THE UNIVERSITY
of ADELAIDE

Robust Parameter Estimation in Computer Vision: Geometric Fitting and Deformable Registration

by

Quoc Huy Tran

A thesis submitted in fulfillment for the
degree of Doctor of Philosophy

in the

Faculty of Engineering, Computer and Mathematical Sciences
School of Computer Science

August 2014

Contents

List of Figures

List of Tables

THE UNIVERSITY OF ADELAIDE

Abstract

Faculty of Engineering, Computer and Mathematical Sciences
School of Computer Science

Doctor of Philosophy

by Quoc Huy Tran

Parameter estimation plays an important role in computer vision. Many computer vision problems can be reduced to estimating the parameters of a mathematical model of interest from the observed data. Parameter estimation in computer vision is challenging, since vision data unavoidably have small-scale measurement noise and large-scale measurement errors (outliers) due to imperfect data acquisition and preprocessing. Traditional parameter estimation methods developed in the statistics literature mainly deal with noise and are very sensitive to outliers. Robust parameter estimation techniques are thus crucial for effectively removing outliers and accurately estimating the model parameters with vision data. The research conducted in this thesis focuses on single structure parameter estimation and makes a direct contribution to two specific branches under that topic: geometric fitting and deformable registration.

In geometric fitting problems, a geometric model is used to represent the information of interest, such as a homography matrix in image stitching, or a fundamental matrix in three-dimensional reconstruction. Many robust techniques for geometric fitting involve sampling and testing a number of model hypotheses, where each hypothesis consists of a minimal subset of data for yielding a model estimate. It is commonly known that, due to the noise added to the true data (inliers), drawing a single all-inlier minimal subset is not sufficient to guarantee a good model estimate that fits the data well; the inliers therein should also have a large spatial extent. This thesis investigates a theoretical reasoning behind this long-standing principle, and shows a clear correlation between the span of data points used for estimation and the quality of model estimate. Based on this finding, the thesis explains why naive distance-based sampling fails as a strategy to maximise the span of all-inlier minimal subsets produced, and develops a novel sampling algorithm which, unlike previous approaches, consciously targets all-inlier minimal subsets with large span for robust geometric fitting.

The second major contribution of this thesis relates to another computer vision problem which also requires the knowledge of robust parameter estimation: deformable registration. The goal of deformable registration is to align regions in two or more images corresponding to a common object that can deform nonrigidly such as a bending piece of paper or a waving flag. The

information of interest is the nonlinear transformation that maps points from one image to another, and is represented by a deformable model, for example, a thin plate spline warp. Most of the previous approaches to outlier rejection in deformable registration rely on optimising fully deformable models in the presence of outliers due to the assumption of the highly nonlinear correspondence manifold which contains the inliers. This thesis makes an interesting observation that, for many realistic physical deformations, the scale of errors of the outliers usually dwarfs the nonlinear effects of the correspondence manifold on which the inliers lie. The finding suggests that standard robust techniques for geometric fitting are applicable to model the approximately linear correspondence manifold for outlier rejection. Moreover, the thesis develops two novel outlier rejection methods for deformable registration, which are based entirely on fitting simple linear models and shown to be considerably faster but at least as accurate as previous approaches.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the Universitys digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed: _____

Date: _____

Publications

This thesis is based on the content of the following conference and journal papers:

- Quoc-Huy Tran, Tat-Jun Chin, Gustavo Carneiro, Michael S. Brown, and David Suter. In Defence of RANSAC for Outlier Rejection in Deformable Registration. In *European Conference on Computer Vision (ECCV)*, Florence, Italy, October 2012.
  (DOI: http://dx.doi.org/10.1007/978-3-642-33765-9_20)

- Quoc-Huy Tran, Tat-Jun Chin, Wojciech Chojnacki, and David Suter. Sampling Minimal Subsets with Large Spans for Robust Estimation. *International Journal of Computer Vision (IJCV)*, July 2013.
  (DOI: http://dx.doi.org/10.1007/s11263-013-0643-y)

- Quoc-Huy Tran, Tat-Jun Chin, Julio Zaragoza, Gustavo Carneiro, and David Suter. Outlier Rejection in Deformable Registration with Moving Least Squares. *Submitted to Journal of Computer Vision and Image Understanding (CVIU)*.
  (under review)

In addition, I have co-authored the below paper:

- Julio Zaragoza, Tat-Jun Chin, Quoc-Huy Tran, Michael S. Brown, and David Suter. As-Projective-As-Possible Image Stitching with Moving DLT. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, November 2013.
  (DOI: http://dx.doi.org/10.1109/TPAMI.2013.247)

Acknowledgements

First of all, I would like to express my deepest gratitude to my principal supervisor Doctor Tat-Jun Chin for his continuous guidance and strong support throughout my time as his student. He has always made himself available to me for discussing any research problems and potential directions that I brought. His intelligence and creativity were always helpful and so inspiring to me. I am sure that this thesis would not have been possible without his guidance and enthusiasm.

I owe my sincere thankfulness to my co-supervisor Professor David Suter for his valuable support and insightful advice over the past three years. I am especially grateful to his countless hours of efforts on reviewing and improving my research papers and thesis. His scientific expertise and advice helped me overcome difficulties during my candidature. I could not have finished this thesis without his support and encouragement.

It is my great pleasure to have had the opportunity to work with Doctor Gustavo Carneiro and Professor Wojciech Chojnacki at the Australian Centre for Visual Technologies, and Associate Professor Michael S. Brown at the National University of Singapore, who were kind and so supportive. I have learned much from valuable discussions during the collaboration with them. Many thanks also go to Professor Ba-Ngu Vo for sharing his priceless research experiences during my visit at the University of Western Australia.

My gratitude is extended to the University of Adelaide for the financial support through the International Postgraduate Research Scholarship and Australian Postgraduate Award over the past few years. I am especially grateful to the School of Computer Science for providing me with excellent facilities during my candidature. Moreover, I am indebted to my colleagues at the Australian Centre for Visual Technologies for their various forms of support during my study in Adelaide. In particular, I thank my friends Julio Zaragoza, Gousheng Lin, Trung Pham, Sim Wong, Alexander Cichowski and Alvaro Parra Bustos for interesting discussions and memories that we had. They certainly made my life in the graduate school much more fun.

Finally, I want to express my great appreciation to my family for their unconditional support and encouragement throughout the past three years. I would especially like to thank my wife Truc Ngo for her constant love, patience and support during my study in Adelaide.

Quoc Huy Tran
August 2014

Dedicated to my family,
for their unconditional love and endless support.

Chapter 1

Introduction

1.1 Introduction

In contrast to humans with the natural ability to observe and understand the images, computers can only see an image as a collection of numbers, which merely represent the color information. The ultimate purpose of *Computer Vision* is thus to mimic human perception, and provide machines with the ability to analyse, extract and understand the semantic information in the images. Research in computer vision also has an important contribution towards the future of intelligent robots which can behave like humans.

Over the last decade, there have been significant advances in the field of computer vision. As the result of the intense efforts of computer vision researchers, numerous computer vision applications have emerged, and made a positive impact on our everyday lives. One representative example of such applications is Microsoft's *Photosynth*. Photosynth allows users to align multiple images of a *planar* scene and stitch them into a single seamless panoramic photo [1]. The tool is very useful in helping users organise and appreciate their photo collections. Another interesting computer vision application is *Videotrace*, which speeds up the generation of three-dimensional structure of a *static* scene from a sequence of images [2]. The generated models offer users a better understanding of the objects, e.g., archaeological remains or architectural structures, compared to just two-dimensional images, and can be smoothly inserted into video games or simulated environments for virtual reality.

The above novel applications depend on methods that belong to the broad topic of *Parameter Estimation*. In parameter estimation problems, the information of interest is represented by a *parameterised model* and the goal is to estimate the model parameters from the observed data. For example, in Photosynth, the transformation between two images is modelled by a homography matrix [3], which is employed to stitch the images together. In Videotrace, a fundamental

matrix [3] is used to establish the relationship between two views, and employed for three-dimensional reconstruction. Since various parameterised models are used in computer vision, parameter estimation plays an integral role in a variety of computer vision applications, such as three-dimensional reconstruction [4, 5], image stitching [6, 7], image registration [8, 9], shape matching [10, 11], object recognition [12, 13] and surveillance [14, 15].

1.2 Background and motivation

Vision data (e.g., images or videos) usually have a huge variability due to several factors. For instance, the projective transformation in the image formation process can severely distort the scene. The lighting conditions can cause extreme brightness variations such as reflections or shadows. In addition, being very high-dimensional, vision data typically contain redundant or spurious information that interferes with the task of parameter estimation. In practice, *feature-based representations* of the images are often used to capture the key elements (features) that are relevant to a particular task and reduce the variability of vision data. For example, scale-invariant feature transform (SIFT) matches [16] are frequently employed as the input data for the estimation of two-view geometry (i.e., fundamental matrix estimation or homography matrix estimation), while shape matching is often performed on shape context (SC) correspondences [10] instead of raw images.

Due to imperfect feature extraction and matching, visual data points invariably include small-scale measurement noise, and are unavoidably contaminated by large-scale measurement errors (*outliers*) that are not explainable by the model (e.g., incorrect SIFT correspondences or false SC matches). Traditional parameter estimation methods proposed in the field of statistics, e.g., least squares regression, mostly deal with noise and are vulnerable to outliers [17]. *Robust Parameter Estimation* hence aims to effectively remove outliers and accurately estimate the model parameters in the presence of outliers.

Despite the immense body of work in both the statistics and the computer vision literatures, robust parameter estimation with vision data remains challenging and there are still many unanswered questions. Progress in this topic is difficult since it requires a thorough understanding of various disciplines, including mathematical analysis, statistical modelling and optimisation. For a specific task in computer vision, a mathematical model is first required for correctly representing the information of interest. The statistical modelling is also employed for properly handling noise in variables. Furthermore, the optimisation needs careful design to effectively tackle outliers and avoid getting stuck at locally optimal solutions.

Given that robust parameter estimation in computer vision is complex and challenging, we concentrate on single structure parameter estimation and set out to make a contribution to two

specific branches under that topic: geometric fitting (Section 1.2.1) and deformable registration (Section 1.2.2).

1.2.1 Geometric fitting

In *Geometric Fitting* problems, the information of interest is represented by a *geometric model*, which has a *fixed* and *known* number of degrees of freedom. For instance, many tasks in multiple-view geometry are geometric fitting problems, such as fundamental matrix estimation with seven degrees of freedom or homography matrix estimation with eight degrees of freedom [3]. Due to the known number of degrees of freedom, many robust techniques for geometric fitting involve generating and analysing a number of model hypotheses, where each hypothesis is fitted on a minimal subset of data, e.g., a fundamental matrix estimate can be obtained by using seven SIFT matches (each match offers one constraint), while four SIFT correspondences are used for yielding a homography matrix estimate (every correspondence provides two constraints) [3].

Intuitively, drawing an all-inlier minimal subset is not sufficient to guarantee a satisfactory model estimate that fits the data well; the inliers therein should also have a large span. To illustrate this notion, we consider the problem of 2D line fitting in Figure 1.1, where the data have been generated without outliers for simplicity. Two particular choices of all-inlier minimal subsets are highlighted; clearly Set A yields a better estimate than Set B, as can be verified by a suitable goodness-of-fit criterion, e.g., [18, 19]. It is also apparent that the points in Set A are separated by a larger distance than those in Set B.

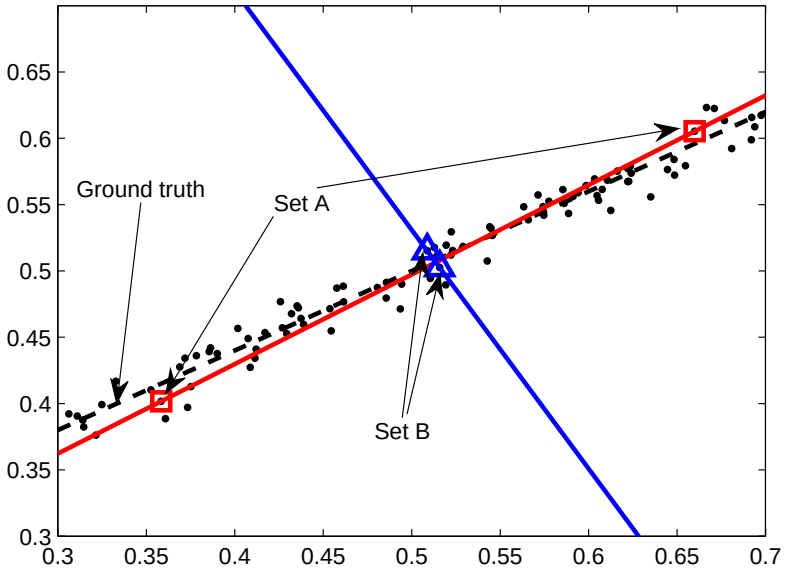


Figure 1.1: Two particular all-inlier minimal subsets with different fitting qualities on 2D line fitting.

The observation above explains, to a large degree, previous findings [17, 20–23] that the number of randomly drawn minimal subsets required before finding a satisfactory model estimate is

often far higher than predicted. The truth is that such predictions assume that retrieving a single all-inlier sample is sufficient, which ignores the fact that all-inlier samples differ in terms of their spatial coverage, and hence intrinsic quality. It has also been remarked that taking minimal subsets amplifies the noise magnitude [17] . However, little study was devoted to this phenomenon and, more importantly, the manner in which the data span affects different minimal subsets. Theoretical and algorithmic studies of this long-standing issue are generally lacking.

1.2.2 Deformable registration

When the scene can deform nonrigidly between the images, another computer vision problem, which also requires the knowledge of robust parameter estimation, frequently arises: *Deformable Registration*. The goal of deformable registration is to align regions in two or more images containing a common object that can deform smoothly, such as a waving flag or a bending piece of paper. The information of interest is the nonlinear transformation that maps points from one image to another, and is represented by a *deformable model*, e.g., a thin plate spline (TPS) [24] warp. Given that visual data points are noisy and invariably contain outliers, robust techniques are thus necessary for effectively rejecting outliers and accurately estimating the model parameters. If no mismatches exist, estimating the model parameters is trivial, e.g., by solving a linear system for a TPS warp [24].

At the first glance it would appear that standard robust techniques for geometric fitting are not suitable for outlier rejection in deformable registration, the fundamental obstacle being that the underlying transformation is of *unknown* and *varying* complexity [25, 26]. Also, it is widely assumed that many realistic deformations (e.g., a bending sheet of paper or a rippling cloth) are too nonlinear to be amenable to simple geometric modelling. Figure 1.2 (taken from [27]) depicts such impressions of the data.

Due to the above assumption of the highly nonlinear correspondence manifold on which the inliers lie, robust loss functions [28] are usually applied to alleviate the effects of outliers, e.g., see [29, Equation (5)] or [30, Equation (6)]. However, such robust loss functions (known as hard redescenders [28]) are often nonconvex and they complicate the warp optimisation. Most of the previous approaches resort to iterative algorithms, such as semi-implicit optimisation [29] or Gauss-Newton approximation [30], whose success depends critically on good parameter initialisations.

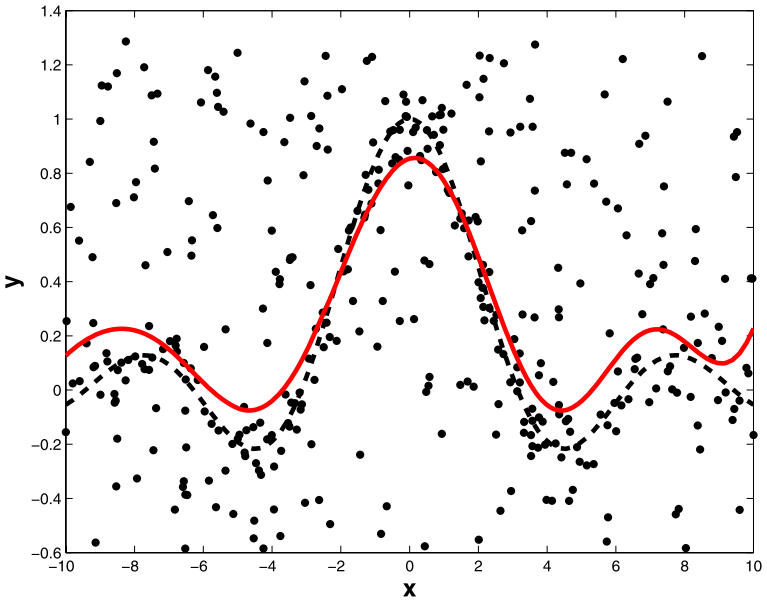


Figure 1.2: The characteristic and level of difficulty of the correspondence manifold targeted in the literature. This figure is taken from [27].

1.3 Contributions of this thesis

In this thesis we focus on parameter estimation of a single structure and make an original contribution to two problems that are typically encountered in the context of robust parameter estimation with vision data. The first problem relates to sampling minimal subsets with large span for robust geometric fitting (Section 1.3.1), while the second one relates to the approximately linear correspondence manifold for robust deformable registration (Section 1.3.2).

1.3.1 Geometric fitting with large span sampling

We provide an important study on the problem of large span sampling in geometric fitting. Specifically, we will show that a theoretical explanation for the intuition in Figure 1.1 lies in a little known result relating to minimal subset expansion for least squares regression. From this result, the quality of a minimal subset estimate is proportional to the span of the associated data points.

One of the main contributions of this thesis is to develop a similar minimal subset expansion for total least squares [31], which is directly related to geometric estimation techniques in computer vision, such as the direct linear transformation [32]. In addition, we will show a clear connection between the span of data points used for estimation and the quality of model estimate in the task of fundamental matrix estimation.

Based on this insight, we will explain why naive distance-based sampling fails as a strategy to maximise the span of all-inlier minimal subsets produced. Moreover, we propose a novel sampling algorithm which, unlike previous approaches [20–23, 33–37], consciously targets all-inlier minimal subsets with large span for robust geometric fitting.

1.3.2 Deformable registration with approximately linear correspondence manifold

In contrast to the common assumption of the highly nonlinear correspondence manifold illustrated in Figure 1.2, we make an interesting observation that, for many types of deformations examined in the literature, the scale of errors of the outliers usually dwarfs the effects of the curvature of the correspondence manifold containing the inliers. Therefore, relative to the outliers, the inliers lie very close to a linear hyperplane.

Based on this insight, standard robust techniques for geometric fitting can be used to model the approximately linear correspondence manifold for outlier rejection. Furthermore, we will present in this thesis two novel outlier rejection methods for deformable registration, which are based entirely on fitting simple linear models. Our techniques are shown to be significantly more efficient and at least as accurate as previous approaches [27, 29, 30, 38–42] that are often based on optimising fully deformable models in the presence of outliers.

1.4 Organisation of this thesis

Chapter 2

In the second chapter, we review previous approaches to robust geometric fitting and robust deformable registration in the current literature, which are relevant to this thesis. We further discuss the problem of large span sampling in robust geometric fitting (Section 1.2.1), and explain the assumption of the highly nonlinear correspondence manifold for outlier rejection in deformable registration (Section 1.2.2). This chapter also provides the reader a theoretical background for better understanding the proposed methods in the next chapters.

Chapter 3

The third chapter presents a theoretical basis for large span sampling in robust geometric fitting (Section 1.2.1). We first describe the minimal subset expansion for least squares regression (previously mentioned in Section 1.3.1), where the quality of a minimal subset estimate is a

function of the span of the associated data points. We then develop an analogous minimal subset expansion for total least squares, which is highly related to the direct linear transformation, a widely used geometric estimation technique in computer vision. Moreover, we analyse the influence of data span on the task of fundamental matrix estimation.

Chapter 4

The forth chapter focuses on an algorithmic solution for large span sampling in robust geometric fitting (Section 1.2.1). We study the disadvantage of proximity sampling which actually limits the span of all-inlier minimal subsets produced, and propose a novel sampling algorithm which actively searches for all-inlier minimal subsets with large span. Our method is benchmarked against previous approaches on synthetic and real datasets for robust fundamental matrix estimation.

Chapter 5

The fifth chapter applies standard robust geometric fitting techniques to outlier rejection in deformable registration (Section 1.2.2). We describe the interesting observation of the approximately linear correspondence manifold (previously mentioned in Section 1.3.2), and present a simple outlier rejection method which approximates the correspondence manifold with a linear hyperplane. We evaluate the performance of our method on synthetic and real datasets, and investigate its application in retexturing deformable surfaces.

Chapter 6

The sixth chapter pushes the linear approximation idea further, for outlier rejection in deformation registration (Section 1.2.2). We tweak the fitted linear hyperplane in the previous chapter to produce a nonlinear manifold that better respects the nonlinearity of the true manifold and improves the separation between inliers and outliers. We analyse the observation of the approximately linear correspondence manifold in the previous chapter, and benchmark the proposed method against previous approaches on different kinds of deformable objects.

Chapter 7

In the last chapter, we summarise the main contributions of this thesis and offer some open problems for future research.

Chapter 2

Robust Parameter Estimation in Computer Vision

2.1 Introduction

Traditional parameter estimation methods developed in the statistics community focus on handling small-scale noise in the data, e.g., the least squares estimator proposed by Gauss in 1809 computes the model parameters that minimise the sum of squared residuals (fitting errors) of all data points. It is well known that traditional methods are very sensitive to outliers [17], e.g., a single outlier in the data can move the least squares estimate arbitrarily far from the true solution. Statisticians have thus attempted to develop methods which are more robust to outliers. One representative effort is the least median squares estimator [43] which was introduced by Rousseeuw in 1984. The least median squares method searches for the model parameters that offer the minimum median of squared residuals of all data points, and hence it can tolerate at most 50% outliers in the data.

As discussed in Chapter 1, the data encountered in computer vision are often heavily contaminated by outliers, which can compose a large percentage (e.g., 90%) of the data. Traditional parameter estimation techniques in the statistics community are not applicable to vision data because of their sensitivity to high outlier rate. Computer vision researchers have thus made significant efforts to develop robust methods which can tackle highly contaminated vision data.

In this chapter, we will review state-of-the-art techniques for robust geometric fitting and robust deformable registration, which are two typical branches under the topic of robust parameter estimation with vision data. Since the focus of this thesis is on parameter estimation of a single structure, most of the methods discussed in this chapter (and more generally this thesis) are for

data containing one structure; although we will discuss the handling of multiple-structure data in Section 7.2.

The rest of this chapter is organised as follows: We first present in Section 2.2 a review of previous robust geometric fitting methods that often rely on sampling and testing minimal subsets of data. We then discuss the problem of sampling minimal subsets with large span (previously mentioned in Section 1.2.1) and current solutions to this issue. Section 2.3 describes the common assumption of the highly nonlinear correspondence manifold in deformable registration (previously mentioned in Section 1.2.2), which precludes the use of standard robust geometric fitting techniques for outlier rejection. Moreover, we survey previous outlier rejection methods for deformable registration that are based on this assumption. Finally, Section 2.4 summarises the issues to be addressed in the next chapters.

2.2 Geometric fitting

2.2.1 Minimal subset sampling for robust geometric fitting

One of the earliest recorded use of minimal subsets in parameter estimation occurred in 1755, when Boscovich attempted to determine the meridian arc near Rome (Italy) from five measurements [44]. He solved for the two unknowns of the arc by using all ten possible pairings (minimal subsets) of the data. Two of the pairs were ignored for yielding what Boscovich considered to be unusual outcomes, and the remaining estimates were simply averaged for his final result. Boscovich's work predated Gauss's paper on least squares which was published in 1809, but did not gain traction due to a lack of analytical basis.

Presently however, the use of minimal subsets has become an integral part of robust geometric fitting, especially in the estimation of multiple-view geometry from noisy image data [3]. This prevalent use of minimal subsets stems from the fact that geometric models have a known number of degrees of freedom, e.g., fundamental matrix, thus an estimate of the model parameters can be obtained from a minimal subset of data. Moreover, many robust criteria, such as least median squares [43], do not have closed-form solutions. Therefore, sampling and testing model hypotheses from minimal subsets is often the only way to obtain a good model estimate in a reasonable time.

The most famous and widely used robust geometric fitting method in the computer vision community is RANSAC [18], which relies on sampling and analysing minimal subsets of data. Compared to traditional parameter estimation methods in the statistics literature, RANSAC is simple but more effective in dealing with outliers. Figure 2.1 illustrates an example on 2D

line fitting data with 60% outliers, where the classical least squares and least median squares methods fail while RANSAC achieves a reasonably good estimate of the line.

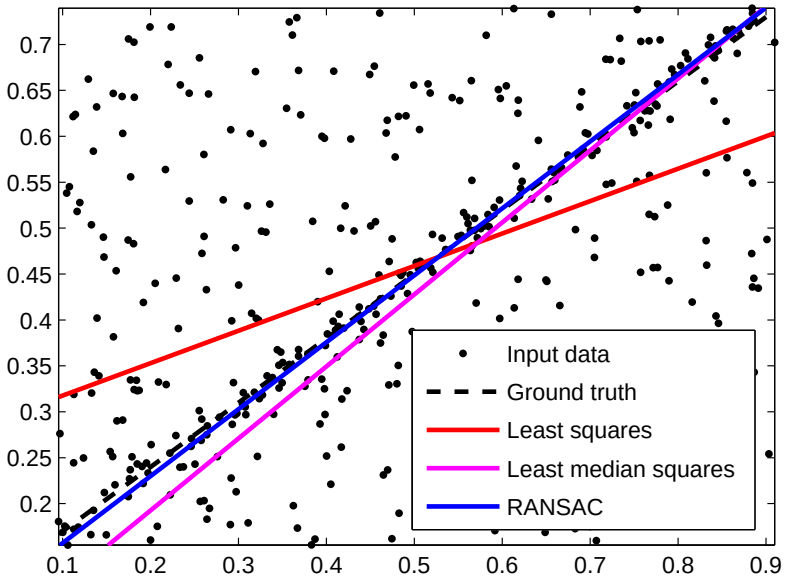


Figure 2.1: Performance comparison of least squares, least median squares and RANSAC on 2D line fitting data with 60% outliers.

RANSAC iteratively performs the two following steps: *(1)* it first generates a model hypothesis by randomly sampling a minimal subset of data (e.g., four data points for a homography matrix estimate), and *(2)* it then scores the model hypothesis by calculating the number of data points with residuals less than a predefined inlier threshold. The set of such data points (inliers) is called the *consensus set* of the hypothesis. The output of RANSAC is the maximum consensus set found among all generated hypotheses, as well as an estimate of the model parameters based on this set of inliers. In particular, the inliers within the maximum consensus set can be used with uniform weights or truncated quadratic weights to compute the model estimate. For example, in Figure 2.1, the line attributed to RANSAC was estimated by using the direct linear transformation technique [45] on the maximum consensus set obtained among 100 randomly generated hypotheses, and the inliers within the maximum consensus set were given the same weight in computing the line estimate. In some cases, the RANSAC estimate is further refined by using iterative algorithms such as Levenberg-Marquardt, whose starting point is the RANSAC estimate and errors are distributed over all data rather than a minimal or small subset.

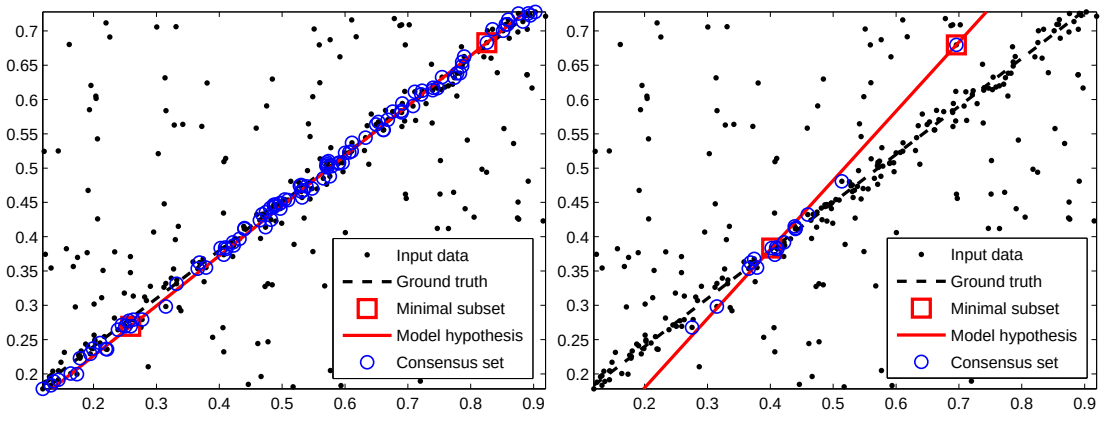
While RANSAC makes a hard distinction between inliers and outliers and maximises the number of inliers achieved, there are variants of RANSAC [46, 47] which make a *soft* decision and employ different criteria. These methods associate a value between 0 and 1 to each data point. In particular, MSAC [46] assigns a weight computed using robust weight functions to each datum, and searches for the hypothesis maximising the sum of weights of all data. In MLESAC [47], the likelihood of each datum being an inlier or outlier is calculated, and the goal is to seek the hypothesis offering the maximum likelihood of all data. These modified criteria can replace the

maximum consensus criterion of RANSAC with little extra cost, and following RANSAC, these methods also rely on random sampling of minimal subsets for generating hypotheses.

The number of iterations in the above *hypothesise-and-verify* procedure of RANSAC is frequently calculated such that with a high probability at least one outlier-free (*all-inlier*) minimal subset is obtained. The underlying assumption is that if the minimal subset contains no outliers then the generated hypothesis is more likely to be a good fit to the data (i.e., with a large consensus set); see Figure 2.2 for an example on 2D line fitting. Due to this assumption, the number of samples required to be drawn is highly related to the size of a minimal subset m and the ratio of outliers over the entire data ϵ. Assuming that the minimal subset size m and the outlier ratio ϵ are known, the probability of selecting m inliers (i.e., an all-inlier minimal subset) at random is $P_I = (1-\epsilon)^m$, and hence the probability of obtaining an outlier-contaminated (*impure*) minimal subset is $P_O = 1 - (1-\epsilon)^m$. Suppose that we draw N samples, the probability of success (i.e., at least one all-inlier sample among N randomly drawn samples) is computed as $\alpha = 1 - (1 - (1-\epsilon)^m)^N$. In practice, the probability of success α is fixed at a value close to 1.0 (e.g., 0.95), thus we can deduce the number of samples needed to draw such that with a probability α at least one all-inlier sample is obtained as

$$N = \frac{\log(1-\alpha)}{\log(1-(1-\epsilon)^m)}. \tag{2.1}$$

Let the probability of success α = 0.95, the estimated N under different settings of ϵ and m is presented in Table 2.1. It is clear that the required number of samples N increases exponentially with the minimal subset size m and the outlier ratio ϵ. In particular, it is very difficult to select an all-inlier minimal subset at random when the minimal subset size m is large or the outlier ratio ϵ is high. Therefore, randomly sampling minimal subsets in which each datum has an equal likelihood of being selected can become intractable in practice.



(a) All-inlier minimal subset **(b)** Impure minimal subset

Figure 2.2: Consensus sets of an all-inlier minimal subset and an impure minimal subset on 2D line fitting.

	$\epsilon = 20\%$	$\epsilon = 35\%$	$\epsilon = 50\%$	$\epsilon = 65\%$	$\epsilon = 80\%$	$\epsilon = 90\%$
$m = 2$	3	6	11	23	74	299
$m = 3$	5	10	23	69	373	2995
$m = 4$	6	16	47	199	1871	29956
$m = 5$	8	25	95	569	9361	299572
$m = 6$	10	39	191	1629	46807	2995731
$m = 7$	13	60	382	4655	234041	29957322
$m = 8$	17	93	766	13302	1170207	299573225

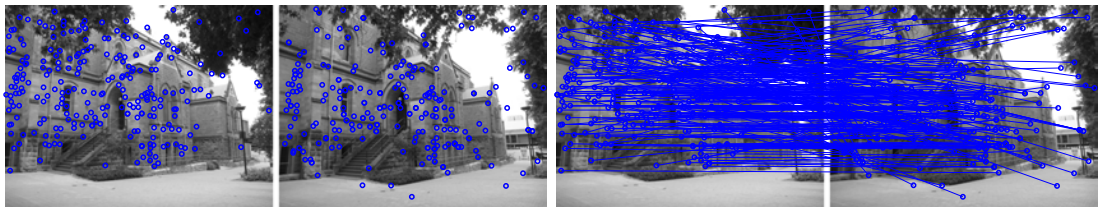
Table 2.1: The number of samples required by RANSAC with different minimal subset sizes and outlier ratios.

Since random sampling is not efficient in cases of high-dimensional models or highly contaminated data, and also the accuracy of the estimated model depends on the ability to obtain good hypotheses which are fitted on all-inlier minimal subsets, many alternative sampling strategies have been proposed in the computer vision literature. In this section, we will review various robust geometric fitting methods which rely on the hypothesise-and-verify framework of RANSAC and aim to accelerate the retrieval of all-inlier minimal subsets or reduce the number of samples required. Such sampling methods (known as *guided sampling methods*) can be categorised into four major groups. In the following, we will discuss the main ideas and some representative methods of each group.

2.2.1.1 Matching scores

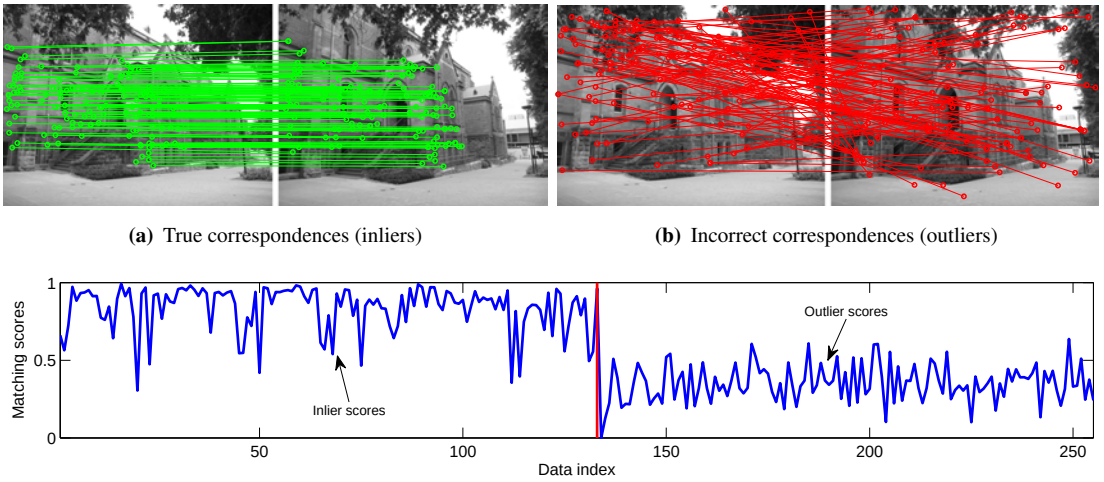
As mentioned in Chapter 1, feature detection and matching algorithms [16, 48, 49] are usually used to reduce the large variability of image data and capture the key elements that are relevant to a particular task in computer vision. More specifically, these methods detect the local salient features (keypoints) in each individual image and establish the correspondences across the images. In numerous parameter estimation problems in computer vision, the input data are a set of keypoint correspondences between the images. For instance, Figure 2.3 shows the correspondences between two images computed by using SIFT [16], which can become the input data for fundamental matrix estimation (the images are taken from the publicly available real datasets of [50]). The quality of a particular correspondence is often measured as a numeric score which indicates how similar its keypoint in one image is compared to the corresponding keypoint in another image.

The first group of guided sampling methods [22, 35, 51, 52] exploits such matching scores to bias the procedure of minimal subset sampling. The assumption behind theses methods is that the inliers (true correspondences) often have higher matching scores than the outliers (incorrect correspondences). For example, Figure 2.4 illustrates the matching scores of the correspondences in Figure 2.3, in which the inlier scores are frequently larger than the outlier scores.



(a) Detected keypoints **(b)** Keypoint correspondences

Figure 2.3: Detected keypoints and keypoint correspondences between two images obtained by using SIFT [16].



(a) True correspondences (inliers) **(b)** Incorrect correspondences (outliers)

(c) Matching scores

Figure 2.4: Matching scores of the keypoint correspondences in Figure 2.3.

Due to the above assumption, Guided-MLESAC [22] uses the keypoint matching scores to compute the prior inlier probabilities so that the data points with higher scores are more likely to be selected for forming minimal subsets, hence increasing the chance of hitting all-inlier samples. In [35], PROSAC sorts the data points according to their descending matching scores. It first generates minimal subsets within a small set of top-ranked data points, which have a high probability of being inliers, and then gradually expands this sampling set by adding data points with lower scores. PROSAC speeds up the sampling process since all-inlier minimal subsets can be drawn easily from top-ranked data points in the early sampling stages. Moreover, it converges towards the traditional RANSAC if the matching scores are incorrect.

Another representative method in this group is SWIGS [52], which further assumes that each correspondence has its own correct and incorrect matching score distributions. Given a keypoint $\mathbf{x}_i$ in one image, SWIGS considers the set of k highest matching scores $\{s_{i,1}, s_{i,2}, \ldots, s_{i,k}\}$ between $\mathbf{x}_i$ and the keypoints in another image. The basic idea of SWIGS is that the first element $s_{i,1}$ has a high chance to be produced by a true correspondence (inlier), while the last $k-1$ matching scores $\{s_{i,2}, \ldots, s_{i,k}\}$ are likely to be samples from the incorrect matching score distribution and hence used to model this distribution (i.e., with a Rayleigh distribution).

The correct matching score distribution is then represented by the Rayleigh's complementary cumulative distribution function and used to predict the inlier probability of $s_{i,1}$. These inlier probabilities are employed to guide the sampling process as in Guided-MLESAC.

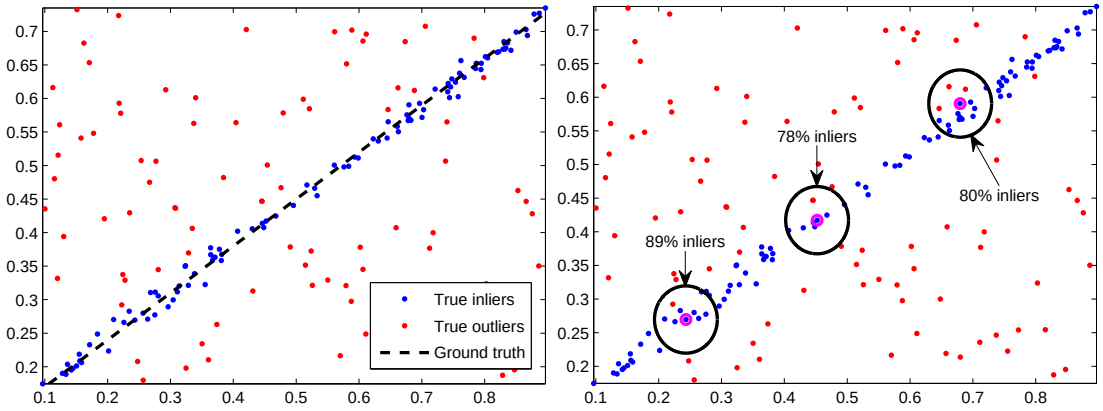
The methods in this group are computationally efficient since they only rely on the prior matching scores. However, in some applications the matching scores may not be available. Moreover, these methods can fail if the assumption of the monotonic relationship between matching scores and inlier probabilities is not satisfied. This situation often happens in urban scenes with repetitive textures, which eventually give rise to many correspondences that are well matched (i.e., with high scores) in terms of their local appearance but incorrect according to the geometric structure of the scene.

2.2.1.2 Spatial information

The second group of methods [33, 34, 53, 54] explores the spatial information of the data to guide the process of minimal subset sampling. The spatial knowledge can be the assumption that the inliers are closer to each other (than the outliers) in the spatial domain; thus a local region around an inlier yields a higher inlier ratio than the set of all data (e.g., [33, 34, 53]). Alternatively, a spatial grouping structure of the data is available, where some groups have a larger percentage of inliers than the entire set of data (e.g., [54]).

In many cases, the inliers inhabit a lower dimensional manifold than the outliers, e.g., in the case of 2D line fitting, the inliers mostly lie on a 1D manifold whereas the outliers spread over a 2D space. This difference in dimensionality leads to the idea that some regions of the spatial domain are more densely populated by inliers, and these regions thus have higher inlier rates than the set of all data. Based on this idea, in [33, 34], various proximity sampling strategies were proposed to sample minimal subsets from a local region, which is defined as a hypersphere of radius r centred on a randomly sampled data point. If the chosen datum (hypersphere centre) is an inlier, its neighbouring data points are expected to be inliers; see Figure 2.5 for an example on 2D line fitting. NAPSAC [33] uniformly selects the data points within the hypersphere to form minimal subsets. Similarly, the method of [34] generates minimal subsets within the hypersphere, but unlike the uniform sampling scheme in NAPSAC, it uses a weighted sampling procedure where the weight of each datum is a Gaussian function of its distance to the hypersphere centre (thus the data points lie closer to the centre have a higher chance of being selected).

These methods [33, 34] often have a high probability of hitting all-inlier minimal subsets. However, they can break down if the scene is ambiguous, e.g., when there are many repeated patterns such as windows on building facades. Due to the scene ambiguities, feature detection and matching algorithms produce many incorrect correspondences (outliers) which can mix closely with the inliers. To address this issue, SCRAMSAC [53] introduces the spatial consistency check to



(a) Input data with 50% inliers **(b)** High inlier ratio in the local region around an inlier

Figure 2.5: High inlier ratio in the local region around an inlier on 2D line fitting.

remove mismatches from the input set of correspondences prior to the sampling process. Figure 2.6 shows an example of the spatial consistency check, where the images are taken from the publicly available real datasets of [50]. In particular, given a reference correspondence $(\mathbf{x}_1, \mathbf{x}_2)$ between two images, SCRAMSAC first defines a circular region C_1 around the keypoint $\mathbf{x}_1$, and a similar region C_2 for the corresponding keypoint $\mathbf{x}_2$. A particular correspondence is considered to have passed the spatial consistency test if its keypoints respectively appear inside C_1 and C_2. This spatial consistency check can identify and remove incorrect matches that do not satisfy the geometric structure of the scene, and hence it returns a smaller set of more confident correspondences which has a high inlier ratio. Moreover, since the reduced set of correspondences is used not only for generating minimal subsets but also for verifying model hypotheses, the run time of SCRAMSAC is further decreased compared to RANSAC.

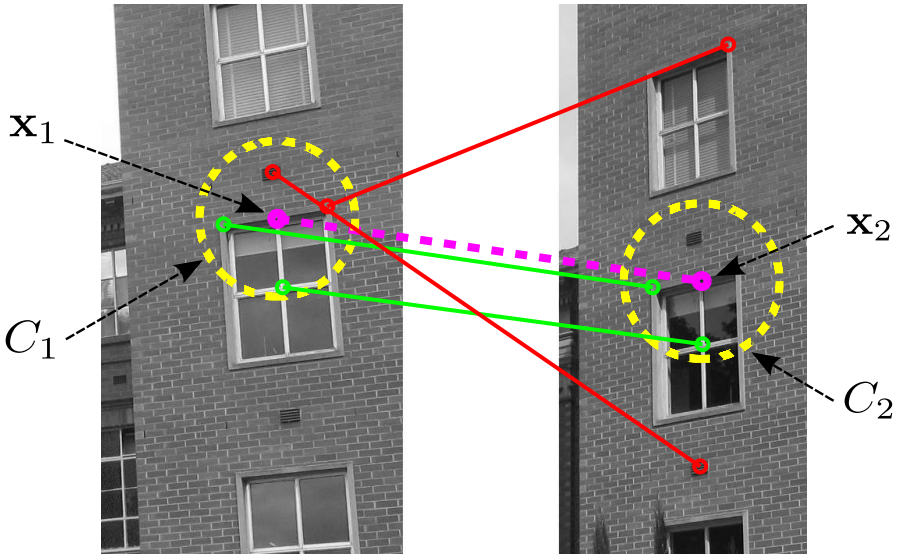
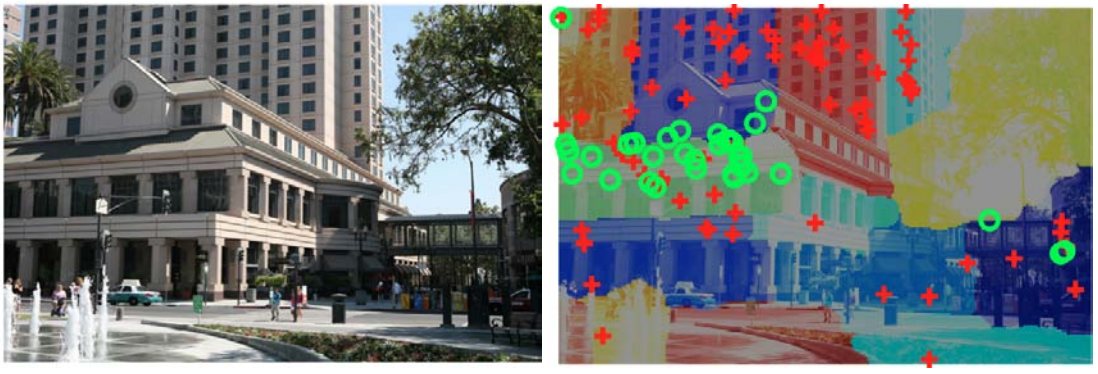


Figure 2.6: Spatial consistency check in SCRAMSAC. A reference correspondence is in magenta. The spatially consistent correspondences are in green, and others are in red.

Another representative method in this group is GroupSAC [54], which assumes that there exists a spatial grouping structure of the data, in which some of the groups have a high percentage of inliers while the others are mostly contaminated by outliers. For example, in two-view geometry estimation, such grouping information can be obtained by clustering the set of correspondences according to the result of image segmentation or optical flow estimation. An example of grouping using image segmentation is presented in Figure 2.7 (the images are taken from [54]), where there is a clear correlation between the image segmentation and inlier/outlier classification, and the segment colored in green has a considerably higher inlier percentage than the others. The basic idea of GroupSAC is that if the minimal subset is drawn from a fewer number of groups, it has a higher probability of containing only inliers. Therefore, GroupSAC focuses on drawing minimal subsets from a single group first, and then progressively increases the number of groups involved by adding more groups to the sampling set. Similar to PROSAC, GroupSAC has a high chance of obtaining all-inlier minimal subsets which involve a fewer number of groups in the early sampling steps, and it degrades to the classical RANSAC if the grouping information is wrong.



(a) One of the two images for fundamental matrix estimation **(b)** Correspondences clustered based on image segmentation

Figure 2.7: Spatial grouping of correspondences by using image segmentation in GroupSAC. True correspondences (inliers) are in green while incorrect correspondences (outliers) are in red.

When the spatial assumptions are correct, the guided sampling methods in this group speed up the retrieval of all-inlier minimal subsets compared to RANSAC. However, these assumptions are not always satisfied in practice. For instance, the neighbourhood of an inlier may unavoidably include many outliers (i.e., in richly textured scenes or heavily contaminated data), or the grouping knowledge is not available or does not provide a meaningful segmentation of the data. In such cases, obtaining all-inlier minimal subsets becomes difficult and time-consuming.

