

Volumetric Image Analysis: Optical Flow, Registration and Segmentation

By
Ruwan Bandara Tennakoon

Submitted in total fulfilment of the requirements of the degree of
Doctor of Philosophy



Faculty of Science, Engineering and Technology
Swinburne University of Technology
Melbourne, Australia
2015

Abstract

Three-dimensional (3D) image acquisition technology has experienced intense development during the last few decades. Use of these images is becoming increasingly popular in the field of medicine for tasks including disease diagnosis, progress monitoring and treatment planning. An image analysis task may involve multiple images that may be acquired using the same or different sensors, obtained at different times or from different viewpoints. In order to make the most informative judgments, medical practitioners often need to align these images and analyse those in one coordinate reference. The task of finding a one-to-one mapping between the coordinates of one image to the other such that the corresponding structures overlap is called image registration and provides a basis for most image analysis tasks. The task of registering medical images can be challenging, especially for three-dimensional medical images, due to large data volume, low resolution of images compared to the structures of interest, complex motions, and image artefacts. Due to the complex nature of the motions involved, medical image registration is also often approached as a non-rigid registration problem.

While the problem of non-rigid registration has been studied extensively in the recent past, due to the constant improvements in acquisition technology and ever-increasing applications in medicine, there is a need to find better solutions in terms of both robustness and computational efficiency. This research involves a study of the fundamental problems of motion estimation and automatic registration of volumetric medical images to find how to improve the performance of registration in terms of both accuracy and efficiency.

Firstly, motion estimation using optical flow is considered, which requires some smoothing due to its ill-posed nature. Here, the relationship between the estimation accuracy and the required amount of smoothness for a general solution from a robust statistics perspective is theorized. Surprisingly small amount of local smoothing is required to satisfy both the necessary and sufficient conditions for accurate optic flow estimation. The theoretical predictions were tested by observing the accuracy of motion estimation in dynamic lung CT image (both synthetic and actual images) sequences using a robust 3D optic flow

method with quantized local smoothing. In line with theoretical predictions, the results showed that only a very small amount of local smoothing is required to achieve high accuracy and its proper implementation has a profound effect on the preservation of local information.

Next, a fast and accurate non-rigid registration method for intra-modality volumetric images is proposed. Non-rigid image registration techniques using intensity-based similarity measures have high computational complexities, particularly for volumetric images. The proposed method exploits the information provided by an order statistics-based segmentation method, to find the important regions for registration and use an appropriate sampling scheme to target those areas and reduce computation time. The experiments on registration of end-inhale to end-exhale lung CT scan pairs, with expert annotated landmarks, show that the new method is both faster and more accurate than the state-of-the-art sampling-based techniques, particularly for registration of images with large deformations.

Next, a new algorithm for modelling visual data using robust thin plate splines (TPS) is proposed. Non-rigid registration algorithms use robust TPS-based data modelling techniques to model transformation in feature-based image registrations, smoothing of motion fields and interpolating intensity field. However, these tend to be inaccurate, especially when the data contain outliers. The proposed fitting algorithm uses a variant of the least k -th order statistics fitting approach with a novel iterative method for performing the inlier/outlier segmentation. The proposed method is tested on synthetic and real data and is shown to be effective, even when numerous structured outliers (pseudo-outliers) are present.

Finally, a new method that improves the model fitting accuracy and efficiency using higher than minimal subset sampling is proposed. Identifying local structures or model-fitting is a fundamental task in many applications, such as motion estimation, segmentation, visual tracking and medical image analysis. The method uses a direct approach similar to Newton's method that starts from a random parameter set and moves towards a dense cluster using the local information. The extended experiments conducted on line/plane fitting and motion segmentation showed that the proposed method is both efficient and accurate compared to the state-of-the-art techniques.

Acknowledgements

Firstly, I would like to acknowledge with particular gratitude my supervisors, Prof Alireza Bab-Hadiashar and Dr Zhenwei Cao, for following my research closely and providing me with valuable guidance and support. I am also grateful for the painstaking efforts made by them to improve the writing of the material presented in my thesis. I would like to thank Prof. Alireza again for introducing and motivating me to explore the field of medical image registration.

I am indebted to a number of other people, in particular, Prof. Marleen de Bruijne and Prof. David Suter who collaborated with me on various research papers and provided much needed support. Dr. Alex McKnight for his efforts in proofreading the final version of this thesis for grammar and style in accordance with the Australian standards for editing practice.

I am grateful to Swinburne University of Technology for awarding me the SUPRA scholarship, which provided me with vital financial support throughout the three and a half years of my study. Swinburne also provided me with financial support to attend international conferences and a comfortable and helpful working environment.

I thank the authors of numerous software packages, codes and datasets used in the testing and comparisons reported in this thesis for making them available and providing me with technical support when needed. I would also like to mention the constructive feedback received from reviewers of journals and conference papers that enabled me to improve my work.

Finally, I would like to thank my family for their loving patience during the long period it has taken me to conduct the research and write up this thesis.

Declaration

I declare that this thesis contains no material that has been accepted for the award of any other degree or diploma and to the best of my knowledge contains no material previously published or written by another person except where due reference is made in the text of this thesis.

The work is based on the joint research and publications; the relative contributions of the respective authors are disclosed.

Ruwan Bandara Tennakoon, 2014

Publications Arising from this Thesis

The following publications are based on the work reported in this thesis:

Refereed journal papers:

1. R. B. Tennakoon, A. Bab-Hadiashar, Z. Cao, and M. de Bruijne, "Nonrigid Registration of Volumetric Images Using Ranked Order Statistics," *IEEE Transactions on Medical Imaging*, vol. 33, pp. 422-432, 2014.
2. A. Bab-Hadiashar, R. B. Tennakoon, and M. de Bruijne, "Quantification of Smoothing Requirement for 3D Optic Flow Calculation of Volumetric Images," *IEEE Transactions on Image Processing*, vol. 22, pp. 2128-2137, 2013.

Refereed conference papers:

1. R. B. Tennakoon, A. Bab-Hadiashar, M. de Bruijne, and Z. Cao, "Efficient non-rigid registration using ranked order statistics," in *2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI)*, 2013, pp. 496-499.
2. R. B. Tennakoon, A. Bab-Hadiashar, D. Suter, and Z. Cao, "Robust Data Modelling Using Thin Plate Splines," in *2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 2013, pp. 1-8.

Book chapter:

1. R. B. Tennakoon, A. Bab-Hadiashar and Z. Cao, "Nonlinear approaches in three dimensional medical image registration," in *Nonlinear Approaches in Engineering Applications volume 3*, L. Dai and R. N. Jazar, eds., Springer New York, In press.

Manuscripts under preparation:

1. R. B. Tennakoon, A. Bab-Hadiashar, Z. Cao and D. Suter, "Robust Model Fitting Using Higher Than Minimal Subset Sampling," To be submitted.

Contents

Publications Arising from this Thesis	vi
1 Introduction	1
2 Non-rigid Image Registration	9
2.1 Introduction	9
2.2 Similarity Measures	13
2.2.1 Voxel information-based similarity measures	14
2.2.2 Image feature based similarity measures	20
2.3 Transformation Model	20
2.3.1 Transformations derived from physical models	23
2.3.2 Transformations derived from interpolation theory	25
2.3.3 Constraints on transformation	30
2.4 Optimization	31
2.4.1 Continuous optimization	32
2.4.2 Discrete optimization	33
2.4.3 Hierarchical methods	35
2.5 Summary	36
3 Quantification of Smoothing Requirement for 3D Optic Flow Calculation	37
3.1 Introduction	37
3.1.1 3D Optic flow estimation	40
3.2 How Much Smoothing is “Just Enough”?	41
3.3 Estimation of 3D Optic Flow	44
3.4 Experimental Results	46
3.4.1 Effect of smoothing on accuracy	47
3.4.2 Optic flow estimation using real 4D CT images	48
3.4.3 Side effects of smoothing increases	52
3.4.4 Detection of local variations in flow fields	54
3.5 Conclusion	57
4 Image Registration Using Ranked Order Statistics	59
4.1 Introduction	59
4.2 Previous Research	61
4.3 The Proposed Method	62
4.3.1 How to sample efficiently?	62
4.3.2 Convergence criterion	66
4.4 Analysis of the Proposed Algorithms	67
4.4.1 Experimental set-up	67

4.4.2	Implementation	68
4.4.3	Analysis with synthetic images	70
4.4.4	Performance variation with noise	70
4.4.5	Experimental results	72
4.5	Discussion	79
4.6	Conclusion	83
5	Robust Data Modelling Using Thin Plate Splines	84
5.1	Introduction	84
5.2	Robust Spline Fitting	87
5.3	Experimental Results	91
5.3.1	Synthetic data simulations	92
5.3.2	Real data experiments	96
5.4	Conclusion	98
6	Robust Model Fitting Using Higher Than Minimal Subset Sampling	101
6.1	Introduction	101
6.2	Higher Than Minimal Sampling	103
6.2.1	Previous work	103
6.2.2	Evidence to support the use of higher than minimal subset sampling	104
6.3	Proposed Method	108
6.3.1	Cost function	108
6.3.2	Stopping criterion	109
6.3.3	Why does the proposed method converge?	111
6.4	Experimental Analysis	113
6.4.1	2D line fitting	114
6.4.2	3D plane fitting	115
6.4.3	Two-view motion segmentation	117
6.4.4	Multi-homography detection	119
6.4.5	3D-motion segmentation of rigid bodies	121
6.5	Discussion	123
6.6	Conclusion	127
7	Conclusion and Future Work	128
	References	132

List of Figures

2.1	In this illustration the moving image needs to be both translated and rotated to align with the fixed image.	10
2.2	(a) CT (b) MRI (c) PET images of the same subject. Images are from the RIRE dataset (http://www.insight-journal.org/rire).	11
2.3	Flow of a typical intensity (voxel information) based registration algorithm (Klein & Staring 2011).	13
2.4	Joint histograms between the original image (top-left) and rotated versions of it. (top-right) no rotation, (bottom-left) rotated by 2 degrees (bottom-right) rotated by 15 degrees.	18
2.5	Inverse transformation vs. forward transformation. The figure shows that using inverse transform is computationally advantageous.	22
2.6	Original image and the image after an affine transform. The transformation of both the grid and the intensities are shown.	23
2.7	B-spline kernels.	27
2.8	Original image and the image after a B-spline transform. The transformation of both the grid and the intensities are shown together with the control point locations.	28
3.1	Finite sample bias of MSSE in the segmentation of a step edge discontinuity with height of either 4 or 6 times the scale of noise in the measured data with 30% inliers. It is important to note here that as shown in Hoseinnezhad et al. (2010), for $\mu > 5\sigma$ the amount of finite sample bias is not affected by either the μ or the inlier ratio.	44
3.2	Sample of axial view of the sinusoidal image (top) and its flow field (bottom). The white arrow identifies the stationary column area.	47
3.3	Change in average error when the smoothing window size is varied, $\sigma = 2.0$	50
3.4	Average registration errors of our method are shown for each case. Corresponding average errors from Castillo et al. (2010), Castillo, Castillo, Guerra, Johnson, McPhail, Garg & Guerrero (2009) are also shown for comparison.	50
3.5	Effect of varying the size of smoothing window on average landmark errors in images of Popi model dataset (Vandemeulebroucke et al. 2011).	52
3.6	The above slices are chosen to depict the position of different landmark points used in our calculations (P1 to P9). Where possible, the fissure is also highlighted in these images.	53
3.7	Variation in estimation error with smoothing window size for landmark points near a fissure.	54

3.8	Variation in estimation error with smoothing window size for landmark points not near a fissure.	54
3.9	Top image is an axial view of the sinusoidal image where the cubic region with two distinct motions is marked by a white rectangle. Bottom image plots the associated voxels in the derivative space (red dots and green asterisks are used for moving voxels while black crosses represent the outliers). The points are segmented by fitting planes using MSSE. Bottom picture is rotated to show how the points are aligned with respect to the affine plane. Since a plane in the derivative space represents constant motion, the points with constant underlying motion (red dots) fit a plane better than the points with affine underlying motion (green asterisks). . . .	56
3.10	Top image is an axial view of a 4D CT image where the two cubic areas with either one (the top rectangle placed entirely on one lobe) or two (the lower rectangle straddles two lobes separated by the fissure - shown by a white line) distinct motions are marked by white rectangles. Middle and bottom images plot the associated voxels in the derivative space (red dots and green asterisks are used for moving voxels while black crosses represent the outliers). The points in both areas are segmented by fitting planes using MSSE. Middle and bottom pictures are rotated to show how the points are aligned with respect to every plane. The bottom picture clearly shows that the difference between the motions of two lobes on different sides of a fissure is detectable using a local approach.	58
4.1	The joint histogram of the fixed image ($I_f(x)$) and the transformed moving image ($I_m(T(x))$) for case 5 of DIR-dataset at the 10^{th} iteration. . . .	63
4.2	Classification of different voxels based on their absolute residuals using the MSSE constraint. The arrows show the boundaries between groups in a sample data set.	64
4.3	The results of the simulation demonstrate the effectiveness of the proposed algorithm in identifying the misregistered regions (a) The predefined motion field together with the surface rendering of the lung. (b) Plot of all voxels classified as the second group in the simulation.	71
4.4	Performance variation of registration error of the proposed algorithms versus the variance of manually added normal noise. The RS method is also included for the sake of comparison.	72
4.5	(a) Rate of the mean landmark error reduction over time for different registration methods for Case 1. (b)-(d) Rate of decrease of mean landmark error vs number of required iterations for different methods in cases where registration involves significant deformations.	78
4.6	Axial, sagittal and coronal slices of the intensity difference image between fixed and moving images (DIR-Case 5) before registration are shown in (a),(b),(c), respectively. (d),(e) and (f) show the axial, sagittal and coronal slices of intensity differences between fixed and transformed moving images after RIS registration.	79
4.7	Mean error of each iteration for algorithms RIS, RRS and RS. The maximum number of iterations for RS has been increased to 100 and 400 in the two registration levels respectively. The image used is DIR-Case 5. . .	80

4.8	The joint histogram of the fixed image ($I_f(x)$) and the transformed moving image ($I_m(T(x))$) for Case 5 of DIR-dataset at the 10^{th} iteration (a) contains all the voxels. (b) contains only the voxels identified as the middle group by the MSSE.	81
5.1	Fitting a TPS model to a dataset with 70% inliers. (a) When initial number of inliers is set to 35% and FLKOS and MSSE are directly applied. (b) When initial number of inliers is set to 35% and a spline is fitted using the proposed iterative method.	90
5.2	The variation of the mean error of various spline estimation methods for different percentages of gross outliers. The data in this simulation is created using (5.3.2) and gross outliers are generated using a zero mean normal distribution with standard deviation of 5.	93
5.3	Resulting TPS fit for the function in (5.3.2) with (a) 30% (b) 40% outliers from distribution $N(5, 0.1^2)$. The number of knots for L1-Norm method was set to 25.	94
5.4	The estimated splines for function (5.3.1) with 30% of outliers using different methods. Numbers of knots for L1-Norm TPS are chosen to be 10 in figure (a), 50 in figure (b) and 100 in figure (c).	95
5.5	The estimated splines for function (5.3.2) where the L1-Norm method is implemented with 10 knots, which was shown to be the optimum number for the first function (5.3.1). This figure shows the L1-Norm method is fairly sensitive to the number of knots chosen.	96
5.6	Estimated splines for function (5.3.2) with: 25% pseudo-outliers plus 5% gross outliers (a) and 10% pseudo-outliers plus 5% gross-outliers (b). To make a meaningful comparison, densities of data points along the x axis for pseudo-outlier and main structure are equalised.	97
5.7	Plot of the function (5.3.4) with no noise and all added pseudo-outliers.	98
5.8	The variation of the mean error of various spline estimation methods for different percentages of pseudo-outliers. The data in this simulation is created using (5.3.4) and outliers are generated using a normal distribution with mean of 15 and standard deviation of 2.	98
5.9	Estimated splines for function (5.3.4) with only 3% localised pseudo-outliers using (a) L1-Norm method and (b) proposed method.	99
5.10	The results of the real data experiment. The measured 3D depth points (point cloud) are shown by red dots and are rotated for better visualisation. (a) The picture of the object used in the range data modelling experiment. (b) Estimated splines for the L1-Norm method. (c) Estimated splines for the proposed method.	100
6.1	Minimal subset sampling vs. higher than minimal subset sampling. The data do not contain outliers (Pham et al. 2012). The figure represents an extreme case where the span of the data samples is made deliberately small.	102
6.2	The variation in minimum estimation error with number of data points n for different sample sizes (h). The figure shows the mean results of 100 experiments for each combination.	105

6.3	The cumulative distribution function of estimation errors for 2D line (a), (b) and 3D plane (c), (d) fitting using different sample sizes (h). The data used for fundamental matrix estimation and the CDF of the median residuals are shown in (e), (f).	107
6.4	The intermediate steps of the proposed method in a simple 2D line fitting example for $h = 4$. The model represented by the current hyper-edge is plotted to make the steps clear. Note that the vertices of the starting point (Step 1) are not members of the structure and the algorithm does not move away from the underlying structure after it is found (Step 6).	110
6.5	2D Line fitting results.	116
6.6	3D plane fitting results.	118
6.7	(a) Variation of clustering accuracy with sampling time (b) Inliers identified by the proposed method (c) Outliers identified by the proposed method.	119
6.8	Two-view motion segmentation results for image sequences in dataset Schindler & Suter (2006). Line 1 shows the ground truth whereas lines 2 and 3 show the final HMSS samples selected and the clustering results of the proposed method. The outliers are marked in red.	120
6.9	Qualitative results of multi-homography estimation. line 1 is the ground truth where as lines 2 and 3 are segmentation results of the proposed method and RCM respectively. The outliers are marked in red.	122
6.10	Segmentation result for several sequence in the Hopkins 155 dataset where the proposed method was successful. The sample used to generate the best hypothesis is also shown.	124
6.11	Segmentation result for several sequences in the Hopkins 155 dataset where the proposed method was not successful. The sample used to generate the best hypothesis is also shown.	125
6.12	Two-view motion segmentation results for image sequences “ <i>poster-checkerboard</i> ” (line 1) and “ <i>poster-keyboard</i> ” (line 2) from Poling & Lerman (2014).	126

List of Tables

2.1	Commonly used radial basis function kernels where $r = x - x_k $	26
3.1	Estimation accuracy of calculating 3D optic flow using different sizes of smoothing windows	48
3.2	Average error of calculating 3D optic flow using different sizes of smoothing window for real 4D CT data	49
3.3	P-values and hypothesis test results (with confidence interval of 99%) of paired t-tests conducted to test the significance of improvements achieved by increasing the size of the smoothing window. $H = 1$ means that the differences are statistically significant.	51
4.1	Comparison of mean landmark error (Euclidean distance) for IS-dataset measured using expertly identified landmark points.	73
4.2	Registration performance for images in DIR-dataset measured using expertly identified landmark points. Each algorithm is repeated 10 times to account for the randomness.	75
4.3	Average running time for each algorithm and the Holms' corrected p-values of Wilcoxon rank-sum test that compare the final mean registration error of competing methods with the RIS method.	76
4.4	Holms' adjusted p-values for the Wilcoxon rank-sum test conducted between the landmark error of RIS and other competing methods.	76
4.5	The improvement of the registration time due to the use of convergence criterion and the results of the TOST (low p-values mean the results are statistically equal).	77
4.6	The measure of correlation between the proposed convergence criterion and the actual registration error measured using expert identified landmarks. A value in excess of 0.5 for Pearson's correlation indicates a strong linear correlation. Images from the DIR dataset are used.	78
4.7	Median of the landmark errors belonging to Group one and Groups two and three.	81
4.8	The total computation time and the time for sorting for each image in DIR-dataset.	83
5.1	Computation time of estimating a 2D spline using both the proposed and the L1-Norm methods. Size of the dataset in these simulations is 101×101	96
6.1	Multi-homography detection results. The time is given in milliseconds. .	120
6.2	Percentage clustering Error of 3D motion segmentation.	123

Chapter 1

Introduction

Three dimensional (3D) medical image acquisition technology has experienced intense development during the last few decades and the use of these images is becoming increasingly popular in the field of medicine for tasks including disease diagnosis, progress monitoring and treatment planning. A 3D image here refers to a collection of scalar intensity values, depicting internal organs or structures of a patient, sampled over a 3D spatial grid. It is important to note that the term 3D image is used in some instances to describe range images, where the depth information of a scene is sampled in a 2D grid, which is different to its use here. A variety of physical principles are used to generate the intensity information regarding objects of interest and based on these principles, images are categorized into different modalities including: Computer tomography (CT), Magnetic resonance (MRI), Positron emission tomography (PET), Single proton emission tomography (SPECT) and Ultra sound (US). Each of these modalities provides a different type of information. For example, a CT image provides better information on the hard (bone) structures, whereas an MRI image contains better information about soft tissues.

An image analysis task may involve more than one image and in order to make the most informed decisions, medical practitioners often need to study two or several images in the same coordinate system. Normally, the coordinates of a medical image are set out by the acquisition device. However, due to the positioning of the patient, internal organ motions, anatomical changes in the patient (e.g. sequential images taken at large intervals) and device characteristics, the object coordinate space may vary even for the same machine. The task associated with this spatial mapping of images from one coordinate system to another such that the corresponding structures overlap is called image

registration and a formal definition of registration is as follows (Maurer Jr. & Fitzpatrick 1993):

Definition 1.1. *Image Registration* is the determination of a one-to-one mapping between the coordinates in one space and those in another, such that points in the two spaces that correspond to the same anatomical point are mapped to each other.

In medical applications, registration is required to be performed on images captured using different techniques, acquired in different times, from different viewpoints or of different subjects (Zitova & Flusser 2003). The registration may also be performed on images with partial or complete overlap. 3D image registration or motion estimation does not involve some of the problems associated with conventional 2D image analysis such as occlusions, etc. However, it poses a new set of problems, such as increasing volume of data due to the continuous advancements in acquisition technologies (a typical CT image usually contains around $512 \times 512 \times 128$ voxels), relatively low resolution compared to the structures of interest, especially for low dose images (voxel spacing along the some axis around 1 – 2.5 mm), complex organ motions (for example in a thoracic CT image the motion of the lungs is affected by the respiratory cycle, the motion of the heart, and the fixed nature of the spine and rib cage (Tustison et al. 2011)) and the high level of image artifacts caused by limitations in the physical principles of acquisition, limitations in current technology and patient motion/anatomical changes (Hsieh 2009).

There are many applications of image registration in the medical imaging field, some of which are listed below. A detailed description of those applications is outside the scope of this exposition and but be found in Holden (2008), Oliveira & Tavares (2014), Rueckert & Aljabar (2010).

Disease Diagnosis Early detection is the most effective way to reduce mortality for many diseases. Computer-aided diagnosis (CAD) systems are developed to enhance the capabilities of medical practitioners by attracting attention to certain areas in images, providing better comparisons and enabling medical practitioners to measure tumor sizes effectively. These systems can speed up the diagnostic process, reduce errors or misses and improve quantitative evaluation. Image registration plays a major role in CAD by enabling the combination of information from multiple views, previous times, and multiple modalities (Giger et al. 2008). CAD is used to diagnose breast cancer (Wei et al.

2009), lung cancer (Guo et al. 2012, Shiraishi et al. 2006, Van Ginneken et al. 2001), Parkinson's (Illán et al. 2012), emphysema (Sluimer et al. 2006), brain tumors (Jaya & Thanushkodi 2011) and many other common diseases.

Progress monitoring Detecting and measuring changes in an anomalous region over time is vital to evaluate the success and to plan the future course of medication. However, due to orientation changes, organ motions and anatomical changes during successive image acquisitions, direct comparisons are not possible. Image registration helps to reduce the effects of these deformations in order to make a realistic comparison (Arzhaeva et al. 2010, Hua et al. 2013, Ko & Betke 2001, Liang et al. 2013, Timp et al. 2007, Yuanjie et al. 2009).

Image-guided surgery Image guidance provides valuable information about the target position and optimal probe manipulations in minimally invasive treatment procedures, such as neurosurgery (Archip et al. 2007, Huang et al. 2012, Mostayed et al. 2013), endoscopy (Burschka et al. 2004, Mirota et al. 2012), interventional cardiology (Linte et al. 2010, Xishi et al. 2009), minimally invasive lung surgery (Sadeghi Naini et al. 2010, Uneri et al. 2013) and robotic microsurgery (Gerber et al. 2013). Here registration plays an essential role in combining and augmenting the high quality pre-operative images (CT, MRI, PET) with the intra-operative images of the patient on the operating table (X-ray, US, interventional cone beam CT). A comprehensive review of registration techniques that are applied specifically for minimal invasive therapy can be found in Rui et al. (2013).

Image-guided radiotherapy In radiotherapy, generous safety margins have to be set around the target due to organ movements and target localization errors. These margins lead to a compromise in the dose delivered to the tumor. As a result, accurately targeted localization (defining the tumor volume) and alignment is very important in order to maximize the radiation dose to the target while minimizing the effects on the surrounding organs. Image-guided radiotherapy techniques use on-line or off-line image registration. Off-line registration of images from different modalities (CT/MRI, CT/PET) can improve target localization, whereas organ motion can be determined by on-line registration of repeatedly acquired images. This information is then used to adapt the treatment (patient positioning, directing radiation beam) (Robertson et al. 2013, van

Herk 2007, Worm et al. 2010).

Atlas building & comparison Image atlases are used to compare individual images to an average image constructed by combining a large number of images. Image registration is used in both atlas building steps as well as in the comparison step (D’Agostino et al. 2004, Ehrhardt et al. 2011, Zhang et al. 2013).

Generally, a registration task is carried out by first establishing a goodness measure that quantifies the alignment between images (also called *similarity measure*) and a *transformation model* that can capture the deformation between the relevant images with adequate accuracy. Next, this information is used to establish an objective function, reflecting both the similarity and prior knowledge of the model, which is maximized (or minimized) using a suitable *optimization method* to find the parameters of the transformation model. Finally, the second image is warped to match the reference image using the estimated transformation model.

The spatial transformation needed to align one image to another depends on the type of motions involved between the images and can be modelled using rigid, affine or non-rigid transformations. Rigid transformations have limited degrees of freedom that only allow rotation and translation. They are used in registering objects that are rigid in nature, such as teeth and skulls. Affine transformations, on the other hand, allow for scaling and shear in addition to translation and rotation and are commonly used as a pre-registration step for non-rigid registration algorithms. The limited degrees of freedom allowed by these simple transformations are not adequate to model the motions involved in medical images and most applications use more sophisticated models called non-rigid models (Tustison et al. 2011). While the additional degrees of freedom in a non-rigid model helps capture complex motions, it also increases susceptibility to image/motion artifacts and computational complexity. To mitigate these effects, non-rigid models typically use some sort of regularization (based on prior knowledge of the motions and anatomical structures involved) that limit the solution domain, making the overall model robust. A given transformation model can also be viewed as either a global or a local transformation. A global transformation affects the entire image while a local transformation affects only a sub-region of the image domain.

Similarity measures used in typical registration algorithms either use only a specific part of the information extracted from the image called features (e.g. edges, anatomical

landmarks) or they use the voxel information directly. The features used can be either intrinsic or extrinsic. Extrinsic methods use artificial markers attached to the patient prior to being imaged, whereas intrinsic methods rely only on readily available information in the image. The intrinsic features can be either picked by a human expert or automatically computed using a technique such as SIFT (Lowe 1999). However, this feature extraction presents an additional step which can introduce errors to the registration algorithm and the estimates become inaccurate further away from the extracted feature points. For these reasons voxel information based registration methods are becoming more popular (Sotiras et al. 2013). The complexity of deriving a voxel information based similarity measure may depend on the type of images involved. If the images are from the same modality, simple measures such as sum of squared intensity differences can be used, but when the images are from different modalities, more complex measures are needed. Accordingly, information theory-based similarity measures have attracted attention for multi-modal image registration (Pluim et al. 2003).

The similarity measure and the transformation model established are then used to build an objective function, which in most non-rigid registration problems is highly non-linear. Hence, iterative optimization techniques have to be used. These optimization techniques can be either continuous or discrete depending on the nature of the objective function. The complexity of the optimisation task in a medical image registration problem is governed by both the volume of information and the complexity of the deformation model.

While the problem of non-rigid registration has been studied extensively in the recent past, due to the constant improvements in acquisition technology and ever increasing applications in medicine, there is the need to find better solutions in terms of both robustness and computational efficiency. This research involves studying the fundamental problems of automatic registration and motion estimation of volumetric medical images and finding how to improve the performance of registration in terms of both accuracy and efficiency. The main contributions of this research are as follows:

- Establishment of a theoretical relationship between the estimation accuracy and the amount of smoothness required for a general solution of the motion estimation problem from a robust statistics perspective.
- Development of a robust registration method for volumetric CT images in cases where the images are affected by significant organ motion and artifacts, that also

significantly improves computational efficiency.

- Exploration of the use of statistical methods to improve robustness in modeling data using splines, which are commonly used in non-rigid registration algorithms for feature matching, smoothing of motion fields and intensity interpolation.
- Development of a new method that improves parametric model fitting accuracy and efficiency by using higher than minimal subsets to generate hypotheses.

The thesis is organized as follows. Chapter 2 provides an overview of non-rigid registration techniques used for medical image analysis applications. Here, the image registration problem is segmented into three key sections (transformation model, objective function, optimization) and the techniques used in each of those sections are analysed in detail.

Optical flow is a technique that is used to approximate the motions between mono-modal images and is capable of recovering fairly local motions. Due to the well-known aperture problem, optical flow is ill-posed and requires some smoothing. Chapter 3 examines the relationship between the estimation accuracy and amount of smoothness required for a general solution from a robust statistics perspective. We show that a surprisingly small amount of local smoothing is required to satisfy both the necessary and sufficient conditions for accurate optic flow estimation. This notion is called “just enough” smoothing and its implementation has a profound effect on the preservation of local information in processing 3D dynamic scans. To demonstrate the effect of “just enough” smoothing, a robust 3D optic flow method with quantized local smoothing is presented and the effect of local smoothing on the accuracy of motion estimation in dynamic lung CT images examined using both synthetic and real image sequences with ground truth.

Non-rigid image registration techniques using intensity-based similarity measures are widely used in medical imaging applications. Due to the high computational complexities of these techniques, particularly for volumetric images, finding appropriate registration methods to both reduce the computation burden and increase the registration accuracy has become an intensive area of research. In Chapter 4, a fast and accurate non-rigid registration method for intra-modality volumetric images is proposed. This approach exploits the information provided by an order statistics-based segmentation method, to find the important regions for registration and use an appropriate sampling scheme to target those

areas and reduce the registration computation time. A unique advantage of the proposed method is its ability to identify the point of diminishing returns and stop the registration process. Experiments on registration of end-inhale to end-exhale lung CT scan pairs, with expert annotated landmarks, show that the new method is both faster and more accurate than the state-of-the-art sampling based techniques, particularly for registration of images with large deformations.

Using splines to model spatio-temporal data is one of the most common methods of data fitting used in a variety of computer vision applications. Despite its ubiquitous applications, particularly for volumetric image registration and interpolation, the existing estimation methods are still sensitive to the existence of noise and outliers. In Chapter 5 a method of robust data modelling using thin plate splines, based upon the well-known least K-th order statistical model fitting, is proposed and compared with the best available robust spline fitting techniques. Our experiments show that existing methods are not suitable for typical computer vision applications where outliers are structured (pseudo-outliers), while the proposed method performs well even when there are numerous pseudo-outliers.

Identifying the underlying model in a set of data contaminated with both noise and outliers is a fundamental task in computer vision and the cost function associated with such tasks is often highly complex. Commonly, an approximate solution is obtained by evaluating the cost function on discrete locations in the parameter space (hypothesis). However, To be successful at least one hypothesis has to be in the vicinity of the solution. Due to noise, the hypothesis generated using minimal subsets can still be far from the underlying model, even when the samples are from the said structure. Chapter 6 investigates the feasibility of using higher than minimal subset sampling for hypothesis generation. The synthetic experiments conducted showed that higher than minimal subsets sampling will increase the probability of generating a hypothesis closer to the true model, given that the subset is selected from inliers. However, the probability of selecting an all inlier sample decreases with the sample size, making direct extension of existing methods unfeasible. Hence, we propose an efficient method for robust model fitting that uses higher than minimal subsets. The method starts from an arbitrary hypothesis (which does not need to be in the vicinity of the solution) and moves towards a structure in data using the local information available at each iteration. The method also has the ability to identify when the algorithm has reached a hypothesis with adequate accuracy and stop

appropriately, thereby saving computational time. The experimental analysis carried out using synthetic and real data shows that the proposed method is both accurate and efficient compared to the state-of-the-art model fitting techniques.

Chapter 2

Non-rigid Image Registration

2.1 Introduction

Image registration is the task of finding a spatial transformation that aligns the objects in two or more images capturing the same or related scene. The registration can be performed on images from different times, captured using different sensors or from different viewpoints (Zitova & Flusser 2003). The registration may also be performed on images with partial or complete overlap. Registration is one of the most crucial problems of computer vision and has been studied for over three decades. The underlying task is very general and has a wide range of applications in the fields of medical imaging (Oliveira & Tavares 2014, Sotiras et al. 2013), computer vision (Peng & Chia-Hsiang 2013), remote sensing (Le Moigne et al. 2011, Richards 2012) and others. In this chapter we focus on the application of image registration in the field of medical imaging.

Although registration can involve two or more images, a typical medical image registration problem comprises two images:

Fixed image This image is used as the reference in the registration and is not changed during the process. It is also called the target or the reference image. In this document it is denoted by $I_F(x)$.

Moving Image This is the image to be transformed into the reference coordinate system and is also called the source, template or test image. In this document its denoted by $I_M(x)$.

Mathematically the problem of image registration can be expressed as finding a spatial

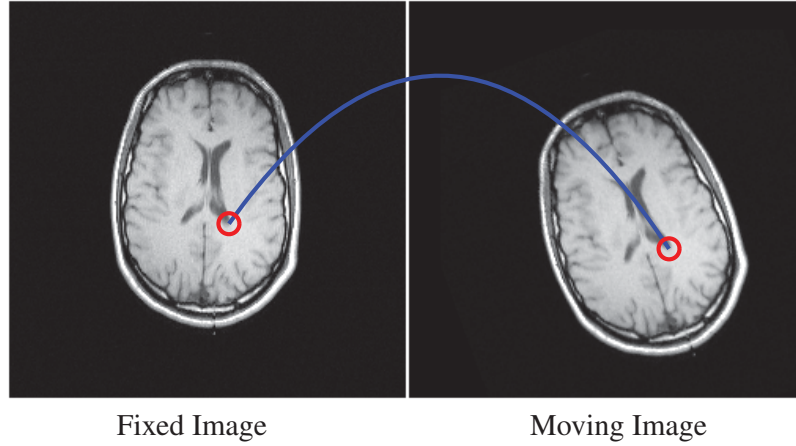


Figure 2.1: In this illustration the moving image needs to be both translated and rotated to align with the fixed image.

mapping

$$T : \Omega_F \rightarrow \Omega_M \quad (2.1.1)$$

such that the following correspondence holds:

$$I_F(x) \leftrightarrow I_M(x) \quad (2.1.2)$$

where Ω_F is the fixed image domain with dimensionality d and Ω_M the moving image domain. The image dimensionality d can either be two (2D) or three (3D). Hence, the transformation function can be a mapping from $2D \rightarrow 2D$, $3D \rightarrow 2D$ or $3D \rightarrow 3D$. Figure 2.1 shows an example of two images with overlap that can be registered in one coordinate system by translating and rotating one image (moving) to be aligned with the other (fixed) image.

Medical imaging consists of many modalities which differ on the physical principles used to capture the images. The common modalities are X-ray, computed tomography (CT), magnetic resonant imaging (MRI), positron emission tomography (PET), single proton emission tomography (SPECT) and ultrasound (US) imaging. These modalities can be categorized into two main groups according to the purpose of the images (Oliveira & Tavares 2014): *anatomical* images and *functional* images. CT, MRI, US and X-ray fall into the first category, where the purpose of the image is to study the anatomy, while functional MRI (fMRI), PET and SPECT fall in to the second category, where the purpose is to study the functions performed by the imaged organ. Figure 2.2 shows brain images

from the same subject captured using three different techniques (CT, MIT and PET). We can observe that each modality provides a specific type of information. For example, MRI captures the soft tissue in detail, whereas the bone structures are clearer in the CT image. A registration task may involve images from the same modality or different modalities. Depending on the number of modalities involved, the registration task can be classified as either *mono-modal* or *multi-modal*. In mono-modal registration, both images are from the same modality (e.g. time series CT image registration for change detection). In multi-modal registration, the images are from two different modalities (e.g. fusion of CT with MRI images for image-guided radiotherapy).

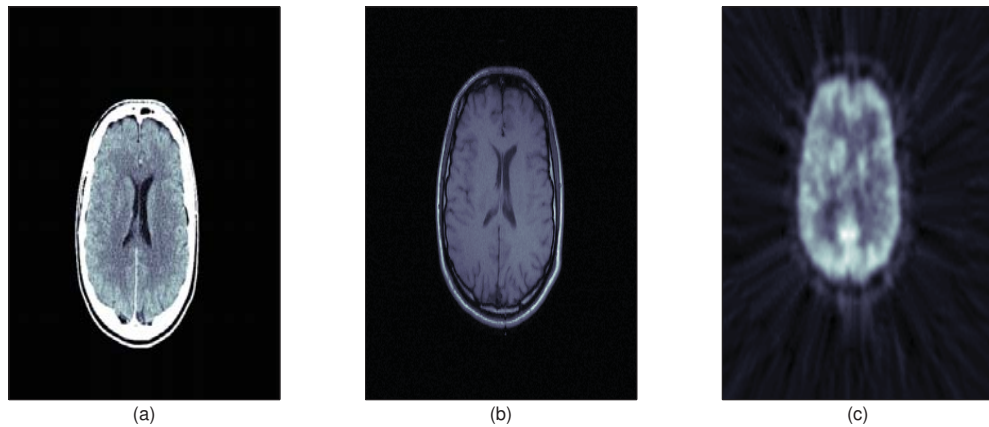


Figure 2.2: (a) CT (b) MRI (c) PET images of the same subject. Images are from the RIRE dataset (<http://www.insight-journal.org/rire>).

The solution to the registration problem generally consists of three main parts: transformation model, objective function and the optimization procedure (Oliveira & Tavares 2014, Sotiras et al. 2013). The techniques used in each module of a given algorithm depend on the specific requirements of the given registration problem.

Transformation model The transformation model defines the nature of the geometric transformation between the two images. The selection of the model depends on the expected deformations in the imaged object and this selection critically affects the accuracy of the final registration outcome.

Objective function The objective function measures the quality of the registration. Typically an objective function consists of two parts: The (dis-)similarity measure $\mathcal{M}$, which quantifies the alignment between images and the regularization term

$\mathcal{R}$, which constrains the solution domain. A typical objective function can be expressed as follows:

$$\mathcal{M}(I_F(x), I_M(T(x))) + \lambda \mathcal{R}(T) \quad (2.1.3)$$

where λ represents the trade-off between similarity and regularization. The choice of appropriate measure in the objective function depends on the modalities involved in the given registration task as well as the image quality, texture and the amount of deformation between the fixed and moving images.

Optimization algorithm The optimization tries to find the parameters of the transformation model that maximize (or minimize) the similarity (or dis-similarity) measure. In rare cases where landmark associations are available the optimal parameters of the transformation model can be directly calculated, but in general more sophisticated optimization algorithms are needed to find the optimal transformation model.

In addition to the three key parts described above, there are several support modules involved in registration, such as image interpolation, hierarchical strategies and pre-processing.

Interpolation The mapped point from one image domain to the other is unlikely to coincide exactly with its grid positions. Hence, interpolation has to be used to calculate the image values of the non-grid positions using grid values. The most common interpolation functions used in medical image registration are nearest neighbor, linear and B-spline interpolators.

Hierarchical strategies The complex and nonlinear objective functions in image registration lead to many local minima. In order to avoid being trapped by one of these local minima, most registration algorithms use hierarchical approaches. These techniques can be classified into three main categories (Lester & Arridge 1999): increasing *data* complexity, increasing *warp* complexity and increasing *model* complexity.

Image preprocessing In order to obtain the best performance, the images for registration may need to be preprocessed before feeding them to the registration algorithm. Smoothing, motion correction, segmentation of the relevant organ and normalization are some of the preprocessing steps that can be performed.

Although the registration algorithms vary widely, in general the flow of an image information based registration algorithm can be summarized by the flowchart shown in Figure 2.3.

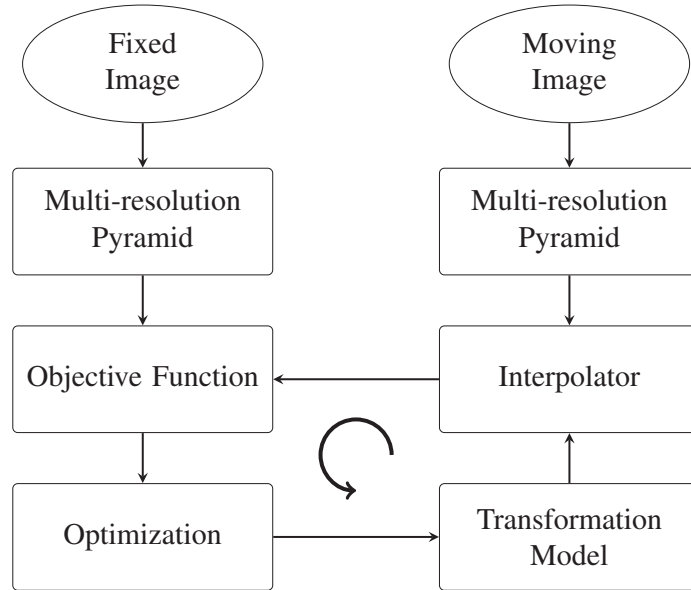


Figure 2.3: Flow of a typical intensity (voxel information) based registration algorithm (Klein & Staring 2011).

The deformations between images involved in medical imaging applications are often non-rigid and the relationships between measured values of the same structure in two different images can be nonlinear (especially for multi-modal registration). For these reasons, the resulting optimization becomes highly nonlinear. Several broad reviews of non-rigid registration techniques have appeared in the literature (Holden 2008, Oliveira & Tavares 2014, Rueckert & Aljabar 2010, Sotiras et al. 2013, Tustison et al. 2011). This chapter explores the non-rigid techniques used in different registration methods of common medical imaging applications.

The rest of this chapter is organized as follows. Section 2.2 explores the nonlinear similarity measures used in medical image registration, while Section 2.3 investigates the transformation models and Section 2.4 the methods used in nonlinear registration problem. Finally Section 2.5 concludes the chapter.

2.2 Similarity Measures

A similarity measure is part of the objective function which quantifies the alignment between the images being registered. Registration algorithms can be broadly classified

into two main groups, depending on the nature of the information used to calculate the similarity measure: *extrinsic* or *intrinsic*. Extrinsic methods use artificial markers attached to the patient, prior to being imaged. These markers are visible in all imaged modalities and will lead to efficient registration methods where parameters can be calculated explicitly. However, due to the invasiveness of attaching markers, restrictions on the number of markers that can be attached and the difficulty in capturing non-rigid motions, these methods are limited to rigid registration and pose estimation (Maintz & Viergever 1998). In contrast, Intrinsic methods rely only on information readily available in the image. Depending on the feature space used in the similarity measure, registration algorithms are broadly classified into three main categories (Sotiras et al. 2013): *voxel information*-based methods, *image feature-based* methods and *hybrid* methods. Feature-based methods use landmarks or features extracted from the images in the registration process, while voxel information based methods use the voxel information (intensity) directly. Hybrid methods are a combination of the above two techniques. These methods will be further analyzed in this chapter.

