1**	**1.8**	**3.1**
	10	2.4±0.3	1.7	3.8
Mulit.	15	2.4±0.3	1.7	3.3
	20	2.4±0.3	1.9	3.1

(e)

Method	m	Mean±s.e.	Med.	90%
	5	6.3±0.9	2.4	26.1
	10	**3.6±0.5**	**1.9**	**5.3**
Opt.	15	6.0±0.6	3.8	19.6
	20	5.0±0.5	3.1	8.5
Voting	31	6.5±0.4	5.3	10.8
	5	3.3±0.4	2.0	6.2
	10	2.5±0.3	1.7	3.4
Mulit.	15	**2.1±0.1**	**1.7**	**3.2**
	20	2.3±0.1	1.9	3.9

(f)

Method	m	Mean±s.e.	Med.	90%
	5	9.1±0.7	5.6	19.9
	10	4.0±0.5	2.4	4.8
Opt.	15	3.3±0.5	1.9	5.1
	20	**3.3±0.5**	**1.6**	**4.1**
Voting	27	6.7±0.5	4.7	15.9
	5	2.7±0.3	1.8	4.4
	10	2.2±0.3	1.7	2.6
Mulit.	15	2.6±0.3	1.9	3.3
	20	**1.9±0.1**	**1.6**	**3.1**

(g)

Method	m	Mean±s.e.	Med.	90%
Opt.	5	4.7±0.6	2.7	9.4
	10	3.7±0.5	2.4	4.0
	15	**3.5±0.5**	**2.1**	**5.9**
	20	4.6±0.4	3.3	6.8
Voting	16	4.8±0.4	3.6	9.2
Mulit.	5	2.2±0.2	1.7	3.0
	10	**2.0±0.3**	**1.6**	**2.6**
	15	2.2±0.3	1.7	2.8
	20	2.1±0.1	1.9	3.1

(h)

Method	m	Mean±s.e.	Med.	90%
Opt.	5	7.4±0.8	3.7	21.1
	10	**4.4±0.5**	**2.6**	**8.5**
	15	4.8±0.5	2.9	10.6
	20	7.0±0.6	4.3	19.9
Voting	15	9.3±0.4	8.2	14.2
Mulit.	5	3.7±0.4	2.5	6.0
	10	**3.6±0.4**	**2.5**	**5.1**
	15	4.8±0.4	3.5	7.1
	20	4.7±0.4	3.1	7.3

(i)

Method	m	Mean±s.e.	Med.	90%
Opt.	5	7.4±0.8	3.3	22.6
	10	**3.6±0.5**	**2.2**	**3.9**
	15	6.5±0.7	3.7	20.2
	20	8.1±0.8	4.2	22.5
Voting	14	13.6±0.6	12.6	20.9
Mulit.	5	3.5±0.4	2.1	5.9
	10	**2.6±0.2**	**2.0**	**3.9**
	15	3.8±0.4	2.4	7.3
	20	3.5±0.3	2.6	5.0

(j)

Method	m	Mean±s.e.	Med.	90%
Opt.	5	4.9±0.7	2.4	8.4
	10	3.3±0.6	1.9	3.5
	15	**2.7±0.4**	**1.8**	**4.3**
	20	2.7±0.3	1.9	4.4
Voting	37	4.4±0.4	3.4	5.7
Mulit.	5	2.3±0.4	1.6	2.7
	10	**1.6±0.1**	**1.5**	**2.5**
	15	1.6±0.1	1.4	2.6
	20	1.6±0.1	1.4	2.8

REFERENCES

[1] S. A. Adeshina and T. F. Cootes. Constructing part-based models for groupwise registration. In *Proceedings of IEEE Symposium on Biomedical Imaging*, pages 1073–1076, 2010.

[2] V. Arsigny, O. Commowick, X. Pennec, and N. Ayache. A log-euclidean polyaffine framework for locally rigid or affine registration. In *Proceedings of International Workshop on Biomedical Image Registration*, pages 120–127, 2006.

[3] K. Babalola and T. Cootes. Using parts and geometry models to initialise active appearance models for automated segmentation of 3d medical images. In *Proceedings of IEEE Symposium on Biomedical Imaging*, pages 1069–1072, 2010.

[4] S. Baker, I. Matthews, and J. Schneider. Automatic construction of active appearance models as an image coding problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1380–1384, 2004.