2.2.1.3 Local exploitation

The third group of guided sampling methods [20, 21, 55–57] exploits the good hypotheses which were generated previously in the sampling process to increase the probability of hitting all-inlier minimal subsets and improve the accuracy of the estimated parameters. These methods typically consist of the two following procedures: *(1)* local exploitation in the parameter space to improve the chance of finding all-inlier samples and refine the model estimate if a nearby solution is better, and *(2)* global exploration in the parameter space to quickly escape from the local minima and seek other promising solutions.

The first method that was proposed in this group is LO-RANSAC [20, 21], in which the random sampling procedure of RANSAC is employed for global exploration. When the global exploration obtains a new hypothesis that has a larger consensus set than those generated previously, the local exploitation is triggered. In LO-RANSAC, the local exploitation first generates a fixed number of model hypotheses by sampling the data points within the largest consensus set found so far, and an iterative least squares process is then conducted on each generated hypothesis to refine the model estimate. Since the consensus set usually has a larger inlier percentage than the set of all data, many all-inlier minimal subsets can be drawn during the local exploitation steps. In [57], an improved version of LO-RANSAC (i.e., LO$^+$-RANSAC) was presented. In particular, the improvements in accuracy and run time are made (respectively) by using a truncated quadratic cost function (instead of the classical 0-1 cost function of RANSAC) and a limit on the number of data points that participate in the iterative least squares process.

The idea of LO-RANSAC is further extended in the recent work of [55, 56]. Motivated by the methods surveyed in Section 2.2.1.1, these methods [55, 56] exploit the matching score information to guide the global exploration. In BEEM [55], the global exploration is based on the distance ratios of the closest to the second closest keypoint of the SIFT correspondences while the local exploitation of LO-RANSAC is employed. Similarly, BLOGS [56] uses the similarity weights of the SIFT correspondences to bias the global exploration while the local exploitation is guided by the conditional probabilities of correspondences given the best hypothesis obtained so far (i.e., computed by using the Joint Feature Distribution [58]). Both BEEM and BLOGS aim to balance between the local exploitation and global exploration steps. More specifically, these methods terminate the local exploitation if no better hypotheses can be found, and go back to the global exploration to search for other possible solutions. Therefore, they can avoid spending too much time on unpromising local minima in the parameter space.

The guided sampling methods in this group speed up the retrieval of all-inlier minimal subsets and improve the accuracy of the estimated model. However, the success of these methods depends on having an optimal inlier threshold to correctly qualify the generated hypotheses. In practice, the inlier threshold is data-dependent and difficult to determine, due to having no prior

knowledge of the inlier noise scale. For BEEM and BLOGS, their performance further relies on the availability of matching scores and the correlation between matching scores and inlier probabilities (see Section 2.2.1.1).

2.2.1.4 Conditional sampling

The last group of guided sampling methods [36, 37, 59, 60] focuses on a conditional sampling strategy for minimal subset generation. More specifically, these methods aim to choose the next datum for forming a minimal subset, by considering the data points that have already been selected into the minimal subset. For instance, suppose that we have chosen two data points into a minimal subset, then the third datum for the minimal subset is sampled conditionally on the first and the second data points.

In BetaSAC [59], to select the next datum for a minimal subset, it first chooses k data points (i.e., $k < n$, where n is the total number of data points) at pure random, and orders these data points with respect to those that have already been selected into the minimal subset (e.g., by their descending distances to the first datum added to the minimal subset). The selection of a datum among the k sorted data points is then modelled as a random variable of the Beta distribution. The advantages of BetaSAC are that any prior information (e.g., matching scores) can be included to order the data points and that it does not require a complete ranking of the data points (e.g., as in PROSAC).

The most representative method in this group is Guided Sampling for Multi-Structure Model Fitting (Multi-GS) [36, 37], which computes the conditional sampling weights based on the pairwise data similarity. In Multi-GS, the data similarity is measured by comparing the data preferences towards the set of hypotheses that have been generated so far in the sampling process. In particular, the preference of a datum is obtained by sorting its residuals (to the hypotheses) in the ascending order. The basic idea of Multi-GS is that the inliers have more overlaps in their preferences (than the outliers), which yields higher similarity values (i.e., preference correlation values) between the inliers than the outliers; see Figure 4.5 for an example. Due to this preference-based similarity measure, Multi-GS can effectively distinguish the inliers from the outliers.

In Multi-GS, the first datum of a minimal subset is chosen purely randomly, and the similarity values between the remaining data points and the first datum are employed as the conditional sampling weights to select the second datum. The third datum is then sampled based on the conditional sampling weights which are computed by multiplying the corresponding similarity values of each datum with respect to the first and the second data points, and likewise for sampling the subsequent data points into the minimal subset. Given that the previously chosen data points in a minimal subset are the inliers, the conditional sampling weights ensure that the

remaining inliers have higher probabilities of being selected for the minimal subset than the outliers. Therefore, all-inlier samples can be drawn easily, even in cases of high-dimensional models (e.g., fundamental matrix).

Compared to the other methods presented above, Multi-GS can effectively generate all-inlier minimal subsets without relying on any prior knowledge (e.g., matching scores, inlier threshold) or data-dependent assumption (e.g., spatial information). However, it is less computationally efficient than the previously discussed methods since the conditional sampling weights are updated on-the-fly during the sampling procedure. Moreover, the conditional sampling strategy in Multi-GS is only worthwhile for obtaining all-inlier samples if the first datum selected into a minimal subset is an inlier. One solution to this issue is to sample the first datum of a minimal subset based on the keypoint matching scores as in Guided-MLESAC/PROSAC (Section 2.2.1.1) while the rest of the minimal subset is sampled using the conditional sampling weights of Multi-GS.

2.2.2 Sampling minimal subsets with large span

The previous section surveyed numerous robust geometric fitting methods which rely on sampling and testing model hypotheses fitted on minimal subsets of data. Many of these methods [20–22, 33] estimate the number of samples required to be drawn by using Equation (2.1) such that with a high probability at least one all-inlier minimal subset is obtained for yielding a satisfactory model estimate. However, it has been observed in [17, 20–23] that the number of samples predicted by RANSAC (Equation (2.1)) is too small for receiving a good model estimate that fits the data well. In particular, the predicted number of samples is merely a lower bound of the number of samples required [17]. In [21], a factor ranging from two to three between the required and the predicted number of samples was reported for models with two parameters. Also, for models consisting of four parameters, a factor of ten was presented in [22].

The reason for the above differences is due to the fact that generating model hypotheses from minimal subsets amplifies the noise magnitude [17]. Moreover, as described intuitively in [22], with noisy data, it is not enough to obtain a sample containing only inliers; they should be the inliers that span the object so that many of the remaining data points are compared with interpolated values. In other words, in the presence of noise in the data, retrieving a single all-inlier minimal subset is not sufficient to guarantee a satisfactory model estimate; the inliers therein should also have a large spatial extent. Figure 1.1 illustrates an example of this principle. Some methods [23, 61] have been proposed in the literature to tackle the problem of large span sampling. In the following, we will review and discuss these techniques.

The first method to address the problem of large span sampling is SURSAC [23], which aims to sample inliers that fully cover the structure for robust geometric fitting. The common stopping criterion of RANSAC (Equation (2.1)), which ensures with a high probability an all-inlier minimal subset is obtained, is not enough; a "relatively loose" sufficient condition employed by SURSAC is that the inliers should also fully span the inlier domain. More specifically, assuming that the inlier domain is divided into m subregions, a minimal subset of m data points is considered to have a sufficient coverage of the inlier domain if each of the subregions contains only one data point (i.e., the inliers equally distribute over the inlier domain) [23].

The basic idea of SURSAC is to modify the predicted number of samples by RANSAC to include the above requirement for a sufficient span of the inlier domain. We first compute the probability of receiving a minimal subset with a sufficient coverage of the inlier domain. The problem can be described as putting a set of m items (data points) into m bins (subregions) where each item is placed sequentially such that the i-th item ($i = 1, 2, \ldots, m$) should be put into one of the remaining $m - i + 1$ empty (available) bins. In SURSAC, assuming that the first $i - 1$ steps of this process are successful (i.e., there exists no two items in the same bin), the probability of success at the i-th step is $\frac{m-i+1}{m}$, and hence the probability that the entire process is successful can be deduced as

$$P_C = \prod_{i=1}^{m} \frac{m-i+1}{m} = \prod_{i=1}^{m} \frac{i}{m} = \frac{m!}{m^m}. \tag{2.2}$$

The probability of retrieving an all-inlier minimal subset that fully covers the inlier domain is computed by multiplying the probability P_I of obtaining an all-inlier minimal subset (see Section 2.2) and the probability P_C that the minimal subset fully spans the inlier domain as $P_{IC} = P_I.P_C = (1-\epsilon)^m.\frac{m!}{m^m}$. Therefore, we can update the predicted number of samples by RANSAC (Equation (2.1)) such that with a probability α at least one all-inlier minimal subset with a sufficient coverage of the inlier domain is received as

$$N' = \frac{\log(1-\alpha)}{\log\left(1-(1-\epsilon)^m.\frac{m!}{m^m}\right)}. \tag{2.3}$$

The predicted number of samples by SURSAC (Equation (2.3)) explains the discrepancies in the required number of samples that were found empirically by [21, 22] very well. Moreover, compared to RANSAC, SURSAC can retrieve all-inlier minimal subsets that fully cover the inlier domain and hence improve the quality of the estimated parameters. However, it dramatically raises the number of samples required for a good model estimate. In particular, with the probability of success $\alpha = 0.95$, the estimated values of N' under different settings of the outlier ratio ϵ and the minimal subset size m are presented in Table 2.2, which shows an exponential increase compared to the corresponding values of N in Table 2.1. Thus, sampling all-inlier minimal subsets with large span by using SURSAC is intractable in practice.

	$\epsilon = 20\%$	$\epsilon = 35\%$	$\epsilon = 50\%$	$\epsilon = 65\%$	$\epsilon = 80\%$	$\epsilon = 90\%$
$m = 2$	8	13	23	48	149	598
$m = 3$	25	48	107	313	1684	13480
$m = 4$	77	178	510	2128	19971	319544
$m = 5$	237	671	2495	14853	243792	7801385
$m = 6$	740	2573	12423	105600	3033178	194123451
$m = 7$	2333	9984	62656	760820	38242719	4895068133
$m = 8$	7429	39119	319110	5535514	486925372	124653097953

Table 2.2: The number of samples required by SURSAC with different minimal subset sizes and outlier ratios.

Another approach to the problem of large span sampling is RCM [61], which samples and fits on large data clusters (instead of minimal subsets). The basic idea of RCM is that larger-than-minimal subsets of inliers usually have a larger span of the inlier domain, and hence a better estimate of the model parameters, compared to all-inlier minimal subsets. Figure 2.8 shows an example on 2D line fitting, in which a larger-than-minimal subset of inliers offers a more accurate estimate of the line than a minimal subset consisting of two inliers (the data in Figure 1.1 were used in this example).

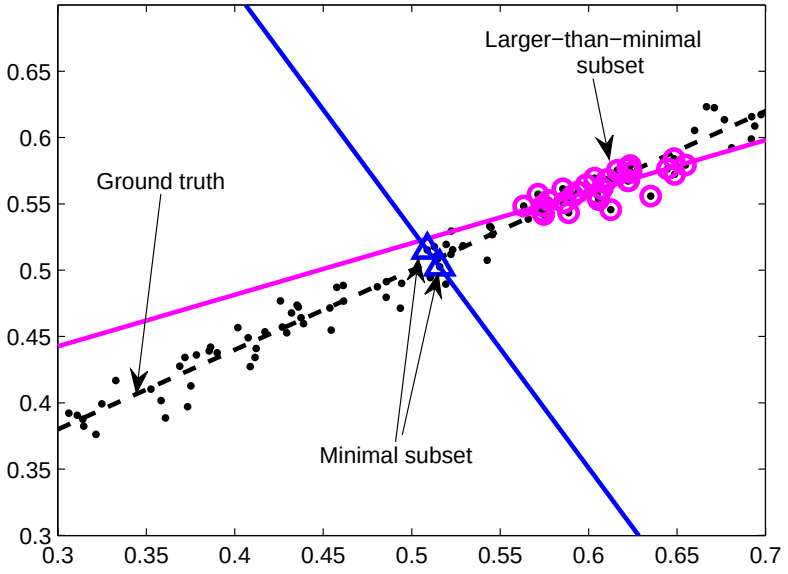


Figure 2.8: A minimal subset and a large-than-minimal subset with different fitting qualities on 2D line fitting.

RCM follows the method of Swendsen and Wang [62], which is often used in statistical mechanics for coupled spin system simulations, to generate large data clusters. In particular, it first builds a sparse adjacency graph on the set of all data points using Delaunay triangulation. A random variable (bond variable) is attached to each of the graph edges, and then sampled to produce the connected components or data clusters with large span which are used for computing the model hypotheses. Due to the spatial smoothness of the data, large clusters of inliers can be found without an exponential increase in sampling efforts. Moreover, the pairwise data similarity values (i.e., preference correlation values) of Multi-GS (Section 2.2.1.4) are employed as the

weights for sampling the bond variables and updated on-the-fly such that meaningful clusters can be obtained efficiently.

RCM can sample large clusters of inliers, which increases the accuracy and efficiency of robust geometric fitting. However, it imposes spatial smoothness on the data (i.e., the inliers must form a single coherent structure). This condition is clearly not the general case in practice (e.g., a single line may consist of multiple distant line segments), and it is more restrictive than the spatial assumption employed by the methods in Section 2.2.1.2 (i.e., which requires the inliers are locally dense). In this thesis, we aim to provide a more fundamental study into the problem of large span sampling in robust geometric fitting using minimal subsets (Chapters 3 and 4).

2.3 Deformable registration

2.3.1 The assumption of highly nonlinear correspondence manifold

We have surveyed in the previous section various methods for robust geometric fitting, where the geometric model of interest has a *fixed* number of degrees of freedom. Due to the known number of degrees of freedom, a model estimate can be computed by using a minimal subset of data, and hence sampling and testing model hypotheses fitted on minimal subsets is often employed to deal with outliers.

In this section, we depart from robust geometric fitting and concentrate on robust deformable registration, which also requires the knowledge of robust parameter estimation with vision data. The information of interest is the nonlinear transformation that aligns regions in two or more images containing a common object that can deform smoothly (e.g., a waving flag), and is represented by a deformable model (e.g., a TPS warp). Various deformable registration methods (e.g., [29, 30, 42]) rely on detecting and matching keypoints between the images, which are then used to compute the transformation parameters. An important issue in such *feature-based methods* is the identification and rejection of outliers (incorrect correspondences), which invariably appear due to imperfect feature detection and matching. Traditional deformable registration methods such as [24, 63] focus on modelling the deformations and handling noise, and thus are vulnerable to outliers.

Despite the fact that robust geometric fitting has a long history in both the statistics and computer vision literatures, research efforts have only recently been invested in robust deformable registration. Similar to robust geometric fitting, an important part of robust deformable registration is to tackle outliers. Common sense suggests that standard robust geometric fitting techniques such as RANSAC are not suitable for outlier rejection in deformable registration because of the substantial nonlinearity of the correspondence manifold on which the inliers (true correspondences)

lie; see Figure 1.2 for such impressions of the data in the literature. In particular, the underlying transformation between the images is assumed to have an unknown and varying number of degrees of freedom, and hence the size of a minimal subset for yielding a model estimate can not be determined. Moreover, it is widely assumed that many realistic deformations (e.g., a bending piece of paper) are too nonlinear to be amendable to simple geometric modelling. Therefore, the assumption of the highly nonlinear correspondence manifold precludes the application of robust geometric fitting techniques for outlier rejection in deformable registration.

2.3.2 Outlier rejection with highly nonlinear correspondence manifold

The input data for deformable registration typically consist of two images corresponding to a target object, i.e., a template in which the object deformations are negligible and an input image in which the object is deformed nonrigidly; Figure 2.9 illustrates an example of the input data, where the images are taken from the publicly available real datasets of [64]. Let $\mathcal{X} = \{(\mathbf{u}_i, \mathbf{v}_i)\}_{i=1}^{n}$ be a set of keypoint correspondences between the template and input image, where point $\mathbf{u}_i = [x_i \; y_i]^T$ in the template is matched with point $\mathbf{v}_i = [x_i' \; y_i']^T$ in the input image. From $\mathcal{X}$, the goal of deformable registration is to learn a nonlinear function $f : \mathbf{p} \mapsto \mathbf{q}$ that maps point $\mathbf{p} = [x \; y]^T$ in the template to point $\mathbf{q} = [x' \; y']^T$ in the input image.

Due to the assumption of the highly nonlinear correspondence manifold described in the previous section, the function f can have a significant nonlinearity. Most of the previous approaches to outlier rejection in deformable registration are based on optimising fully deformable models, which are used to represent f. In the following, we will review and discuss these methods.

2.3.2.1 Triangulated mesh

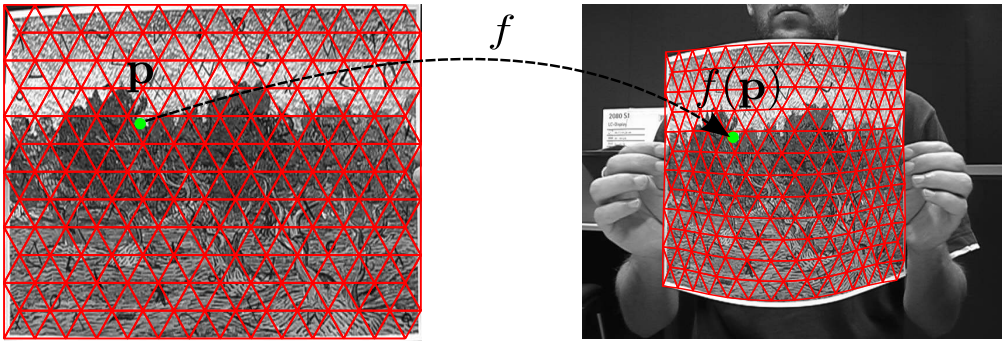
The first approach [29, 30, 38] employs a triangulated mesh to explicitly model the deformations and reject outliers. In particular, the template is modelled by a 2D triangulated mesh with hexagonally connected vertices; see Figure 2.9 for an example of the template mesh. Each vertex $\mathbf{m}_i$ is specified by its image coordinates, and the mesh shape is thus controlled by a shape vector $\mathbf{S}$ which is formed by stacking the coordinates of all vertices. Given an image point $\mathbf{p}$ that lies in a triangle $(\mathbf{m}_i, \mathbf{m}_j, \mathbf{m}_k)$ of the undeformed mesh $\mathbf{S}$ (the template), and its barycentric coordinates (b_1, b_2, b_3) that are computed with respect to $(\mathbf{m}_i, \mathbf{m}_j, \mathbf{m}_k)$, the transformation f that maps $\mathbf{p}$ in the undeformed mesh $\mathbf{S}$ to the deformed mesh $\mathbf{S}'$ (the input image) is defined as $f(\mathbf{p}) = b_1.\mathbf{m}_i' + b_2.\mathbf{m}_j' + b_3.\mathbf{m}_k'$, where $(\mathbf{m}_i', \mathbf{m}_j', \mathbf{m}_k')$ in $\mathbf{S}'$ correspond respectively to $(\mathbf{m}_i, \mathbf{m}_j, \mathbf{m}_k)$ in $\mathbf{S}$.

Given a set of keypoint correspondences $\mathcal{X} = \{(\mathbf{u}_i, \mathbf{v}_i)\}_{i=1}^{n}$, the goal of [29, 30, 38] is to deform the mesh $\mathbf{S}'$ such that for the subset $\mathcal{I} \subset \mathcal{X}$ of true correspondences the sum of squared residuals

of the correspondences in $\mathcal{I}$ is minimised while the deformations of $\mathbf{S}'$ remain as smooth as possible; see Figure 2.9 for an example of the deformed mesh. More specifically, $\mathbf{S}'$ is deformed by minimising the following objective function

$$E(\mathbf{S}') = E_d(\mathbf{S}') + \lambda_r E_r(\mathbf{S}'), \tag{2.4}$$

where E_d is the data term that takes into account the residuals of the correspondences, E_r is the regularisation term that preserves the smoothness of the mesh, and $\lambda_r > 0$ is a constant.



Template mesh Deformed mesh

Figure 2.9: Examples of a template mesh and a deformed mesh.

Denote $(\overline{i,j,k})$ as a vertex index triple such that $(\mathbf{m}_i, \mathbf{m}_j, \mathbf{m}_k)$ form two colinear connected edges in the undeformed mesh $\mathbf{S}$; see Figure 2.10(a). Since $\mathbf{S}$ has equidistant vertices, we have $\mathbf{m}_j - \mathbf{m}_i = \mathbf{m}_k - \mathbf{m}_j, \forall(\overline{i,j,k})$. The regularisation term E_r approximates the squared directional curvatures of the deformed mesh $\mathbf{S}'$ as long as the vertices remain roughly equidistant and increases with the length difference of every two colinear connected edges as

$$E_r(\mathbf{S}') = \frac{1}{2} \sum_{(\overline{i,j,k})} \left\| (\mathbf{m}'_j - \mathbf{m}'_i) - (\mathbf{m}'_k - \mathbf{m}'_j) \right\|^2, \tag{2.5}$$

where $(\mathbf{m}'_i, \mathbf{m}'_j, \mathbf{m}'_k)$ in $\mathbf{S}'$ correspond respectively to $(\mathbf{m}_i, \mathbf{m}_j, \mathbf{m}_k)$ in $\mathbf{S}$; see Figures 2.10(b) and 2.10(c). Therefore, E_r prevents the deformed mesh $\mathbf{S}'$ from overfitting the data in the presence of outliers.

The data term E_d tends to deform the mesh $\mathbf{S}'$ so that the keypoints in the template are matched with their correspondences in the input image. In particular,

$$E_d(\mathbf{S}') = -\sum_{i=1}^{n} \rho(\|\mathbf{v}_i - f(\mathbf{u}_i)\|, r), \tag{2.6}$$

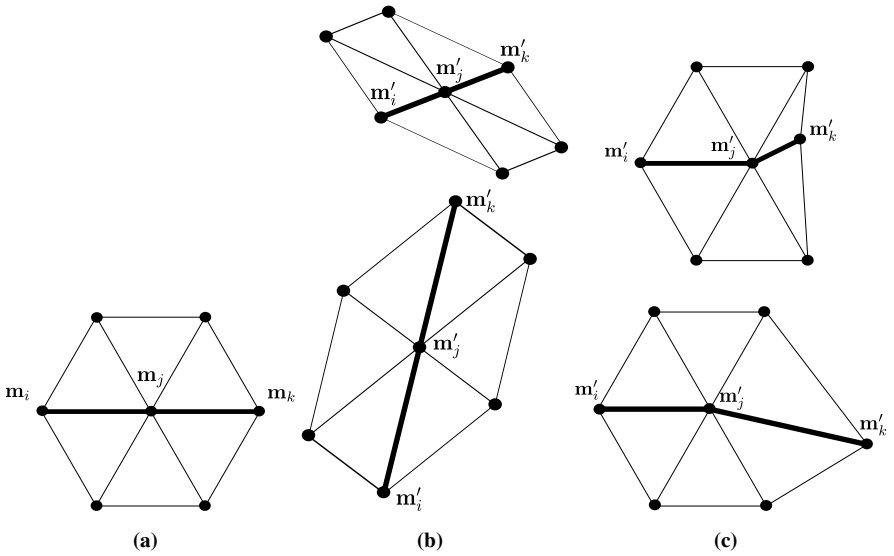


Figure 2.10: Regularisation for a 2D triangulated mesh with hexagonally connected vertices. (a) Two colinear connected edges. (b) Two deformations that are not penalised. (c) Two penalised deformations. Note that the deformations resulting from perspective projection in (b) are much less penalised than those caused by outliers in (c).

where ρ is a robust estimator which is defined by

$$\rho(\delta, r) = \begin{cases} \frac{3(r^2-\delta^2)}{4r^3} & \delta \leq r \\ 0 & \delta > r \end{cases}. \tag{2.7}$$

Here, the parameter r acts as an inlier threshold where the data points with residuals δ greater than r are eliminated as outliers and hence have no effects on the optimisation. When r is large, most of the correspondences are included as inliers and as r reduces, the robust estimator becomes narrower and more selective; see Figure 2.11.

Since the robust estimator introduced in the data term is not convex, the optimisation problem in Equation (2.4) becomes hard to solve. In [29, 38], a deterministic annealing procedure is proposed to tackle the optimisation. In particular, r is initialised with a large value and then progressively decayed at a constant rate. For each value of r, the objective function E is minimised by using the semi-implicit scheme (see [29, 38] for details), and the resultant mesh is employed as the initial state for the next iteration. The annealing procedure stops when r reaches a value close to the expected threshold. At the beginning when r is large, the gradients of E_r are comparatively larger than those of E_d, thus preventing the outliers from wrinkling the mesh while allowing the inliers to choose the global deformation. As r decreases, the gradients of E_d become larger hence the mesh starts bending and the influences of the outliers gradually reduce.

In [30], two contributions are made to speed up the above optimisation procedure. Unlike the semi-implicit scheme in [29, 38] which requires a few iterations for convergence at each value

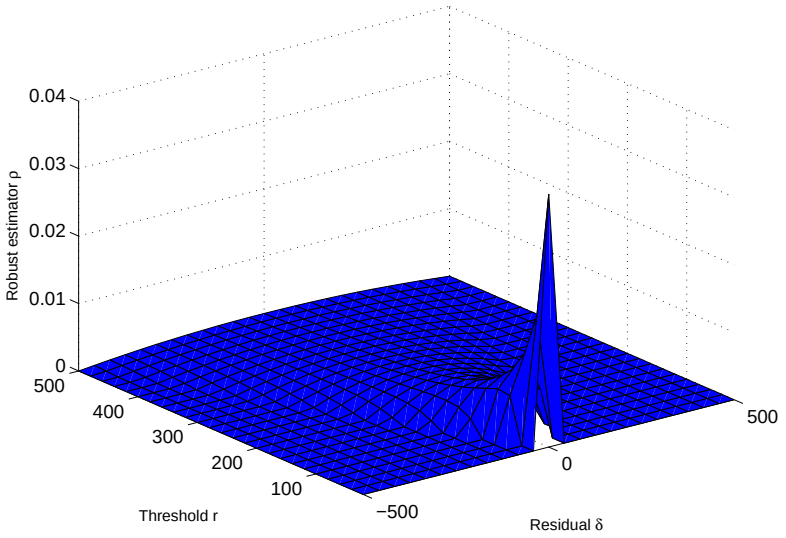


Figure 2.11: The robust estimator ρ is zero for residuals δ larger than threshold r and quadratic for elsewhere.

of r, the Gauss-Newton approximation in [30] takes a single iteration to achieve the convergence at each r. Moreover, the mesh is randomly initialised in [29, 38], which requires a sufficiently large initial value of r to avoid getting stuck at local minima. In contrast, [30] employs PROSAC (Section 2.2.1.1) which exploits the matching score information to provide a better initialisation, which reduces the initial value of r and hence the number of iterations.

In summary, these methods [29, 30, 38] can jointly optimise the triangulated mesh and eliminate outliers. However, the energy function is nonconvex due to the robust estimator. Local optimisation techniques such as semi-implicit scheme and Gauss-Newton approximation are employed to solve the problem but they only guarantee a locally optimal solution and require a relatively good parameter initialisation.

2.3.2.2 Gaussian process regression

The second approach [39] which also jointly estimates the deformations and removes outliers models the transformation f as a Gaussian process. The problem of deformable registration is thus formulated as a nonlinear regression problem, which aims to fit a Gaussian process f onto the set of keypoint correspondences $\mathcal{X} = \{(\mathbf{u}_i, \mathbf{v}_i)\}_{i=1}^n$. The keypoints $\{\mathbf{u}_i\}_{i=1}^n$ in the template and their correspondences $\{\mathbf{v}_i\}_{i=1}^n$ in the input image are taken, respectively, as the independent and dependent measurements for computing f, i.e., $\mathbf{v}_i = f(\mathbf{u}_i) + \boldsymbol{\epsilon}$; where $\boldsymbol{\epsilon}$ is an additive Gaussian noise with zero mean and variance σ^2.

Using a Bayesian approach to regression, the posterior distribution over the latent function f is

$$P(f|\mathcal{X}) \propto P(f)P(\mathcal{X}|f). \tag{2.8}$$

In [39], the likelihood probability $P(\mathcal{X}|f)$ represents how well the function f explains the observed data $\mathcal{X}$, and the prior probability $P(f)$ is specified by a Gaussian process with a mean function $m(\mathbf{u}) = \mathbf{u}$ and a covariance function $k(\mathbf{u}, \mathbf{u}') = \exp\left(-\frac{1}{2}\|\mathbf{u} - \mathbf{u}'\|^2\right)$. Therefore, we can rewrite the posterior distribution in (2.8) as

$$P(f|\mathcal{X}) \propto \mathcal{N}\left(f|m(\mathbf{u}), k(\mathbf{u}, \mathbf{u}')\right) \prod_{i=1}^{n} \mathcal{N}\left(\mathbf{v}_i|f(\mathbf{u}_i), \sigma^2\right). \tag{2.9}$$

Following the detailed derivations in [39], the function f that maps a point $\mathbf{p}$ in the template to the input image is expressed as

$$f(\mathbf{p}) = \mathbf{p} + \sum_{i=1}^{n} \boldsymbol{\alpha}_i^T k(\mathbf{p}, \mathbf{u}_i), \tag{2.10}$$

where $\boldsymbol{\alpha}_i$ is the i-th row of the coefficient matrix $\boldsymbol{\alpha}$ which is computed by

$$\boldsymbol{\alpha} = (\mathbf{WK} + \sigma^2\mathbf{I})^{-1}[\mathbf{W}(\mathbf{V} - \mathbf{U})]. \tag{2.11}$$

Here, $\mathbf{V} = [\mathbf{v}_1 \ \ \mathbf{v}_2 \ \ \ldots \ \ \mathbf{v}_n]^T$, $\mathbf{U} = [\mathbf{u}_1 \ \ \mathbf{u}_2 \ \ \ldots \ \ \mathbf{u}_n]^T$, $\mathbf{I}$ is the identity matrix, $\mathbf{K}$ is the kernel matrix with $\mathbf{K}_{ij} = k(\mathbf{u}_i, \mathbf{u}_j)$, and $\mathbf{W}$ is the diagonal weight matrix with $\mathbf{W}_{ii} = w_i$. The weight w_i is introduced to each correspondence to alleviate the influences of outliers and defined as

$$w_i = \frac{\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{v}_i - f(\mathbf{u}_i)\|^2\right)}{\sum_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{v}_i - f(\mathbf{u}_i)\|^2\right)}. \tag{2.12}$$

To jointly eliminate outliers and estimate the transformation f, the deterministic annealing procedure of [29, 30, 38] is employed. The parameter σ which acts as an inlier threshold is initialised with a large value, and then progressively decayed at a constant rate. At the beginning, all data points in $\mathcal{X}$ are used for initialising f. For each value of σ, the data points which have residuals less than σ are selected into the inlier set, which is used for estimating f by using Equation (2.10) until the inlier set no longer changes or the maximum number of iterations is exceeded. The resultant f is employed as the initial state for the next iteration. The annealing procedure terminates when σ reaches a value close to the expected threshold. As previously discussed in Section 2.3.2.1, this optimisation process is complicated due to the use of the robust weight in dealing with outliers. Moreover, it only ensures a locally optimal solution while being relatively sensitive to parameter initialisation.

2.3.2.3 Support vector regression

The third approach [27, 40] formulates the problem of outlier detection in deformation registration as robust learning of the correspondence manifold. Unlike the previous approaches which estimate a single function f, the correspondence manifold is represented by a *pair* of functions: *(1)* f that maps points from the template to the input image (i.e., $f : \mathbf{p} \mapsto \mathbf{q}$), and *(2)* f' that maps points from the input image to the template (i.e., $f' : \mathbf{q} \mapsto \mathbf{p}$) [27, 40]; see Figure 2.12 for an illustration.

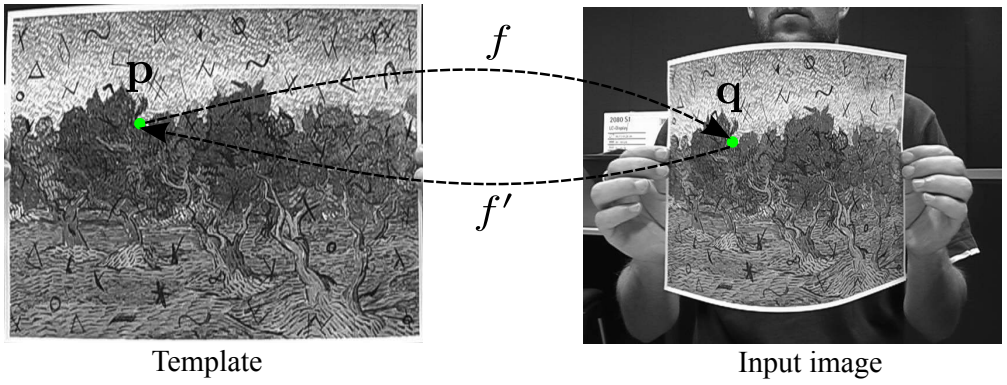


Figure 2.12: A pair of correspondence functions for representing the correspondence manifold.

Since both f and f' are computed in the same way, we will discuss estimating only f in the following. The function f is further separated into two components f_x and f_y with $f_x : \mathbf{p} \mapsto x'$ and $f_y : \mathbf{p} \mapsto y'$. Given a set of keypoint correspondences $\mathcal{X} = \{(\mathbf{u}_i, \mathbf{v}_i)\}_{i=1}^n$, the sets $\mathcal{X}_x = \{(x_i, y_i, x_i')\}_{i=1}^n$ and $\mathcal{X}_y = \{(x_i, y_i, y_i')\}_{i=1}^n$ extracted from $\mathcal{X}$ are considered respectively as the samples of f_x and f_y. Support vector regression is then employed separately for computing f_x from $\mathcal{X}_x$ and f_y from $\mathcal{X}_y$ (see [27, 40] for details).

To robustly learn the function f (i.e., two components f_x and f_y) in the presence of outliers, an iterative procedure is proposed. At the beginning, the sample sets $\mathcal{X}_x$ and $\mathcal{X}_y$ are used for initialising the components f_x and f_y. At each iteration, the data points in $\mathcal{X}$ with residuals greater than a threshold τ are selected into a suspect influential set (outlier set) $\mathcal{S}$, and their corresponding samples in $\mathcal{X}_x$ and $\mathcal{X}_y$ are eliminated. The remaining samples of $\mathcal{X}_x$ and $\mathcal{X}_y$ are then used to compute the new estimates of f_x and f_y. The resultant f_x and f_y are employed as the initial states for the next iteration. The iterative procedure stops when the difference in the estimated data deviations according to the current and new estimates of f_x and f_y (i.e., before and after removing the suspect influential set S) is smaller than a threshold ϵ_{INFL} or the computed mean squared error is less than a threshold ϵ_{MSE}.

After computing the functions f and f' by using the above iterative procedure, the data points that have residuals larger than a threshold ξ are rejected as outliers. In practice, most of the

outliers can be removed by using either f or f'. However, the method involves various parameters, i.e., τ, ϵ_{INFL}, ϵ_{MSE} for estimating the functions and ξ for performing outlier rejection, its success depends critically on optimal settings of the parameters.

2.3.2.4 Local smoothness test

A more recent approach [41, 42] to outlier rejection in deformation registration relies on the assumption that the correspondence manifold is locally smooth. Due to this assumption of local smoothness, the local topology around a correspondence should be preserved during the deformations between the images. The basic idea of [41, 42] is that a particular correspondence is likely to be an inlier if it is consistent with the local transformation, which is computed by using its neighbouring correspondences.

Let $\mathcal{C}_u = \{\mathbf{u}_i\}_{i=1}^n$ and $\mathcal{C}_v = \{\mathbf{v}_i\}_{i=1}^n$ be the sets of keypoints in the template and input image respectively, where point $\mathbf{u}_i \in \mathcal{C}_u$ is matched with point $\mathbf{v}_i \in \mathcal{C}_v$. In [41, 42], a Delaunay triangulation T is first defined on the set of keypoints $\mathcal{C}_u$ in the template, and used as the reference for connecting the correspondences. In particular, two correspondences $(\mathbf{u}_i, \mathbf{v}_i)$ and $(\mathbf{u}_j, \mathbf{v}_j)$ are neighbours if there exists an edge in the triangulation T that connects between the keypoints $\mathbf{u}_i$ and $\mathbf{u}_j$. For a specific correspondence $(\mathbf{u}_i, \mathbf{v}_i)$, the set of its neighbouring correspondences $Q(\mathbf{u}_i, \mathbf{v}_i)$ induced by T is used to estimate the local transformation f'_Q (see [41, 42] for details), which is then employed to measure the residual for $(\mathbf{u}_i, \mathbf{v}_i)$ as

$$d_i = \left\| \mathbf{u}_i - f'_Q(\mathbf{v}_i) \right\|, \tag{2.13}$$

where $f'_Q(\mathbf{v}_i)$ is the warped point of $\mathbf{v}_i$ computed by using the local transformation f'_Q; see Figure 2.13. The residual d_i represents how well the correspondence $(\mathbf{u}_i, \mathbf{v}_i)$ is consistent with the local transformation f'_Q. Due to the above assumption of local smoothness, a correspondence is considered as an inlier if it passes the local smoothness test, i.e., its residual is less than a threshold d_{TH}.

The local smoothness test can easily select strong inliers which are surrounded by inliers but it usually mislabels the inliers that are close to the outliers. The strong inliers are first obtained by applying the local smoothness test on the sets of keypoints $\mathcal{C}_u$ and $\mathcal{C}_v$, and an iterative procedure is then proposed to include the inliers that are nearby to the outliers. Let $\mathcal{G}_u$ and $\mathcal{G}_v$ be the sets of keypoints from the strong inliers, while $\mathcal{B}_u$ and $\mathcal{B}_v$ denote the sets of keypoints from the remaining correspondences (i.e., $\mathcal{C}_u = \mathcal{G}_u \cup \mathcal{B}_u$ and $\mathcal{C}_v = \mathcal{G}_v \cup \mathcal{B}_v$). For each correspondence $(\mathbf{u}_i, \mathbf{v}_i)$ from $\mathcal{B}_u$ and $\mathcal{B}_v$, a new Delaunay triangulation $\tilde{T}$ is first defined on the extended set of keypoints $\mathcal{G}_u \cup \{\mathbf{u}_i\}$ in the template, and the updated residual d_i is then computed based on the new triangulation $\tilde{T}$. The correspondence $(\mathbf{u}_i, \mathbf{v}_i)$ is considered as an inlier and added to $\mathcal{G}_u$

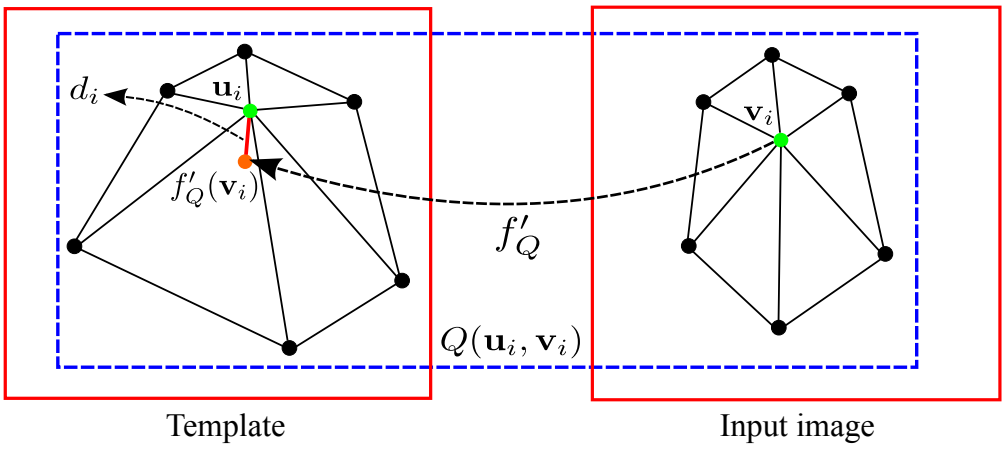


Figure 2.13: Local smoothness test by using a local transformation estimated on neighbouring correspondences.

and $\mathcal{G}_v$ if the updated residual d_i is smaller than d_{TH}. The iterative procedure is repeated until $\mathcal{G}_u$ and $\mathcal{G}_v$ stop growing (no more inliers can be found).

Similar to the methods surveyed in Section 2.2.1.2, the local smoothness test is effective when the inliers are locally dense. Compared to the above approaches which are often based on the assumption of the highly nonlinear correspondence manifold, this approach assumes that the correspondence manifold is locally smooth, which is equivalent to recognising that the nonrigid deformations have low degrees of variation. In this thesis, we further show that the previously discussed methods overestimate the difficulty of the data; and moreover, the correspondence manifold on which the inliers lie is approximately linear (Chapters 5 and 6).

2.4 Summary

We have reviewed in this chapter numerous methods for robust geometric fitting and robust deformable registration. In the following, we summarise the issues that we particularly address in the next chapters of this thesis.

As presented in Section 2.2, the problem of large span sampling has been observed in a large number of previous approaches to robust geometric fitting. However, little research has been invested in this issue. This fact has motivated us to investigate and develop a theoretical basis for this long-standing problem (Chapter 3), which shows how the data span affects different minimal subsets and provides a better understanding of previous work.

None of the previous guided sampling methods for robust geometric fitting actively searches for all-inlier minimal subsets with large span which are important to yielding a good model estimate

that fits the data well. To address this limitation, we present a novel sampling algorithm (Chapter 4) that consciously targets all-inlier samples as well as with large span for robust geometric fitting.

As discussed in Section 2.3, due to the assumption of the highly nonlinear correspondence manifold, most of the previous approaches to outlier rejection in deformable registration employ fully nonlinear models to represent the correspondence manifold, and rely on iterative optimisation procedures to remove outliers. This fact has inspired us to propose and investigate the interesting observation of the approximately linear correspondence manifold (Chapter 5), which facilitates the application of standard robust geometric fitting techniques for outlier rejection in deformable registration. Our proposed methods (Chapters 5 and 6), which are based entirely on fitting simple linear models, are significantly more efficient and at least as accurate as previous approaches.

Chapter 3

Minimal Subset Expansion

3.1 Introduction

In Chapter 2, we described a common method for robust geometric fitting in computer vision - RANSAC [18], which randomly generates and verifies a number of minimal subsets such that with a high probability at least one all-inlier minimal subset is obtained. This aim, for an all-inlier sample, stems from the assumptions that all-inlier minimal subsets have the same intrinsic quality and that any all-inlier sample yields a good model estimate that fits the data well. However, such assumptions are not satisfied when there exists noise in the data, which is invariably the case in real scenarios due to imperfect data acquisition and preprocessing. As illustrated in Figure 1.1, with noisy data, all-inlier minimal subsets are different in terms of their spatial coverage, and hence intrinsic quality. Therefore, drawing a single all-inlier sample is not sufficient to guarantee a good estimate of the model parameters; the inliers therein should also have a large spatial extent.

This chapter aims to show that a more principled reasoning exists for the above intuition. We first show that an explanation lies in a little known result relating to minimal subset expansion for least squares regression, which expresses the least squares regression estimate as a weighted linear combination of all possible minimal subset estimates. From this result, it is clear that the quality of a minimal subset estimate is proportional to the span of the associated data points. As the main contribution of this chapter, we derive an analogous minimal subset expansion for *total least squares* (TLS) [31], which unlike ordinary least squares, accounts for noise in both dependent and independent variables. Our result is of interest to the computer vision community, since TLS is directly related to geometric estimation techniques in computer vision, such as the direct linear transformation [32] that is frequently used in the estimation of multiple-view geometry. Furthermore, we investigate the effect of data span on various fundamental matrix

estimation algorithms, which shows a clear correlation between the span of data points used for estimation and the quality of model estimate.

The main content of this chapter is based on the author's publication in [65]. The rest of this chapter is organised as follows: Section 3.2 introduces the minimal subset expansion for least squares regression. We describe TLS in Section 3.3 and develop an equivalent minimal subset expansion for TLS. In Section 3.4, we explore the connection between our result with geometric fitting problems in computer vision, and analyse the influence of data span on the task of fundamental matrix estimation. We then draw conclusions in Section 3.5. The main content of this chapter is based on the authors publication in [65].

3.2 Minimal subset expansion for least squares

We present in this section a little known result relating to minimal subset expansion for least squares regression, which provides a theoretical explanation for the intuition of large span sampling in robust geometric fitting. The problem of linear regression involves deducing an unknown parameter vector $\boldsymbol{\beta} \in \mathbb{R}^m$ such that the distance between the multiplication of $\boldsymbol{\beta}$ by the design matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ and the observation vector $\mathbf{y} \in \mathbb{R}^n$ is minimised. Given an overdetermined system ($n > m$), the least squares approach assumes there is no noise in $\mathbf{X}$ and corrects for errors in $\mathbf{y}$ by solving

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{y} - \hat{\mathbf{y}}\|^2 \quad \text{s.t.} \quad \mathbf{X}\boldsymbol{\beta} = \hat{\mathbf{y}}, \tag{3.1}$$

where $\hat{\boldsymbol{\beta}}$ is the least squares estimate and $\hat{\mathbf{y}}$ is the corrected version of $\mathbf{y}$. The solution $\hat{\boldsymbol{\beta}}$ for (3.1) can be obtained in closed form by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{3.2}$$

The problem in (3.1) is often called *ordinary least squares* (OLS) to discern it from other types of estimation problems. A geometrical interpretation of (3.1) is that $\mathbf{X}\hat{\boldsymbol{\beta}}$ is the orthogonal projection of $\mathbf{y}$ onto the column span of $\mathbf{X}$.

It turns out that $\hat{\boldsymbol{\beta}}$ can be expanded as a linear combination of minimal subset estimates. To derive this result, first we apply Cramer's rule on (3.2) to write the j-th value of $\hat{\boldsymbol{\beta}}$ as

$$\hat{\boldsymbol{\beta}}_j = \frac{|(\mathbf{X}^T\mathbf{X})_j|}{|\mathbf{X}^T\mathbf{X}|} = \frac{|\mathbf{X}^T\mathbf{X}_j|}{|\mathbf{X}^T\mathbf{X}|}, \tag{3.3}$$

where $|\cdot|$ calculates the determinant of a matrix. We define $(\mathbf{X}^T\mathbf{X})_j$ as $\mathbf{X}^T\mathbf{X}$ with its j-th column replaced by $\mathbf{X}^T\mathbf{y}$, and $\mathbf{X}_j$ as $\mathbf{X}$ with its j-th column replaced by $\mathbf{y}$. Via the Binet-Cauchy formula [66, Page 29], we can expand (3.3) as

$$\hat{\boldsymbol{\beta}}_j = \frac{\sum_\lambda |\mathbf{X}(\lambda)||\mathbf{X}_j(\lambda)|}{\sum_\lambda |\mathbf{X}(\lambda)||\mathbf{X}(\lambda)|}, \tag{3.4}$$

where λ indicates a combination of m integers from the set $\{1, \ldots, n\}$, and $\mathbf{X}(\lambda)$ and $\mathbf{X}_j(\lambda)$ are square matrices formed by the m rows of $\mathbf{X}$ and $\mathbf{X}_j$ indexed by λ. The summations in (3.4) are taken over all $\binom{n}{m}$ possibilities of λ.