2.2.1 Voxel information-based similarity measures

In voxel information based registration, the similarity measure takes the intensity information of the two images as input and outputs a value that relates to how well the two images are aligned. Images for registration can be from the same modality (mono-modal registration) or from different modalities (multi-modal registration) and the type of similarity measure that best matches the given set of images has to be carefully chosen.

Mono-modal registration

Firstly, we consider the mono-modal registration case where the images are acquired using the same imaging technique. An example of this type of registration would be the registration of a time-series CT or MRI images in change detection applications. The simplest similarity measure for mono-modal registration is the sum of square difference (SSD), which is based on the assumption that corresponding anatomical structures in two images have similar voxel values, except for Gaussian noise. This measure can be mathematically represented as follows:

$$\mathcal{M}_{SSD}(\theta) = \frac{1}{N} \sum_{x \in \Omega_F \cap \Omega_M} (I_F(x) - I_M(T(x)))^2 \quad (2.2.1)$$

where N is the number of voxels in the overlapping region of these images. SSD would fail in cases when the images contain some voxels with large errors (outliers), as these high error values skew the objective function. Outliers can commonly occur in medical imaging due to the injection of contrasting agents or position changes of the patient in two time-lapse images (Hill et al. 2001). This problem can be mitigated by using the sum of absolute difference (SAD) similarity measure, which is more robust to outliers. However, it should be noted that unlike SSD, SAD is not generally differentiable, which complicates the optimization part of the registration process.

$$\mathcal{M}_{SAD}(\theta) = \frac{1}{N} \sum_{x \in \Omega_F \cap \Omega_M} |I_F(x) - I_M(T(x))| \quad (2.2.2)$$

The assumption that the voxel values are similar except for Gaussian noise is very restrictive and a more realistic assumption would be to hypothesize a linear relationship between intensities. For this assumption the optimal similarity measure is cross-correlation or normalized cross-correlation of image intensities, which can be represented as:

$$\mathcal{M}_{CC}(\theta) = \frac{\sum_x (I_F(x) - \mu_F)(I_M(T(x)) - \mu_M)}{\sum_x \sqrt{\sum_x (I_F(x) - \mu_F)^2 \sum_x (I_M(T(x)) - \mu_M)^2}}; x \in (\Omega_F \cap \Omega_M). \quad (2.2.3)$$

In the above measure, μ_F and μ_M are the mean intensities of the fixed and moving images, respectively. It is important to note here that, even though the relationship between the intensity values of the two images is assumed to be linear, the resulting cost functions will be nonlinear.

The assumption of linear relationships between corresponding voxels becomes unrealistic in some mono-modal medical image registration applications where the organ of interest is made up of soft tissue and undergoes elastic deformations. An example of this scenario would be the registration of lung CT images acquired at different states of the respiratory cycle. The density of tissue, which is measured by CT, may vary depending on the volume of the air inside the lung. Hence, the intensity of the same anatomical structure would change during the breathing cycle. In order to overcome this problem specifically in lung image registration, there have been methods developed that assume a nonlinear relationship between voxel values, using the conservation of mass principle. Gorbunova et al. (2012) assumed that the mass of parenchyma remains constant and the density of the lung tissue is inversely proportional to the amount of air. The change in

volume was calculated using the determinant of the Jacobian (Yin et al. 2009) and this was integrated into the SSD model to obtain the following cost function:

$$\mathcal{M}_{MP}(\theta) = \frac{1}{N} \sum_{x \in \Omega_F \cap \Omega_M} (I_F(x) - |J_T(x)| \cdot I_M(T(x)))^2 \quad (2.2.4)$$

where $|J_T(x)|$ is the determinant of the Jacobian of the transform model.

Castillo et al. (2012) and Castillo, Castillo, Zhang & Guerrero (2009) also used the principle of mass preservation but in an optical flow framework. They modeled the mass preservation assumption ($\int_{\Omega} I_F(x) dx = \int_{\Omega} I_M(x) dx$) using the following optical flow equation:

$$I_F(x) + \nabla I_M(x) \cdot v + I_M(x) \cdot \text{div}(v) = 0 \quad (2.2.5)$$

where v is the velocity field in the image. The above model is equivalent to the normal optical flow model ($I_F(x) + \nabla I_M(x) \cdot v = 0$) when the flow field divergence is zero or in other words the object is incompressible. The integrated form of Equation 2.2.5 (Corpetti et al. 2002) leads to the following compressible flow similarity measure:

$$\mathcal{M}_{CF} = \frac{1}{N} \sum_{x \in N(x)} (\ln(I_F(x)) - \ln(I_M(T(x))) + \text{div}(T(x) - x))^2. \quad (2.2.6)$$

It is important to note that most similarity measures such as SSD and SAD assume independence between intensities from voxel to voxel and do not incorporate any spatial information. This means that if the same permutation of voxel indices is applied to both images the resulting similarity measure will be unchanged. To account for the spatially varying intensity distortions, Myronenko & Xubo (2010) developed a new similarity measure named residual complexity. To arrive at the correct form of this measure, they first introduced an intensity correction field that represents the differences of intensities of the registered images. By analytically solving for this correction field and using adaptive regularization, they derived the following similarity measure:

$$\mathcal{M}_{RC} = \sum_{n=1}^N \log \left(\frac{(q_n^T r)^2}{\alpha} + 1 \right) \quad (2.2.7)$$

where $r = (I_F(x) - I_M(T(x)))$ and q_i is an eigenvector of Q where $Q \Lambda Q^T$ is the spectral

decomposition of $P^T P$. P is the regularization operator of the correlation field.

Multi-modal registration

Secondly, we will consider the similarity measures for multi-modal image registration where the voxel intensities of the images to be registered are generated through different acquisition methods (physical processes). In this type of application, the relationships between the voxel intensities of the same structure are generally complex and information theory measures such as Mutual Information (MI) proposed by Viola & Wells III (1997) and Collignon et al. (1995) have attracted much attention. These MI-based methods only assume a probabilistic relationship between the voxel values belonging to the same structure. It is important to note that, although these methods are specifically developed for multi-modal registration, they can also be directly applied to mono-modal applications.

Mutual information is a measure of how well one image explains the other image. If the images are well registered, the knowledge of one image will significantly reduce the uncertainty about the other image. This reduction of uncertainty can be quantified using conditional entropy, which measures how well the intensity of a voxel in one image predicts the intensity of the corresponding voxel in the other image. The resulting similarity measure can be written as:

$$\mathcal{M}_{MI} = H(I_M^T) - H(I_M^T | I_F) \quad (2.2.8)$$

where $H(I)$ is the Shannon entropy of the image intensities and $H(I_M^T | I_F)$ is the conditional entropy of image I_M^T given I_F . MI can also be represented in two alternative forms using the joint entropy and Kullback-Leibler distance as follows (Pluim et al. 2003):

$$\mathcal{M}_{MI} = H(I_F) + H(I_M^T) - H(I_F, I_M^T) \quad (2.2.9)$$

$$= \sum_{x \in I_F, y \in I_M^T} p(x, y) \log \frac{p(x, y)}{p(x) \cdot p(y)} \quad (2.2.10)$$

Figure 2.4 shows the joint histograms between a brain MRI image and a rotated version of itself. These plots show that when the images are perfectly registered the joint histogram is concentrated, whereas when the misalignment increases the joint histogram becomes dispersed. The mutual information value for each case is also shown in this figure and we can see that the more the images are aligned, the larger the mutual information

values.

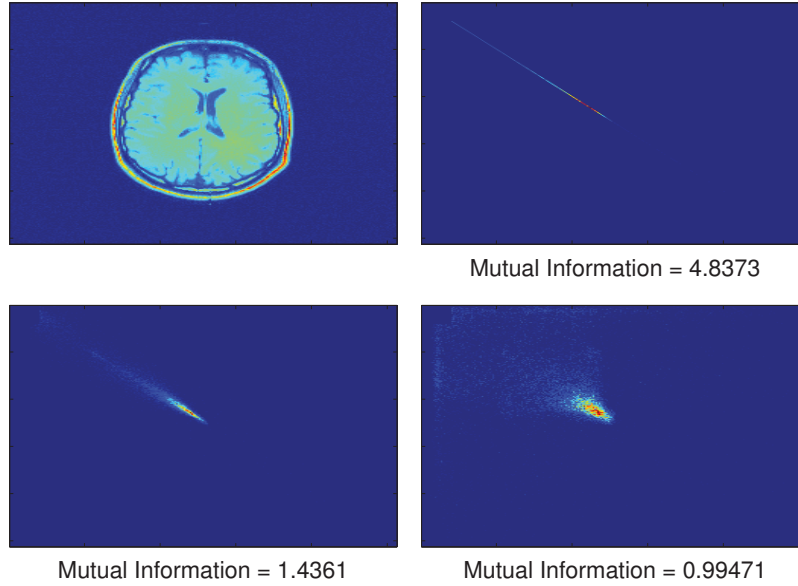


Figure 2.4: Joint histograms between the original image (top-left) and rotated versions of it. (top-right) no rotation, (bottom-left) rotated by 2 degrees (bottom-right) rotated by 15 degrees.

The above MI measure is sensitive to the overlap between the images i.e. it is not overlap invariant. To address this problem, Studholme et al. (1999) proposed the normalized mutual information (NMI) similarity measure given in Equation 2.2.11.

$$\mathcal{M}_{NMI} = \frac{H(I_F) + H(I_M^T)}{H(I_M, I_M^T)} \quad (2.2.11)$$

Cahill et al. (2008) further analyzed the overlap invariance of several information theory-based similarity measures including NMI and found that in some cases those methods would still fail. To make the NMI measure overlap invariant, they introduced a modified form of NMI.

Another drawback of mutual-information based methods is that they do not consider the relevant spatial information. As a result, MI-based similarity measures may face problems in the presence of local structures or spatially-varying intensity distortions. Various techniques have been proposed that integrate spatial information in the MI cost function. For instance, Pluim et al. (2000) proposed a new measure that combined mutual information with spatial information given by gradient magnitude and orientation. The new measure was constructed by simply multiplying the NMI cost function with a proposed

gradient measure. In contrast, Rueckert et al. (2000) incorporated the spatial information using the second order entropy and joint entropy in the calculation of NMI.

Another approach to include spatial information in the MI cost function is to use local estimation of the mutual information by progressively subdividing the images (Andronache et al. 2006, Likar & Pernu 2001). Studholme et al. (2006) combined these regional terms into one global cost function called Regional Mutual Information (RMI). RMI incorporates the regional information describing the global relationship between intensities as a third channel of information (in addition to the two intensity channels) in the MI calculation. Using additional channels in MI calculation is computationally expensive and Klein et al. (2008) proposed to reduce the computational complexity by using random sampling. Loeckx et al. (2010) also incorporated a third spatial dimension to create the conditional mutual information (cMI) similarity measure.

Local structural information can also be added to MI computation using graph-based α -MI techniques (Hero et al. 2002, Staring et al. 2009). Recently, Rivaz et al. (2014) and Rivaz & Collins (2012) used self-similarity, which measures the similarity of one image patch to another patch in the same image in a weighted graph based α -MI implementation for non-rigid image registration.

Calculation of mutual information involves some nontrivial and sensitive smoothing parameters for density approximations and the resulting cost function is highly non-convex and contains many local minima. In order to overcome these limitations, Haber & Modersitzki (2006) and Modersitzki (2009) proposed the use of Normalized Gradient Fields (NGF). The rationale behind this is that if the images are mis-registered, the angle between the two gradient fields at a given point will be non-zero. The resulting similarity measure can be expressed as:

$$\mathcal{M}_{NGF} = \frac{1}{2} \sum_{x \in \Omega} \|n_\varepsilon(I_F, x) \times n_\varepsilon(I_M, T(x))\| \quad (2.2.12)$$

where

$$n_\varepsilon(I, x) = \frac{\nabla I(x)}{\sqrt{\nabla I(x)' \nabla I(x) + \varepsilon^2}} \quad (2.2.13)$$

Ruhaak et al. (2013) have used this similarity measure in registration of full-body CT and PET images.

2.2.2 Image feature based similarity measures

The next class of image registration algorithms is based on matching image features. These methods first identify anatomically important points, curves or surface landmarks of the images. Once the correspondences are established between the landmarks, a similarity measure (SSD) that minimizes the Euclidean distance between points can be used to calculate the model parameters. A detailed review of feature-based registration methods was provided by Audette et al. (2000). The interest in the use of pure feature-based methods has gradually diminished, since the landmark or feature extraction processes introduce some errors that propagate through the algorithm. Moreover, the need for extrapolation/interpolation from a sparse set of landmarks to recover the dense deformation field means that the estimation of those fields away from the landmarks will be unreliable (Sotiras et al. 2013).

Intensity-based methods only account for intensity patterns and do not take into account the structure, while feature-based methods only use sparse landmarks. The hybrid algorithms combine the two in order to achieve more robust registration results (Rueckert & Aljabar 2010).

2.3 Transformation Model

The transformation model in the image registration context is a spatial mapping from one image domain to another. The transformation model that best matches a given registration application depends on the underlying deformations that are to be recovered. For example, if the object of interest is rigid in nature (skull, teeth), a transformation model with few parameters can be used to model the deformations involved. On the other hand, if the object of interest is made up of soft tissue (lung, heart) more complex models with high degrees of freedom are needed. Many transformation models are used in medical image processing and these can be either linear or nonlinear. The definition of a linear transformation function is as follows:

Definition 2.1. Let V and W be two vector spaces over the same scalar field $\mathcal{F}$. A function

$$T : V \rightarrow W \tag{2.3.1}$$

is said to be a linear transformation if the following two properties hold for any two points

$x, y \in V$ and scalar $\lambda \in \mathcal{F}$ (Rynne & Youngson 2008):

$$T(x + y) = T(x) + T(y) \quad (2.3.2)$$

$$T(\lambda x) = \lambda T(x). \quad (2.3.3)$$

In general, linear transformations map a line to a line or a point. In contrast, nonlinear transformations map lines to curves. In most cases of medical image registration the nature of the deformations is nonlinear due to the non-rigid nature of organs and muscles. Hence, transformation functions that are used are mostly nonlinear.

The objective of image registration is generally to map the moving image to the coordinate system of the fixed image, which is called the forward transformation. However, most algorithms due to the practical advantages calculate the inverse transformation (transformation from fixed image to moving image). The rationale behind this choice (Sotiras et al. 2013) is that if a forward transformation is calculated, in order to generate the transformed image each grid point in Ω_M needs to be transformed using the calculated transformation function followed by a scattered data interpolation. Due to the large number of points, the computational cost of the interpolation would be huge. On the other hand, if the inverse transformation is calculated, the values of the transformed image can be calculated by transforming each point in the new grid to Ω_M and picking the relevant value from the moving image (when the transformation leads to a non-grid position in the moving image, intensity interpolation can be done using neighboring voxels), which has far less computational workload. Figure 2.5 further elaborates this point.

Transformation models can be grouped into two main categories: *rigid* and *non-rigid*. Rigid transformation, which is the simplest type of transformation function used in medical imaging, can be used to match images that are of bone structures or enclosed by bone structures. Rigid transformation is a combination of a rotation and a translation, which in three dimensions can be represented by three rotation and three translation parameters. The rigid transformation falls within the broader category of affine transformations that can be expressed using the following equation:

$$x'_{[d \times 1]} = A_{[d \times d]} \cdot x_{[d \times 1]} + b_{[d \times 1]} \quad (2.3.4)$$

Affine transformation consists of a linear transformation ($A \cdot x$) representing rotation, scal-

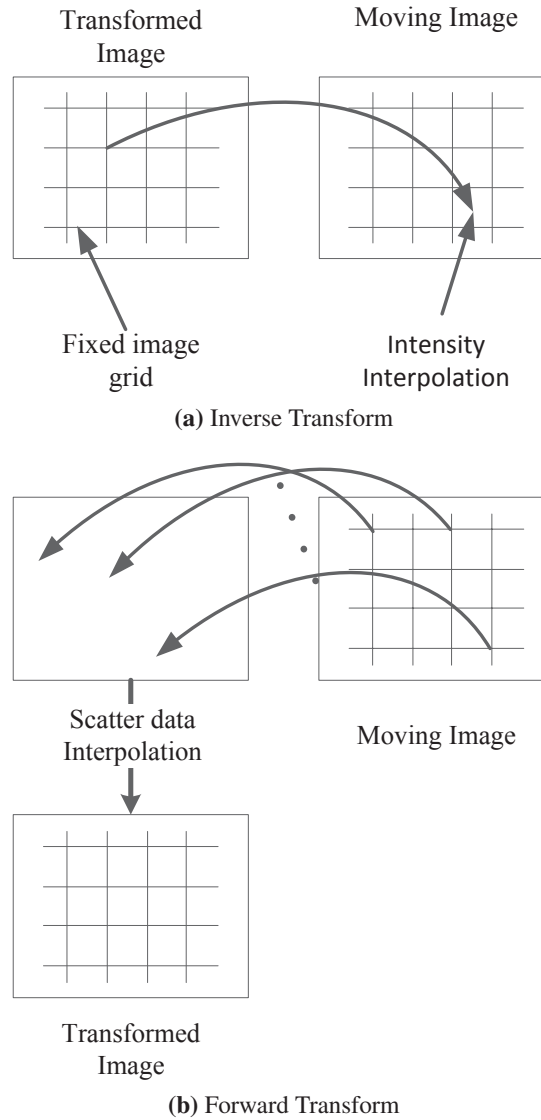


Figure 2.5: Inverse transformation vs. forward transformation. The figure shows that using inverse transform is computationally advantageous.

ing and shear and a translation vector (b). Except for the special case where the translation part is zero, the affine transformation in general is nonlinear, contrary to its appearance which has caused some authors to classify it as a linear transform (Glocker et al. 2011, Lester & Arridge 1999). Due to the scaling and shear components, the general affine transformation is also categorized as non-rigid. Affine registration is sometimes used in cases where images are distorted due to the physical characteristics of the acquisition system. It is also extensively used as a pre-registration technique in complex registration algorithms with large degrees of freedom.

The limited degrees of freedom allowed by affine or rigid transformation is not ad-

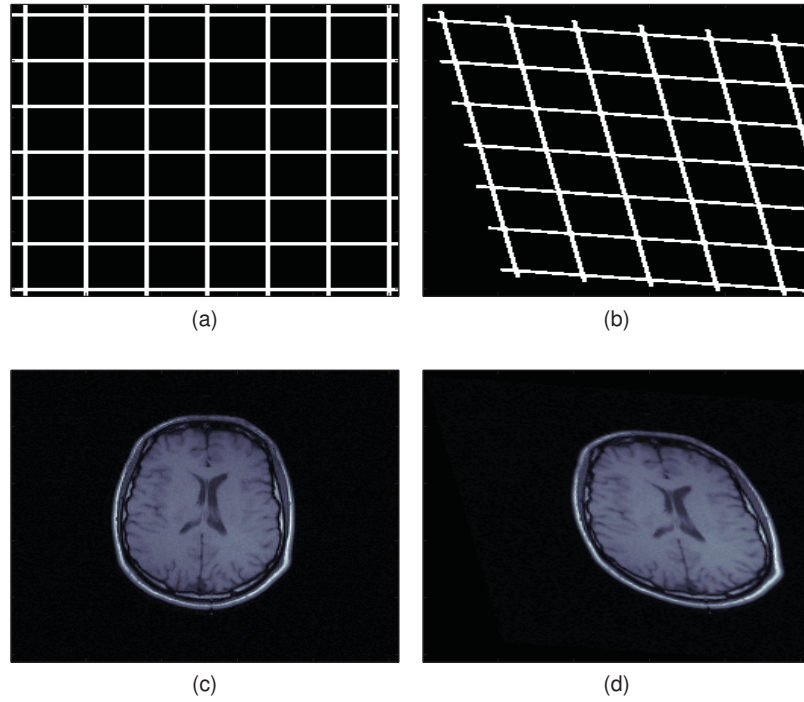


Figure 2.6: Original image and the image after an affine transform. The transformation of both the grid and the intensities are shown.

equate to represent the complex deformations involved in medical image registration. Hence, non-rigid registration models with much higher degrees of freedom have been introduced. Holden (2008) and Sotiras et al. (2013) categorized those non-rigid registration algorithms into three main groups: transformations derived from *physical models*, transformations derived from *interpolation theory* and *constraints on transformation*.

2.3.1 Transformations derived from physical models

These models use the characteristics of physical phenomena such as elastic material and viscous fluid flows to model the deformations in medical images.

Elastic models

Elastic models assume that the images behave similar to elastic bodies under deformation. Here, the external forces due to the mismatch in images (driving forces in registration are calculated using the similarity measure) are balanced with the internal forces that impose smoothness in the deformation field. The elastic deformations of an object

can be represented using Navier-Cauchy equation:

$$\mu \nabla^2 u(x) + (\lambda + \mu) \nabla(\nabla^T u(x)) + f(x) = 0 \quad (2.3.5)$$

where, $f(x)$ is the external force that drives the registration, which is the similarity measure, $u(x)$ is the deformation at point x , μ quantifies the stiffness of the material and λ is the Lamé's first coefficient.

This Navier-Cauchy partial differential equation leads to an optimization problem which can be solved in many ways, including variational, finite difference, finite element, basis function expansions and Fourier transform methods (Broit 1981, Christensen et al. 1994, Gee et al. 1994, Modersitzki 2004).

One of the drawbacks in using this model is that the resulting transformation is not inversely consistent. This can be corrected, either by adding a constraint that penalizes the inverse inconsistency (after calculating forward and backward transformations), or by coupling the forward and backward transformations in deriving the final transformation. A detailed description of this is provided later when constraint on transformation is explained.

In the linear elastic model explained above, the second order terms of the displacement field gradients are ignored. This means that the relationship between stress and strain is assumed to be linear, which makes this model incapable of handling large deformations. To overcome these issues, nonlinear elastic models are developed. These models were based on hyper-elastic material properties (Rabbitt et al. 1995) and St Venant-Kirchhoff elasticity energy (Pennec et al. 2005).

Fluid flow

In fluid flow models the images are assumed to behave as viscous fluids under deformation. Since these models do not assume small deformations, they can be used to model images that have undergone large non-rigid deformations.

The fluid flow transformation model can be represented by the following Navier-Stokes partial differential equation.

$$\mu \nabla^2 v(x) + (\lambda + \mu) \nabla(\nabla^T v(x)) + f(x, u) = 0 \quad (2.3.6)$$

This equation is very similar to the elastic model equation in the previous section. The only difference is that in the fluid flow model a velocity field is used instead of the deformation field. In the above equation, the first term imposes spatial smoothness of the velocity field while the second term allows for expansions and contractions. λ and μ are the viscosity coefficients.

Finding the solution to fluid flow models involves solving a large set of partial differential equations. The earliest implementations used successive over-relaxation, which is computationally expensive (Christensen et al. 1994). Christensen et al. (1996) used parallel computations to reduce the computational time. Some faster ways of solving differential equations such as scale-space convolution with a filter (impulse response of the regularization operator) (Bro-Nielsen & Gramkow 1996) have also been proposed.

2.3.2 Transformations derived from interpolation theory

In the transformation models that belong to this category the displacement field is modeled using functions derived from interpolation and approximation theories. The requirement to intercept sample values is relaxed in approximation, assuming errors in sample values (Holden 2008). In these methods firstly the basis function centre points (or control points) are selected, and then the parameters of each basis function are adjusted (or calculated) to fit the deformation field.

Radial basis functions

Radial basis functions are the most commonly-used interpolation theory-based transformation models in landmark-based image registration. The value of a radial basis function (RBF) depends only on the distance from its control point. Several such basis functions placed on the spatial domain of the image can be used to model the deformation field of the image effectively. The transformation function based on RBF can be defined as follows:

$$T(x) = x + \sum_{k=1}^N \theta_k R(\|x - x_k\|) \quad (2.3.7)$$

where θ_k are the parameters that control the shape of the deformation field, $x_k, k \in [1 \dots N]$ are the RBF control points and $R(\cdot)$ is the radial basis function kernel. One of the advantages of radial basis functions is the ability to place the control points in an

irregular grid, which enables modeling non-uniform deformations with fewer parameters than a grid-based control point placement scheme.

In landmark-based registration, the radial basis function is combined with a polynomial basis. The polynomial basis $\phi(x)$ is added to account for the more global deformations while the RBF part highlights the local changes. This transformation function can be expressed as follows:

$$T(x) = x + \sum_{k=1}^N \theta_k R(\|x - x_k\|) + \sum_{j=1}^m \theta_{(N+j)} \phi_j(x) \quad (2.3.8)$$

Once the problem is in this setting, if the landmarks and their correspondences are available, a closed form solution can be found as the RBFs are positive definite. Some of the radial basis functions commonly used in the literature are given in Table 2.1.

Table 2.1: Commonly used radial basis function kernels where $r = \|x - x_k\|$

Thin-plate spline	$r_i^2 \log r_i$
Multi-quadratic	$\sqrt{r_i^2 + d^2}$
Inverse multi-quadratic	$(r_i^2 + d^2)^{-1/2}$
Generalized multi-quadratic	$(r_i + c_1^2)^{c_2}$
Gaussian	$e^{-(r_i/c_1)^2}$

The radial basis function described above has a global support region, or in other words a given control point affects the entire deformation field. This property has adverse affects when there are outliers or the desired deformation field is local.

In medical image registration a given image may contain several local motions due to the presence of different anatomical structures. In order to recover these types of local motions RBFs that have compact spatial support are proposed. Examples of these functions are the Wendland Function and Wu's compact support RBF (Rohde et al. 2003).

Free-form deformation (FFD)

The main idea behind free-form deformation (FFD)-based image registration is to match two images by deforming one image by manipulating a grid of control points superimposed over that image. FFD is the most widely used transformation model in medical image registration. Typically, the control point grid is a rectangular grid with arbitrary resolution which is coarser than that of the image. The control points are moved in such

a way that a similarity measure is maximized and the deformation at each voxel is then calculated by interpolation using locally supported kernels such as B-splines.

A B-spline function is a compact support kernel which makes it suitable for modeling deformation fields with local motions. The B-spline kernel of order zero is a rectangle function and the higher order kernels are calculated by convolution of the zero order kernel with itself N times (N is the degree of the kernel). The first four B-spline kernels are shown in Figure 2.7.

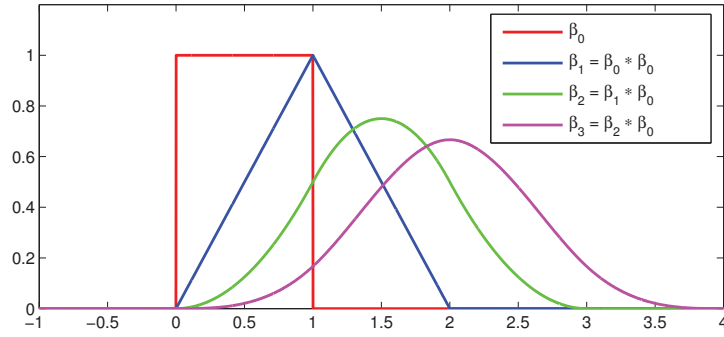


Figure 2.7: B-spline kernels.

These spline kernels are stitched together over the image domain in a rectangular grid in order to model the deformations of the image. Given the compact support of the B-spline kernel, the deformation of a particular point represented by a B-spline would only depend on several control points in the immediate neighborhood of that point. Using this property, for a 3D control point grid with spacing $[\delta_x, \delta_y, \delta_z]$ the multivariate extension of a cubic B-splines-based FFD transformation model can be written as follows (Rueckert et al. 1999):

$$T(x) = x + \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 B_{l,3}(u) B_{m,3}(v) B_{n,3}(w) d_{i+l, j+m, k+n} \quad (2.3.9)$$

where $i = \lfloor x/\delta_x \rfloor - 1$, $j = \lfloor y/\delta_y \rfloor - 1$, $k = \lfloor z/\delta_z \rfloor - 1$, $u = x/\delta_x - \lfloor x/\delta_x \rfloor$, $v = y/\delta_y - \lfloor y/\delta_y \rfloor$ and $w = z/\delta_z - \lfloor z/\delta_z \rfloor$.

The cubic B-spline functions in the above equation are:

$$B_{0,3}(u) = \frac{(1-u)^3}{6} \quad (2.3.10)$$

$$B_{1,3}(u) = \frac{3u^3 + 6u^2 + 4}{6} \quad (2.3.11)$$

$$B_{2,3}(u) = \frac{-3u^3 + 3u^2 + 3u + 1}{6} \quad (2.3.12)$$

$$B_{3,3}(u) = \frac{u^3}{6} \quad (2.3.13)$$

The B-spline kernels are also $\mathcal{C}^2$ smooth at the control points which provides continuity when the control points are moved. This enables the B-spline transformation model to generate smooth deformations. An example of a B-spline transformation is shown in Figure 2.8 where the control points are placed in a regular grid.

Rueckert et al. (1999) used the B-spline motion model after an affine pre-registration step for non-rigid registration of contrast-enhanced breast MRI images. In their work, NMI was used as the similarity measure. Kybic & Unser (2003) used the B-spline deformation modal coupled with a SSD similarity measure to register images from MRI, SPECT, CT and ultrasound modalities. Their method also included a soft landmark-based constraint to guide registration in difficult scenarios.

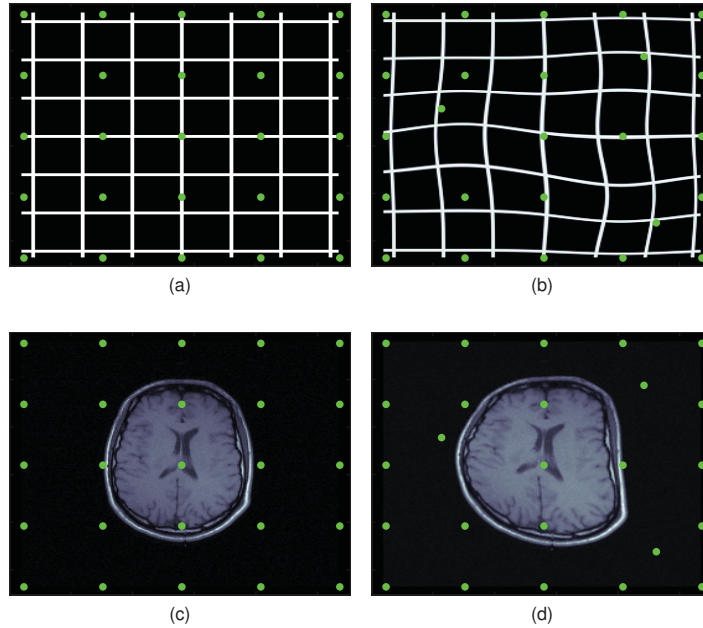


Figure 2.8: Original image and the image after a B-spline transform. The transformation of both the grid and the intensities are shown together with the control point locations.

The control node displacements act as parameters in a FFD. Hence, the details captured by the deformation field depend on the resolution of the control point grid. Large spacing enables the modeling of global deformations, whereas small spacing represents highly local deformations. The number of control points also directly affects the computational complexity of the resulting optimization problem. The optimal control point grid is decided based on the model's flexibility and computational complexity.

The B-spline-based FFD framework does not guarantee the preservation of topology and there are several methods that can be employed to rectify this issue. Noblet et al. (2005) enforced topology preservation by controlling the Jacobian of the transformation and Kim (2004) suggested a sufficient condition for local invariability. Chun & Fessler (2008) extended this and proposed a new penalty-based approach for topology preservation B-spline transformation.

Although in the basic implementation of B-spline-based FFD the control point grid is uniform, there has been some work that employs a non-uniform grid (Jacobson & Murphy 2011, Wang & Jiang 2007).

Wavelet and Fourier

Fourier series and wavelet functions are extensively used to represent signals. In Fourier representation the signal is characterized as a combination of sinusoidal signals which localize the signal in the frequency domain. On the other hand the wavelet representation localizes the signal in both frequency and spatial domains, making it ideal to represent localized deformations (Wu et al. 2000).

One of the major attractions of these techniques in modeling deformation fields is their ability to decompose into multi-resolution.

Local affine

In locally affine transformation models, the image domain is partitioned into independent blocks and for each block a rigid or affine transformation is calculated by maximizing a voxel-based similarity measure (Pitiot et al. 2003). An issue with this local transformation is that it may not be continuous at the block edges. In order to overcome this issue, regularization can be performed using either nonlinear interpolation or filtering. Another possible solution is to use fuzzy regions instead of the definitive independent blocks (Arsigny et al. 2005).

One of the parameters which needs to be selected is the size of the block. This parameter controls the degrees of freedom in the registration. If a small block size is selected there will be inadequate information inside the block to find the registration, while a large block size may not capture local deformations. Several algorithms are proposed in the literature that can automatically calculate the appropriate block size using image information such as structural information, residual error and motion.

2.3.3 Constraints on transformation

Constraints on transformation models limit the available solution space by imposing constraints based on prior knowledge of the physical properties of the objects in the registration problem.

Inverse consistency

Given two images I_F and I_M to be registered, a basic assumption would be that the transformations that map moving to fixed images $T_{MF} : (I_M \leftarrow I_F)$ and fixed to moving images $T_{FM} : (I_F \leftarrow I_M)$ are directly related:

$$T_{MF} = T_{FM}^{-1} \quad (2.3.14)$$

However, most of the existing algorithms are asymmetric, which violate the above condition. Hence, the accuracy of these algorithms may vary, depending on which image is selected as the fixed image.

One option in overcoming this issue is to add an additional term to the objective function that penalizes the difference between the forward and reverse transformation (Christensen & He 2001).

$$R_{IC} = \sum_{x \in \Omega} (T_{MF}(x) - T_{FM}^{-1}(x)) \quad (2.3.15)$$

Another solution is to transform both images to the same domain and construct the final transform by taking the inverse of one part and adding it to the other.

Topology preservation

The aim of topology preservation is to keep the connected structures together and maintain the neighborhood relationship between structures. This will prevent the disap-

pearance of existing structures and the appearance of new structures in the image (Musse et al. 2001). In order to preserve topology, the transformation function must be continuous, bijective and the determinant of the Jacobian must be positive.

Diffeomorphic transformation

A diffeomorphic transformation maps a differentiable manifold to another differentiable manifold such that both the function and its inverse are differentiable. This means that a diffeomorphic transformation has a unique inverse which maps that fixed image back to the moving image. The diffeomorphic assumption is violated in some cases such as images before a surgery and after, where some anatomical structures are changed (Christensen 1999). Diffeomorphic mapping also guarantees topology preservation.

Continuum mechanical methods such as linear elasticity and viscous fluid were used to enforce diffeomorphism by Christensen (1999) and Christensen et al. (1996).

Vercauteren et al. (2007) used the demons algorithm combined with the Lie group framework on diffeomorphism and an optimization procedure on Lie groups to provide non-parametric free-form diffeomorphic transformation.

2.4 Optimization

The optimization procedure tries to find the transformation model that best aligns the images according to an objective function. The objective function Ψ employed by most image registration scenarios is nonlinear as well as non-convex.

$$\Psi(\theta) = \mathcal{M}(I_F(x), I_M(T(x))) + \mathcal{R}(T_\theta) \quad (2.4.1)$$

Here $\mathcal{M}$ is the similarity measure which quantifies the alignment between the two images and $\mathcal{R}$ is a regularization term that incorporates a prior known property of the deformation. The similarity form for most problems is organized such that when the alignment is at the highest, the value of the similarity measure is the minimum (used like a dissimilarity measure). This leads to a nonlinear optimization problem of the form:

$$\hat{\theta} = \arg \min_{\theta} [\Psi(\theta)] \quad (2.4.2)$$

where θ represents the parameters of the transformation and $\hat{\theta}$ stands for the optimal parameters that best align the images.

In some cases the parameters of the registration can be calculated directly by solving a set of linear equations. An example of such a scenario is landmark (with correspondences) based registration using radial basis function. Other than these specific cases, most registration problems need an iterative approach to find the optimum transformation model. These optimization algorithms can be categorized into two main parts: *continuous optimization* and *discrete optimization*.

2.4.1 Continuous optimization

Continuous optimization methods can be used in problems where the parameters are real valued and the objective function is differentiable. Generally, these methods update the parameters using an iterative method that follows the equation:

$$\theta_{n+1} = \theta_n + \alpha_n g_n(\theta_n) \quad (2.4.3)$$

where θ are the model parameters, n is the index of the current iteration, α_n is the step length and $g_n(\theta_n)$ is the search direction. There are several methods that belong to this category and they differ based on the method used to calculate step length and search direction.

In gradient descent (or steepest descent) the cost function gradient at the current position is selected as the search direction. Once the direction is set, a line search method or a gradually decreasing function can be used as the step length (Klein et al. 2007). The gradient descent has a low convergence rate and a better method in terms of convergence is the nonlinear conjugate gradient method, where the search direction is along the gradient component from which the previous gradient direction is excluded. In both of the above algorithms the knowledge of a mathematical model is needed in order to calculate the gradients. However, in Powell's conjugate gradient method no such information is needed as it does not calculate gradients.

In quasi-Newton methods, the Hessian of the objective function is set as the step length. Because the calculation and inversion of the Hessian are computationally expensive, an approximation of the Hessian is commonly used. Some well-known methods to approximate the Hessian are Symmetric-Rank-1, Davidon-Fletcher-Powell and Broyden-

Fletcher-Goldfarb-Shanno.

The Gauss-Newton method is a gradient-based optimization method that can be employed when the objective function is in the form of sum of squares. Calculation of the Hessian, which is computationally expensive, is not needed for this method. Here Hessian is approximated using the Jacobian and the search direction is given by $(J^T(\theta)J(\theta))^{-1}\nabla_{\theta}$. The Levenberg-Marquardt method is also used to optimize a function defined as the sum of squares but here the parameter updates are adaptively changed between gradient descent update and Gauss-Newton update. The search direction is given by $J^T(\theta)J(\theta) + \lambda \text{diag}(J^T(\theta)J(\theta))^{-1}\nabla_{\theta}$. The advantage of this method is that it can converge to the solution starting from a location far away from the final solution.

Another very popular method in medical imaging is the stochastic optimization, where an approximation of the gradient is used as the search direction together with a diminishing step size. The gradient approximation can be calculated using either analytical gradients or only a subsample of points, the Kiefer-Wolfowitz method or simultaneous perturbation. Klein et al. (2007) showed that of the above methods, analytical gradients calculated using only a subset of points selected by uniform sampling is the most efficient. Uniform sampling gives equal significance to all the voxels in an image and since the true function in this case is not distributed uniformly, using only a subset of voxels produces a biased estimate of the gradients. To reduce the sampling bias, they proposed renewing the set of chosen samples at every iteration of the optimization routine, giving equal significance to all the voxels. To improve the registration accuracy, Bhagalia et al. (2009) introduced the idea of using the importance sampling (IS) technique to select a subset of voxels. Tennakoon et al. (2014) proposed the use of rank ordered statistics-based robust segmentation technique on the intensity difference image to identify the relevant voxels for registration and used these in calculating the gradients.

2.4.2 Discrete optimization

Discrete optimization schemes are used when the parameters of the optimization problem are in discrete forms. For instance, the image registration problem can be seen as a discrete optimization problem when posed as a metric labeling problem. However, generally this will result in a non-convex energy with many local minima and the optimization would involve a large set of parameters in which finding a solution is computationally

expensive. Kleinberg & Tardos (2002) have shown that the original registration problem when posed as a metric labeling problem is equivalent to a discrete Markov random fields (MRF) minimization problem, where the aim is to assign each node (voxel or parameter) a label $l \in \mathcal{L}$ (where $\mathcal{L}$ is a discrete label set) which minimizes an energy function $E(l)$ of the following form:

$$E(l) = \underbrace{\sum_{p \in P} D_p(l_p)}_{\text{Data Term}} + \lambda \underbrace{\sum_{\{p,q\} \in N} V_{p,q}(l_p, l_q)}_{\text{Smoothness Term}} \quad (2.4.4)$$

In this equation, $D_p(l_p)$ is the cost of assigning label l_p to node p and $V_{p,q}(l_p, l_q)$ is the cost of assigning labels l_p, l_q to neighboring pixels p, q . Recently many techniques have emerged that can solve this problem efficiently and these methods can be segmented into three main groups (Glocker et al. 2011): graph-cut based methods, message passing methods and linear programming methods.

Graph cut-based methods use the max-flow min-cut principle (Ford & Fulkerson 2010). Here a graph is constructed where each node is connected to its neighbours and two special nodes called the source and the sink. Once the graph is constructed, the minimum cut that segments this graph into two partitions is given by the saturated edges at the maximum flow from source to sink. Boykov et al. (2001) extended this idea to multiple labels. They introduced two algorithms: the $\alpha - \beta$ -swap and α -expansion for energies where $V_{p,q}(\cdot)$ is semi-metric and metric, respectively. These methods use large moves compared to standard algorithms making them computationally efficient and there is also a theoretical guarantee of convergence within a known factor (vicinity) of the global minimum. So & Chung (2009) and So et al. (2011) formulated the image registration problem as a discrete labeling problem and assigned each voxel a discrete displacement label. They used an energy function which comprised of SAD as the data term and a smoothness term that penalizes the difference in displacement in adjacent voxels. They used the α -expansion algorithm to derive the optimal labeling.

Graph cut methods can only be applied to a limited set of energy functions (Kolmogorov & Zabini 2004). However, the Markov random fields problem discussed above can be at least approximately solved using the max-product belief propagation (BP) technique introduced by Pearl (1988). BP uses local messages passed along the edges of a graphical model to derive the solution. The solution will only be exact in the case when

the graph is a tree, but if the graph contains loops only an approximate solution can be found. The generalized BP algorithm proposed by Yedidia et al. (2000) achieves better convergence than the ordinary BP. Max-product tree-reweighed message passing (TRW) introduced by Wainwright et al. (2005) is a more generalized version that efficiently finds the solution to any type of graph and this approach was further modified by Kolmogorov (2006). A detailed description of the BP algorithm is provided in Kschischang et al. (2001), Szeliski et al. (2008).

Shekhovtsov et al. (2008) and Lee et al. (2008) used the TRW algorithm in 2D and 3D intensity-based image registration respectively. They decomposed the graph into single dimensional graphs in order to reduce the computations. Kwon et al. (2011) used TRW to optimize an energy that contains a dense local descriptor as data term and higher order smoothness prior.

Linear programming-based methods try to improve efficiency by solving a linear programming relaxation of the original labeling problem. These methods are generally faster than graph cut-based methods and the regularization term does not have to be metric as in α -expansion. Glocker et al. (2008) used the free-form deformation model with B-splines to reduce the dimensionality of the problem and used an energy function that is projected onto the control points (which comprised of an arbitrary similarity measure and a smoothness measure that penalizes the difference in adjacent parameter labels). Then the problem is modeled as an equivalent integer program which was subjected to linear program relaxation. The optimal labeling is then found by using the fast primal-dual algorithm (Komodakis et al. 2007). Additionally, they used a hierarchical warping scheme to account for the large deformation and imposed a hard constraint to preserve the diffeomorphism. In Glocker et al. (2011) the authors extended this work to integrate both iconic and geometric registration into a unified formulation and proposed two algorithms. The first is a computationally-efficient implementation while the second uses a tight relaxation.