[5] S. K. Balci, P. Golland, M. Shenton, and W. M. Wells. Free-form B-spline deformation model for groupwise registration. In *Proceedings of MICCAI Statistical Registration Workshop*, pages 23–30, 2007.

[6] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape context. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.

[7] K. K. Bhatia, J. V. Hajnal, B. K. Puri, A. D. Edwards, and D. Rueckert. Consistent groupwise non-rigid registration for atlas construction. In *Proceedings of IEEE Symposium on Biomedical Imaging*, pages 908–911, 2004.

[8] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In A. Rockwood, editor, *Proceedings of SIGGRAPH*, pages 187–194, Los Angeles, 1999. Addison Wesley Longman.

[9] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989.

[10] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

[11] M. Burl, M. Weber, and P. Perona. A probabilistic approach to object recognition using local photometry and global geometry. In H. Burkhardt and B. Neumann, editors, *Proceedings of European Conference on Computer Vision*, volume 1407, pages 628–641, 1998.

[12] M. C. Burl, T. K. Leung, and P. Perona. Face localization via shape statistics. In *Proceedings of International Workshop on Automatic Face and Gesture Recognition*, Zurich, Switzerland, 1995.

[13] M. C. Burl and P. Perona. Recognition of planar object classes. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 223–230, 1996.

[14] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[15] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In H. Burkhardt and B. Neumann, editors, *Proceedings of European Conference on Computer Vision*, volume 1407, pages 484–498, 1998.

[16] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[17] T. F. Cootes, S. Marsland, C. J. Twining, K. Smith, and C. J. Taylor. Groupwise diffeomorphic non-rigid registration for automatic model building. In T. Pajdla and J. Matas, editors, *Proceedings of European Conference on Computer Vision*, volume 3024, pages 316–327, 2004.

[18] T. F. Cootes and C. J. Taylor. Active shape models - 'smart snakes'. In D. Hogg and R. Boyle, editors, *Proceedings of British Machine Vision Conference*, pages 266–275, 1992.

[19] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[20] T. F. Cootes, C. J. Twining, V. Petrović, K. O. Babalola, and C. J. Taylor. Computing accurate correspondences across groups of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):1994–2005, 2010.

[21] T. F. Cootes, C. J. Twining, V. Petrović, R. Schestowitz, and C. J. Taylor. Groupwise construction of appearance models using piece-wise affine deformations. In

W. Clocksin, A. Fitzgibbon, and P. Torr, editors, *Proceedings of British Machine Vision Conference*, volume 2, pages 879–888, 2005.

[22] D. J. Crandall, P. Felzenszwalb, and D. P. Huttenlocher. Spatial priors for part-based recognition using statistical models. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 10–17, 2005.

[23] D. J. Crandall and D. P. Huttenlocher. Weakly supervised learning of part-based spatial models for visual object recognition. In A. Leonardis, H. Bischof, and A. Pinz, editors, *Proceedings of European Conference on Computer Vision*, volume 3951, pages 16–29, 2006.

[24] R. H. Davies, T. F. Cootes, and C. J. Taylor. A minimum description length approach to statistical shape modelling. In M. Insana and R. Leahy, editors, *Proceedings of International Conference on Information Processing in Medical Imaging*, volume 2082, pages 50–63, 2001.

[25] R. H. Davies, T. F. Cootes, C. J. Twining, and C. J. Taylor. An information theoretic approach to statistical shape modelling. In T. F. Cootes and C. J. Taylor, editors, *Proceedings of British Machine Vision Conference*, pages 3–11, 2001.

[26] R. H. Davies, T. F. Cootes, J. C. Waterton, and C. J. Taylor. An efficient method of constructing optimal statistical shape models. In W. Niessen and M. Viergever, editors, *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 2208, pages 57–65, 2001.

[27] R. H. Davies, C. Twining, T. F. Cootes, and C. J. Taylor. A minimum description length approach to statistical shape modelling. *IEEE Transactions on Medical Imaging*, 21(5):525–537, 2002.

[28] R. H. Davies, C. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. 3D statistical shape models using direct optimisation of description length. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Proceedings of European Conference on Computer Vision*, volume 2352, pages 1–17, 2002.

[29] F. De la Torre Frade and M. Nguyen. Parameterized kernel principal component analysis: Theory and applications to supervised and unsupervised image alignment. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[30] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[31] R. Donner, B. Micusik, G. Langs, and H. Bischof. Sparse MRF appearance models for fast anatomical structure localisation. In A. Bhalerao and N. Rajpoot, editors,

Proceedings of British Machine Vision Conference, volume 2, pages 1080–1089, 2007.