Picking the rows of $\mathbf{X}$ and $\mathbf{y}$ according to a λ amounts to choosing a minimal subset, since m cases are sufficient to uniquely determine $\boldsymbol{\beta}$. The minimal estimate from λ is

$$\hat{\boldsymbol{\beta}}^{(\lambda)} = \mathbf{X}(\lambda)^{-1}\mathbf{y}(\lambda), \tag{3.5}$$

where $\mathbf{y}(\lambda)$ is the vector formed by the m rows of $\mathbf{y}$ indexed by λ. Via Cramer's rule again, the j-th value of $\hat{\boldsymbol{\beta}}^{(\lambda)}$ is

$$\hat{\boldsymbol{\beta}}_j^{(\lambda)} = \frac{|\mathbf{X}_j(\lambda)|}{|\mathbf{X}(\lambda)|}. \tag{3.6}$$

By substituting $|\mathbf{X}_j(\lambda)| = |\mathbf{X}(\lambda)|\, \hat{\boldsymbol{\beta}}_j^{(\lambda)}$ in (3.4) we obtain

$$\hat{\boldsymbol{\beta}}_j = \frac{\sum_\lambda |\mathbf{X}(\lambda)||\mathbf{X}(\lambda)|\hat{\boldsymbol{\beta}}_j^{(\lambda)}}{\sum_\lambda |\mathbf{X}(\lambda)||\mathbf{X}(\lambda)|}, \tag{3.7}$$

or in vectorial form for the full parameter vector

$$\boxed{\hat{\boldsymbol{\beta}} = \sum_\lambda w_\lambda \hat{\boldsymbol{\beta}}^{(\lambda)}, \quad w_\lambda = \frac{|\mathbf{X}(\lambda)|^2}{\sum_\lambda |\mathbf{X}(\lambda)|^2}}, \tag{3.8}$$

where $0 \leq w_\lambda \leq 1$ and $\sum_\lambda w_\lambda = 1$. The quantity w_λ is the *weight* or *importance* of minimal subset λ towards estimating $\hat{\boldsymbol{\beta}}$. This little known result is due to Jacobi [67], and later rediscovered by others, e.g., [68].

Jacobi's result provides an algebraic justification to the intuition of maximising the span of minimal subsets. To illustrate this idea, consider the problem of 2D line fitting where we have

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \tag{3.9}$$

and $\boldsymbol{\beta}$ contains the usual parameters of intercept and slope. The weight of the estimate corresponding to a minimal subset $\lambda = \{k_1, k_2\}$ is proportional to

$$|\mathbf{X}(\lambda)|^2 = \left| \begin{bmatrix} 1 & x_{k_1} \\ 1 & x_{k_2} \end{bmatrix} \right|^2 = (x_{k_1} - x_{k_2})^2, \tag{3.10}$$

i.e., widely separated points provide better line estimates. More generally, $|\mathbf{X}(\lambda)|$ is the hypervolume of the parallelotope whose vertices are the rows of $\mathbf{X}(\lambda)$, a quantity that is directly related to the relative span of the data indexed by λ.

3.2.1 Generalising to non-minimal subsets

Can the weighted expansion in (3.8) be generalised to using subsets of size greater than m? Let ν index a subset of size $m + i \leq n$ ($i > 0$), and $\mathbf{X}(\nu)$ be the (non-square) submatrix of $\mathbf{X}$ containing the $m + i$ rows selected according to ν. The OLS estimate can actually also be expanded as

$$\boxed{\hat{\boldsymbol{\beta}} = \sum_{\nu} w_\nu \hat{\boldsymbol{\beta}}^{(\nu)}, \quad w_\nu = \frac{|\mathbf{X}(\nu)^T \mathbf{X}(\nu)|}{\sum_\nu |\mathbf{X}(\nu)^T \mathbf{X}(\nu)|}}, \tag{3.11}$$

where the summation is over all $\binom{n}{m+i}$ choices of ν; we refer the reader to [69] for the proof. Observe that (3.8) is a special case of (3.11). To gain a geometrical understanding of the weights, applying the Binet-Cauchy formula again yields

$$w_\nu \propto |\mathbf{X}(\nu)^T \mathbf{X}(\nu)| = \sum_{\lambda} |\mathbf{X}(\lambda|\nu)|^2, \tag{3.12}$$

where λ indexes over all $\binom{m+i}{m}$ *minimal* subsets of $\mathbf{X}(\nu)$, and we define $\mathbf{X}(\lambda|\nu)$ as the submatrix of $\mathbf{X}(\nu)$ indexed by λ. In other words, the weight of $\mathbf{X}(\nu)$ is proportional to the sum of the weights of all the minimal subsets from $\mathbf{X}(\nu)$.

3.3 The case of total least squares

In this section we describe TLS and develop the main contribution of this chapter — a minimal subset expansion for TLS. In contrast to OLS presented in the previous section, TLS (also called *errors-in-variables* modelling [17] or *orthogonal regression* [70]) corrects for errors in both the independent and dependent measurements $(\mathbf{X}, \mathbf{y})$. The TLS estimate is

$$\breve{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \min_{\breve{\mathbf{X}}} \left\| [\mathbf{X}\ \mathbf{y}] - [\breve{\mathbf{X}}\ \breve{\mathbf{y}}] \right\|_F^2 \quad \text{s.t.} \quad \breve{\mathbf{X}}\boldsymbol{\beta} = \breve{\mathbf{y}}, \tag{3.13}$$

where we use the breve accent ($\breve{}$) to distinguish TLS results from those of OLS. The solution $\breve{\boldsymbol{\beta}}$ for (3.13) can be reasoned as follows [70]: Assume $[\mathbf{X}\ \mathbf{y}] \in \mathbb{R}^{n\times(m+1)}$ to be full rank, i.e., rank $m+1$ since $n > m$. To make the following system

$$[\mathbf{X}\ \mathbf{y}] \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} \approx \mathbf{0} \tag{3.14}$$

compatible, we must reduce the rank of $[\mathbf{X}\ \mathbf{y}]$ by one. Let

$$[\mathbf{X}\ \mathbf{y}] = \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^T \tag{3.15}$$

be the singular value decomposition (SVD) of $[\mathbf{X}\ \mathbf{y}]$. From the Eckart-Young Theorem the closest rank-m matrix to $[\mathbf{X}\ \mathbf{y}]$ in the Frobenius sense is

$$[\breve{\mathbf{X}}\ \breve{\mathbf{y}}] = \mathbf{A}\breve{\boldsymbol{\Sigma}}\mathbf{B}^T, \tag{3.16}$$

where $\breve{\boldsymbol{\Sigma}}$ is obtained by setting the $(m+1)$-th singular value σ_{m+1} in $\boldsymbol{\Sigma}$ to zero. Let $\mathbf{b}_{m+1}$ be the $(m+1)$-th right singular vector of $[\mathbf{X}\ \mathbf{y}]$. Then the optimal approximation to (3.14) is

$$[\breve{\mathbf{X}}\ \breve{\mathbf{y}}]\mathbf{b}_{m+1} = \mathbf{0}, \tag{3.17}$$

which holds for arbitrary scalings of $\mathbf{b}_{m+1}$. The TLS estimate $\breve{\boldsymbol{\beta}}$ is obtained by scaling $\mathbf{b}_{m+1}$ as

$$\begin{bmatrix} \breve{\boldsymbol{\beta}} \\ -1 \end{bmatrix} = -\frac{1}{\mathbf{b}_{m+1}(m+1)}\mathbf{b}_{m+1} \tag{3.18}$$

such that the $(m+1)$-th element of the vector is -1. Here, $\mathbf{b}_{m+1}(m+1)$ denotes the $(m+1)$-th element of $\mathbf{b}_{m+1}$.

Clearly, for $\breve{\boldsymbol{\beta}}$ to exist the element $\mathbf{b}_{m+1}(m+1)$ can not be zero. A sufficient condition for a nonzero $\mathbf{b}_{m+1}(m+1)$ is that the m-th singular value s_m of $\mathbf{X}$ to be strictly greater than σ_{m+1}:

$$s_m > \sigma_{m+1} \Longrightarrow \mathbf{b}_{m+1}(m+1) \neq 0 \text{ and } \sigma_m > \sigma_{m+1}. \tag{3.19}$$

See [70, Chapter 2] for the proof. As argued in [71], this condition is not restrictive and is usually satisfied. Moreover, as we will explain in Section 3.4, convenient pre- and postprocessing are available to prevent the condition from becoming debilitating for geometric fitting problems in computer vision.

3.3.1 Minimal subset expansion for TLS

Following the above derivations, the eigenvector identity

$$[\mathbf{X}\ \mathbf{y}]^T[\mathbf{X}\ \mathbf{y}]\mathbf{b}_{m+1} = \sigma_{m+1}^2\mathbf{b}_{m+1} \tag{3.20}$$

holds for $\mathbf{b}_{m+1}$ and arbitrary scalings of $\mathbf{b}_{m+1}$. We can thus rewrite and expand the eigenvector identity as

$$\begin{bmatrix}\mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{y} \\ \mathbf{y}^T\mathbf{X} & \mathbf{y}^T\mathbf{y}\end{bmatrix}\begin{bmatrix}\breve{\boldsymbol{\beta}} \\ -1\end{bmatrix} = \sigma_{m+1}^2\begin{bmatrix}\breve{\boldsymbol{\beta}} \\ -1\end{bmatrix}. \tag{3.21}$$

Multiplying through the top part and rearranging yields

$$\breve{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X} - \sigma_{m+1}^2\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}. \tag{3.22}$$

See also [70, Chapter 2] for the derivation of (3.22). To develop our TLS minimal subset expansion we define

$$\mathbf{Z} := \mathbf{X} - \sigma_{m+1}^2(\mathbf{X}^T)^\dagger, \tag{3.23}$$

where $(\mathbf{X}^T)^\dagger$ is the Moore-Penrose generalised inverse of $\mathbf{X}^T$, and rewrite (3.22) as

$$\breve{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{Z})^{-1}\mathbf{X}^T\mathbf{y}. \tag{3.24}$$

We first prove the following intermediate result which appears to be novel and is of interest in its own right.

Proposition 3.1. *The solution to the following OLS problem*

$$\arg\min_{\boldsymbol{\beta}} \|\mathbf{y} - \hat{\mathbf{y}}\|^2 \quad \text{s.t.} \quad \mathbf{Z}\boldsymbol{\beta} = \hat{\mathbf{y}} \tag{3.25}$$

coincides with the TLS estimate $\breve{\boldsymbol{\beta}}$.

Proof. If $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ is the SVD of $\mathbf{X} \in \mathbb{R}^{n\times m}$, then

$$(\mathbf{X}^T)^\dagger = \mathbf{U}\mathbf{S}^{-1}\mathbf{V}^T \tag{3.26}$$

is the SVD of $(\mathbf{X}^T)^\dagger$, where we define $\mathbf{S}^{-1}$ as taking the reciprocal of the diagonal elements of $\mathbf{S}$, while leaving the other elements unchanged. From (3.23) we can rewrite $\mathbf{Z}$ as

$$\mathbf{Z} = \mathbf{U}(\mathbf{S} - \sigma_{m+1}^2\mathbf{S}^{-1})\mathbf{V}^T := \mathbf{U}\tilde{\mathbf{S}}\mathbf{V}^T, \tag{3.27}$$

where $\mathbf{U}\tilde{\mathbf{S}}\mathbf{V}^T$ is a valid SVD of $\mathbf{Z}$ since the diagonal values in $\tilde{\mathbf{S}}$ are still in descending order; recall that we impose condition (3.19) such that σ_{m+1} is strictly smaller than the smallest singular value s_m of $\mathbf{X}$. Therefore, since $\mathbf{X}$, $(\mathbf{X}^T)^\dagger$ and $\mathbf{Z}$ share the same left singular vectors $\mathbf{U}$, we have

$$\mathcal{R}(\mathbf{X}) = \mathcal{R}((\mathbf{X}^T)^\dagger) = \mathcal{R}(\mathbf{Z}), \tag{3.28}$$

i.e., the column spans of the three matrices are equal.

If $\tilde{\boldsymbol{\beta}}$ is the solution of (3.25), then by invoking the closed-form expression for the OLS estimate

$$\tilde{\boldsymbol{\beta}} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y}, \tag{3.29}$$

which can be further manipulated to yield

$$\mathbf{X}^T\mathbf{Z}\tilde{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{y} + \sigma_{m+1}^2(\mathbf{X}^T)^{\dagger T}(\mathbf{Z}\tilde{\boldsymbol{\beta}} - \mathbf{y}). \tag{3.30}$$

Since $\tilde{\boldsymbol{\beta}}$ is the solution to the OLS problem (3.25), $\mathbf{Z}\tilde{\boldsymbol{\beta}}$ is the projection of $\mathbf{y}$ onto $\mathcal{R}(\mathbf{Z})$, and consequently $\mathbf{Z}\tilde{\boldsymbol{\beta}} - \mathbf{y}$ is orthogonal to $\mathcal{R}(\mathbf{Z})$. Hence, the equality (3.30) reduces to

$$\mathbf{X}^T\mathbf{Z}\tilde{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{y} \tag{3.31}$$

since $\mathbf{Z}\tilde{\boldsymbol{\beta}} - \mathbf{y}$ will also be orthogonal to $\mathcal{R}((\mathbf{X}^T)^\dagger)$. Comparing (3.24) with (3.31) yields the result $\breve{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}$. □

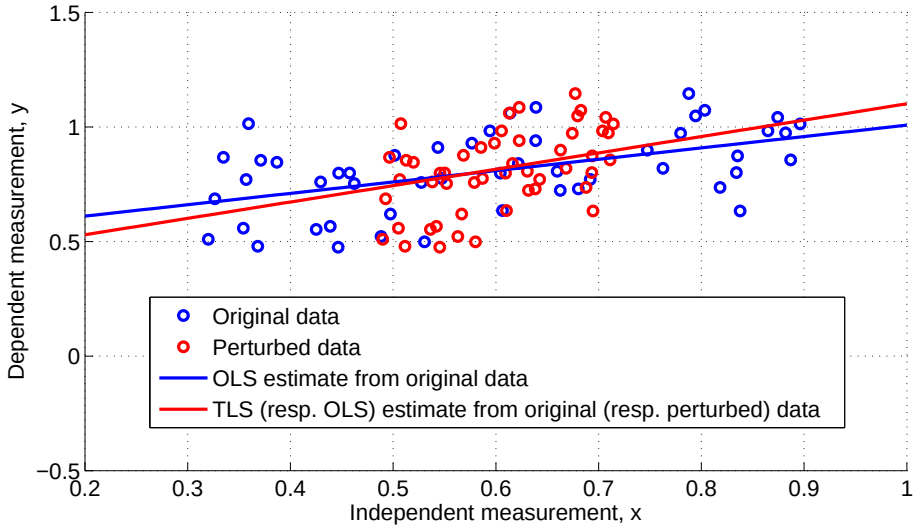
Proposition 3.1 states that, given the measurements $[\mathbf{X}\ \mathbf{y}]$, the TLS estimate can be calculated as

$$\breve{\boldsymbol{\beta}} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y}, \tag{3.32}$$

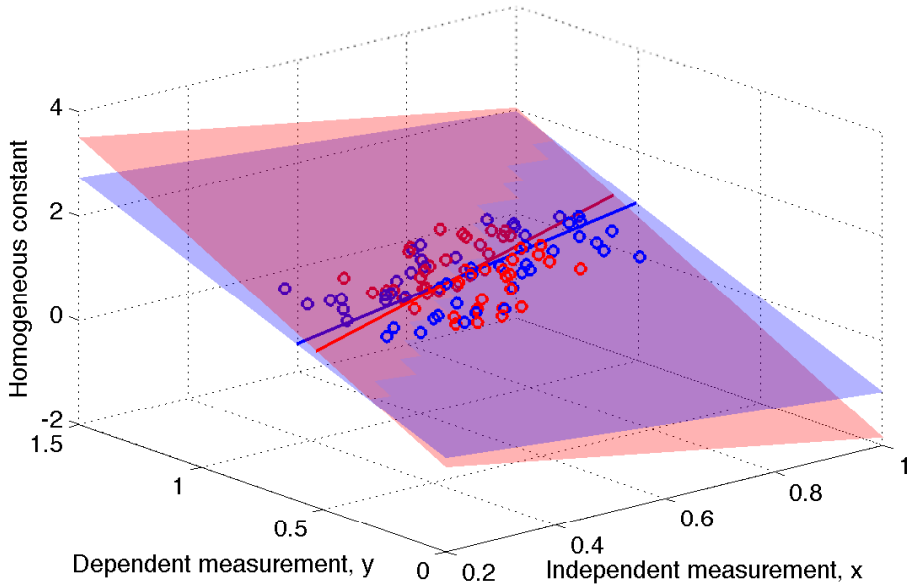
where $\mathbf{Z}$ is as defined in (3.23). In other words, we perturb $\mathbf{X}$ by the amount $-\sigma_{m+1}^2(\mathbf{X}^T)^\dagger$ to become $\mathbf{Z}$, while leaving $\mathbf{y}$ unchanged, such that TLS on $(\mathbf{X}, \mathbf{y})$ can be solved as OLS on $(\mathbf{Z}, \mathbf{y})$. Figure 3.1 illustrates this idea on 2D line fitting.

To exploit Proposition 3.1, we define

$$\breve{\boldsymbol{\beta}}^{(\lambda)} := \mathbf{Z}(\lambda)^{-1}\mathbf{y}(\lambda) \tag{3.33}$$



(a) In 2D line fitting, the measurements $\{x_i, y_i\}_{i=1}^n$ are collected into matrix $\mathbf{X} \in \mathbb{R}^{n\times 2}$ and vector $\mathbf{y} \in \mathbb{R}^n$ as defined in (3.9), where the first column of $\mathbf{X}$ contains the homogeneous constants 1. Following (3.23), $\mathbf{X}$ is perturbed to become $\mathbf{Z}$. From Proposition 3.1 the TLS estimate on $(\mathbf{X}, \mathbf{y})$ is the same as the OLS estimate on $(\mathbf{Z}, \mathbf{y})$. The data after perturbation are plotted in red — note that the homogeneous constants in $\mathbf{Z}$ would also have been perturbed away from 1, which mirrors the fact that TLS (3.13) will correct all columns in $\mathbf{X}$. To plot $(\mathbf{Z}, \mathbf{y})$ here, we "dehomogenise" each datum by dividing the measurements with the corresponding homogeneous constant. See also panel (b) below.



(b) Line fitting on 2D data $\{x_i, y_i\}_{i=1}^n$ is really accomplished as 2D subspace fitting on 3D data $(\mathbf{X}, \mathbf{y})$, where points in the subspace satisfies $[c\ x]\boldsymbol{\beta} - y = 0$, and c is an auxiliary variable corresponding to the homogeneous constant. Proposition 3.1 states that TLS on $(\mathbf{X}, \mathbf{y})$ is the same as OLS on $(\mathbf{Z}, \mathbf{y})$, where $\mathbf{Z}$ is perturbed from $\mathbf{X}$ according to (3.23). Here, we plot $(\mathbf{X}, \mathbf{y})$ and $(\mathbf{Z}, \mathbf{y})$ in $\mathbb{R}^3$, where the homogeneous constants in $\mathbf{X}$ and $\mathbf{Z}$ are plotted in the vertical axis. The 2D subspaces fitted on $(\mathbf{X}, \mathbf{y})$ and $(\mathbf{Z}, \mathbf{y})$ by OLS are also plotted. The lines fitted on $\{x_i, y_i\}_{i=1}^n$ are 1D *affine* subspaces within the 2D subspaces. Panel (a) is the projection of (b) onto the plane $c = 1$.

Figure 3.1: Illustration of Proposition 3.1 on the problem of 2D line fitting.

as the minimal subset estimate for $\breve{\boldsymbol{\beta}}$ based on the m rows of $\mathbf{Z}$ and $\mathbf{y}$ as indexed by λ. Applying the steps taken in Section 3.2 on (3.32), we obtain a minimal subset expansion for $\breve{\boldsymbol{\beta}}$

$$\boxed{\breve{\boldsymbol{\beta}} = \sum_{\lambda} w_\lambda \breve{\boldsymbol{\beta}}^{(\lambda)}, \quad w_\lambda = \frac{|\mathbf{Z}(\lambda)|^2}{\sum_\lambda |\mathbf{Z}(\lambda)|^2}}, \tag{3.34}$$

where again w_λ is the weight of minimal subset λ, with $0 \leq w_\lambda \leq 1$ and $\sum_\lambda w_\lambda = 1$. Extending (3.34) to using non-minimal subsets also follows easily from (3.11).

3.3.2 Comparing weights of minimal subsets under TLS

In many cases (e.g., in robust estimation) we are mainly interested in comparing the goodness (from the aspect of span) of two minimal subsets and not the actual weight values. This fact motivates the following result on weight comparison.

Proposition 3.2. *Given two minimal subsets λ_1 and λ_2,*

$$|\mathbf{X}(\lambda_1)|^2 > |\mathbf{X}(\lambda_2)|^2 \implies |\mathbf{Z}(\lambda_1)|^2 > |\mathbf{Z}(\lambda_2)|^2. \tag{3.35}$$

Proof. Recall the SVD of $\mathbf{X}$ and $\mathbf{Z}$ in Proposition 3.1. Since both $\mathbf{X}$ and $\mathbf{Z}$ are of size $n \times m$ with $n > m$, they can be expanded using the first-m left singular vectors $\mathbf{U}_m$ of $\mathbf{X}$ as

$$\mathbf{X} = \mathbf{U}_m \mathbf{S}_m \mathbf{V}^T, \quad \mathbf{Z} = \mathbf{U}_m \tilde{\mathbf{S}}_m \mathbf{V}^T, \tag{3.36}$$

where $\mathbf{S}_m$ is the first $m \times m$ submatrix of $\mathbf{S}$ (similarly for $\tilde{\mathbf{S}}$). The determinants $|\mathbf{X}(\lambda)|$ and $|\mathbf{Z}(\lambda)|$ can be obtained as

$$|\mathbf{X}(\lambda)| = |\mathbf{U}_m(\lambda)\|\mathbf{S}_m \mathbf{V}^T|, \quad |\mathbf{Z}(\lambda)| = |\mathbf{U}_m(\lambda)\|\tilde{\mathbf{S}}_m \mathbf{V}^T|,$$

where we define $\mathbf{U}_m(\lambda)$ as the m rows of $\mathbf{U}_m$ selected according to λ. It is clear that

$$|\mathbf{X}(\lambda)| = \kappa |\mathbf{Z}(\lambda)|, \quad \kappa := |\mathbf{S}_m \mathbf{V}^T| / |\tilde{\mathbf{S}}_m \mathbf{V}^T|, \tag{3.37}$$

where κ is a constant independent of λ. Therefore, given two minimal subsets λ_1 and λ_2, if $|\mathbf{X}(\lambda_1)|^2 > |\mathbf{X}(\lambda_2)|^2$ then

$$|\mathbf{X}(\lambda_1)|^2 / |\mathbf{X}(\lambda_2)|^2 = |\mathbf{Z}(\lambda_1)|^2 / |\mathbf{Z}(\lambda_2)|^2 > 1, \tag{3.38}$$

or $|\mathbf{Z}(\lambda_1)|^2 > |\mathbf{Z}(\lambda_2)|^2$. □

Proposition 3.2 states that the *relative* weights of two minimal subsets λ_1 and λ_2 are equivalent under TLS and OLS; it is thus sufficient to compute $|\mathbf{X}(\lambda)|^2$ and there is no need to obtain

$|\mathbf{Z}(\lambda)|^2$ (which requires the singular value σ_{m+1}). In fact, Proposition 3.2 implies that the weight w_λ of a minimal subset λ is *equal* under (3.8) and (3.34), although the associated minimal estimates $\hat{\boldsymbol{\beta}}^{(\lambda)}$ and $\breve{\boldsymbol{\beta}}^{(\lambda)}$ may be different.

Extending Proposition 3.2 to compare the weights of non-minimal subsets is straightforward; we refer the reader to Appendix A.

3.3.3 TLS with frozen columns

In some applications it is useful to constrain the TLS correction to occur only on some of the columns of $\mathbf{X}$ while leaving the other *known* columns unchanged or "frozen" [72] (in Section 3.4 we explain why this case is relevant for geometric fitting problems in computer vision). The problem is also known as *mixed OLS-TLS* [70]. Here we show that our result extends easily to this case.

Let $\mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2]$ be rearranged such that $\mathbf{X}_1 \in \mathbb{R}^{n\times m_1}$ are the frozen columns while $\mathbf{X}_2 \in \mathbb{R}^{n\times m_2}$ are to be corrected, and $m = m_1 + m_2$. The task is to estimate

$$\breve{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta},\breve{\mathbf{X}}_2} \left\| [\mathbf{X}_2\ \mathbf{y}] - [\breve{\mathbf{X}}_2\ \breve{\mathbf{y}}] \right\|_F^2 \quad \text{s.t.} \quad [\mathbf{X}_1\ \breve{\mathbf{X}}_2]\boldsymbol{\beta} = \breve{\mathbf{y}}.$$

We first perform the QR factorisation

$$[\mathbf{X}_1\ \mathbf{X}_2\ \mathbf{y}] = \mathbf{Q}\mathbf{R} \quad \text{with} \quad \mathbf{R} = \begin{array}{c} \\ m_1 \\ n-m_1 \end{array} \begin{array}{c} \begin{array}{ccc} m_1 & m_2 & 1 \end{array} \\ \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \mathbf{r}_1 \\ \mathbf{0} & \mathbf{R}_{22} & \mathbf{r}_2 \end{bmatrix} \end{array},$$

where $\mathbf{R}_{11}$ is an upper triangular matrix. Basic TLS is invoked on $\mathbf{R}_{22}$ and $\mathbf{r}_2$ to solve for the last-m_2 elements of $\breve{\boldsymbol{\beta}}$. These are then substituted back into the system to allow the first-m_1 variables to be obtained using OLS.

Let σ_{m_2+1} be the smallest singular value of $[\mathbf{R}_{22}\ \mathbf{r}_2]$. The mixed OLS-TLS estimate can be expressed as [70]

$$\breve{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X} - \sigma_{m_2+1}^2\mathbf{L})^{-1}\mathbf{X}^T\mathbf{y}, \tag{3.39}$$

where $\mathbf{L}$ is a "selector matrix" defined as

$$\mathbf{L} := \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m_2} \end{bmatrix} \in \mathbb{R}^{m\times m}, \tag{3.40}$$

with $\mathbf{I}_{m_2}$ the $m_2 \times m_2$ identity matrix. In effect, $\mathbf{L}$ chooses which columns of $\mathbf{X}$ get corrected. Defining $\mathbf{Z}$ now as

$$\mathbf{Z} := \mathbf{X} - \sigma_{m_2+1}^2 (\mathbf{X}^T)^\dagger \mathbf{L}, \tag{3.41}$$

we may also re-express (3.39) like (3.24) as

$$\breve{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{Z})^{-1} \mathbf{X}^T \mathbf{y}. \tag{3.42}$$

Observe now that in (3.41) $\mathbf{L}$ chooses the columns of $\mathbf{X}$ that are perturbed. It turns out that Propositions 3.1 and 3.2 also hold for mixed OLS-TLS (see Appendix A for the proof), i.e., $\breve{\boldsymbol{\beta}}$ can be obtained as the solution of the OLS in (3.25), which motivates the expansion (3.34) for mixed OLS-TLS with minimal or non-minimal subsets.

3.3.4 Orthogonal distance fitting

We explore the consequence of our result on orthogonal distance fitting of lines onto 2D data. Given data $\{x_i, y_i\}_{i=1}^n$, we wish to estimate the $\boldsymbol{\beta} = [\boldsymbol{\beta}_1 \ \boldsymbol{\beta}_2]^T \in \mathbb{R}^2$ that minimises

$$\sum_{i=1}^{n} \frac{(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 x_i - y_i)^2}{\boldsymbol{\beta}_2^2 + 1}, \tag{3.43}$$

i.e., the sum of squared *orthogonal distances* to the line $\boldsymbol{\beta}$. Creating matrix $\mathbf{X}$ and vector $\mathbf{y}$ as in (3.9), the problem is equivalent to minimising the generalised Rayleigh quotient

$$\underset{\theta}{\operatorname{argmin}} \ \frac{\theta^T [\mathbf{X} \ \mathbf{y}]^T [\mathbf{X} \ \mathbf{y}] \theta}{\theta^T \begin{bmatrix} \mathbf{L} & 0 \\ 0 & 1 \end{bmatrix} \theta}, \tag{3.44}$$

where $\theta = [\boldsymbol{\beta}^T \ -1]^T$ and $\mathbf{L}$ is as defined in (3.40) with $m = 2$ and $m_2 = 1$. The solution $\breve{\theta} = [\breve{\boldsymbol{\beta}}^T \ -1]^T$ satisfies the generalised eigenvector equation

$$[\mathbf{X} \ \mathbf{y}]^T [\mathbf{X} \ \mathbf{y}] \begin{bmatrix} \breve{\boldsymbol{\beta}} \\ -1 \end{bmatrix} = \eta \begin{bmatrix} \mathbf{L} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \breve{\boldsymbol{\beta}} \\ -1 \end{bmatrix}, \tag{3.45}$$

with η the smallest generalised eigenvalue. Multiplying through the top part and rearranging yields

$$\breve{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X} - \eta \mathbf{L})^{-1} \mathbf{X}^T \mathbf{y}, \tag{3.46}$$

which resembles (3.39). Identifying the first column in $\mathbf{X}$ as the frozen column (the homogeneous constants in $\mathbf{X}$ are prevented from being corrected), it turns out that orthogonal distance fitting is an instance of mixed OLS-TLS.

Figure 3.2 illustrates the implications of Proposition 3.1 on orthogonal distance line fitting, using the same data as in Figure 3.1. Here $\mathbf{Z}$ is perturbed from $\mathbf{X}$ following (3.41) which preserves the values of the homogeneous constants.

It is worthwhile to clarify why the literature also calls standard TLS defined in (3.13) "orthogonal regression" [70]. If there are no homogeneous constants in $\mathbf{X}$ (or equivalently no intercept parameter in $\boldsymbol{\beta}$), (3.13) estimates the *linear* subspace which minimises the sum of squared orthogonal distances to the points. When $\mathbf{X}$ contains homogeneous constants (or $\boldsymbol{\beta}$ involves an intercept), the solution to (3.13) no longer minimises the sum of squared orthogonal distances. See [73] for details.

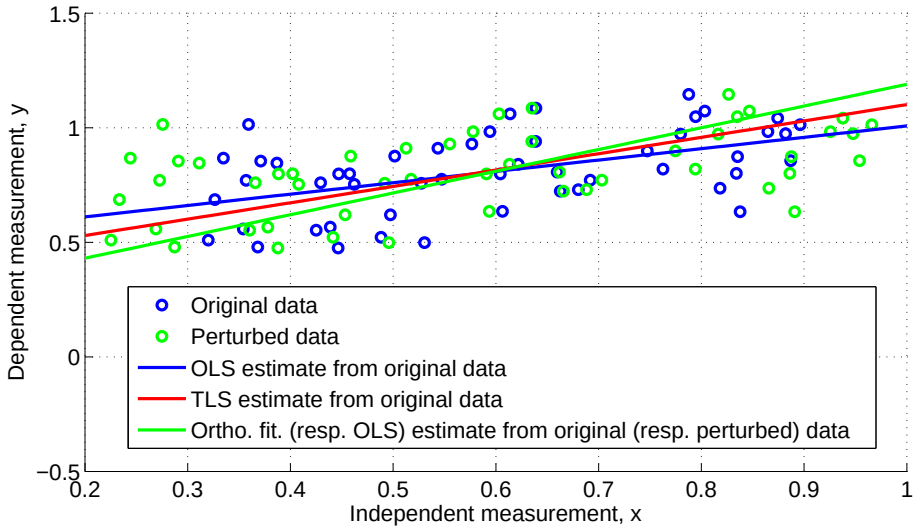
3.4 TLS in geometric fitting problems

In this section we establish the relevance of our result in the previous section to geometric fitting problems in computer vision. In particular, we relate TLS to the direct linear transformation (DLT) [45] technique, which is usually employed in the estimation of multiple-view geometry [3]. The intimate connection between TLS and DLT has been established elsewhere, e.g., [32, 74]. In the following we concentrate on the task of fundamental matrix estimation, which is a common geometric fitting problem in computer vision, and discuss the equivalence between TLS and DLT in that context.

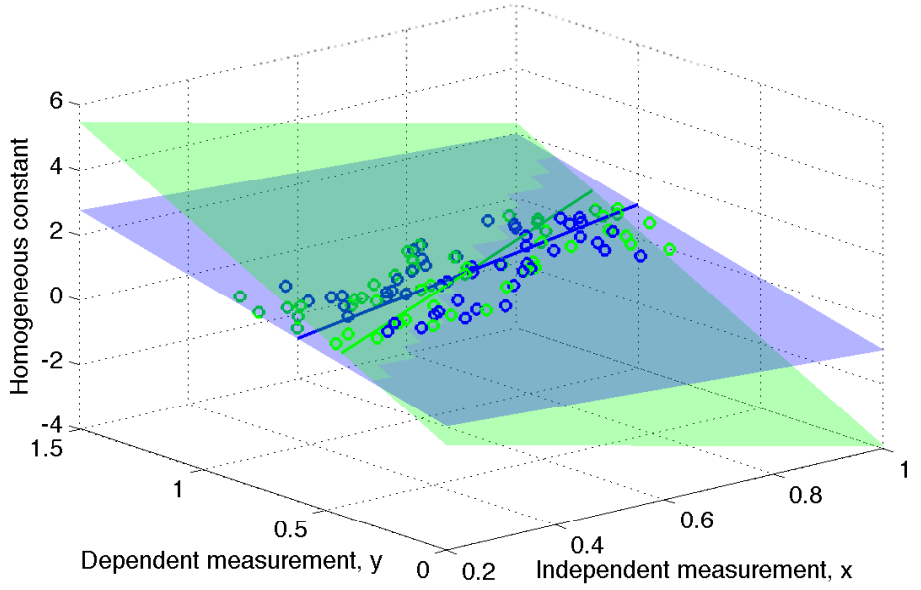
3.4.1 The equivalence between TLS and DLT on fundamental matrix estimation

We first describe the epipolar geometry and the fundamental matrix. Suppose that we have two images of a static scene taken from two different views. We may ask "How do the coordinates of the points in one image relate to the coordinates of the points in another?". The answer to this question paves the way for developing many important applications in computer vision, since it turns out that once the correspondences between two views are established, the three-dimensional structure of the scene can be recovered from the images.

The *epipolar geometry* studies the intrinsic projective geometry between two views and provides the answer to the above question. Suppose that a point $\mathbf{X}$ in three-dimensional space is projected through the centre of projection $\mathbf{C}_1$ (of the first camera) to the point $\mathbf{u}$ in the first image, and through the centre of projection $\mathbf{C}_2$ (of the second camera) to the point $\mathbf{v}$ in the second image; see Figure 3.3 for an illustration. The plane determined by $\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{X}$ is called an *epipolar*



(a) Comparing estimates from OLS, TLS and orthogonal line fitting. Here, $\mathbf{Z}$ is perturbed from $\mathbf{X}$ following (3.41) which preserves the homogeneous constants at 1. Therefore, dehomogenisation of $(\mathbf{Z}, \mathbf{y})$ is not required to plot the perturbed data here, unlike in Figure 3.1(a).



(b) Here $(\mathbf{X}, \mathbf{y})$ and $(\mathbf{Z}, \mathbf{y})$ are plotted in $\mathbb{R}^3$, where the homogeneous constants in $\mathbf{X}$ and $\mathbf{Z}$ are plotted in the vertical axis. The 2D subspaces fitted on $(\mathbf{X}, \mathbf{y})$ and $(\mathbf{Z}, \mathbf{y})$ by OLS are also plotted. Unlike Figure 3.1(b), where under standard TLS the homogeneous constants in $\mathbf{X}$ are not preserved in $\mathbf{Z}$, here the homogeneous constants remain at 1.

Figure 3.2: Illustration of Proposition 3.1 on the problem of orthogonal distance line fitting in 2D.

plane. There are numerous epipolar planes according to different three-dimensional points $\mathbf{X}$ but all epipolar planes will pass through the baseline which connects $\mathbf{C}_1$ and $\mathbf{C}_2$. The point of intersection between the baseline and the image plane is called an *epipole*.

Assume that $\mathbf{C}_1$, $\mathbf{C}_2$ and the point $\mathbf{u}$ in the first image are known, we now search for the corresponding point $\mathbf{v}$ in the second image to answer the above question. Recall that three-dimensional points are projected by straight lines through the centre of projection (of the camera) to points in the image, which means that the epipolar plane formed by $\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{X}$ is co-planar with the plane formed by $\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{u}$, and hence the epipolar plane on which $\mathbf{X}$ and $\mathbf{v}$ must lie can be determined. Furthermore, the search for $\mathbf{v}$ can be restricted to the line of intersection between that epipolar plane with the image plane of the second image, which is called an *epipolar line* (i.e., $\mathbf{l}_2$ in Figure 3.3).

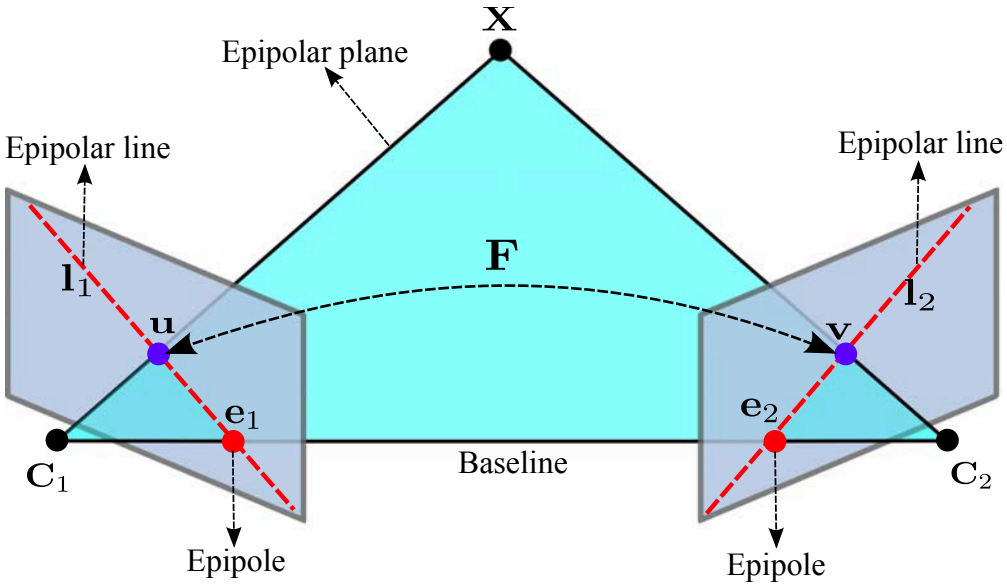


Figure 3.3: The epipolar geometry and the fundamental matrix.

The *fundamental matrix* is the algebraic representation of the above constraint in epipolar geometry. In particular, let $\mathbf{u} = [p\ q]^T$ and $\mathbf{v} = [p'\ q']^T$ be a pair of matching points between two views of a static scene, the epipolar geometric constraint between $\mathbf{u}$ and $\mathbf{v}$ is expressed as

$$[p\ q\ 1]\mathbf{F}[p'\ q'\ 1]^T = 0, \tag{3.47}$$

where $\mathbf{F}$ is a 3×3 matrix of rank 2, called the fundamental matrix. The constraint is linearised by multiplying through to yield

$$\begin{bmatrix} p'p & p'q & p' & q'p & q'q & q' & p & q & 1 \end{bmatrix} \mathbf{f} = 0, \tag{3.48}$$

where $\mathbf{f}$ is a column vector containing the nine elements of $\mathbf{F}$. Given a set of noisy keypoint correspondences $\mathcal{X} = \{(\mathbf{u}_i, \mathbf{v}_i)\}_{i=1}^n$ across two images of a fixed scene, where point $\mathbf{u}_i =$

$[p_i\ q_i]^T$ in one image is matched with point $\mathbf{v}_i = [p'_i\ q'_i]^T$ in another, the goal is to estimate the $\mathbf{F}$ of $\mathbf{f}$ corresponding to the scene.

The algorithm DLT [45] computes (3.48) for each keypoint correspondence $(\mathbf{u}_i, \mathbf{v}_i)$, and stacks the rows to yield a matrix $\mathbf{D} \in \mathbb{R}^{n\times 9}$. The solution is obtained by minimising the sum of squared algebraic errors as

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \|\mathbf{D}\mathbf{f}\|^2 \quad \text{s.t.} \quad \|\mathbf{f}\| = 1, \tag{3.49}$$

where the quadratic constraint avoids the trivial solution $\mathbf{f} = \mathbf{0}$. It can be shown that $\mathbf{f}^*$ is the least significant right singular vector of $\mathbf{D}$ [45]. Equating $\mathbf{D}$ with $[\mathbf{X}\ \mathbf{y}]$ under TLS, it is clear that $\mathbf{f}^* = \mathbf{b}_{m+1}$ in (3.17), and apart from the rescaling step (3.18) the TLS and DLT estimates are equal. Note that since $\mathbf{f}$ is homogeneous it has only eight degrees of freedom, mirroring the fact that the TLS parameter vector $\boldsymbol{\beta}$ has eight elements in the task of fundamental matrix estimation. Therefore, a minimal subset of eight rows of $\mathbf{D}$ or $[\mathbf{X}\ \mathbf{y}]$ corresponding to eight keypoint matches are sufficient to instantiate $\mathbf{f}$ or $\boldsymbol{\beta}$[1].

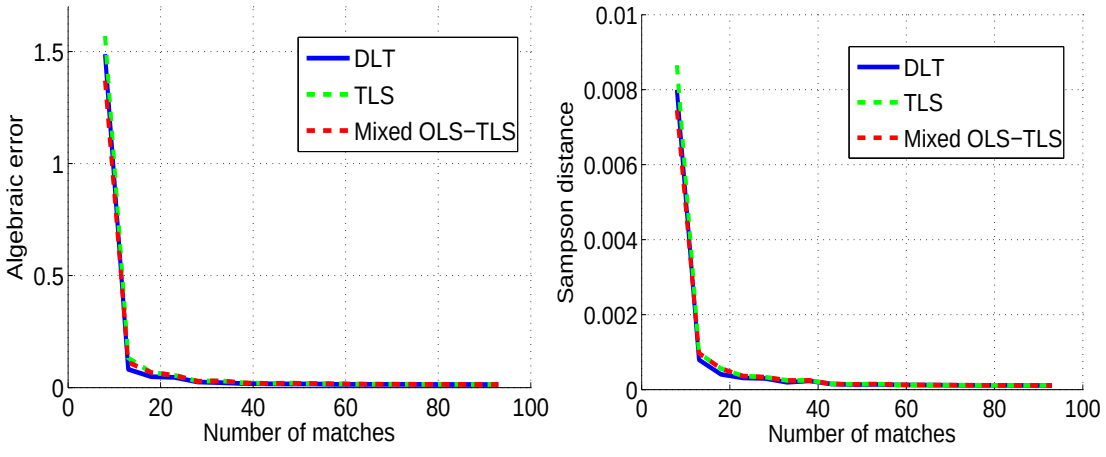
Our derivations in Section 3.3.1, however, require that the constraint (3.47) is dehomogenised. This amounts to fixing an element in $\mathbf{f}$ to -1, and moving the corresponding column in $\mathbf{D}$ to the RHS of (3.47) to become the observation vector $\mathbf{y}$. Such a modification is prevalent in the literature, e.g., [76–79], and is usually applied to bring to bear the framework of regression onto homogeneous estimation. If the element fixed to -1 in $\mathbf{f}$ has the true value 0, the result may be numerically unstable. A frequent solution (e.g., see [76]) is to detect when this case occurs, and then change another element in $\mathbf{f}$ to fix at -1 (see also the handling of such *non-generic collinearities* under TLS [70]). In some cases preprocessing of the data is available to ensure that an element in the homogeneous vector is nonzero [77, 79]. From (3.48), it is also clear that one column in $\mathbf{D}$ consists of the homogeneous constant 1. This column should be frozen under TLS, and not be moved to the RHS to yield the observation vector $\mathbf{y}$; see [32, 80]. Our result in Section 3.3.3 for mixed OLS-TLS is relevant to this notion. In the following experiments, we dehomogenise by fixing the first element in $\mathbf{f}$ to -1.

It is important to normalise the keypoint coordinates by using the method of [75] before applying the above algorithms for computing $\mathbf{f}$. This requirement is due to the fact that some columns in $\mathbf{D}$ have quadratic terms in p, p', q, and q', while others have linear terms. Without prior normalisation, the columns with quadratic terms in $\mathbf{D}$ would have much larger values than others with linear terms, and hence are subject to proportionally smaller perturbations because the sum of squared algebraic errors is minimised. By contrast, if normalised coordinates are employed, the columns of $\mathbf{D}$ will have similar ranges of values, and hence are treated approximately equally. In practice, with normalised coordinates, the dominant influence of quadratic

[1] Actually 7 matches are sufficient since $\mathbf{F}$ is rank deficient by one, but the estimation process from 7 matches is more complicated. In any case, the rank constraint can be imposed post-estimation from 8 matches [75].

terms in the algebraic error (i.e, heteroscedastic noise in the columns of $\mathbf{D}$) is significantly reduced [75]. Note that mixed OLS-TLS does not aim to deal with the above issue, except for the purpose of preventing the column with the homogeneous constant 1 in $\mathbf{D}$ from being rectified.

We now compare the performances of DLT, TLS and mixed OLS-TLS in the task of fundamental matrix estimation. On the *Hartley* image pair in Figure 4.8, we detect and match SIFT keypoints [16] using the VLFeat toolbox [81] and manually identify the correct correspondences (there are 123). The keypoint coordinates are then normalised such that they are centred at the origin and the mean distance to the origin is $\sqrt{2}$ [75]. The number of keypoint matches, n, used to compute the fundamental matrix ranges between 8 and 93. For each value of n, 100 distinct samples of size n are randomly drawn from the original set of 123 inliers, and the three algorithms are then run respectively on these samples (as discussed above, for TLS and mixed OLS-TLS we dehomogenise by fixing the first element in $\mathbf{f}$ to -1 and move the corresponding column in $\mathbf{D}$ to the RHS to yield the observation vector $\mathbf{y}$). The errors (i.e., algebraic error and Sampson distance [3]) are computed using all 123 inliers, and for each value of n, the average errors over 100 runs of the algorithms are shown in Figure 3.4. The results show that the three algorithms have very comparable performances, which shows the equivalence between DLT and TLS techniques. In fact, when the number of matches is small, the performances of both TLS and mixed OLS-TLS are slightly worse than that of DLT, which is likely due to the dehomogenisation step. Moreover, by constraining the constant column in $\mathbf{D}$ from being modified, mixed OLS-TLS obtained slighly lower errors than TLS in cases of small number of matches.



(a) Algebraic error **(b)** Sampson distance

Figure 3.4: Performance comparison of DLT, TLS and mixed OLS-TLS on fundamental matrix estimation.