2.4.3 Hierarchical methods

The hyper surface that represents the objective function in medical imaging is complex and may contain many local minima and the gradient-based approaches explained earlier may be trapped in these local minima and return a sub-optimum solution. Hi-

erarchical strategies are often used to overcome the above problem in the optimization process. These methods can be broadly categorized into three main parts: increasing data complexity, Gaussian scale space and increasing model complexity.

In increasing data complexity, the image is down-sampled and a coarser version is used in the initial registration levels and gradually finer images are used in subsequent levels. In Gaussian scale space the initial levels use an image convolved with a Gaussian kernel with large standard deviation and the standard deviation is gradually reduced in subsequent steps. In most practical implementations these two methods are combined.

In the increasing model complexity method, the initial registration is performed using a transformation model that has low degrees of freedom such as the affine or rigid motion and then the degrees of freedom are increased in subsequent steps.

2.5 Summary

Image registration of medical images is a fundamental task in image processing and it has been studied for over three decades. Due to the nonrigid motions associated with soft tissue, the different types of modalities involved in capturing these images and the large number of parameters needed to model the associated deformation image registration problem, the associated objective function, transformation models and optimization procedures are highly nonlinear. This chapter has outlined the nonlinear methods used in the key stages of solving the medical image registration problem.

Chapter 3

Quantification of Smoothing

Requirement for 3D Optic Flow Calculation

3.1 Introduction

Motion estimation is one of the most crucial and well-studied problems of computer vision. The underlying task is very general and has a wide range of applications. In the context of medical imaging, the estimation of motion has received substantial attention. With the advance of 3D dynamic imaging by MR, CT and ultrasound, motion estimation has become important in diagnostics, for instance, to assess localized abnormalities in heart wall motion or vessel distensibility, as well as in radiation therapy planning and to compensate for soft tissue motion during image-guided interventions.

From the beginning, two distinct approaches to the estimation of apparent motion (or optic flow) have emerged. The first approach, described by Horn & Schunck (1981), views estimation as a *global* optimization problem and attempts to find globally smooth warping regimes that relate sequential images to each other. The second approach, presented by Lucas & Kanade (1981), views the problem as a *local* correspondence problem.

In both of the above approaches, smoothing plays a crucial role. The optic flow, due to the well-known aperture problem, is ill-posed and cannot be solved for a single data point (a pixel in 2D or a voxel in 3D). Therefore, some degree of smoothing or regularization is always required.

From a historical perspective, the *global* and *local* approaches were conceptually much closer at the beginning (in the late 1970s and early 1980s) than they are now. A typical test image sequence of the time consisted mostly of a few thousand gray level pixels (small digitized TV signal) of usually a single flat moving object (see the Results section of Horn & Schunck (1981)). Therefore, it is important to interpret their assertion that apparent velocity “varies smoothly almost everywhere in the image” (Horn & Schunck 1981) in its context, which is very different to today’s concept of an image sequence typically containing a large amount of detail including several motions.

It is also important to note here that, although some excellent results have been produced by refining global methods (Bruhn, Weickert, Kohlberger & Schnörr 2005, Modersitzki 2008, Papenberg et al. 2006) and the accuracy of optical flow estimation, as measured by the Middlebury benchmark (Baker et al. 2011), has been improving, the effect of local smoothing on the estimation accuracy is yet to be fully understood. To examine the underlying cause of the success of those methods, a baseline method similar to Horn & Schunck (1981) was used in Sun et al. (2010) to study the influence of different choices of how to model an appropriate objective function and its approximation (for computational tractability) and optimization on the overall accuracy. Their comparisons showed that applying a median filter to optical flow estimates in different iterations of those algorithms produced the most significant improvements. Although this is a form of local smoothing, a theory of how much spatial smoothing is required, at a given scale, is yet to be developed. In particular, the above study does not consider the effect of smoothing imposed by the Lucas & Kanade (1981) formulation of the optical flow problem. This chapter is an attempt to address this important question by examining how much smoothing would be sufficient from a local perspective. In this context, Bruhn, Weickert & Schnörr (2005) address the question of how to merge global and local approaches, while we aim to answer the question of where the meeting should take place. In other words, the desire is to study the concept of *just enough* smoothing: the least amount of smoothing that both overcomes the ill-posed nature of this problem and ensures the desired accuracy of the estimation process.

In 2D optical flow estimation, the work of Xu et al. (2012) has shown that the imposition of global smoothing can be partial toward global changes and highlights those changes (motions) at the expense of localized variations. In volumetric images, where

there is no 3D to 2D projection, the bias introduced by imposing more than just enough smoothness is particularly undesirable. For example, in 4D chest CT scans the lungs display a complex deformation pattern during the respiratory cycle, with motion boundaries where the lungs slide along the rib cage and the lung lobes move relative to each other. The elastic characteristics of airways and large blood vessels differ from those of lung parenchyma, and pathologic tissue such as pulmonary nodules can be expected to deform differently from its surroundings. Global smoothing may lead to errors in the estimation of tumor motion, and thus to inaccuracies in derived treatment plans.

In cases where some prior knowledge does exist, this can simply be included in the motion models. However, the analysis presented here is based on the distribution of residuals near a motion boundary and the result does not depend upon the type of model used.

The concept of *just enough* smoothing has, in the past, been considered in the context of scale space theory and in particular, for finding the appropriate scale to terminate a multiscale hierarchical algorithm (Lindeberg 1998). However, the scale space approach differs from the robust statistics approach presented in this chapter in a fundamental way. In the scale space approach, the discontinuities (motion boundaries) are modeled by transitions toward higher scales, in which small discontinuities, with the exception of the main motion, are smooth enough to be considered continuous. The problem is then solved at that scale and the solution is then propagated to the finer scales to recover smaller motions. The question of *just enough* smoothing therefore refers to finding the scale at which the estimation is most reliable.

In robust statistics, the discontinuities are explicitly modeled as separate instances of a single model (or a finite set of plausible models) and the scale of each instance and the shape of the smoothing window are estimated concurrently with the model parameters (Meer 2004). In this context, just enough smoothing refers to the size of the locality (size of population, bandwidth, etc. depending on the type of estimator) on which the estimation is based. To our knowledge, the relationship between the amount of smoothing and accuracy of the estimation has thus far not been quantified either by analysis or experiment, and more importantly, no specific link between the amount of sufficient smoothing and the accuracy of estimation has yet been established.

Comparison of existing results for the calculation of 2D optic flow reveals that for

real images with discontinuous flow (like the Otte image sequence (Otte & Nagel 1994) in contrast to synthetic images like the Yosemite sequence with fairly smooth flow), local robust approaches perform as well as global approaches (for instance, compare the results presented in Table 4 of Bab-Hadiashar & Suter (1998) with those presented in Table 8 of Papenberg et al. (2006) and Table 5 of Bruhn, Weickert & Schnörr (2005)) and the best available result thus far (Bruhn, Weickert & Schnörr 2005) is achieved by modeling the local discontinuities using a robust estimator. Therefore, the question of what should be the extent of local smoothing is also important in fine-tuning hybrid (combining local and global) methods.

3.1.1 3D Optic flow estimation

In contrast to 2D optic flow calculation, the use of 3D methods, particularly for dynamic CT images, has only started to attract the attention of practitioners in recent years (Boldea et al. 2008, Yang et al. 2008, Zhang et al. 2008) and its issues and potential are yet to be fully analyzed. In the biomedical imaging area, 3D optic flow calculation was first used to capture heart 3D motion using CT (Song & Leahy 1991) and MRI (Andreopoulos & Tsotsos 2008) images. The pioneering work of Song & Leahy (1991) assumes that the image is conserved and incompressible and therefore the velocity field satisfies the divergence-free and the incompressibility constraints. However, the computation of the flow using these two constraints is an ill-posed problem and the solution was found by adding a smoothness term to regularize the penalty function of the weighted sum of the two constraining terms. The velocity field was then calculated by minimizing this penalty function using variational calculus (the minimizing solution generally satisfies the Euler-Lagrange equations) and the solution was found by solving a set of simultaneous coupled elliptic partial differential equations. The differential equations were further discretized, resulting in a system of linear equations where the solution is an approximation of the velocity field. This method, in essence, is very similar to some of the contemporary approaches for both calculating the 2D optic flow (Bruhn et al. 2003, Bruhn, Weickert, Kohlberger & Schnörr 2005, Bruhn, Weickert & Schnörr 2005, Weickert & Schnörr 2001) and energy minimization based image registration (Fischer & Modersitzki 2003, Haber & Modersitzki 2007, Modersitzki 2008) techniques.

With the exception of the 3D generalizations of the original version of the Lucas &

Kanade optic flow method (Barron & Thacker 2005, Castillo, Castillo, Zhang & Guerrero 2009), to our knowledge, no other Lucas & Kanade-based 3D optic flow method has yet appeared in the computer vision literature and the potentials and difficulties associated with using such methods on dynamic 3D data such as 4D CT are yet to be explored. The Lucas & Kanade optic flow based method, in contrast to variational-based methods, is of particular interest here because it allows the effect of smoothing to be directly controlled and measured.

In this chapter, we first examine the theoretical relationship between the estimation accuracy and the amount of smoothness required for a general solution from a robust statistics perspective. The analysis leads to a guideline for the sufficient amount of smoothing for the 3D optic flow estimation.

A robust 3D optic flow in which the imposition of smoothing can be locally quantized is then devised to test the proposed hypothesis on the smoothing requirement. The hypothesis is then tested using a geometrically-realistic synthetic CT image sequence of the breathing lung and five cases of real 4D CT lung images with extensive sets of expert annotated landmarks. We further examine, both quantitatively and qualitatively, the suitability of imposing “just enough” smoothing on real 4D CT scans in calculating motions near fissures (borders of lobes) and show the deteriorating effect of unnecessary smoothing on the estimation of motion, particularly in those areas.

3.2 How Much Smoothing is “Just Enough”?

To answer the above question, we first need to establish an explicit relationship between the smoothing requirement and the estimation accuracy. To ascertain this relationship, we first consider the 3D optic flow estimation problem. The optic flow constraint in 3D is generally written as (Song & Leahy 1991):

$$I_x u + I_y v + I_z w + I_t = 0 \quad (3.2.1)$$

where I_x , I_y and I_z are the spatial derivatives and I_t is the temporal derivative of the image brightness function I and u , v and w are unknown components of the local flow along x , y and z axes respectively. The imposition of this constraint implies that for every voxel, there is only one equation for three unknowns and therefore it is not possible to solve this

without adding the extra assumption that is generally referred to as smoothing.

The simplest form of smoothing is imposed by assuming the flow to be constant in a local neighborhood and therefore, the above equation is turned into a system of linear equations that can be solved for the unknown components of the flow. In this context, the question of how much smoothing is imposed directly relates to the size of this neighborhood and the number of voxels included in the calculation of velocity components. We use this scenario as the basis of our analysis.

To measure the effect of smoothing on accuracy, we then need to examine how the estimation is performed. We already know that in the absence of noise one needs to apply the smoothing assumption to only three voxels to be able to calculate the flow in 3D (this is the necessary condition). However, noise is always present and therefore substantially more voxels need to be included. The *extra smoothing assumption is therefore required solely for the sake of accuracy*. More importantly, as the motion boundaries and unmodeled data are unavoidable, some form of outlier rejection (robust statistics) is also required to ensure accuracy. It is worth noting here that assuming other models of motion (for example affine) only affects the amount of necessary smoothing (would be 12 voxels for full 3D affine). Regardless of the chosen model, one would need to include significantly more points than the necessary number in order to obtain an accurate estimate and therefore the sufficiency requirement is not directly affected by the type of motion model.

Having formulated the problem in this setting, for a given level of noise, the amount of smoothing required now directly relates to the number of included voxels and in turn, the number of included voxels directly relates to the accuracy of estimation that is expressed by a measure called *finite sample bias*. Following Hoseinnezhad et al. (2010), the finite sample bias of a robust estimator is defined as:

$$\lambda(n; \Theta^*, H) \equiv \frac{|E[\hat{\sigma}_n^2; \Theta^*, H] - \sigma^2|}{\sigma^2} \quad (3.2.2)$$

where σ is the true scale and $E[\hat{\sigma}_n^2; \Theta^*, H]$ is the statistical mean of the square of an estimated scale for a given hypothesized fit Θ^* and a specific data distribution H . The arguments n , Θ^* and H are to show that this is a scale-invariant definition of bias and the above measure only depends on the number of data samples, the hypothesized fit and the data population, but not on the scale σ .

Using the above definition, we propose an approach that generates a straightforward

relationship between the required smoothing and the estimation accuracy by which the minimum *sufficient* smoothing requirement can be evaluated based on the desired level of accuracy. The analysis here does not include the relationship between the estimation bias and the level of noise, as the latter is fixed in a given set of data. Interested readers are referred to Hoseinnezhad et al. (2010) for a detailed discussion of that relationship.

For the sake of quantifying the amount of “just enough” smoothing, we also need to choose the estimator. Based on the analysis of Hoseinnezhad & Bab-Hadiashar (2007), we implemented the Modified Selective Statistical Estimator (MSSE) (Bab-Hadiashar & Suter 1999), as the most consistent¹ robust estimator for the estimation task. The analysis presented in Hoseinnezhad et al. (2010) shows that all estimators, including MSSE, are heavily biased when the distance between the two parallel structures (e.g. two very similar motions in a selected region), μ , is less than 5σ . In this scenario, structures in the data are too close to be separated from each other and existence of one structure heavily biases the estimation of the other structure. The analysis also showed that the estimation accuracy would not improve by increasing the amount of data. However, for $\mu \geq 5\sigma$, the MSSE finite sample bias is always less than 20%, and the finite sample bias does not change with the inlier percentage or the distance between the two structures (μ). Importantly, the finite sample bias of the MSSE as well as a number of other robust estimators analyzed in Hoseinnezhad et al. (2010) do not significantly change as the number of data samples increases beyond a relatively small sample size of around 100. Therefore, the minimum amount of required smoothing would not be significantly different if any other high breakdown estimators were used instead.

We also assume that the smoothing support is a cube (as the data is discretized in a Cartesian grid) centered on the voxel of interest. In the absence of any prior information, the cube is to be constructed symmetrically around the target. Having specified the estimation parameters, we are now able to hypothesize the minimum size requirement of the Gaussian window by analyzing the finite sample bias of the estimator in cases where there are multiple motions in the region of interest. Since we use a locally constant velocity model, the motion boundary is modeled by a step edge in the velocity space.

The finite sample bias of various robust estimators, including MSSE, for segmenting a step edge structure, as the worst-case scenario, has already been analyzed (Hoseinnezhad

¹An estimator is said to be consistent if its estimated value approaches the true value as the number of data approaches infinity.

et al. 2010). The regression framework used in Hoseinnezhad et al. (2010), as shown in Section 3.4.4, is identical to the flow estimation presented here. Part of the results presented in Figure 9 of Hoseinnezhad et al. (2010) are reproduced here and shown in Figure 3.1. The above results show that for cases where the estimator is consistent, less

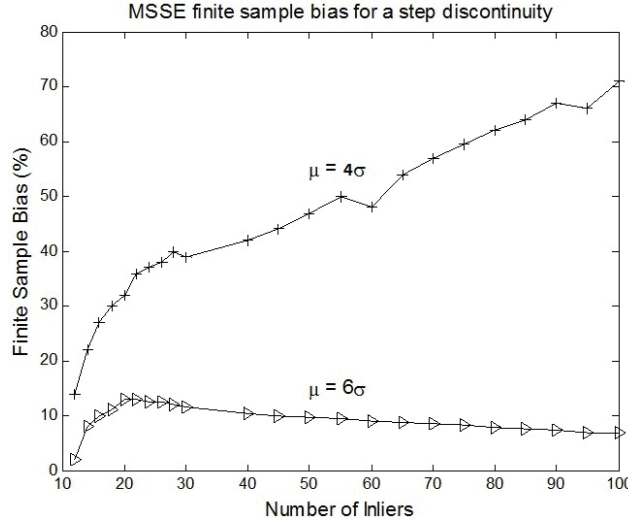


Figure 3.1: Finite sample bias of MSSE in the segmentation of a step edge discontinuity with height of either 4 or 6 times the scale of noise in the measured data with 30% inliers. It is important to note here that as shown in Hoseinnezhad et al. (2010), for $\mu > 5\sigma$ the amount of finite sample bias is not affected by either the μ or the inlier ratio.

than 100 data points are required to minimize the effect of the finite sample bias of the robust estimator. Therefore, a cube with sides as small as 5 or 7 voxels should provide the minimum required smoothing, even in cases where the target voxel is in the vicinity of a motion boundary and only a fraction of voxels actually belong to the motion of interest. The minimum requirement in 3D data is therefore delightfully small (a CT scan typically has more than 10 million voxels). This means that very localized changes should be directly observable. We examine this hypothesis both in terms of the average accuracy using synthetic and real data (with known ground truth) and its implication for segmenting lung motion based on known anatomical features in real 4D CT data.

3.3 Estimation of 3D Optic Flow

To examine the full implications of the smoothing requirements in the context of robust optic flow calculation, we present a straightforward optic flow method. In this method, we estimate the 3D flow by assuming that the majority of voxels in a local cube can be explained by velocity perturbations around a constant value (inlier group). Voxels

with motions that cannot be explained by the above model (based on the MSSE criteria) would be considered outliers and would not be included in the estimation process. The velocity of the inlier group is calculated using the least square method.

We have implemented the above regularization approach and solved the estimation (MSSE (Bab-Hadiashar & Suter 1999)) step using random sampling. The spatial and temporal derivatives are calculated as prescribed by Barron & Thacker (2005) using either Simoncelli's (Simoncelli 1994) or Gaussian derivative masks.

The implementation of MSSE, as described in Algorithm1, is very straightforward and only involves taking N number of random samples of 3 voxels from within the support volume. The 3D flow is then calculated for each 3-tuple by solving the system of three linear equations for the three unknown components of the flow at that point. The flow vectors for all samples are then used to calculate square residuals (the algebraic distance between voxels' optic flow constraints and the calculated flow of a given 3-tuple) of all voxels inside the support window. The sample that has the minimum sorted square residuals at the K^{th} order index of those residuals is selected as the best estimate (in our experiments we set $K = 0.5$ which gives the median). By starting from the K^{th} residual of this best estimate, the point where the condition: $|r_{i+1}| < T\sigma_i$ based on $\sigma_i^2 = \sum_{j=1}^i r_j^2 / (i - 3)$ is no longer true is found and voxels up to this index are considered as inliers. In the above, r is the residual and is given by: $r^2 = (I_x u + I_y v + I_z w + I_t)^2$ where i is the sorted index, T is a constant threshold and numbers around 1.5-2.5 is usually used to indicate an inclusion of around 93-99% of inliers based on a normal distribution for noise. For a general noise distribution Chebyshevs inequality² states that no more than $\frac{1}{T^2}$ points can be more than $T\sigma$ distance away from the mean. This indicates an inclusion of around 40-66% for the above range of T values. The least squares solution of all the inlier voxels is considered as the final estimate. Since the final scale estimate is calculated by using least squares and including all inliers, the variation of T has little effect on the final results (Bab-Hadiashar & Suter 1999).

²For a random variable X with expected value μ and a variance σ^2 , $Pr(|X - \mu| \geq T\sigma) \leq \frac{1}{T^2}$ where T is a constant real number.

Algorithm 1 The Step-by-Step Algorithm for MSSE

Inputs: Spatial and Temporal Derivatives, Number of repetitive epochs N

- 1: Repeat steps 2-6 N times;
- 2: Choose an elemental subset(3-tuple) by random sampling;
- 3: Compute the corresponding velocity vector using

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = - \begin{bmatrix} I_{x1} & I_{y1} & I_{z1} \\ I_{x2} & I_{y2} & I_{z2} \\ I_{x3} & I_{y3} & I_{z3} \end{bmatrix}^{-1} \begin{bmatrix} I_{t1} \\ I_{t2} \\ I_{t3} \end{bmatrix}$$

- 4: Calculate the square residuals $r^2 = (I_x u + I_y v + I_z w + I_t)^2$;
 - 5: Sort the square residuals in ascending order;
 - 6: Find the sample with the least K^{th} square residuals;
 - 7: Recalculate the square residuals using the velocity vector with the least median square residual and sort them;
 - 8: Find the first point starting from the median where $|r_{i+1}| > 2.5\sigma_i$. The data up to this point are considered inliers;
 - 9: Calculate the final velocity vector using all the inlier points.
-

3.4 Experimental Results

To investigate the effect of local smoothing on the accuracy of optic flow, we created a sequence of synthetic 3D images with a variety of known motions. The sequence is designed to mimic typical changes in lung CT sequences with several objects having different motions and irregularly-shaped motion boundaries. The geometry of lungs is generated from the segmentation of a real human lung CT image and the textures are created by superimposing three 3D sinusoidal patterns similar to those used in Barron et al. (1992).

In this sequence, the lungs have affine motions while the background has constant 3D motion. A small stationary column between two lungs has also been included to simulate the motionless parts of a real image. A sample 2D slice (axial view) and its associated flow field are shown in Figure 3.2. Although the texture in this sequence is synthetic, the geometry and types of motions are quite realistic and exhibit issues similar to those encountered in real images, including issues associated with the existence of quantization noise and deterioration of the estimation of derivatives near the motion boundaries. However the lungs are not segmented into different lobes (with different motions) because the analysis is based on using a spatially small area and therefore having different segments does not improve the relevance of the experiment.

For error measurement, we have extended and used the Fleet & Jepson (1990) angular

measure of error. In the extension of this scheme to 3D, the flow at every voxel is represented by a 4D vector of its Cartesian components in a homogeneous coordinate system $(u, v, w, 1)$ and the error is measured as an angular deviation of the calculated flow from the true velocity. The error is therefore calculated by finding the inverse cosine of the dot product of two vectors in the above 4D homogeneous coordinate system. A detailed analysis of this measure is provided in Barron et al. (1992) and will not be repeated here.

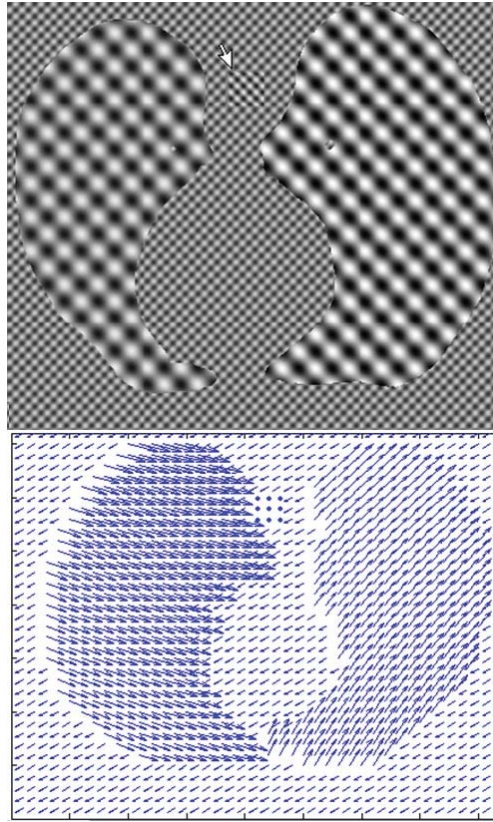


Figure 3.2: Sample of axial view of the sinusoidal image (top) and its flow field (bottom). The white arrow identifies the stationary column area.

3.4.1 Effect of smoothing on accuracy

To demonstrate the effect of the size of the Gaussian window on estimation accuracy, we varied the size from $3 \times 3 \times 3$ to $11 \times 11 \times 11$ voxels. The results shown in Table 3.1 are in full agreement with our predictions based on the finite sample bias (see the last part of Section 3.2). The accuracy is significantly enhanced when the size of the window increases from $3 \times 3 \times 3$ to $5 \times 5 \times 5$ (which is only around 0.00028% of the whole data). However, any more increases in the amount of smoothing result in only small changes to the final accuracy.

Table 3.1: Estimation accuracy of calculating 3D optic flow using different sizes of smoothing windows

Size of Smoothing Window	Average Error (degrees)	Standard Deviation (degrees)
3x3x3	4.49	16.04
5x5x5	2.74	11.10
7x7x7	2.07	8.88
9x9x9	1.84	8.24
11x11x11	1.75	8.12

3.4.2 Optic flow estimation using real 4D CT images

Both cardiac and respiratory motions severely affect the quality of lung CT images. Four-dimensional CT images are developed to facilitate the analysis of respiratory motion by using spirometer signals to align and synchronize data acquired during different breathing cycles. However the data contain a large amount of noise owing to the low-dose protocols used in dynamic CT imaging as well as various types of artifacts due to the fact that lung deformation is not exactly the same in each respiratory cycle (4D CT is typically acquired during several cycles) and cardiac motion is out of sync with the breathing cycle. However, 4D CT data is increasingly used for tracking lung motion and helps clinicians minimize and better target radiation in oncology treatments.

Currently there are two commonly used and openly available thoracic 4D CT data sets with ground truth in terms of landmark motions (Castillo, Castillo, Guerra, Johnson, McPhail, Garg & Guerrero 2009, Vandemeulebroucke et al. 2011) that can be used to validate the accuracy of dense flow calculations. In our experiments, we used both datasets to show real data applications of the proposed theory. The data provided by the University of Texas M.D. Anderson Cancer Center (Castillo, Castillo, Guerra, Johnson, McPhail, Garg & Guerrero 2009) contains 300 manually-identified landmark points per case. The results of nine deformable registration algorithms using these data have also been provided in Castillo et al. (2010), Castillo, Castillo, Guerra, Johnson, McPhail, Garg & Guerrero (2009) and Gu et al. (2010). To provide context, results of similar approaches have also been included here. The dataset provided by the Léon Bérard Cancer Center & CREATIS laboratory, Lyon, France (Vandemeulebroucke et al. 2011) has three images with 100 manually annotated landmarks in all frames.

The first five cases of Castillo, Castillo, Guerra, Johnson, McPhail, Garg & Guerrero (2009) are used here to study the effect of the size of the smoothing window on estimation accuracy. We varied the size from $3 \times 3 \times 3$ to $13 \times 13 \times 13$ voxels and the results are shown in Table 3.2.

Table 3.2: Average error of calculating 3D optic flow using different sizes of smoothing window for real 4D CT data

Gaussian Window size	Case 1	Case 2	Case 3	Case 4	Case 5
	Av. Error (SE) ^a	Av. Error (SE)	Av. Error (SE)	Av. Error (SE)	Av. Error (SE)
Average Disp.	4.01 (2.91)	4.65 (4.09)	9.42 (4.81)	6.73 (4.21)	7.10 (5.14)
Maximum Displacement	12.65	17.8	21.0	18.46	24.78
3x3x3	5.028 (0.416)	4.496 (0.371)	6.383 (0.374)	9.193 (0.450)	9.018 (0.853)
5x5x5	2.945 (0.221)	2.916 (0.192)	4.349 (0.205)	6.062 (0.257)	5.986 (0.381)
7x7x7	2.269 (0.138)	2.315 (0.141)	3.528 (0.176)	4.773 (0.207)	5.387 (0.314)
9x9x9	1.936 (0.119)	1.897 (0.114)	3.353 (0.160)	4.188 (0.181)	4.969 (0.288)
11x11x11	1.710 (0.097)	1.828 (0.109)	3.156 (0.150)	3.949 (0.165)	4.742 (0.261)
13x13x13	1.661 (0.099)	1.641 (0.098)	3.187 (0.150)	3.785 (0.158)	4.536 (0.252)

^a Standard error as defined in Castillo, Castillo, Guerra, Johnson, McPhail, Garg & Guerrero (2009).

The above results, similar to those obtained using synthetic data, show that the increase in overall accuracy due to the increase in the size of the smoothing window plateaus when window sizes are larger. To illustrate this further, average errors are plotted against smoothing window size in Figure 3.3. To bring all the results to the same scale, the average errors are divided by the average error of $13 \times 13 \times 13$ window size for each case.

To show that slight improvements in the mean accuracy of larger window sizes are not statistically significant, a paired t-test was conducted and the results are shown in Table 3.3. For a given set of images, we calculated the p-value using the landmark regis-

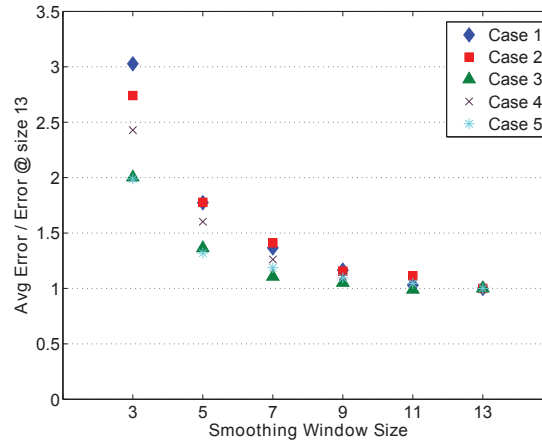


Figure 3.3: Change in average error when the smoothing window size is varied, $\sigma = 2.0$.

tration errors of two adjacent smoothing window sizes. These results were then used to compare the mean errors against a null hypothesis (results are not significantly different) based on a confidence interval of 99%. A p-value $< 1\%$ means that the null hypothesis is rejected ($H = 1$) and the difference between the two results is statistically significant. As expected we can see that in all cases the results improve when the smoothing size increases from $3 \times 3 \times 3$ to $5 \times 5 \times 5$, but increases beyond $7 \times 7 \times 7$ are not statistically significant.

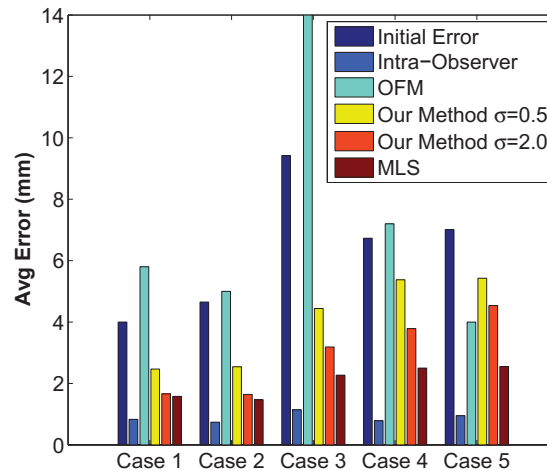


Figure 3.4: Average registration errors of our method are shown for each case. Corresponding average errors from Castillo et al. (2010), Castillo, Castillo, Guerra, Johnson, McPhail, Garg & Guerrero (2009) are also shown for comparison.

The average errors of our method for all these cases were also compared with the

Table 3.3: P-values and hypothesis test results (with confidence interval of 99%) of paired t-tests conducted to test the significance of improvements achieved by increasing the size of the smoothing window. $H = 1$ means that the differences are statistically significant.

		3x3x3 ↓ 5x5x5	5x5x5 ↓ 7x7x7	7x7x7 ↓ 9x9x9	9x9x9 ↓ 11x11x11	11x11x11 ↓ 13x13x13
Case 1	P	9.56E-06	0.008	0.068	0.142	0.722
	H	1	1	0	0	0
Case 2	P	2.00E-04	0.012	0.021	0.660	0.201
	H	1	0	0	0	0
Case 3	P	2.35E-06	0.003	0.460	0.370	0.738
	H	1	1	0	0	0
Case 4	P	2.77E-09	1.00E-04	0.033	0.330	0.473
	H	1	1	0	0	0
Case 5	P	1.20E-03	0.226	0.328	0.560	0.570
	H	1	0	0	0	0

results in Castillo et al. (2010), Castillo, Castillo, Guerra, Johnson, McPhail, Garg & Guerrero (2009)³ and the results are shown in Figure 3.4. In particular, the optical flow method (OFM) in this figure (from Castillo, Castillo, Guerra, Johnson, McPhail, Garg & Guerrero (2009)) represents the results of the Horn & Schunck-based flow calculation method, which is a global method. The average error in landmark registration across all five cases for OFM is 7.3mm, whereas for our method with $\sigma = 0.5$ it is 4.051mm and with $\sigma = 2.0$ the average error is 2.962mm.

The moving least-square (MLS) in Figure 3.4 (from Castillo, Castillo, Guerra, Johnson, McPhail, Garg & Guerrero (2009)) is a landmark-based registration algorithm which uses external information (expert knowledge) to identify the landmarks. MLS achieved an average error of 2.074mm across all five cases. The difference between the average errors of our method ($\sigma = 2.0$) and the MLS is less than the intra-observer error (this is the error in identifying the landmark points by an expert in repeated experiments) for the first 3 cases. This shows that for small displacements, the local optic flow-based method can achieve similar results to those of hand-tuned registration methods.

³The mean errors for our method is calculated using the publicly available landmark set (containing 300 landmarks per image) which is a subset of the full landmarks set used to calculate the results in Castillo et al. (2010), Castillo, Castillo, Guerra, Johnson, McPhail, Garg & Guerrero (2009)

It is important to note that a hierarchical approach would have yielded more accurate results for the last two cases which have higher displacements. We have not implemented this because improving accuracy by itself is not the aim here. In addition, the landmark dataset used in evaluating our method consists of 300 publicly available landmarks.

In the next set of experiments, we used the Popi model dataset provided by Vandemeulebroucke et al. (2011) to demonstrate the effect of smoothing window size on the accuracy of optic flow calculations. The flow field was calculated using two adjacent frames in each 4D CT dataset. All landmarks were used to calculate the mean landmark error for each time step and the results are shown in Figure 3.5. In this figure, each time frame of the 4D CT image is identified by a frame number (T00, T10, T20, etc.) and the mean landmark errors of all landmarks at each time step are averaged over the three cases. These results, similar those achieved for DIR data, are consistent with our theoretical predictions.

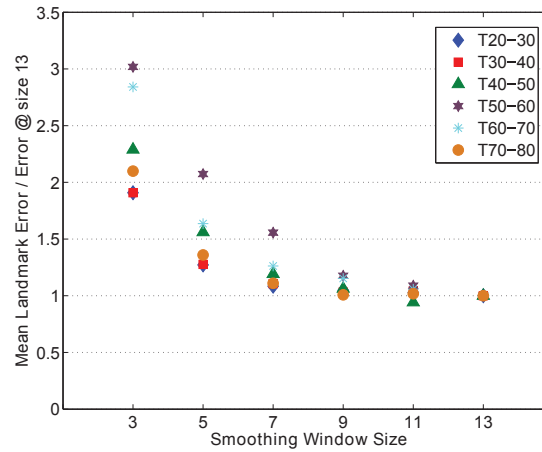


Figure 3.5: Effect of varying the size of smoothing window on average landmark errors in images of Popi model dataset (Vandemeulebroucke et al. 2011).

3.4.3 Side effects of smoothing increases

A lung contains separate lobes with different deformation patterns and the motion at the lobes boundaries (fissure) is discontinuous. The juxtaposition of lobes in a typical lung is shown in Figure 3.6.

To see the effect of the size of the smoothing window on the accuracy of estimation near the fissure, we compared the estimated error for different smoothing window sizes for a small selection of points near the fissure and points in the same image, but further away

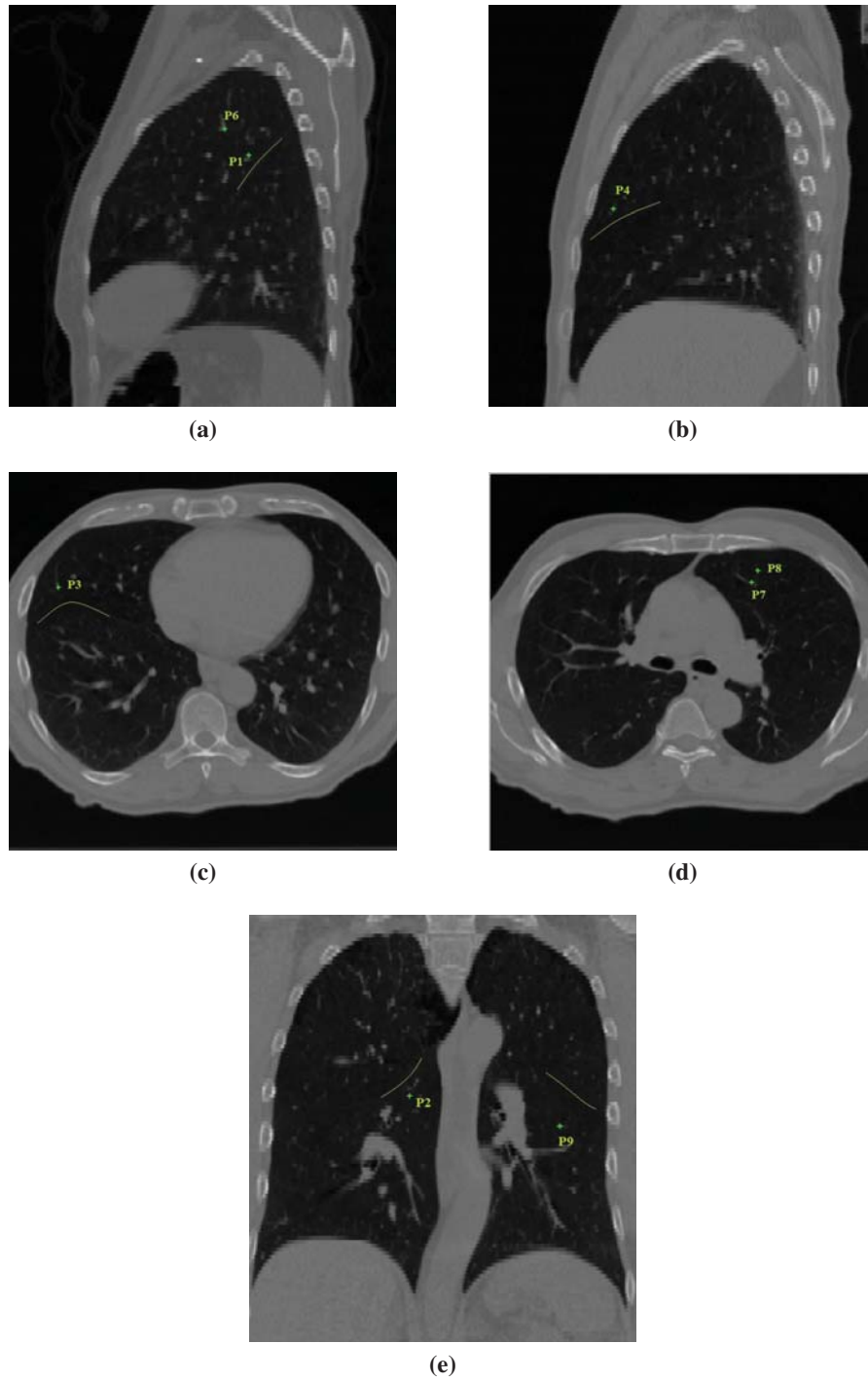


Figure 3.6: The above slices are chosen to depict the position of different landmark points used in our calculations (P1 to P9). Where possible, the fissure is also highlighted in these images.

from the fissure. The results for the estimation error for different smoothing window sizes are shown in Figure 3.7. As predicted, the results show that the increase in smoothing window can indeed reduce the overall accuracy.

Four points (P6, P7 ,P8 & P9) were selected from cases 1 and 2, where P6 is in the same sagittal slice as P1 and P7 & P8 are in the same axial slice as P2, but are further apart from the fissure than P1 and P2. The errors are plotted against the smoothing window size in Figure 3.8.

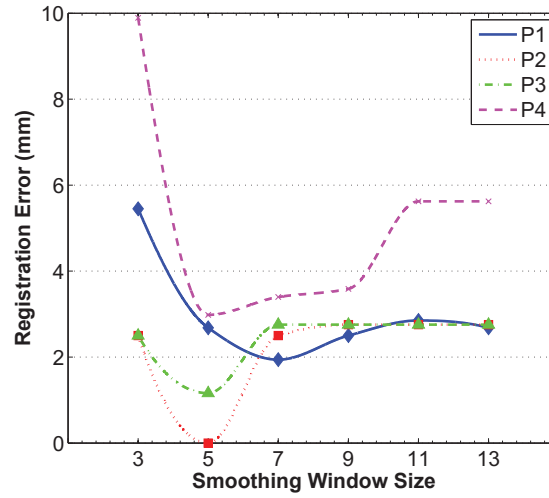


Figure 3.7: Variation in estimation error with smoothing window size for landmark points near a fissure.

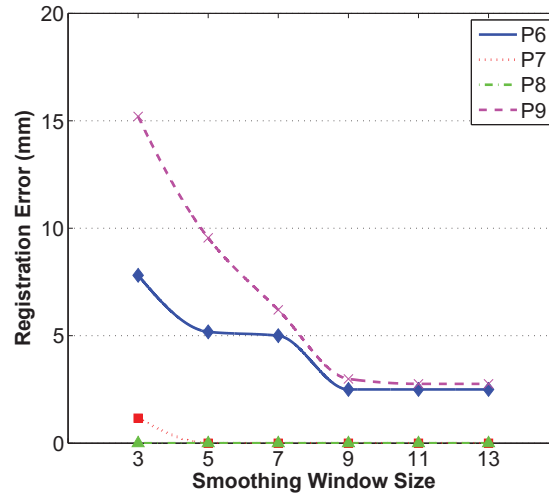


Figure 3.8: Variation in estimation error with smoothing window size for landmark points not near a fissure.

3.4.4 Detection of local variations in flow fields

To highlight the importance of controlling the size of the smoothing window, we concentrate on the detection of anatomically-induced variation in flow field. We first intro-

duce a novel way of visualizing the variations of 3D flow fields in 3D space and use that method to show the detectability of anatomical variation of flow fields on both synthetic and real 4D CT images. To visualize the variation, we rewrite the 3D optic flow constraints (Equation 3.2.1) as:

$$\frac{I_x}{I_t}u + \frac{I_y}{I_t}v + \frac{I_z}{I_t}w = -1 \quad (3.4.1)$$

The optic flow constraint is commonly viewed as a plane in the velocity space of (u, v, w) . However, using the duality of points and planes in perspective geometry, we can view the above equation as a point in the derivative space $(I_x/I_t, I_y/I_t, I_z/I_t)$ and therefore all the points with the same (constant) velocity will form a 3D plane parameterized by (u, v, w) . The importance of this transformation stems from the fact that, while it is very difficult to visualize a number of 3D planes crossing each other around the vicinity of a single point, we can easily visualize a set of points scattering around a plane in a 3D Cartesian space.

To demonstrate the effectiveness of the above transformation to visualize the existence of different motions in a local area, we consider a cubic region in the synthetic lung image introduced earlier. The cubic region is shown in the top part of Figure 3.9 by a white rectangle in the axial view and includes voxels with two different motions: constant and affine. A regular subset (one in nine) of voxels from this area is plotted in the derivative space and the points are segmented by applying MSSE sequentially (i.e. a fit and remove application of MSSE, as explained in Bab-Hadiashar & Suter (1999)).

The resulting planes are shown in the bottom part of Figure 3.9, while the plot is rotated to show the spread of points around one of those planes. As expected, voxels with constant underlying motion fit well to a plane in the derivative space and therefore form a very thin scatter of points around a plane. The voxels with affine motions, on the other hand, are scattered wider while they still generally spread near a plane as long as they are spatially close to each other in the actual image. These results support our earlier hypothesis that good estimates of motion can be achieved using localized small smoothing area.