[32] R. Donner, H. Wildenauer, H. Bischof, and G. Langs. Weakly supervised group-wise model learning based on discrete optimization. In G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, editors, *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 5762, pages 860–868, 2009.

[33] I. Dryden and K. Mardia. *Statistical Shape Analysis*. JohnWiley & Sons, 1998.

[34] J. Duchon. Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. *R.A.I.R.O. Analyse Numérique*, 10:5–12, 1976.

[35] P. F. Felzenszwalb and D. P. Huttenlocher. Pictorial structures for object recognition. *International Journal of Computer Vision*, 61(1):55–79, 2005.

[36] P. F. Felzenszwalb and D. P. Huttenlocher. Representation and detection of deformable shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2):208–220, 2005.

[37] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2003.

[38] R. Fergus, P. Perona, and A. Zisserman. A visual category filter for google images. In T. Pajdla and J. Matas, editors, *Proceedings of European Conference on Computer Vision*, volume 3021, pages 242–256, 2004.

[39] R. Fergus, P. Perona, and A. Zisserman. A sparse object category model for efficient learning and exhaustive recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 380–387, 2005.

[40] V. Ferrari, F. Jurie, and C. Schmid. From images to shape models for object detection. *International Journal of Computer Vision*, 87(3):284–303, 2010.

[41] M. A. Fischler and R. A. Elschlager. The representation and matching of pictorial structures. *IEEE Transactions on Computer*, 22(1):67–92, 1973.

[42] A. F. Frangi, D. Rueckert, J. A. Schnabel, and W. J. Niessen. Automatic 3D ASM construction via atlas-based landmarking and volumetric elastic registration. In M. Insana and R. Leahy, editors, *Proceedings of International Conference on Information Processing in Medical Imaging*, volume 2082, pages 78–91, 2001.

[43] 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 modeling. *IEEE Transactions on Medical Imaging*, 21(9):1151–1166, 2002.

[44] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine Vision and Applications*, 6:35–49, 1993.

[45] S. Gu, Y. Zheng, and C. Tomasi. Critical nets and beta-stable features for image matching. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Proceedings of European Conference on Computer Vision*, volume 6313, pages 663–676, 2010.

[46] A. Guimond, J. Meunier, and J.-P. Thirion. Automatic computation of average brain models. In W. Wells, A. Colchester, and S. Delp, editors, *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 1496, pages 631–640, 1998.

[47] A. Guimond, J. Meunier, and J.-P. Thirion. Average brain models: A convergence study. Technical Report RR-3731, INRIA, Sophia Antipolis, July 1999.

[48] G. Huang, V. Jain, and E. G. Learned-Miller. Unsupervised joint alignment of complex images. In *Proceedings of IEEE Conference on International Conference on Computer Vision*, pages 1 –8, 2007.

[49] M. Jones and T. Poggio. Multidimensional morphable models: A framework for representing and matching object classes. In *Proceedings of IEEE Conference on International Conference on Computer Vision*, pages 683–688, 1998.

[50] S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23(Supplement 1):151–160, 2004.

[51] T. Kadir and M. Brady. Saliency, scale and image description. *International Journal of Computer Vision*, 45(2):83–105, 2001.

[52] J. Karlsson and K. Åström. MDL patch correspondence on unlabeled images with occlusions. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[53] I. Kokkinos and P. Maragos. Synergy between object recognition and image segmentation using the expectation-maximization algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(8):1486–1501, 2009.

[54] I. Kokkinos and A. Yuille. Unsupervised learning of object deformation models. In *Proceedings of IEEE Conference on International Conference on Computer Vision*, pages 1–8, 2007.

[55] G. Langs, R. Donner, P. Peloschek, and H. Bischof. Robust autonomous model learning from 2D and 3D data sets. In N. Ayache, S. Ourselin, and A. Maeder, editors, *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 4791, pages 968–976, 2007.

[56] G. Langs, P. Peloschek, R. Donner, and H. Bischof. Annotation propagation by MDL based correspondences. In O. Chum and V. Franc, editors, *Proceedings of Computer Vision Winter Workshop*, 2006.

[57] A. Lanitis, C. J. Taylor, and T. F. Cootes. Automatic interpretation and coding

of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):743–756, 1997.

[58] E. G. Learned-Miller. Data driven image models through continuous joint alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):236–250, 2006.