Note that DLT (3.49) minimises the sum of squared *algebraic* errors. It is also common to minimise *geometric* errors in geometric fitting problems in computer vision, such as the Sampson distance [3]. Such errors are typically nonlinear functions of the parameters, thus extending our TLS expansion (3.34) to accommodate such errors is nontrivial. In some cases, geometric errors

are minimised during the refinement stage (i.e., after robust estimation) using iterative methods (e.g., Levenberg-Marquardt), and since the error contributions of all data are considered simultaneously, the issue of data span is not as critical.

3.4.2 The influence of data span on fundamental matrix estimation

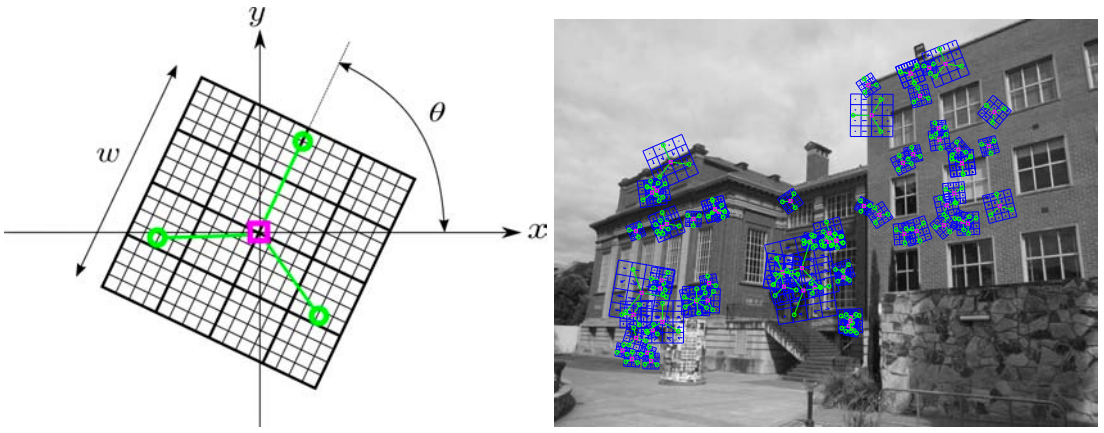
We have spent the bulk of this chapter on quantifying the ways in which large span minimal subsets are weighted in OLS and TLS (or equivalently DLT) techniques to solving problems (or computer vision problems in particular). What the results in previous sections show is that such methods automatically weight large span minimal subsets accordingly. In approaches that rely on using a fraction of the data to derive a model estimate (e.g., RANSAC), it is obvious that to achieve statistical efficiency one needs to look towards preferring larger span samples than smaller span ones (this is obvious as we acknowledged at the outset). However, as intuitive as it is, it may be perhaps overlooked as we now illustrate. In the following illustration, we take the fundamental matrix estimation algorithms of [21, 55] which require only a few number of keypoint matches as an example and we show that the difference in the performance is likely due to data span. This result appears not to have been known (at the very least not stated) by the previous work of [21, 55].

Many keypoint detectors used in multi-view geometry are really *2D region* detectors. Some of the more common detectors are designed to detect affine invariant regions [49], including the SIFT method [16]. This fact has been exploited in [21, 55] to reduce the minimum number of keypoint matches for yielding a fundamental matrix estimate, since two unique planar correspondences contain the sufficient degrees of freedom.

In practice, these methods [21, 55] generate extra matches from the keypoint matches to make up the required data for estimation. Given three keypoint matches, Chum et al. [21] generate two extra matches per keypoint match, yielding in total nine keypoint matches. Given two keypoint matches, Goshen and Shimshoni [55] generate three extra matches per keypoint match, producing eight keypoint matches. Henceforth, we call the two methods *2-* and *3-point* to highlight the number of unique keypoint matches used. The premise behind these methods is that under a sampling scheme for robust estimation (e.g., RANSAC), only two or three inliers (correct correspondences) need to be sampled to get an all-inlier sample (instead of eight or nine inliers).

To generate the extra matches, the 2- and 3-point methods exploit the scale and orientation information output by keypoint detectors. In particular, the extra keypoints are produced within the region (of size proportional to the scale) of the original keypoint; see Figure 3.5(a) for an illustration of this idea. However, since the keypoint regions are usually small (e.g., see Figure 3.5(b)), the 2- and 3-point methods effectively limit the span of data used for estimation.

Here, we compare these estimation methods against those that actually use 8 or 9 unique keypoint matches (which we respectively call *8-* and *9-point* methods in the following).



(a) Illustration of extra keypoints produced within a keypoint region with orientation θ and width w

(b) Extra keypoints generated on one of the *Barr-Smith* image pair in Figure 4.8

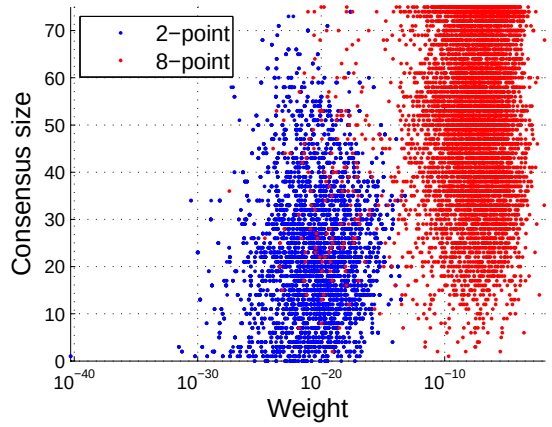
Figure 3.5: Extra keypoints generated by using the 2-point method [55]. Original keypoints are in magenta while extra keypoints are in green.

On the *Barr-Smith* image pair in Figure 4.8, we detect and match SIFT keypoints [16] using the VLFeat toolbox [81] and manually identify the correct correspondences (there are 75). The keypoint coordinates are also normalised by using the method of [75]. We randomly draw 10,000 samples of sizes 2, 3, 8 and 9 over the set of 75 inliers. For each sample, a fundamental matrix is estimated using DLT, where extra matches are produced for the 2- and 3-point methods by following [21, 55]. We calculate the consensus size (for a tuned inlier threshold) and weight for each sample, based on the following rules:

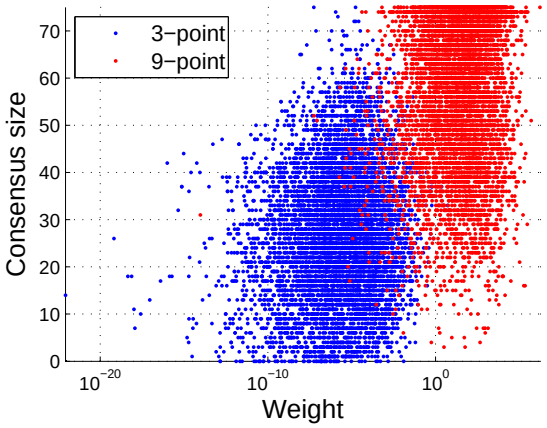
- For the 8- and 9-point methods, we simply compute the weight as $|\mathbf{X}(\nu)^T\mathbf{X}(\nu)|$, where ν indexes a subset of 8 or 9 rows sampled from the data matrix $\mathbf{X} \in \mathbb{R}^{n\times 8}$.
- For the 2- and 3-point methods, we calculate the weight as $|\bar{\mathbf{X}}^T\bar{\mathbf{X}}|$, where $\bar{\mathbf{X}}$ contains a subset of 2 or 3 rows sampled from $\mathbf{X}$ and the rows corresponding to the *extra* generated matches.

We plot consensus size versus weight of the samples. Figure 3.6(a) contains the results of the 2- and 8-point methods, while Figure 3.6(b) presents the results of the 3- and 9-point methods. We separated these plots since the subset weights calculated from using 8 and 9 rows of data have different *units*. For clearer comparisons, Figures 3.6(c) to 3.6(f) show the same results using kernel density estimates of weight and consensus size.

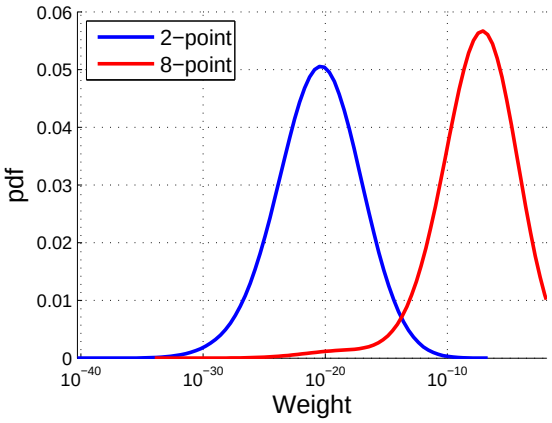
From the results, we can see that the 2- and 3-point methods do have a tendency to produce data with smaller span. Furthermore, there is a clear correlation between the subset span and the quality of model estimate, indicating the danger posed by using data with small span for



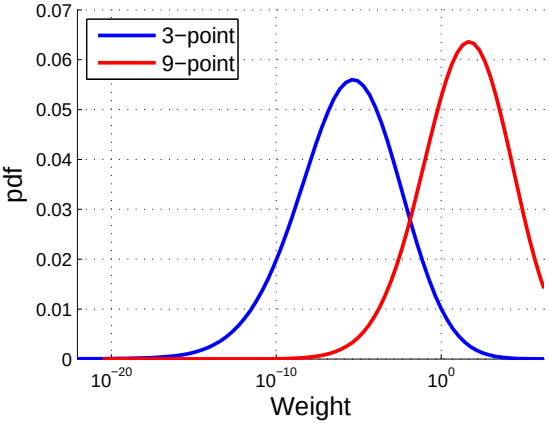
(a) Clusters of sampled subsets



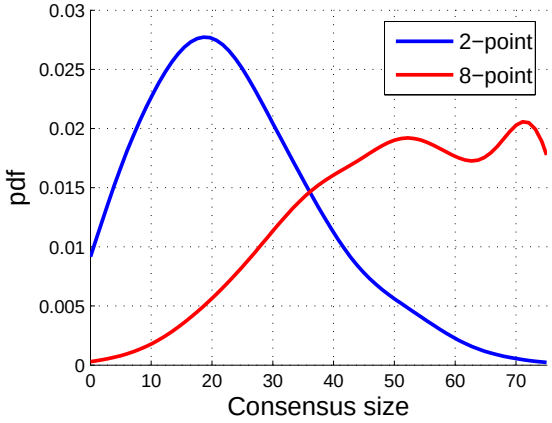
(b) Clusters of sampled subsets



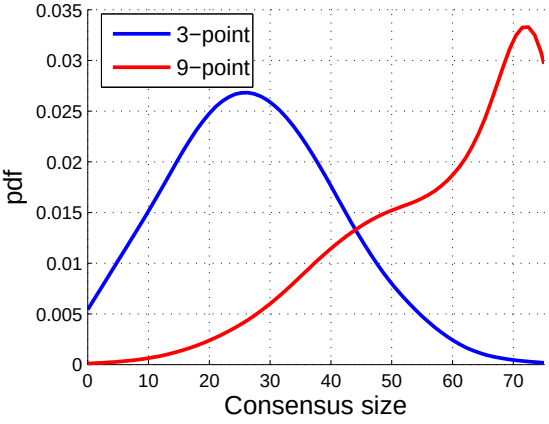
(c) Distribution of weights



(d) Distribution of weights



(e) Distribution of consensus sizes



(f) Distribution of consensus sizes

Figure 3.6: Performance comparison of fundamental matrix estimation algorithms. First column: 2- and 8-point methods. Second column: 3- and 9-point methods. Note that in (a)–(d), the horizontal (weight) axis is in logarithmic scale.

estimation. Goshen and Shimshoni [55] have experimentally shown that the model estimates from the 2- and 3-point methods are usually inferior to those from the 8- and 9-point methods. Our results conclusively show that this phenomenon is due, in a large part, to data span.

Note that another common estimation approach is the 7-point method [3]. In theory, $\mathbf{F}$ has only 7 degrees of freedom since it is homogeneous and only rank-2, thus 7 unique keypoint matches are actually sufficient. However, the estimation of the 7-point method involves solving nonlinear equations whereby the rank constraint is imposed *during* computation. This approach is unlike the 2-, 3-, 8- and 9-point methods whose estimation (via DLT) is linear and where the rank constraint is imposed afterwards. Therefore, we did not test the 7-point method from the aspect of span. In any case, Goshen and Shimshoni [55] have shown that the 7-point method is usually inferior to the 8- or 9-point methods (see [55, Figure 5]).

3.5 Summary

When sampling minimal subsets for robust geometric fitting, it is commonly known that obtaining an all-inlier sample is not sufficient to guarantee a good model estimate; the points therein should also have a large spatial coverage. We started this chapter by showing that a theoretical explanation for this principle lies in a result by Jacobi, where the least squares regression estimate can be expanded as a weighted sum of all possible minimal subset estimates. It turns out that the weight of a minimal subset estimate is a function of the span of the associated data points. Our main contribution in this chapter is to show that an analogous minimal subset expansion can be developed for total least squares. We also highlighted the equivalence between total least squares and the direct linear transformation, which is a common geometric estimation technique in computer vision. Moreover, we studied the influence of data span on various fundamental matrix estimation algorithms. The results showed a clear correlation between the span of data points used for estimation and the quality of model estimate.

Chapter 4

Large Span Sampling for Robust Geometric Fitting

4.1 Introduction

The previous chapter provided a theoretical basis for large span sampling in robust geometric fitting, and showed that the span of data points used for estimation is highly related to the quality of model estimate. In Chapter 2, we reviewed a variety of guided sampling methods [20–23, 33–37] which aim to accelerate the retrieval of all-inlier minimal subsets. However, none of those methods *actively* searches for all-inlier samples with large span which are important to yielding a good model estimate that fits the data well.

In this chapter, we first investigate the true performance of distance-based guided sampling that is frequently employed to speed up the generation of all-inlier minimal subsets. In particular, we highlight the danger posed by *proximity* sampling [33, 34] which actually *limits* the span of all-inlier minimal subsets produced, and we explain why sampling *based on distances alone* is not a generally reliable strategy for generating all-inlier samples with large span. Based on these insights, we develop the main contribution of this chapter — a novel sampling algorithm which, unlike previous approaches, consciously targets all-inlier minimal subsets with large span for robust geometric fitting.

It is also vital to avoid degeneracies in robust geometric fitting, specifically, in two-view geometry estimation [82, 83] when some scenes contain a dominant plane on which most of the keypoints lie. Degenerate parameters are estimated from a minimal subset containing entirely keypoint matches from the dominant plane. Previous research efforts have tackled this issue by detecting when a degenerate estimate is produced, then recovering from the degenerate estimate

by using the plane-and-parallax method [82, 83]. Theoretically, degenerate minimal subsets correspond to data with very small span or insufficient degrees of freedom. Since our new sampling algorithm directly targets samples with large spatial extent, it *actively* prevents degeneracies; see results in Section 4.3. Our sampling method can also be complemented with the degeneracy detection and recovery techniques of [82, 83].

The rest of this chapter is organised as follows: Section 4.2 explains why naive distance-based sampling fails as a strategy to maximise the span of all-inlier minimal subsets produced, and presents our novel sampling scheme for all-inlier samples with large span. In Section 4.3, we benchmark the effectiveness of our proposed method against state-of-the-art approaches on synthetic and real datasets, and analyse the performance under degeneracies. We then summarise this chapter in Section 4.4. The main content of this chapter is based on the author's publication in [65].

4.2 Guided sampling with large span

We develop in this section a new guided sampling algorithm that consciously seeks all-inlier minimal subsets with large span for robust geometric fitting. Following Chapter 3, we focus on the task of robust fundamental matrix estimation (see Section 3.4.1), and use minimal subsets of size eight to avoid limiting the data span (see Section 3.4.2).

4.2.1 Distance-based sampling

First, we study the performance of distance-based guided sampling (Section 2.2.1.2) which exploits distances between data points in the spatial domain to increase the chance of hitting all-inlier minimal subsets. Given a set of keypoint correspondences $\mathcal{X} = \{(\mathbf{u}_i, \mathbf{v}_i)\}_{i=1}^{n}$, the keypoint coordinates are first normalised by using the method of [75]. The linearisation (3.48) is then performed to yield the data matrix $\mathbf{D}$, which we partition into $[\mathbf{X}\ \mathbf{y}] \in \mathbb{R}^{n\times 9}$; see Section 3.4.1. Note that each row of $[\mathbf{X}\ \mathbf{y}]$ is a "datum", which consists of dependent measurements $\mathbf{x}_i \in \mathbb{R}^8$ and an independent measurement y_i arising from the keypoint correspondence $(\mathbf{u}_i, \mathbf{v}_i)$.

Distance-based guided sampling algorithms such as [33, 34] require a notion of distance between two data points. For example, NAPSAC [33] uses the Euclidean distance of the keypoint coordinates prior to linearisation as

$$d_{ij} = \left\| [p_i\ q_i\ p'_i\ q'_i]^T - [p_j\ q_j\ p'_j\ q'_j]^T \right\|, \tag{4.1}$$

while the method of [34] simply uses the distance in the first view

$$d_{ij} = \left\| [p_i \; q_i]^T - [p_j \; q_j]^T \right\| . \tag{4.2}$$

Here, to be consistent with our results in Section 3.3.1, we use the Euclidean distance of the dependent measurement vectors as

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|, \tag{4.3}$$

since it is the span of the rows of the design matrix $\mathbf{X}$ that matter. In any case, we have verified that using (4.3) gives similar results as (4.1) and better results than (4.2).

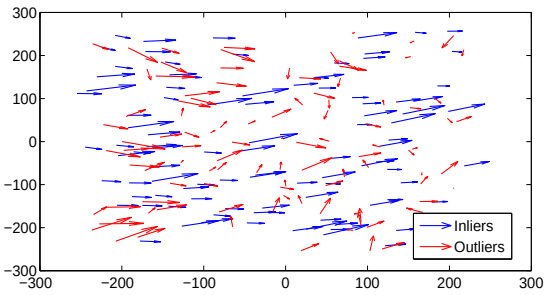
As discussed in Section 2.2.1.2, the inliers establish a lower dimensional manifold than the outliers. In particular, for robust fundamental matrix estimation, the inliers lie on a 3D manifold which can be parameterised by using the coordinates of the keypoint $\mathbf{u}$ in the first view and the location of the corresponding keypoint $\mathbf{v}$ on the epipolar line $\mathbf{l}_2$ in the second view; see Figure 3.3. By contrast, the outliers mostly spread over the 4D joint image space which is modelled by using the coordinates of the keypoints $\mathbf{u}$ and $\mathbf{v}$ in both views. Due to this distinction in dimensionality, some regions of the spatial domain are more densely populated by inliers, and the inlier rates of these regions are thus higher than the global inlier rate of the set of all data. Based on this observation, proximity sampling [33, 34] focuses on sampling neighbouring data points within local regions to increase the probability of obtaining all-inlier minimal subsets. Algorithm 4.1 summarises proximity sampling.

Figure 4.1(a) shows 100 synthetic point matches (inliers) arising from an underlying fundamental matrix generated by using Torr's SfM Toolkit[1], on top of which we add 100 uniformly sampled outliers; in total $n = 200$ matches. To show the benefit of proximity sampling, we calculate the percentage of inliers that are within a radius r of every inlier, where r is varied by changing a multiplier k in

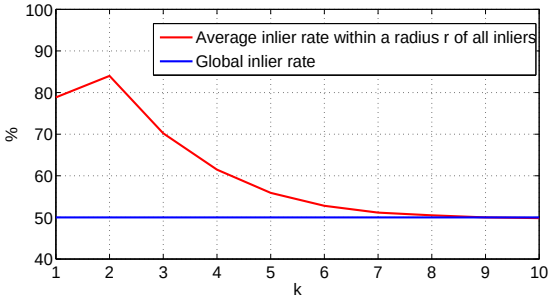
$$r = \frac{k}{n} \sum_{i=1}^{n} \min_{j \neq i} \|\mathbf{x}_i - \mathbf{x}_j\|, \tag{4.4}$$

i.e., r is k times the average nearest neighbour distance. Figure 4.1(b) shows the average inlier rate within a radius r of all inliers. The results show that as k is small, the inlier rate is much higher than the global inlier rate of 50%, and as k increases, the inlier rate decreases gradually and approaches the global inlier rate. Note that when $k = 1$, within distance r there are only a few data points, which makes the inlier rate lower than the highest inlier rate at $k = 2$. Moreover, a randomly selected inlier is highlighted in Figure 4.1(c), where the data $\mathbf{X}$ are projected to 2D using PCA for visualisation, and the set of data points within distance r of the chosen inlier is

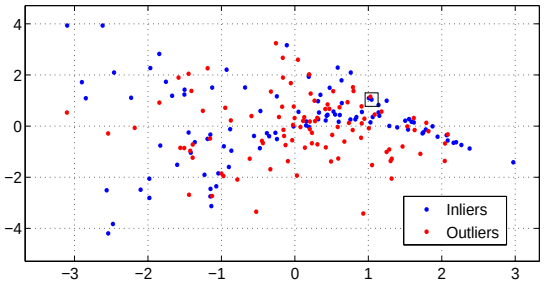
[1] Obtained from http://cms.brookes.ac.uk/research/visiongroup/



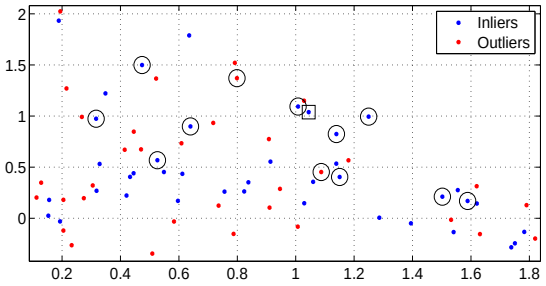
(a) Synthetically generated point matches with 100 inliers arising from an underlying fundamental matrix, and 100 outliers sampled uniformly within the image domain.



(b) The average inlier rate within a radius r of all inliers (r is varied by changing a multiplier k as in (4.4)).



(c) An inlier (highlighted with a square) is chosen as the first datum for a minimal subset (the data $\mathbf{X}$ are projected to 2D using PCA for visualisation).



(d) Zoom-in of (d) showing data (circled) within a radius r of the chosen inlier (r is 2 times the average nearest neighbour distance), and 10 out of 12 (83.3%) are inliers.

Figure 4.1: Proximity sampling for fundamental matrix estimation.

Algorithm 4.1: Proximity sampling

Input: Design matrix $\mathbf{X} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]^T$, radius r.

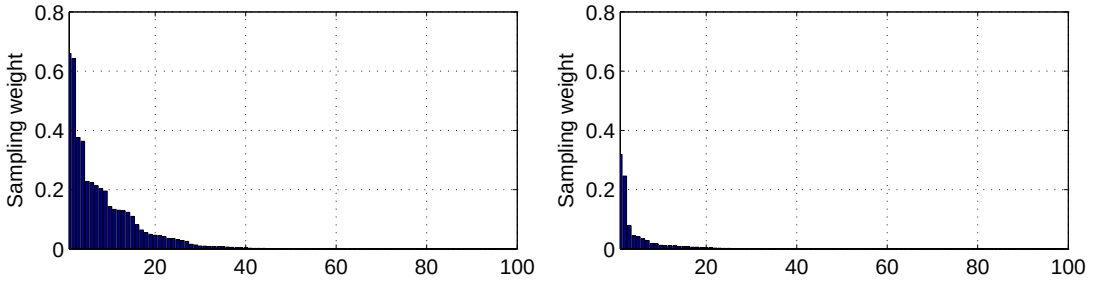
1: $\mathbf{d}_1 \leftarrow$ a row of $\mathbf{X}$ sampled randomly.
2: **for** $j := 2, \ldots, m$ **do**
3: $\quad \mathbf{d}_j \leftarrow$ a row of $\mathbf{X}$ sampled based on the weights

$$P(\mathbf{x}_i) \sim \exp\left(-\|\mathbf{x}_i - \mathbf{d}_1\|^2 / 2r^2\right). \tag{4.5}$$

4: **end for**

identified in Figure 4.1(d), where r is 2 times the average nearest neighbour distance. Observe that within distance r of the chosen inlier, the inlier percentage is 83.3%, which is much higher than the global inlier percentage. Figure 4.2 shows the sampling weights (4.5) centred on the chosen inlier in Figure 4.1(c), where r is also 2 times the average nearest neighbour distance. It can be seen that inliers have a higher probability of being selected than outliers.

However, it is clear that proximity sampling does not encourage minimal subsets with large span. One may use a larger radius r to increase the data span, but this change impacts the ability to retrieve all-inlier samples. Figure 4.3 illustrates the effects of increasing r on the number of all-inlier minimal subsets and maximum consensus (for a suitably tuned inlier threshold) achievable within 300 iterations of Algorithm 4.1. These are median results over 100 synthetic data generated à la Figure 4.1(a), and radius r is varied by changing a multiplier k as in (4.4). It can be seen that as k increases, the ability to find all-inlier minimal subsets decreases exponentially,



(a) Inlier weights **(b)** Outlier weights

Figure 4.2: Weights for the data in Figure 4.1(a) computed according to (4.5) (Algorithm 4.1) when it is centred on the inlier chosen in Figure 4.1(c). The data are first sorted in increasing distance to the chosen inlier. Radius r is 2 times the average nearest neighbour distance.

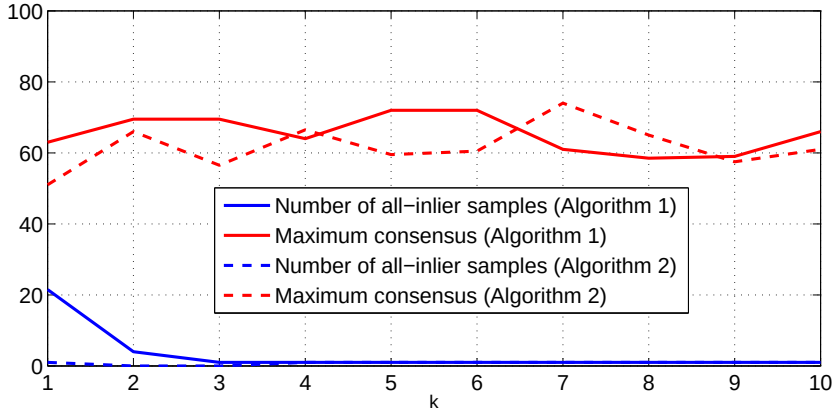


Figure 4.3: Performance of proximity sampling and proximity sampling with offset as a function of multiplier k (radius r).

without a commensurate increase in maximum consensus. Note that due to the possibility of further refinement, it is not necessary for the algorithm to obtain all 100 inliers for yielding a good model estimate. However, the algorithm still probably needs to achieve around 80% of the inliers, but it can not obtain that. Moreover, in some cases where efficiency is important and further refinement is not available, the algorithm should achieve as close as possible to 100 inliers to guarantee a satisfactory model estimate.

Instead of placing the mode of the sampling distribution on the first datum of the minimal subset, one might suggest offsetting the distribution to favour data points that are farther away. This method is summarised in Algorithm 4.2, which requires an extra offset parameter t. Ideally t should be proportional to the maximum achievable span among *all the inliers* in the data. We repeat the above experiment using Algorithm 4.2 for increasing k. The offset t is set as 2 times the average pairwise distance among the inlier portion of the data (this knowledge is not available in practice). Figure 4.3 shows that the performance (median over 100 unique datasets) of Algorithm 4.2 is not better than Algorithm 4.1. In fact the ability to retrieve all-inlier minimal subsets is significantly decreased.

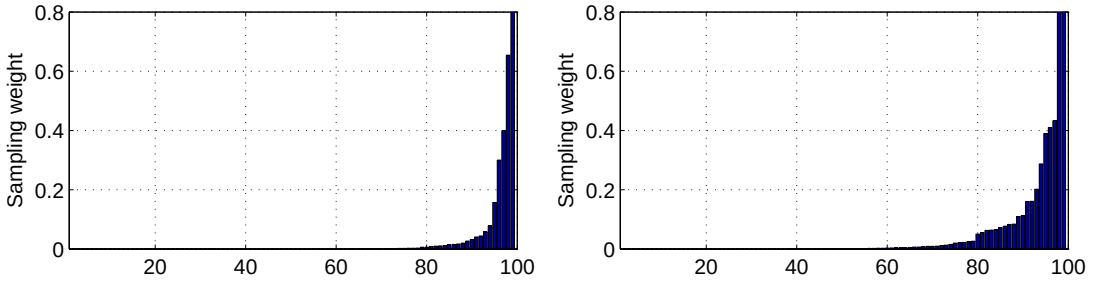
Algorithm 4.2: Proximity sampling with offset

Input: Design matrix $\mathbf{X} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]^T$, radius r, offset t.

1: $\mathbf{d}_1 \leftarrow$ a row of $\mathbf{X}$ sampled randomly.
2: **for** $j := 2, \ldots, m$ **do**
3: $\quad \mathbf{d}_j \leftarrow$ a row of $\mathbf{X}$ sampled based on the weights

$$P(\mathbf{x}_i) \sim \exp\left(-(\|\mathbf{x}_i - \mathbf{d}_1\| - t)^2 / 2r^2\right). \tag{4.6}$$

4: **end for**



(a) Inlier weights **(b)** Outlier weights

Figure 4.4: Weights for the data in Figure 4.1(a) computed according to (4.6) (Algorithm 4.2) when it is centred on the inlier chosen in Figure 4.1(c). The data are first sorted in increasing distance to the chosen inlier. Radius r is 2 times the average nearest neighbour distance, while offset t is 2 times the average pairwise distance among *all the inliers*.

One might suspect that the offset t is not properly tuned. Figure 4.4 plots the weights (4.6) of the data in Figure 4.1(a) with the inlier chosen in Figure 4.1(c) as the first datum. Again t is 2 times the average pairwise distance among all the inliers. Observe that the distribution favours the inliers which are as far away as possible, i.e., t is optimal in this case (in the general case the optimal offset is a priori unknown). However equally favoured are many outliers, rendering the distribution ineffective in generating all-inlier minimal subsets.

Contrary to popular opinion, the results indicate that simple distance-based sampling can not be relied upon to retrieve all-inlier minimal subsets with large span. For Algorithm 4.2 to work, the inliers have to form a *single tight cluster*, a condition that can not be guaranteed in general (cf. Algorithm 4.1 only requires that inliers are *locally dense*). In Section 4.3 we provide further results on synthetic and real data that support the findings in this section.

4.2.2 Combining Multi-GS with distance-based sampling

We employ the recently proposed conditional sampling algorithm Guided Sampling for Multi-Structure Model Fitting - Multi-GS (Section 2.2.1.4) as the basis of our new method. Instead of analysing distances between data points as distance-based sampling, Multi-GS [36, 37] represents each datum using a preference and analyses data preferences to increase the chance of

hitting all-inlier minimal subsets. Preferences (or rankings) have been widely used in our everyday lives. For example, for many prizes and awards such as the Nobel prize and Oscars award, committee members are asked to rank nominated candidates based on their preferences, and these preferences are then analysed for selecting recipients. Various nonparametric statistical methods [84, 85] have made use of preferences for analysing data since obtaining and comparing preferences does not require any data-dependent knowledge. Based on this observation, Multi-GS has applied nonparametric analysis using preferences for solving the problem of geometric fitting.

Preferences (or rankings) arise when *judges* evaluate a set of *objects* and rank them based on their preferences. In Multi-GS, data points and model hypotheses are considered respectively as judges and objects, and the preference of a datum towards a set of hypotheses is derived from sorting its residuals (to the hypotheses) in ascending order, since the residual is essentially the fitting error, i.e., the smaller the residual is, the better the datum fits the hypothesis. In particular, let $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\}$ be a set of N model hypotheses generated thus far, and for each datum $\mathbf{x}_i$, its residuals to the hypotheses are computed as

$$\mathbf{r}^{(i)} = \begin{bmatrix} r_1^{(i)} & r_2^{(i)} & \ldots & r_N^{(i)} \end{bmatrix}, \tag{4.7}$$

where $r_l^{(i)} = \|y_i - \boldsymbol{\beta}_l \mathbf{x}_i\|$ is the residual of datum $\mathbf{x}_i$ to hypothesis $\boldsymbol{\beta}_l$. The residuals are then sorted in increasing order to yield the permutation

$$\mathbf{a}^{(i)} = \begin{bmatrix} a_1^{(i)} & a_2^{(i)} & \ldots & a_N^{(i)} \end{bmatrix}, \tag{4.8}$$

where $r^{(i)}_{a_u^{(i)}} \leq r^{(i)}_{a_v^{(i)}}, \forall\, u < v$. The permutation $\mathbf{a}^{(i)}$ is called the *preference* of datum $\mathbf{x}_i$ towards the set of hypotheses $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\}$. To examine the similarity/dissimilarity of data preferences, the *preference correlation* between two data points $\mathbf{x}_i$ and $\mathbf{x}_j$ is defined as

$$f(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{h} \left| \mathbf{a}^{(i)}_{1:h} \cap \mathbf{a}^{(j)}_{1:h} \right|, \tag{4.9}$$

where $\mathbf{a}^{(i)}_{1:h}$ are the first-h elements of $\mathbf{a}^{(i)}$, and $\cap$ is set intersection. The bandwidth h is typically set as $\lceil 0.1N \rceil$ [36, 37]. Intuitively, the preference correlation (4.9) measures the degree of overlaps among the top-h most preferred hypotheses of $\mathbf{x}_i$ and $\mathbf{x}_j$.

In Multi-GS, preference correlations are employed to compute the conditional sampling weights for generating all-inlier minimal subsets. Algorithm 4.3 summarises Multi-GS[2]. Figure 4.5 shows the pairwise matrix of preference correlation values for the data in Figure 4.1(a), based on

[2]Note that unlike Algorithms 4.1 and 4.2, Algorithms 4.3 and 4.4 update the sampling distribution (Step 5) according to the data available so far in the minimal subset. This modification can also be done for Algorithms 4.1 and 4.2, e.g., by recentring (4.5) and (4.6) on the datum last sampled. However our experiments suggest that this produces worse performance.

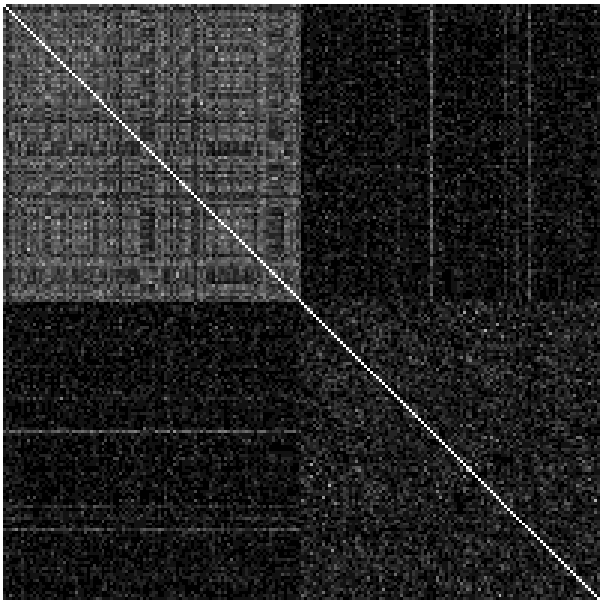


Figure 4.5: Pairwise matrix of preference correlation values (4.9) for the data in Figure 4.1(a), where the first-100 rows/columns correspond to inliers and the rest to outliers. The preferences are based on 150 fundamental matrix hypotheses.

Algorithm 4.3: Multi-GS

Input: Design matrix $\mathbf{X} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]^T$, preferences $\{\mathbf{a}^{(i)}\}_{i=1}^n$ towards a set of model hypotheses $\{\boldsymbol{\beta}_l\}_{l=1}^N$.

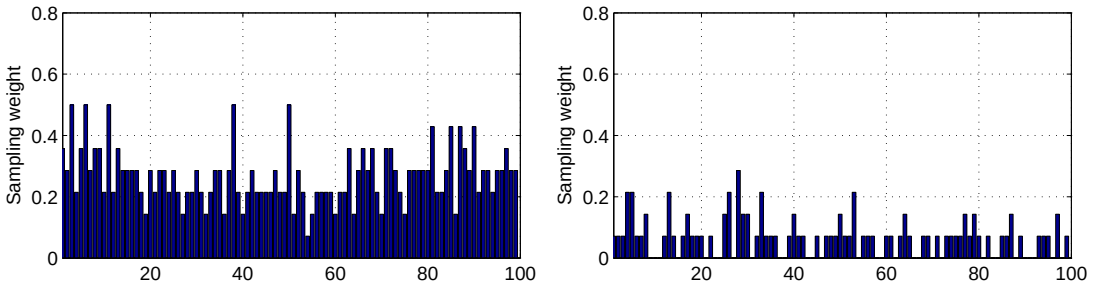
1: $\mathbf{d}_1 \leftarrow$ a row of $\mathbf{X}$ sampled randomly.
2: Initialise $\mathbf{w} = [f(\mathbf{x}_1, \mathbf{d}_1) \ \ldots \ f(\mathbf{x}_n, \mathbf{d}_1)]$.
3: **for** $j := 2, \ldots, m$ **do**
4: Sample a row $\mathbf{d}_j$ from $\mathbf{X}$ based on the weights

$$P(\mathbf{x}_i) \sim \mathbf{w}(i). \tag{4.10}$$

5: $\mathbf{w} = \mathbf{w} \odot [f(\mathbf{x}_1, \mathbf{d}_j) \ \ldots \ f(\mathbf{x}_n, \mathbf{d}_j)]$.
6: /* $\odot$ means element-wise product.*/
7: **end for**

150 previously generated fundamental matrix hypotheses. It can be seen that inliers have higher preference correlations than outliers. Figure 4.6 illustrates the sampling weights when (4.9) is centred on the inlier chosen in Figure 4.1(c) (these would be taken from the row corresponding to the chosen inlier in Figure 4.5). Observe that inliers generally have much higher sampling weights than outliers. Moreover, this effect was achieved without having a model hypothesis in $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\}$ fitted on an all-inlier minimal subset; see [36, 37] for details.

Although it does not favour all-inlier minimal subsets with small spatial extent (observe that the inlier weights in Figure 4.6 are evenly distributed), Multi-GS does not explicitly aim for large inlier span either. This fact is a source of inefficiency. Furthermore, the experiments in [36, 37] only examined the ability of Multi-GS to retrieve all-inlier minimal subsets, and not the *actual quality* of the fitted model hypotheses.



(a) Inlier weights **(b)** Outlier weights

Figure 4.6: Weights for the data in Figure 4.1(a) computed according to (4.9) (Algorithm 4.3) when it is centred on the inlier chosen in Figure 4.1(c). The data are first sorted in increasing distance to the chosen inlier. Preferences are induced from 150 previously generated fundamental matrix hypotheses.

We propose an extension to Multi-GS to rectify this inadequacy. Our idea is to combine Multi-GS with distance-based sampling. Specifically, we multiply the preference correlations with the distance-based distribution (4.6) to yield

$$P(\mathbf{x}_i) \sim f(\mathbf{x}_i, \mathbf{d}).\exp\left(-(\|\mathbf{x}_i - \mathbf{d}\| - t)^2/2r^2\right), \tag{4.11}$$

where $\mathbf{d}$ is the current centre of the distribution (e.g., the datum previously added to the minimal subset). Our method is summarised in Algorithm 4.4. Figure 4.7 illustrates the idea by multiplying the weights in Figures 4.6 and 4.4; the outlier weights are severely attenuated, while the weights corresponding to inliers with *maximum distances* remain high.

A practical difficulty is determining the appropriate value for the offset t. In our experiments we obtain t as 2 times the average pairwise distance among the consensus set of the model hypothesis $\boldsymbol{\beta}_*$, which has the largest consensus S_* thus far, i.e.,

$$t = \frac{4}{|S_*|(|S_*|-1)} \sum_{\mathbf{x}_i, \mathbf{x}_j \in S_*} \|\mathbf{x}_i - \mathbf{x}_j\|, \tag{4.12}$$

where $S_* = \{\mathbf{x}_i \mid \mathbf{x}_i \in \mathbf{X},\ r_*^{(i)} = \|y_i - \boldsymbol{\beta}_* \mathbf{x}_i\| \leq \theta\}$, and θ is the inlier threshold. The idea is to iteratively gauge the maximum span based on the evidence available on-the-fly.

4.3 Experimental results

In this section we evaluate the performance of our proposed method (Algorithm 4.4, henceforth, *Multi-GS-Offset*) on synthetic and real datasets. We benchmark Multi-GS-Offset against the following state-of-the-art approaches:

1. Pure random sampling [18] (Random);

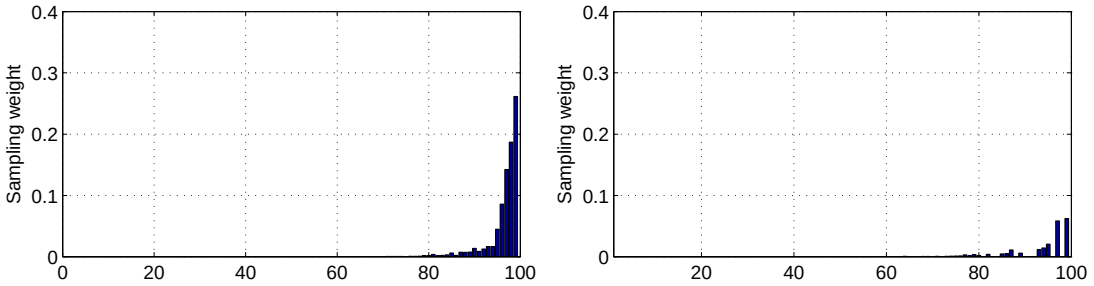
Algorithm 4.4: Multi-GS with offset

Input: Design matrix $\mathbf{X} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]^T$, preferences $\{\mathbf{a}^{(i)}\}_{i=1}^n$ towards a set of model hypotheses $\{\boldsymbol{\beta}_l\}_{l=1}^N$, radius r, offset t.

1: $\mathbf{d}_1 \leftarrow$ a row of $\mathbf{X}$ sampled randomly.
2: Initialise $\mathbf{w} = [f(\mathbf{x}_1, \mathbf{d}_1) \ \ldots \ f(\mathbf{x}_n, \mathbf{d}_1)]$.
3: **for** $j := 2, \ldots, m$ **do**
4: Sample a row $\mathbf{d}_j$ from $\mathbf{X}$ based on the weights

$$P(\mathbf{x}_i) \sim \mathbf{w}(i).\exp\left(-(\|\mathbf{x}_i - \mathbf{d}_{j-1}\| - t)^2/2r^2\right). \tag{4.13}$$

5: $\mathbf{w} = \mathbf{w} \odot [f(\mathbf{x}_1, \mathbf{d}_j) \ \ldots \ f(\mathbf{x}_n, \mathbf{d}_j)]$.
6: /* $\odot$ means element-wise product.*/
7: **end for**



(a) Inlier weights **(b)** Outlier weights

Figure 4.7: Weights for the data in Figure 4.1(a) computed according to (4.11) (Algorithm 4.4) when it is centred on the inlier chosen in Figure 4.1(c). The data are first sorted in increasing distance to the chosen inlier. Radius r is 2 times the average nearest neighbour distance, while offset t is 2 times the average pairwise distance *among all inliers*.

2. Proximity sampling [33, 34] (Proxim);
3. Proximity sampling with offset (Proxim-Offset);
4. Multi-GS [36, 37];
5. LO-RANSAC [20, 21];
6. Guided-MLESAC [22]; and
7. PROSAC [35].

As surveyed in Chapter 2, LO-RANSAC [20, 21] introduces the local exploitation steps, which generate model hypotheses from the largest consensus obtained thus far. This method bears some similarity with Multi-GS-Offset, which iteratively estimates the maximum span of the inliers within the largest consensus received so far. State-of-the-art methods like PROSAC [35] and Guided-MLESAC [22] employ the matching score information to bias the sampling process. These matching scores are difficult to simulate realistically for synthetic data. Therefore, we only evaluate PROSAC and Guided-MLESAC on real data.

4.3.1 Stopping criterion

For all methods, we use the common RANSAC stopping criterion which estimates the number of samples by using Equation (2.1). This stopping criterion requires an estimate of the outlier ratio, which is progressively decreased based on the largest consensus found thus far. Although this stopping criterion assumes pure random sampling, following [20–22, 33] we also apply it in the guided sampling methods. Under guided sampling, the stopping criterion *overestimates* the number of samples to retrieve. This does not necessarily mean that guided sampling will take as long as pure random sampling, since guided sampling methods increase the consensus size (hence decrease the outlier ratio) faster.

In theory, given a sufficient amount of time, any sampling method (including pure random sampling) can obtain the maximum achievable consensus. An objective performance measure is therefore how soon the method hits a sufficiently high consensus to satisfy the stopping criterion. A lower run time thus reflects a higher accuracy in generating minimal subsets with good model estimates. Consequently, we will use run time as the main benchmark in our experiments.

4.3.2 Synthetic data

We first test on synthetic data generated by using Torr's SfM Toolkit; a sample of the synthetic data is in Figure 4.1(a). For each data instance, a unique fundamental matrix is generated, from which 100 inlying point matches are sampled and then contaminated with additive Gaussian noise of standard deviation of 5 pixels. A number of outliers are created by randomly sampling points in the image domain (500×500 pixels). The points are normalised by using the method of [75].

Parameter settings for all methods are as follows: Offset t for Proxim-Offset and Multi-GS-Offset is computed as in (4.12), while radius r (used also in Proxim) is set as $t/2$. The inlier threshold θ is set as 0.0005; this parameter is used in calculating the consensus size, the consensus set in the local exploitation steps of LO-RANSAC, and in evaluating (4.12) for t. Finally, for Multi-GS and Multi-GS-Offset, the preferences $\{\mathbf{a}^{(i)}\}_{i=1}^{n}$ are updated only after every 10 hypotheses; see [36, 37] for details.

We set the number of outliers as 50, 100, 150 and 200 (respectively, 33%, 50%, 60% and 67% outlier rates). For each outlier rate we generate 100 unique data instances and run each method. For each run we record the following measures:

1. The number of minimal subsets sampled;
2. The number of *all-inlier* minimal subsets sampled;

3. The maximum and median span among *all-inlier* minimal subsets sampled, where the span of a minimal subset λ is measured as $|\mathbf{X}(\lambda)|^2$;

4. The maximum consensus among *all* hypotheses;

5. The number of true inliers within the consensus set of the maximum consensus hypothesis;

6. The classification error (number of mislabelled data points, i.e., inliers labelled as outliers and vice versa) of the maximum consensus hypothesis;

7. The run time based on the common RANSAC stopping criterion.

Note that LO-RANSAC involves the local exploitation steps that generate larger-than-minimal subsets (following [20, 21], the subset size for the local exploitation steps is 14). To allow all sampled subsets from LO-RANSAC to be accounted for, we count a local exploitation subset as a "minimal subset" so as to standardise notations. Furthermore, we also approximate the span of a local exploitation subset ν as

$$\max_{\lambda} |\mathbf{X}(\lambda|\nu)|^2, \tag{4.14}$$

where λ ranges over all minimal subsets in ν, i.e., the largest span among all minimal subsets within ν. Although (3.12) is theoretically more justified, the value from (4.14) is more useful for comparisons with spans from other minimal subsets.