We can now apply the same transformation to real 4D CT images and examine the possibility of detecting anatomical motion boundaries. We examined the possibility of detecting the differences in motion between individual lung lobes by considering two

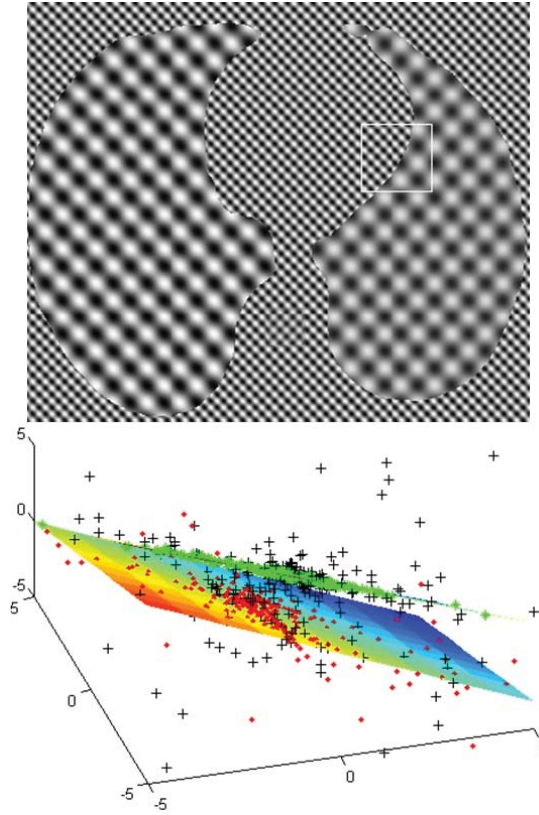


Figure 3.9: Top image is an axial view of the sinusoidal image where the cubic region with two distinct motions is marked by a white rectangle. Bottom image plots the associated voxels in the derivative space (red dots and green asterisks are used for moving voxels while black crosses represent the outliers). The points are segmented by fitting planes using MSSE. Bottom picture is rotated to show how the points are aligned with respect to the affine plane. Since a plane in the derivative space represents constant motion, the points with constant underlying motion (red dots) fit a plane better than the points with affine underlying motion (green asterisks).

small (local) cubic areas in a 4D CT image. As shown in the top part of Figure 3.10, one of these cubes (shown by a white rectangle in the axial view) is entirely within a lobe and the other straddles two different lobes (where the fissure - highlighted by a white line - is located).

We have plotted all the points associated with the voxels in these areas and applied the MSSE (Bab-Hadiashar & Suter 1999) (with 30% minimum ratio and $T=2$) to segment the motions. The segmentation results for both areas are shown in the middle and bottom parts of Figure 3.10. While only one motion is detected in the upper area, the one straddling two lobes includes two distinct motions. This evidence again shows that the imposition of “just enough” smoothing is important in preserving the local flow information.

3.5 Conclusion

A new approach to quantifying the minimum required smoothing based on the concept of the finite sample bias of a robust estimator has been presented. The proposed approach is very general and makes predictions about the amount of smoothness required to satisfy the sufficiency condition for a broad range of visual estimation tasks such as optic flow calculation. We particularly showed that smoothing over a cubic area as small as 5 to 7 voxels wide is sufficient to achieve the highest practical accuracy. This is a significant observation, as it proves that very localized changes in motion in 3D data are directly observable. The predictions were tested for 3D optic flow estimation of 4D lung CT images and consistent with the theoretical predictions the results showed that only a very small amount of local smoothing is required to achieve high accuracy and observe local anatomically-induced motion variations. In some cases, increasing the amount of smoothing reduce the quality of the results.

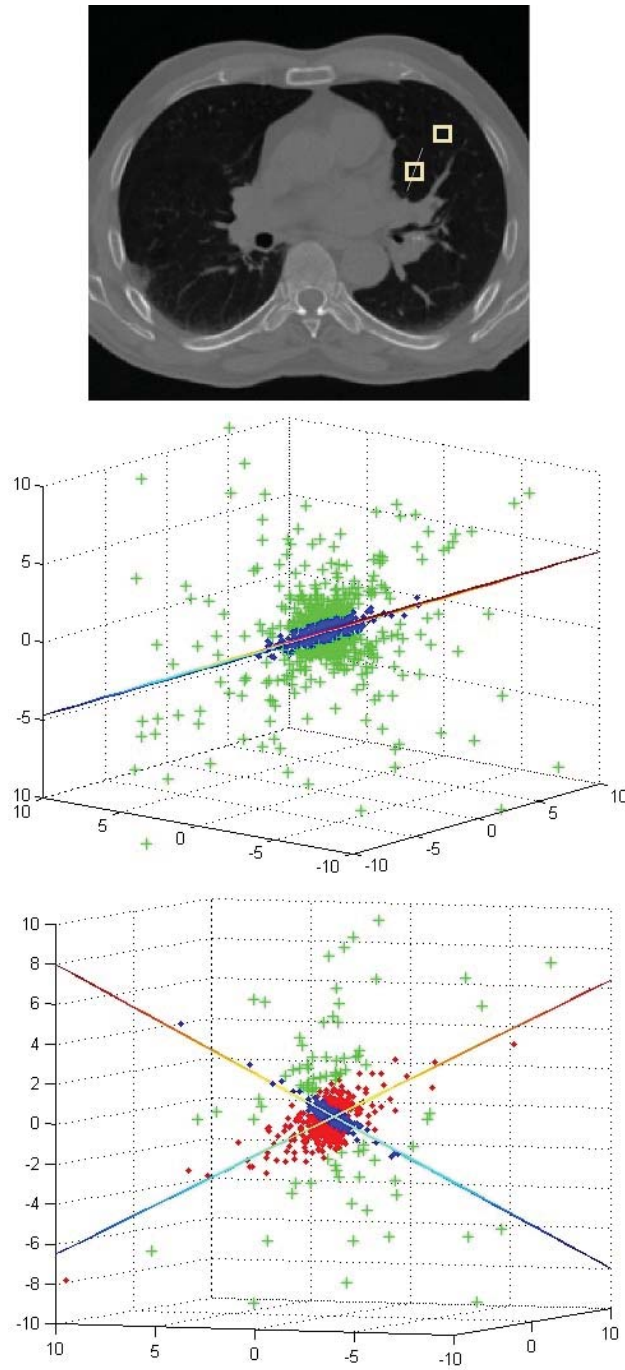


Figure 3.10: Top image is an axial view of a 4D CT image where the two cubic areas with either one (the top rectangle placed entirely on one lobe) or two (the lower rectangle straddles two lobes separated by the fissure - shown by a white line) distinct motions are marked by white rectangles. Middle and bottom images plot the associated voxels in the derivative space (red dots and green asterisks are used for moving voxels while black crosses represent the outliers). The points in both areas are segmented by fitting planes using MSSE. Middle and bottom pictures are rotated to show how the points are aligned with respect to every plane. The bottom picture clearly shows that the difference between the motions of two lobes on different sides of a fissure is detectable using a local approach.

Chapter 4

Image Registration Using Ranked Order Statistics

4.1 Introduction

Non-rigid image registration is the task of finding a transformation that spatially aligns two images with globally non-uniform differences. The non-rigid registration problem has received substantial attention in previous years and several methods to solve this problem have been proposed. Non-rigid registration methods are broadly classified into three categories based on their transformation models (Holden 2008): physical models, basis function expansions and constraints on transformation. Physical models are based on physical phenomena such as viscous fluids or elasticity and are described using partial differential equations (Christensen et al. 1994, Thirion 1998). Basis function expansions use parametric models such as B-splines (Rueckert et al. 1999), radial basis functions (Buhmann 2003) and wavelets (Wu 1995) to represent the transformation (Holden 2008). Among all these models, the B-spline transformation model is commonly used for non-rigid registration, largely due to its lower computational complexity compared to physical models.

Although the implementation of non-rigid intensity-based registration varies in practice, the intensity-based registration framework can generally be segregated into three modules (Damas et al. 2011): transformation space, similarity measure and optimization algorithm, to find a parametrized transformation by minimizing an intensity-based cost function. This is essentially a nonlinear optimization problem: Given the intensity values

of two images, namely fixed ($I_f(x)$) and moving ($I_m(x)$) images, the transform parameters vector θ is found by solving the following minimization problem (Klein et al. 2007):

$$\hat{\theta} = \arg \min_{\theta} [\Psi(\theta)] \quad (4.1.1)$$

where Ψ is the cost function (or dissimilarity measure) and $\hat{\theta}$ is the optimized vector of the transformation parameters.

To determine the optimal set of parameters that satisfies the above equation, gradient-based iterative optimization strategies are commonly used (Klein et al. 2007). However, for non-rigid registration of medical images with large numbers of voxels and transformation parameters, calculation of the cost function gradient ($\nabla_{\theta}\Psi(\theta)$) takes a significant amount of time (Bhagalia et al. 2009). There are two commonly-used techniques to reduce the computational complexity of this task. The first technique, based on the stochastic sub-sampling method introduced by Robbins & Monroe (1951), uses only a subset of voxels to estimate the model parameters (Bhagalia et al. 2009, Klein et al. 2010), while the second technique reduces the number of registration parameters by carrying out registration only on selected regions of the image (Rohde et al. 2003) which affect the registration outcome most.

In this chapter we propose a fast and accurate non-rigid registration method for images of the same modality that combines the properties of the above two techniques using a robust statistical approach. Our approach exploits the information available in the difference image, by using an order statistics-based segmentation method (Bab-Hadiashar & Suter 1999), to find the important regions for registration and use an intricate sampling scheme to target those areas and reduce the registration computation time. Our comparative experiments on registration of end-inhale end-exhale lung CT scan pairs, with expert annotated landmarks, show that the new method is faster and more accurate than the state-of-the-art sampling based techniques (Bhagalia et al. 2009, Rohde et al. 2003), particularly for registration of images with large deformations.

The rest of this chapter is organized as follows. Section 4.2 provides a brief description of previous research in this area. Section 4.3 explains the proposed registration scheme. The experimental setup and the comparative registration results of the proposed methods are presented in Section 4.4. The performance of the proposed methods in comparison with the best available techniques and the advantages and drawbacks of those techniques

are discussed in Section 4.5. Section 4.6 concludes the chapter.

4.2 Previous Research

There are two commonly-used techniques to reduce the computational cost of parametric non-rigid image registration. Rohde et al. (2003) first introduced the idea of carrying out the registration only on the regions that are not so well registered. Their method, called adaptive registration (AR), uses a hierarchical framework, where at each level a set of radial basis functions are placed on an irregular grid. The nodes of the irregular grid are the centres of the misregistered regions in the image at a particular level. Using such an irregular grid reduces the number of parameters compared to a fixed grid registration method, leading to a reduction in the computational cost. In order to identify the misregistered regions, the AR method uses an additional step, in which another set of basis functions are placed on a fixed grid and the gradient magnitudes of the cost function with respect to the parameters of the fixed grid are evaluated. The locations of the remaining grid points after eliminating those with low gradients are then considered as centres of the regions of misregistration (nodes of the irregular grid at that level).

A careful analysis of the above method reveals that it is necessary to recalculate the cost function derivatives, involving all the voxels, at the beginning of each registration step to determine misregistered regions. Since this calculation is computationally expensive, the overall registration process takes a relatively long time to complete. Our inspiration for the proposed method stems from the fact that the registration time will be significantly reduced if the misregistered regions are identified more efficiently.

The second technique to reduce computational time is to use only a subset of voxels to estimate the parameters on a fixed grid. Klein et al. (2010) were the first to advocate the use of uniform sampling for non-rigid medical image registration. Uniform sampling gives equal significance to all the voxels in an image and since the true function in this case is not distributed uniformly, using only a subset of voxels produces a biased estimate of the gradients. To reduce the sampling bias, they proposed to renew the set of chosen samples at every iteration of the optimization routine, giving equal significance to all the voxels. To improve registration accuracy Bhagalia et al. (2009) introduced the idea of using the importance sampling (IS) technique to select a subset of voxels. Importance sampling is a way of obtaining the properties of the desired distribution using another but

related distribution. In their method, they used the properties of the edge magnitude to approximate the probability distribution of the gradients.

Here we propose a new method that combines the benefits of both techniques mentioned above. Although our method focuses on misregistered parts in the image, the method used for finding them is completely different to that of the method used by AR. Here, we move away from trying to define the misregistered regions based on cost function gradients and directly use the information carried by the misregistered points to improve the optimization outcome.

We propose a new approach for the identification of misregistered points, in which the square intensity differences of fixed and transformed moving image voxels guide the identification process using a rank ordered statistics-based robust segmentation technique. In addition to improve the accuracy of registration, our method provides a computationally efficient means of stopping the registration process once the improvements become small. In practice it is very important to know precisely when to end the registration process (Damas et al. 2011). Our proposed method is described in the following section.

4.3 The Proposed Method

4.3.1 How to sample efficiently?

The main idea of our proposed method is to turn the registration focus to the regions of the image which are not so well registered. The challenge is to efficiently identify the regions of misregistration at each step of the registration process. Our contribution here is to devise an efficient method that exploits the information available in the difference image (the image intensity difference between the fixed image and the transformed moving image) to identify the regions of misregistration. We note that the sum of square difference (SSD) cost function is not significantly influenced by the small intensity differences. A very small difference between two corresponding voxels means that the voxels are either almost registered or they are both in a low textured region. Therefore, the significance of these voxels to the registration process is minimal (these voxels are referred to as group one voxels). On the other hand, very high variations are often caused by occasional differences such as misregistered vessels, organ boundaries or gross measurement errors (outliers). Although these are almost always present in every image and may be relevant for registration, their inclusion has the potential to skew the optimization and bias the

final results (those are referred to as group three voxels). To visualize how the intensity differences are distributed for a given image, the joint histogram between the fixed image and the transformed moving image during a registration process is shown in Figure 4.1. The joint histogram shows that a large number of voxels are clustered around the centre line. These are the voxels that correspond to an almost registered group of voxels (group one).

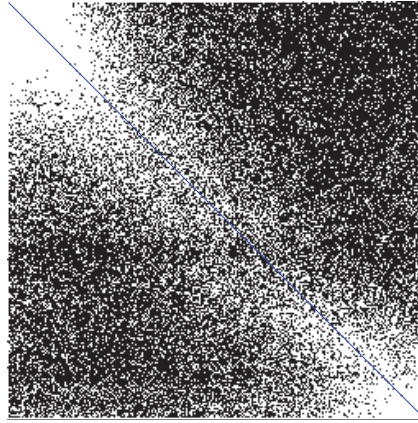


Figure 4.1: The joint histogram of the fixed image ($I_f(x)$) and the transformed moving image ($I_m(T(x))$) for case 5 of DIR-dataset at the 10^{th} iteration.

The challenge is to find computationally efficient ways of excluding the group one and three voxels and use only the voxels in the middle group (group two) for registration at every step. This is, to an extent, similar to multi-structural data segmentation and many different methods to dichotomize data based on the distribution of residuals have been developed. The emphasis of these methods is to group data points that are most similar and the analysis presented in Hoseinnezhad et al. (2010) showed that most of these methods have similar performance. Following recommendations of Hoseinnezhad et al. (2010), we used the modified selective statistical estimator (MSSE) (Bab-Hadiashar & Suter 1999) to segment voxels. The implementation of MSSE is very straightforward (computationally efficient), the estimator has nice asymptotic properties (Hoseinnezhad & Bab-Hadiashar 2007) and the value of its required parameter (i.e. the minimum size of an acceptable group) is known in this case. Having said this, we expect the use of other similar robust estimators to produce comparable results.

In our implementation of MSSE, the members of the first group are found by using the following criterion starting from the median of absolute residuals (Bab-Hadiashar &

Suter 1999):

$$|r_{i+1}| < T\sigma_i \quad (4.3.1)$$

here i is the index after sorting, $|r_{i+1}|$ is the absolute image intensity difference (residual) at index $i + 1$, T is a constant threshold (2.5 is used to include 99% of inliers based on a normal distribution (Bab-Hadiashar & Suter 1999)) and $\sigma(i)$ is the standard deviation of sorted residuals up to index i . Figure 4.2 shows the classification of different voxels based on their absolute residuals using the MSSE constraint (4.3.1).

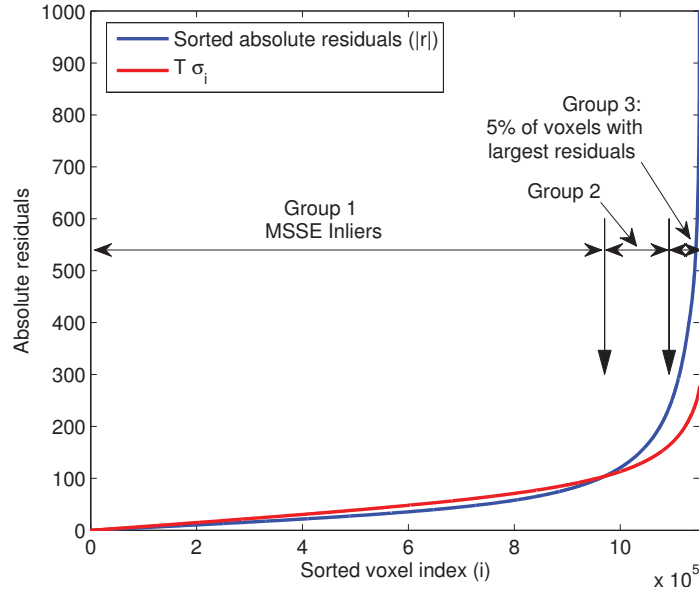


Figure 4.2: Classification of different voxels based on their absolute residuals using the MSSE constraint. The arrows show the boundaries between groups in a sample data set.

As mentioned earlier, voxels with very large differences have the potential to bias the registration process. Therefore, the last five per cent of the largest residuals are also discarded from the group identified as outliers by the MSSE. Once the voxels that belong to the second group are identified, we modify the sum of squared differences (SSD) cost function to include only voxels that belong to this group at every iteration. Since the group memberships are decided at every iteration, voxels with large differences that have the potential to improve their registration will eventually be included in the second group. The proposed cost function is written as:

$$\Psi(\theta) = \frac{1}{(i_2 - i_1)} \sum_{n=i_1}^{i_2} r_{\theta}^2(x_i) \quad (4.3.2)$$

where i_1 and i_2 are the first and last sorted indices of the squared residuals that belong to the middle group and $r_\theta^2(x_i) = (I_f(x_i) - I_m(T_\theta(x_i)))^2$. $T_\theta(x_i)$ is the B-spline transformation with parameter θ .

Having defined the cost function, the next step is to estimate the gradient of the cost function. The estimation of the gradient using all the chosen points would still be computationally expensive. Adopting sampling-based stochastic optimization techniques (Robbins & Monro 1951) enables us to perform the registration using a computationally cheaper estimate of the cost function gradients. To estimate the gradients using this approach we can either use random sampling or importance sampling techniques (Bhagalia et al. 2009). We have developed registration algorithms using both techniques and these are named robust random sampling (RRS) and robust importance sampling (RIS), respectively. Importance sampling is way of obtaining the properties of the desired distribution using another but related distribution (Hesterberg 2003). For example if X is a random variable with distribution f and $\theta(X)$ is a function of x (Hesterberg 2003):

$$E_f(\theta(X)) = \int \theta(X) f(x) dx \quad (4.3.3)$$

$$= \int \theta(X) \frac{f(x)}{g(x)} g(x) dx \quad (4.3.4)$$

Assuming $Y = \theta(X)f(x)/g(x)$, we have:

$$E_f(\theta(X)) = E_g(Y(X)) \quad (4.3.5)$$

The above shows that sampling $\theta(X)$ with distribution f gives the same expected value as sampling $Y(X)$ with distribution g . If the shape of the new sampling distribution $g(x)$ is relatively similar to the original function, the variance of the approximation will be reduced Bhagalia et al. (2009).

In the importance sampling algorithm, the gradient magnitudes of both the fixed ($|\nabla I_f(x)|$) and the transformed moving ($|\nabla I_m^\theta(x)|$) images are first used to compute the following probability distribution function (PDF):

$$P_s^\theta(i) = \frac{e_i^\theta}{\sum_{j=1}^n e_j^\theta}; i = 1, \dots, n \quad (4.3.6)$$

where $e_i^\theta = \frac{|\nabla I_f(i)|}{\sum_{j=1}^N |\nabla I_f(j)|} + \frac{|\nabla I_m^\theta(i)|}{\sum_{j=1}^N |\nabla I_m^\theta(j)|}$. The inverse cumulative distribution function (CDF) sampling (Devroye 1986) is then used to choose the appropriate voxels for gradient estimation. To explain how the above method works, we first note that when a random variable X has a cumulative distribution function F , then the values of $F(X)$ are uniformly distributed in the range $[0, 1]$. As a result, a set of uniformly distributed samples of the calculated probabilities can be used to generate samples of the fixed image that have the desired distribution P .

4.3.2 Convergence criterion

Another important challenge in successful implementation of stochastic optimization schemes is to find an appropriate stopping criteria. Simple criterion used for gradient descent algorithms such as the magnitude of the gradient or the change in cost function are inappropriate for stochastic optimization, as the gradient approximate is noisy and may not vanish near the optimal solution. To overcome this limitation, most practical algorithms simply use a fixed number of iterations (Bhagalia et al. 2009, Klein et al. 2007, Rohde et al. 2003). In practice this is not an appropriate solution and the required number of iterations is not known a priori. Fixed numbers of iterations for all cases would either lead to inadequate accuracy or waste of computation resources.

An important advantage of the proposed algorithm is that the degree of convergence of the registration can be directly measured by looking at the number of voxels identified by the MSSE as inliers (the voxels that are almost registered). The increase of voxels in this group at every iteration is proportional to the increase in the registration accuracy. To take advantage of this characteristic, we devised a stopping criterion by developing a normalized measure of this attribute. The measure defined is as follows:

$$C_k = \frac{N_k - N_{k-1}}{N_{k-1}} \times 100\% \quad (4.3.7)$$

where N_k denotes the number of voxels in the inlier group at iteration k . The registration is then deemed successful once C_k is less than a specific threshold (0.1% in all of our experiments).

An added advantage of this approach is that the computational cost associated with calculating this criterion is negligible. A complete description of the proposed registration algorithms is provided in Algorithm 2.

Algorithm 2 Step-by-Step Algorithm of Proposed Methods**Inputs:** Fixed image, moving image, transform parameters

- 1: Apply the transform to the moving image and recalculate the intensity values at grid points (transformed image).
- 2: Calculate the absolute differences between the fixed and the transformed images ($|r|$).
- 3: Sort the $|r|$ values and keep their orders (ascending).
- 4: Find K , K is the first index of the sorted residuals where $|r_{i+1}| < T\sigma_i$ is not satisfied.
- 5: Find $K_L = 0.95 \times \text{Total number of voxels}$.
- 6: **if** sampling scheme = RRS **then**
- 7: Randomly select N voxels starting from the K^{th} position in the sorted array up to the K_L^{th} position as the samples .
- 8: **else if** sampling scheme = RIS **then**
- 9: Select N voxels starting from the K^{th} position in the sorted array up to the K_L^{th} position using inverse CDF (CDF of the edge magnitudes of the fixed and transformed moving images) sampling, as the samples.
- 10: **end if**
- 11: **if** ($C_k \leq \text{THRESHOLD}$) **then** Stop Optimization

4.4 Analysis of the Proposed Algorithms

4.4.1 Experimental set-up

To compare the performance of the proposed registration methods with the best available techniques, an extensive set of experiments using lung CT images were conducted. The elastix registration toolbox (Klein et al. 2010) based on the Insight Segmentation and Registration Toolkit (ITK) was used as the basis of the registration framework that was developed to implement both the importance sampling and the proposed robust sampling methods. The code provided by the first author of Rohde et al. (2003) was used to obtain the results for the AR method. All experiments were conducted using an HP Z400 workstation with a single Intel Xeon W3550 3.06GHz processor.

Two CT datasets were used to estimate the registration accuracy of the two proposed methods and to compare these with the state-of-the-art methods. The first was the CT dataset used by Bhagalia et al. (2009) (referred to as the IS-dataset). This dataset consists of 8 pairs (breath-hold) of maximum inhale to maximum exhale CT images with $1.87 \times 1.87 \times 5.0 \text{ mm}^3$ voxels (Coselmon et al. 2004). The second data set, referred to as the DIR-dataset, was provided by the University of Texas MD Anderson Cancer Center (Castillo, Castillo, Guerra, Johnson, McPhail, Garg & Guerrero 2009). This dataset consists of 4D-CT images of ten different patients each consisting of a sequence of images taken

through a full respiratory cycle. In our experiments we registered the maximum inhale and maximum exhale images in each case. Every case of the 4D-CT dataset has 300 expert identified landmarks and their associations at full inhale and exhale images. In our experiments all images were cropped and segmented to include only the lungs. The segmentation was done automatically using ITK-SNAP (Yushkevich et al. 2006) software.

4.4.2 Implementation

The first step in our registration framework is to implement an appropriate hierarchical scheme to ensure that the registration process is not trapped in a local minimum. To make our results comparable with those published in Bhagalia et al. (2009), we also implemented a two-level Gaussian pyramid scheme (Lester & Arridge 1999). In this scheme, the amount of data in the initial level of the registration process is down-sampled by a factor of two and smoothed using a Gaussian kernel $\mathcal{N}(0, 1)$, while in the second level, the original image is used.

For the transformation model the cubic B-spline described by the following equation (Kybic & Unser 2003) was used:

$$T_\theta(x) = x + \sum_{x_k \in N_x} \theta_k \beta^3\left(\frac{x - x_k}{s}\right) \quad (4.4.1)$$

where θ_k is the k^{th} parameter, $\beta^3(x)$ is the 3rd order B-spline polynomial given by Unser (1999):

$$\beta^3(x) = \begin{cases} \frac{2}{3} - |x|^2 + \frac{|x|^3}{3} & 0 \leq |x| \leq 1 \\ \frac{(2-|x|)^3}{6} & 1 \leq |x| \leq 2 \\ 0 & 2 \leq |x| \end{cases} \quad (4.4.2)$$

and N_x is the set of all control points (x_k) within the compact support of the B-spline at point x . The control points for each level were organized in a fixed grid with voxel spacing of $16 \times 16 \times 8$ for the first level and $8 \times 8 \times 4$ for the second level. In order to optimize the cost function, the stochastic gradient descent described by the following equation was used:

$$\theta_{k+1} = \theta_k + a_k \hat{g}(\theta_k) \quad (4.4.3)$$

where a_k is a vector that controls the step size in different directions at every iteration of the algorithm and $\hat{g}(\theta_k)$ is an appropriate estimate of the gradient. For the stochastic

optimization solution to converge, the step size a_k should satisfy the following conditions: $a_k \geq 0$; $a_k \rightarrow 0$ as $k \rightarrow \infty$; $\sum_{k=0}^{\infty} a_k = \infty$; $\sum_{k=0}^{\infty} a_k^2 < \infty$. A common way to ensure that the step size satisfies the above criterion is to use the following step size formula (Spall 1998):

$$a_k = \frac{a_0}{(A + k)^\alpha} \quad (4.4.4)$$

where k is the iteration number and a_0 and A are empirically determined constants. However, for problems with large numbers of parameters, the optimal step size for every parameter may vary widely and therefore a specific step size for each parameter needs to be calculated. In this case, to ensure convergence, the following equation is commonly used to calculate the step size for each parameter (Kesten 1958):

$$a_k^i = \frac{a_0}{(A + Q_k^i)} \quad (4.4.5)$$

where Q_k^i is the number of sign changes in the i^{th} parameter up to the k^{th} iteration.

Another commonly used method to ensure convergence in stochastic approximations is to increase the sample size progressively during the iterations (Dupuis & Simha 1991). In our work we have adopted a combination of these two methods, as advocated by Bhagalia et al. (2009). Using the combination scheme not only ensures the convergence of our proposed methods but also makes the final results directly comparable with Bhagalia et al. (2009). There are four key parameters to be selected in this optimization scheme: number of iterations, A , a_0 and number of samples in each iteration.

All competing registration methods, in contrast to those proposed here, need a predefined measure to stop the registration process. For these methods the maximum number of iterations at each level of hierarchy was set to 60 and 100, respectively. These represent the best case scenario (in terms of timely registration with the best accuracy) for those methods determined by manual inspection of mean landmark errors in registered test images. In practice, this information is not known for each image and can significantly increase the overall computation time required for those registration methods. To make all the results comparable, in all our experiments involving sampling-based methods, we used 4096 and 8192 samples at the sequential registration levels. We observed that the registration accuracy does not change significantly with small variations of the number of samples. The parameters A and a_0 are manually tuned to achieve the overall

best registration outcomes. Following Bhagalia et al. (2009), the value of A was set to 10 and the mean landmark errors for different values of a_0 were calculated. Overall, the $a_0 = 500$ appeared to be the most suitable, and this value was kept constant for all the registration methods in order to make the results comparable. Our experiments showed that the overall result is not significantly affected by the small variations of a_0 (increased or decreased by two or three times).

In MSSE implementation the basic quick-sort algorithm was used in our experiments and the timing of different runs showed that on average around 15% of the total computation time was spent on the sorting step.

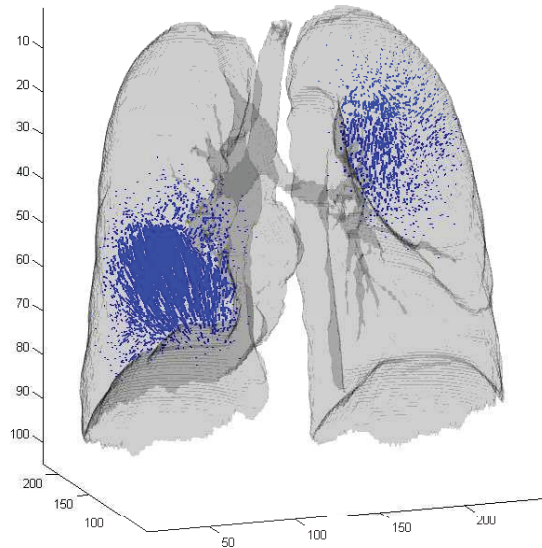
4.4.3 Analysis with synthetic images

To demonstrate the effectiveness of the proposed algorithm in identifying the mis-registered regions, a simulation study involving a pair of synthetic images (constructed by deforming a lung CT image with known deformations and adding normal noise) was conducted. The second image was designed to exhibit localized motions: changes are confined to two specific regions of the image.

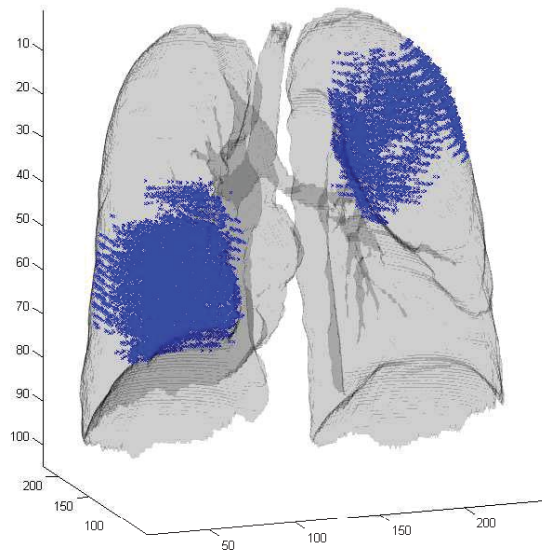
The above images were then registered using the proposed method and the voxels classified as the second group (used in the actual registration process) were separated and plotted in Figure 4.3(b) together with the surface rendering of the lung. To show the correspondence between actual motion and identified voxels, the actual motion field together with the surface rendering of the lung image is shown in Figure 4.3(a). These images show that the proposed method is able to correctly identify misregistered voxels. The proposed methods will in turn place more emphasis on these voxels during the registration process.

4.4.4 Performance variation with noise

The impact of noise on the proposed registration algorithm was studied by creating images with known deformations and additive normal noise. In those simulations a lung CT image was first deformed using a predefined B-spline transform that was calculated in a real image experiment using the DIR dataset. The intensities of both fixed and deformed images were normalized between zero and one. Zero mean normally-distributed noise with standard deviation ranging from 0.0001 to 0.05 was added to both images. It is important to note that, since the lung densities cover only around 25% of the entire range



(a)



(b)

Figure 4.3: The results of the simulation demonstrate the effectiveness of the proposed algorithm in identifying the misregistered regions (a) The predefined motion field together with the surface rendering of the lung. (b) Plot of all voxels classified as the second group in the simulation.

of Hounsfield values in the CT images, the maximum noise of 0.05, which is added to the entire range of normalized density values of both images, represents a substantial amount of actual noise in the lung regions. The noisy images were then registered with the proposed algorithms. Finally, registration errors for each image pair were calculated using a set of randomly-distributed points and averaged over five repeated experiments. The results of those experiments are shown in Figure 4.4. To show the actual amount of added noise, the registration errors are plotted in terms of the variance of added noise in Hounsfield units. The figure shows that for relatively large amounts of additive noise (compared to image intensities), the registration error remains unaffected.

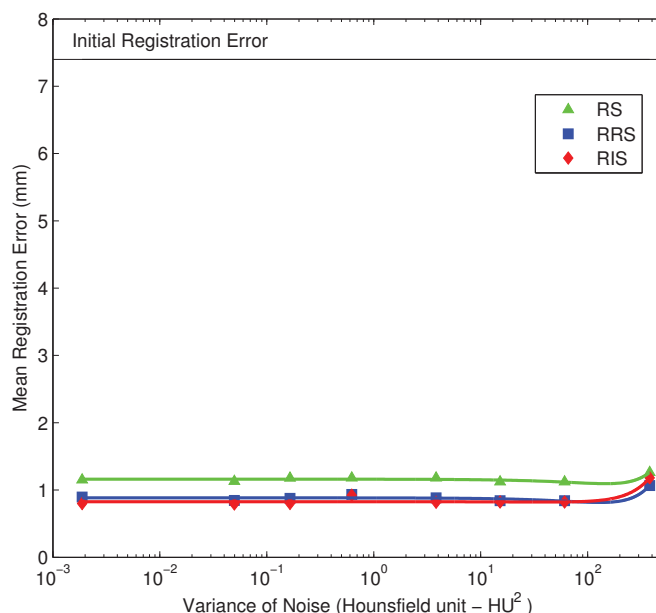


Figure 4.4: Performance variation of registration error of the proposed algorithms versus the variance of manually added normal noise. The RS method is also included for the sake of comparison.

4.4.5 Experimental results

The first set of experiments was conducted using the IS-dataset. The registration accuracy of the proposed methods in comparison with the competing methods using the above dataset is shown in Table 4.1. The result shows that the proposed RIS method always achieves better results than the IS and AR. Following the recommendations by Demšar (2006) on statistical comparisons of classifiers over multiple datasets, the final mean registration errors achieved using different algorithms were compared with the RIS method

using the Wilcoxon rank-sum test (Wilcoxon 1945). The p-values, after Holms' (Holm 1979) correction, are given in Table 4.3. The results show that RIS produces significantly better results than the competing methods.

Table 4.1: Comparison of mean landmark error (Euclidean distance) for IS-dataset measured using expertly identified landmark points.

		IE¹	AR²	RS³	IS⁴	RRS⁵	RIS⁶
IS-C1	Mean	15.1	4.5	4.39	3.69	3.61	3.39
	Std		2.32	1.28	2.04	2.10	2.14
	Min/Max			4.39/4.86	3.55/3.81	3.37/3.79	2.98/3.88
IS-C2	Mean	14.52	6.49	4.53	2.56	2.27	2.51
	Std		9.81	3.64	3.00	2.67	2.38
	Min/Max			4.34/4.65	2.11/3.05	1.59/2.79	2.48/2.58
IS-C3	Mean	13.31	4.57	2.74	2.62	2.31	1.77
	Std		3.50	1.47	2.60	1.94	2.33
	Min/Max			2.38/3.30	2.45/2.76	1.90/2.94	1.64/2.03
IS-C4	Mean	11.73	1.69	3.21	2.46	2.01	1.73
	Std		0.89	2.70	2.02	1.46	1.10
	Min/Max			2.52/3.96	2.26/2.81	1.69/2.39	1.38/2.00
IS-C5	Mean	9.13	2.66	3.46	3.24	2.14	1.62
	Std		2.08	2.30	2.29	1.70	2.02
	Min/Max			3.35/3.60	3.18/3.29	1.77/2.45	1.20/1.82
IS-C6	Mean	8.62	3.18	1.16	0.94	0.96	0.86
	Std		2.23	1.30	0.99	1.17	1.19
	Min/Max			1.06/1.20	0.62/1.25	0.62/1.19	0.75/1.06
IS-C7	Mean	7.77	1.82	2.43	1.86	1.86	1.39
	Std		1.95	2.10	2.05	2.05	1.61
	Min/Max			1.51/3.18	1.46/2.35	1.46/2.35	0.94/1.77
IS-C8	Mean	6.89	3.29	2.34	2.70	2.73	2.32
	Std		2.35	1.93	2.17	2.33	1.96
	Min/Max			2.08/2.79	2.53/2.79	2.48/2.92	2.15/2.47
Overall	Mean	10.88	3.53	3.03	2.51	2.24	1.95

¹ Initial error without any registration

² Adaptive registration

³ Random sampling

⁴ Importance sampling

⁵ Robust random sampling

⁶ Robust importance sampling

The limitation of the IS-dataset is that it has only six annotated landmarks. To provide more compelling evidence, we also conducted experiments using the DIR-dataset (Castillo, Castillo, Guerra, Johnson, McPhail, Garg & Guerrero 2009), which has 300 annotated landmarks. Those landmarks were used to evaluate the registration performance of all methods based on their mean landmark errors. The results of these experiments

are shown in Table 4.2. The table also includes the standard deviation of the landmark errors and the maximum/minimum of the mean errors of ten repeated experiments. A Wilcoxon-based statistical analysis similar to the IS-dataset was performed on these results as well and the results are given in Table 4.3. The results show that the proposed RIS algorithm performs significantly better than the competing algorithms. Table 4.3 also shows the results for the Wilcoxon test using all 18 test images, which are consistent with the previous results.

To compare the performance of RIS in comparison with the competing algorithms for each image of the DIR-dataset, a Wilcoxon rank-sum test (Wilcoxon 1945) was used (this test uses the error at each landmark point averaged over ten repeated runs). The p-values after correcting for multiplicity using Holms' procedure are given in Table 4.4. These results show that the proposed RIS method in comparison to all competing methods (AR, RS, IS) produces significantly better results in cases with large deformations (DIR-C4 to C8). It is important to note that the first few cases include very small motions and all methods achieve very good registration results.

To compare the computational complexity of these methods, the average computation times of all competing methods for registration of all images in both datasets were measured and the results are shown in Table 4.3. We note that both the RS and IS methods use a fixed number of iterations as the stopping criterion and their computation times are heavily dependent on the number of required iterations, which is considered an input. The experiments showed that the computation times of our methods are significantly lower than those of the competing methods. The registration time for the AR method is significantly higher than other competing methods. This point is shown in Table 4.3 and Figure 4.5 (a), where the mean errors of all the algorithms (RS, IS, RIS, RRS and AR) are compared against the time taken to achieve that level of accuracy for a sample image (case 1). We also note that the proposed RRS method is more than twice as fast as the IS in both datasets.

Figure 4.5 (b)-(d) show the mean errors of all the sampling-based algorithms with the number of iterations taken to achieve that level of accuracy. These figures show that both proposed methods reduce the mean error faster than RS and IS. We also observed in our experiments that there are fewer sign changes in the RIS estimated gradient of the cost function compared to RS and IS. This implies that the step-size reductions in RIS

Table 4.2: Registration performance for images in DIR-dataset measured using expertly identified landmark points. Each algorithm is repeated 10 times to account for the randomness.

		IE	OE	AR	RS	IS	RRS	RIS
DIR-C1	Mean	4.01	0.85	0.98	1.11	1.03	1.00	0.94
	Std	2.91	1.24	1.51	1.01	1.00	0.99	0.96
	Min				1.06	0.99	0.96	0.90
	Max				1.14	1.12	1.03	0.99
DIR-C2	Mean	4.65	0.70	1.10	1.15	0.97	0.92	0.88
	Std	4.09	0.99	1.87	1.23	1.06	0.97	0.97
	Min				1.11	0.9	0.85	0.84
	Max				1.25	1.00	0.98	0.96
DIR-C3	Mean	6.73	0.77	1.26	1.39	1.16	1.08	1.03
	Std	4.21	1.01	1.68	1.25	1.15	1.10	1.08
	Min				1.32	1.11	1.04	0.96
	Max				1.48	1.21	1.12	1.10
DIR-C4	Mean	9.42	1.13	1.93	1.84	1.68	1.64	1.55
	Std	4.81	1.27	2.24	1.51	1.32	1.33	1.26
	Min				1.77	1.62	1.61	1.48
	Max				1.89	1.74	1.69	1.64
DIR-C5	Mean	7.10	0.92	3.06	2.91	2.38	2.12	1.70
	Std	5.14	1.16	4.50	3.03	2.66	2.29	1.68
	Min				2.70	2.26	2.02	1.59
	Max				3.09	2.56	2.23	1.84
DIR-C6	Mean	11.10	0.97	5.27	2.05	1.73	1.65	1.58
	Std	6.98	1.38	8.29	1.70	1.26	1.21	1.19
	Min				1.95	1.66	1.54	1.48
	Max				2.22	1.85	1.70	1.71
DIR-C7	Mean	11.59	0.81	4.16	3.31	2.31	2.16	1.71
	Std	7.87	1.32	6.03	3.40	2.35	2.08	1.27
	Min				3.03	2.14	1.95	1.63
	Max				3.55	2.58	2.58	1.81
DIR-C8	Mean	15.16	1.03	5.93	3.79	2.42	2.21	1.76
	Std	9.11	2.19	7.12	4.54	2.74	2.55	1.74
	Min				3.33	2.27	1.96	1.64
	Max				4.05	2.62	2.49	1.86
DIR-C9	Mean	7.82	0.75	2.85	1.51	1.48	1.47	1.43
	Std	3.99	1.09	2.64	1.22	1.23	1.17	1.09
	Min				1.47	1.42	1.41	1.34
	Max				1.57	1.51	1.54	1.53
DIR-C10	Mean	7.63	0.86	3.26	1.89	1.64	1.65	1.74
	Std	6.54	1.45	5.40	1.84	1.46	1.55	2.03
	Min				1.79	1.53	1.5	1.54
	Max				1.93	1.71	1.73	2.13

Table 4.3: Average running time for each algorithm and the Holms' corrected p-values of Wilcoxon rank-sum test that compare the final mean registration error of competing methods with the RIS method.

	IS-dataset		DIR-dataset		Combined
	Time (s)	p-value	Time (s)	p-value	p-values
AR	348.0	0.03	1,757.0	7.81E-03	7.00E-04
RS	55.0	0.03	24.8	7.81E-03	7.86E-04
IS	112.0	0.03	100.0	0.03	1.07E-03
RRS	43.0	0.04	41.5	0.03	2.47E-03
RIS	65.0	-	86.5	-	-

Table 4.4: Holms' adjusted p-values for the Wilcoxon rank-sum test conducted between the landmark error of RIS and other competing methods.

	AR	RS	IS	RRS
DIR-C1	0.34	1.36E-04	0.04	0.39
DIR-C2	0.07	1.64E-05	0.10	0.14
DIR-C3	0.10	9.54E-09	4.00E-04	0.03
DIR-C4	0.05	1.56E-04	1.37E-03	0.04
DIR-C5	1.46E-05	5.41E-16	7.61E-08	2.43E-04
DIR-C6	8.48E-09	1.09E-07	3.00E-04	0.17
DIR-C7	3.66E-08	7.65E-20	3.35E-08	1.09E-09
DIR-C8	2.47E-19	1.64E-08	1.37E-04	0.01
DIR-C9	3.67E-14	0.50	0.65	0.61
DIR-C10	2.27E-04	0.18	0.73	0.26

are slower, which contributes to its faster convergence. The graphs also show that the convergence criterion described in Section 4.3.2 is successful in stopping the registration process at appropriate points (stable results).