[59] V. Lepetit and P. Fua. Keypoint recognition using randomized trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1465–1479, 2006.

[60] B. Likar and F. Pernu*vs*. A hierarchical approach to elastic registration based on mutual information. *Image and Vision Computing*, 19(1-2):33–44, 2001.

[61] X. Liu, Y. Tong, F. Wheeler, and P. Tu. Facial contour labeling via congealing. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Proceedings of European Conference on Computer Vision*, volume 6311, pages 354–368, 2010.

[62] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *ACM Siggraph Computer Graphics*, 21(4):163–169, 1987.

[63] P. Lorenzen, B. Davis, and S. Joshi. Unbiased atlas formation via large deformations metric mapping. In J. Duncan and G. Gerig, editors, *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 3750, pages 411–418, 2005.

[64] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[65] C. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing.* MIT Press, 1999.

[66] S. Marsland, C. Twining, and C. J. Taylor. Groupwise non-rigid registration using polyharmonic clamped-plate splines. In R. Ellis and T. Peters, editors, *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 2879, pages 771–779, 2003.

[67] M. A. Martín-Fernández, R. Cárdenes, E. Muñoz Moreno, R. de Luis-García, M. Martín-Fernández, and C. Alberola-López. Automatic articulated registration of hand radiographs. *Image and Vision Computing*, 27:1207–1222, 2009.

[68] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.

[69] M. Minoux. *Mathematical programming: theory and algorithms.* Wiley, 1986.

[70] M. Özuysal, M. Calonder, V. Lepetit, and P. Fua. Fast keypoint recognition using random ferns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):448–461, 2010.

[71] X. Papademetris, D. Dione, L. Dobrucki, L. Staib, and A. Sinusas. Articulated rigid

registration for serial lower-limb mouse imaging. In J. Duncan and G. Gerig, editors, *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 3750, pages 919–926, 2005.

[72] A. Pitiot, G. Malandain, E. Bardinet, and P. Thompson. Piecewise affine registration of biological images. In J. Gee, J. Maintz, and M. Vannier, editors, *Proceedings of International Workshop on Biomedical Image Registration*, volume 2717, pages 91–101, 2003.

[73] J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, editors. *Towards Category-Level Object Recognition.* Springer-Verlag, 2006.

[74] V. Potesil, T. Kadir, G. Platsch, and M. Brady. Personalization of pictorial structures for anatomical landmark localization. In G. Székely and H. Hahn, editors, *Proceedings of International Conference on Information Processing in Medical Imaging*, volume 6801, pages 333–345, 2011.

[75] J. Rissanen. *Stochastic Complexity in Statistical Inquiry.* World Scientific Publishing Co., Inc., 1989.

[76] D. Rueckert, A. Frangi, and J. Schnabel. Automatic construction of 3d statistical deformation models using non-rigid registration. In W. Niessen and M. Viergever, editors, *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 2208, pages 77–84, 2001.

[77] D. Rueckert and A. F. Frangi. Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging*, 22(8):1014–1025, 2003.

[78] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.

[79] S. Schmidt, J. Kappes, M. Bergtholdt, V. Pekar, S. Dries, D. Bystrov, and C. Schnörr. Spine detection and labeling using a parts-based graphical model. In N. Karssemeijer and B. Lelieveldt, editors, *Proceedings of International Conference on Information Processing in Medical Imaging*, volume 4584, pages 122–133, 2007.

[80] S. Sclaroff and J. Isidoro. Active blobs. In *Proceedings of IEEE Conference on International Conference on Computer Vision*, pages 1146–1153, 1998.

[81] D. Seghers, J. Hermans, D. Loeckx, F. Maes, D. Vandermeulen, and P. Suetens. Model-based segmentation using graph representations. In D. Metaxas, L. Axel, G. Fichtinger, and G. Székely, editors, *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 5241, pages 393–400, 2008.

[82] C. E. Shannon. Communication in the Presence of Noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

[83] K. Sidorov, D. Marshall, and S. Richmond. An efficient stochastic approach to groupwise non-rigid image registration. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2208–2213, 2009.

[84] C. Studholme, D. L. G. Hill, and D. J. Hawkes. Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. *Medical Physics*, 24(1):25–35, 1997.

[85] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999.

[86] J.-P. Thirion. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998.

[87] M. Toews and T. Arbel. A statistical parts-based model of anatomical variability. *IEEE Transactions on Medical Imaging*, 26(4):497–508, 2007.