Table 4.1 illustrates the median results over 100 repetitions for each outlier rate. Note that although only 100 inliers were generated, it is possible that a few randomly produced outliers may align closely with the underlying model, thus contributing to consensus sizes above 100. A common trend is that all methods deteriorate with the increase in outlier rates. The top-3 methods are LO-RANSAC, Multi-GS and Multi-GS-Offset. At low outlier rates, the three methods give comparable sampling accuracy, although Multi-GS and Multi-GS-Offset appear to stop later due to their more involved computations. Nonetheless, as the outlier rate increases, Multi-GS and Multi-GS-Offset become much faster than the other methods, indicating their significant accuracy in sampling minimal subsets with good model estimates (see Section 4.3.1).

The crux of this chapter, however, is the maximisation of the span of *all-inlier* minimal subsets. Table 4.1 shows that Multi-GS-Offset is the most successful in this respect, as it consistently obtains all-inlier minimal subsets with the highest span. This result also translates into better sampling accuracy and run time compared to the closest competitor (Multi-GS), which does not deliberately maximise span.

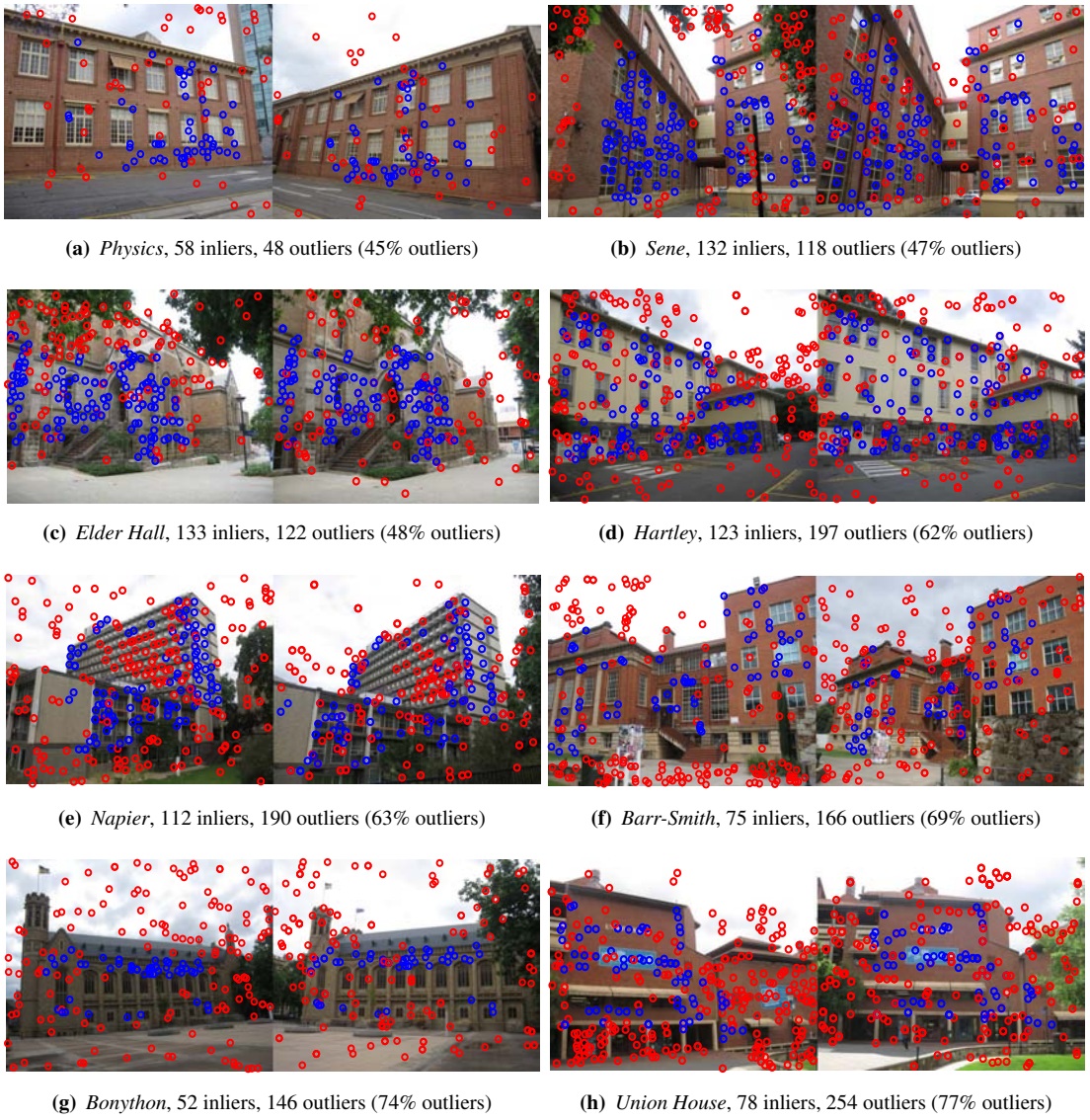
Table 4.1: Performance comparison of sampling algorithms on synthetic data. The best score on each measure is bolded.

Data		Random	Proxim	Proxim -Offset	LO-RAN SAC	Multi-GS	Multi-GS -Offset
100 inliers 50 outliers (33% outliers)	# min. subsets	268	326	248	**211**	252	232
	# all-inlier min. subsets	10	13	8	23	**101**	90
	Max. all-inlier span	1.58e0	4.93e-1	1.76e0	6.17e+1	8.01e0	**7.33e+1**
	Med. all-inlier span	1.08e-2	5.67e-4	1.16e-2	7.08e-2	3.25e-2	**1.97e-1**
	Max. consensus	93	90	94	100	99	**100**
	# of true inliers	91	88	92	98	98	**99**
	Classification error	11	14	10	4	3	**2**
	Run time	0.41	0.59	0.43	**0.28**	0.76	0.73
100 inliers 100 outliers (50% outliers)	# min. subsets	2395	2295	2319	1087	1031	**929**
	# all-inlier min. subsets	7	11	7	14	**255**	252
	Max. all-inlier span	5.68e-1	2.48e0	2.13e0	2.41e+1	3.84e+1	**6.94e+1**
	Med. all-inlier span	1.71e-3	4.53e-3	3.22e-3	2.79e-2	2.94e-2	**1.36e-1**
	Max. consensus	93	96	94	99	100	**101**
	# of true inliers	88	92	91	96	97	**99**
	Classification error	17	12	12	7	6	**3**
	Run time	4.43	5.11	5.06	**2.71**	7.38	7.24
100 inliers 150 outliers (60% outliers)	# min. subsets	11,339	13,425	10,256	5545	1056	**934**
	# all-inlier min. subsets	5	10	6	12	116	**120**
	Max. all-inlier span	5.97e-1	3.61e-1	7.27e-1	3.59e0	4.32e0	**3.89e+1**
	Med. all-inlier span	1.63e-2	5.17e-3	1.89e-2	2.64e-2	2.90e-2	**1.50e-1**
	Max. consensus	96	94	97	99	100	**102**
	# of true inliers	91	89	92	96	97	**99**
	Classification error	14	16	13	7	6	**4**
	Run time	26.95	31.37	24.46	13.09	8.14	**7.82**
100 inliers 200 outliers (67% outliers)	# min. subsets	46,622	51,694	62,870	36,527	1152	**947**
	# all-inlier min. subsets	4	9	5	16	74	**77**
	Max. all-inlier span	9.37e-2	5.61e-3	4.78e-4	5.92e0	7.40e0	**1.90e+1**
	Med. all-inlier span	4.16e-3	9.94e-4	7.24e-5	3.16e-2	5.54e-2	**2.08e-1**
	Max. consensus	95	94	92	97	98	**100**
	# of true inliers	90	89	87	93	94	**98**
	Classification error	15	16	18	11	10	**4**
	Run time	115.75	126.89	156.47	92.37	16.15	**14.53**

4.3.3 Real data

We now test on real data. Given two views of a static scene, SIFT keypoints [16] are first detected and matched across the views (using VLFeat toolbox [81]). These keypoint correspondences are manually identified as true correspondences (inliers) and false correspondences (outliers). The keypoint coordinates are also normalised by using the method of [75]. Parameter settings follow from the experiments on synthetic data, except for inlier threshold θ which is manually tuned for each dataset. The same threshold is given to all methods. SIFT matching scores are provided for PROSAC and Guided-MLESAC.

The datasets used are shown in Figure 4.8. We use a subset of the publicly available real dataset of [50], where the more difficult datasets (e.g., higher outlier rates, larger inlier noise magnitudes) are chosen (the methods perform similarly on the other easier datasets). Each method is



(a) *Physics*, 58 inliers, 48 outliers (45% outliers)

(b) *Sene*, 132 inliers, 118 outliers (47% outliers)

(c) *Elder Hall*, 133 inliers, 122 outliers (48% outliers)

(d) *Hartley*, 123 inliers, 197 outliers (62% outliers)

(e) *Napier*, 112 inliers, 190 outliers (63% outliers)

(f) *Barr-Smith*, 75 inliers, 166 outliers (69% outliers)

(g) *Bonython*, 52 inliers, 146 outliers (74% outliers)

(h) *Union House*, 78 inliers, 254 outliers (77% outliers)

Figure 4.8: Real datasets used in the experiments. Features are detected and matched across the two views using SIFT [16]. True matches (inliers) are in blue while incorrect matches (outliers) are in red (correspondence lines are not drawn for clarity).

given 100 runs on each dataset. We record the same performance measures used in the synthetic data experiments. Table 4.2 shows the median results over 100 runs. Due to space constraints we omit the results from pure random sampling.

On datasets with low outlier rates, PROSAC and Guided-MLESAC are the fastest due to the benefits of SIFT matching scores. On datasets with higher outlier rates ($> 60\%$), the substantial run time (and hence sampling accuracy) of Multi-GS and Multi-GS-Offset become apparent — on the three hardest datasets, Multi-GS and Multi-GS-Offset are almost 6 times faster than the other methods. In terms of maximising the span of all-inlier minimal subsets, it is clear that Multi-GS-Offset is the most successful in all datasets. This allows it to achieve a faster run time than Multi-GS.

Table 4.2: Performance comparison of sampling algorithms on real data. The best score on each measure is bolded.

Data		Proxim	Proxim-Offset	LO-RAN SAC	Guided-MLESAC	PROSAC	Multi-GS	Multi-GS -Offset
Physics (Figure 4.8(a))	# min. subsets	571	572	504	500	**498**	522	504
	# all-inlier min. subsets	18	11	25	14	26	**215**	188
	Max. all-inlier span	2.27e-7	2.40e-8	5.53-6	3.12e-6	1.34e-6	1.49e-5	**2.46e-5**
	Med. all-inlier span	3.37e-10	4.78e-11	5.88e-8	4.25e-8	2.90e-8	1.57e-6	**2.90e-6**
	Max. consensus	58	58	59	59	59	**60**	**60**
	# of true inliers	56	56	57	57	57	**58**	**58**
	Classification error	4	4	3	3	3	**2**	**2**
	Run time	0.89	0.91	0.70	0.70	**0.68**	1.85	1.81
Sene (Figure 4.8(b))	# min. subsets	614	666	564	562	**559**	652	615
	# all-inlier min. subsets	8	5	31	48	143	**206**	185
	Max. all-inlier span	1.51e-8	5.76e-7	1.28e-6	1.85e-6	3.75e-6	1.35e-6	**8.75e-5**
	Med. all-inlier span	2.07e-11	1.13e-10	5.32e-9	8.13e-8	9.29e-8	1.45e-7	**3.39e-6**
	Max. consensus	133	133	134	134	134	134	**135**
	# of true inliers	129	130	130	131	131	131	**132**
	Classification error	7	5	6	4	4	4	**3**
	Run time	1.67	1.77	1.43	1.42	**1.39**	5.30	5.12
Elder Hall (Figure 4.8(c))	# min. subsets	662	675	595	592	589	630	**585**
	# all-inlier min. subsets	9	4	26	44	143	**188**	172
	Max. all-inlier span	7.12e-9	6.13e-10	2.82e-7	4.22e-7	1.30e-6	4.62e-7	**4.91e-5**
	Med. all-inlier span	1.52e-11	2.40e-12	2.26e-8	3.24e-8	1.14e-7	1.19e-8	**1.37e-6**
	Max. consensus	132	131	132	132	132	132	**134**
	# of true inliers	127	126	128	131	131	131	**133**
	Classification error	11	12	9	3	3	3	**1**
	Run time	1.90	1.93	1.68	1.60	**1.56**	6.20	5.71
Hartley (Figure 4.8(d))	# min. subsets	4078	4325	3632	3850	3848	474	**446**
	# all-inlier min. subsets	5	2	15	19	66	**84**	79
	Max. all-inlier span	9.39e-9	4.53e-10	4.15e-7	3.81e-7	5.56e-7	7.36e-7	**2.47e-6**
	Med. all-inlier span	1.11e-10	2.31e-12	1.02e-8	1.13e-8	2.48e-8	4.04e-7	**1.07e-6**
	Max. consensus	127	127	128	128	128	128	**129**
	# of true inliers	118	118	120	121	121	121	**123**
	Classification error	14	14	11	9	9	9	**6**
	Run time	12.92	14.02	11.67	12.47	12.22	3.73	**3.67**
Napier (Figure 4.8(e))	# min. subsets	15,277	13,145	6491	7409	8774	724	**688**
	# all-inlier min. subsets	7	4	18	12	16	**119**	107
	Max. all-inlier span	4.62e-7	7.48e-7	5.98e-6	1.46e-6	9.71e-6	9.92e-6	**5.15e-5**
	Med. all-inlier span	8.77e-11	4.55e-10	1.75e-8	2.59e-9	6.35e-8	7.51e-8	**8.31e-7**
	Max. consensus	112	114	119	118	119	120	**121**
	# of true inliers	104	106	110	109	110	111	**112**
	Classification error	16	14	11	12	11	10	**9**

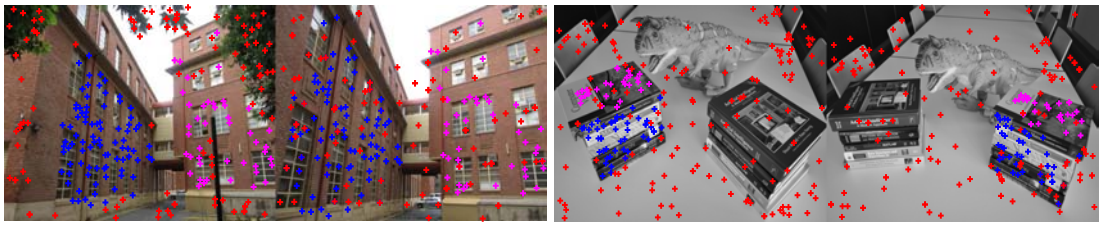
Table 4.2: continue from the previous page.

Data		Proxim	Proxim-Offset	LO-RANSAC	Guided-MLESAC	PROSAC	Multi-GS	Multi-GS-Offset
	Run time	44.39	38.54	18.35	21.19	24.26	6.45	**5.90**
Barr-Smith (Figure 4.8(f))	# min. subsets	47,453	54,462	42,178	39,423	38,578	639	**583**
	# all-inlier min. subsets	8	5	24	19	22	85	**88**
	Max. all-inlier span	2.42e-10	4.19e-11	6.09e-8	7.25e-8	8.86e-8	1.21e-7	**2.84e-7**
	Med. all-inlier span	6.22e-11	8.82e-12	5.15e-9	6.07e-9	6.89e-9	3.29e-8	**7.23e-8**
	Max. consensus	76	75	78	78	78	79	**80**
	# of true inliers	71	70	72	73	73	74	**75**
	Classification error	9	10	9	7	7	6	**5**
	Run time	80.45	91.87	70.75	66.18	64.53	4.28	**4.09**
Bonython (Figure 4.8(g))	# min. subsets	462,871	470,628	330,352	394,366	388,741	830	**713**
	# all-inlier min. subsets	7	5	33	21	35	113	**118**
	Max. all-inlier span	5.92e-15	1.12e-15	5.35e-13	7.82e-14	8.77e-14	9.35e-13	**1.62e-11**
	Med. all-inlier span	2.89e-17	1.25e-17	3.09e-15	3.37e-16	4.59e-16	4.01e-15	**4.94e-13**
	Max. consensus	47	47	49	48	48	49	**51**
	# of true inliers	44	43	46	46	46	47	**50**
	Classification error	11	13	9	8	8	7	**3**
	Run time	911.32	927.68	650.64	778.16	765.29	5.48	**5.19**
Union House (Figure 4.8(h))	# min. subsets	374,852	415,178	307,632	286,186	277,379	696	**584**
	# all-inlier min. subsets	5	4	17	26	48	73	**76**
	Max. all-inlier span	6.31e-13	2.05e-13	4.41e-12	9.22e-12	7.83e-11	1.47e-9	**4.32e-9**
	Med. all-inlier span	7.28e-15	3.13e-15	2.83e-14	6.19e-14	5.35e-13	1.39e-11	**9.61e-10**
	Max. consensus	81	80	83	84	84	85	**86**
	# of true inliers	73	72	74	75	75	76	**78**
	Classification error	13	14	13	12	12	11	**8**
	Run time	1096.52	1213.76	900.35	837.74	811.72	6.98	**5.51**

4.3.4 Performance under degeneracies

We also examine the benefits of sampling with large span for avoiding degeneracies. For the 8-point fundamental matrix estimation algorithm, a degenerate estimate is obtained when more than six keypoint matches among eight in the minimal subset lie on the same plane. This case occurs frequently when there exists a dominant plane in the scene [82, 83]. Theoretically, a degenerate minimal subset has a very small span or insufficient degrees of freedom ($|\mathbf{X}(\lambda)|$ is close to zero). It is crucial for sampling methods to avoid degenerate minimal subsets.

We follow the experimental methodology of [37]. Scenes with two large planes are chosen, specifically *Sene* in Figure 4.8 and *Dinobooks* in [37]. We detect and match SIFT keypoints which we then manually categorise into inliers and outliers. On each dataset, we keep the



(a) *Sene*, $Set\ A$: 86 inlier matches (blue), $Set\ B$: 46 inlier matches (magenta), and 118 outlier matches (red).

(b) *Dinobooks* [37], $Set\ A$: 49 inlier matches (blue), $Set\ B$: 29 inlier matches (magenta), and 155 outlier matches (red).

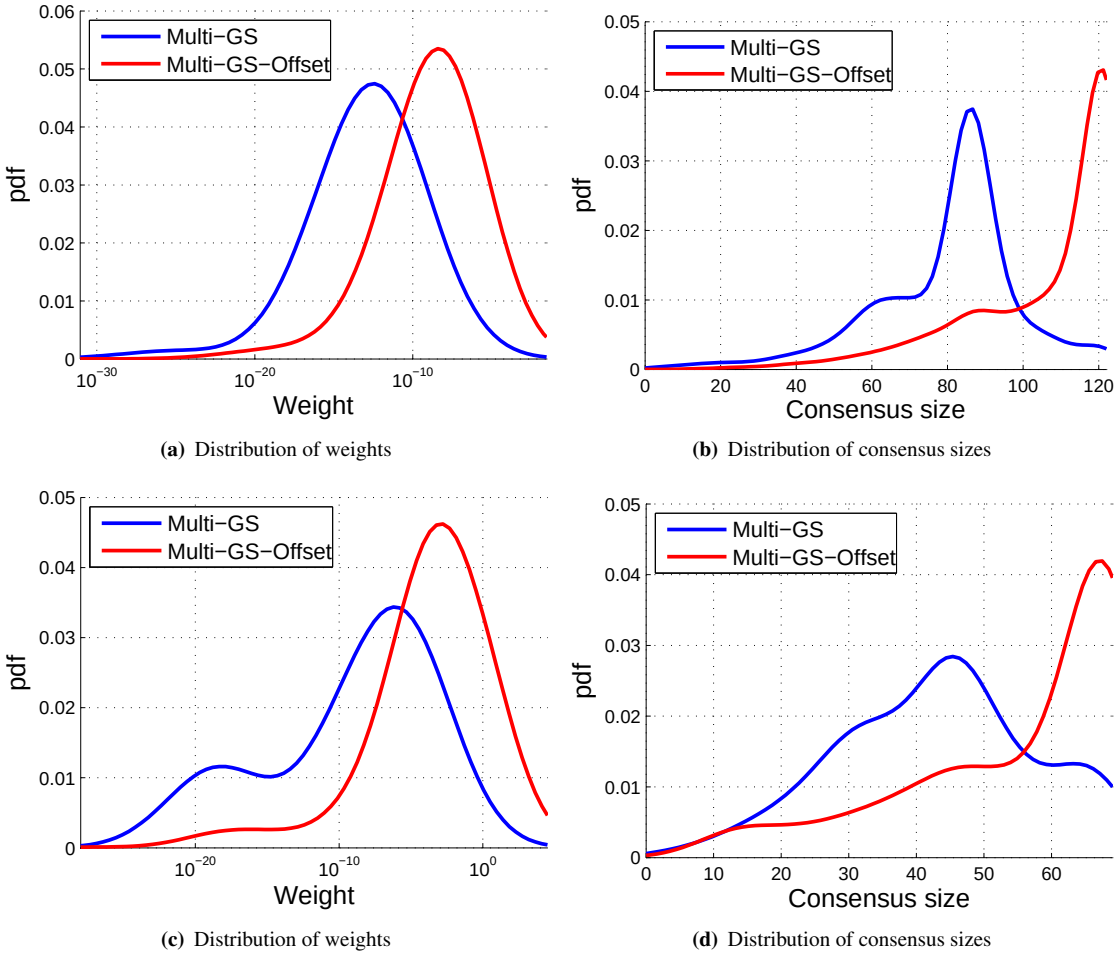
Figure 4.9: Image pairs of a static scene (thus supporting one fundamental matrix structure) where the inlier matches lie on two distinct planes.

outliers and separate the inliers into two sets ($Set\ A$ and $Set\ B$) based on the plane on which they lie; see Figure 4.9. We then create different levels of degeneracies by maintaining $Set\ A$ (the dominant plane inliers) while controlling the number of inliers in $Set\ B$ (the "off-plane" inliers). This yields the ratio between the dominant plane inliers and the total inliers as

$$\gamma = |Set\ A|/(|Set\ A| + |Set\ B|). \tag{4.15}$$

First, we focus on comparing Multi-GS and Multi-GS-Offset. The ratio γ is fixed at 0.7, and for each method, 10,000 minimal subsets are drawn from the set $Set\ A \cup Set\ B$. The distributions (obtained via kernel density estimation) of the weights $|\mathbf{X}(\lambda)|^2$ and consensus sizes of the minimal subsets are illustrated in Figure 4.10. Observe that in both datasets, the consensus size distributions of Multi-GS tend to peak at $|Set\ A|$, indicating that Multi-GS tends to fit the dominant plane — this trend has been observed in [37]. In contrast, the peak of the consensus size distributions of Multi-GS-Offset is close to $|Set\ A \cup Set\ B|$, indicating that Multi-GS-Offset is more capable of capturing all of the inliers in the scene. This result represents a clear evidence of the ability of Multi-GS-Offset to avoid degeneracies by sampling minimal subsets with large span.

Next, we analyse with different levels of degeneracies. The ratio γ is varied within $[0.7, 0.9]$, For each γ, 100 instances of the data are generated; here, besides $Set\ A$ and $Set\ B$ we also include all the outliers. Each sampling method in Section 4.3 is invoked using the common RANSAC stopping criterion. We record the number of all-inlier minimal subsets, and the number of *non-degenerate* all-inlier minimal subsets achieved — the latter is calculated by checking the plane on which each inlier lies. Figure 4.11 shows the median results of Multi-GS and Multi-GS-Offset, with Random also included as the baseline (the performances of the other methods are significantly worse than Multi-GS and Multi-GS-Offset, so we did not plot them in the figure). Expectedly all methods deteriorate with the increase in γ, since this entails the increase in outlier rates. Although obtaining fewer all-inlier minimal subsets than Multi-GS, Multi-GS-Offset achieves more non-degenerate all-inlier minimal subsets; see the dashed lines in Figures 4.11(a)



(a) Distribution of weights

(b) Distribution of consensus sizes

(c) Distribution of weights

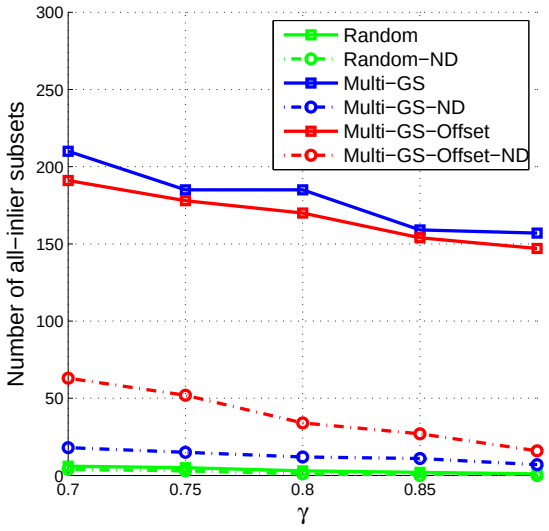
(d) Distribution of consensus sizes

Figure 4.10: Performance comparison of Multi-GS and Multi-GS-Offset on degenerate configurations with $\gamma = 0.7$. Row 1: *Sene*, Row 2: *Dinobooks*. Note that in (a) and (c), the horizontal (weight) axis is in logarithmic scale.

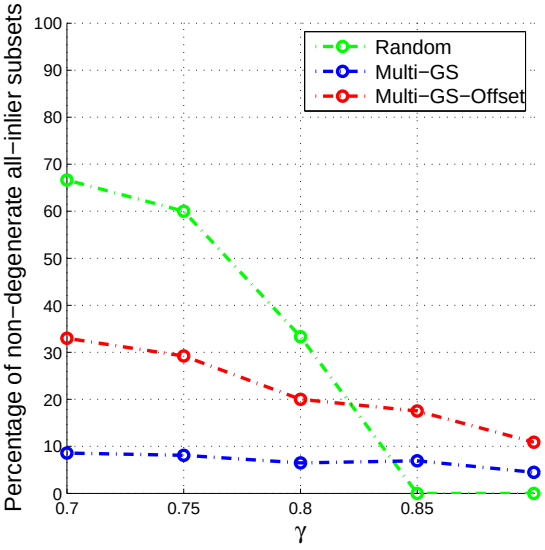
and 4.11(c). This naturally leads to a higher percentage of non-degenerate all-inlier samples over all-inlier samples produced; see Figure 4.11(b) and 4.11(d). Note that since Random hits very few all-inlier minimal subsets, its percentage of non-degenerate all-inlier minimal subsets is sometimes high by chance.

4.4 Summary

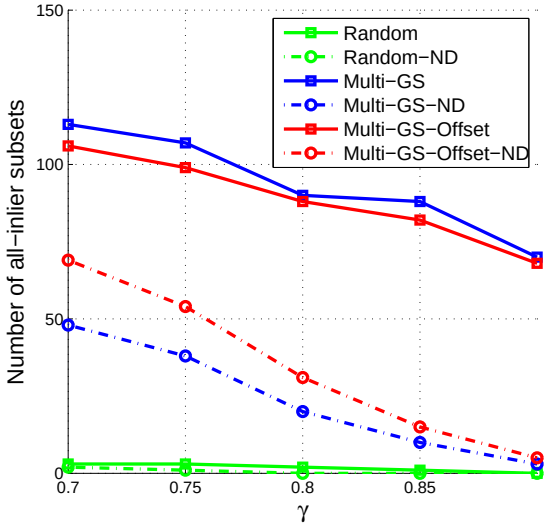
The previous chapter presented a theoretical basis for large span sampling in robust geometric fitting, and suggested that sampling all-inlier minimal subsets with large span is important to yielding a satisfactory model estimate. Our main contribution in this chapter is to develop a novel guided sampling algorithm that can retrieve all-inlier minimal subsets with large span for robust geometric fitting. We showed that simple distance-based sampling is not effective for seeking all-inlier samples with large span. In contrast, our proposed algorithm consciously



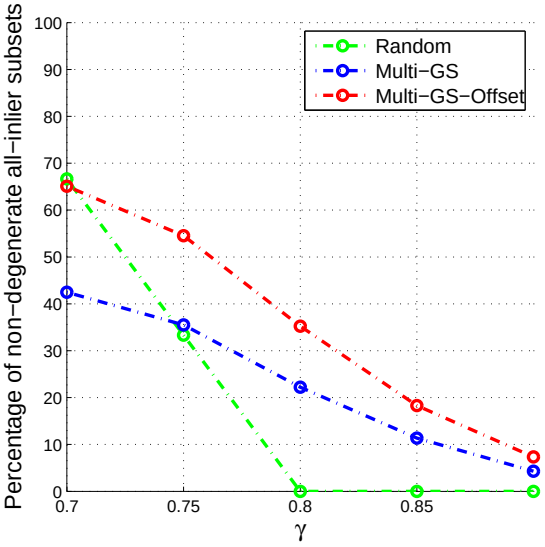
(a) Number of all-inlier minimal subsets (solid lines) and non-degenerate all-inlier minimal subsets (dashed lines).



(b) Percentage of non-degenerate all-inlier minimal subsets.



(c) Number of all-inlier minimal subsets (solid lines) and non-degenerate all-inlier minimal subsets (dashed lines).



(d) Percentage of non-degenerate all-inlier minimal subsets.

Figure 4.11: Performance comparison of sampling algorithms on degenerate configurations with $\gamma \in [0.7, 0.9]$. Row 1: *Sene*. Row 2: *Dinobooks*.

targets minimal subsets with large span without a decreased accuracy in finding all-inlier samples. This also permits an ability to avoid sampling degenerate minimal subsets. The significant performance of our proposed method is demonstrated on synthetic and real datasets.

Chapter 5

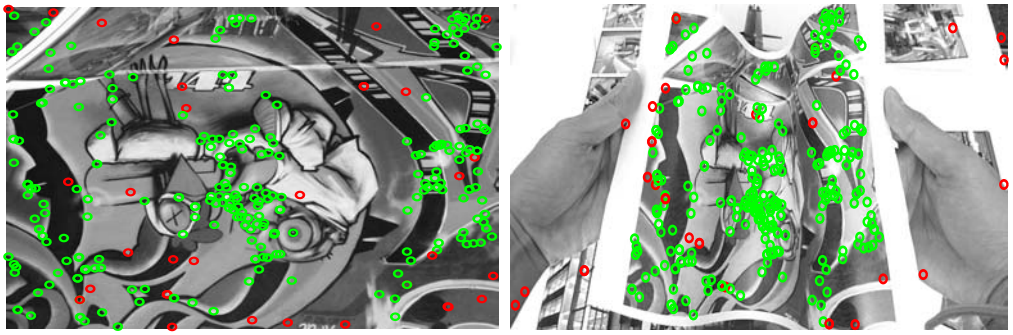
Outlier Rejection in Deformable Registration with RANSAC

5.1 Introduction

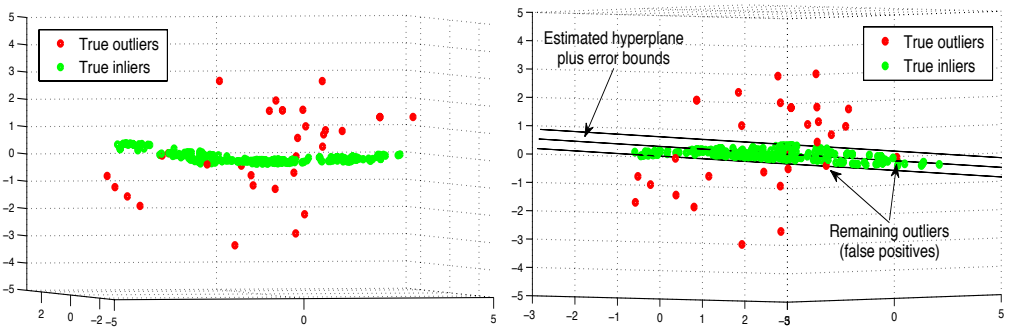
So far we have discussed various methods for robust geometric fitting, where the geometric model of interest has a *fixed* number of degrees of freedom; and sampling and testing model hypotheses from minimal subsets is often employed in dealing with outliers. The last two chapters provided a theoretical basis and algorithmic solution to the problem of maximising the span of all-inlier minimal subsets for robust geometric fitting.

In this chapter, we depart from robust geometric fitting and focus on robust deformable registration. Similar to robust geometric fitting, an important part of robust deformable registration is to tackle outliers. If no outliers exist, estimating the deformable model of interest is trivial, e.g., by solving a linear system for a TPS warp [24]. As illustrated in Figure 1.2, common sense suggests that the correspondence manifold is of unknown and varying complexity, and hence the size of a minimal subset can not be determined. This assumption precludes the application of standard robust geometric fitting techniques for outlier rejection in deformable registration. As we can see from our survey in Chapter 2, most of the current approaches that are capable of outlier rejection in deformable registration [27, 29, 30, 38–42] represent the correspondence manifold by using fully nonlinear models, and employ iterative optimisation procedures whose success depends critically on good parameter initialisations.

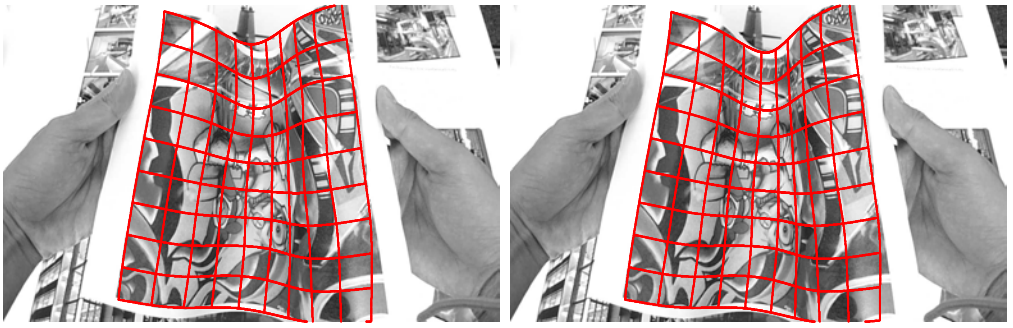
This chapter explores the use of standard robust geometric fitting techniques such as RANSAC for a *simple outlier rejection method* in deformable registration. We introduce a surprising view that, in practice, the scale of errors of the mismatches (outliers) is orders of magnitude larger than the effects of the curvature of the correspondence manifold containing the correct



(a) Template **(b)** Input image



(c) Keypoint correspondences from (a) and (b) plotted using the first-3 principal components. **(d)** Another view of (c) with the fitted hyperplane shown in its "side view".



(e) TPS estimated using keypoint matches lying within the hyperplane bounds in (d). **(f)** TPS estimated using the true correspondences identified manually.

Figure 5.1: Robust deformable registration using RANSAC. In (a) and (b), true matches (inliers) are in green while incorrect matches (outliers) are in red (correspondence lines are not drawn for clarity).

matches (inliers). Figure 5.1 illustrates what we mean, with images showing a bending sheet of paper — this kind of data is typically used in the literature, e.g., see [29, 30, 42]. SIFT [16] is first invoked on the template (Figure 5.1(a)) and input image (Figure 5.1(b)) to yield a set of keypoint correspondences $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^{n}$ with $\mathbf{x}_i = [x_i\ y_i\ x_i'\ y_i']^T \in \mathbb{R}^4$, i.e., point $[x_i\ y_i]^T$ in the template is matched with point $[x_i'\ y_i']^T$ in the input image. Projecting the data onto the first-3 principal components reveals that the inliers are actually distributed compactly on a 2D affine hyperplane, *relative* to the gross errors of the outliers; see Figures 5.1(c) and 5.1(d). This means that we can robustly fit a linear hyperplane onto the data to dichotomise the inliers and outliers; Figure 5.1(e) shows the TPS warp estimated using the keypoint matches returned by

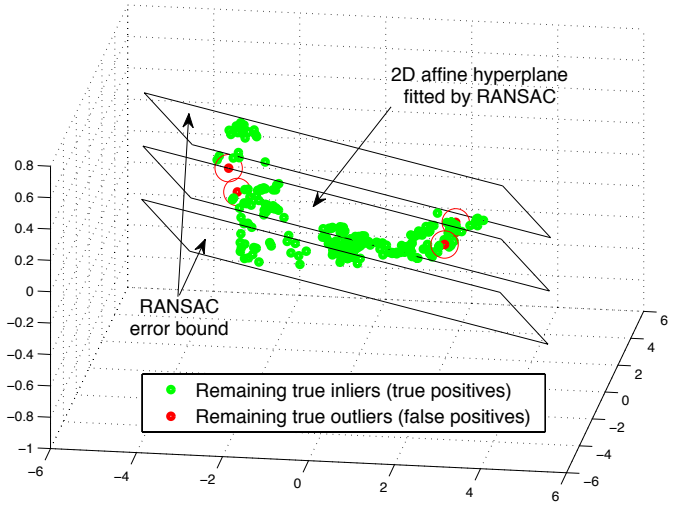


Figure 5.2: Data remaining after RANSAC, shown in the "local" scale of the manifold.

RANSAC, which models the underlying deformations quite well. As we show later, this is the characteristic of many realistic physical deformations examined in the literature.

Our observation motivates the conclusion that, for many types of deformations, a linear hyperplane is adequate to model the correspondence manifold for outlier rejection in deformable registration. After the outliers are removed, a fully nonlinear warp (e.g., a TPS warp) can be fitted easily on the remaining inliers to achieve the registration result. Most outliers that remain after outlier rejection (i.e., false positives) are relatively benign rather than outright mismatches, and can usually be smoothened out by the warp's regulariser; see Figure 5.2. Observe that the TPS in Figure 5.1(e) is very similar to the "ground truth" TPS in Figure 5.1(f) estimated using only the true inliers[1].

In a sense our observation is not surprising, since PROSAC has been used as preprocessing to remove egregious mismatches and provide affine initialisations for warp estimation in [30] (although it was not used in [29, 38], there are few obstacles to initialise with PROSAC there); see Section 2.3.2.1. However, it has always been assumed that due to the unknown complexity of the inlier distribution, significant outliers will remain and it is vital to further optimise the warp robustly, e.g., by a deterministic annealing procedure. Our aim is to show that such procedures overestimate the difficulty of the data, and basic RANSAC followed directly by (non-robust) warp estimation is sufficient.

The main content of this chapter is based on the author's publication in [86]. The rest of this chapter is organised as follows: Section 5.2 describes our observation of the approximately linear correspondence manifold. We then present in Section 5.3 our RANSAC-based approach to outlier rejection in deformable registration. Section 5.4 evaluates the performance of our

[1] It is worth noting that without further pixel-based refinement [42], warps estimated from keypoint matches alone can not extrapolate well to correspondence-poor or occluded regions; see the bottom right corner of Figure 5.1(f).

method on synthetic and real datasets, and investigates its application in retexturing deformable surfaces. Finally, we draw conclusions in Section 5.5.

5.2 The correspondence manifold

Radial basis function (RBF) warps have been applied extensively to model the deformations of various kinds of deformable objects [24]. For deformations of 2D image features, it is common to use *two* separate RBF components f_x, f_y that share the same centres $\{\mathbf{c}_k\}_{k=1}^K$ as

$$\begin{aligned} f_x : [x\ y]^T \mapsto x' &= \mathbf{a}_1^T [x\ y\ 1]^T + \mathbf{w}_1^T \boldsymbol{l}(x,y) \\ f_y : [x\ y]^T \mapsto y' &= \mathbf{a}_2^T [x\ y\ 1]^T + \mathbf{w}_2^T \boldsymbol{l}(x,y) \end{aligned}, \tag{5.1}$$

where $\boldsymbol{l}(x,y)$ is a lifting function containing the centres

$$\boldsymbol{l}(x,y) = \begin{bmatrix} \phi(\|[x\ y]^T - \mathbf{c}_1\|) & \phi(\|[x\ y]^T - \mathbf{c}_2\|) & \dots & \phi(\|[x\ y]^T - \mathbf{c}_K\|) \end{bmatrix}^T, \tag{5.2}$$

and $\phi(\cdot)$ is the RBF, e.g., Gaussian or TPS [24]. Given a set of keypoint matches $\mathcal{X} = \{(\mathbf{u}_i, \mathbf{v}_i)\}_{i=1}^n$, the centres are taken as $\{\mathbf{u}_i\}_{i=1}^n$. Learning an RBF warp (5.1) involves estimating the affine parameter vectors $\mathbf{a}_1, \mathbf{a}_2$ and the coefficient vectors $\mathbf{w}_1, \mathbf{w}_2$ with regularisation to control the warp's bending energies ν_x, ν_y, where ν_x, ν_y denote the bending energies of f_x, f_y respectively. For a TPS warp, the solution can be achieved by solving a linear system [24].

By regarding each keypoint correspondence as a single point $\mathbf{x}_i = [x_i\ y_i\ x_i'\ y_i']^T \in \mathbb{R}^4$, it can be shown that $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^n$ are samples from a smooth manifold [40]. Furthermore, the manifold is 2D due to the two degrees of freedom of x_i and y_i. Assuming that the underlying warp is a TPS warp, we can express each point on the manifold as

$$\mathbf{x}_i = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \mathbf{a}_{11} & \mathbf{a}_{12} \\ \mathbf{a}_{21} & \mathbf{a}_{22} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \mathbf{a}_{13} \\ \mathbf{a}_{23} \end{bmatrix}}_{\text{2D affine subspace}} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \mathbf{w}_1^T \boldsymbol{l}(x_i, y_i) \\ \mathbf{w}_2^T \boldsymbol{l}(x_i, y_i) \end{bmatrix}}_{\text{Nonlinear deviation}}, \tag{5.3}$$

where $\mathbf{a}_{pq}$ is the q-th component of the p-th affine parameter vector. In other words, the correspondence manifold "undulates" around a 2D affine subspace, and the deviation of each $\mathbf{x}_i$ from the subspace is due to the data-dependent nonlinear terms $\mathbf{w}_p^T \boldsymbol{l}(x_i, y_i)$.

Since outliers invariably appear in $\mathcal{X}$, the task is thus to robustly estimate the mapping function. Let $\mathcal{X}$ be divided into an inlier set $\mathcal{I}$ and outlier set $\mathcal{O}$. Our premise is that the deviations of $\mathcal{I}$ from the affine component of the correspondence manifold are quite small compared to the

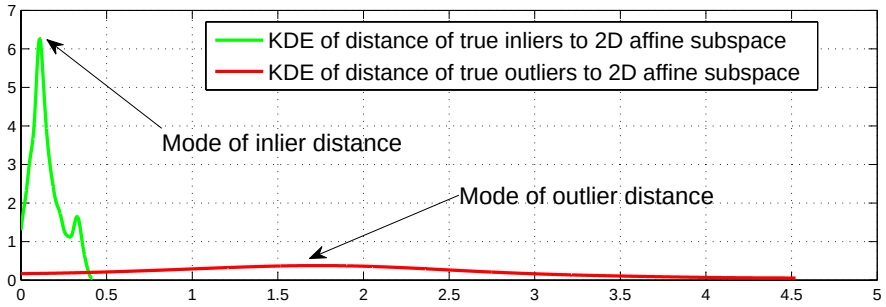


Figure 5.3: Distribution of distances of the keypoint matches in Figures 5.1(a) and 5.1(b) to the RANSAC-fitted 2D affine hyperplane in Figure 5.1(d).

distances of O. To illustrate this point, Figure 5.3 plots the distribution of orthogonal distances of the keypoint matches in Figures 5.1(a) and 5.1(b) to the RANSAC-fitted 2D affine hyperplane in Figure 5.1(d). It is apparent that a clear separation exists between the inlier and outlier distribution.

5.3 Outlier rejection as robust hyperplane fitting

Our observations suggest that RANSAC (Section 2.2) is sufficient for outlier rejection in deformable registration. The goal of our RANSAC-based approach is to robustly fit a 2D affine subspace onto the data $\mathcal{X}$. An estimate of the affine subspace can be derived from a minimal subset of three data points randomly sampled from $\mathcal{X}$ (recall that each datum $\mathbf{x}_i = [x_i\ y_i\ x_i'\ y_i']^T \in \mathcal{X}$ is a particular keypoint correspondence). Let $\mathbf{S} = [\mathbf{x}_{s_1}\ \mathbf{x}_{s_2}\ \mathbf{x}_{s_3}] \in \mathbb{R}^{4\times 3}$ be a randomly drawn minimal subset with the data concatenated horizontally. First, the mean of the sample $\boldsymbol{\mu}_\mathbf{S}$ is subtracted from each column to yield $\hat{\mathbf{S}}$, whose the first-2 left singular vectors $\mathbf{A}_{\hat{\mathbf{S}}} \in \mathbb{R}^{4\times 2}$ are then obtained. The pair $(\boldsymbol{\mu}_\mathbf{S}, \mathbf{A}_{\hat{\mathbf{S}}})$ is sufficient to characterise the affine subspace. The residual of each datum $\mathbf{x}_i$ is defined as its orthogonal distance to the fitted subspace

$$d(\mathbf{x}_i|\boldsymbol{\mu}_\mathbf{S}, \mathbf{A}_{\hat{\mathbf{S}}}) = \left\| \mathbf{x}_i - \left[\mathbf{A}_{\hat{\mathbf{S}}} \mathbf{A}_{\hat{\mathbf{S}}}^T (\mathbf{x}_i - \boldsymbol{\mu}_\mathbf{S}) + \boldsymbol{\mu}_\mathbf{S} \right] \right\|. \tag{5.4}$$

Following Section 2.2, RANSAC iteratively generates a set of N 2D affine subspace hypotheses, where each hypothesis is fitted on a randomly drawn minimal subset of data. The consensus of a hypothesis is the number of data with residuals less than a threshold θ from the associated affine subspace, and the hypothesis with the maximum consensus is returned for outlier rejection. The inliers of the best hypothesis are then used to estimate the TPS warp for aligning the images.

A crucial parameter in RANSAC is the threshold θ. Firstly, to allow the use of a constant θ for all datasets, we normalise the data such that the centroid of $\{\mathbf{u}_i\}_{i=1}^n$ lies at the origin and the mean distance of all points to the origin is $\sqrt{2}$ (the same normalisation is applied on the points $\{\mathbf{v}_i\}_{i=1}^n$) [75]. The threshold parameter is then manually tuned and used for input images. Note

that an equivalent threshold on the error is also required in the other methods [27, 29, 30, 38–42]; see Section 2.3.2.

A second important parameter is the number of samples N. Following the common stopping criterion of RANSAC, N is estimated by using Equation (2.1) with the minimal subset size $m = 3$. For instance, for the probability of success $\alpha = 0.99$ and the outlier ratio $\epsilon = 0.5$ (50% outliers), the predicted number of samples N is approximately 35. In practice, the required number of samples is much larger than the predicted N. In our experiments, we consistently set $N = 100$ for all datasets; as we show in Section 5.4 our method is still at least as accurate as current approaches.

The threshold θ can be estimated automatically by using scale estimation techniques or the number of samples N can be reduced by using guided sampling methods. However, these options are considered as future directions since our intent is not to complicate the message of this chapter (i.e., a simple outlier rejection method for deformable registration). We will discuss these directions of future work in Section 7.2.