To further analyse the performance of the proposed convergence criterion, we compared the accuracy and computation time of the proposed RIS method with and without the proposed convergence criterion. Here, without convergence criterion refers to RIS registration with a fixed number of iterations (similar to IS and RS registrations). The results are shown in Table 4.5. The p-values (Holms corrected) of Table 4.5 are the results of a two one-sided test (TOST) between the final landmark errors achieved with and without the convergence criterion. These numbers show that the mean landmark errors achieved with and without the proposed convergence criterion are equivalent (within a margin of 5% of the initial landmark error). Calculating the average of the last column of Table 4.5 shows that an average time saving of around 44% is achieved by using the proposed convergence criterion.

Table 4.5: The improvement of the registration time due to the use of convergence criterion and the results of the TOST (low p-values mean the results are statistically equal).

	Mean LM Error With CC	Mean LM Error Without CC	TOST p-value	Time Without CC	Time saved %
DIR-C1	0.94	0.96	5.6E-04	85.3	13.4
DIR-C2	0.88	0.93	1.3E-03	123.6	43.9
DIR-C3	1.03	1.11	2.3E-04	119.2	64.7
DIR-C4	1.55	1.60	2.4E-07	115.4	34.0
DIR-C5	1.70	1.62	6.7E-04	124.3	62.1
DIR-C6	1.58	1.66	5.1E-10	142.7	64.0
DIR-C7	1.71	1.79	5.1E-10	138.7	87.7
DIR-C8	1.88	1.85	5.3E-14	153.4	20.9
DIR-C9	1.43	1.45	9.8E-07	99.23	40.6
DIR-C10	1.74	1.74	4.4E-05	124	12.3

To show that there is a linear correlation between the proposed convergence criterion C_k and the registration accuracy throughout the registration process, the Pearson's correlation coefficients between the actual registration accuracy measured using mean landmark error and C_k at the end of each iteration for the top level are enumerated in Table 4.6. The results show that there is a strong linear correlation between the measure for the convergence criterion and the actual registration error.

Table 4.6: The measure of correlation between the proposed convergence criterion and the actual registration error measured using expert identified landmarks. A value in excess of 0.5 for Pearson's correlation indicates a strong linear correlation. Images from the DIR dataset are used.

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
0.93	0.96	0.76	0.86	0.8	0.89	0.7	0.9	0.87	0.82

To visualize the registration quality of the proposed methods, the intensity difference image of three slices before and after RIS registration are shown in Figure 4.6. These slices show that the registration errors reduce significantly after the registration with the proposed method.

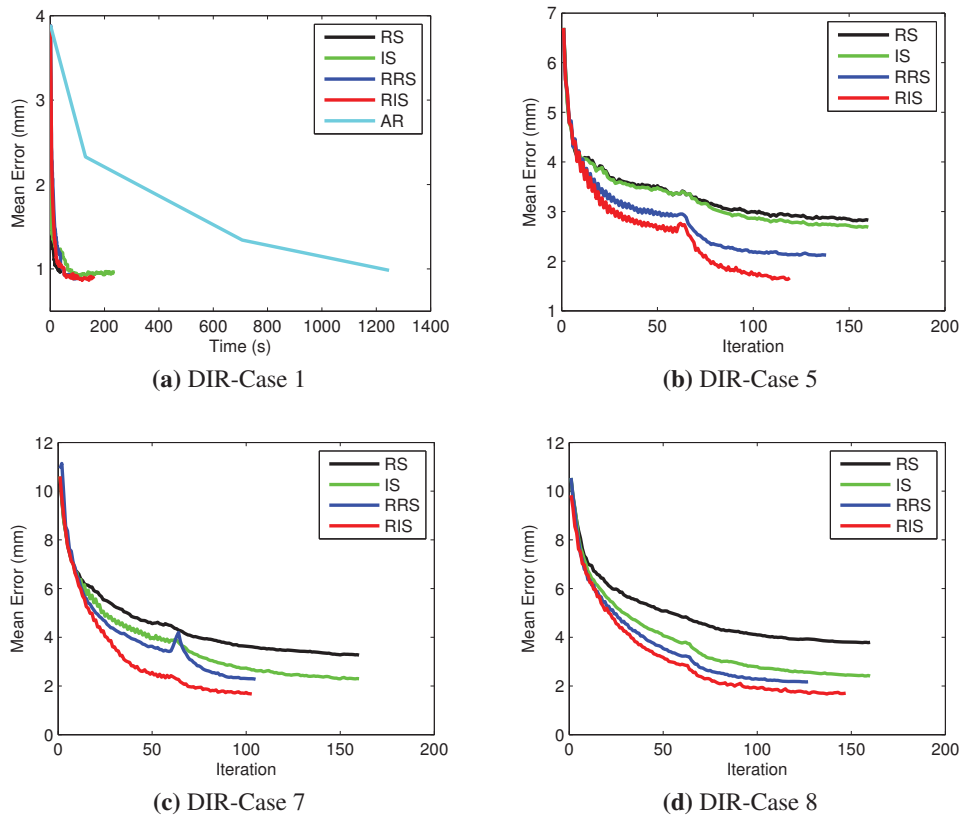


Figure 4.5: (a) Rate of the mean landmark error reduction over time for different registration methods for Case 1. (b)-(d) Rate of decrease of mean landmark error vs number of required iterations for different methods in cases where registration involves significant deformations.

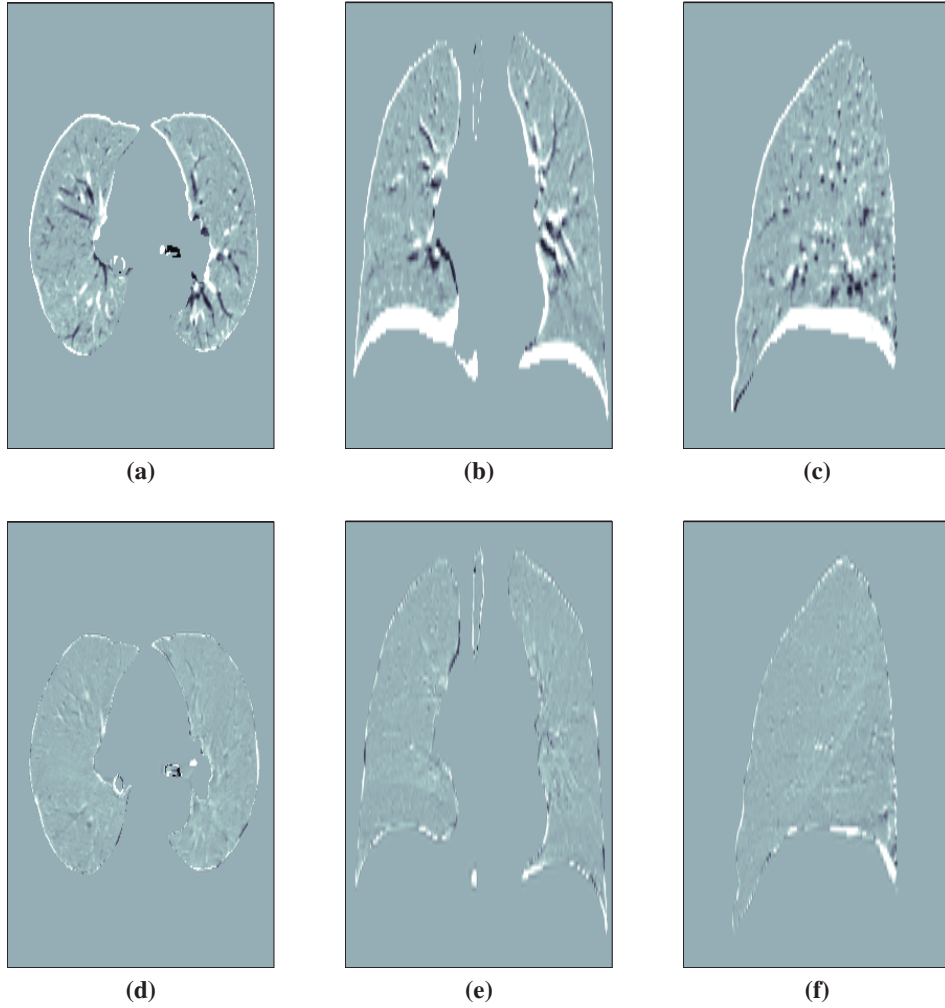


Figure 4.6: Axial, sagittal and coronal slices of the intensity difference image between fixed and moving images (DIR-Case 5) before registration are shown in (a),(b),(c), respectively. (d),(e) and (f) show the axial, sagittal and coronal slices of intensity differences between fixed and transformed moving images after RIS registration.

4.5 Discussion

This chapter has presented a new approach to focus the registration on misregistered points identified using a rank-ordered statistics-based robust segmentation technique. Experiments using lung CT images showed that the proposed method achieved high registration accuracies compared to similar methods. The registration time for the proposed method was also significantly reduced using a new convergence criterion.

It is not surprising to see that the RS method, due to its simplicity, is computationally more efficient than all of the above methods. However, the accuracy of the random sampling method is poor. Our experiments showed that running the RS with more itera-

tions would not improve its registration accuracy. To provide an example, the mean errors of Case five using the random sampling algorithm with significantly higher numbers of iterations (100 and 400 for two registration levels compared to 60 and 100 originally) are shown in Figure 4.7. This figure shows that the accuracy of the RS method does not increase by increasing the number of iterations. This is due to the fact that in stochastic optimization, to ensure convergence, the step sizes are incrementally reduced and when the step size is very small, no significant improvement is gained from additional iterations. A two one-sided test (TOST) was conducted with the null hypothesis: mean registration errors of RS with low and high numbers of iterations are not equal (the margin is set to 5% of the error before registration). The p-values after correcting for multiplicity using Holms' procedure were also calculated and the results showed that the accuracy of the RS does not improve significantly as the number of iterations increases. The RS algorithm was also tested with three times more samples per iteration. Again, the RS final mean error did not significantly change by increasing the sample size.

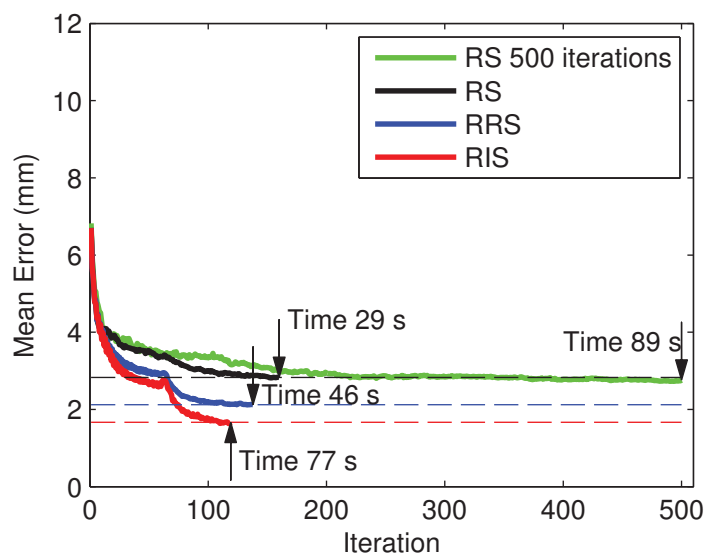


Figure 4.7: Mean error of each iteration for algorithms RIS, RRS and RS. The maximum number of iterations for RS has been increased to 100 and 400 in the two registration levels respectively. The image used is DIR-Case 5.

To show the effect of using rank-order statistics-based segmentation during registration, joint histograms of fixed and moving images before and after the application of MSSE are shown in Figure 4.8. The figure shows that the segmentation strategy has been successful in removing the high density clusters that appeared in the joint histogram and focusing the optimization on the target voxels.

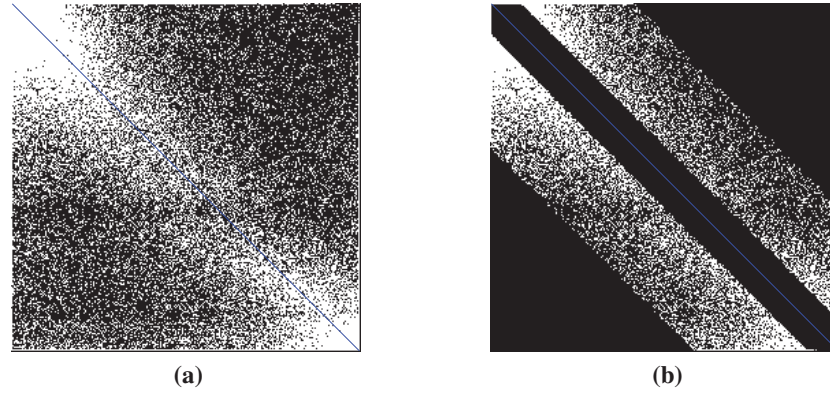


Figure 4.8: The joint histogram of the fixed image ($I_f(x)$) and the transformed moving image ($I_m(T(x))$) for Case 5 of DIR-dataset at the 10th iteration (a) contains all the voxels. (b) contains only the voxels identified as the middle group by the MSSE.

To test the hypothesis that misregistered regions can be identified using robust segmentation of voxels in the intensity difference image, the median landmark errors were calculated for the landmarks that either belong to inliers (Group one) or outliers (Groups two and three). The median is robust to the influence of outliers and presents a rigorous performance measure. The median landmark errors, presented in Table 4.7, show that landmark errors of Group one are significantly lower than those of Groups two and three.

Table 4.7: Median of the landmark errors belonging to Group one and Groups two and three.

	Median Inlier Error	Median Outlier Error
DIR-C1	1.94	2.68
DIR-C2	2.76	4.12
DIR-C3	2.98	3.45
DIR-C4	3.39	5.13
DIR-C5	3.33	5.18
DIR-C6	4.00	13.12
DIR-C7	5.01	12.50
DIR-C8	6.40	7.93
DIR-C9	3.71	5.09
DIR-C10	2.85	4.62

Registration results for images in the DIR-dataset are reported at www.dir-lab.com and according to these results the best registration is achieved with least median of square filtered compressible flow (LFC) (Castillo et al. 2012) and 4-dimensional local trajectory modeling (4DLTM) (Castillo et al. 2010). Comparison of these results shows that the

above methods achieved accuracies comparable to or better than the proposed methods. It should however be noted that the errors of LFC and 4DLTM cannot be directly compared with the errors we computed for RIS and RRS, as those reported in the chapter were on 1200 landmarks per image, while a subset of 300 landmarks per image was available publicly. It is important to mention that the 4DLTM method uses all the images from the 4D-CT dataset, whereas the proposed methods use only the two images corresponding to the extremes of the respiratory cycle. The use of intermediate images by the 4DLTM method means that the algorithm has to process significantly more data than methods using only two extreme images. The execution time of the LFC method (reported in Castillo et al. (2012)) is an indication of the large amount of computation required in this method. The above methods were specifically developed to register lung CT images and use more information than the proposed methods, making them computationally expensive.

A drawback of the proposed methods is that, since these methods rely on the intensity difference for sample selection, the change of intensity with a change in inspiration level in case of large deformations can affect the performance of these methods. Other authors have shown the benefit of using a mass-preserving intensity model that adjusts intensity according to volume changes (Gorbunova et al. 2012). Such a model can be incorporated in the proposed approach and may improve results over those presented here. However, extension to multimodal registration is less straightforward. In such cases, one would need to find appropriate clustering schemes to remove the almost registered data from the images of different modality for the application of the proposed registration methods.

Several recent registration algorithms have been implemented using graphics processing units (GPUs) to accelerate computation (Castillo et al. 2012, Eklund et al. 2013, Fluck et al. 2011, Gu et al. 2010). An algorithm has to be parallelizable in order to take advantage of these computers. In our implementation we use a B-spline transformation to represent the deformation field and the optimization is performed using analytical gradients of the SSD cost function with respect to the B-spline coefficients. In our approach the computation time is reduced by using only a carefully selected subset of voxels in the calculations. Another method to decrease the computation time would be to use a GPU to calculate the gradients. This is because the B-spline coefficients have only a compact support and the memory on the GPU is shared by all the cores. The derivative calculation step can therefore be easily parallelized, as illustrated by Rohlfing & Maurer Jr (2003).

An issue for the GPU implementation of this method is that the sorting step of the proposed methods cannot be easily parallelized. Having said this, we have observed that the sorting step on average takes only around 15% of the overall computation time, even when the basic “qsort” algorithm is used for sorting. The total computation time and the time for sorting for each image in DIR-dataset are shown in Table 4.8.

Table 4.8: The total computation time and the time for sorting for each image in DIR-dataset.

	Total Mean Time Per Iteration (ms)		Mean Sort Time Per Iteration (ms)		Sort Time / Total Time) %	
	Level 1	Level 2	Level 1	Level 2	Level 1	Level 2
DIR-C1	201.78	732.46	17.53	73.96	8.69	10.1
DIR-C2	330.63	1037.56	28.08	220.77	8.49	21.28
DIR-C3	225.52	1056.83	21.52	178.75	9.54	16.91
DIR-C4	279.3	987.03	15.5	124.82	5.55	12.65
DIR-C5	293.92	1066.75	19.47	154.29	6.62	14.46
DIR-C6	361.05	1210.64	20.4	157.13	5.65	12.98
DIR-C7	351.27	1176.73	27.25	213.24	7.76	18.12
DIR-C8	384.02	1304.08	39.2	311	10.21	23.85
DIR-C9	254.45	840.86	15.75	125.72	6.19	14.95
DIR-C10	278.9	1072.36	24.47	191.92	8.77	17.9
Average	296.08	1048.53	22.92	175.16	7.75	16.32

4.6 Conclusion

This chapter has presented a new approach to the non-rigid registration of medical images based on using robust segmentation of squared intensity differences to intelligently guide the sampling part of the stochastic optimization. The performance of the proposed approach in terms of registration accuracy and computation time was compared with several existing methods via an extensive set of experiments on images with ground truth. The experiments showed that the proposed method substantially improves both the accuracy and the computational complexity of the registration task. More importantly, the method incorporates a computationally-efficient means of measuring the quality of registration. Since the registration schemes are all iterative, this measure is highly useful for deciding when to stop the registration process.

Chapter 5

Robust Data Modelling Using Thin Plate Splines

5.1 Introduction

Fitting a model to a set of data contaminated with noise is a well-studied problem and has many applications in the field of computer vision. These applications include: interpolation of missing data, smoothing of motion fields and estimation of derivatives. Modelling using thin plate splines (TPSs) is one of the most common methods of scattered data fitting. TPSs are commonly used in image registration applications to model transformation in feature-based registration algorithms (Bookstein 1999, Coselman et al. 2004, Johnson & Christensen 2001, Zhang et al. 2014) and in smoothing of motion fields (Ehrhardt & Lorenz 2013). It is also used in motion estimation in tomographic reconstruction (Muller et al. 2013).

TPS modeling, as formulated by Meinguet (1979) and Wahba (1979), involves the minimisation of the following cost function (Stals & Roberts 2006):

$$\frac{1}{n} \sum_{i=1}^n (\|f(x_i) - y_i\|)^2 + \lambda \int_{\Omega} \sum_{|v|=2} \binom{2}{v} (D^v f(x))^2 dx \quad (5.1.1)$$

where $x_i \in R^d$ is the location of the i^{th} data point in the d-dimensional space, y_i is the corresponding data, n is the number of data points and v is a d-dimensional multi-index.

The first term in this equation is referred to as the data fidelity term and the second term as the smoothness term. The smoothing parameter λ controls the trade-off between

data fidelity and smoothness. When λ is very small, the resultant function tends to go through all the data points and is highly sensitive to noise. On the other hand, when λ is high, the result will be very smooth and may not follow the curves of the underlying data. The selection of an appropriate smoothing parameter is not a trivial problem. However, solutions based on cross-validation (Wahba 1990) or generalized cross-validation (Hutchinson 1995) principles have been proposed.

The function $f(x_i)$ of the above equation can be any radial basis function (RBF) and several commonly used RBFs are discussed in Karri et al. (2009). For thin plate splines (TPSs), the spline function is as follows:

$$f(x) = \sum_{i=1}^n \theta_i K_i(x) + \sum_{j=1}^m \theta_{n+j} \phi_j(x) \quad (5.1.2)$$

where

$$K_i(x) = \frac{1}{16\pi} \|x - x_i\|^2 \ln(\|x - x_i\|^2) \quad (5.1.3)$$

is the TPS radial basis function, θ_i is the i^{th} parameter and $\phi_j(x)$ is a polynomial basis. The parameters a of the resulting system can be calculated by solving the following linear set of simultaneous equations¹ (Wang 2011):

$$A\theta = y \quad (5.1.4)$$

where

$$A = \begin{bmatrix} K - n\lambda I_{n \times n} & P \\ P^T & 0_{m \times m} \end{bmatrix} \quad (5.1.5)$$

$$\theta = [\theta_1, \theta_2, \dots, \theta_{n+m}]^T; y = [y_1, y_2, \dots, 0_{1 \times m}]^T \quad (5.1.6)$$

$$K = \begin{bmatrix} K_1(x_1) & \cdots & K_n(x_1) \\ \vdots & \ddots & \vdots \\ K_1(x_n) & \cdots & K_n(x_n) \end{bmatrix}; P = \begin{bmatrix} 1 & x_1 & \cdots \\ \vdots & \vdots & \\ 1 & x_n & \cdots \end{bmatrix} \quad (5.1.7)$$

and P^T denotes the transpose of P .

One of the major problems associated with TPS smoothing is its high sensitivity to outliers. In computer vision problems, the data are often contaminated with both the gross outliers and/or pseudo-outliers (Stewart 1997). The gross outliers are produced by

¹The dimensionality is assumed to be two.

errors in the data generation process, while pseudo-outliers are correct measurements of another structure away from the structure of interest in a multi-structure scenario.

In the computer vision literature, many methods for fitting parametric models to data contaminated with noise and outliers have been proposed (Meer 2004). Some well-known robust statistical methods that can tolerate relatively high number of outliers include random sample consensus (RANSAC) (Fischler & Bolles 1981), least median of squares (LMS) (Rousseeuw & Leroy 2005) and many variations of these techniques (Chin et al. 2012). These methods perform efficiently when the number of parameters in the model is both low (below 10) and known a priori. In Tran et al. (2012), it is even shown that RANSAC based approaches can be used for outlier rejections in nonlinear fitting problems (such as deformable registration applications) when the errors associated with outliers are significantly less than the error associated with ignoring the curvature of the underlying function or manifold. However, in applications where the data contain more than one manifold, the scale of pseudo-outliers (outliers associated with other structures), would be similar to the scale of errors associated with linearisation of underlying manifolds. In such applications (e.g. motion field smoothing, 3D modelling using range data), the use of RANSAC-based methods for spline type data fitting will encounter the following two hurdles.

1. In a general case the number of parameters needed for spline representation of a structure of interest (n) is not known a priori and likely to vary even within different parts of a single dataset (e.g. an image). Using a smaller value than needed may lead to a suboptimal fitting of the data, whereas using a higher number will increase the probability of including outliers in the calculation, leading to an incorrect fit.
2. In the robust fitting algorithms mentioned above, the optimization is conducted using random sampling by selecting a finite number of subsets of data, with the hope that at least one of the selected subsets will be outlier-free. For a spline representation with large number of parameters on useful data with a high percentage of outliers (multi-structural scenarios), the probability of picking such a subset is extremely small. This would make these algorithms computationally intractable, even for off-line applications.

In recent years, there has been growing interest in addressing the above issues. A

method based on orthogonal least squares (OLS) was proposed by Chen & Li (2012) to select a suitable set of knots (parameters) that is adequate to represent a function with sufficient accuracy. However, this method fails in the presence of outliers as it gives high significance to outliers. The same authors later developed a robust method for multi-quadratic spline fitting to address the issue of the existence of outliers (Chen & Li 2013). This method first selects a subset of points as knot points using a space-filling design and then minimizes the L1 norm of the cost function instead of the L2 norm to provide robustness². The method was tested for data with gross outliers and the results showed significant improvements compared to non-robust methods. However, our experiments, detailed in Section 5.3, showed that this method performs poorly where the data contains pseudo-outliers. The method also does not include a scheme to automatically calculate the number of knots required to give the best accuracy and the required number of knots has been chosen empirically in their experiments. Again, our examples will show that the performance of this method is very sensitive to the choice of this number.

In the present study, a computationally-tractable robust method for fitting TPSs to contaminated data (including large numbers of gross and pseudo-outliers) is presented. Similar to the FLkOS (Bab-Hadiashar & Hoseinnezhad 2008), we use the k -th order square of residual as the cost function and its derivatives to minimise this function, and find an initial fit to the data and employ MSSE (Bab-Hadiashar & Suter 1999) to dichotomise the data and separate inliers from outliers. Finally, the resulting TPS is calculated using all the inliers. The results of the proposed method are compared with the performance of the robust spline fitting method using L1-Norm (Chen & Li 2013).

The rest of the chapter is organized as follows. Section 5.2 provides a detailed description of the proposed method while Section 5.3 presents the results achieved with the proposed method and compares them with those using competing methods. Section 5.4 concludes the chapter.

5.2 Robust Spline Fitting

To solve the TPS fitting problem in the presence of outliers, we propose an iterative two-step approach. In this method, the data is iteratively classified into inliers and outliers and the TPS fitting is performed using only inliers to improve the earlier classification.

²For a vector x with n elements Lp norm is defined as: $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$

The first challenge in this approach is the selection of a suitable method to dichotomise inliers and outliers. RANSAC and its variants are the most popular high-breakdown model fitting techniques used in computer vision applications. Notwithstanding the fact that these methods need the knowledge of inlier noise to perform segmentation, the main issue with using them to solve the spline fitting problem is that in order to guarantee correct inlier/outlier dichotomisation, at least one candidate model (selected using random sampling) has to be outlier free. Due to the relatively large size of the parameter space involved in TPS fitting, selecting a subset that contains inliers alone (from a set of data that is contaminated by a relatively high percentage of outliers) is extremely time-consuming. To overcome this issue, we propose the use of fast least k-th order statistics (FLkOS) (Bab-Hadiashar & Hoseinnezhad 2008) for finding a good fit and MSSE (Bab-Hadiashar & Suter 1999) to perform the segmentation. FLkOS uses the derivatives of a modified least k-th order statistics cost function in a Newton-based optimisation scheme to derive the model parameters. MSSE uses a putative fit and iteratively identifies inliers using the following criterion (Bab-Hadiashar & Suter 1999):

$$r_{(i+1)}^2 < T^2 \sigma_i^2 \quad (5.2.1)$$

where i is the index after sorting, r_{i+1} is the squared residual at index $i+1$, T is a constant threshold ($T = 2.5$ is used to include 99% of inliers based on a normal distribution (Bab-Hadiashar & Suter 1999)) and $\sigma_{(i)}$ is the standard deviation of sorted residuals up to index i calculated as $\sigma_i^2 = \frac{1}{i} \sum_{j=1}^i r_{(j)}^2$.

The spline fitting problem in Equation (5.1.4) can be set up as a standard regression model as shown below:

$$r_i = A_i \theta - y_i \quad (5.2.2)$$

where A_i is the i^{th} row of the matrix A in (5.1.4) and y_i is the i^{th} element of y . For this representation the least k-th order statistics cost function would be a single term given by:

$$r_k^2 = \left(A_k \hat{\theta} - y_k \right)^2. \quad (5.2.3)$$

Since using this single term cost function in Newton-type optimisation would lead to a non-invertible Hessian, the FLkOS algorithm uses a modified form of this cost function,

as shown below (Bab-Hadiashar & Hoseinnezhad 2008).

$$J(\hat{\theta}) = \sum_{m=-\lceil \frac{p-1}{2} \rceil}^{\lfloor \frac{p-1}{2} \rfloor} r_{k+m}^2 \quad (5.2.4)$$

Assuming that there is a small neighbourhood where the index of sorted residuals would not change due to the change in parameters, the derivatives of the above cost function are as follows:

$$\nabla J(\hat{\theta})|_{\hat{\theta}=\hat{\theta}_l} = 2 \sum_{m=-\lceil \frac{p-1}{2} \rceil}^{\lfloor \frac{p-1}{2} \rfloor} A_{k+m} \left(A_{k+m}^T \hat{\theta}_l - y_{k+m} \right) \quad (5.2.5)$$

$$\nabla^2 J(\hat{\theta})|_{\hat{\theta}=\hat{\theta}_l} = 2 \sum_{m=-\lceil \frac{p-1}{2} \rceil}^{\lfloor \frac{p-1}{2} \rfloor} A_{k+m} (A_{k+m})^T \quad (5.2.6)$$

and using the above calculated derivatives in a Newton optimisation scheme would lead to the following iterative solution:

$$\hat{\theta}_{l+1} = \begin{bmatrix} A_{[k-(p-1)/2]} \\ \vdots \\ A_{[k+(p-1)/2]} \end{bmatrix}^{-1} \begin{bmatrix} y_{[k-(p-1)/2]} \\ \vdots \\ y_{[k+(p-1)/2]} \end{bmatrix}. \quad (5.2.7)$$

This shows that the next parameter estimate is given by the solution to the TPS equation using only the p data points that correspond to the p sorted residuals around the r_k^2 . The complete implementation of FLkOS for TPS fitting is provided in Algorithm 3. In order to minimise the chance of converging to a local minimum, the FLkOS algorithm is initialized at several random starting points. It is important to note here that this random initialisation does not require all the points to be from a group of inliers.

One of the parameters needing to be initialized is the index value of the k -th residual ($k - th$). Ideally, in TPS fitting, this value should be the number of inliers in the structure of interest in the given data. Since this value is not known in advance, selecting a suitable value can be a challenge. Using a value larger than the ideal value would lead to a fitting that always includes some outliers and using a smaller value may lead to a suboptimal fit.

To demonstrate the effect of K -th value on the final segmentation outcome, we present a simulation that uses a set of contaminated data with 70% inliers. In this study, first the K -th parameter is set to a lower value (35%) than the actual number of inliers and the

Algorithm 3 Pseudo-code for FLkOS Implementation**Inputs:** Data, p , k , k -th

```

1: repeat
2:   Randomly Select a set of  $p$  data points.
3:   Compute the TPS parameters ( $\hat{\theta}_l$ ) using the  $p$  points and calculate the residuals for
   all data points.
4:   Sort the Residuals ( $|r|$ ).
5:   calculate  $MAD = 1.4826 \times \text{median}(|r - \text{median}(r)|)$ 
6:   Set  $J_{resh} = 10^{-5} \times MAD$ ,  $J_{prev} = r_{k-th}^2$ ,  $J_{min} = 10^7$ .
7:   repeat
8:     Select  $p$  points corresponding to the  $p$  sorted residuals centred around  $k$ .
9:     Compute the TPS parameters ( $\hat{\theta}_{l+1}$ ) using (5.2.7).
10:    Calculate the Residuals and Sort ( $r$ ).
11:    Set  $J = r_{K^{th}}^2$ ,  $\Delta J = |J - J_{prev}|$ 
12:    if  $J_{min} > J$  then
13:       $J_{min} \leftarrow J$ 
14:       $\theta^* \leftarrow \hat{\theta}_{l+1}$ .
15:    end if
16:  until  $\Delta J < J_{resh}$  or maximum number of iterations
17:  if  $J_{min} < J_{min}^*$  then
18:     $J_{min}^* \leftarrow J_{min}$ 
19:     $\theta_{final}^* \leftarrow \theta^*$ 
20:  end if
21: until maximum number of iterations

```

results are shown in Figure 5.1(a). This figure shows that the direct application of the FLkOS and MSSE is sensitive to the choice of the k value and can underfit the data.

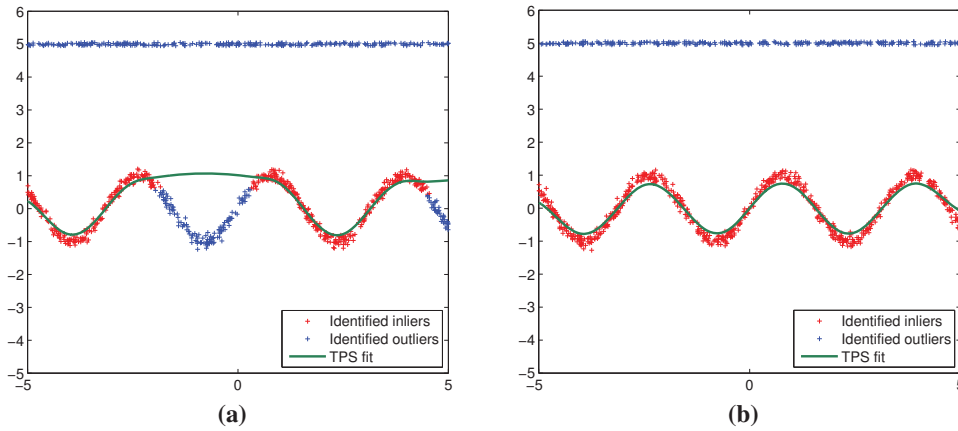


Figure 5.1: Fitting a TPS model to a dataset with 70% inliers. (a) When initial number of inliers is set to 35% and FLkOS and MSSE are directly applied. (b) When initial number of inliers is set to 35% and a spline is fitted using the proposed iterative method.

To overcome this problem, we use an iterative scheme in which we start with a lower

Algorithm 4 Pseudo-code for proposed method**Inputs:** Data, N_{inlier}^{init}

```

1:  $n = \text{length}(\text{Data})$ .
2:  $N_{in} = N_{inlier}^{init}$ 
3: repeat
4:    $p \leftarrow N_{in}/5$ 
5:    $k \leftarrow N_{in}/2$ 
6:    $K^{th} \leftarrow N_{in}$ 
7:   Compute FLkOS Model parameters  $\hat{\theta}$  using  $p, k, K^{th}$ .
8:   Calculate the residuals for TPS with parameters  $\hat{\theta}$ 
9:   Sort the residuals
10:  for  $i = n/2 \rightarrow n$  do
11:    calculate  $\sigma_{i-1}^2 = \frac{1}{i-2} \sum_{j=1}^{i-1} r_j^2$ 
12:    if  $r_i^2 > T^2 \times \sigma_{i-1}^2$  then
13:       $N_{in}^* = i$ 
14:      Break
15:    end if
16:  end for
17:   $\Delta N = |N_{in} - N_{in}^*|$ 
18:   $N_{in} \leftarrow N_{in}^*$ 
19: until  $\Delta N < \text{Threshold}$ 

```

value for variable k-th and compute the best estimate using the FLkOS algorithm. Next, this estimate is used as the candidate model and the number of inliers for this model candidate is identified using the MSSE criterion (5.2.1). After the inliers are identified, a new set of TPS parameters is calculated by using the FLkOS algorithm with k-th equal to the number of inliers, as calculated in the first step. This process is repeated several times until the number of inliers does not change. The proposed method is described in Algorithm 4. The result of applying the proposed scheme to the above data is shown in Figure 5.1(b). The figure shows that the proposed iterative scheme is able to correctly fit a TPS, even when the initial value for k-th is much lower than the actual number of inliers.

It is important to note that the proposed method always returns the largest structure in the given data first. However, the largest structure does not necessarily have to have the absolute majority of data points. In our implementation, we have assumed that the underlying signal can be represented by around 20% of the data points.

5.3 Experimental Results

The proposed method was implemented using MATLAB and the results are compared with the non-robust TPS (L2-Norm) and robust L1-Norm TPS (Chen & Li 2013) methods

using both synthetic and read data experiments.

5.3.1 Synthetic data simulations

In order to visualise how the proposed method works, we have used one-dimensional signals first followed by some two-dimensional results. To analyse the performance of different spline fitting methods, the following one-dimensional functions are commonly used:

$$g_1(x) = \sin(x) + \epsilon ; \quad -5 < x < 5 \quad (5.3.1)$$

$$g_2(x) = \sin(\pi(1 - x)^2) + \epsilon ; \quad -2 < x < 2 \quad (5.3.2)$$

where ϵ is the inlier noise and in our experiments is set to: $\epsilon \sim N(0, 0.1^2)$ and data values are normalised between $[-1, 1]$. In addition to points generated by these functions, some outliers (either unstructured or structured) were also added to the data. Mean error values for every fit were calculated as follows (Chen & Li 2013):

$$Mean\ Error = \sqrt{\frac{\sum_{i=1}^n (f(x_i) - \bar{y}_i)^2}{n}} \quad (5.3.3)$$

where $\bar{y}_i$ are the true values of the underlying function.

In the first simulation, following Chen & Li (2013), we conducted experiments with sets of data that had only gross outliers from the distribution $N(0, 5^2)$. The percentage of gross outliers was changed from 10 to 40 and the results are shown in Figure 5.2. These results show that the proposed method achieved best results across all percentages of outliers followed by the L1-Norm method. For this type of outliers the advantage of using the proposed method is limited, because methods using L1-Norm also perform well on this type of data.

In the second simulation, we introduced some pseudo-outliers (evenly distributed data around a constant value away from the signal) in place of normally-distributed gross outliers. The results are shown in Figure 5.3. These results show that for lower levels of outliers the L1-Norm method also achieves good results. However, when the outlier percentage is increased the performance of the L1-Norm method degrades, while the proposed method still performs well. It is not surprising to note that the results of L1-Norm degrade even faster when the second structure is moved farther apart, while the proposed method is not affected by the location of distant structures. When the structures are very

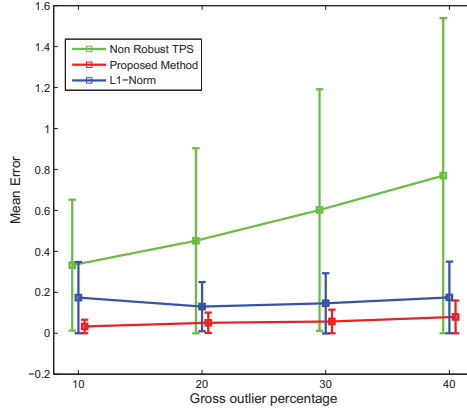


Figure 5.2: The variation of the mean error of various spline estimation methods for different percentages of gross outliers. The data in this simulation is created using (5.3.2) and gross outliers are generated using a zero mean normal distribution with standard deviation of 5.

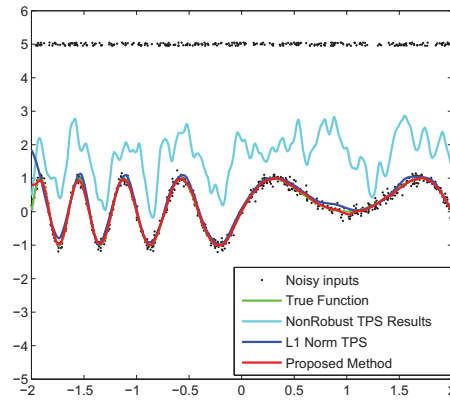
close or are overlapping, both methods had similar results with some visible errors.

The performance variation of the L1-Norm algorithm with different numbers of knot points is shown in Figure 5.4. These results show that the performance of the L1-Norm method is highly dependent on the number of knot points chosen. For instance, for the function (5.3.1), ten knot points would give the best results. However, as shown in Figure 5.5, the same number of knot points results in a poor representation of function (5.3.2). This indicates that the optimum number of knots is dependent on the shape of the structure of interest, and it would be difficult to choose an appropriate value without some knowledge of the underlying shape.

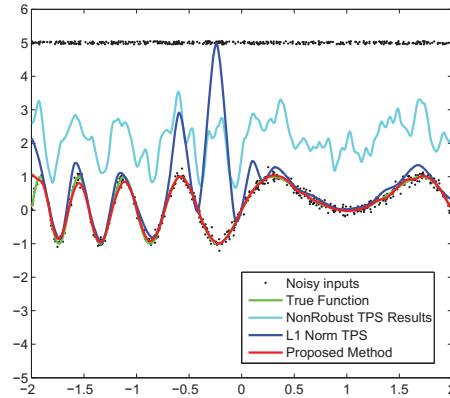
Even for smaller amounts of pseudo-outliers, if the data is locally (spatially) concentrated, the L1-Norm would fail, whereas the proposed method performed well. An example of such a scenario is shown in Figure 5.6. The case of having spatially localised outliers is commonly encountered in many computer vision applications, particularly in parametric segmentation tasks such as range and motion segmentation.

To demonstrate and compare the performance of the proposed method for two-dimensional cases, we used the following 2D test function, as proposed by Chen & Li (2013):

$$g_3(x, y) = 3(1-x)^2 e^{-x^2-(y+1)^2} - 10\left(\frac{x}{5} - x^3 - y^5\right) e^{-x^2-y^2} - \frac{1}{3}e^{-(x+1)^2-y^2} + \epsilon \quad (5.3.4)$$



(a)

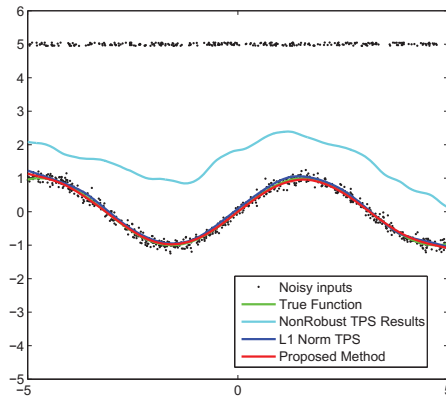


(b)

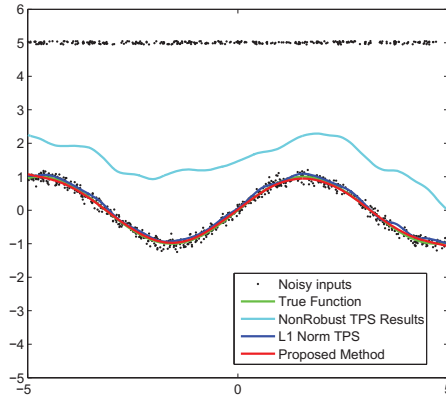
Figure 5.3: Resulting TPS fit for the function in (5.3.2) with (a) 30% (b) 40% outliers from distribution $N(5, 0.1^2)$. The number of knots for L1-Norm method was set to 25.

where $\epsilon \sim N(0, 1^2)$. Different numbers of pseudo-outliers from $N(15, 2^2)$ distribution were then added to test the robustness of the competing methods. The shape of the above function (with no noise) with added outliers is shown in Figure 5.7. The mean error results for the proposed method and the L1-Norm method with different number of outliers are presented in Figure 5.8. For L1-Norm implementation, as suggested in Chen & Li (2013), 100 knot points were used to estimate the underlying spline.

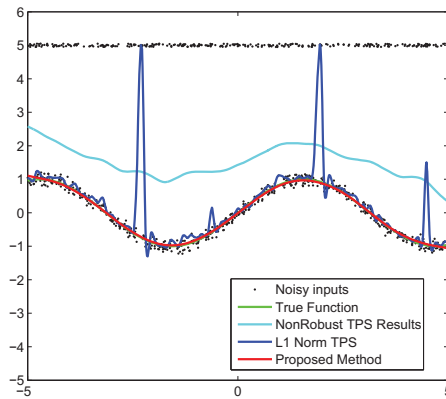
The results show that the mean error of the L1-Norm method increases exponentially when the outlier percentage is increased, whereas the mean error of the proposed method increases linearly at a low gradient. The linear increase in the mean error for the proposed method is explained by the fact that the data points available for fitting are decreased when some points are replaced by outliers.



(a)



(b)



(c)

Figure 5.4: The estimated splines for function (5.3.1) with 30% of outliers using different methods. Numbers of knots for L1-Norm TPS are chosen to be 10 in figure (a), 50 in figure (b) and 100 in figure (c).