[88] Y. Tong, X. Liu, F. W. Wheeler, and P. Tu. Automatic facial landmark labeling with minimal supervision. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2097–2104, 2009.

[89] C. Twining, S. Marsland, and C. Taylor. Measuring geodesic distances on the space of bounded diffeomorphisms. In D. Marshall and P. L. Rosin, editors, *Proceedings of British Machine Vision Conference*, pages 847–856, 2002.

[90] C. J. Twining, T. F. Cootes, S. Marsland, V. Petrović, R. Schestowitz, and C. J. Taylor. A unified information-theoretic approach to groupwise non-rigid registration and model building. In G. Christensen and M. Sonka, editors, *Proceedings of International Conference on Information Processing in Medical Imaging*, volume 3565, pages 167–198, 2005.

[91] T. Vetter, M. J. Jones, and T. Poggio. A bootstrapping algorithm for learning linear models of object classes. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 40–46, 1997.

[92] T. Vetter and T. Poggio. Linear object classes and image synthesis from a single example image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):733–741, 1997.

[93] K. N. Walker, T. F. Cootes, and C. J. Taylor. Automatically building appearance models from image sequences using salient features. In T. Pridmore and D. Elliman, editors, *Proceedings of British Machine Vision Conference*, pages 463–472, 1999.

[94] K. N. Walker, T. F. Cootes, and C. J. Taylor. Determining correspondences for

statistical models of appearance. In D. Vernon, editor, *Proceedings of European Conference on Computer Vision*, volume 1842, pages 829–843, 2000.

[95] K. N. Walker, T. F. Cootes, and C. J. Taylor. Automatically building appearance models from image sequences using salient features. *Image and Vision Computing*, 20(5-6):435–440, 2002.

[96] F. Wang, B. C. Vemuri, and A. Rangarajan. Groupwise point pattern registration using a novel CDF-based Jensen-Shannon divergence. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1283–1288, 2006.

[97] M. Weber, W. Einhäuser, M. Welling, and P. Perona. Viewpoint-invariant learning and detection of human heads. In *Proceedings of International Conference on Automatic Face and Gesture Recognition*, pages 20–27, Grenoble, France, 2000.

[98] M. Weber, M. Welling, and P. Perona. Towards automatic discovery of object categories. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 101–108, 2000.

[99] M. Weber, M. Welling, and P. Perona. Unsupervised learning of models for recognition. In D. Vernon, editor, *Proceedings of European Conference on Computer Vision*, volume 1842, pages 18–32, 2000.

[100] T. Werner. A linear programming approach to max-sum problem: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(7):1165–1179, 2007.

[101] G. Wu, H. Jia, Q. Wang, and D. Shen. Groupwise registration with sharp mean. In T. Jiang, N. Navab, J. Pluim, and M. Viergever, editors, *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 6362, pages 570–577, 2010.

[102] C. Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3):359–369, 1998.

[103] P. Zhang, S. A. Adeshina, and T. F. Cootes. Automatic learning sparse correspondences for initialising groupwise registration. In T. Jiang, N. Navab, J. Pluim, and M. Viergever, editors, *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 6362, pages 635–642, 2010.

[104] P. Zhang and T. F. Cootes. Automatic part selection for groupwise registration. In G. Székely and H. Hahn, editors, *Proceedings of International Conference on Information Processing in Medical Imaging*, volume 6801, pages 636–647, 2011.

[105] L. Zhu, Y. Chen, and A. Yuille. Unsupervised learning of a probabilistic grammar for object detection and parsing. In B. Schölkopf, J. Platt, and T. Hofmann, editors, *Advances in Neural Information Processing Systems 19*, pages 827–834. MIT Press,

2007.

[106] L. Zhu, Y. Chen, and A. Yuille. Unsupervised learning of probabilistic grammar-markov models for object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):114–128, 2009.

[107] L. Zhu, Y. Chen, and A. Yuille. Learning a hierarchical deformable template for rapid deformable object parsing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1029–1043, 2010.

[108] L. Zhu, C. Lin, H. Huang, Y. Chen, and A. Yuille. Unsupervised structure learning: Hierarchical recursive composition, suspicious coincidence and competitive exclusion. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Proceedings of European Conference on Computer Vision*, volume 5303, pages 759–773, 2008.

[109] L. Zhu and A. Yuille. A hierarchical compositional system for rapid object detection. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 1633–1640. MIT Press, 2006.