5.4 Experimental results

5.4.1 Synthetic data

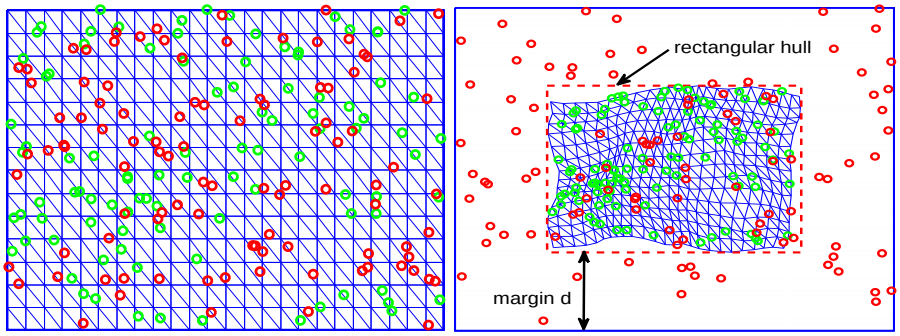
We first test the performance of RANSAC (Section 5.3) on synthetic data. In the following, we describe the procedure for generating synthetic data, and then benchmark RANSAC against current approaches for outlier rejection in deformable registration.

5.4.1.1 Data generation

A rectangular mesh is created with control points (TPS centres) distributed on a grid. Using the control points, a TPS warp is randomly generated following the method proposed in [87]. Inliers are produced by randomly sampling 100 positions on the template mesh and mapped using the synthesised warp. The mapped points are then perturbed with additive Gaussian noise of standard deviation of 5 pixels. We cover the warped mesh with the *smallest rectangular hull*, which is then surrounded by a margin of d pixels to construct the input "image". This yields the ratio between the areas of the added margin and the entire input "image" as

$$r_d = (A_I - A_R)/A_I, \tag{5.5}$$

where A_R and A_I are respectively the areas of the rectangular hull containing the warped mesh and the input "image". We then randomly sample positions on the template and input "image"



(a) Template mesh **(b)** Input "image"

Figure 5.4: Example of a synthetically generated TPS warp.

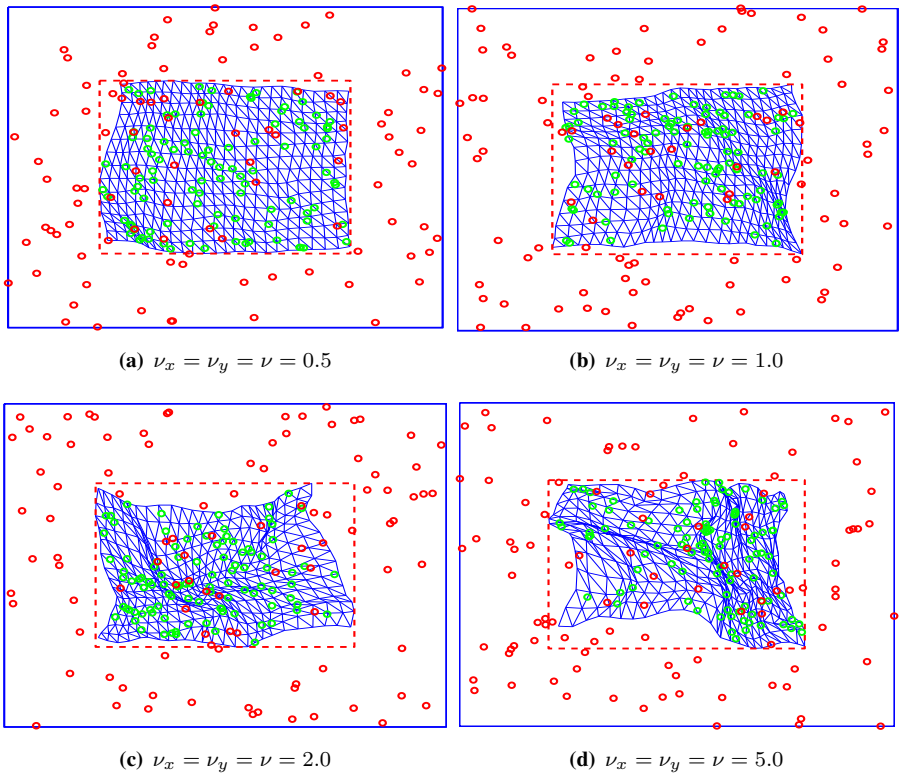
to form outliers. Figure 5.4 shows an example of data generated in this manner, with $r_d = 0.68$ ($d = 150$ pixels) and $\epsilon = 0.5$ (50% outliers).

Parameter ν in the random warp generator controls the warp's bending energies (see [87] for details). We use the same bending energies for *both* components f_x, f_y of the synthesised warp, i.e., $\nu_x = \nu_y = \nu$. The effects of different values of ν are illustrated in Figure 5.5 (in all these images, $r_d = 0.68$ and $\epsilon = 0.5$). Observe that for $\nu \geq 2.0$ the mesh is deformed seriously with self-occlusions. Note that this does not affect our experiments, since our synthetic keypoints will not be occluded or "unmatched". Our purpose is to highlight that the deformations synthesised are visually highly nonlinear.

5.4.1.2 Comparison of methods

We benchmark RANSAC (Section 5.3) against state-of-the-art outlier rejection methods for deformable registration: Iterative local smoothness test [42, Section 3] and SVM regression with resampling [27]. We also compare against the class of Annealed M-estimation methods [29, 30, 38, 39]; since these methods are comparable in accuracy, it is sufficient to compare against [30] which offers the most efficient algorithm. Note that [29, 30, 38, 39] can simultaneously optimise the warp and identify outliers; here we concentrate on the aspect of outlier rejection/identification; see Section 2.3.2.

The ROC curve of each method is obtained by varying the threshold parameter and recording the resultant true positive rate (number of true inliers recovered over all true inliers) and false positive rate (number of true outliers misidentified as inliers over all true outliers). We set $\nu = 0.5, 1.0, 2.0$ and 5.0, and for each ν, the outlier rate ϵ is set as 0.33 and 0.5 (in all these cases, the margin ratio $r_d = 0.68$). For each combination of ν and ϵ, 100 random (and distinct) TPS warps are generated, and the ROC curves for each method are averaged over the 100 warps. Figure 5.6 presents the results.



(a) $\nu_x = \nu_y = \nu = 0.5$ **(b)** $\nu_x = \nu_y = \nu = 1.0$

(c) $\nu_x = \nu_y = \nu = 2.0$ **(d)** $\nu_x = \nu_y = \nu = 5.0$

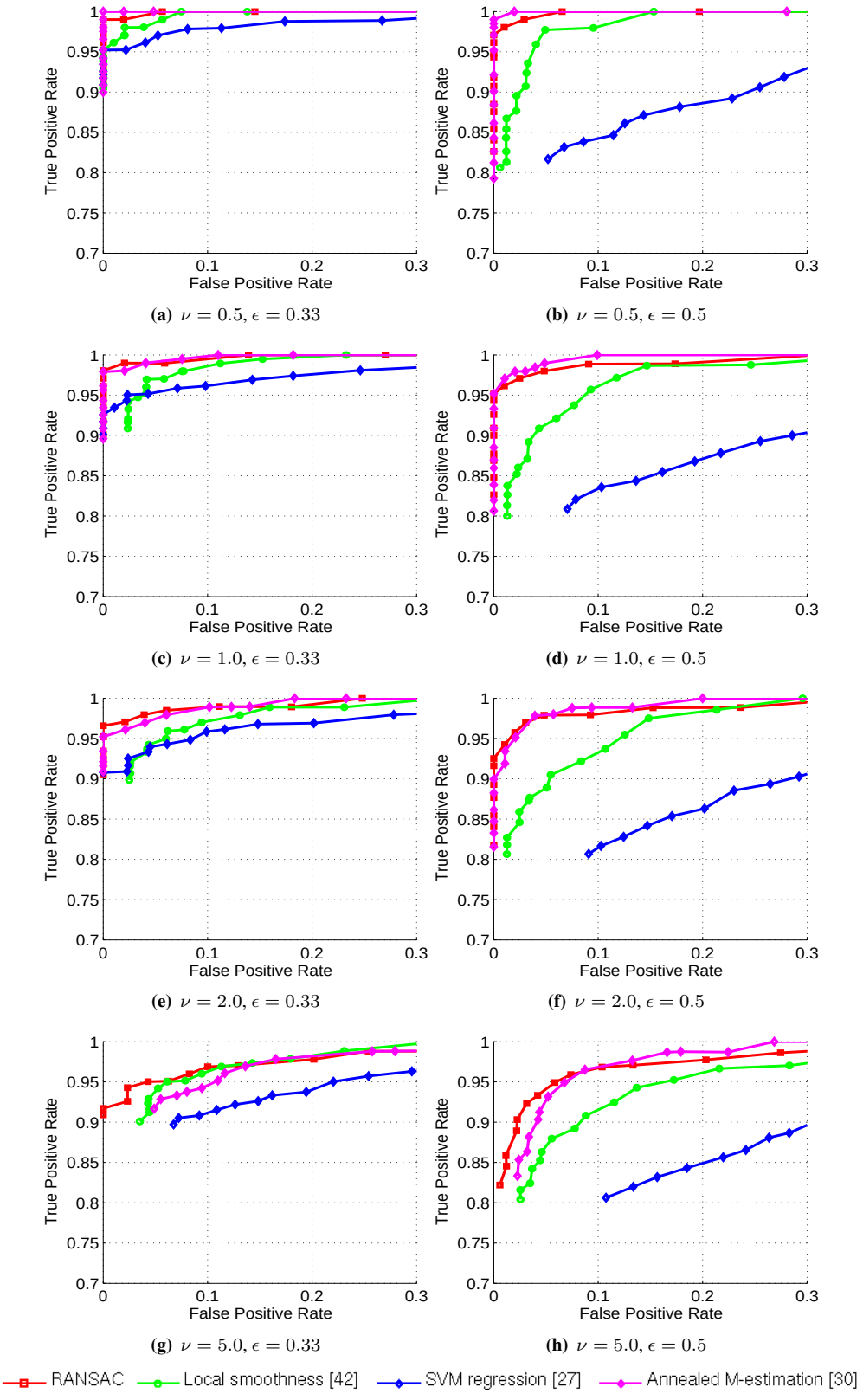
Figure 5.5: Examples of warped meshes using randomly generated TPS warps with increasing bending energies.

	$\epsilon = 0.33$ (total 150 matches)	$\epsilon = 0.5$ (total 200 matches)
RANSAC	0.04	0.04
Local smoothness [42]	0.26	0.29
SVM regression [27]	0.06	0.09
Annealed M-estimation [30]	1.15	0.29

Table 5.1: Average run time (in seconds) for outlier rejection on synthetic data.

An apparent and expected trend is that as ν and ϵ increase, the accuracies of all methods decrease, with the method of [27] deteriorating the fastest, followed by [42]. The other two methods, Annealed M-estimation [30] and RANSAC, provide very comparable accuracies. The strength of RANSAC, however, lies in its simplicity and efficiency. Table 5.1 presents the average running time of all methods for $\epsilon = 0.33$ and 0.5, where RANSAC is clearly the fastest — all methods were implemented and run in Matlab, while the subroutines *svmtrain*, *svmpredict* in [27] and *selectentries* in [30] were implemented in C++[2]. The major factors affecting the speed of RANSAC are the outlier rate ϵ and the minimal subset size m — since only three data are required for a minimal solution, RANSAC can tolerate large ϵ's without significant sampling efforts. On the other hand, the algorithms of [27, 30, 42] are more complicated and their run times scale faster with the data size.

[2]Code of all methods can be obtained from `http://cs.adelaide.edu.au/~huy/`



(a) $\nu = 0.5, \epsilon = 0.33$ **(b)** $\nu = 0.5, \epsilon = 0.5$

(c) $\nu = 1.0, \epsilon = 0.33$ **(d)** $\nu = 1.0, \epsilon = 0.5$

(e) $\nu = 2.0, \epsilon = 0.33$ **(f)** $\nu = 2.0, \epsilon = 0.5$

(g) $\nu = 5.0, \epsilon = 0.33$ **(h)** $\nu = 5.0, \epsilon = 0.5$

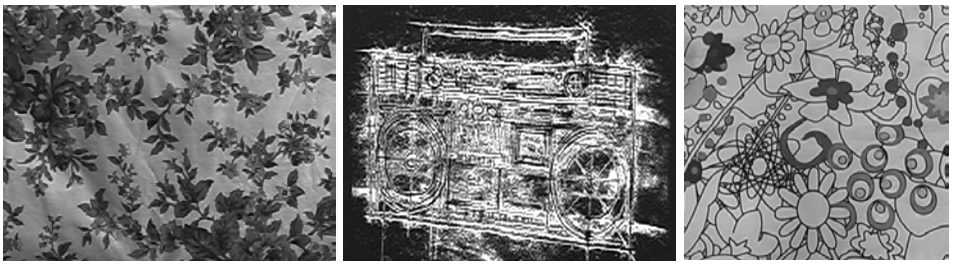
Figure 5.6: Outlier rejection results on synthetic data.

Sequence name	*Bedsheet*			*Tshirt*			*Cushion*		
Frame number	140	160	178	407	720	784	160	175	190
RANSAC	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Local smoothness [42]	0.26	0.29	0.28	0.21	0.17	0.19	0.28	0.26	0.22
SVM regression [27]	0.06	0.06	0.08	0.06	0.04	0.05	0.12	0.10	0.06
Annealed M-estimation [30]	0.13	0.15	0.33	0.17	0.14	0.13	0.31	0.24	0.17

Table 5.2: Average run time (in seconds) for outlier rejection on real data.

5.4.2 Flat surface deformations

We now test our method (Section 5.3) on real images. We used the publicly available image sequences previously used for nonrigid structure from motion (e.g., see [64]). In this experiment we chose three representative frames from the three hardest sequences (*Bedsheet*, *Tshirt* and *Cushion*) as input images for outlier rejection. A subimage which encapsulates a large portion of the surface was cropped from the first image of each sequence to form the template image; see Figure 5.7. SIFT [16] was first invoked on the images to produce keypoint correspondences, which we then manually categorised as true inliers and outliers. For RANSAC, 100 repetitions were performed on each input image and the average results (ROC curves) are reported. Figures 5.8, 5.9 and 5.10 illustrate the results.



(a) *Bedsheet* **(b)** *Tshirt* **(c)** *Cushion*

Figure 5.7: Templates of the *Bedsheet*, *Tshirt* and *Cushion* sequences.

The low-dimensional visualisations of all data show that again, relative to the outliers, the inliers are distributed compactly within a 2D affine hyperplane. Based on the ROC curves, a similar conclusion can be made on the accuracy of outlier rejection, i.e., Annealed M-estimation [30] and RANSAC are the most accurate, followed by Iterative local smoothness test [42][3] and SVM regression with resampling [27]. The run times of all methods are depicted in Table 5.2. Again, RANSAC is the fastest method, with almost constant run times across all images.

The data in which the gap in accuracy between Annealed M-estimation [30] and RANSAC is the largest is Frame 190 of *Cushion* (Figure 5.10). In the next section we investigate the practical difference due to this disparity in accuracy. We provide outlier rejection results on all frames of the sequences (and on other sequences) at `http://cs.adelaide.edu.au/~huy/`.

[3]The non-monotonicity of the ROC curve of [42] is most likely due to using an iterative method. Similar curves were observed in [42].

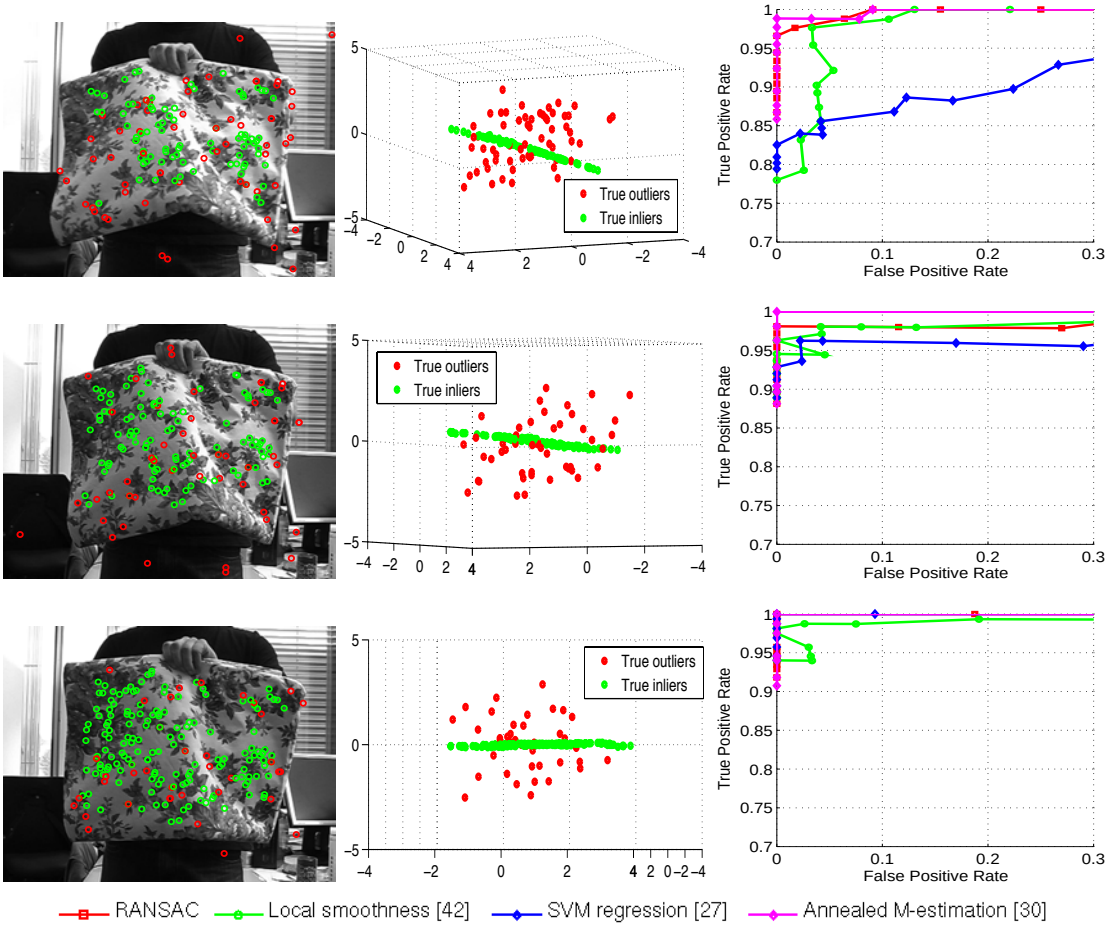


Figure 5.8: Outlier rejection results on Frame 140 (145 matches, 41.38% outliers), Frame 160 (152 matches, 31.58% outliers) and Frame 178 (196 matches, 19.90% outliers) from the *Bedsheet* sequence. Column 1: SIFT matches. Column 2: Data after PCA. Column 3: ROC curves.

5.4.3 Retexturing deformable surfaces

We also investigate the application of our method in retexturing deformable surfaces. Figures 5.11, 5.12 and 5.13 provide the comparison results of the two best performing methods in the previous section. The warps for the meshes (for images used in Section 5.4.2) are obtained by first using RANSAC and Annealed M-estimation [30] to remove outliers, and then using the remaining matches to estimate a TPS warp. The ground truth warp is obtained by estimating a TPS warp using only true inliers. The threshold value for RANSAC and Annealed M-estimation is optimised using the ROC curves in Section 5.4.2. Note that Annealed M-estimation can jointly identify outliers and estimate warps, however to yield comparable parameters (a different kind of warp and bending energy are used in [30] — see Section 2.3.2.1) we simply estimate a TPS warp using the inliers returned.

Both methods yield very close results to the ground truth, including Frame 190 of *Cushion* in which the disparity in outlier rejection accuracy between RANSAC and Annealed M-estimation

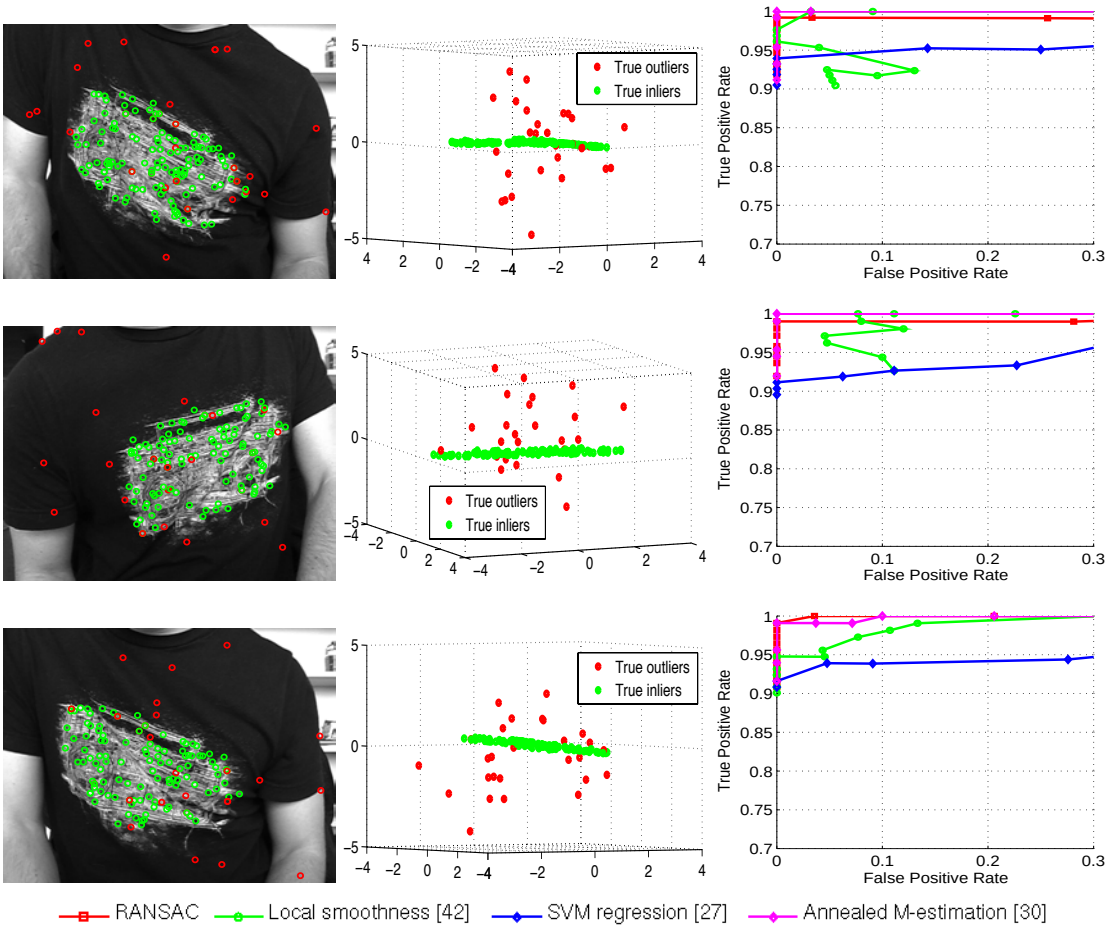


Figure 5.9: Outlier rejection results on Frame 407 (154 matches, 19.48% outliers), Frame 720 (127 matches, 18.90% outliers) and Frame 784 (136 matches, 19.85% outliers) from the *Tshirt* sequence. Column 1: SIFT matches. Column 2: Data after PCA. Column 3: ROC curves.

is the largest (see Row 3 of Figure 5.10). As mentioned in Section 5.1, false positives produced by RANSAC are normally benign outliers which can be smoothened out by the warp's regulariser.

We further compute the goodness of each estimated warp as the number of vertices in the warped mesh which are within 3 pixels away from the corresponding vertices in the ground truth mesh. The results in Table 5.3 show that on several images Annealed M-estimation is better than RANSAC in this measure — however, [30] imposes local smoothness constraints which help to "pin down" the position of each vertex relative to the others and this is beneficial for the goodness measure. This additional information is not provided to RANSAC. In any case, as shown in Figures 5.11, 5.12 and 5.13, the practical differences between the two methods are minuscule.

A general problem for feature-based methods however is the lack of keypoint correspondences in certain areas of the surface. To deal with this issue, we track and propagate features in an image sequence. First, the template is divided into rectangular regions (e.g., 5×5 grid). If the

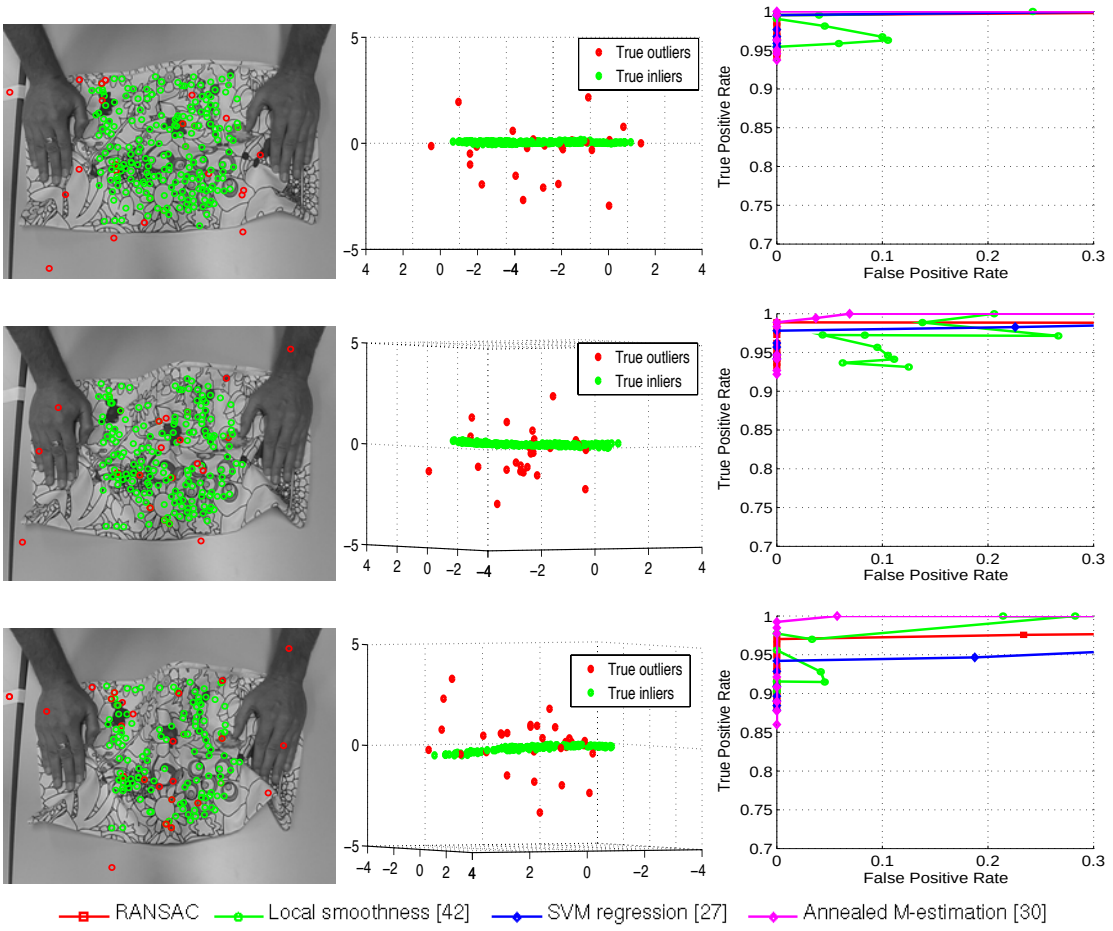
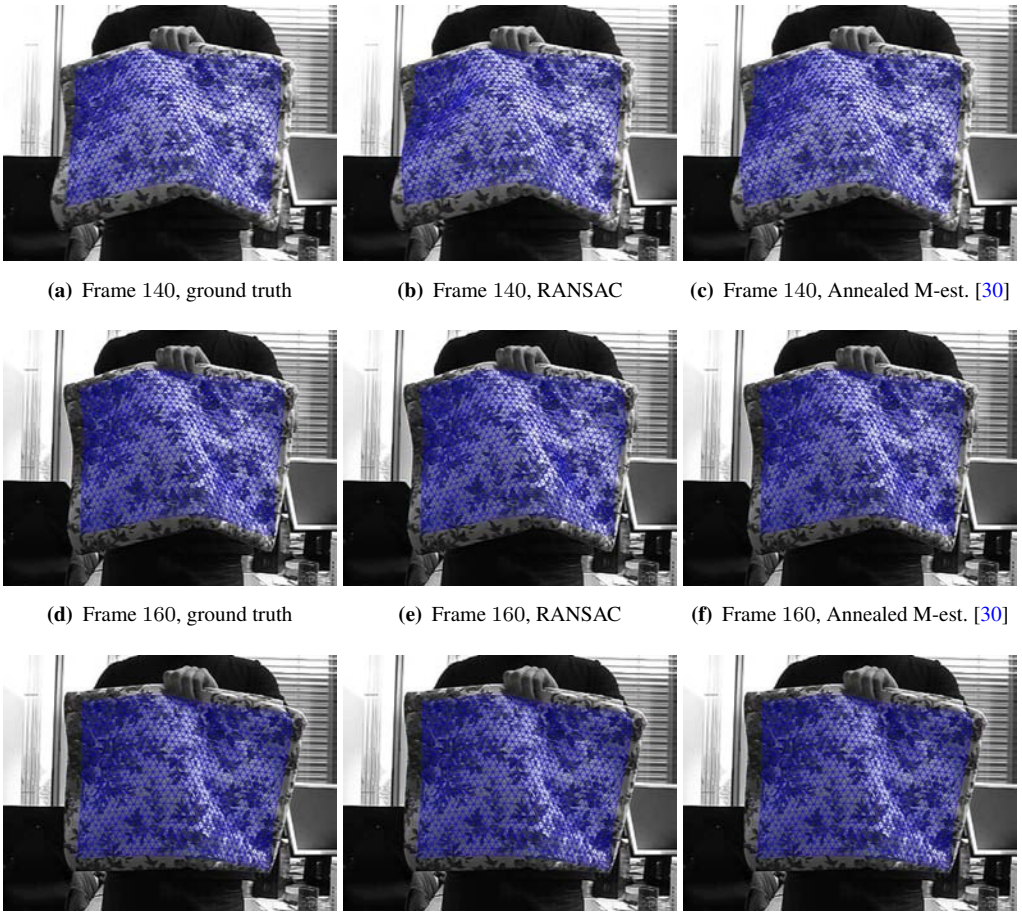


Figure 5.10: Outlier rejection results on Frame 160 (234 matches, 10.68% outliers), Frame 175 (205 matches, 13.17% outliers) and Frame 190 (163 matches, 20.25% outliers) from the *Cushion* sequence. Column 1: SIFT matches. Column 2: Data after PCA. Column 3: ROC curves.

Sequence name	*Bedsheet*			*Tshirt*			*Cushion*		
Frame number	140	160	178	407	720	784	160	175	190
RANSAC	603	728	810	667	660	653	667	666	649
Local smoothness [42]	400	518	675	475	339	217	667	552	645
SVM regression [27]	17	146	702	294	138	220	659	564	473
Annealed M-estimation [30]	648	810	810	667	667	663	667	666	667

Table 5.3: Quantitative retexturing results on real data.

number of keypoint matches in a region between the current frame and the template falls below a threshold, Mean Shift [88, 89] is initiated to track (pre-matched) features from the previous frame. All keypoint matches are then vetted by RANSAC before TPS warp estimation. Note that feature tracking and propagation benefits all feature-based methods [27, 29, 30, 38–42]. The warp estimation results on all frames of the sequences (and on other sequences) can be obtained at `http://cs.adelaide.edu.au/~huy/`.



(a) Frame 140, ground truth **(b)** Frame 140, RANSAC **(c)** Frame 140, Annealed M-est. [30]

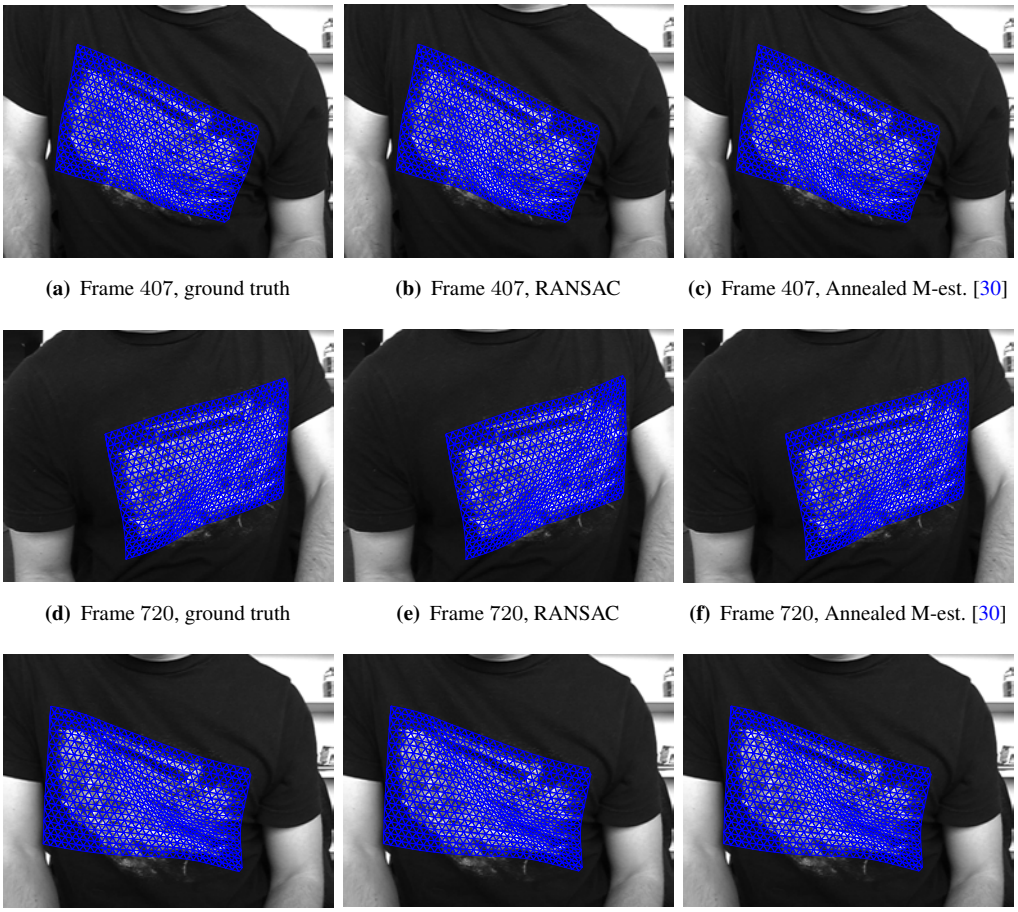
(d) Frame 160, ground truth **(e)** Frame 160, RANSAC **(f)** Frame 160, Annealed M-est. [30]

(g) Frame 178, ground truth **(h)** Frame 178, RANSAC **(i)** Frame 178, Annealed M-est. [30]

Figure 5.11: Qualitative retexturing results on the *Bedsheet* images.

5.5 Summary

We provided in this chapter substantial results supporting RANSAC as a viable and simple alternative for outlier rejection in deformable registration, compared to more sophisticated approaches. Our premise and observation is that, relative to the extreme scale of the gross mismatches, the distribution of the inliers usually resembles a low-dimensional affine subspace. While we focus in this chapter on RANSAC, there are many approaches to robust fitting of linear manifolds. Some may have advantages over RANSAC and, in that regard, an important message of this chapter is that outlier detection with nonlinear warping can likely be done with relatively cheap schemes.

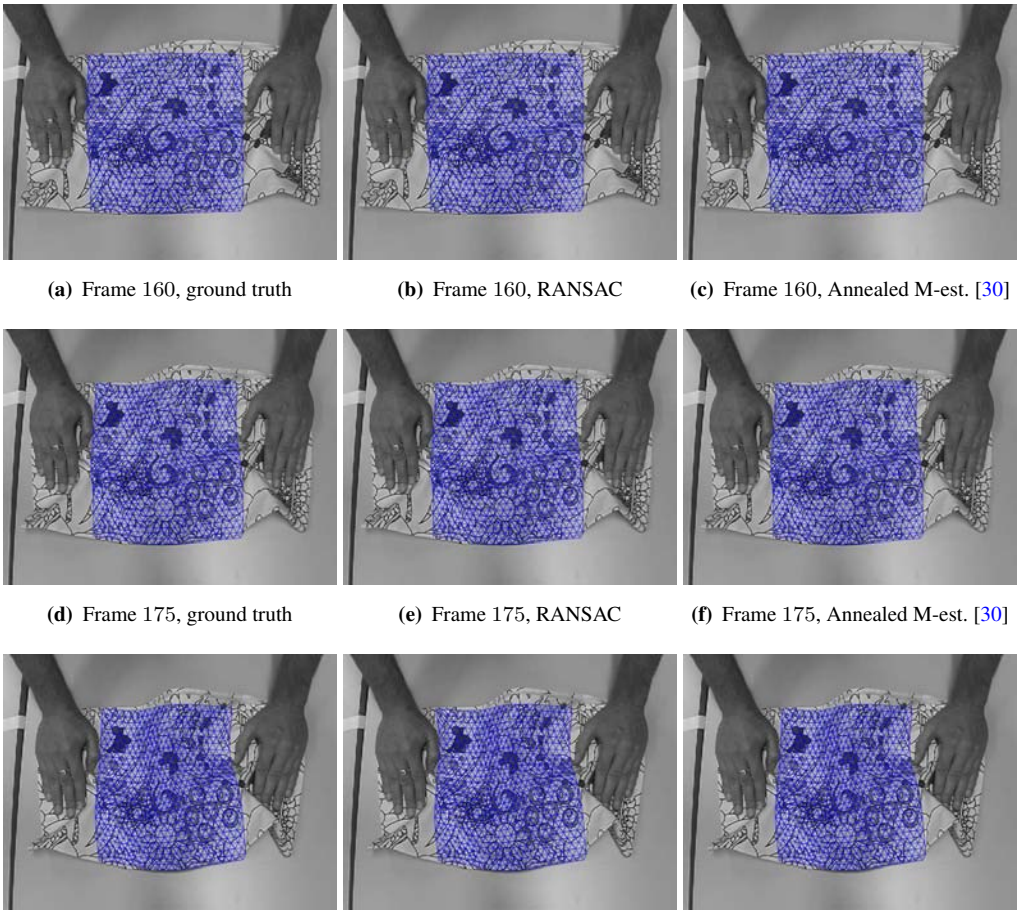


(a) Frame 407, ground truth **(b)** Frame 407, RANSAC **(c)** Frame 407, Annealed M-est. [30]

(d) Frame 720, ground truth **(e)** Frame 720, RANSAC **(f)** Frame 720, Annealed M-est. [30]

(g) Frame 784, ground truth **(h)** Frame 784, RANSAC **(i)** Frame 784, Annealed M-est. [30]

Figure 5.12: Qualitative retexturing results on the *Tshirt* images.



(a) Frame 160, ground truth **(b)** Frame 160, RANSAC **(c)** Frame 160, Annealed M-est. [30]

(d) Frame 175, ground truth **(e)** Frame 175, RANSAC **(f)** Frame 175, Annealed M-est. [30]

(g) Frame 190, ground truth **(h)** Frame 190, RANSAC **(i)** Frame 190, Annealed M-est. [30]

Figure 5.13: Qualitative retexturing results on the *Cushion* images.

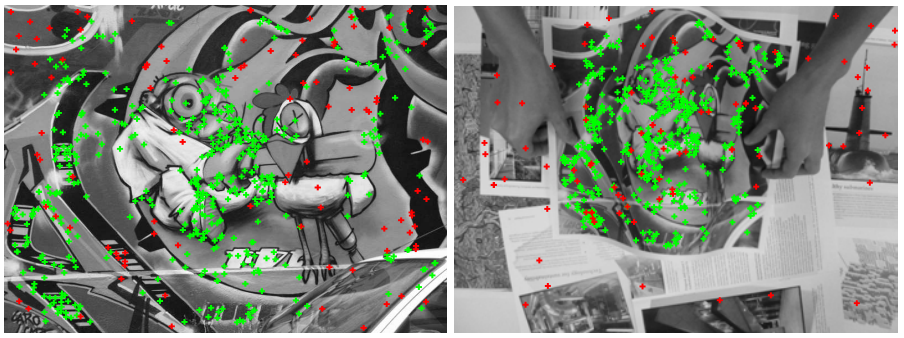
Chapter 6

Outlier Rejection in Deformable Registration with Moving Least Squares

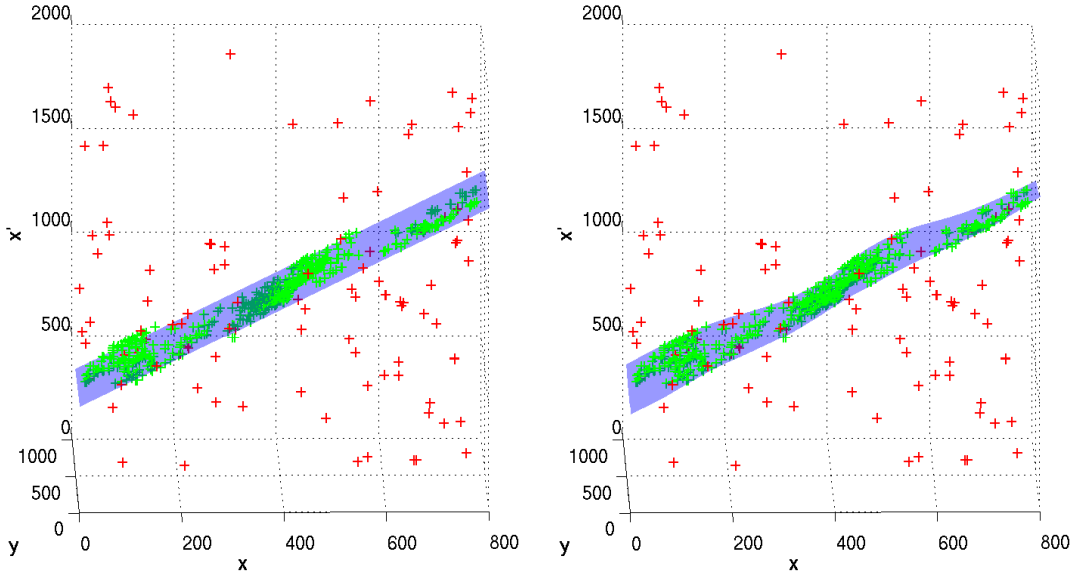
6.1 Introduction

The previous chapter established that, in practice, the scale of errors of the outliers usually dwarfs the nonlinear effects of the correspondence manifold. Therefore, relative to the outliers, the inliers lie very close to a linear hyperplane that corresponds to an affine warp. Furthermore, it suggested the use of RANSAC to robustly fit a linear hyperplane onto the data and separate the outliers. It was shown in the previous chapter that this simple outlier rejection method is at least as accurate as current approaches [27, 29, 30, 38–42] which are often based on optimising fully nonlinear warps in the presence of outliers.

Our observation paves the way for using simple linear models for outlier rejection in deformable registration. Compared to optimising fully nonlinear warps with outlier-contaminated data, the robust fitting of linear models is less cumbersome and can be implemented by many effective algorithms [17, 28]. However, there is a weakness in the simple RANSAC-based approach (Section 5.3): although the inliers do mostly lie within a "band" around the linear hyperplane, estimating the thickness of this band (i.e., inlier threshold) can be nontrivial since the complexity of the correspondence manifold is data-dependent. In particular, if the threshold is too large the estimate is affected by outliers, whereas a too small threshold may accidentally remove valid inliers.



(a) Template **(b)** Input image



(c) Linear M-estimation **(d)** Nonlinear surface approximation with MLS

Figure 6.1: Overview of our two-step approach to outlier rejection in deformable registration. First row shows the template (a) and input image (b) with SIFT correspondences. True matches (inliers) are in green while incorrect matches (outliers) are in red (correspondence lines are not drawn for clarity). Second row shows the stages of our approach: linear M-estimation (c) and nonlinear surface approximation with MLS (d), where the fitted function ($f_x : [x\ y]^T \mapsto x'$) is refined to better model the inliers (the same procedure is applied for $f_y : [x\ y]^T \mapsto y'$).

In this chapter, we seek to alleviate the above weakness. We introduce an approach that actually pushes the linear approximation idea further, so as to leverage on the effectiveness and computational ease of linear models. More specifically, we propose the following two-step method:

1. we first perform *linear* M-estimation [28] to quickly capture the correspondence manifold;
2. we then tweak the hyperplane using *nonlinear* surface approximation with moving least squares (MLS) [90–92] to refine the fit.

Outlier rejection is performed by thresholding the distances of keypoint matches to the fitted nonlinear warp; Figure 6.1 illustrates the proposed algorithm. As we will show in the experiments, this two-step method leads to significant improvements in accuracy. Being a well-known technique in computer graphics, MLS has received relatively little attention in computer vision. At their core, M-estimators and MLS solve a sequence of weighted linear least squares problems. Our two-step approach exploits this commonality such that outlier rejection for deformable registration can be accomplished as a sequence of simple linear least squares problems.

The rest of this chapter is organised as follows: Section 6.2 presents our two-step approach to outlier rejection in deformable registration, from linear initialisation to nonlinear refinement. Section 6.3 analyses our observation of the approximately linear correspondence manifold (Section 5.2), and benchmarks our two-step method against current approaches on various kinds of deformable objects. Finally, we summarise this chapter in Section 6.4. The main content of this chapter is based on the author's work in [93].

6.2 Outlier rejection as a sequence of linear least squares problems

We present in this section an effective outlier rejection method for deformable registration whoses core computation is a sequence of simple linear least squares problems.

6.2.1 Robust hyperplane fitting with M-estimators

Following our observation of the approximately linear correspondence manifold (Section 5.2), we first conduct robust hyperplane fitting to quickly capture the overall linear trend of the inliers. Due to the random sampling heuristic, the RANSAC-based approach in Section 5.3 does not guarantee good solutions unless a large number of samples are generated. In this section, we instead rely on M-estimators [28], which have good convergence properties and ensure at least locally optimal solutions given a fixed number of samples.

Given a set of keypoint matches $\mathcal{X} = \{(\mathbf{u}_i, \mathbf{v}_i)\}_{i=1}^n$, we adopt the standard regression setting by taking $\{\mathbf{u}_i\}_{i=1}^n$ as the independent measurements, and $\{x'_i\}_{i=1}^n$ and $\{y'_i\}_{i=1}^n$ separately from $\{\mathbf{v}_i\}_{i=1}^n$ as the dependent measurements. Due to the two dependent measurements, our technique robustly estimates *two* linear hyperplanes $\boldsymbol{\theta}_x$ and $\boldsymbol{\theta}_y$ (i.e., $\boldsymbol{\theta}_x$ for $\{x'_i\}_{i=1}^n$ and $\boldsymbol{\theta}_y$ for $\{y'_i\}_{i=1}^n$). Note that with a slight abuse of terminology, we will use the "shorthand" 2D linear hyperplane/nonlinear manifold as a description for the entire warp when we really mean for each component of the warp.

Since both $\boldsymbol{\theta}_x$ and $\boldsymbol{\theta}_y$ are computed in the same way, we will discuss estimating only $\boldsymbol{\theta}_x$ in the following. Denote $\tilde{\mathbf{u}}_i = [x_i \; y_i \; 1]^T$ as the homogeneous point of $\mathbf{u}_i$. The *M-estimate* is defined

as

$$\hat{\boldsymbol{\theta}}_x = \underset{\boldsymbol{\theta}_x \in \mathbb{R}^3}{\operatorname{argmin}} \sum_{i=1}^{n} \rho(\tilde{\mathbf{u}}_i^T \boldsymbol{\theta}_x - x_i'), \tag{6.1}$$

where $\tilde{\mathbf{u}}_i^T \boldsymbol{\theta}_x - x_i'$ is the residual of the i-th datum according to $\boldsymbol{\theta}_x$, and $\rho(u)$ is the robust loss function which is nonnegative, even symmetric, and nondecreasing with $|u|$ [28].