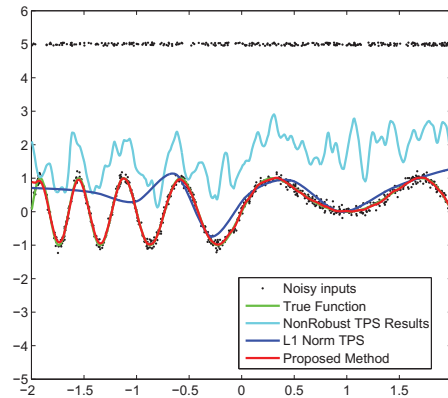


Figure 5.5: The estimated splines for function (5.3.2) where the L1-Norm method is implemented with 10 knots, which was shown to be the optimum number for the first function (5.3.1). This figure shows the L1-Norm method is fairly sensitive to the number of knots chosen.

To provide an indication of the time required to estimate the 2D splines of function (5.3.4) using both the proposed and the L1-Norm methods, their average running times, using our MATLAB implementations, are shown in Table 5.1. It is important to note here that since the proposed method uses all the inliers to construct the TPS the times would be different for different percentages of inliers.

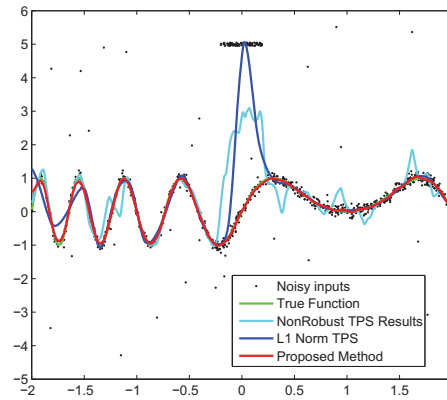
Table 5.1: Computation time of estimating a 2D spline using both the proposed and the L1-Norm methods. Size of the dataset in these simulations is 101×101 .

	L1-Norm	Proposed Method
Outlier Rejection	-	94.75
Fitting TPS to inliers	-	19.25
Total Time	135.5 (s)	114.0 (s)

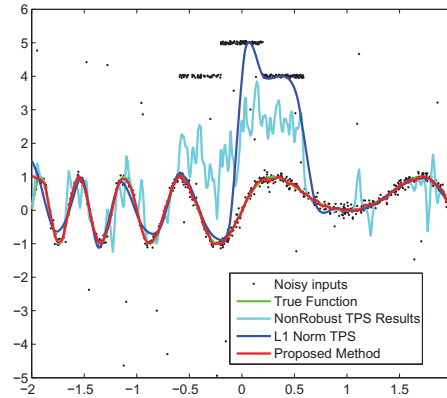
The performance of the L1-Norm method degrades even for a much lower number of outliers if they are concentrated in space. An example of such behaviour is shown in Figure 5.9. In the example, the percentage of outliers is around only 3%. The results show that the L1-Norm method failed to exclude outliers and returned a wrong fit. The proposed method is, however, able to properly recover the intended function.

5.3.2 Real data experiments

To demonstrate that the proposed method can be used to model surfaces in range data, the 3D cloud of points produced by a structured light range scanner of an object with a



(a)



(b)

Figure 5.6: Estimated splines for function (5.3.2) with: 25% pseudo-outliers plus 5% gross outliers (a) and 10% pseudo-outliers plus 5% gross-outliers (b). To make a meaningful comparison, densities of data points along the x axis for pseudo-outlier and main structure are equalised.

number of parallel surfaces was used in our experiment. Although range data are usually clean and accurate, the existence of multiple close structures means that there are significant numbers of pseudo-outliers, no matter which surface is modelled. Figure 5.10 depicts the test data as well as the results of applying both the L1-Norm and the proposed method to model these objects. The outcomes show that, similar to the simulation results, the L1-Norm method bridges different surfaces while the proposed method discards outliers (data associated with other surfaces as well as gross outliers) correctly and fits a spline to the largest surface.

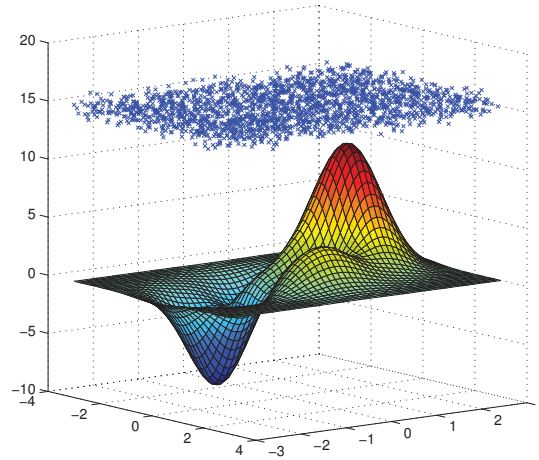


Figure 5.7: Plot of the function (5.3.4) with no noise and all added pseudo-outliers.

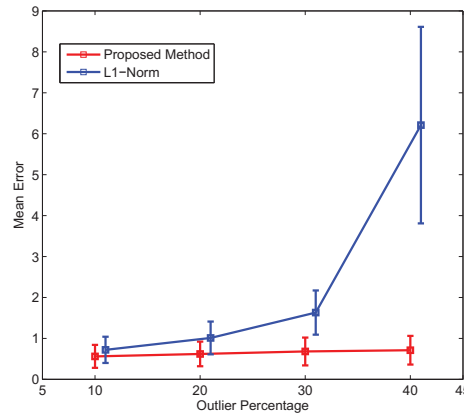
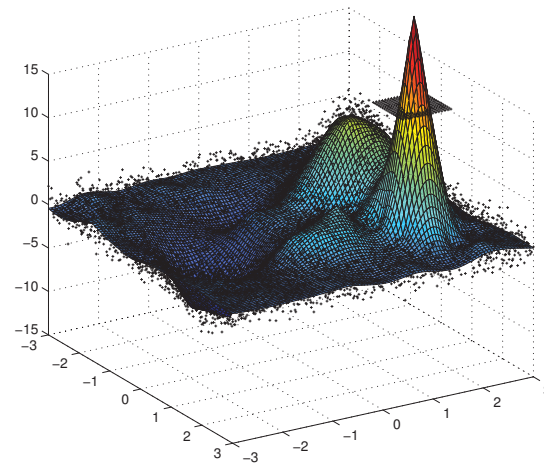


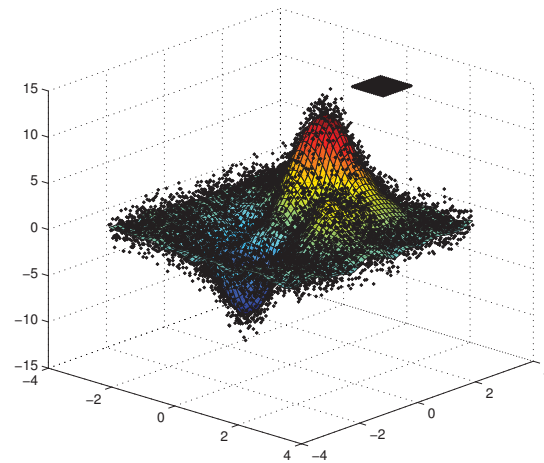
Figure 5.8: The variation of the mean error of various spline estimation methods for different percentages of pseudo-outliers. The data in this simulation is created using (5.3.4) and outliers are generated using a normal distribution with mean of 15 and standard deviation of 2.

5.4 Conclusion

A new algorithm for modelling visual data using robust thin plate spline fitting has been proposed in this chapter. The fitting algorithm uses a variant of the least k -th order statistics fitting approach with a novel iterative method for performing the inlier/outlier segmentation. The proposed method was tested on synthetic and real data and was shown to be effective, even when numerous structured outliers (pseudo-outliers) are present.

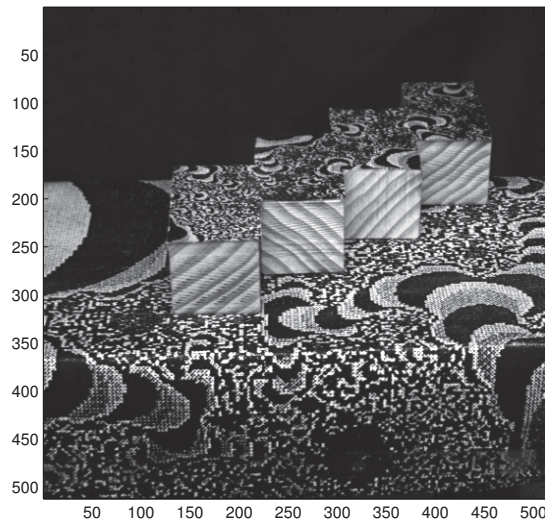


(a)

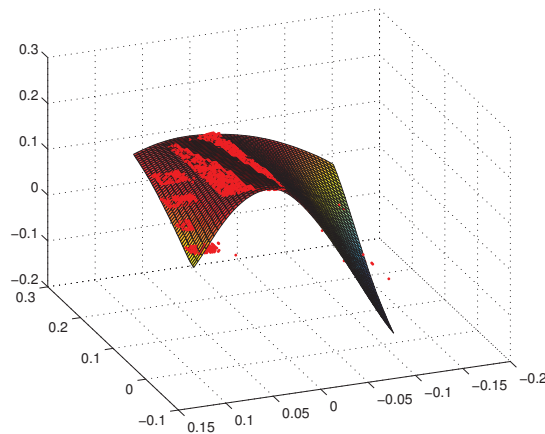


(b)

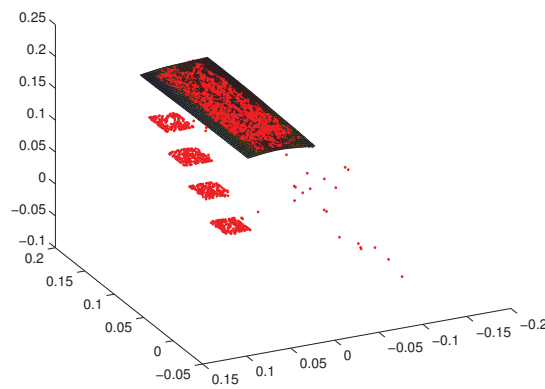
Figure 5.9: Estimated splines for function (5.3.4) with only 3% localised pseudo-outliers using (a) L1-Norm method and (b) proposed method.



(a)



(b)



(c)

Figure 5.10: The results of the real data experiment. The measured 3D depth points (point cloud) are shown by red dots and are rotated for better visualisation. (a) The picture of the object used in the range data modelling experiment. (b) Estimated splines for the L1-Norm method. (c) Estimated splines for the proposed method.

Chapter 6

Robust Model Fitting Using Higher Than Minimal Subset Sampling

6.1 Introduction

The task of identifying the underlying model in a set of data contaminated with both noise and outliers is a highly researched area in computer vision. This task has many applications, including motion segmentation (Elhamifar & Vidal 2013, Poling & Lerman 2014, Torr 1995), range image segmentation (Hesami et al. 2010, Lara-Alvarez et al. 2009, Wang & Suter 2004), medical image analysis (Hevia-Montiel et al. 2007) and visual tracking (Lucena et al. 2010, Shen et al. 2010). In computer vision problems, the data often comprise multiple structures that result in pseudo-outliers (correct measurements of another structure away from the structure of interest) in addition to gross-outliers (Stewart 1997) that are produced by errors in the data generation process.

There are a large number of robust model-fitting techniques that can be used in recovering the underlying models in the presence of both gross and pseudo-outliers, and many of these techniques involve optimization of highly complex cost functions. A commonly used approach is to discretise the parameter space using sampling and evaluate the cost function on these discrete points to find the optimum (*hypothesize and verify strategy*). The assumption here is that at least one of the hypotheses selected will be sufficiently close to the true structure that is to be recovered.

In hypothesize and verify approach, the hypotheses are generated by sampling subsets of p data points and estimating the model represented by those data points. Here p refers to

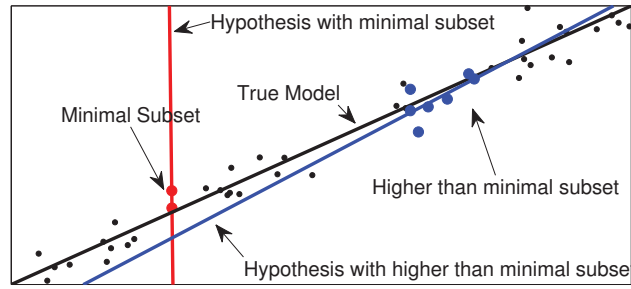


Figure 6.1: Minimal subset sampling vs. higher than minimal subset sampling. The data do not contain outliers (Pham et al. 2012). The figure represents an extreme case where the span of the data samples is made deliberately small.

the number of parameters needed to represent the model and using p as the sample size is known as minimal subset sampling (MSS) (Pham et al. 2012). In the presence of outliers and multiple structures, the chance of generating a hypothesis close to a true structure using random sampling is small. Recently several methods have been proposed to bias the sampling process towards selecting points from the same structure in consecutive steps (Chin et al. 2010, Fan & Pylvänäinen 2009). Estimating how many samples need to be selected in order to guarantee success with a high probability is not trivial. The main assumption made in setting the number of samples is that one sample with only inliers (clean sample) will be adequate to generate a hypothesis close to the true structure. However, as shown in Figure 6.1, in the presence of noise even a clean sample may result in a hypothesis that is far from the true model, particularly if the span of the sample data points, in one dimension, is not much larger than the scale of noise (Pham et al. 2012). For this reason, most implementations generate a higher number of samples than necessary, which in applications involving high-dimensional model fitting, can make them computationally inefficient. More importantly, there is usually no measure to indicate the success, at a given number of samples.

Methods such as LO-RANSAC (Chum et al. 2003) try to improve the closeness of an initial hypothesis to the true model by local search. However, this only works when the initial guess is in the vicinity of the true structure. Pham et al. (2012) proposed higher than minimal subset sampling to generate better hypotheses instead of refining the inaccurate ones. As shown in Figure 6.1 using higher than minimal subset sampling for hypothesis generation has a greater chance of generating a hypothesis closer to the true model, given that all the points in the sample are inliers (this claim will be considered in more detail

in the next section). It is important to note here, that selecting an all-inlier sample in the presence of multiple structures and gross outliers, becomes increasingly difficult when the sample size is increased (due to the multiplication of probabilities). This makes direct extension of MSS sampling methods to higher than minimal subset sampling computationally inefficient.

In this chapter, similar to Bab-Hadiashar & Hoseinnezhad (2008), we propose an efficient robust model-fitting method that uses higher than minimal subset sampling to generate hypotheses. In this method, one starts from an arbitrary hypothesis (which does not need to be in the vicinity of the solution) and moves towards a structure in the data using local information available at each iteration. The method also has the ability to identify when the algorithm has reached a hypothesis with adequate accuracy and stop appropriately, saving computational time.

The rest of this chapter is organized as follows. Section 6.2 discuss the implications of using higher than minimal subset sampling and prior work in that direction. Section 6.3 describes the proposed method in detail and Section 6.4 presents experimental results involving synthetic and real data, and comparisons with state-of-the-art model-fitting techniques. Section 6.6 concludes the chapter.

6.2 Higher Than Minimal Sampling

6.2.1 Previous work

There are two approaches in the literature that use higher than minimal subsets to solve the model-fitting problem.

The first approach is to use higher than minimal subset samples to generate affinities between those points and represent them using a hyper-graph, which is then partitioned to obtain clusters in the data. Agarwal et al. (2005) proposed a two-step algorithm to cluster the higher order affinities. In the first step, they constructed a hyper-graph with $h = p + 1$ vertices per edge. This hyper-graph was then approximated with a pairwise graph using the clique averaging technique and the resulting pairwise graph was segmented using a spectral partitioning algorithm. A method that partitioned the hyper-graph directly without converting it to a pairwise graph was introduced by Liu et al. (2010). Their approach requires the hyper-graph and the weights to be calculated at the start. However, calculating the weights of a full hyper-graph with all the nodes having adjacencies with

each other is a very expensive task and a large memory space is needed to store the weights (space needed $\propto n^h$ where n is the number of data). Hairong & Shuicheng (2012) proposed a computationally efficient hyper-graph clustering method that used a hypothesize and verify strategy to approximately construct the hyper-graph called the random consensus graph which is then converted to a pairwise graph that approximately retains the affinity information. However, instead of constructing the hyper-graph and then converting it to the pairwise graph, the authors directly compute the pairwise edge weights from the consensus information. Once the pairwise graph is approximated it is partitioned using a slightly modified version of the robust ensemble clustering approach proposed in Liu et al. (2010). Since this algorithm relies on a RANSAC like method to construct the consensus information, it inherits the problems that come with RANSAC: such as identifying structures with multiple noise levels.

The second approach is to use higher than minimal subset sampling to improve the quality of the hypothesis in a hypothesize and verify strategy. In LO-RANSAC method, Chum et al. (2003) proposed a local optimization step that uses higher than minimal subset samples. However, this step would only work if an initial estimate in the vicinity of the true solution is provided using MSS. Pham et al. (2012), used higher than minimal subsets obtained using random cluster models (RCM) to generate hypotheses and used those to initialize a metric labeling problem that clusters the data points and recovers the underlying model. However, their method relies on the spatial contiguity of structures in the dataset, which may not be true for some model-fitting problems. Due to the need to generate spatial relationships, this method is also computationally expensive.

6.2.2 Evidence to support the use of higher than minimal subset sampling

In robust model-fitting techniques that utilize the hypothesize and verify strategy, at least p points need to be sampled at a time to generate a valid hypothesis. The number of parameters p constitutes the necessary condition for the sample size to derive a unique hypothesis.

It is said that increasing the number of points in a sample beyond p (higher than minimal subset sampling - HMSS) will increase the quality of the hypothesis (closer to the true model), given that all the sampled points are from the structure of interest (Pham

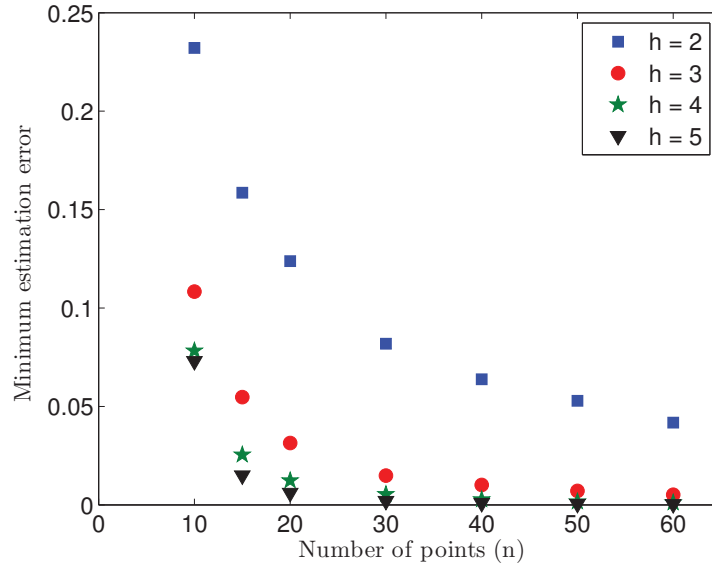


Figure 6.2: The variation in minimum estimation error with number of data points n for different sample sizes (h). The figure shows the mean results of 100 experiments for each combination.

et al. 2012). This assertion is examined in using a Monte Carlo simulation of a 2D line fitting. In this test, n data points representing a line in 2D space with Gaussian noise of $N(0, \sigma^2)$ were generated (no outliers) and all the possible tuples (each with h data points) were used to generate the hypotheses¹. Next, to measure the closeness of a given hypothesis (h_i) to the original line (h_t) we used the following criterion: given that p_{h_i} and p_{h_t} are the two points on each line that is closest to the origin, the distance between these two lines is calculated as $\|p_{h_i} - p_{h_t}\|_2$. Since this distance measure is proportional to the scale of noise we normalize by σ to get the final measure $ED(h_i, h_t) = \|p_{h_i} - p_{h_t}\|_2 / \sigma$. The minimum of these estimation errors ($\min_{i=1..(\binom{n}{h})} ED(h_i, h_t)$) indicates the closeness of the best hypothesis to the true model for a given number of points (n) and sample size (h). The average results of 100 such experiments for each n and h , are shown in Figure 6.2.

The results show that the quality of the hypotheses generated increases with the sample size. However, the improvement becomes very small after a few additional points ($h = p + 2$), particularly for data structures with a high number of points ($n > 20$). Importantly, we observed similar patterns when the dimensionality of the data is varied.

To analyze the probability of generating a good hypothesis, given that the data are

¹For $h = p = 2$, the hypothesis is generated by finding line connecting the two points. For $h > p$, the hypothesis is given by the least squares method.

from the inlier set, we generated n data points from a 2D line model with noise $N(0, \sigma^2)$. For each dataset, the estimation errors ($ED(h_i, h_t)$) of all possible hypotheses generated with h points were computed. These estimation error values (repeated over 100 such experiments) were then used to draw the cumulative distribution function (CDF) of estimation errors, shown in Figures 6.3 (a) and (b). These results show that the probability of obtaining a good hypothesis (given that the data are inliers) increases with the sample size h . Similar to the previous results, the improvement is only significant up to a few additional points ($h = p + 2$). The pattern remains similar when the dimensionality of the model is increased, as shown by an equivalent 3D plane fitting experiment in Figures 6.3 (c) and (d).

To investigate the behavior on real data, we carried out a fundamental matrix estimation task. The points that belong to one group were segmented manually, as shown in Figure 6.3(e) and the fundamental matrix was estimated using 10^5 random samples with sample size h . Since there is no information on the true model, the estimation error was calculated using median Sampson's distance from each hypothesis to the data points. The CDF of estimation error for different sample sizes is shown in Figure 6.3(f) and similar observations can be made from these results.

The above results show that “Higher than Minimal Subset Sampling” increases the likelihood of the closeness of the generated hypothesis to the true model. However, in the presence of outliers, increasing the number of points in a sample will decrease the probability of selecting a clean sample exponentially. Therefore, identifying how many more points one should include beyond the necessary condition is not straightforward. The above presented Monte Carlo simulation and experimental results show that the advantage of increasing the sample size beyond $h = p + 2$ would be limited. Direct extension of RANSAC-like methods that use random sampling for higher than minimal subset sampling will be computationally inefficient due to the decreasing probability of selecting a clean sample. In addition, these methods provide no means of knowing that a good hypothesis has been reached, making it necessary to generate a predefined number of hypotheses even if a good hypothesis is obtained early.

In the next section, we propose a new efficient method that carries out higher than minimal subset sampling using a direct approach similar to a greedy algorithm, which starts from an arbitrary sample and moves towards a good hypothesis using the available

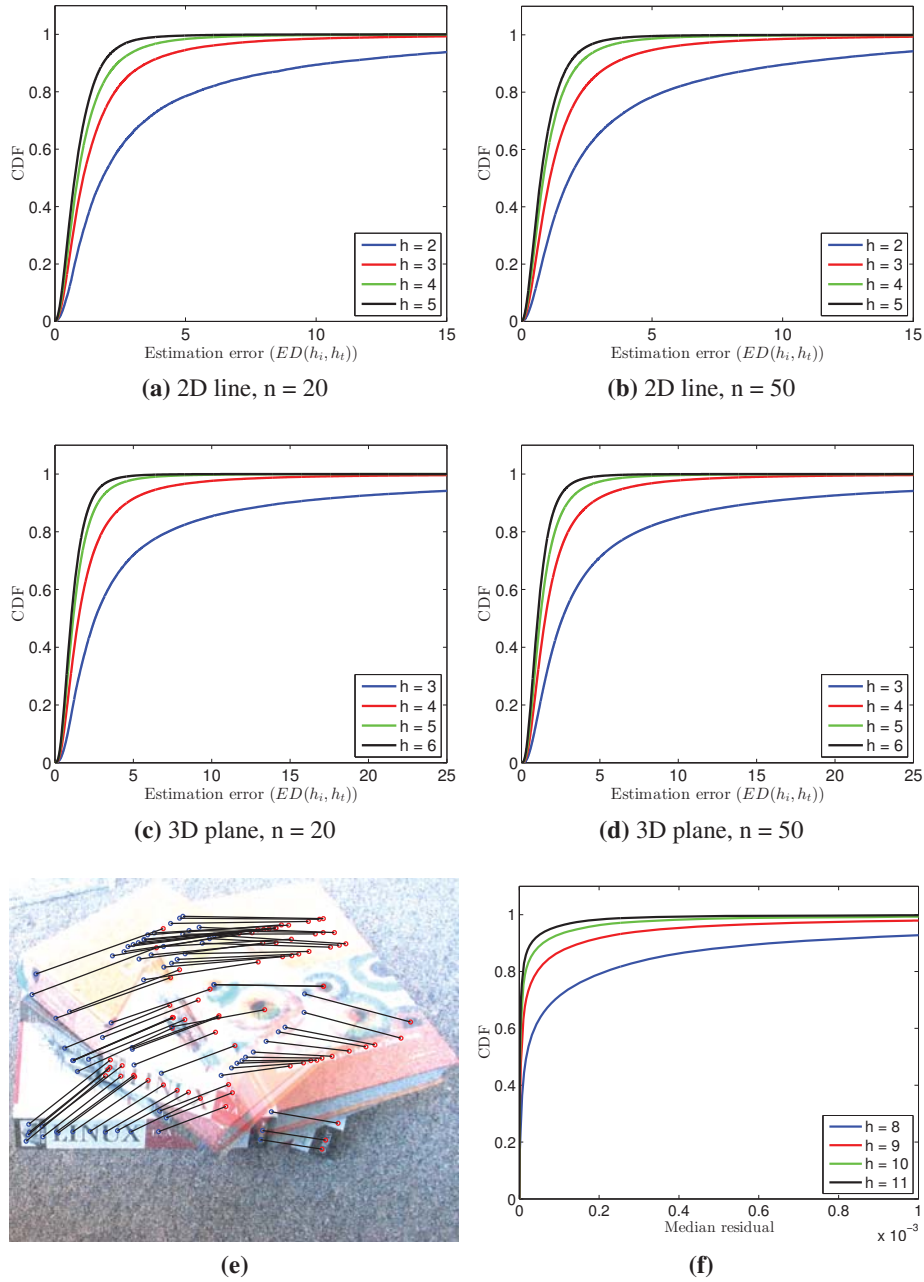


Figure 6.3: The cumulative distribution function of estimation errors for 2D line (a), (b) and 3D plane (c), (d) fitting using different sample sizes (h). The data used for fundamental matrix estimation and the CDF of the median residuals are shown in (e), (f).

local information. The method also has the capability to stop once a good estimate that is likely to be a true structure is reached, saving computation time.

6.3 Proposed Method

In this section we describe the proposed method for model fitting. The intention here is to cluster data points $([x_i]_{i=1}^n \in \mathbb{R}^d)$ into sub-groups, based on the existence of underlying models $([\theta^{(j)}]_{j=0}^{n_c})$. Here n is the number of data points and n_c is the number of structures in the dataset with $j = 0$ assigned to outliers.

6.3.1 Cost function

The first task is to establish a cost function that quantifies the suitability of a given model to represent a structure in data. Here we select the Least k -th order statistics (LkOS) estimator, which is well known for its stability and high breakdown point (Rousseeuw & Leroy 2005). LkOS cost function is as follows:

$$F(\theta) = r_{i_{k,\theta}}^2(\theta) \quad (6.3.1)$$

where $r_i^2(\theta)$ is the i -th squared residual with respect to model θ and $i_{k,\theta}$ is the index of the k -th sorted square residual with respect to model θ . Here k refers to the minimum acceptable size of a structure in a given application. It should be noted here that the value of k is almost always much larger than the dimensions of the parameter space ($k \gg p$).

Optimizing this cost function is highly complex. Hence, the parameter space is commonly discretized using randomly generated hypotheses and the cost function is evaluated at each of these points to find the best solution (hypothesize and verify). As was previously discussed, hypotheses generated using minimal subset sampling may not generate a hypothesis close enough to the true model. The intention in the proposed method is to generate more appropriate hypotheses using higher than minimal subset sampling. However, due to the multiplication of probabilities, it is not efficient to generate accurate hypotheses using random sampling of higher than minimal subsets.

To find a way of conducting HMSS efficiently, we present a greedy algorithm that starts from a random location on the parameter space and moves towards a solution that is within the basin of attraction of a local minimum using the locally available information. The algorithm starts by generating a hypothesis (θ_0) via random sampling of h data points which are then used to generate the residuals $r^2(\theta_0)$. Next, these residuals are sorted and the h data points around the k -th sorted residual are then used to generate a new set of

parameters as follows:

$$\theta_{l+1} = \text{LeastSquareFit} \left([x_{i_m, \theta_l}]_{m=k-h+1}^k \right) \quad (6.3.2)$$

where θ_l is the parameters at iteration l and the sample size h is set as $h = p + 2$ following the analysis presented in Section 6.2.2. This step (equation 6.3.2) is then repeated until the stopping criterion described in the next section is reached. The convergence of the proposed iterative scheme at a good estimate will be further analyzed in Section 6.3.3.

To visualize the operation of the above algorithm, the intermediate steps of the scheme in a simple 2D line fitting problem are shown in Figure 6.4. The figure shows that the correct structure is recovered by going through only a few iterations, even when the starting samples are not members of that structure (outliers).

6.3.2 Stopping criterion

The next main challenge is to identify a method to stop the algorithm once it has reached a good estimate that is likely to be a true structure (having at least k points as inliers). Once the algorithm arrives at a vicinity of a local minima representing a structure in data, the first k sorted points should be from that structure. As the proposed algorithm picks the points for the next iteration around the k -th sorted index, they too would be from the same structure. This leads to a situation where the consecutive samples are from the same structure. This property is utilized here to devise a stopping criterion by which we can detect if the algorithm has found a structure in data.

The stopping criterion is as follows:

$$F_{stop} = \left(r_{i_k, \theta_l}^2(\theta_l) < \frac{1}{h} \sum_{j=k-h+1}^k \underbrace{r_{i_j, \theta_{(l-1)}}^2(\theta_l)}_{(a)} \right) \wedge \left(r_{i_k, \theta_l}^2(\theta_l) < \frac{1}{h} \sum_{j=k-h+1}^k \underbrace{r_{i_j, \theta_{(l-2)}}^2(\theta_l)}_{(b)} \right) \quad (6.3.3)$$

Here (a) and (b) are the residuals of the sampled points in iterations $l - 1$ and $l - 2$ with respect to the current parameters θ_l . This criterion checks the data points associated with the two previous samples to see if the average residuals of those points (with respect to

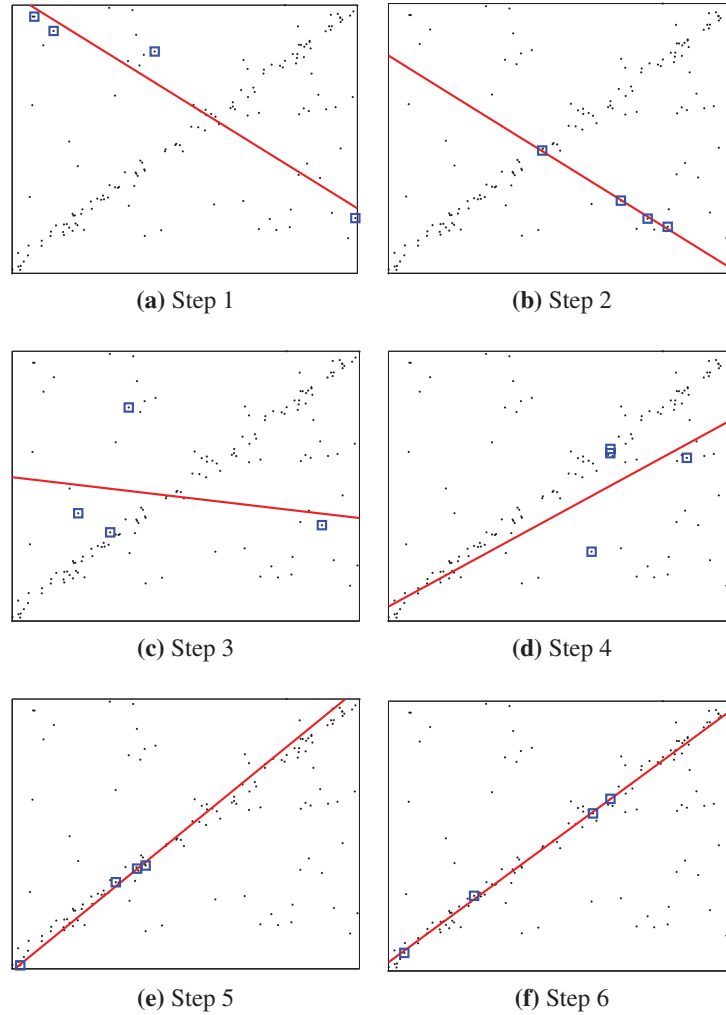


Figure 6.4: The intermediate steps of the proposed method in a simple 2D line fitting example for $h = 4$. The model represented by the current hyper-edge is plotted to make the steps clear. Note that the vertices of the starting point (Step 1) are not members of the structure and the algorithm does not move away from the underlying structure after it is found (Step 6).

the current parameters) are still lower than the inclusion threshold associated with having k points (assuming that a structure has at least k points implies that data points with residuals less than $r_{i_k, \theta_l}^2(\theta_l)$ are inliers). This indicates that the samples selected in the last three iterations are likely to be from the same structure hence the algorithm has converged.

A challenging problem in multi structural data segmentation is the existence of some local minima due to accidental alignment of outliers and true structures. A common solution, also used by competing methods ((Hairong & Shuicheng 2012, Pham et al. 2012, Yu et al. 2011)) is to reinitialize the hypothesis generation a number of times. Where this number is determined by the user. To quantify the number of required re-initializations,

in our method this process is carried out until there is no further improvement in the cost function in consecutive runs and the algorithm is then stopped automatically. In our experiments, it was observed that the number of random initializations required was always smaller than ten across different types of problems, which was far less than the number of random samples needed by the RANSAC-based methods.

For problems with multiple structures, once a reliable local minimum is returned, the core data points that correspond with that minimum are segmented out and the process is repeated until all the structures are recovered. In this implementation, we selected the Modified Selective Statistical Estimator (MSSE) (Bab-Hadiashar & Suter 1999) to segment points due to its low computational cost, high level of consistency, and small bias in applications involving close data structures (Hoseinnezhad et al. 2010). This step is aimed at removing the majority of points associated with already identified models to prevent the algorithm from returning the same structure in subsequent iterations. This may be particularly relevant to problems with one structure with a significantly lower level of noise compared to other structures. It should be noted that MSSE does not require any additional information (such as noise level) for the segmentation and if such information is available, a separate segmentation strategy utilizing those information would result in a better segmentation. The more one knows a priori, the better the outcome will be.

The complete algorithm of the proposed higher than minimal subset sampling based model-fitting method is given in Algorithm 5.

6.3.3 Why does the proposed method converge?

To analyze the local convergence of the proposed method, we first show that the parameter update used is equivalent to that of a generalized Newton method.

First we slightly modify the cost function in (6.3.1) to suit HMSS. The new cost function is:

$$\hat{F}(\theta) = \sum_{m=0}^{m=h-1} r_{i_{(k-m)},\theta}^2 \quad (6.3.4)$$

where $r_{i_{m,\theta}} = \theta^T x_{i_{m,\theta}} - y_{i_{m,\theta}}$ and $k \gg h \geq p$. The proposed method is based on finding the roots of this cost function

$$H(\theta) = \frac{\partial \hat{F}}{\partial \theta} = 0 \quad (6.3.5)$$

via Newton method. Bab-Hadiashar & Hoseinnezhad (2008) have shown that, although

Algorithm 5 Step-by-step algorithm of proposed higher than minimal subset sampling based model-fitting methods

Inputs: Data Points ($X \in [x_i]_{i=1}^N$), minimum cluster size (k), Number of clusters (n_c)

```

1:  $l_{max} \leftarrow 50, t_{max} \leftarrow 10, h \leftarrow p + 2$ 
2:  $l \leftarrow 0, t \leftarrow 1, j \leftarrow 1, [\hat{F}_{min}^{(j)}]_{j=1}^{n_c} \leftarrow \infty$ 
3: repeat
4:   repeat
5:     Select a random  $h$ -tuple from the data points.
6:     Generate model  $\theta_0$  using the  $h$ -tuple.
7:     repeat
8:        $[r^2(\theta_l), i_{\theta_l}] = \text{SortedRes}(X, \theta_l)$  .
9:       Calculate the cost function  $\hat{F}(\theta_l)$ .
10:       $\theta_{l+1} \leftarrow \text{LSFit}\left([x_{i_m, \theta_l}]_{m=k-h+1}^k\right)$ 
11:      Evaluate equation (6.3.3)
12:      if  $F_{stop}$  then
13:        break;
14:      end if
15:    until  $(l++ < l_{max})$ 
16:    if  $\hat{F}(\theta_l) < \hat{F}_{min}^{(j)}(t-1)$  then
17:       $\hat{F}_{min}^{(j)}(t) \leftarrow \hat{F}(\theta_l)$ 
18:       $\theta_{best}^{(j)} \leftarrow \theta_l$ 
19:    end if
20:    if  $\hat{F}_{min}^{(j)}(t) = \hat{F}_{min}^{(j)}(t-1) = \hat{F}_{min}^{(j)}(t-2)$  then
21:      break;
22:    end if
23:  until  $t++ < t_{max}$ 
24:   $[\text{outliers}] = \text{GetOutliers}(X, \theta_{best}^{(j)})$ .
25:   $X \leftarrow X(\text{outliers})$ 
26: until  $j++ < n_c$ 
27: Cluster data using  $[\theta_{best}^{(j)}]_{j=1}^{n_c}$ .
```

the index i of the k -th sorted residual cost function in equation (6.3.5) depends on the parameter estimates, there is a finite neighborhood (D_H) around the estimate in which the sorted indices do not change². Inside this neighborhood, the cost function has a quadratic relationship to the model parameters. Therefore, the cost function and its derivatives are continuous and differentiable. The generalized Newton method for solving this can be defined as follows (Qi & Sun 1999):

$$\theta_{l+1} = \theta_l - V_l^{-1} H(\theta_l) \quad (6.3.6)$$

²Only exception is when two residuals are exactly the same, in which case, the remedy is to merge those as one.

where $V_l \in \partial H(\theta_l)$. For the cost function (6.3.5) we can derive the following:

$$H(\theta) = 2 \sum_{m=0}^{h-1} x_{i_{(k-m),\theta_l}} (\theta^T x_{i_{(k-m),\theta_l}} - y_{i_{(k-m),\theta_l}}) = 2J^\top r \quad (6.3.7)$$

$$H'(\theta) = 2 \sum_{m=0}^{h-1} x_{i_{(k-m),\theta_l}} x_{i_{(k-m),\theta_l}}^\top = 2J^\top J \quad (6.3.8)$$

where $J^\top = [x_{i_{k,\theta_l}}, \dots, x_{i_{(k-h+1),\theta_l}}] = \mathcal{X}^\top$ and $r^\top = [(\theta_l^\top x_{i_{k,\theta_l}} - y_{i_{k,\theta_l}}), \dots, (\theta_l^\top x_{i_{(k-h+1),\theta_l}} - y_{i_{(k-h+1),\theta_l}})]$. Equations (6.3.6), (6.3.7) and (6.3.8) can be simplified into:

$$\theta_{l+1} = [\mathcal{X}^\top \mathcal{X}]^{-1} \mathcal{X}^\top \mathcal{Y} \quad (6.3.9)$$

which is equal to the proposed parameter update in equation (6.3.2).

The classical Newton method that is used to optimize in case where $H(\theta)$ is continuously differentiable, is locally quadratically convergent. This property is not directly applicable to a piecewise continuous function. However, theorem 2.2 in Qi & Sun (1999) shows that the generalized Newton method is quadratically convergent in a neighborhood of θ^* if H is strongly semi-smooth at θ^* . Equation (6.3.8) shows that $H'(\theta)$ does not depend on θ inside the neighborhood where the index does not change, hence $H'(\theta^* + \delta) - H'(\theta^*) = 0$. This satisfies the criterion of strong semi-smoothness³ (Qi & Sun 1999), which makes the proposed method quadratically convergent in a neighborhood of θ^* . The experimental analysis of both synthetic and real data presented in the next section shows that the proposed method either lands at this neighborhood and consequently converges to a solution, or the process is re-initiated.

6.4 Experimental Analysis

We evaluated the proposed method using both synthetic and real data experiments. The results of the proposed method were then compared in terms of both accuracy and computational time with RCM (Pham et al. 2012), which uses higher than minimal subset sampling as well as the following state-of-the-art model fitting techniques: QP-MF (Yu et al. 2011), MultiGS (Chin et al. 2010), LO-RANSAC (Chum et al. 2003) and RGC (Hairong & Shuicheng 2012).

³ H is strongly semi-smooth at x if $V_d - H'(x; d) = \mathcal{O}(\|d\|^2)$, $d \rightarrow 0$ where $V \in \partial H(x + d)$.

The code for the proposed algorithm was developed in MATLAB and the codes provided by the authors were used to generate the results for competing methods with parameters either set as instructed by the authors or tuned to give the best results. It should be noted that the RCM, MultiGS and LO-RANSAC methods have some part of their code implemented in C (MEX) and the QP-MF method uses the MOSEK quadratic solver whereas the proposed method is implemented in MATLAB by simple scripts.

The experiments were run on a HP Z400 workstation with an Intel Xeon W3550 processor. For each instance, the experiments were repeated 100 times and the averages of the results are reported.

The accuracy of all methods was evaluated using the commonly used clustering accuracy measure (Hairong & Shuicheng 2012, Yu et al. 2011) given as:

$$CA = \frac{\sum_{i=0}^{n_c} n_{tp,i}}{\sum_{i=0}^{n_c} n_i} \quad (6.4.1)$$

where $n_{tp,i}$ is the number of true positives in group i and n_i is the total number of points in that group.

The MultiGS and LO-RANSAC methods are intended at hypothesis generation and can be paired with any clustering method to generate the clustering accuracy measure. In this chapter, we have used consensus information (similar to RANSAC) as the clustering method for those techniques, noting that it requires inlier noise as an input. This parameter was manually set to the true inlier noise in our synthetic data experiments. Hence, the clustering accuracy for those methods reflects the best possible value and in practice, lower accuracies can be expected.

Since the MultiGS and LO-RANSAC methods do not have any explicit stopping criteria, in our experiments, the sampling times of these methods are limited to the average run time required by the proposed method in each problem.

6.4.1 2D line fitting

First, we evaluated our algorithm in detecting lines in a 2D point set contaminated with both noise and gross outliers using the standard regression model. The data points were generated by combining four intersecting lines, with each line containing n_i points with Gaussian noise $N(0, \sigma_i^2)$. Furthermore, uniformly distributed $n_0 = 100$ gross outliers were added. An example of a point set ($[n_i]_{i=0}^4 = 100, \sigma = 0.025$) is shown in

Figure 6.5(a). The structures returned by the proposed method are also shown.

We examined the performance of the algorithm for varying levels of inlier noise in the interval $\sigma_i \in [0.001, 0.05]$. The number of data points was fixed at $[n_i]_{i=0}^4 = 100$. The clustering accuracies are shown in Figure 6.5(b) together with those of competing methods.

The results show that the proposed method produces the best accuracy, specially for higher noise levels, closely followed by MultiGS and LO-RANSAC. It should be noted here that the threshold value for clustering in MultiGS and LO-RANSAC was manually set to the true value, giving these methods an advantage over the proposed method which uses MSSE for clustering (automatically estimating the scale of noise). The model complexity penalty (β) is an external parameter that is needed by the RCM method and during our experiments we found that in order to recover the correct model, this parameter needs to be manually tuned for each noise level. Figure 6.5(e) shows the variation of the clustering accuracy with β for each noise level. The RCM clustering accuracy in Figure 6.5(b) is the best achieved for a given noise level across all tested parameter values. None of the other methods required parameter tuning for each step.

Figure 6.5(c) shows the total computation time for each method. The results show that the proposed method is the fastest of the tested methods and is more than an order of magnitude faster than RCM.

Next, the number of total data points was varied in the range 250-2500 while the noise was fixed at $\sigma_i = 0.01$. The total computation times are shown in Figure 6.5(d). The results show that the proposed method is very efficient in terms of computation time with a slight linear increase in time with the number of points. As expected clustering accuracy remained constant.

Through these experiments the parameter k for the proposed method was fixed at $k = 0.1 * N$, where N is the total number of points.

6.4.2 3D plane fitting

In the second set of experiments we detected planes in a 3D point set contaminated with both noise and gross outliers using standard regression model. The data points were generated by combining four planes, with each plane containing n_i points with Gaussian noise $N(0, \sigma_i)$. Furthermore, uniformly distributed n_0 gross outliers were also added. An

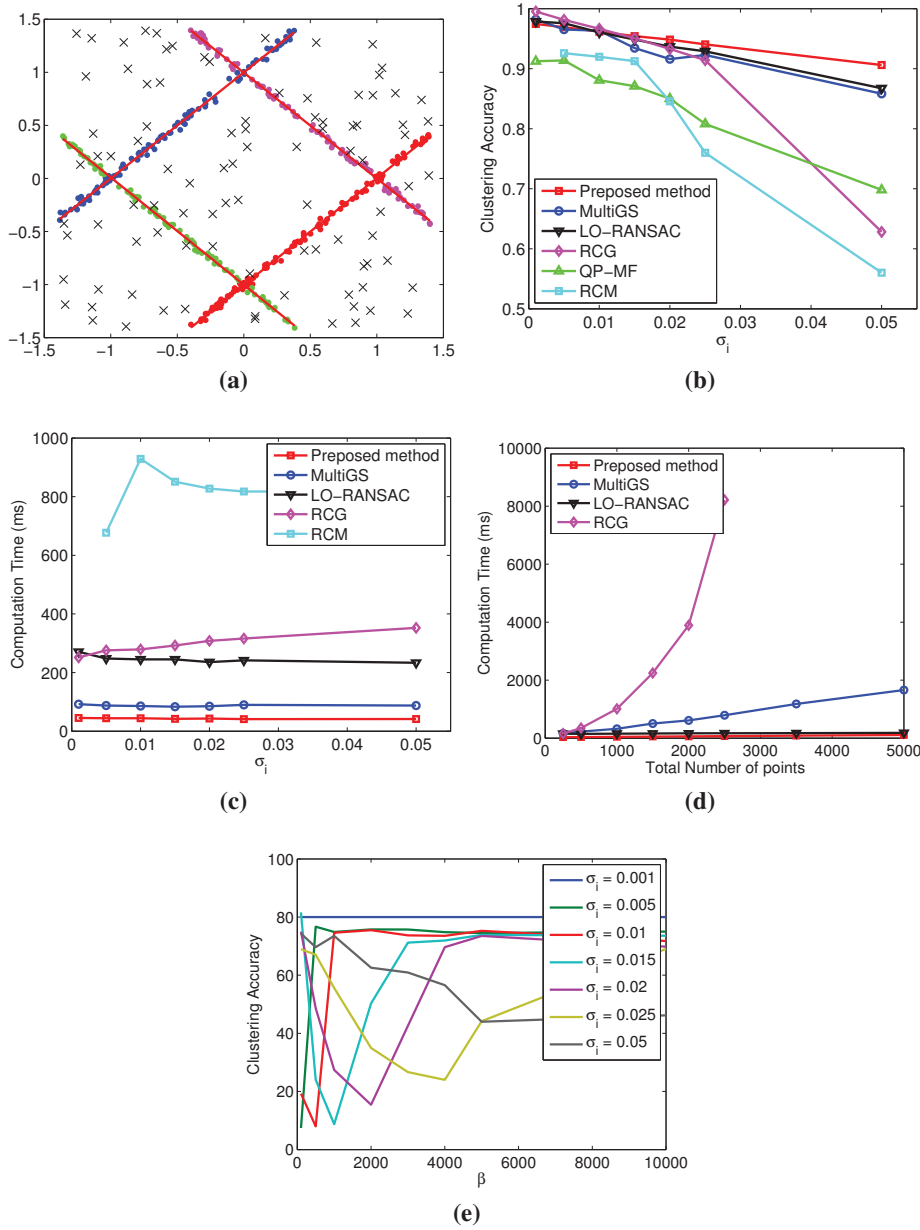


Figure 6.5: 2D Line fitting results.

example of a points set ($[n_i]_{i=0}^4 = 100, \sigma = 0.5$) is shown in Figure 6.6(a) with the clusters returned by the proposed method.

We assessed the performance of the proposed algorithm for varying levels of inlier noise in the interval $\sigma_i \in [0.1, 5.0]$. The number of data points was fixed at $[n_i]_{i=0}^4 = 100$. The clustering accuracy and the total computation time of the proposed method are shown in Figures 6.6(b) and 6.6(c) respectively, together with those of the competing methods. The proposed method again produced the best clustering accuracy with the

lowest computation time, and the improvement over MultiGS and LO-RANSAC is much larger than that of the 2D line fitting case.

As shown in Figure 6.6(e), the clustering accuracy for MultiGS and LO-RANSAC does not exceed that of the proposed method (for high noise values) even when the sampling times for those methods are increased to 10 to 25 times that of the proposed method.

In the next experiment, the number of total data points was varied in the range 250-2500 while the noise level was fixed at $\sigma_i = 0.2$. The results shown in Figure 6.6(d), indicate that the proposed method is very efficient in terms of computation time with slight linear increase in time with number of points, whereas the computation time of MultiGS increased exponentially with the size of the data.

6.4.3 Two-view motion segmentation

Two-view motion segmentation is the task of identifying the points corresponding to each object in two views of a dynamic scene that contains multiple independently moving objects. Provided that the point matches between the two views are given as , each motion can be modeled using the fundamental matrix $F \in \mathbb{R}^{3 \times 3}$ as Torr & Murray (1997):

$$X_1^\top F X_2 = 0 \quad (6.4.2)$$

The distance from a given model to a point pair can be measured using the Sampson distance (Hartley & Zisserman 2003).

First, we used the “box-book-mag” image pair from Schindler & Suter (2006) to evaluate the performance of the proposed and competing methods. The “box-book-mag” has two images of three independently moving objects together with feature correspondences. The performance of each algorithm was evaluated using clustering accuracy and computation time.

The clustering accuracy of the proposed method was observed while limiting the number of allowed random initializations to a specific value. The results are presented in Figure 6.7(a), and show the proposed method achieves high accuracy without the need for many random re-initializations. Then, we set the sampling times of the MultiGS and LO-RANSAC to the time taken by the proposed method at each step and recorded their accuracy. The results, plotted on the same figure, show that these algorithms take longer to achieve the same level of accuracy as the proposed method. For comparison, we also

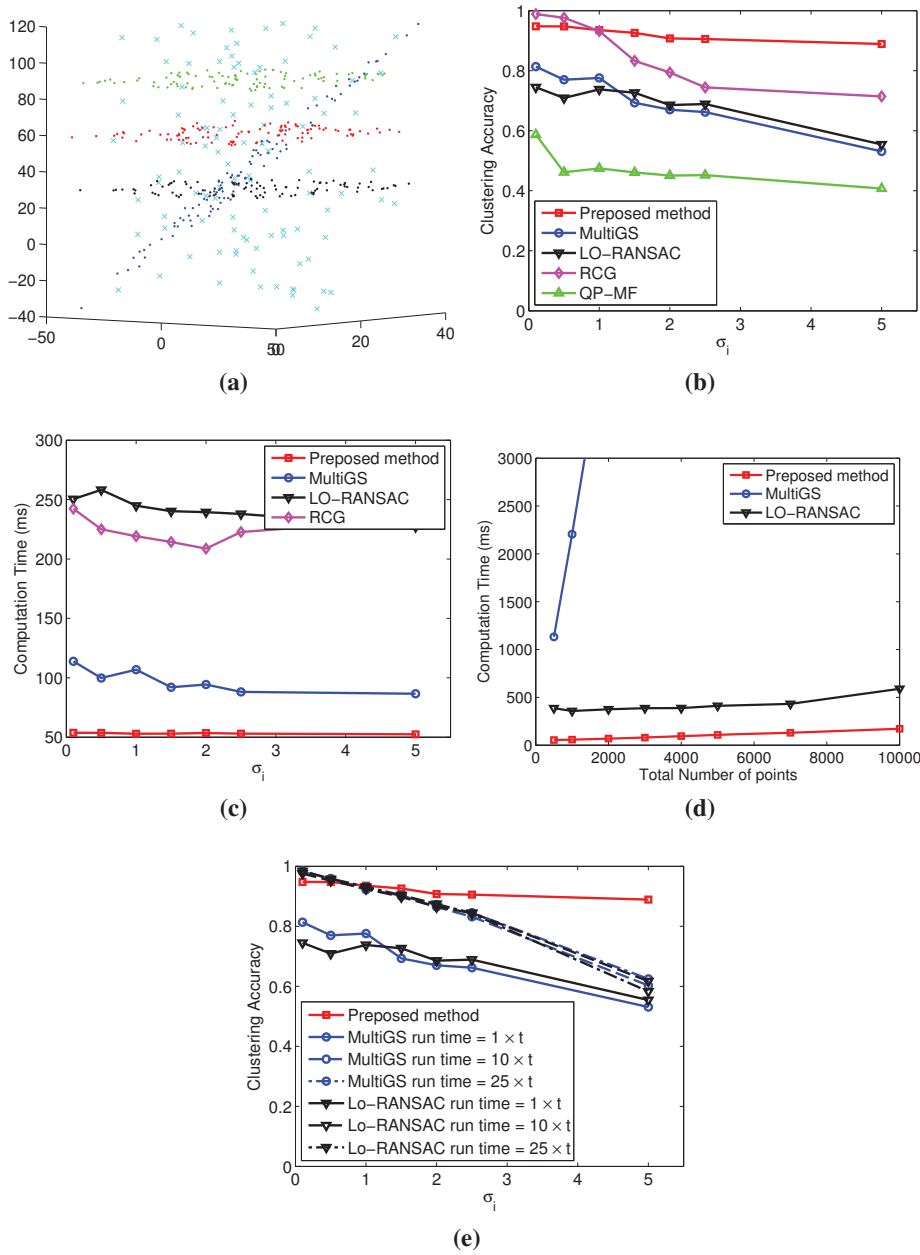


Figure 6.6: 3D plane fitting results.

included the results given by RCM. The results show that the proposed method is fast and can achieve better accuracy. Figures 6.7(b) and 6.7(c) show the inlier and outlier points identified by the proposed method, respectively.

The clustering results of the proposed method for the remaining sequences in dataset Schindler & Suter (2006) is shown in Figure 6.8. The results show that the proposed method has successfully identified the structures present in data with multiple structures and gross outliers.

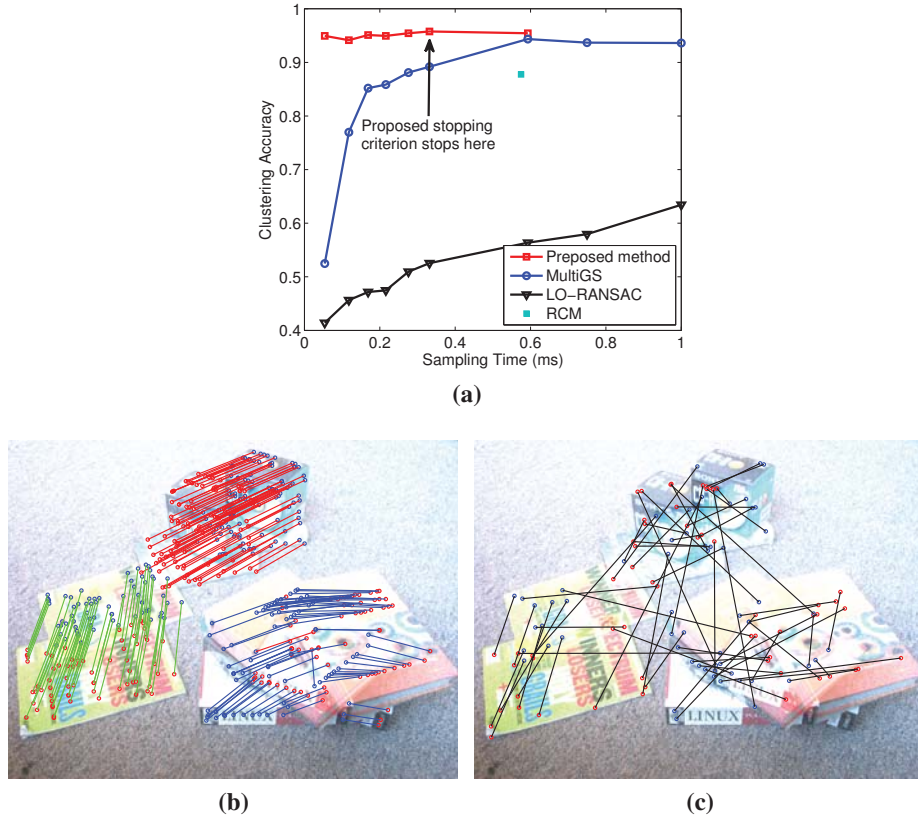


Figure 6.7: (a) Variation of clustering accuracy with sampling time (b) Inliers identified by the proposed method (c) Outliers identified by the proposed method.

6.4.4 Multi-homography detection

Assume that the point matches between two views of a static scene with multiple planar surfaces are given as $[X_1, X_2]$. Multi-homography detection aims to detect point matches arising from the same planar surface using a homography matrix $H \in \mathbb{R}^{3 \times 3}$ that relates the matching points via $X_1 \sim HX_2$. The distance from a data point to a given model can be measured using the Sampson's distance.

Similar to Pham et al. (2012) we test the performance of the proposed method on the AdelaideRMF dataset (Wong et al. 2011). The clustering accuracy of the proposed method together with RCM and multiGS is given in Table 6.1. Here, the sampling time of MultiGS method was set to ten times that of the proposed method.

The results show that the proposed method achieved high accuracy in a very short time compared to other methods. It should be noted here that the proposed method was not able to detect the two smallest (in terms of number of points) structures in *Johnsonb* dataset. This is due to the small number of points in each of those structures, which was

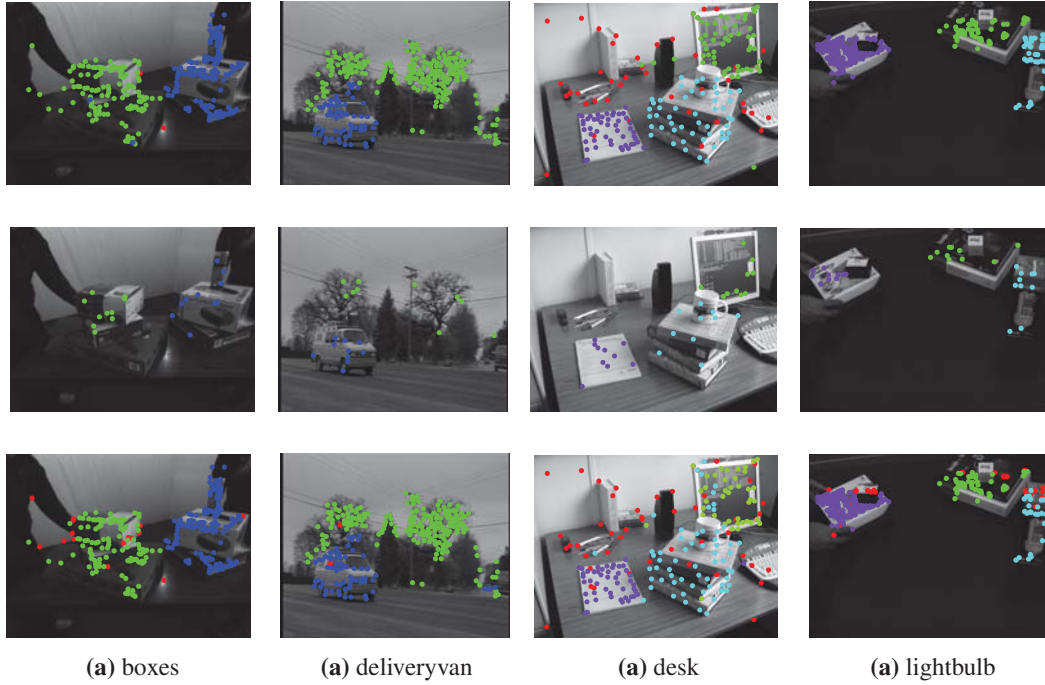


Figure 6.8: Two-view motion segmentation results for image sequences in dataset Schindler & Suter (2006). Line 1 shows the ground truth whereas lines 2 and 3 show the final HMSS samples selected and the clustering results of the proposed method. The outliers are marked in red.

20 and 15 respectively. RCM was also not able to detect these structures reliably. The analysis of Hoseinnezhad et al. (2010) showed that if the scale of inlier noise is not known a priori, its estimation requires at least 20 data points to limit the effects of finite sample bias.

To provide a qualitative measure of the performance of those methods, clustering results of the proposed method and RCM are compared with the ground truth in Figure 6.9. The first column shows that both methods were able to achieve good results on *union-*

Table 6.1: Multi-homography detection results. The time is given in milliseconds.

	RCM		MultiGS		Proposed Method	
	CA	Time	CA	Time	CA	Time
Johnsona	0.91	1300	0.69	5023	0.94	473
Johnsonb	0.89	2150	0.73	6958	0.85	663
Ladysymon	0.91	1010	0.89	2371	0.94	219
Neem	0.92	1020	0.74	3565	0.93	327
Oldclassicswing	0.98	950	0.92	2699	0.97	252
Sene	0.98	1220	0.99	2590	0.98	237

house having five structures. However, the proposed method was able to detect all six structures in the *bonhall* image (some incorrect classification of outliers as inliers can be seen) whereas RCM was not able to detect one of those structures. Due to the use of spatial consistency, RCM results are affected if the same structure is separated by either outliers or another structure as seen in images of *elderhall*, *barsmith* and *napiera* buildings. The proposed method does not assume the existence of spatial contiguity and was therefore able to achieve good clustering in those cases.

In these experiments, the parameter k of the proposed method was set to a value between 20 to 40 based on the number of points in each dataset.

6.4.5 3D-motion segmentation of rigid bodies

The objective of 3D motion segmentation is to identify multiple moving objects using point trajectories through a video sequence. If the projections (to the image plane) of N points tracked through F frames are available, $[x_{f\alpha}]_{\alpha=1\dots N}^{f=1\dots F} \in \mathbb{R}^2$ then Sugaya & Kanatani (2004) has shown that the point trajectories $P_\alpha = [x_{1\alpha}, y_{1\alpha}, x_{2\alpha}, \dots, x_{F\alpha}, y_{F\alpha}]^\top \in \mathbb{R}^{2F}$ that belong to a single rigid moving object are contained within a subspace of $rank \leq 4$, under the affine camera projection model. Hence, the problem of 3D motion segmentation can be reduced to a subspace clustering problem.

We utilized the commonly used “checkerboard” image sequence in the Hopkins 155 dataset (Tron & Vidal 2007) to evaluate our algorithm. This dataset contains trajectory information of 104 video sequences that are categorized into two main groups depending on the number of motions in each sequence (two or three motions).

One of the characteristics in subspace segmentation is that the dimension of the subspaces may vary between two and four, depending on the nature of the motions. The proposed method, which was not specifically developed to solve this problem (similar to most competing techniques) is not capable of identifying the number of dimensions of a given motion and requires this information as an input. In our implementation we have taken two approaches. In the first approach (PM T1), we set the dimensions of the subspaces to four and in the second we set the dimensions to three or four based on the average ground truth knowledge (not set for each sequence but to the whole sequence i.e. the subspace dimensions for all checkerboard three object sequences were set to $[3, 4, 4]$). The second approach (PM T2) is intended to demonstrate the accuracy of the method in

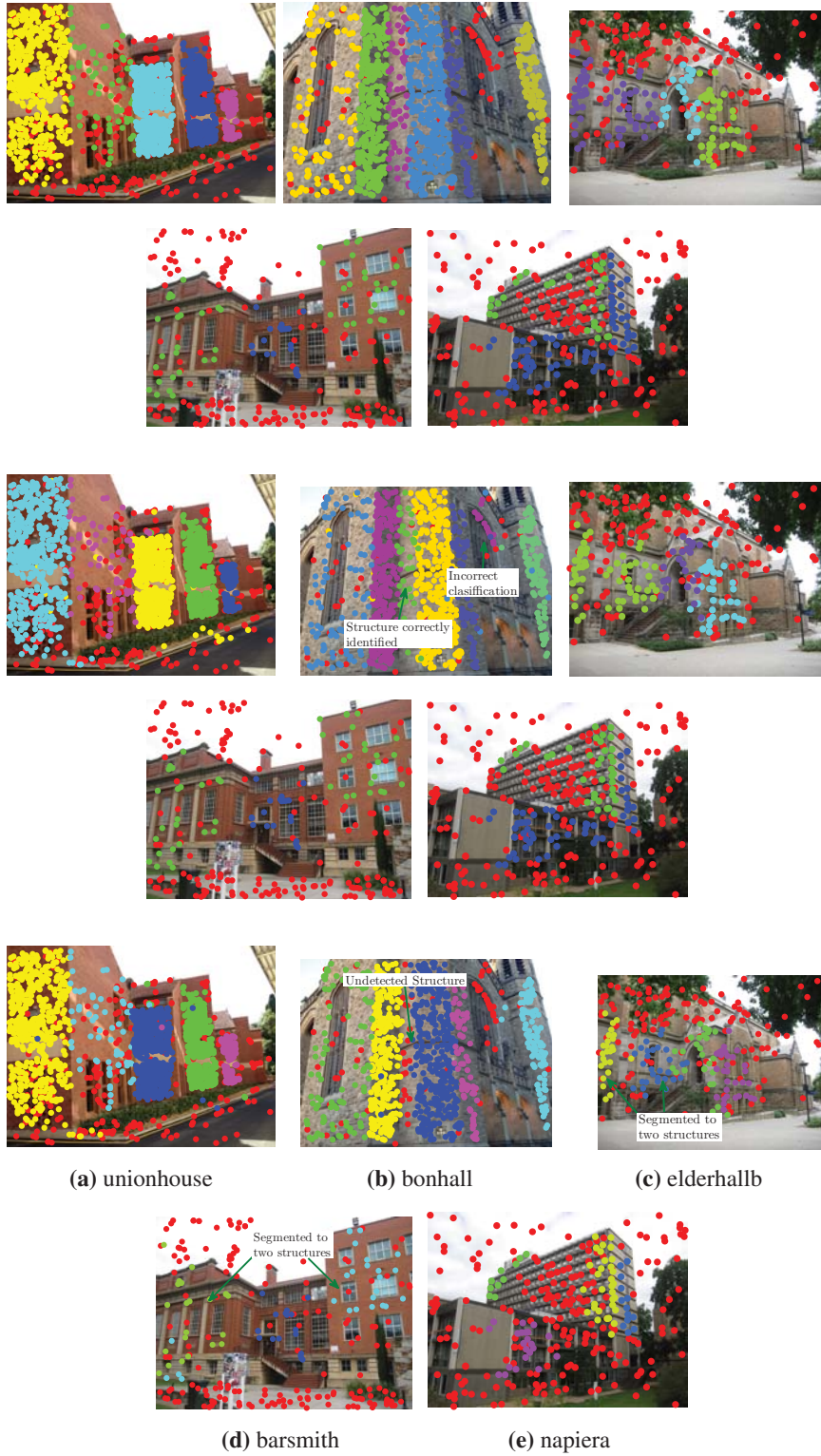


Figure 6.9: Qualitative results of multi-homography estimation. line 1 is the ground truth where as lines 2 and 3 are segmentation results of the proposed method and RCM respectively. The outliers are marked in red.

cases where an estimate of the subspace dimensions is available.

We compared our results with energy minimization and QP-MF⁴. For completeness we also included the results of the sparse subspace clustering (SSC) (Elhamifar & Vidal 2013) method which does not rely on dimensionality information. The results are shown in Table 6.2. The proposed method with fixed subspace dimensions ($\text{dim} = 4$) achieved better results than the competing model-fitting techniques (QP-MF, RANSAC, Energy minimization) but these results are not as good as SSC. However, when some information about the dimensionality of the subspaces was provided to the algorithm, our proposed algorithm achieved results comparable with the SSC.

Table 6.2: Percentage clustering Error of 3D motion segmentation.

	Reference	RANSAC	Enargy	QP-MF	SSC	PM T1	PM T2
2 Objects							
Mean	2.76	6.52	5.28	9.98	2.23	3.98	3.88
Median	0.49	1.75	1.83	1.38	0.00	0.00	0.00
3 Objects							
Mean	6.28	25.78	21.38	15.61	5.77	11.06	6.81
Median	5.06	26.01	21.14	8.82	0.95	1.20	1.04

To provide a qualitative measure of the performance the final segmentation results of several sequences in the Hopkins 155 dataset, where the proposed method was successful, are shown in Figure 6.10. These results show that the proposed method has been successful in a range of problems with different types of motion. We have also included some cases where the proposed method was not successful (Figure 6.11). Figures 6.11(a) and 6.11(b) show instances where a single motion has been segmented into two. This is because some motions are segmented in to multiple degenerate motions. Figures 6.11(c) and 6.11(d) show instances where a structure has very low number of points relative to the other structures and in these cases the proposed method would be expected to fail due to the finite sample bias in the inlier noise estimation.

6.5 Discussion

This chapter presents an accurate and efficient method that can be used to detect underlying structures in data contaminated with noise and outliers. The method uses higher than minimal samples to generate hypothesis and moves from one sample to another using

⁴The results published in Yu et al. (2011) are used in this comparison.

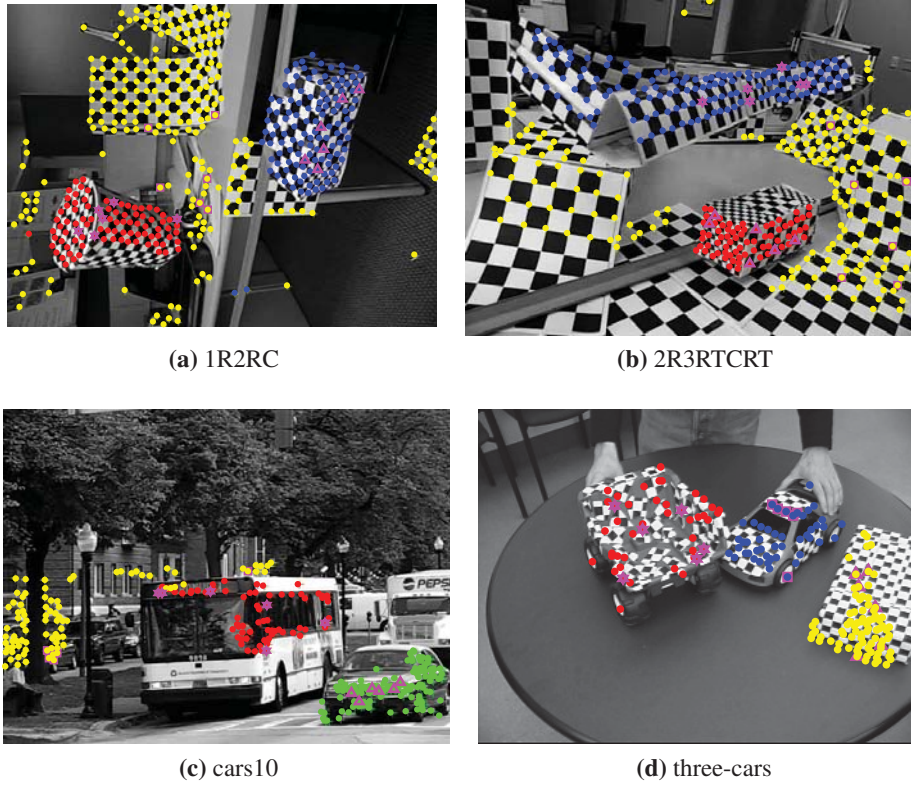


Figure 6.10: Segmentation result for several sequence in the Hopkins 155 dataset where the proposed method was successful. The sample used to generate the best hypothesis is also shown.

locally available information at each step while trying to minimize the cost function. The proposed method is very general and requires only a very few input parameters compared to the competing methods.

One of the parameters required by the proposed method is the value of k , which defines the minimal acceptable size for a structure in a given application. Any robust model fitting method needs to establish the minimal acceptable structure size (either explicitly or implicitly), or else it may result in a trivial solution. for example in 2D line fitting any two points will result in a perfect solution. The proposed method also assumes that the number of structures is known a prior. This is one of the weakness in the proposed method, however the problem of identifying the number of structures present and the scale of noise simultaneously is still a highly researched area with no good solutions. Remaining outliers can always be seen as members of a model with large noise values. Some model fitting methods that are based on energy minimization (DeLong et al. 2012, Pham et al. 2012) are devised to estimate the number of structures given the scale of noise.

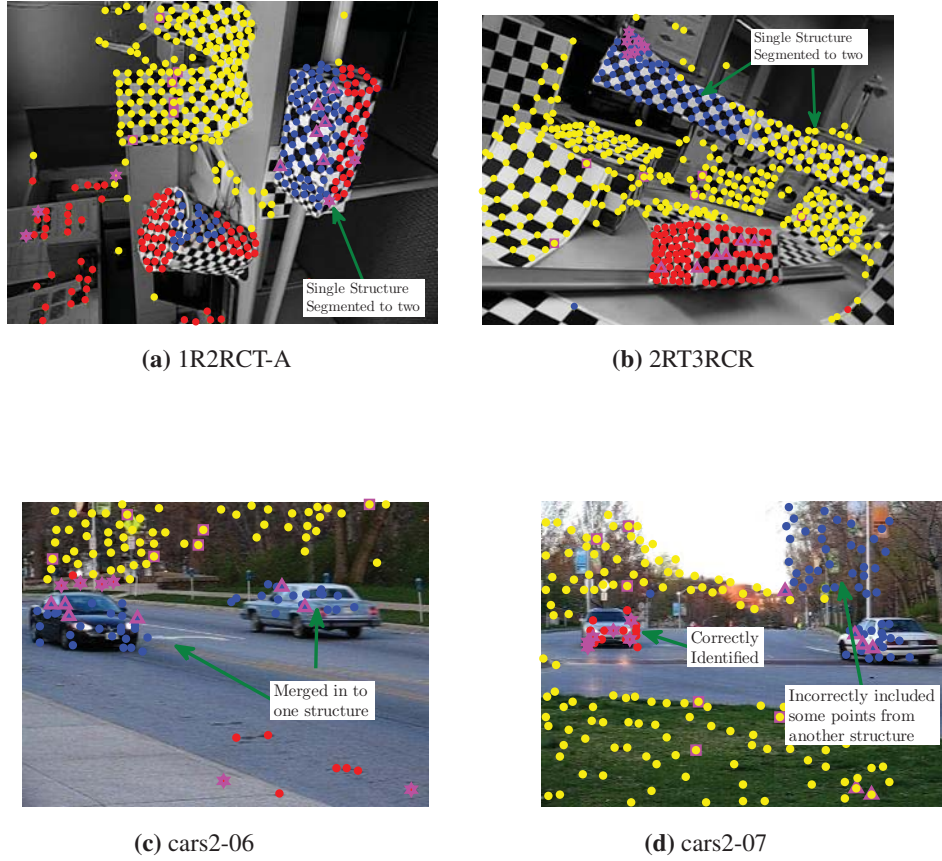


Figure 6.11: Segmentation result for several sequences in the Hopkins 155 dataset where the proposed method was not successful. The sample used to generate the best hypothesis is also shown.

They achieve this by adding a model complexity term to the cost function that penalize additional structures in a given solution. However, these methods require an additional parameter that balances the data fidelity cost with the model complexity (number of structures in RCM). Our experiments on RCM showed that the output of these methods were heavily dependent on this parameter and required hand tuning to generate reliable results (see Figure 6.5). The RCG method which is a hyper-graph-based method does not take the number of structures directly as an input. However, it may also result in multiple models representing the same structure and in order to remove these repetitions that method uses an additional pruning step that require the knowledge of separation between the structures (another parameter).

Knowledge on the scale of noise is essential for any robust model-fitting system to work. While most competing methods require this as an input, the proposed method

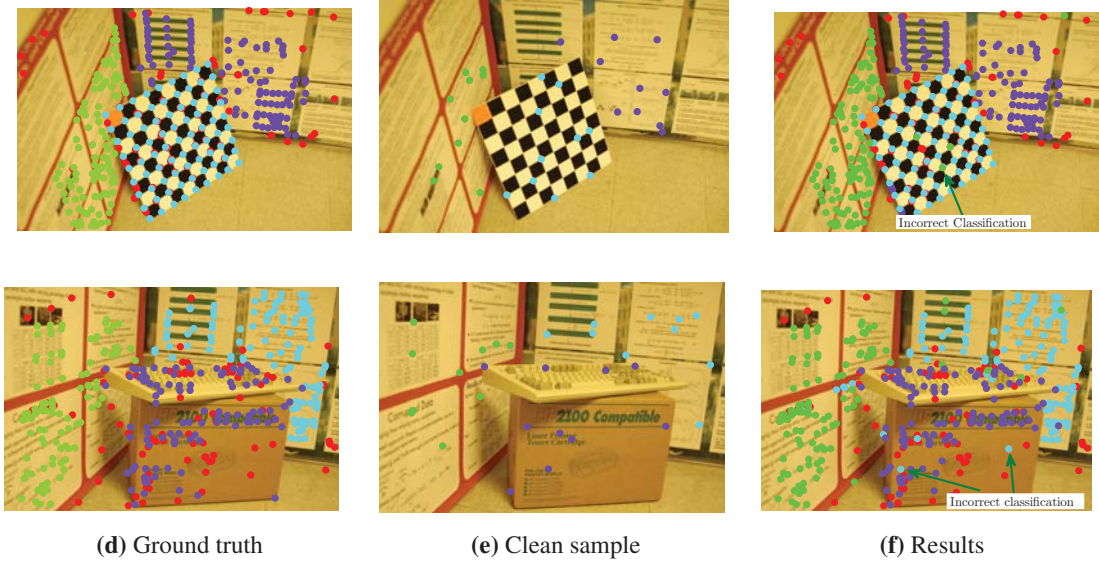


Figure 6.12: Two-view motion segmentation results for image sequences “*poster-checkerboard*” (line 1) and “*poster-keyboard*” (line 2) from Poling & Lerman (2014).

estimates the noise scale from data. The analysis of Hoseinnezhad et al. (2010) showed that if the scale of inlier noise is not known a priori, its estimation requires at least 20 data points to limit the effects of finite sample bias. Due to this the proposed method would fail to identify the underlying model if that structure contains only a small number of data points (< 20).

The MSSE algorithm which is used to estimate the scale of noise, requires the constant threshold T as an input. T defines the inclusion percentage of inliers based on a normal distribution for noise which is a number around 2.5, i.e. $T = 2.5$ will include 99% of inliers.

The proposed method does not employ any additional information such as the spatial contiguity of data in clustering. While this is advantageous in cases where the spatial contiguity is violated (see Figure 6.9), in some problems where a structure has spatial contiguity, not using such priors will result in slight degradation in the clustering. This problem is elaborated in Figure 6.12. The clean samples in the figure shows that the proposed algorithm has identified the underlying model correctly. However some points within one structure is clustered into another as these points are closer to that model. This problem can be eliminated by coupling the segmentation step with partial contiguity prior where applicable. However, we have not implemented such scheme in this chapter as the main aim is to detect the a underlying models.

6.6 Conclusion

In this chapter we first studied the usefulness of using higher than minimal subsets for hypothesis generation in parametric model fitting. The synthetic data experiments showed that using higher than minimal subset samples for hypothesis generation increases the probability of generating a good hypothesis close to the true model, given that it is a clean sample. The experiments also showed that the probability increase is only significant up to few additional data points. However, the probability of selecting a clean sample decreases with the increase of the sample size, making it improbable to extend the commonly used sampling methods to accommodate higher than minimal samples.

This chapter presents a new approach to parametric model fitting that uses higher than minimal subset sampling to generate hypotheses. The proposed method, equivalent to Newton's optimization method, starts from an arbitrary set of parameters and moves towards a dense structure in data using the local information present at each iteration. The algorithm also consists of a stopping criterion that identifies when a dense cluster in data is reached and stops the iterations saving computation time.

The performance of the algorithm in terms of accuracy and computational efficiency was evaluated on several models-fitting problems including line/plane fitting, two-view motion segmentation and 3D motion segmentation. The results of the proposed method were compared with state-of-the-art model fitting techniques. The comparisons showed that the proposed method is both highly accurate and computationally efficient.

Chapter 7

Conclusion and Future Work

The main objectives of this dissertation are to study the fundamental problem of automatic registration and motion estimation of volumetric medical images and improve the performance of registration in terms of both robustness and computational efficiency. Registration is a powerful tool used in many medical image analysis tasks such as disease diagnosis, progress monitoring and treatment planning. Non-rigid registration techniques are commonly used for medical images as they involve complex motions which cannot be captured using rigid techniques with low degrees of freedom.

In Chapter 2, we studied the non-rigid registration techniques applied in medical imaging applications. Although the methods vary, a typical registration algorithm consists of a similarity measure, transformation model and an optimization strategy. Once the overall problem is segmented into these categories, techniques used in each of them were analysed extensively.

Next, in Chapter 3 we explored the smoothness requirement of motion estimation from a robust statistical perspective. Here, smoothing refers to the size of the locality upon which the estimation is based. In this study a new approach to quantifying the minimum required smoothing based on the concept of the finite sample bias of a robust estimator is proposed. The proposed approach is very general and makes predictions about the amount of smoothness required to satisfy the sufficiency condition for a broad range of visual estimation tasks such as optic flow calculation. The theoretical analysis predicts that smoothing over a cubic area as small as 5 to 7 voxels wide is sufficient to achieve the highest practical accuracy. A robust 3D optic flow, in which the imposition of smoothing can be locally quantized, was then devised to test the proposed hypothesis

on the smoothing requirement. The hypothesis was then tested using a geometrically realistic synthetic CT image sequence of the breathing lung and five cases of real 4D CT lung images with extensive sets of expert annotated landmarks. As predicted, the results showed that only a very small amount of local smoothing is required to achieve high accuracy and that in some cases increasing the amount of smoothing reduces the quality of the results.

In Chapter 4 we focused on image information-based non-rigid image registration methods. Due to the enormous amount of information to be processed and the complexity in the transformation models (the large number of parameters) these methods tend to be computationally expensive. We propose a new approach that uses robust segmentation of squared intensity differences to intelligently identify the most relevant information for the registration and use them in a stochastic optimization framework. The method also incorporates a computationally-efficient means of measuring the quality of registration. Since the registration schemes are all iterative, this measure is highly useful for deciding when to stop the registration process. The performance of the proposed registration algorithm in terms of both accuracy and efficiency was extensively evaluated using two real CT image sets with ground truth and the results were compared with those from several available techniques. The experiments showed that the proposed method substantially improves both the accuracy and the computational complexity of the registration task.

Non-rigid registration algorithms use TPS-based data modelling techniques to model transformation in feature-based image registrations, smoothing of motion fields and interpolating intensity fields. However, they tend to be inaccurate, especially when the data contain outliers. Chapter 5 presents a new algorithm for modelling visual data using robust thin plate splines. The fitting algorithm uses a variant of the least kth order statistics fitting approach with a novel iterative method for performing the inlier/outlier segmentation. The proposed method was tested on synthetic and real data and was shown to be effective even when numerous structured outliers (pseudo-outliers) are present.

In Chapter 6 we studied the usefulness of using higher than minimal subsets for hypothesis generation in parametric model fitting. The synthetic data experiments presented showed that using higher than minimal subset samples for hypothesis generation increases the probability of generating a good hypothesis close to the true model, given that it is a clean sample. These experiments also showed that the probability increase is only

significant up to few additional points. However, the probability of selecting a clean sample decreases with the increase of the sample size, making it improbable to extend the commonly-used sampling methods to accommodate higher than minimal samples. Hence, we presented an efficient approach for parametric model fitting that uses higher than minimal subset sampling to generate hypotheses. The proposed method equivalent to a Newton optimization method starts from an arbitrary set of parameters and moves towards a dense structure in data using the local information present at each iteration. The algorithm also consists of a stopping criterion that identifies when a dense cluster in data is reached and stops the iterations, saving computation time. The performance of the algorithm in terms of accuracy and computational efficiency was evaluated on several model-fitting problems including line/plane fitting, two-view motion segmentation and 3D motion segmentation. The results of the proposed method were compared with the state of the art model fitting techniques. The comparisons showed that the proposed method is both highly accurate and computationally efficient.

Future Work

- Optical flow estimation is known to be not very accurate in cases with large deformations due to the existence of local optima. Hierarchical strategies are commonly used to overcome this limitation. Implementation of a hierarchical strategy such as coarse to fine estimation or affine preregistration for the local optical flow calculations in Chapter 3 would have improved the registration accuracy of the images with large deformations (Cases 3-5 in the DIR dataset). However, this was not carried out in the present study as the objective was not to propose a new registration method but to quantify the smoothing requirement from a robust statistical perspective.
- In terms of improving the accuracy of the local optical flow estimation, Sun et al. (2010) found that spatial smoothing of the estimates using a median filter was the single most significant step. Using such a technique on the estimated deformation vectors would have further improved the results presented in Chapter 3. However, such an approach would have made it impossible to directly compare the relationship between the smoothing and estimation accuracy. Hence, it was not implemented in our study.

- The non-rigid registration method proposed in Chapter 4 uses the intensity difference in selecting the samples for derivative estimations. However, changes in intensity of lung tissue due to breathing motion can affect the performance of this method. The use of a mass-preserving intensity model that adjusts intensity according to the volume expansion may improve results over those presented in this thesis.
- Due to the use of intensity differences in the sample selection process, the registration method proposed in Chapter 4 is limited to the registration of mono-modal images. Development of appropriate clustering methods to identify the almost registered group of voxels from a joint histogram may enable this method to be extended to registering images from different modalities. However, this path is yet to be explored.
- A common technique used to reduce the computational time of an algorithm is to implement it using a graphical processing unit (GPU). One of the main requirements for accelerating an algorithm using GPUs is that it has to be parallelizable. The non-rigid registration algorithm presented in Chapter 4 uses B-spline as the transformation model and because of the compact support region of these spline kernels and the shared memory on GPUs, the gradient estimation can be easily parallelized to be implemented on a GPU. Such implementation would further reduce the computation time of the proposed algorithm.
- The experiments in Chapter 4 were run on thoracic CT images where no significant anomalies exist within the region of interest. It would be beneficial to evaluate the proposed registration method on images acquired from patients with anomalies in the lung. To our knowledge, such datasets with adequate ground truth do not exist and defining the registration accuracy in such cases would be challenging.
- The model fitting technique presented in Chapter 5 has been evaluated only on synthetic data and range images. The contribution of such a method in image registration-related applications has not yet been explored.