Differentiating (6.1) with respect to $\boldsymbol{\theta}_x$ and setting to zero yields the system of equations

$$\sum_{i=1}^{n} \tilde{\mathbf{u}}_i \psi(\tilde{\mathbf{u}}_i^T \boldsymbol{\theta}_x - x_i') = 0, \tag{6.2}$$

where $\psi(u) = \rho'(u)$ is the first-order derivative of $\rho(u)$. It is customary to use robust loss functions with derivatives of the form $\psi(u) = w(u) \cdot u$, where $w(u)$ is the *weight function*. In our work, we use Tukey's biweight function [28], which gives rise to the weight function

$$w(u) = \begin{cases} \left[1 - \left(\frac{u}{\tau}\right)^2\right]^2 & u \leq \tau \\ 0 & u > \tau \end{cases}, \tag{6.3}$$

where τ is the error scale.

Defining $w_i = w(\tilde{\mathbf{u}}_i^T \boldsymbol{\theta}_x - x_i')$ allows us to rewrite (6.2) as

$$\sum_{i=1}^{n} \tilde{\mathbf{u}}_i w_i \tilde{\mathbf{u}}_i^T \boldsymbol{\theta}_x = \sum_{i=1}^{n} \tilde{\mathbf{u}}_i w_i x_i', \tag{6.4}$$

which has the equivalent matrix form

$$\mathbf{X}^T \mathbf{W} \mathbf{X} \boldsymbol{\theta}_x = \mathbf{X}^T \mathbf{W} \mathbf{y}, \tag{6.5}$$

where $\mathbf{X} = [\tilde{\mathbf{u}}_1 \ \tilde{\mathbf{u}}_2 \ \ldots \ \tilde{\mathbf{u}}_n]^T$, $\mathbf{y} = [x_1' \ x_2' \ \ldots \ x_n']^T$, and $\mathbf{W} = \operatorname{diag}([w_1 \ w_2 \ \ldots \ w_n]^T)$ is the weight matrix ($\operatorname{diag}()$ yields a diagonal matrix from a given vector).

Given $\mathbf{W}$, solving for $\boldsymbol{\theta}_x$ based on (6.5) is a weighted linear least squares problem which can be accomplished using singular value decomposition (SVD). Given $\boldsymbol{\theta}_x$, $\mathbf{W}$ can be computed using $w_i = w(\tilde{\mathbf{u}}_i^T \boldsymbol{\theta}_x - x_i')$. This gives rise to the procedure known as *iterative reweighted least squares* (IRLS), which alternates the computation of $\boldsymbol{\theta}_x$ and $\mathbf{W}$ until convergence. The quality of the final estimate, however, depends on the initialisation. We perform multiple initialisations of $\boldsymbol{\theta}_x$ by randomly generating hypotheses in the manner of RANSAC. Each generated hypothesis is updated by IRLS until convergence, and the best hypothesis that minimises (6.1) is selected for the final estimate. In practice, the number of random hypotheses and IRLS iterations required

are relatively small (respectively about 100 and 20 in our experiments). Figure 6.1(c) shows a sample result of linear M-estimation on the data in Figures 6.1(a) and 6.1(b).

At convergence, data with larger weights are more likely to be inliers, and vice versa; see (6.3). As we will explain next, the M-estimator weights will be utilised to construct a robust version of MLS for outlier rejection in deformable registration.

6.2.2 Tweaking the hyperplane with MLS

The M-estimation in the previous section yields two sets of weights, $\{w_i^x\}_{i=1}^n$ and $\{w_i^y\}_{i=1}^n$, which arose respectively from the estimation of $\boldsymbol{\theta}_x$ and $\boldsymbol{\theta}_y$. We combine the two sets into a single one by multiplying the corresponding weights between the two sets as

$$w_i = w_i^x . w_i^y. \tag{6.6}$$

Therefore, for a keypoint correspondence $(\mathbf{u}_i, \mathbf{v}_i)$ to be an inlier, *both* its weights w_i^x and w_i^y according to $\boldsymbol{\theta}_x$ and $\boldsymbol{\theta}_y$ must be high. Henceforth, we regard $\mathbf{W}$ to contain the combined weights w_i from the estimation of $\boldsymbol{\theta}_x$ and $\boldsymbol{\theta}_y$.

Unlike the RANSAC-based approach in Section 5.3 where outlier rejection is performed after linear hyperplane fitting (i.e., by thresholding the weights in $\mathbf{W}$), we further tweak the hyperplane to produce a nonlinear manifold that better respects the correspondence data. Mismatch removal is then performed by thresholding the distances of keypoint matches to the fitted nonlinear warp. As we will show in Section 6.3, this modification leads to significant improvements in accuracy with a slight increase in computational cost.

Given a set of keypoint correspondences $\mathcal{X} = \{(\mathbf{u}_i, \mathbf{v}_i)\}_{i=1}^n$, for *each* point $\mathbf{p}_*$ in the template, MLS [90] approximates the warp f at point $\mathbf{p}_*$ as the *affine* transformation

$$\mathbf{A}_* = \underset{\mathbf{A}\in\mathbb{R}^{2\times 3}}{\operatorname{argmin}} \sum_{i=1}^{n} v_{i*} \|\mathbf{v}_i - \mathbf{A}\tilde{\mathbf{u}}_i\|^2. \tag{6.7}$$

The weights v_{i*} have the following form

$$v_{i*} = \exp\left(-\|\mathbf{u}_i - \mathbf{p}_*\|^2/\sigma^2\right), \tag{6.8}$$

where σ is the neighbourhood scale. Intuitively, since (6.8) assigns higher weights to data closer to $\mathbf{p}_*$, the affine transformation $\mathbf{A}_*$ better respects the local structure around $\mathbf{p}_*$. Furthermore, as $\mathbf{p}_*$ is *moved* continuously in its domain (the template), the parameters $\mathbf{A}_*$ also vary smoothly. Asymptotically this produces a *smooth* nonlinear warp [91] that adapts flexibly to the manifold represented by the samples $\mathcal{X}$.

The problem (6.7) is also a weighted linear least squares problem which has the matrix form

$$\mathbf{X}^T\mathbf{V}_*\mathbf{X}\mathbf{A}^T = \mathbf{X}^T\mathbf{V}_*\mathbf{Y}, \tag{6.9}$$

where $\mathbf{X} = [\tilde{\mathbf{u}}_1\ \tilde{\mathbf{u}}_2\ \dots\ \tilde{\mathbf{u}}_n]^T$, $\mathbf{Y} = [\mathbf{v}_1\ \mathbf{v}_2\ \dots\ \mathbf{v}_n]^T$, and $\mathbf{V}_* = \mathrm{diag}\,([v_{1*}\ \dots\ v_{n*}]^T)$.

To account for outliers, we incorporate the (converged) weights from M-estimation. Specifically, we modify (6.9) to become

$$\mathbf{X}^T\mathbf{W}\mathbf{V}_*\mathbf{X}\mathbf{A}^T = \mathbf{X}^T\mathbf{W}\mathbf{V}_*\mathbf{Y}, \tag{6.10}$$

where $\mathbf{W}$ reduces the influence of outlying data. In effect, this robust form of MLS tweaks the linear hyperplane fitted by M-estimation to yield a nonlinear manifold that better respects the nonlinearity of the true manifold and improves the separation between inliers and outliers. Again the solution can be obtained by means of SVD. Figure 6.1(d) illustrates the refined fitted function by MLS on the data in Figures 6.1(a) and 6.1(b).

6.2.3 Outlier rejection with MLS

Since we focus on mismatch removal, there is no need to solve (6.10) for every location $\mathbf{p}_*$ in the template. Instead, for each tentative match $(\mathbf{u}_i, \mathbf{v}_i)$ in $\mathcal{X}$, we estimate the warped point $\bar{\mathbf{v}}_i = \mathbf{A}_i\tilde{\mathbf{u}}_i$, where $\mathbf{A}_i$ is computed from (6.10) with $\mathbf{p}_* = \mathbf{u}_i$. The residual of the pair $(\mathbf{u}_i, \mathbf{v}_i)$ is defined as the Euclidean distance between $\bar{\mathbf{v}}_i$ and $\mathbf{v}_i$

$$r_i = \|\bar{\mathbf{v}}_i - \mathbf{v}_i\|. \tag{6.11}$$

Finally, outliers are identified and removed by thresholding the residuals of keypoint matches in $\mathcal{X}$. Hence, the number of instances of (6.10) to solve is equal to the number of keypoint matches in $\mathcal{X}$. Moreover, the affine parameters of all keypoint matches can be estimated independently, a straightforward way to speed up the algorithm is to solve multiple instances of (6.10) in parallel or multi-threads. Our method is summarised in Algorithm 6.1.

6.3 Experimental results

In this section we analyse our observation of the approximately linear correspondence manifold (Section 5.2) and evaluate the performance of our two-step method on various kinds of deformable objects ranging from synthetic to real datasets.

Algorithm 6.1: Outlier rejection with MLS

Input: set of correspondences $\mathcal{X} = \{(\mathbf{u}_i, \mathbf{v}_i)\}_{i=1}^{n}$, error scale τ, neighbourhood scale σ, inlier threshold γ.

Output: set of inliers $\mathcal{I} \subseteq \mathcal{X}$.

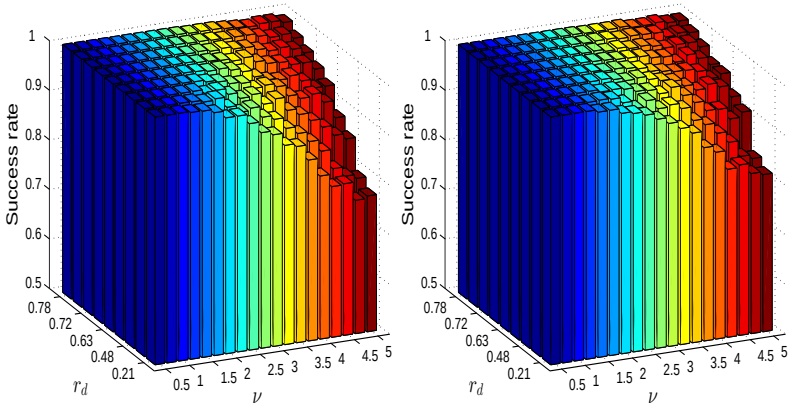
1: /* Linear M-estimation */
2: Construct $\mathbf{X} = [\tilde{\mathbf{u}}_1\ \tilde{\mathbf{u}}_2\ \ldots\ \tilde{\mathbf{u}}_n]^T$, $\mathbf{y}_x = [x'_1\ x'_2\ \ldots\ x'_n]^T$, $\mathbf{y}_y = [y'_1\ y'_2\ \ldots\ y'_n]^T$.
3: Estimate $\boldsymbol{\theta}_x$ and $\boldsymbol{\theta}_y$ respectively from ($\mathbf{X}$,$\mathbf{y}_x$) and ($\mathbf{X}$,$\mathbf{y}_y$) (Section 6.2.1).
4: Compute $\{w_i^x\}_{i=1}^n$ and $\{w_i^y\}_{i=1}^n$ according to $\boldsymbol{\theta}_x$ and $\boldsymbol{\theta}_y$ (Equation (6.3)).
5: Form $\mathbf{W} = \mathrm{diag}\left([w_1\ w_2\ \ldots\ w_n]^T\right)$ with $w_i = w_i^x.w_i^y$, $i = 1, 2, \ldots, n$.
6: /* Outlier rejection with MLS */
7: Initialise $\mathcal{I} = \emptyset$.
8: **for** each $(\mathbf{u}_i, \mathbf{v}_i)$ in $\mathcal{X}$ **do**
9: Estimate $\mathbf{A}_i$ with $\mathbf{p}_* = \mathbf{u}_i$ (Section 6.2.2).
10: Compute r_i (Equation (6.11)).
11: If $r_i \leq \gamma$ then $\mathcal{I} = \mathcal{I} \cup \{(\mathbf{u}_i, \mathbf{v}_i)\}$.
12: **end for**

6.3.1 Synthetic data

6.3.1.1 Data analysis

We first investigate our observation of the approximately linear correspondence manifold (Section 5.2) on synthetic data. More specifically, we aim to test the *conditions* under which the linear hyperplane approximation is effective. To this end, we vary the margin d in $[25, 225]$ pixels (or the margin ratio r_d in $[0.21, 0.78]$) and the bending energy ν in $[0.5, 5.0]$; see Section 5.4.1.1 for the generation procedure. For each pair of r_d and ν, we generate 100 random synthetic image pairs, and compute for each image pair the maximum inlier deviation (over 100 random inliers) and the average outlier distance (over 100 random outliers) to the randomly generated TPS warp's affine component. If the maximum of inlier deviations is less than the mean of outlier distances, we consider the test successful. The success rate over 100 random image pairs for each r_d and ν combination is illustrated in Figure 6.2(a). In addition, Figure 6.2(b) presents the results in which the distances of keypoint matches are computed with respect to the nonlinear MLS warp obtained using our two-step technique.

Unsurprisingly, the results show that for high ν (high bending energy), the linear hyperplane approximation is not effective; this is because the correspondence manifold will be highly distorted and the inliers will be "mixed" closely with the outliers. Interestingly, however, for $\nu \leq 2.0$ the linear hyperplane approximation is quite valid. The question is therefore what kind of bending energies are exhibited by real-life objects. Figure 6.3 shows TPS warps estimated using (manually annoted) true inliers and (manually added) extra inliers on deformable objects typically examined in the literature (we use the publicly available real datasets of [64] that were used in Section 5.4.2, and our real datasets in Section 6.3.2). In each video sequence of [64], we select



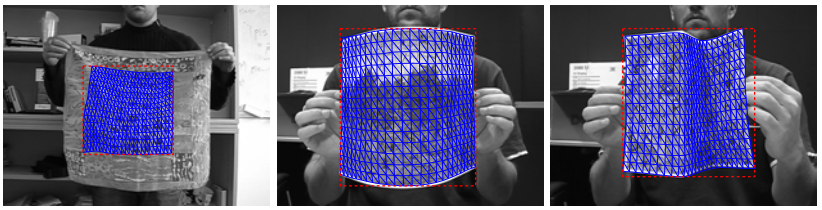
(a) Linear hyperplane approximation **(b)** Our novel MLS method

Figure 6.2: Success rate with different bending energies and margin ratios.

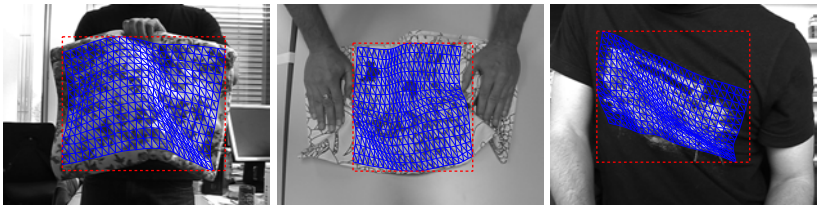
the frame corresponding to the object under the *highest estimated bending energies*. The warps in Figure 6.3 all show bending energies less than 2.0, hence justifying the proposed method. Note that without extra inliers, the estimated bending energies in Figure 6.3 will be slightly decreased.

Another factor influencing the results in Figure 6.2 is the margin ratio r_d. More specifically, the validity of the linear hyperplane approximation reduces as r_d decreases. This has the simple explanation in that, since the outliers are randomly sampled points in the template and input image, if the margin ratio is small the outliers will be sampled mostly from the same image regions as the inliers, thus decreasing the average distance of the outliers to the true correspondence manifold.

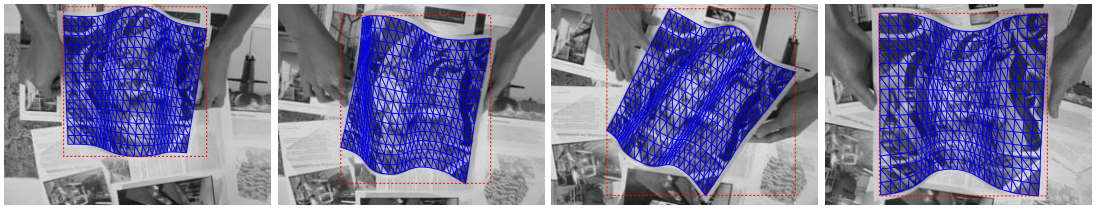
We further analyse the data as follows. For each random synthetic image pair, we consider the randomly generated TPS warp as the *ground truth* warp. To obtain a warp resulting purely from a linear hyperplane approximation, we first set the inlier threshold as the maximum deviation of the true inliers to the RANSAC-fitted linear hyperplane (obtained using the RANSAC-based approach in Section 5.3). All keypoint matches having distances larger than this threshold are removed and a TPS warp is estimated using the remaining matches. We compute the root mean squared error (RMSE) on pixel coordinates (sampled at the mesh vertices) between the warped "image" (mapped by the estimated warp) and the ground truth "image" (mapped by the ground truth warp). The average RMSE over 100 random image pairs for each r_d and ν combination is shown in Figure 6.4(a). Similarly, Figure 6.4(b) presents the results in which the MLS warp estimated by our two-step method is used instead for outlier rejection. An expected trend in both figures is that the registration error increases with the increasing bending energy and is less sensitive to the margin ratio. Furthermore, we obtain the *baseline* warp by estimating a TPS warp on the true inliers (note that the baseline warp may differ from the ground truth warp since the inliers are contaminated with additive Gaussian noise). Figure 6.4(c) shows the



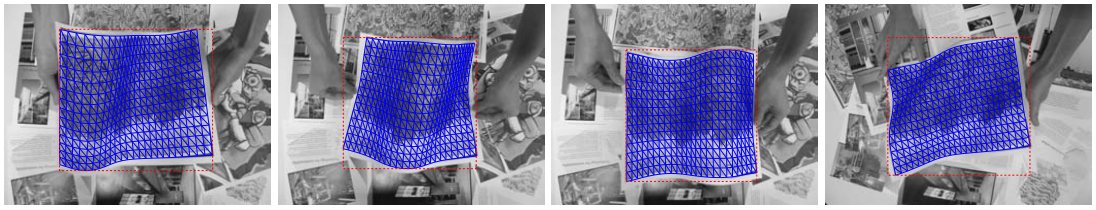
(a) $\nu(0.08, 0.14)$,$r_d = 0.85$ **(b)** $\nu(0.22, 0.21)$,$r_d = 0.60$ **(c)** $\nu(0.25, 0.21)$,$r_d = 0.63$



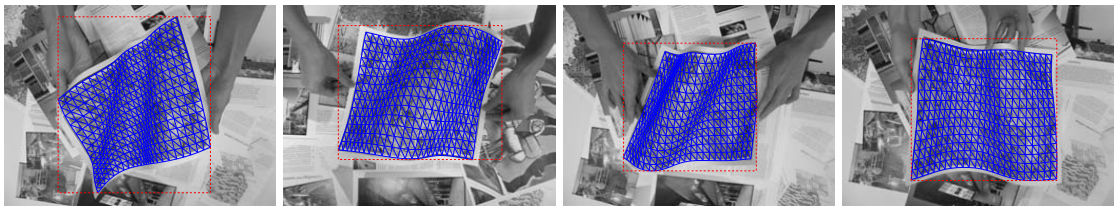
(d) $\nu(0.27, 0.38)$,$r_d = 0.59$ **(e)** $\nu(0.39, 0.33)$,$r_d = 0.71$ **(f)** $\nu(0.54, 0.57)$,$r_d = 0.62$



(g) $\nu(0.36, 0.15)$,$r_d = 0.60$ **(h)** $\nu(0.55, 0.32)$,$r_d = 0.53$ **(i)** $\nu(0.63, 0.53)$,$r_d = 0.34$ **(j)** $\nu(0.66, 0.39)$,$r_d = 0.42$

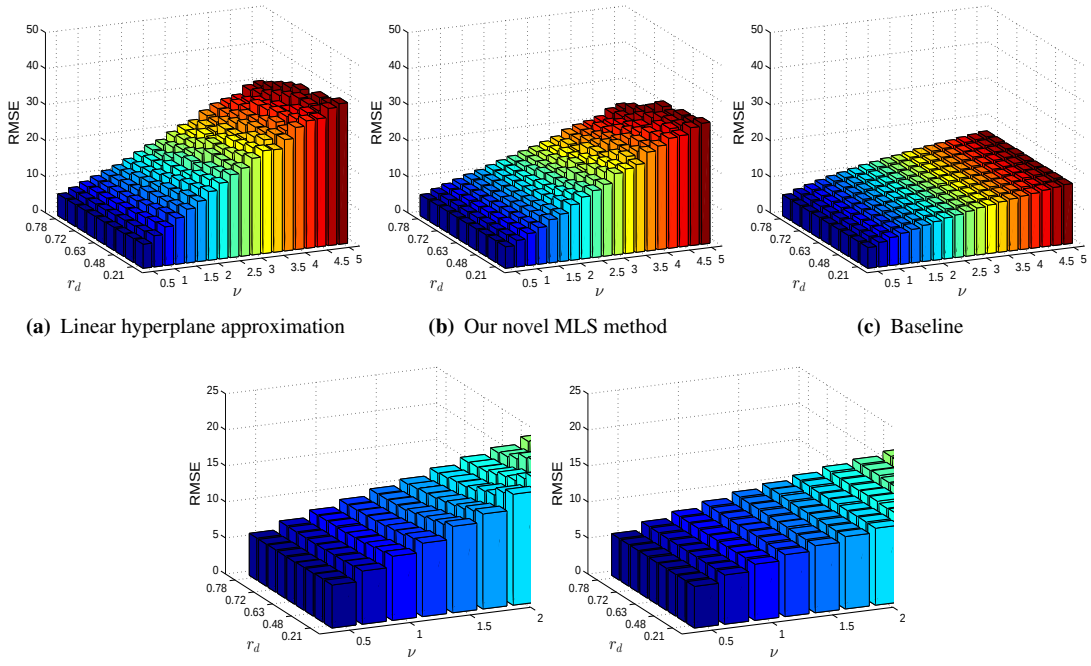


(k) $\nu(0.46, 0.37)$,$r_d = 0.59$ **(l)** $\nu(0.55, 0.20)$,$r_d = 0.67$ **(m)** $\nu(0.24, 0.14)$,$r_d = 0.68$ **(n)** $\nu(0.18, 0.12)$,$r_d = 0.63$



(o) $\nu(0.90, 0.32)$,$r_d = 0.51$ **(p)** $\nu(0.53, 0.36)$,$r_d = 0.60$ **(q)** $\nu(0.93, 0.47)$,$r_d = 0.69$ **(r)** $\nu(0.35, 0.20)$,$r_d = 0.62$

Figure 6.3: Examples of bending energies and margin ratios on publicly available real datasets of [64] (first and second rows) and our real datasets in Section 6.3.2 (third, forth and fifth rows). Under each image are the estimated bending energies $\nu(\nu_x, \nu_y)$ and margin ratio r_d.



(a) Linear hyperplane approximation **(b)** Our novel MLS method **(c)** Baseline

(d) Zoom-in view of (b) for $\nu \leq 2.0$ **(e)** Zoom-in view of (c) for $\nu \leq 2.0$

Figure 6.4: RMSE with different bending energies and margin ratios.

average RMSE between the baseline "image" (mapped by the baseline warp) and the ground truth "image". Note that, comparing Figure 6.4(b) with Figure 6.4(c), there is just a small difference for $\nu \leq 2.0$ (see zoom-in views in Figures 6.4(d) and 6.4(e)).

The above results justify our ideas that outlier rejection on deformable objects can likely be accomplished using simple linear models.

6.3.1.2 Comparison of methods

We benchmark our method in Section 6.2, namely *ME+MLS*, against the following state-of-the-art outlier rejection methods for deformable registration: Iterative local smoothness test [42, Section 3], SVM regression with resampling [27] and Annealed M-estimation [30]. We also compare against the RANSAC-based approach in Section 5.3. For ME+MLS, following previous methods, e.g., [29, 30, 38, 39], we normalise the data (using [75]) to allow the use of a constant error scale τ and neighbourhood scale σ for all datasets. These parameters were manually tuned and then used for input images (note that σ acts as a regularisation coefficient, however following previous methods, we fixed σ in our experiments). As we show later, our method still provides the best overall accuracy. The ROC curve of each method is obtained by varying the inlier threshold and recording the resultant true positive rate and false positive rate. We set $\nu = 0.5, 1.0, 2.0, 5.0$, and for each ν, the outlier rate ϵ is set as 0.33 and 0.5 (in all these cases, the margin ratio $r_d = 0.68$). For each combination of ν and ϵ, 100 random TPS warps

Outlier rate	$\epsilon = 0.33$ (total 150 matches)	$\epsilon = 0.5$ (total 200 matches)
RANSAC	0.04	0.04
Local smoothness [42]	0.26	0.31
SVM regression [27]	0.06	0.09
Annealed M-estimation [30]	0.15	0.29
ME+MLS	0.06	0.06

Table 6.1: Average run time (in seconds) for outlier rejection on synthetic data.

are generated, and the ROC curves for each method are averaged over the 100 warps. Figure 6.5 presents the results.

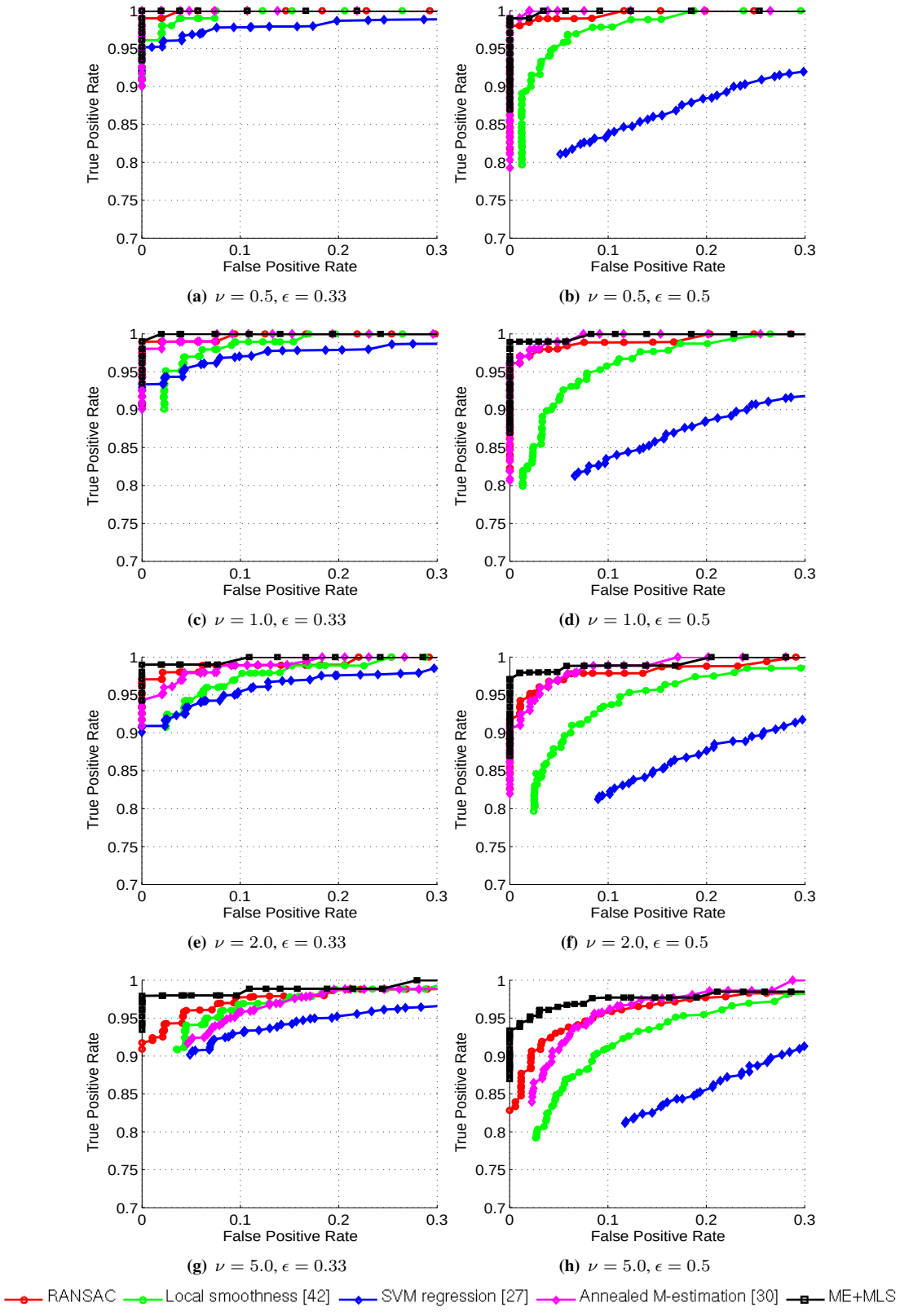
An apparent and expected trend is that as ν and ϵ increase, the accuracies of all methods decrease, with the method of [27] deteriorating the fastest, followed by [42]. RANSAC and Annealed M-estimation [30] provide very comparable results, and both outperform the methods of [27, 42]. ME+MLS provides the best overall accuracy. Comparing RANSAC (pure linear hyperplane fitting) and ME+MLS shows the advantage of the nonlinear refinement using MLS after linear hyperplane fitting.

Table 6.1 presents the average running time of all methods for ϵ = 0.33 and 0.5. From the results, RANSAC and ME+MLS are the most efficient methods, owing to taking advantage of the *overall linear trend* of the inliers. In particular, RANSAC is the fastest due to its lowest complexity. ME+MLS has a small increase in run time compared to RANSAC due to the MLS surface approximation. On the other hand, the methods of [27, 30, 42] are relatively much slower, and their run times also scale faster with the data size.

6.3.2 Flat surface deformations

We now test the performance of ME+MLS (Section 6.2) on real surfaces. All methods were run on 3 different datasets, namely *Graffiti*, *Paper* and *Model*, which we created by physically deforming the template images; see Figure 6.6 for the templates. The template of *Graffiti* has size of 640×800 pixels, while the *Paper*'s and *Model*'s templates are of size 480×600 pixels. Each dataset contains 4 input images of size 1500×2000 pixels. The datasets are considered to be more difficult than the publicly available real datasets of [64] that were used in Section 5.4.2, due to strong deformations and background clutter. SIFT matches [16] were obtained using VLFeat toolbox [81], and we then manually categorised them as true inliers and outliers. For RANSAC and ME+MLS, 100 repetitions were performed on each input image, and the average results (ROC curves) are reported in Figures 6.7, 6.8 and 6.9. To observe the correspondence manifold, we project the joint keypoint matches $\mathbf{x}_i = [x_i \; y_i \; x'_i \; y'_i]^T$ to 3D space using PCA.

The low-dimensional visualisations of all data indicate that the inliers do mostly lie within a "band" around a linear hyperplane. However, by explicitly modelling the deviations of inliers



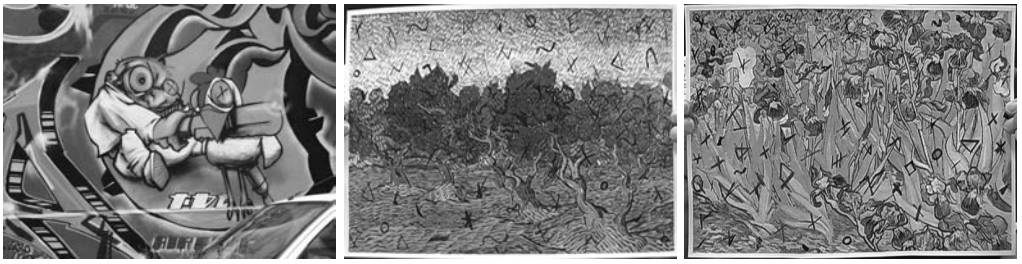
(a) $\nu = 0.5, \epsilon = 0.33$ **(b)** $\nu = 0.5, \epsilon = 0.5$

(c) $\nu = 1.0, \epsilon = 0.33$ **(d)** $\nu = 1.0, \epsilon = 0.5$

(e) $\nu = 2.0, \epsilon = 0.33$ **(f)** $\nu = 2.0, \epsilon = 0.5$

(g) $\nu = 5.0, \epsilon = 0.33$ **(h)** $\nu = 5.0, \epsilon = 0.5$

Figure 6.5: Outlier rejection results on synthetic data.



(a) *Graffiti* **(b)** *Paper* **(c)** *Model*

Figure 6.6: Templates of the *Graffiti*, *Paper* and *Model* datasets.

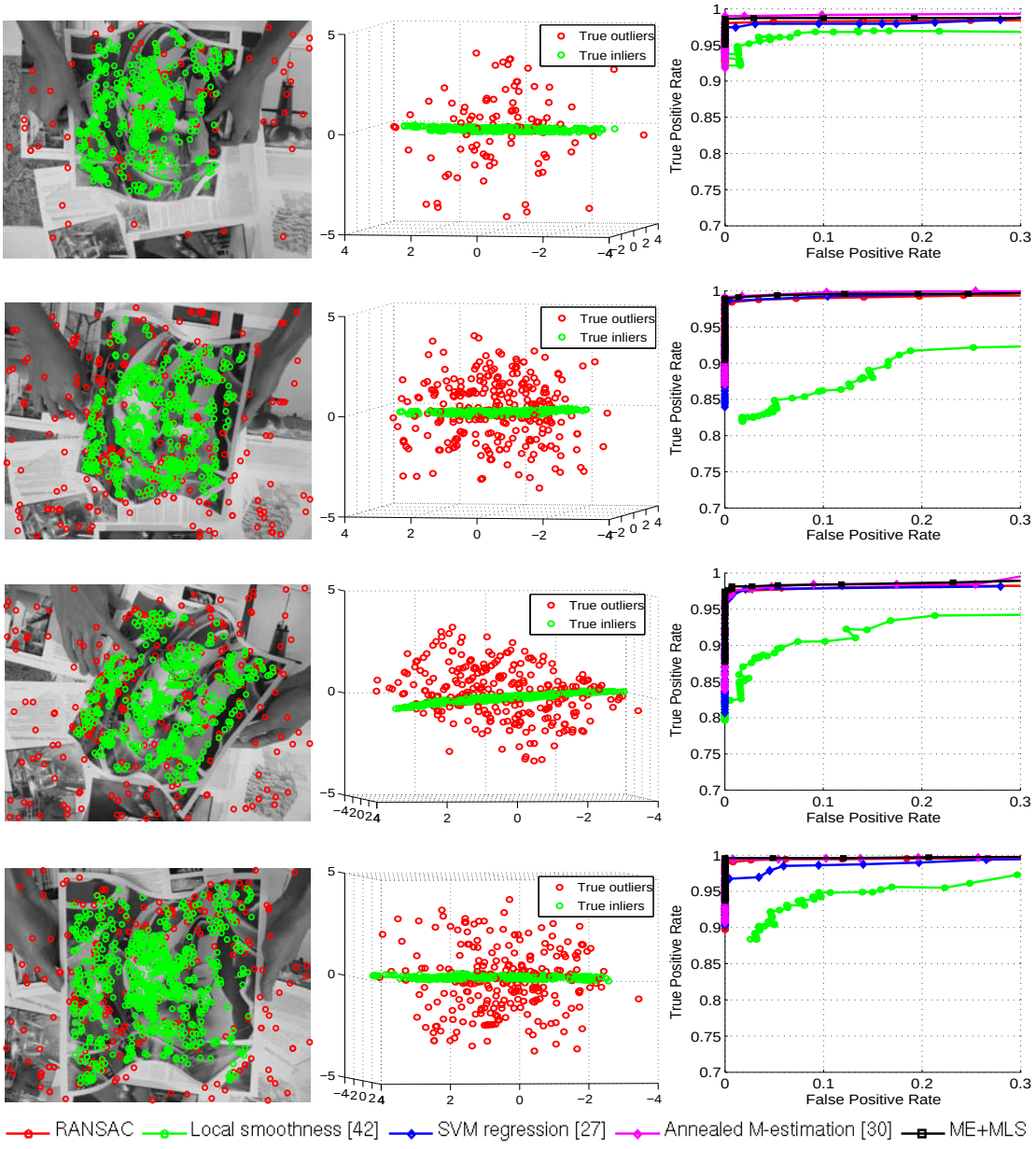


Figure 6.7: Outlier rejection results on the *Graffiti* dataset. Column 1: Input images (top to bottom): $G1$ (599 matches, 17.53% outliers), $G2$ (853 matches, 33.65% outliers), $G3$ (826 matches, 35.23% outliers), and $G4$ (989 matches, 25.08% outliers). Column 2: Data after PCA. Column 3: ROC curves.

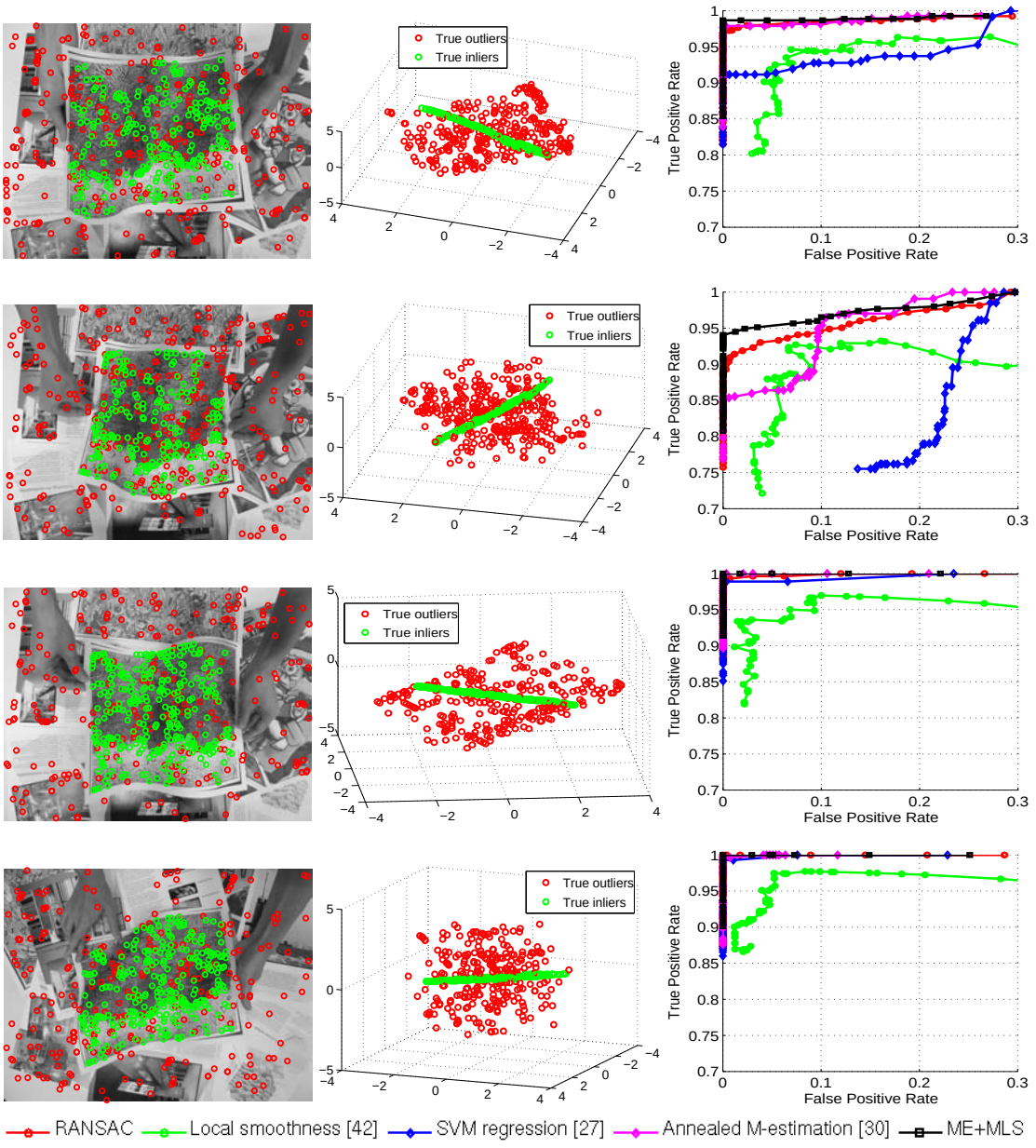


Figure 6.8: Outlier rejection results on the *Paper* dataset. Column 1: Input images (top to bottom): $P1$ (572 matches, 56.12% outliers), $P2$ (571 matches, 63.92% outliers), $P3$ (573 matches, 50.09% outliers), and $P4$ (563 matches, 49.73% outliers). Column 2: Data after PCA. Column 3: ROC curves.

within the linear band, ME+MLS achieves the highest accuracy. This is followed by Annealed M-estimation [30] and RANSAC, which usually give comparable results. The methods of [27] and [42] give the lowest accuracies. Note that the accuracy of [27] seems to improve when the inliers are dense, such as in *Graffiti*. Also, the accuracy of [42] shows a marked decrease as the outlier rate increases. Table 6.2 depicts the run times of all methods, which are similar to the results in Table 6.1.

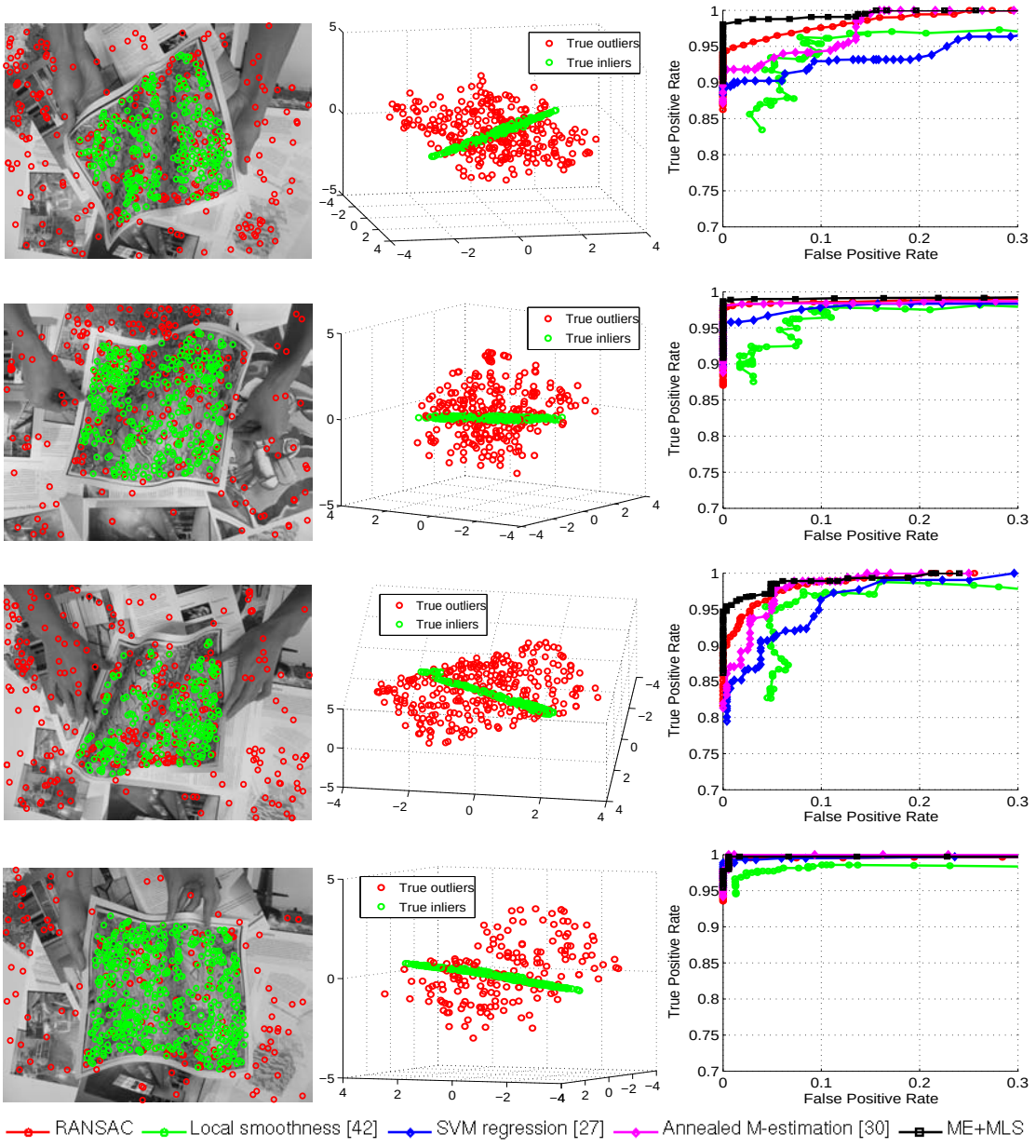


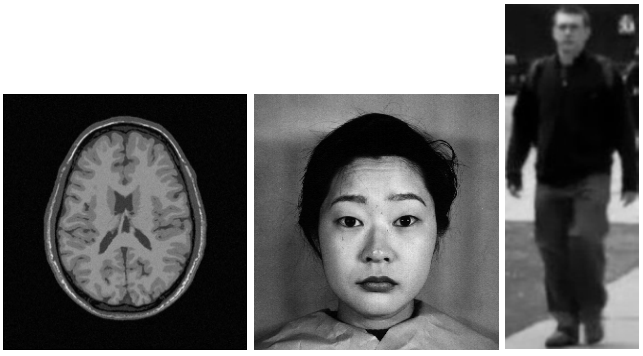
Figure 6.9: Outlier rejection results on the *Model* dataset. Column 1: Input images (top to bottom): $M1$ (547 matches, 53.02% outliers), $M2$ (563 matches, 46.36% outliers), $M3$ (515 matches, 57.67% outliers), and $M4$ (612 matches, 28.76% outliers). Column 2: Data after PCA. Column 3: ROC curves.

Dataset name	*Graffiti*				*Paper*				*Model*			
Input image	$G1$	$G2$	$G3$	$G4$	$P1$	$P2$	$P3$	$P4$	$M1$	$M2$	$M3$	$M4$
RANSAC	0.09	0.11	0.11	0.13	0.09	0.09	0.09	0.09	0.09	0.09	0.08	0.10
Local smoothness [42]	2.44	3.74	3.92	4.56	2.14	2.27	1.98	1.83	1.86	1.74	1.56	2.64
SVM regression [27]	0.70	1.05	1.09	1.28	0.59	0.64	0.53	0.49	0.55	0.49	0.39	0.78
Annealed M-est. [30]	0.87	1.32	1.37	1.62	0.81	0.84	0.79	0.77	0.77	0.75	0.71	0.89
ME+MLS	0.16	0.19	0.19	0.21	0.15	0.15	0.15	0.15	0.15	0.15	0.14	0.17

Table 6.2: Average run time (in seconds) for outlier rejection on the *Graffiti*, *Paper* and *Model* datasets.

6.3.3 Medical images, facial expressions and human actions

We also test the performance of our method (Section 6.2) on medical images. Brain MRI images from [94, 95] are used in this experiment, where the brain deforms due to artificial deformations which resemble different intra-subject and inter-subject mapping behaviours in the real world; see Figure 6.10 for the template. VLFeat library was first used to produce SIFT correspondences, which we then manually labelled as true inliers or outliers. The ROC results of all methods are illustrated in Figure 6.11, while the run times are reported in Table 6.3. From the results, ME+MLS outperforms all other methods in term of accuracy and is also the second fastest.