References

- Agarwal, S., Lim, J., Zelnik-Manor, L., Perona, P., Kriegman, D. & Belongie, S. (2005), Beyond pairwise clustering, *in* 'IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings CVPR 2005.', Vol. 2, IEEE, pp. 838–845.
- Andreopoulos, A. & Tsotsos, J. K. (2008), 'Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI', *Medical Image Analysis* **12**(3), 335–357.
- Andronache, A., Cattin, P. & Székely, G. (2006), Local intensity mapping for hierarchical non-rigid registration of multi-modal images using the cross-correlation coefficient, *in* J. Pluim, B. Likar & F. A. Gerritsen, eds, 'Biomedical Image Registration', Vol. 4057 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 26–33.
- Archip, N., Clatz, O., Whalen, S., Kacher, D., Fedorov, A., Kot, A., Chrisochoides, N., Jolesz, F., Golby, A., Black, P. M. & Warfield, S. K. (2007), 'Non-rigid alignment of pre-operative MRI, fMRI, and DT-MRI with intra-operative MRI for enhanced visualization and navigation in image-guided neurosurgery', *NeuroImage* **35**(2), 609 – 624.
- Arsigny, V., Pennec, X. & Ayache, N. (2005), 'Polyrigid and polyaffine transformations: a novel geometrical tool to deal with non-rigid deformations-application to the registration of histological slices', *Medical Image Analysis* **9**(6), 507–523.
- Arzhaeva, Y., Prokop, M., Murphy, K., Van Rikxoort, E. M., De Jong, P. A., Gietema, H. A., Viergever, M. A. & Van Ginneken, B. (2010), 'Automated estimation of progression of interstitial lung disease in CT images', *Medical Physics* **37**(1), 63–73.
- Audette, M. A., Ferrie, F. P. & Peters, T. M. (2000), 'An algorithmic overview of surface registration techniques for medical imaging', *Medical Image Analysis* **4**(3), 201–217.
- Bab-Hadiashar, A. & Hoseinnezhad, R. (2008), Bridging parameter and data spaces for fast robust estimation in computer vision, *in* 'Digital Image Computing: Techniques and Applications, 2008. DICTA'08.', IEEE, pp. 1–8.
- Bab-Hadiashar, A. & Suter, D. (1998), 'Robust optic flow computation', *International Journal of Computer Vision* **29**(1), 59–77.
- Bab-Hadiashar, A. & Suter, D. (1999), 'Robust segmentation of visual data using ranked unbiased scale estimate', *Robotica* **17**(6), 649–660.
- Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M. & Szeliski, R. (2011), 'A database and evaluation methodology for optical flow', *International Journal of Computer Vision* **92**(1), 1–31.

- Barron, J. L., Fleet, D. J., Beauchemin, S. S. & Burkitt, T. A. (1992), Performance of optical flow techniques, in 'IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92.', pp. 236–242.
- Barron, J. L. & Thacker, N. A. (2005), 'Tutorial: Computing 2D and 3D optical flow', *Imaging Science and Biomedical Engineering Division, Medical School, University of Manchester*.
- Bhagalia, R., Fessler, J. A. & Boklye, K. (2009), 'Accelerated nonrigid intensity-based image registration using importance sampling', *IEEE Transactions on Medical Imaging* **28**(8), 1208–1216.
- Boldea, V., Sharp, G. C., Jiang, S. B. & Sarrut, D. (2008), '4D-CT lung motion estimation with deformable registration: Quantification of motion nonlinearity and hysteresis', *Medical Physics* **35**(3), 1008–1018.
- Bookstein, F. L. (1999), Linear methods for nonlinear maps: Procrustes fits, thin-plate splines, and the biometric analysis of shape variability, in A. W. Toga, ed., 'Brain Warping', Academic Press, San Diego, chapter 10, pp. 157–182.
- Boykov, Y., Veksler, O. & Zabih, R. (2001), 'Fast approximate energy minimization via graph cuts', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**(11), 1222–1239.
- Bro-Nielsen, M. & Gramkow, C. (1996), Fast fluid registration of medical images, in K. H. Höhne & R. Kikinis, eds, 'Visualization in Biomedical Computing', Vol. 1131 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 265–276.
- Broit, C. (1981), Optimal Registration of Deformed Images, PhD thesis, Philadelphia, PA, USA. AAI8207933.
- Bruhn, A., Weickert, J., Feddern, C., Kohlberger, T. & Schnörr, C. (2003), *Real-Time Optic Flow Computation with Variational Methods*, Vol. 2756 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 222–229.
- Bruhn, A., Weickert, J., Kohlberger, T. & Schnörr, C. (2005), *Discontinuity-Preserving Computation of Variational Optic Flow in Real-Time*, Vol. 3459 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 279–290.
- Bruhn, A., Weickert, J. & Schnörr, C. (2005), 'Lucas/kanade meets horn/schunck: Combining local and global optic flow methods', *International Journal of Computer Vision* **61**(3), 211–231.
- Buhmann, M. D. (2003), *Radial basis functions: theory and implementations*, Vol. 12, Cambridge University Press.
- Burschka, D., Li, M., Taylor, R. & Hager, G. (2004), Scale-invariant registration of monocular endoscopic images to CT-scans for sinus surgery, in C. Barillot, D. Haynor & P. Hellier, eds, 'Medical Image Computing and Computer-Assisted Intervention MICCAI 2004', Vol. 3217 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 413–421.

- Cahill, N. D., Schnabel, J. A., Noble, J. A. & Hawkes, D. J. (2008), Revisiting overlap invariance in medical image alignment, in 'IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08.', IEEE, pp. 1–8.
- Castillo, E., Castillo, R., Martinez, J., Shenoy, M. & Guerrero, T. (2010), 'Four-dimensional deformable image registration using trajectory modeling', *Physics in Medicine and Biology* **55**(1), 305.
- Castillo, E., Castillo, R., White, B., Rojo, J. & Guerrero, T. (2012), 'Least median of squares filtering of locally optimal point matches for compressible flow image registration', *Physics in Medicine and Biology* **57**(15), 4827–4833.
- Castillo, E., Castillo, R., Zhang, Y. & Guerrero, T. (2009), 'Compressible image registration for thoracic computed tomography images', *Journal of Medical and Biological Engineering* **29**(5), 222–233.
- Castillo, R., Castillo, E., Guerra, R., Johnson, V. E., McPhail, T., Garg, A. K. & Guerrero, T. (2009), 'A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets', *Physics in Medicine and Biology* **54**(7), 1849.
- Chen, C. & Li, Y. (2012), 'A robust method of thin plate spline and its application to DEM construction', *Computers and Geosciences* **48**(0), 9 – 16.
- Chen, C. & Li, Y. (2013), 'A robust multiquadric method for digital elevation model construction', *Mathematical Geosciences* **45**, 297–319.
- Chin, T.-J., Yu, J. & Suter, D. (2010), Accelerated hypothesis generation for multi-structure robust fitting, in K. Daniilidis, P. Maragos & N. Paragios, eds, 'Computer Vision - ECCV 2010', Vol. 6315 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 533–546.
- Chin, T.-J., Yu, J. & Suter, D. (2012), 'Accelerated hypothesis generation for multistructure data via preference analysis', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**, 625–638.
- Christensen, G. (1999), Consistent linear-elastic transformations for image matching, in A. Kuba, M. Šámal & A. Todd-Pokropek, eds, 'Information Processing in Medical Imaging', Vol. 1613 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 224–237.
- Christensen, G. E., Rabbitt, R. D. & Miller, M. I. (1994), '3D brain mapping using a deformable neuroanatomy', *Physics in Medicine and Biology* **39**(3), 609.
- Christensen, G. E., Rabbitt, R. D. & Miller, M. I. (1996), 'Deformable templates using large deformation kinematics', *IEEE Transactions on Image Processing* **5**(10), 1435–1447.
- Christensen, G. & He, J. (2001), Consistent nonlinear elastic image registration, in 'IEEE Workshop on Mathematical Methods in Biomedical Image Analysis, 2001. MMBIA 2001.', IEEE, pp. 37–43.

- Chum, O., Matas, J. & Kittler, J. (2003), Locally optimized RANSAC, in B. Michaelis & G. Krell, eds, 'Pattern Recognition', Vol. 2781 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 236–243.
- Chun, S. Y. & Fessler, J. A. (2008), Regularized methods for topology-preserving smooth nonrigid image registration using B-spline basis, in '5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008.', IEEE, pp. 1099–1102.
- Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P. & Marchal, G. (1995), 'Automated multi-modality image registration based on information theory', *Information Processing in Medical Imaging* **3**, 263–274.
- Corpetti, T., Mmin, t. & Prez, P. (2002), 'Dense estimation of fluid flows', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(3), 365–380.
- Coselmon, M. M., Balter, J. M., McShan, D. L. & Kessler, M. L. (2004), 'Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines', *Medical Physics* **31**(11), 2942–2948.
- D'Agostino, E., Maes, F., Vandermeulen, D. & Suetens, P. (2004), Non-rigid atlas-to-image registration by minimization of class-conditional image entropy, in C. Barillot, D. Haynor & P. Hellier, eds, 'Medical Image Computing and Computer-Assisted Intervention - MICCAI 2004', Vol. 3216 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 745–753.
- Damas, S., Cordón, O. & Santamaría, J. (2011), 'Medical image registration using evolutionary computation: An experimental survey', *IEEE Computational Intelligence Magazine* **6**(4), 26–42.
- DeLong, A., Osokin, A., Isack, H. N. & Boykov, Y. (2012), 'Fast approximate energy minimization with label costs', *International Journal of Computer Vision* **96**(1), 1–27.
- Demšar, J. (2006), 'Statistical comparisons of classifiers over multiple data sets', *The Journal of Machine Learning Research* **7**, 1–30.
- Devroye, L. (1986), General principles in random variate generation, in 'Non-Uniform Random Variate Generation', Springer New York, pp. 27–82.
- Dupuis, P. & Simha, R. (1991), 'On sampling controlled stochastic approximation', *IEEE Transactions on Automatic Control* **36**(8), 915–924.
- Ehrhardt, J. & Lorenz, C. (2013), *4D modeling and estimation of respiratory motion for radiation therapy*, Biological and Medical Physics, Biomedical Engineering, Springer Berlin Heidelberg.
- Ehrhardt, J., Werner, R., Schmidt-Richberg, A. & Handels, H. (2011), 'Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration', *IEEE Transactions on Medical Imaging* **30**(2), 251–265.
- Eklund, A., Dufort, P., Forsberg, D. & LaConte, S. M. (2013), 'Medical image processing on the GPU-past, present and future', *Medical Image Analysis* **17**(8), 1073 – 1094.

- Elhamifar, E. & Vidal, R. (2013), 'Sparse subspace clustering: Algorithm, theory, and applications', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**(11), 2765–2781.
- Fan, L. & Pylvänäinen, T. (2009), Adaptive sample consensus for efficient random optimization, in G. Bebis, R. Boyle, B. Parvin, D. Koracin, Y. Kuno, J. Wang, R. Pajarola, P. Lindstrom, A. Hinkenjann, M. Encarnação, C. T. Silva & D. Coming, eds, 'Advances in Visual Computing', Vol. 5876 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 252–263.
- Fischer, B. & Modersitzki, J. (2003), FLIRT: A flexible image registration toolbox, in J. Gee, J. Maintz & M. Vannier, eds, 'Biomedical Image Registration', Vol. 2717 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 261–270.
- Fischler, M. A. & Bolles, R. C. (1981), 'Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography', *Communications of the ACM* **24**(6), 381–395.
- Fleet, D. J. & Jepson, A. D. (1990), 'Computation of component image velocity from local phase information', *International Journal of Computer Vision* **5**(1), 77–104.
- Fluck, O., Vetter, C., Wein, W., Kamen, A., Preim, B. & Westermann, R. (2011), 'A survey of medical image registration on graphics hardware', *Computer Methods and Programs in Biomedicine* **104**(3), e45–e57.
- Ford, D. & Fulkerson, D. R. (2010), *Flows in networks*, Princeton University Press.
- Gee, J. C., Haynor, D. R., Reivich, M. & Bajcsy, R. K. (1994), Finite element approach to warping of brain images, in 'Medical Imaging', International Society for Optics and Photonics, pp. 327–337.
- Gerber, N., Gavaghan, K. A., Bell, B. J., Williamson, T. M., Weisstanner, C., Caversaccio, M. D. & Weber, S. (2013), 'High-accuracy patient-to-image registration for the facilitation of image-guided robotic microsurgery on the head', *IEEE Transactions on Biomedical Engineering* **60**(4), 960–968.
- Giger, M. L., Chan, H.-P. & Boone, J. (2008), 'Anniversary paper: history and status of CAD and quantitative image analysis: the role of medical physics and AAPM', *Medical Physics* **35**(12), 5799–5820.
- Glocker, B., Komodakis, N., Tziritas, G., Navab, N. & Paragios, N. (2008), 'Dense image registration through MRFs and efficient linear programming', *Medical Image Analysis* **12**(6), 731–741.
- Glocker, B., Sotiras, A., Komodakis, N. & Paragios, N. (2011), 'Deformable medical image registration: Setting the state of the art with discrete methods', *Annual Review of Biomedical Engineering* **13**, 219–244.
- Gorbunova, V., Sporring, J., Lo, P., Loeve, M., Tiddens, H. A., Nielsen, M., Dirksen, A. & de Bruijne, M. (2012), 'Mass preserving image registration for lung CT', *Medical Image Analysis* **16**(4), 786–795.

- Gu, X., Pan, H., Liang, Y., Castillo, R., Yang, D., Choi, D., Castillo, E., Majumdar, A., Guerrero, T. & Jiang, S. B. (2010), 'Implementation and evaluation of various demons deformable image registration algorithms on a GPU', *Physics in Medicine and Biology* **55**(1), 207.
- Guo, W., Li, Q., Boyce, S. J., McAdams, H. P., Shiraishi, J., Doi, K. & Samei, E. (2012), 'A computerized scheme for lung nodule detection in multiprojection chest radiography', *Medical Physics* **39**(4), 2001–2012.
- Haber, E. & Modersitzki, J. (2006), Intensity gradient based registration and fusion of multi-modal images, in R. Larsen, M. Nielsen & J. Sporring, eds, 'Medical Image Computing and Computer-Assisted Intervention - MICCAI 2006', Vol. 4191 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 726–733.
- Haber, E. & Modersitzki, J. (2007), 'Image registration with guaranteed displacement regularity', *International Journal of Computer Vision* **71**(3), 361–372.
- Hairong, L. & Shuicheng, Y. (2012), Efficient structure detection via random consensus graph, in 'IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012', pp. 574–581.
- Hartley, R. & Zisserman, A. (2003), *Multiple view geometry in computer vision*, Cambridge university press.
- Hero, A., Ma, B., Michel, O. J. & Gorman, J. (2002), 'Applications of entropic spanning graphs', *IEEE Signal Processing Magazine* **19**(5), 85–95.
- Hesami, R., Bab-Hadiashar, A. & HosseinNezhad, R. (2010), 'Range segmentation of large building exteriors: A hierarchical robust approach', *Computer Vision and Image Understanding* **114**(4), 475–490.
- Hesterberg, T. C. (2003), Advances in importance sampling, PhD thesis, Dept. of Statistics, Stanford University.
- Hevia-Montiel, N., Jimenez-Alaniz, J. R., Medina-Banuelos, V., Yanez-Suarez, O., Rosso, C., Samson, Y. & Baillet, S. (2007), Robust nonparametric segmentation of infarct lesion from diffusion-weighted MR images, in '29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007. EMBS 2007.', IEEE, pp. 2102–2105.
- Hill, D. L., Batchelor, P. G., Holden, M. & Hawkes, D. J. (2001), 'Medical image registration', *Physics in Medicine and Biology* **46**(3), R1.
- Holden, M. (2008), 'A review of geometric transformations for nonrigid body registration', *IEEE Transactions on Medical Imaging* **27**(1), 111–128.
- Holm, S. (1979), 'A simple sequentially rejective multiple test procedure', *Scandinavian Journal of Statistics* pp. 65–70.
- Horn, B. K. P. & Schunck, B. G. (1981), 'Determining optical flow', *Artificial Intelligence* **17**, 185–203.

- Hoseinnezhad, R. & Bab-Hadiashar, A. (2007), 'Consistency of robust estimators in multi-structural visual data segmentation', *Pattern Recogn.* **40**(12), 3677–3690.
- Hoseinnezhad, R., Bab-Hadiashar, A. & Suter, D. (2010), 'Finite sample bias of robust estimators in segmentation of closely spaced structures: A comparative study', *Journal of Mathematical Imaging and Vision* **37**(1), 66–84.
- Hsieh, J. (2009), *Computed tomography: principles, design, artifacts, and recent advances*, SPIE Bellingham, WA.
- Hua, X., Hibar, D. P., Ching, C. R., Boyle, C. P., Rajagopalan, P., Gutman, B. A., Leow, A. D., Toga, A. W., Jr., C. R. J., Harvey, D., Weiner, M. W. & Thompson, P. M. (2013), 'Unbiased tensor-based morphometry: Improved robustness and sample size estimates for Alzheimer's disease clinical trials', *NeuroImage* **66**, 648 – 661.
- Huang, C. H., Hsieh, C. H., Lee, J. D., Huang, W. C., Lee, S. T., Wu, C. T., Sun, Y. N. & Wu, Y. T. (2012), 'A CT-ultrasound-coregistered augmented reality enhanced image-guided surgery system and its preliminary study on brain-shift estimation', *Journal of Instrumentation* **7**(08), P08016.
- Hutchinson, M. (1995), 'Interpolating mean rainfall using thin plate smoothing splines', *International Journal of Geographical Information Systems* **9**(4), 385–403.
- Illán, I. A., Grriz, J. M., Ramrez, J., Segovia, F., Jimnez-Hoyuela, J. M. & Ortega Lozano, S. J. (2012), 'Automatic assistance to parkinsons disease diagnosis in DaTSCAN SPECT imaging', *Medical Physics* **39**(10), 5971–5980.
- Jacobson, T. J. & Murphy, M. J. (2011), 'Optimized knot placement for B-splines in deformable image registration', *Medical Physics* **38**(8), 4579–4582.
- Jaya, J. & Thanushkodi, K. (2011), 'Implementation of computer aided diagnosis system based on parallel approach of Ant based medical image segmentation', *Journal of Computer Science* **7**(2), 291–297.
- Johnson, H. & Christensen, G. (2001), Landmark and intensity-based, consistent thin-plate spline image registration, in M. Insana & R. Leahy, eds, 'Information Processing in Medical Imaging', Vol. 2082 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 329–343.
- Karri, S., Charonko, J. & Vlachos, P. P. (2009), 'Robust wall gradient estimation using radial basis functions and proper orthogonal decomposition (POD) for particle image velocimetry (PIV) measured fields', *Measurement Science and Technology* **20**(4), 045401.
- Kesten, H. (1958), 'Accelerated stochastic approximation', *The Annals of Mathematical Statistics* **29**(1), 41–59.
- Kim, J. (2004), Intensity based image registration using robust similarity measure and constrained optimization: applications for radiation therapy, PhD thesis, The University of Michigan.
- Klein, S. & Staring, M. (2011), 'elastix the manual'.

- Klein, S., Staring, M., Murphy, K., Viergever, M. A. & Pluim, J. (2010), 'elastix: A toolbox for intensity-based medical image registration', *IEEE Transactions on Medical Imaging* **29**(1), 196–205.
- Klein, S., Staring, M. & Pluim, J. P. W. (2007), 'Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-Splines', *IEEE Transactions on Image Processing* **16**(12), 2879–2890.
- Klein, S., van der Heide, U. A., Lips, I. M., van Vulpen, M., Staring, M. & Pluim, J. P. (2008), 'Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information', *Medical Physics* **35**(4), 1407–1417.
- Kleinberg, J. & Tardos, E. (2002), 'Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields', *Journal of the ACM (JACM)* **49**(5), 616–639.
- Ko, J. P. & Betke, M. (2001), 'Chest CT: Automated nodule detection and assessment of change over time-preliminary experience 1', *Radiology* **218**(1), 267–273.
- Kolmogorov, V. (2006), 'Convergent tree-reweighted message passing for energy minimization', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(10), 1568–1583.
- Kolmogorov, V. & Zabini, R. (2004), 'What energy functions can be minimized via graph cuts?', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(2), 147–159.
- Komodakis, N., Tziritas, G. & Paragios, N. (2007), Fast, approximately optimal solutions for single and dynamic MRFs, in 'IEEE Conference on Computer Vision and Pattern Recognition, CVPR '07.', pp. 1–8.
- Kschischang, F. R., Frey, B. J. & Loeliger, H.-A. (2001), 'Factor graphs and the sum-product algorithm', *IEEE Transactions on Information Theory* **47**(2), 498–519.
- Kwon, D., Lee, K. J., Yun, I. D. & Lee, S. U. (2011), Nonrigid image registration using higher-order MRF model with dense local descriptor, in 'Computer Vision and Pattern Recognition Workshop'.
- Kybic, J. & Unser, M. (2003), 'Fast parametric elastic image registration', *IEEE Transactions on Image Processing* **12**(11), 1427–1442.
- Lara-Alvarez, C., Romero, L., Flores, J. & Gomez, C. (2009), A simple sample consensus algorithm to find multiple models, in E. Bayro-Corrochano & J.-O. Eklundh, eds, 'Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications', Vol. 5856 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 918–925.
- Le Moigne, J., Netanyahu, N. S. & Eastman, R. D. (2011), *Image registration for remote sensing*, Cambridge University Press.
- Lee, K. J., Kwon, D., Yun, I. D. & Lee, S. U. (2008), Deformable 3D volume registration using efficient MRFs model with decomposed nodes, in 'BMVC', pp. 1–10.

- Lester, H. & Arridge, S. R. (1999), 'A survey of hierarchical non-linear medical image registration', *Pattern Recognition* **32**(1), 129–149.
- Liang, Y., Ryan, N. S., Schott, J. M. & Fox, N. C. (2013), 'Imaging the onset and progression of Alzheimer's disease: Implications for prevention trials', *Journal of Alzheimer's Disease* **33**(0), S305–S312.
- Likar, B. & Pernu, F. (2001), 'A hierarchical approach to elastic registration based on mutual information', *Image and Vision Computing* **19**(1), 33–44.
- Lindeberg, T. (1998), 'A scale selection principle for estimating image deformations', *Image and Vision Computing* **16**(14), 961–977.
- Linté, C. A., White, J., Eagleson, R., Guiraudon, G. M. & Peters, T. M. (2010), 'Virtual and augmented medical imaging environments: Enabling technology for minimally invasive cardiac interventional guidance', *IEEE Reviews in Biomedical Engineering* **3**, 25–47.
- Liu, H., Latecki, L. J. & Yan, S. (2010), Robust clustering as ensembles of affinity relations, in 'Advances in Neural Information Processing Systems', pp. 1414–1422.
- Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D. & Suetens, P. (2010), 'Nonrigid image registration using conditional mutual information', *IEEE Transactions on Medical Imaging* **29**(1), 19–29.
- Lowe, D. G. (1999), Object recognition from local scale-invariant features, in 'The Proceedings of the Seventh IEEE International Conference on Computer vision, 1999.', Vol. 2, IEEE, pp. 1150–1157.
- Lucas, B. D. & Kanade, T. (1981), 'An iterative image registration technique with an application to stereo vision', *Proceedings of Imaging Understanding Workshop* pp. 121–130.
- Lucena, M., Fuertes, J. M., de la Blanca, N. P. & Marín-Jiménez, M. J. (2010), 'Tracking people in video sequences using multiple models', *Multimedia Tools and Applications* **49**(2), 371–403.
- Maintz, J. B. A. & Viergever, M. A. (1998), 'A survey of medical image registration', *Medical Image Analysis* **2**(1), 1–36.
- Maurer Jr., C. R. & Fitzpatrick, J. M. (1993), 'A review of medical image registration', *Interactive image-guided neurosurgery* **17**.
- Meer, P. (2004), 'Robust techniques for computer vision', *Emerging Topics in Computer Vision* pp. 107–190.
- Meinguet, J. (1979), 'Multivariate interpolation at arbitrary points made simple', *Zeitschrift für angewandte Mathematik und Physik ZAMP* **30**(2), 292–304.
- Mirota, D. J., Hanzi, W., Taylor, R. H., Ishii, M., Gallia, G. L. & Hager, G. D. (2012), 'A system for video-based navigation for endoscopic endonasal skull base surgery', *IEEE Transactions on Medical Imaging* **31**(4), 963–976.

- Modersitzki, J. (2004), *Numerical Methods for Image Registration (Numerical Mathematics and Scientific Computation)*, Oxford University Press USA.
- Modersitzki, J. (2008), 'FLIRT with rigidity-image registration with a local non-rigidity penalty', *International Journal of Computer Vision* **76**(2), 153–163.
- Modersitzki, J. (2009), *FAIR: flexible algorithms for image registration*, Vol. 6, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.
- Mostayed, A., Garlapati, R. R., Joldes, G. R., Wittek, A., Roy, A., Kikinis, R., Warfield, S. K. & Miller, K. (2013), 'Biomechanical model as a registration tool for image-guided neurosurgery: Evaluation against bspline registration', *Annals of Biomedical Engineering* **41**(11), 2409–2425.
- Muller, K., Schwemmer, C., Hornegger, J., Zheng, Y., Wang, Y., Lauritsch, G., Rohkohl, C., Maier, A. K., Schultz, C. & Fahrig, R. (2013), 'Evaluation of interpolation methods for surface-based motion compensated tomographic reconstruction for cardiac angiographic C-arm data', *Medical Physics* **40**(3), 031107.
- Musse, O., Heitz, F. & Armspach, J.-P. (2001), 'Topology preserving deformable image matching using constrained hierarchical parametric models', *IEEE Transactions on Image Processing* **10**(7), 1081–1093.
- Myronenko, A. & Xubo, S. (2010), 'Intensity-based image registration by minimizing residual complexity', *IEEE Transactions on Medical Imaging* **29**(11), 1882–1891.
- Noblet, V., Heinrich, C., Heitz, F. & Armspach, J.-P. (2005), '3-D deformable image registration: a topology preservation scheme based on hierarchical deformation models and interval analysis optimization', *IEEE Transactions on Image Processing* **14**(5), 553–566.
- Oliveira, F. P. & Tavares, J. M. (2014), 'Medical image registration: a review', *Comput Methods Biomech Biomed Engin* **17**(2), 73–93.
- Otte, M. & Nagel, H.-H. (1994), Optical flow estimation: Advances and comparisons, in J.-O. Eklundh, ed., 'Computer Vision - ECCV '94', Vol. 800 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 49–60.
- Papenberg, N., Bruhn Andrès, B., Thomas Didas, S. & Weickert, J. (2006), 'Highly accurate optic flow computation with theoretically justified warping', *International Journal of Computer Vision* **67**(2), 141–158.
- Pearl, J. (1988), *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- Peng, C. & Chia-Hsiang, M. (2013), 'Real-time continuous image registration enabling ultraprecise 2-D motion tracking', *IEEE Transactions on Image Processing* **22**(5), 2081–2090.
- Pennec, X., Stefanescu, R., Arsigny, V., Fillard, P. & Ayache, N. (2005), Riemannian elasticity: A statistical regularization framework for non-linear registration, in J. Duncan & G. Gerig, eds, 'Medical Image Computing and Computer-Assisted Intervention MICCAI 2005', Vol. 3750 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 943–950.

- Pham, T. T., Chin, T.-J., Yu, J. & Suter, D. (2012), The random cluster model for robust geometric fitting, in '2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)', IEEE, pp. 710–717.
- Pitiot, A., Malandain, G., Bardinet, E. & Thompson, P. (2003), Piecewise affine registration of biological images, in J. Gee, J. Maintz & M. Vannier, eds, 'Biomedical Image Registration', Vol. 2717 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 91–101.
- Pluim, J. P. W., Maintz, J. B. A. & Viergever, M. A. (2000), Image registration by maximization of combined mutual information and gradient information, in S. Delp, A. DiGoia & B. Jaramaz, eds, 'Medical Image Computing and Computer-Assisted Intervention MICCAI 2000', Vol. 1935 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 452–461.
- Pluim, J. P. W., Maintz, J. B. A. & Viergever, M. A. (2003), 'Mutual-information-based registration of medical images: a survey', *IEEE Transactions on Medical Imaging* **22**(8), 986–1004.
- Poling, B. & Lerman, G. (2014), 'A new approach to two-view motion segmentation using global dimension minimization', *International Journal of Computer Vision* **108**(3), 165–185.
- Qi, L. & Sun, D. (1999), A survey of some nonsmooth equations and smoothing newton methods, in 'Progress in Optimization, volume 30 of Applied Optimization', Kluwer Academic Publishers, pp. 121–146.
- Rabbitt, R. D., Weiss, J. A., Christensen, G. E. & Miller, M. I. (1995), Mapping of hyperelastic deformable templates using the finite element method, in 'SPIE's 1995 International Symposium on Optical Science, Engineering, and Instrumentation', International Society for Optics and Photonics, pp. 252–265.
- Richards, J. A. (2012), *Remote Sensing Digital Image Analysis: An Introduction*, Springer Berlin Heidelberg.
- Rivaz, H. & Collins, D. (2012), Self-similarity weighted mutual information: A new nonrigid image registration metric, in N. Ayache, H. Delingette, P. Golland & K. Mori, eds, 'Medical Image Computing and Computer-Assisted Intervention - MICCAI 2012', Vol. 7512 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 91–98.
- Rivaz, H., Karimaghloo, Z. & Collins, D. L. (2014), 'Self-similarity weighted mutual information: A new nonrigid image registration metric', *Medical Image Analysis* **18**(2), 343 – 358.
- Robbins, H. & Monro, S. (1951), 'A stochastic approximation method', *The Annals of Mathematical Statistics* **22**(3), 400–407.
- Robertson, S., Weiss, E. & Hugo, G. D. (2013), 'Deformable mesh registration for the validation of automatic target localization algorithms', *Medical Physics* **40**(7), 1721.

- Rohde, G. K., Aldroubi, A. & Dawant, B. M. (2003), 'The adaptive bases algorithm for intensity-based nonrigid image registration', *IEEE Transactions on Medical Imaging* **22**(11), 1470–1479.
- Rohlfing, T. & Maurer Jr, C. R. (2003), 'Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees', *Information Technology in Biomedicine, IEEE Transactions on* **7**(1), 16–25.
- Rousseeuw, P. J. & Leroy, A. M. (2005), *Robust regression and outlier detection*, Vol. 589, John Wiley & Sons.
- Rueckert, D. & Aljabar, P. (2010), 'Nonrigid registration of medical images: Theory, methods, and applications [applications corner]', *IEEE Signal Processing Magazine* **27**(4), 113–119.
- Rueckert, D., Clarkson, M., Hill, D. & Hawkes, D. J. (2000), Non-rigid registration using higher-order mutual information, in 'Medical Imaging 2000', International Society for Optics and Photonics, pp. 438–447.
- Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O. & Hawkes, D. J. (1999), 'Nonrigid registration using free-form deformations: application to breast MR images', *IEEE Transactions on Medical Imaging* **18**(8), 712–721.
- Ruhaak, J., Konig, L., Hallmann, M., Papenberg, N., Heldmann, S., Schumacher, H. & Fischer, B. (2013), A fully parallel algorithm for multimodal image registration using normalized gradient fields, in '2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI)', pp. 572–575.
- Rui, L., Li, Z., Ying, S., Shun, M. & Chafd'hotel, C. (2013), 'A review of recent advances in registration techniques applied to minimally invasive therapy', *IEEE Transactions on Multimedia* **15**(5), 983–1000.
- Rynne, B. P. & Youngson, M. A. (2008), *Linear functional analysis*, Vol. 283, Springer-Verlag London.
- Sadeghi Naini, A., Patel, R. V. & Samani, A. (2010), 'CT-enhanced ultrasound image of a totally deflated lung for image-guided minimally invasive tumor ablative procedures', *IEEE Transactions on Biomedical Engineering* **57**(10), 2627–2630.
- Schindler, K. & Suter, D. (2006), 'Two-view multibody structure-and-motion with outliers through model selection', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(6), 983–995.
- Shekhovtsov, A., Kovtun, I. & Hlavá č, V. (2008), 'Efficient MRF deformation model for non-rigid image matching', *Computer Vision and Image Understanding* **112**(1), 91–99.
- Shen, C., Kim, J. & Wang, H. (2010), 'Generalized kernel-based visual tracking', *IEEE Transactions on Circuits and Systems for Video Technology* **20**(1), 119–130.
- Shiraishi, J., Abe, H., Li, F., Engelmann, R., MacMahon, H. & Doi, K. (2006), 'Computer-aided diagnosis for the detection and classification of lung cancers on chest radiographs. ROC analysis of radiologists' performance', *Academic Radiology* **13**, 995–1003.

- Simoncelli, E. P. (1994), Design of multi-dimensional derivative filters, in 'IEEE International Conference on Image Processing, 1994. Proceedings. ICIP-94.', Vol. 1, pp. 790–794 vol.1.
- Sluimer, I., Schilham, A., Prokop, M. & Van Ginneken, B. (2006), 'Computer analysis of computed tomography scans of the lung: A survey', *IEEE Transactions on Medical Imaging* **25**(4), 385–405.
- So, R. & Chung, A. (2009), Multi-level non-rigid image registration using graph-cuts, in 'IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009.', pp. 397–400.
- So, R. W. K., Tang, T. W. H. & Chung, A. C. S. (2011), 'Non-rigid image registration of brain magnetic resonance images using graph-cuts', *Pattern Recognition* **44**(1011), 2450–2467.
- Song, S. M. & Leahy, R. M. (1991), 'Computation of 3-D velocity fields from 3-D cine CT images of a human heart', *IEEE Transactions on Medical Imaging* **10**(3), 295–306.
- Sotiras, A., Davatzikos, C. & Paragios, N. (2013), 'Deformable medical image registration: A survey', *IEEE Transactions on Medical Imaging* **32**(7), 1153–1190.
- Spall, J. C. (1998), 'Implementation of the simultaneous perturbation algorithm for stochastic optimization', *IEEE Transactions on Aerospace and Electronic Systems* **34**(3), 817–823.
- Stals, L. & Roberts, S. (2006), 'Smoothing large data sets using discrete thin plate splines', *Computing and Visualization in Science* **9**(3), 185–195.
- Staring, M., van der Heide, U. A., Klein, S., Viergever, M. A. & Pluim, J. P. (2009), 'Registration of cervical MRI using multifeature mutual information', *IEEE Transactions on Medical Imaging* **28**(9), 1412–21.
- Stewart, C. V. (1997), 'Bias in robust estimation caused by discontinuities and multiple structures', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(8), 818–833.
- Studholme, C., Drapaca, C., Iordanova, B. & Cardenas, V. (2006), 'Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change', *IEEE Transactions on Medical Imaging* **25**(5), 626–639.
- Studholme, C., Hill, D. L. & Hawkes, D. J. (1999), 'An overlap invariant entropy measure of 3D medical image alignment', *Pattern Recognition* **32**(1), 71–86.
- Sugaya, Y. & Kanatani, K. (2004), Geometric structure of degeneracy for multi-body motion segmentation, in D. Comaniciu, R. Mester, K. Kanatani & D. Suter, eds, 'Statistical Methods in Video Processing', Vol. 3247 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 13–25.
- Sun, D., Roth, S. & Black, M. (2010), Secrets of optical flow estimation and their principles, in '2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)', IEEE, pp. 2432–2439.

- Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M. & Rother, C. (2008), 'A comparative study of energy minimization methods for markov random fields with smoothness-based priors', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(6), 1068–1080.
- Tennakoon, R. B., Bab-Hadiashar, A., Cao, Z. & de Bruijne, M. (2014), 'Nonrigid registration of volumetric images using ranked order statistics', *IEEE Transactions on Medical Imaging* **33**(2), 422–432.
- Thirion, J. P. (1998), 'Image matching as a diffusion process: an analogy with Maxwell's demons', *Medical Image Analysis* **2**(3), 243–260.
- Timp, S., Varela, C. & Karssemeijer, N. (2007), 'Temporal change analysis for characterization of mass lesions in mammography', *IEEE Transactions on Medical Imaging* **26**(7), 945–953.
- Torr, P. H. & Murray, D. W. (1997), 'The development and comparison of robust methods for estimating the fundamental matrix', *International Journal of Computer Vision* **24**(3), 271–300.
- Torr, P. H. S. (1995), Motion segmentation and outlier detection, PhD thesis, University of Oxford England.
- Tran, Q.-H., Chin, T.-J., Carneiro, G., Brown, M. & Suter, D. (2012), In defence of RANSAC for outlier rejection in deformable registration, in A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato & C. Schmid, eds, 'Computer Vision - ECCV 2012', Vol. 7575 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 274–287.
- Tron, R. & Vidal, R. (2007), A benchmark for the comparison of 3-D motion segmentation algorithms, in 'IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07.', IEEE, pp. 1–8.
- Tustison, N. J., Cook, T. S., Song, G. & Gee, J. C. (2011), 'Pulmonary kinematics from image data: a review', *Academic Radiology* **18**(4), 402–417.
- Uneri, A., Nithiananthan, S., Schafer, S., Otake, Y., Stayman, J. W., Kleinszig, G., Sussman, M. S., Prince, J. L. & Siewerdsen, J. H. (2013), 'Deformable registration of the inflated and deflated lung in cone-beam CT-guided thoracic surgery: Initial investigation of a combined model- and image-driven approach', *Medical Physics* **40**(1).
- Unser, M. (1999), 'Splines: a perfect fit for signal and image processing', *IEEE Signal Processing Magazine* **16**(6), 22–38.
- Van Ginneken, B., Ter Haar Romeny, B. M. & Viergever, M. A. (2001), 'Computer-aided diagnosis in chest radiography: A survey', *IEEE Transactions on Medical Imaging* **20**(12), 1228–1241.
- van Herk, M. (2007), 'Different styles of image-guided radiotherapy', *Semin. Radiat. Oncol.* **17**(4), 258–67.
- Vandemeulebroucke, J., Rit, S., Kybic, J., Clarysse, P. & Sarrut, D. (2011), 'Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs', *Medical Physics* **38**(1), 166–178.

- Vercauteren, T., Pennec, X., Perchant, A. & Ayache, N. (2007), Non-parametric diffeomorphic image registration with the demons algorithm, *in* N. Ayache, S. Ourselin & A. Maeder, eds, 'Medical Image Computing and Computer-Assisted Intervention - MICCAI 2007', Vol. 4792 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 319–326.
- Viola, P. & Wells III, W. M. (1997), 'Alignment by maximization of mutual information', *International Journal of Computer Vision* **24**(2), 137–154.
- Wahba, G. (1979), *How to smooth curves and surfaces with splines and cross-validation*, Defense Technical Information Center.
- Wahba, G. (1990), *Spline models for observational data*, Vol. 59, Society for Industrial and Applied Mathematics.
- Wainwright, M. J., Jaakkola, T. S. & Willsky, A. S. (2005), 'MAP estimation via agreement on trees: message-passing and linear programming', *IEEE Transactions on Information Theory* **51**(11), 3697–3717.
- Wang, H. & Suter, D. (2004), 'MDPE: A very robust estimator for model fitting and range image segmentation', *International Journal of Computer Vision* **59**(2), 139–166.
- Wang, J. & Jiang, T. (2007), 'Nonrigid registration of brain MRI using NURBS', *Pattern Recognition Letters* **28**(2), 214–223.
- Wang, Y. (2011), *Smoothing splines: Methods and applications*, Taylor & Francis US.
- Wei, J., Chan, H. P., Sahiner, B., Zhou, C., Hadjiiski, L. M., Roubidoux, M. A. & Helvie, M. A. (2009), 'Computer-aided detection of breast masses on mammograms: Dual system approach with two-view analysis', *Medical Physics* **36**(10), 4451–4460.
- Weickert, J. & Schnörr, C. (2001), 'Variational optic flow computation with a spatio-temporal smoothness constraint', *Journal of Mathematical Imaging and Vision* **14**(3), 245–255.
- Wilcoxon, F. (1945), 'Individual comparisons by ranking methods', *Biometrics Bulletin* **1**(6), 80–83.
- Wong, H. S., Chin, T.-J., Yu, J. & Suter, D. (2011), Dynamic and hierarchical multi-structure geometric model fitting, *in* '2011 IEEE International Conference on Computer Vision (ICCV)', pp. 1044–1051.
- Worm, E. S., Hansen, A. T., Petersen, J. B., Muren, L. P., Prestegard, L. H. & Hoyer, M. (2010), 'Inter and intrafractional localisation errors in cone-beam CT guided stereotactic radiation therapy of tumours in the liver and lung', *Acta Oncologica* **49**(7), 1177–1183.
- Wu, Y.-T., Kanade, T., Li, C.-C. & Cohn, J. (2000), 'Image registration using wavelet-based motion model', *International Journal of Computer Vision* **38**(2), 129–152.
- Wu, Z. (1995), 'Compactly supported positive definite radial functions', *Advances in Computational Mathematics* **4**(1), 283–292.

- Xishi, H., Jing, R., Guiraudon, G., Boughner, D. & Peters, T. M. (2009), 'Rapid dynamic image registration of the beating heart for diagnosis and surgical navigation', *IEEE Transactions on Medical Imaging* **28**(11), 1802–1814.
- Xu, L., Jia, J. & Matsushita, Y. (2012), 'Motion detail preserving optical flow estimation', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(9), 1744–1757.
- Yang, D., Lu, W., Low, D. A., Deasy, J. O., Hope, A. J. & Naqa, I. E. (2008), '4D-CT motion estimation using deformable image registration and 5D respiratory motion modeling', *Medical Physics* **35**(10), 4577–4590.
- Yedidia, J. S., Freeman, W. T. & Weiss, Y. (2000), Generalized belief propagation, in 'NIPS', Vol. 13, MIT Press, pp. 689–695.
- Yin, Y., Hoffman, E. A. & Lin, C.-L. (2009), 'Mass preserving nonrigid registration of CT lung images using cubic B-spline', *Medical Physics* **36**, 4213.
- Yu, J., Chin, T.-J. & Suter, D. (2011), A global optimization approach to robust multi-model fitting, in '2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)', IEEE, pp. 2041–2048.
- Yuanjie, Z., Kambhamettu, C., Bauer, T. & Steiner, K. (2009), Accurate estimation of pulmonary nodule's growth rate in CT images with nonrigid registration and precise nodule detection and segmentation, in 'IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009.', pp. 101–108.
- Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C. & Gerig, G. (2006), 'User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability', *NeuroImage* **31**(3), 1116–1128.
- Zhang, G. G., Huang, T. C., Guerrero, T., Lin, K. P., Stevens, C., Starkschall, G. & Forster, K. (2008), 'Use of three-dimensional (3D) optical flow method in mapping 3D anatomic structure and tumor contours across four-dimensional computed tomography data', *Applied Clinical Medical Physics / American College of Medical Physics* **9**, 2738.
- Zhang, M., Singh, N. & Fletcher, P. (2013), Bayesian estimation of regularization and atlas building in diffeomorphic image registration, in J. Gee, S. Joshi, K. Pohl, W. Wells & L. Zöllei, eds, 'Information Processing in Medical Imaging', Vol. 7917 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 37–48.
- Zhang, Y., Zhang, X., Zhang, Z., Wu, D. & Zhou, B. (2014), 'Registration of hepatic CT images based on thin plate spline (TPS) deformation using automatically searching landmarks on binary maps', *Human Health and Medical Engineering* **18**, 471.
- Zitova, B. & Flusser, J. (2003), 'Image registration methods: a survey', *Image and Vision Computing* **21**(11), 977–1000.