(a) *Brain* **(b)** *Face* **(c)** *Walk*

Figure 6.10: Templates of the *Brain*, *Face* and *Walk* datasets.

We also applied the methods on images involving deformations due to facial expressions and human actions. Four different expressions (fear, happiness, sadness and surprise) of the sample person taken from the Japanese Female Facial Expression Database [96], namely *Face*, and four video frames capturing a walking man taken from the Collective Activity Dataset [97], namely *Walk*, are input images for this experiment; see Figure 6.10 for the templates of these datasets. Outlier rejection results on these datasets are illustrated respectively in Figures 6.12 and 6.13 while the run times of all methods are presented in Table 6.3. The results show that ME+MLS is again the most accurate and the second fastest.

Dataset name	*Brain*				*Face*				*Walk*			
Input image	$B1$	$B2$	$B3$	$B4$	$F1$	$F2$	$F3$	$F4$	$W1$	$W2$	$W3$	$W4$
RANSAC	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Local smoothness [42]	0.38	0.35	0.36	0.36	0.46	0.45	0.41	0.43	0.27	0.22	0.23	0.25
SVM regression [27]	0.09	0.08	0.08	0.08	0.12	0.11	0.09	0.10	0.08	0.07	0.07	0.08
Annealed M-est. [30]	0.25	0.21	0.22	0.23	0.37	0.34	0.30	0.32	0.18	0.11	0.12	0.15
ME+MLS	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06

Table 6.3: Average run time (in seconds) for outlier rejection on the *Brain*, *Face* and *Walk* datasets.

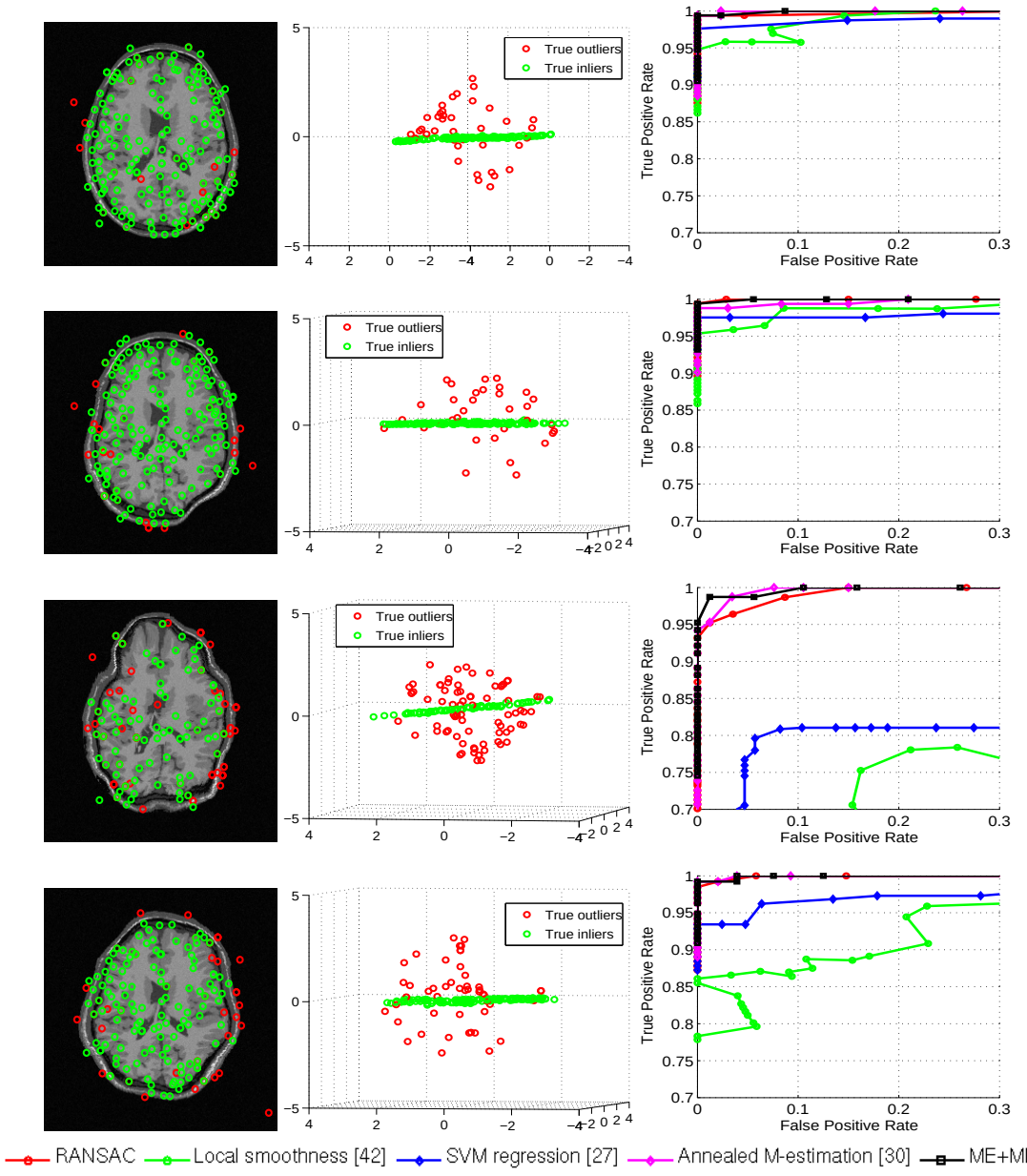


Figure 6.11: Outlier rejection results on the *Brain* dataset. Column 1: Input images (top to bottom): $B1$ (204 matches, 20.59% outliers), $B2$ (198 matches, 17.17% outliers), $B3$ (167 matches, 50.90% outliers) and $B4$ (179 matches, 27.37% outliers). Column 2: Data after PCA. Column 3: ROC curves.

6.3.4 Performance with decreasing number of matches

We now examine the performance of our method with decreasing number of matches. Two datasets ($G3$ and $P3$) with different outlier ratios (respectively, 35.23% and 50.09%) are selected from the datasets in Section 6.3.2. For each chosen dataset, the number of inliers is decreased by 25%, 50% and 75% respectively, and we then control the number of outliers to keep the original outlier ratio (by randomly sampling inliers and outliers from the original data). Therefore, the total number of matches is reduced by 25%, 50% and 75% respectively, while fixing the outlier ratio. The ROC curves of all methods on these new datasets are plotted in Figure 6.14. The

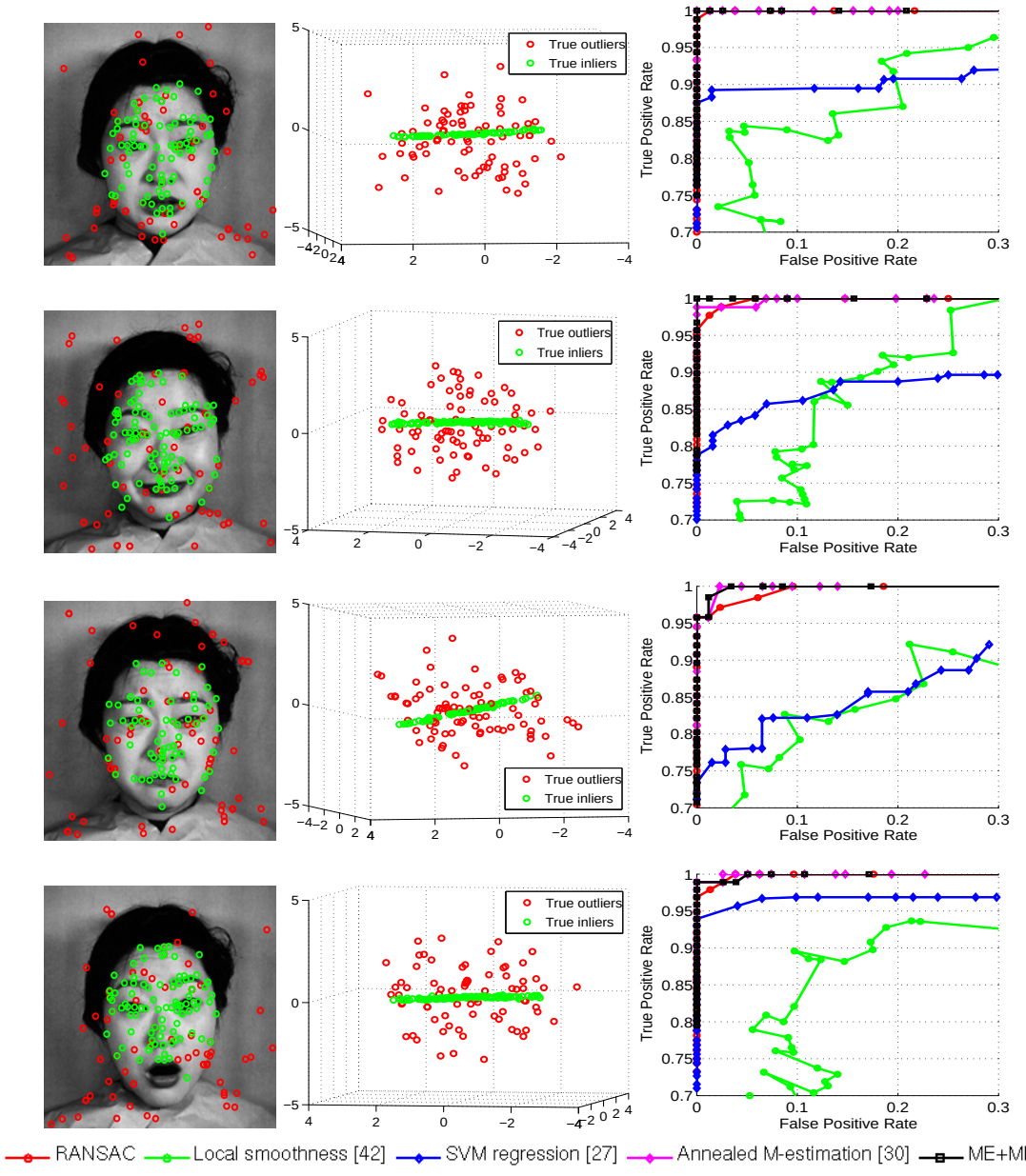


Figure 6.12: Outlier rejection results on the *Face* dataset. Column 1: Input images (top to bottom): $F1$/fear (160 matches, 47.50% outliers), $F2$/happiness (170 matches, 47.65% outliers), $F3$/sadness (155 matches, 55.48% outliers) and $F4$/surprise (168 matches, 44.64% outliers). Column 2: Data after PCA. Column 3: ROC curves.

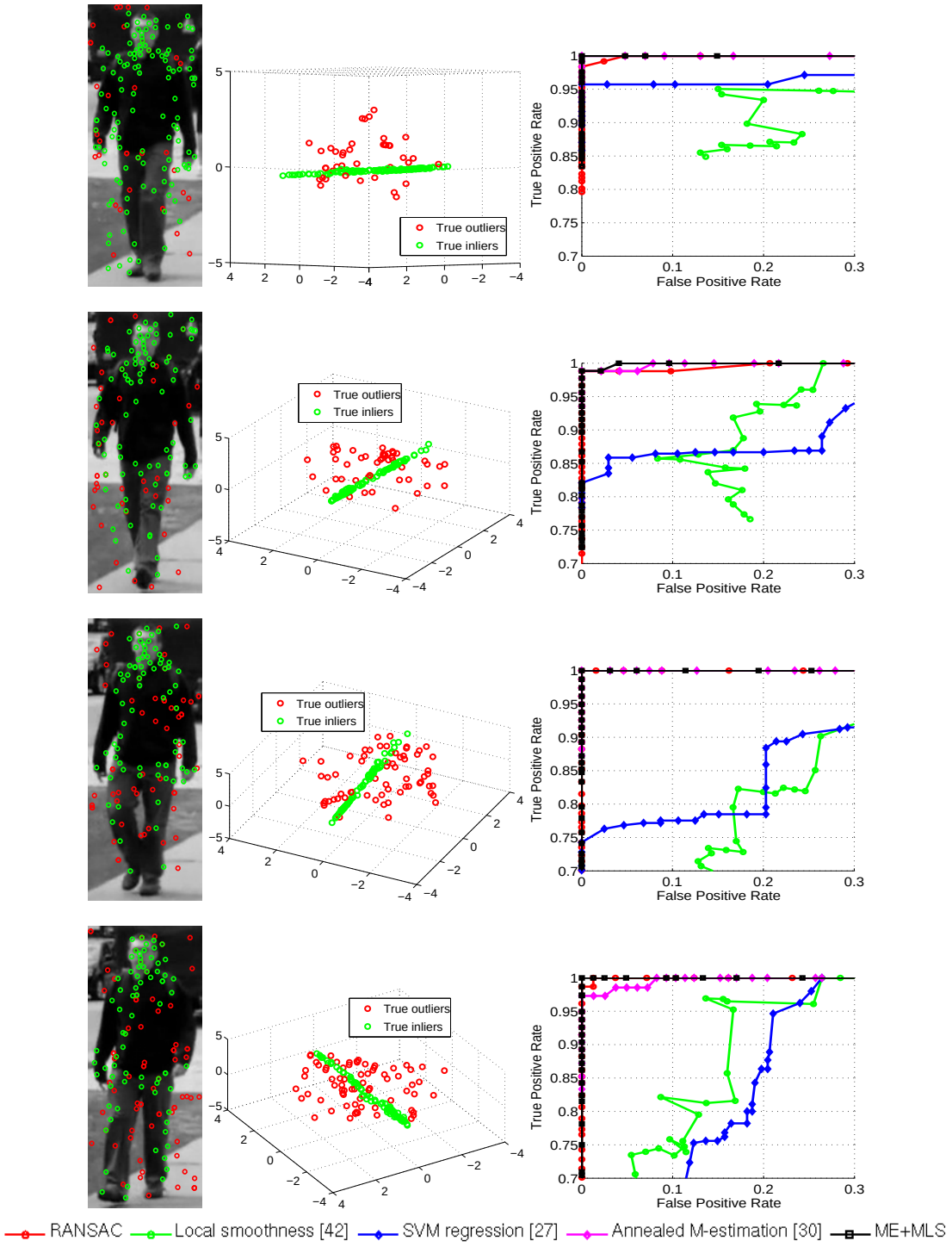
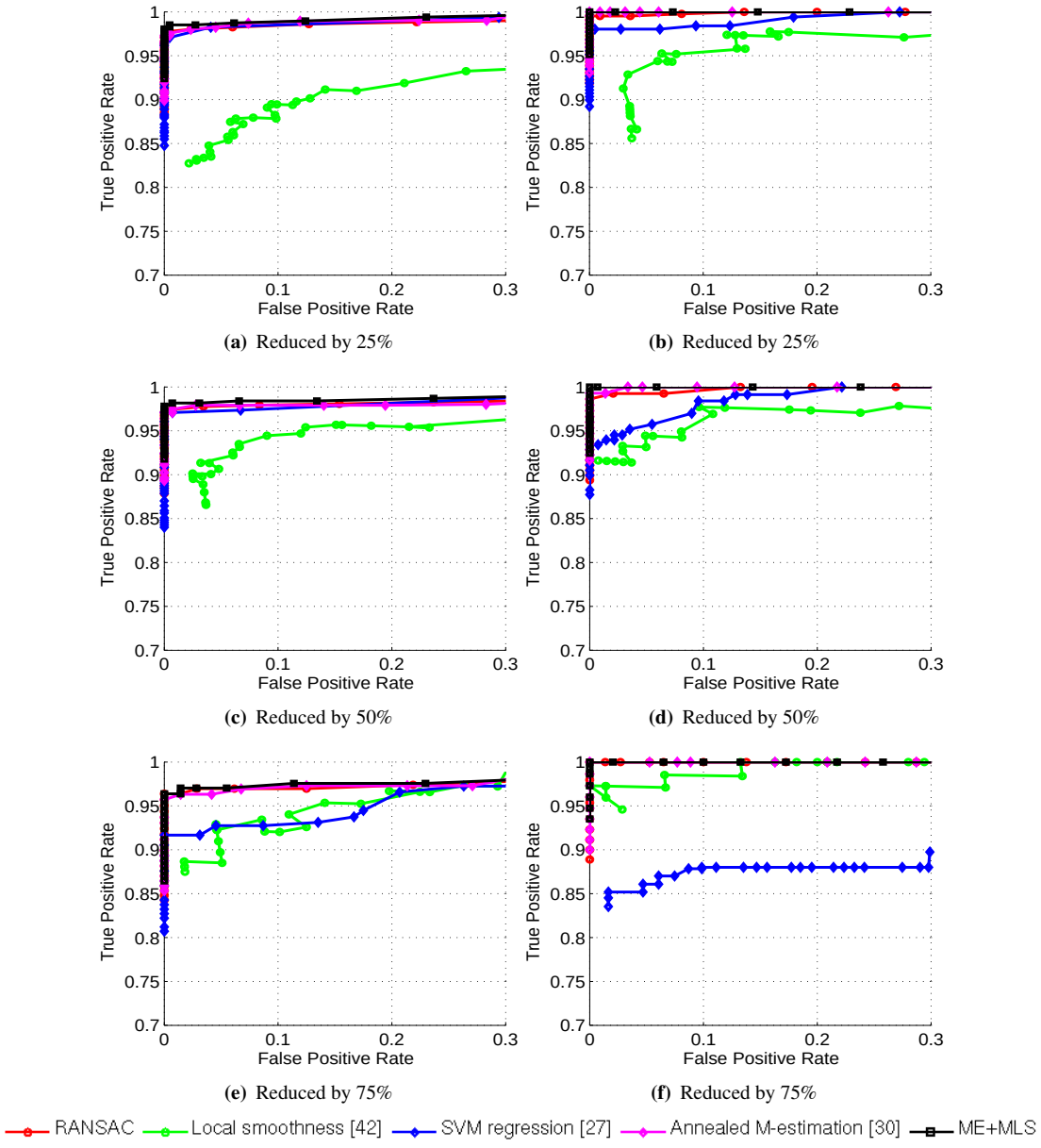


Figure 6.13: Outlier rejection results on the *Walk* dataset. Column 1: Input images (top to bottom): $W1$ (161 matches, 24.84% outliers), $W2$ (134 matches, 35.07% outliers), $W3$ (137 matches, 45.26% outliers) and $W4$ (153 matches, 50.98% outliers). Column 2: Data after PCA. Column 3: ROC curves.



(a) Reduced by 25%

(b) Reduced by 25%

(c) Reduced by 50%

(d) Reduced by 50%

(e) Reduced by 75%

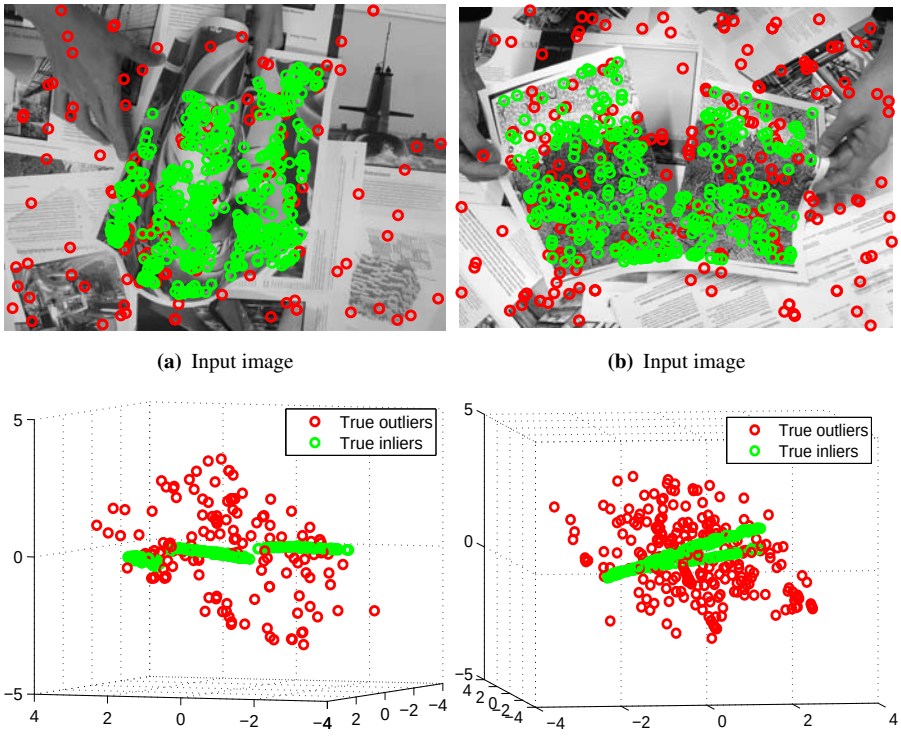
(f) Reduced by 75%

Figure 6.14: Performance with decreasing number of matches. First column: $G3$ (initially, 826 matches, 35.23% outliers). Second column: $P3$ (initially, 573 matches, 50.09% outliers).

results show that the performances of Annealed M-estimation [30], RANSAC, and ME+MLS are stable under decreasing number of matches. In contrast, the accuracy of [27] improves when there are a large number of matches (inliers), while the accuracy of [42] seems to increase in cases of small number of matches (outliers). These results are consistent with our comments in Section 6.3.2.

6.3.5 Failure cases on real data

Figure 6.15 shows failure cases of our method on real data. As discussed in Section 6.3.1.1, our method relies on the observation of the approximately linear correspondence manifold, which becomes less effective in the presence of strong deformation and self-occlusion; see Figure 6.15(a) for an example. Under these conditions, the correspondence manifold is distorted significantly, and outliers are mixed easily among inliers; see Figure 6.15(c). Moreover, our method may fail due to changes in object topology; see Figure 6.15(b) for an example of a flat surface being torn. In this case, the correspondence manifold is likely to be broken into various segments, and outliers are blended easily among these segments; see Figure 6.15(d). Finally, although we have not tested on Laparoscopic images, it is predictable for our method to find them challenging since the objects involved in those images are not inextensible flat surfaces. Note that the above situations also affect other methods [27, 30, 42].



(a) Input image **(b)** Input image

(c) Data after PCA **(d)** Data after PCA

Figure 6.15: Examples of failure cases on real data. First column: strong deformation and self-occlusion. Second column: strong variation in object topology.

6.4 Summary

Motivated by the interesting observation of the approximately linear correspondence manifold in the previous chapter, we have presented in this chapter an effective method for outlier rejection

in deformable registration, which is based entirely on fitting simple linear models: first, linear M-estimation is employed for quickly capturing the correspondence manifold, and to better fit the inliers, MLS is then conducted as a sequence of simple linear least squares problems. Outliers are robustly removed by thresholding the distances of keypoint matches to the fitted nonlinear warp. Comparisons against state-of-the-art approaches on synthetic and real datasets have illustrated the effectiveness of our method. Most importantly, the significant results of our method on various kinds of deformable objects empirically validate our message that outlier rejection on deformable objects can be accomplished using simple linear models.

Chapter 7

Conclusion and Future Directions

Parameter estimation underpins a variety of applications in computer vision, including three-dimensional reconstruction, image stitching, image registration and shape matching to name a few. Due to imperfect data acquisition and preprocessing, it is usually unavoidable that vision data are contaminated by noise and outliers. Traditional parameter estimation methods mostly deal with noise and are vulnerable to outliers. To facilitate computer vision applications, robust parameter estimation techniques are thus necessary for effectively rejecting outliers and accurately estimating the model parameters.

7.1 Summary of contributions of this thesis

In this thesis we concentrated on single structure parameter estimation and made a direct contribution to two specific branches under that topic: geometric fitting and deformable registration.

7.1.1 Geometric fitting

It has been observed in numerous previous approaches to robust geometric fitting that retrieving a single all-inlier minimal subset is not sufficient to guarantee a satisfactory model estimate that fits the data well; the inliers therein should also have a large spatial extent. We investigated in Chapter 3 a theoretical basis behind this long-standing principle. The starting point was a result by Jacobi relating to minimal subset expansion for least squares regression, where the quality of a minimal subset estimate is proportional to the span of the associated data points. We then developed a similar minimal subset expansion for total least squares, and highlighted the equivalence between total least squares and the direct linear transformation, a commonly used geometric fitting technique in computer vision. Moreover, we analysed the influence of data

span on various fundamental matrix estimation algorithms, which showed a clear correlation between the span of data points used for estimation and the quality of model estimate.

So far, a number of sampling methods have been proposed in the literature to speed up the retrieval of all-inlier minimal subsets. However, none of them actively searches for all-inlier minimal subsets with large span which are important to yielding a good model estimate. We developed in Chapter 4 a novel sampling algorithm, which unlike previous approaches, consciously targets all-inlier minimal subsets as well as with large span. Thanks to this ability, the proposed method can also avoid sampling degenerate minimal subsets. Experimental results on synthetic and real datasets for robust fundamental matrix estimation have illustrated the significant performance of our method.

7.1.2 Deformable registration

Most of the previous approaches to outlier rejection in deformable registration employ fully deformable models to represent the correspondence manifold, and rely on iterative optimisation procedures to remove outliers. This prevalent use of fully deformable models stems from the assumption of the highly nonlinear correspondence manifold. In contrast, we made in Chapter 5 the interesting observation that, for many realistic physical deformations tested in the literature, the errors of the outliers are usually so extreme that they dwarf the nonlinear effects of the correspondence manifold. Therefore, relative to the outliers, the distribution of inliers often resembles a low-dimensional affine subspace.

Our observation motivates the view that standard robust geometric fitting techniques are applicable to model the approximately linear correspondence manifold. Furthermore, we developed in Chapters 5 and 6 two novel outlier rejection methods for deformable registration, which are based entirely on fitting simple linear models. Experimental results on synthetic and real datasets demonstrated that the proposed methods are considerably more efficient but at least as accurate as previous approaches. Most importantly, the substantial results of our methods on various kinds of deformable objects empirically validated our message that outlier detection in deformable registration can be accomplished using simple linear models rather than optimising fully deformable models.

7.2 Directions of future work

We presented in Chapter 5 a simple RANSAC-based approach to outlier rejection in deformable registration. Although this method achieves the state-of-the-art performance, there are some issues that may be explored in future studies. Firstly, the performance of the method relies

on having an optimal setting of the inlier threshold. Methods for scale estimation [98, 99] can be applied to automatically estimate the inlier threshold and alleviate the dependence on an optimal threshold. Secondly, random sampling (RANSAC) becomes inefficient in cases of highly contaminated data. To reduce the number of minimal subsets required to be drawn, guided sampling methods which either speed up the generation of all-inlier minimal subsets (Section 2.2) or maximise the span of all-inlier minimal subsets produced (Section 4.2) can be employed.

As we mentioned at the outset, this thesis focused on parameter estimation of a single structure. However, vision data can contain multiple structures of a model, e.g., multiple motions of rigid objects often give rise to multiple fundamental matrices. In multiple-structure data, the inliers of a particular structure behave as pseudo-outliers to the other structures and the number of structures is usually a priori unknown. Another avenue for future research is to extend both our work on geometric fitting and deformable registration to handle data with multiple structures. A natural solution is to apply our methods sequentially [34, 100], i.e., estimate one single structure first, remove its corresponding inliers and continue on the remaining data until no more structures can be found. A major drawback of this simple approach is that the errors in the estimation of earlier structures will be propagated and amplified in the estimation of subsequent structures. Methods that simultaneously estimate the number of structures and the parameters of each structure [101, 102] have been recently proposed in the literature. However, these methods do not consider the role of data span and are used for estimating multiple geometric structures only. Thus, it is desirable to develop new methods that can exploit the benefits of large span sampling in multiple-structure geometric estimation and investigate the application of these methods in multiple-structure deformable estimation.

Appendix A

Proof of Propositions 3.1 and 3.2

A.1 Proof of Proposition 3.2 for TLS with non-minimal subsets

From Sec. 3.2.1, the weight of a non-minimal subset ν is proportional to $|\mathbf{X}(\nu)^T\mathbf{X}(\nu)|$ which, from (3.36), is equal to

$$|\mathbf{X}(\nu)^T\mathbf{X}(\nu)| = |\mathbf{V}\mathbf{S}_m^T\mathbf{U}_m(\nu)^T\mathbf{U}_m(\nu)\mathbf{S}_m\mathbf{V}^T| \tag{A.1}$$

$$= |\mathbf{U}_m(\nu)^T\mathbf{U}_m(\nu)||\mathbf{S}_m\mathbf{V}^T|^2. \tag{A.2}$$

Similarly,

$$|\mathbf{Z}(\nu)^T\mathbf{Z}(\nu)| = |\mathbf{U}_m(\nu)^T\mathbf{U}_m(\nu)||\tilde{\mathbf{S}}_m\mathbf{V}^T|^2. \tag{A.3}$$

Therefore, $|\mathbf{X}(\nu)^T\mathbf{X}(\nu)| = \kappa|\mathbf{Z}(\nu)^T\mathbf{Z}(\nu)|$ where κ is a constant which does not depend on ν. This proves that

$$|\mathbf{X}(\nu_1)^T\mathbf{X}(\nu_1)| > |\mathbf{X}(\nu_2)^T\mathbf{X}(\nu_2)| \tag{A.4}$$

$$\implies |\mathbf{Z}(\nu_1)^T\mathbf{Z}(\nu_1)| > |\mathbf{Z}(\nu_2)^T\mathbf{Z}(\nu_2)|. \tag{A.5}$$

A.2 Proof of Propositions 3.1 and 3.2 for mixed OLS-TLS

We aim to prove that Proposition 3.1 holds for the mixed OLS-TLS problem (Sec.3.3.3), i.e., the solution to the OLS problem

$$\arg\min_{\boldsymbol{\beta}} \|\mathbf{y} - \hat{\mathbf{y}}\|^2 \quad \text{s.t.} \quad \mathbf{Z}\boldsymbol{\beta} = \hat{\mathbf{y}} \tag{A.6}$$

coincides with the mixed OLS-TLS estimate

$$\breve{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X} - \sigma^2_{m_2+1}\mathbf{L})^{-1}\mathbf{X}^T\mathbf{y} = (\mathbf{X}^T\mathbf{Z})^{-1}\mathbf{X}^T\mathbf{y} \tag{A.7}$$

where

$$\mathbf{Z} := \mathbf{X} - \sigma^2_{m_2+1}(\mathbf{X}^T)^\dagger\mathbf{L}. \tag{A.8}$$

See Sec. 3.3.3 for the definition of the other symbols involved.

Let $\tilde{\boldsymbol{\beta}}$ be the solution to (A.6). Then

$$\tilde{\boldsymbol{\beta}} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y} \tag{A.9}$$

which, following the proof of Proposition 3.1, can be rearranged to become

$$\mathbf{X}^T\mathbf{Z}\tilde{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{y} + \sigma^2_{m+1}\mathbf{L}^T(\mathbf{X}^T)^{\dagger T}(\mathbf{Z}\tilde{\boldsymbol{\beta}} - \mathbf{y}). \tag{A.10}$$

As shown in the proof of Proposition 3.1, the column spans of $\mathbf{Z}$ and $(\mathbf{X}^T)^\dagger$ are equal. Since vector $(\mathbf{Z}\tilde{\boldsymbol{\beta}} - \mathbf{y})$ is orthogonal to $\mathcal{R}(\mathbf{Z})$, it is also orthogonal to $\mathcal{R}((\mathbf{X}^T)^\dagger)$, thus the second component on the RHS of (A.10) equates to zero, yielding

$$\mathbf{X}^T\mathbf{Z}\tilde{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{y}. \tag{A.11}$$

Comparing (A.11) to (A.7) proves $\tilde{\boldsymbol{\beta}} = \breve{\boldsymbol{\beta}}$, i.e., the mixed OLS-TLS estimate $\breve{\boldsymbol{\beta}}$ coincides with the solution of the OLS (A.6).

To prove that Proposition 3.2 also holds for mixed OLS-TLS, it is sufficient to show that, given a non-minimal data subset ν of size $m + i \leq n$ ($i \geq 0$),

$$|\mathbf{X}(\nu)^T\mathbf{X}(\nu)| \propto |\mathbf{Z}(\nu)^T\mathbf{Z}(\nu)|. \tag{A.12}$$

To begin, let $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ be the SVD of $\mathbf{X}$. Then $(\mathbf{X}^T)^\dagger = \mathbf{U}\mathbf{S}^{-1}\mathbf{V}^T$ is the SVD of $(\mathbf{X}^T)^\dagger$. Also, since $n > m$

$$\mathbf{X} = \mathbf{U}_m\mathbf{S}_m\mathbf{V}^T \quad (\mathbf{X}^T)^\dagger = \mathbf{U}_m\mathbf{S}_m^{-1}\mathbf{V}^T. \tag{A.13}$$

Then, from (A.8)

$$\mathbf{Z} = \mathbf{U}_m(\mathbf{S}_m\mathbf{V}^T - \sigma^2_{m_2+1}\mathbf{S}_m^{-1}\mathbf{V}^T\mathbf{L}) = \mathbf{U}_m\boldsymbol{\Gamma} \tag{A.14}$$

where we define the square matrix

$$\mathbf{\Gamma} := (\mathbf{S}_m \mathbf{V}^T - \sigma^2_{m_2+1} \mathbf{S}_m^{-1} \mathbf{V}^T \mathbf{L}). \tag{A.15}$$

Also, observe that

$$\mathbf{Z}(\nu) = \mathbf{U}_m(\nu)\mathbf{\Gamma}. \tag{A.16}$$

In (A.2), the following determinant has been established

$$|\mathbf{X}(\nu)^T \mathbf{X}(\nu)| = |\mathbf{U}_m(\nu)^T \mathbf{U}_m(\nu)||\mathbf{S}_m \mathbf{V}^T|^2. \tag{A.17}$$

The determinant $|\mathbf{Z}(\nu)^T \mathbf{Z}(\nu)|$ is then

$$|\mathbf{Z}(\nu)^T \mathbf{Z}(\nu)| = |\mathbf{U}_m(\nu)^T \mathbf{U}_m(\nu)||\mathbf{\Gamma}|^2 \tag{A.18}$$

which implies $|\mathbf{X}(\nu)^T \mathbf{X}(\nu)| \propto |\mathbf{Z}(\nu)^T \mathbf{Z}(\nu)|$. Note that this result also holds for minimal subsets by setting $i = 0$.

Bibliography

[1] W. Uricchio. The algorithmic turn: Photosynth, augmented reality and the changing implications of the image. In G. Bolin, editor, *Cultural Technologies: The Shaping of Culture in Media and Society*. Routledge, 2012.

[2] A. Van Den Hengel, A. Dick, T. Thormählen, B. Ward, and P. H. S. Torr. Videotrace: Rapid interactive scene modelling from video. In *ACM SIGGRAPH*, 2007.

[3] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[4] P. H. S. Torr and A. Zisserman. Feature based methods for structure and motion estimation. In *International Conference on Computer Vision (ICCV) Workshops*, 1999.

[5] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch. Visual modeling with a hand-held camera. *International Journal of Computer Vision (IJCV)*, 2004.

[6] R. Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends in Computer Graphics and Vision*, 2006.

[7] M. Brown and D. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision (IJCV)*, 2007.

[8] L. Zagorchev and A. Goshtasby. A comparative study of transformation functions for nonrigid image registration. *IEEE Transactions on Image Processing (TIP)*, 2006.

[9] A. Goshtasby. *Image Registration: Principles, Tools and Methods*. Advances in Computer Vision and Pattern Recognition. Springer, 2012.

[10] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2002.

[11] G. Mori, S. Belongie, and J. Malik. Efficient shape matching using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2005.

[12] T. Okabe and Y. Sato. Object recognition based on photometric alignment using RANSAC. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003.

[13] C. Papazov and D. Burschka. An efficient RANSAC for 3D object recognition in noisy and occluded scenes. In *Asian Conference on Computer Vision (ACCV)*, 2010.

[14] E. Brau, K. Barnard, R. Palanivelu, D. Dunatunga, T. Tsukamoto, and P. Lee. A generative statistical model for tracking multiple smooth trajectories. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[15] A. Andriyenko and K. Schindler. Multi-target tracking by continuous energy minimization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[16] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 2004.

[17] P. Meer. Robust techniques for computer vision. In G. Medioni and S. B. Kang, editors, *Emerging Topics in Computer Vision*. Prentice Hall, 2004.

[18] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 1981.

[19] P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*. Wiley Series in Probability and Statistics. Wiley, 1987.

[20] O. Chum, J. Matas, and J. Kittler. Locally optimized RANSAC. In *Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM)*, 2003.

[21] O. Chum, J. Matas, and S. Obdrzakek. Enhancing RANSAC by generalized model optimization. In *Asian Conference on Computer Vision (ACCV)*, 2004.

[22] B. J. Tordoff and D. W. Murray. Guided-MLESAC: Faster image transform estimation by using matching priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2005.

[23] N. Scherer-Negenborn and R. Schaefer. Model fitting with sufficient random sample coverage. *International Journal of Computer Vision (IJCV)*, 2010.

[24] F. L. Bookstein. Principal warps: Thin plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 1989.

[25] G. Carneiro and A. D. Jepson. Flexible spatial configuration of local image features. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2007.

[26] A. Bartoli. Maximizing the predictivity of smooth deformable image warps through cross-validation. *Journal of Mathematical Imaging and Vision (JMIV)*, 2008.

[27] X. Li and Z. Hu. Rejecting mismatches by correspondence function. *International Journal of Computer Vision (IJCV)*, 2010.

[28] P. J. Huber. *Robust Statistics*. Wiley Series in Probability and Statistics. Wiley, 1981.

[29] J. Pilet, V. Lepetit, and P. Fua. Real-time non-rigid surface detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[30] J. Zhu and M. R. Lyu. Progressive finite Newton approach to real-time nonrigid surface detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[31] G. H. Golub and C. F. Van Loan. An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis*, 1980.

[32] M. Mühlich and R. Mester. The role of total least squares in motion analysis. In *European Conference on Computer Vision (ECCV)*, 1998.

[33] D. R. Myatt, P. H. S. Torr, S. J. Nasuto, J. M. Bishop, and R. Craddock. NAPSAC: High noise, high dimensional robust estimation - It's in the bag. In *British Machine Vision Conference (BMVC)*, 2002.

[34] Y. Kanazawa and H. Kawakami. Detection of planar regions with uncalibrated stereo using distributions of feature points. In *British Machine Vision Conference (BMVC)*, 2004.

[35] O. Chum and J. Matas. Matching with PROSAC- Progressive sample consensus. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[36] T.-J. Chin, J. Yu, and D. Suter. Accelerated hypothesis generation for multi-structure robust fitting. In *European Conference on Computer Vision (ECCV)*, 2010.

[37] T.-J. Chin, J. Yu, and D. Suter. Accelerated hypothesis generation for multi-structure data via preference analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2012.

[38] J. Pilet, V. Lepetit, and P. Fua. Fast non-rigid surface detection, registration and realistic augmentation. *International Journal of Computer Vision (IJCV)*, 2008.

[39] J. Zhu, C. H. Hoi, and M. R. Lyu. Nonrigid shape recovery by Gaussian process regression. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[40] X. Li, X. Li, H. Li, and M. Cao. Rejecting outliers based on correspondence manifold. *Acta Automatica Sinica*, 2009.

[41] D. Pizarro and A. Bartoli. Feature-based deformable surface detection with self-occlusion reasoning. In *International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, 2010.

[42] D. Pizarro and A. Bartoli. Feature-based deformable surface detection with self-occlusion reasoning. *International Journal of Computer Vision (IJCV)*, 2012.

[43] P. J. Rousseeuw. Least median of squares regression. *Journal of the American Statistical Association*, 1984.

[44] S. M. Stigler. *The History of Statistics: The Measurement of Uncertainty before 1900*. Belknap Press, 2000.

[45] Z. Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. *Journal of Image and Vision Computing (IVC)*, 1997.

[46] P. Torr and A. Zisserman. Robust computation and parametrization of multiple view relations. In *International Conference on Computer Vision (ICCV)*, 1998.

[47] P. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Journal of Computer Vision and Image Understanding (CVIU)*, 2000.

[48] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2004.

[49] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision (IJCV)*, 2005.

[50] H. S. Wong, T.-J. Chin, J. Yu, and D. Suter. Dynamic and hierarchical nulti-structure geometric model fitting. In *International Conference on Computer Vision (ICCV)*, 2011.

[51] P. Mcllroy, E. Rosten, S. Taylor, and T. Drummond. Deterministic sample consensus with multiple match hypotheses. In *British Machine Vision Conference (BMVC)*, 2010.

[52] V. Fragoso and M. Turk. SWIGS: A swift guided sampling method. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[53] T. Sattler, B. Leibe, and L. Kobbelt. SCRAMSAC: Improving RANSAC's efficiency with a spatial consistency filter. In *International Conference on Computer Vision (ICCV)*, 2009.

[54] K. Ni, H. Jin, and F. Dellaert. GroupSAC: Efficient consensus in the presence of groupings. In *International Conference on Computer Vision (ICCV)*, 2009.

[55] L. Goshen and I. Shimshoni. Balanced exploration and exploitation model search for efficient epipolar geometry estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2008.

[56] A. S. Brahmachari and S. Sarkar. BLOGS: Balanced local and global search for non-degenerate two view epipolar geometry. In *International Conference on Computer Vision (ICCV)*, 2009.

[57] K. Lebeda, J. Matas, and O. Chum. Fixing the locally optimized RANSAC. In *British Machine Vision Conference (BMVC)*, 2012.

[58] B. Triggs. Joint feature distributions for image correspondence. In *International Conference on Computer Vision (ICCV)*, 2001.

[59] A. Meler, M. Decrouez, and J. L. Crowley. BetaSAC: A new conditional sampling for RANSAC. In *British Machine Vision Conference (BMVC)*, 2010.

[60] H. S. Wong, T.-J. Chin, J. Yu, and D. Suter. Efficient multi-structure robust fitting with incremental top-k lists comparison. In *Asian Conference on Computer Vision (ACCV)*, 2010.

[61] T. T. Pham, T.-J. Chin, J. Yu, and D. Suter. The random cluster model for robust geometric fitting. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[62] D. J. C. Mackay. *Information Theory, Inference, and Learning Algorithms*, chapter The Swendsen-Wang method. Cambridge University Press, 2003.

[63] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 1999.

[64] M. Salzmann and P. Fua. Reconstructing sharply folding surfaces: A convex formulation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[65] Q.-H. Tran, T.-J. Chin, W. Chojnacki, and D. Suter. Sampling minimal subsets with large spans for robust estimation. *International Journal of Computer Vision*, 2013.

[66] C. R. Rao. *Linear Statistical Inference and Its Applications*. Wiley Series in Probability and Statistics. Wiley, 1965.

[67] C. G. J. Jacobi. De formatione et proprietatibus determinantium. *Journal für die Reine und Angewandte Mathematik*, 1841.

[68] M. Subrahmanyam. A property of simple least squares estimates. *Sankhya*, 1972.

[69] A. E. Hoerl and R. W. Kennard. A note on least squares estimates. *Communications in Statistics: Simulation and Computation*, 1980.

[70] S. Van Huffel and J. Wandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis*. Frontiers in Applied Mathematics. SIAM, 1991.

[71] S. Van Huffel and J. Wandewalle. Algebraic connections between the least squares and total least squares problem. *Numerische Mathematik*, 1989.

[72] G. H. Golub, A. Hoffman, and G. W. Stewart. A generalization of the Eckart-Young-Mirksy matrix approximation theorem. *Linear Algebra and Its Applications*, 1987.

[73] P. De Groen. An introduction to total least squares. *Nieuw Archief voor Wiskunde*, 1996.

[74] A. Kukush, I. Markovsky, and S. Van Huffel. Consistent fundamental matrix estimation in a quadratic measurement error model arising in motion analysis. *Computational Statistics and Data Analysis*, 2002.

[75] R. Hartley. In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 1997.

[76] F. Kahl and D. Henrion. Globally optimal estimates for geometric reconstruction problems. In *International Conference on Computer Vision (ICCV)*, 2005.

[77] F. Kahl and R. Hartley. Multiple-view geometry under the l_∞-norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2008.

[78] C. Olsson, A. Eriksson, and R. Hartley. Outlier removal using duality. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[79] O. Chum and J. Matas. Planar affine rectification from change of scale. In *Asian Conference on Computer Vision (ACCV)*, 2010.

[80] M. Harker and P. O'Leary. Direct estimation of homogeneous vectors: An ill-solved problem in computer vision. In *Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)*, 2006.

[81] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/, 2008.

[82] O. Chum, T. Werner, and J. Matas. Two-view geometry estimation unaffected by a dominant plane. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[83] J.-M. Frahm and M. Pollefeys. RANSAC for (quasi-)degenerate data (QDEGSAC). In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006.

[84] D. E. Critchlow. *Metric Methods for Analyzing Partially Ranked Data*. Lecture Notes in Statistics. Springer, 1985.

[85] J. I. Marden. *Analyzing and Modeling Rank Data*. Monographs on Statistics and Applied Probability. Chapman and Hall, 1996.

[86] Q.-H. Tran, T.-J. Chin, G. Carneiro, M. S. Brown, and D. Suter. In defence of RANSAC for outlier rejection in deformable registration. In *European Conference on Computer Vision*, 2012.

[87] G. Donato and S. Belongie. Approximate thin plate spline mappings. In *European Conference on Computer Vision (ECCV)*, 2001.

[88] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2000.

[89] D. Comaniciu, V. Ramesh, and P. Meer. Mean shift: A robust approach towards feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2002.

[90] D. Levin. The approximation power of moving least squares. *Mathematics of Computation*, 1998.

[91] M. Alexa, J. Behr, D. Cohen-Or, S. Fleishman, D. Levin, and C. T. Silva. Computing and rendering point set surfaces. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 2003.

[92] S. Schaefer, T. McPhail, and J. Warren. Image deformation using moving least squares. In *ACM SIGGRAPH*, 2006.

[93] Q.-H. Tran, T.-J. Chin, J. Zaragoza, G. Carneiro, and D. Suter. Outlier rejection in deformable registration with moving least squares. *Submitted to Journal of Computer Vision and Image Understanding*, under review.

[94] T. W. H. Tang and A. C. S. Chung. Non-rigid image registration using graph-cuts. In *Medical Image Computing and Computer Assisted Intervention Conference (MICCAI)*, 2007.

[95] R. W. K. So and A. C. S. Chung. Multi-level non-rigid image registration using graph-cuts. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2009.

[96] M. J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. In *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 1998.

[97] W. Choi, K. Shahid, and S. Savarese. What are they doing? : Collective activity classification using spatio-temporal relationship among people. In *International Conference on Computer Vision (ICCV) Workshops*, 2009.

[98] H. Chen and P. Meer. Robust regression with projection based M-estimators. In *International Conference on Computer Vision (ICCV)*, 2003.

[99] S. Rozenfeld and I. Shimshoni. The modified pbM-estimator method and a runtime analysis technique for the RANSAC family. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[100] E. Vincent and R. Laganiere. Detecting planar homographies in an image pair. In *International Symposium on Image and Signal Processing and Analysis*, 2001.

[101] R. Toldo and A. Fusiello. Robust multiple structures estimation with J-Linkage. In *European Conference on Computer Vision (ECCV)*, 2008.

[102] T.-J. Chin, H. Wang, and D. Suter. Robust fitting of multiple structures: The statistical learning approach. In *International Conference on Computer Vision (ICCV)*, 2009.